INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Data assimilation of OMI NO_2 observations for improving air quality forecast over Europe

Xiaoni Wang — Vivien Mallet — Jean-Paul Berroir — Isabelle Herlin

N° 6884

March 2008

Thème NUM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--6884--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
PARIS - ROCQUENCOURT

Data assimilation of OMI NO_2 observations for improving air quality forecast over Europe

Xiaoni Wang , Vivien Mallet , Jean-Paul Berroir , Isabelle Herlin

Thème NUM — Systèmes numériques
Équipes-Projets Clime

Rapport de recherche n° 6884 — March 2008 — 17 pages

Abstract: This paper concerns the improvements of NO_2 forecast due to satellite data assimilation. The Ozone Monitoring Instrument (OMI) aboard NASA Aura satellite provides observations of NO_2 columns for air quality study. These satellite observations are assimilated, with the optimal-interpolation method, in an air quality model from Polyphemus, in order to improve NO_2 forecasts in Europe. Good consistency is seen in the comparisons of model simulations, satellite data and ground observations before assimilation. The model results with and without assimilation are then compared with ground observations for evaluating the assimilation effects. It is found that in winter the errors between model data and ground observations have been reduced after assimilation, indicating a better NO_2 forecast can be obtained using satellite observations. Such improvements are not found in summer, which is probably due to the shorter life time and higher temporal variability of NO_2 in the warmer season.

Key-words: air quality, data assimilation, numerical simulation, satellite observations, troposphere

Centre de recherche INRIA Paris – Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

Assimilation d'observations de NO_2 acquises par OMI dans un modèle de prévision de la qualité de l'air

Résumé : Cette étude présente des résultats d'assimilation de données satellite pour améliorer la prévision de la concentration de NO_2 par un modèle de qualité de l'air. L'instrument OMI (*Ozone Monitor Instrument*, à bord du satellite Aura, NASA) acquiert des colonnes troposphériques de NO_2 compatibles avec les modèles régionaux de prévision de la qualité de l'air. Ces observations sont assimilées par Interpolation Optimale dans la plate-forme de prévision Polyphemus. Les résultats de prévision avec et sans assimilation sont évalués par comparaison avec les données au sol. Les erreurs de prévision en hiver ont été signifactivement réduite par assimilation de données satellite. Ce résultat n'est pas obtenu en été, probablement en raison de la durée de vie plus courte et de la plus grande variabilité du NO_2 en été.

Mots-clés : qualité de l'air, assimilation de données, simulation numérique, données satellite

1 Introduction

Nitrogen dioxide (NO_2) plays an important role in tropospheric chemistry (Logan, 1983). Anthropogenic NOx emissions are key precursors for tropospheric ozone (O_3) (Murphy et al., 1993) and can affect the formation of aerosol nitrate (NH_4NO_3) in the presence of ammonia (NH_3)(Logan, 1983). NO_2 also has a direct impact on the public health. Long-term NO_2 exposure may affect lung function and increase the risk of respiratory symptoms (Panella et al., 2000; Smith et al., 2000). Strong associations exist between NO_2 and nonaccidental mortality in daily time series studies (Steib et al., 2003; Burnett et al., 2004). Therefore, a better knowledge of NO_2 concentration is important to many issues related to air quality. In such context this study aims to improve the NO_2 forecast using assimilation of new observations.

Data assimilation is effective in the optimal estimation and forecast of atmosphere (Kalnay, 2003). By using observations of parts of the state it can estimate the less known model parameters or the state itself. Data assimilation techniques have been applied in the numerical weather prediction for a long time (Daley, 1993; Kalnay, 2003) and recently been introduced to air quality study (Austin, 1992; Fisher and Lary, 1995; Riishojgaard, 1996). Pollutant measurements of diverse nature (ground, airborne *etc.*) have been assimilated for different species (Jeuken et al., 1999; Levelt et al., 1999). For nitrogen dioxide satellite data were used to improve the estimation of NOx emissions (Konovalov et al., 2006; Napelenok et al., 2008). Few studies have been made to directly adjust the NO_2 concentrations with data assimilation techniques, due to the difficult modeling of NO_2.

Satellite observations of tropospheric NO_2 columns began in 1995 with the Global Ozone Monitoring Experiment (GOME) spectrometer aboard the European Research Satellite (ERS-2) (Burrows et al., 1999), and was followed by the Scanning Imaging Absorption spectroMeter for Atmospheric ChartograpHY (SCIAMACHY) (Bovensmann et al., 1999). Studies on GOME and SCIAMACHY observations have demonstrated the capacity of observing air pollution from space (Leue et al., 2001; Richter et al., 2005). More recently, the Ozone Monitoring Instrument (OMI) aboard the Earth Observing System (EOS) Aura satellite was launched in 2004 (Levelt et al., 2006). OMI provides high quality measurements of NO_2 columns with better spatial resolution and in daily global coverage, which are important advantages compared to earlier instruments of GOME or SCIAMACHY. The OMI data are then expected to improve the day-to-day monitor and forecast of air quality, and are used in this study.

Polyphemus is a modeling system for forecast and assessment of air quality (Mallet et al., 2007). It can deal with applications on photochemistry and aerosol dynamics at continental scales using Eulerian models. Polyphemus has been involved in operational testing for photochemistry forecasts on the French Prév'air platform from summer 2004 (`http://www.prevair.org/`). In this study the system is used with a similar configuration to forecast NO_2 concentrations in Europe.

In section 2 we introduce the Polyphemus system, its configuration in the present case and the comparisons between model simulations and ground observations. The NO_2 column observations from OMI and their comparisons with model results are then detailed in section 3. In section 4 the consistency

of model results, satellite observations and ground measurements is examined. The assimilation method is then evaluated and its results in a cold season are presented in section 5. In section 6 the seasonal differences of NO_2 assimilation results are presented with a summer study included. Conclusions can be found in section 7.

2 Polyphemus system and comparisons with ground observations

2.1 Polyphemus system and experiment configuration

Polyphemus system can deal with applications of passive tracers, radioactive decay, photochemistry and aerosol dynamics from local to continental scales (Quélo et al., 2007; Sartelet et al., 2007; Korsakissok and Mallet, 2008). It is made of four distinct components. i) Preprocessing of input fields (meteorological fields, emissions, ...) in the first stage, notably relying on physical parameterizations available in the AtmoData library. ii) Eulerian chemistry-transport models (Polair3D and Castor) and local-scale plume and puff Gaussian models. iii) Drivers for handling the models, *e.g.,* a driver for Monte Carlo simulations, drivers for data assimilation. iv) Postprocessing tools (*i.e.,* the Python module AtmoPy) for statistical analysis of the results. Detailed descriptions of the system and source code are available at `http://cerea.enpc.fr/polyphemus/`. The recent version, Polyphemus 1.3.1, is used for study.

The configuration of the model can be described as follows. The land use data are obtained with 1 km resolution from GLCF (Global Land Cover Facility) data base (`http://glcf.umiacs.umd.edu`). Meteorological input data are computed from ECMWF (European Centre for Medium-Range Weather Forecasts) fields (12-hour forecast cycles starting from analyses) with 60 vertical levels, a horizontal resolution of 0.36° by 0.36° and a time step of 3 hours. The initial conditions and daily boundary conditions are extracted from outputs of the global Chemistry-Transport Model Mozart 2. Anthropogenic emissions (*e.g.*, NOx, SOx, *etc.*) are provided by EMEP (Co-operative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe) inventory, converted according to Middleton et al. (1990). Biogenic emissions are parameterized with the method proposed in Simpson et al. (1999). Deposition velocities are from Zhang et al. (2003) for gaseous species.

The Chemistry-Transport Model (CTM) Polair3D (Boutahar et al., 2004) is in charge of the time integration of the reactive transport equation. The model domain covers the west and central Europe, from 10.5° W to 23° E and from 25° N to 58° N, with a resolution of 0.5° by 0.5° in horizontal directions. The maximum vertical height is 5000 meters with 10 levels from the ground. The simulation time step is 10 minutes. A cold season (November and December) of year 2005 is studied here.

2.2 Comparisons with ground observations

Ground observations are useful for validating the model simulations. They can be further used for quantifying the improvements due to satellite data assimilation (Sect. 5). In this study the ground observations are from the EMEP

network. It provides the hourly and daily measurements of NO_2 near surface in Europe. The important advantages of EMEP measurements are the common quality control standards applied and their station locations; the latter makes the measurements representative of regional background conditions, relatively unaffected by local emissions. Several other monitoring networks (urban, suburban, *etc.*) are not used here, since they would pose serious problems concerning criteria of data selection and treatment of uncertainties. An obvious disadvantage of the EMEP network is the rather sparse distribution of stations for NO_2 (totally 25 in 2005).

The model results are then compared with the hourly EMEP data and their daily statistics are computed, *i.e.,* mean ($\mu g\,m^{-3}$), correlation and root mean square error (RMSE) ($\mu g\,m^{-3}$). The statistics are computed station per station and then averaged. The mean RMSE is 6.9 $\mu g\,m^{-3}$ between model and ground observations; the mean correlation is 0.62, indicating a good consistency between them. Model results are higher than ground observations on average: 12.4 $\mu g\,m^{-3}$ versus 8.1 $\mu g\,m^{-3}$. This difference is now discussed. The ground NO_2 is measured by chemiluminescence monitors and the values can be systematically overestimated because of interferences of other compounds (Dunlea et al., 2007; Steinbacher et al., 2007). However, with model results being larger than ground observations, these measurement errors can not explain the discrepancy. Konovalov et al. (2006) have discussed the effects of emission uncertainties on the modeling of NO_2 concentrations and showed that they may contribute significantly to the differences between model and ground observations. The other possible explanation is that heterogeneous reactions, which may reduce the NO_2 concentration in winter (Sartlet et al., 2008; Roustan et al., 2008), are not included in the current simulation.

3 Satellite observations and its comparisons with model columns

The Ozone Monitoring Instrument (OMI) aboard NASA's EOS Aura satellite was launched on 15 July 2004. It traces a Sun-synchronous polar orbit at an altitude of approximately 705 km. It has a period of 100 minutes and an equator crossing time between 13:40 and 13:50 local time. OMI has been equipped with two-dimensional CCD detectors and provides a wide field of view for one orbit (*i.e.*, a 2600 km wide spatial swath on the Earth's surface), large enough to obtain a complete global coverage in one day. OMI measures NO_2 columns with pixels of 13 km (along track) by 24 km (cross track) at nadir and 13 by 28 km^2 for the outermost swath angles (Boersma et al., 2006). Therefore, it has a much finer spatial resolution than GOME (320 by 40 km^2) (Richter and Burrows, 2002) and a significant improvement in both spatial and temporal resolutions than SCIAMACHY (60 by 30 km^2 and six days to complete a global coverage). OMI data can therefore provide high-quality observations for air quality study (Bucsela et al., 2006; Boersma et al., 2006).

OMI NO_2 columns can be downloaded from the NASA website (`http://disc.sci.gsfc.nasa.gov/Aura/OMI/omno2g_v003.shtml`). The standard data product (version 3 of daily gridded data) is used in this study. It is the daily

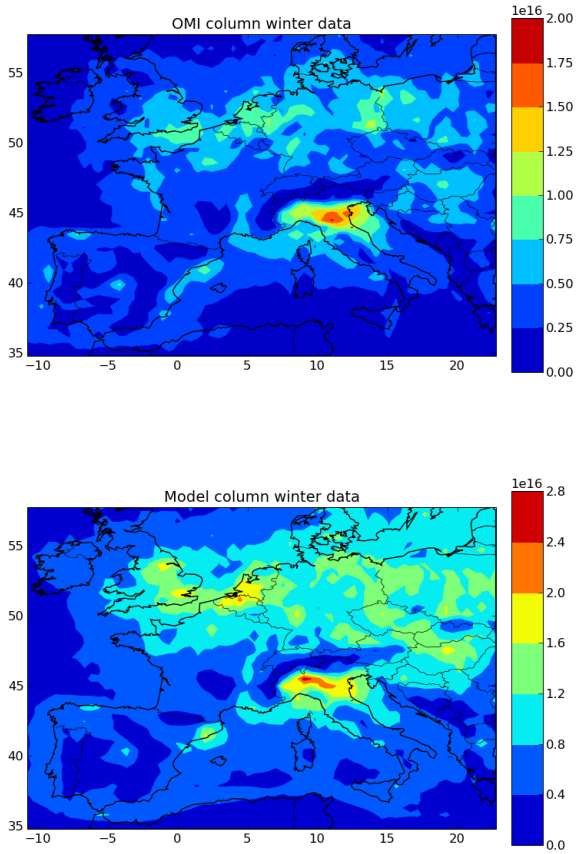

Figure 1: The temporal average of column data in Europe over November and December of 2005. Top: OMI column data. Bottom: model simulation of columns. Good consistency is seen in Spain, Italy, Great Britain, North Europe. Model simulations have higher values than satellite observations.

gridded product which bins the 24 UTC hours of data to the longitude-latitude grid.

Several procedures of data processing have been applied to prepare the data for assimilation. A general quality control is first used to select data with a quality flag which concerns solar zenith angles, the missing data, *etc.* The columns are then filtered out when the cloud fraction reaches 20% since clouds can obscure gases below and increase the measurement sensitivity above them. Finally, we consider to match the satellite data to model grids. The satellite observation resolution is 0.25° by 0.25° in longitude and latitude and the model is 0.5° by 0.5°. The available satellite data from the same orbit (4 or less observations) are collected in each model grid; their average is used as the observation for that grid of model.

These data are then collected for the same domain as the model simulation (see section 2) over November and December of 2005. They are used for the data assimilation in section 5.

OMI data errors come from instruments and retrieval algorithms. The details of retrieval algorithms are described in Bucsela et al. (2006). The errors can be represented by the standard deviations available on the OMI data website. These values are collected with the same processing procedures and for the same domain as the column data. The mean of the data errors is 1.68 10^{15}molecules cm^{-2} (or 1283 $\mu g\, m^{-2}$), which is 50% of the mean column value. This error is consistent with that in previous studies (Bucsela et al., 2006).

Comparisons between satellite data and model simulated columns are useful for understanding their general consistency and differences. Since the model simulation usually saves the hourly results, we collect the satellite and model columns at the same hours and locations, and these collected data are then compared. Their spatial-temporal correlation has a value of 0.57, suggesting a reasonable consistency. By averaging the available data (from model or satellite) within each grid cell over the studied period, we get the average spatial distribution of NO_2 columns shown in Figure 1 (a and b). Good consistency is seen between model and satellite data, with strong agreement in England, north Europe, Italy and Spain. Satellite observations generally have lower values than model columns in the studied period, which is consistent with previous reports that OMI may underestimates tropospheric columns in winter (Kramer et al., 2008; Lamasal et al., 2008).

4 Consistency of model, satellite and ground data

The consistency between satellite and ground data is first briefly discussed, with data being collected at the same hours and locations. Their spatial-temporal correlation is computed, with a value of 0.52. Since satellite columns and ground observations do not measure the same quantity, this indicates a reasonable consistency between them.

Data from model, satellite and ground stations (with the same locations and hours) are now compared together to understand their consistency, *i.e.,* how the satellite and model data can help retrieve ground observations. Let $\mathbf{x_{gd}}$ represent the model ground concentrations, $\mathbf{y_{cl}}$ the satellite observations and $\mathbf{y_{gd}}$ ground observations. $\mathbf{x_{cl}}$ (the column) is an integration of model concentrations from all the levels. If the vertical profile has the right shape (i.e., the ratio between the concentrations at two levels is correct), and if there are no errors in $\mathbf{y_{cl}}$ or $\mathbf{y_{gd}}$, then $\mathbf{y_{gd}} = \mathbf{y_{cl}}\ \mathbf{x_{gd}}/\mathbf{x_{cl}}$. This suggests that ground observations may be retrieved from satellite columns and model simulations. However, given the errors in both model and observations, such relationship does not exist in the real study.

One may try to find a more efficient retrieval of ground observations with the model and satellite data. This is realized by adjusting the ratio $\mathbf{x_{gd}}/\mathbf{x_{cl}}$ (correction of the model vertical profile), and the improved ground concentration

Table 1: Consistency experiment using model simulation, EMEP ground observations and OMI data. Results with parameter A of 1, 1.4, 2 and 5 listed.

statistics	Reference	$A = 1$	$A = 1.4$	$A = 2$	$A = 5$
Model mean	7.66	6.22	6.48	6.72	7.19
Ground mean	6.93	6.93	6.93	6.93	6.93
Correlation	0.7	0.7	0.71	0.71	0.71
RMSE	5.8	5.1	5.09	5.14	5.38

$\hat{\mathbf{x}}_{\mathbf{gd}}$ can be defined as:

$$\hat{\mathbf{x}}_{\mathbf{gd}} = \mathbf{x}_{\mathbf{gd}} \left(1 + \frac{\mathbf{y}_{\mathbf{cl}} - \mathbf{x}_{\mathbf{cl}}}{(1+A)\mathbf{x}_{\mathbf{cl}}} \right) \tag{1}$$

where A is a scaler. $\mathbf{x}_{\mathbf{gd}}$ and $\hat{\mathbf{x}}_{\mathbf{gd}}$ can then be compared with $\mathbf{y}_{\mathbf{gd}}$; their statistics, i.e., correlation and RMSE, can represent the consistency among the model, satellite and ground data. In this experiment the A values range from 1 to 5 and the statistical results are listed in Table 1. It is shown that the RMSE is reduced when comparing $\hat{\mathbf{x}}_{\mathbf{gd}}$ and $\mathbf{y}_{\mathbf{gd}}$ than that of $\mathbf{x}_{\mathbf{gd}}$ and $\mathbf{y}_{\mathbf{gd}}$. The minimum value of RMSE is found when A is equal to 1.4. Their correlation however remains similar as A varies. Therefore, A with a value of 1.4 can be used to provide the best direct retrieval of the ground observations from satellite and model data.

5 Data assimilation

Now the assimilation of satellite data is applied with the CTM model. The optimal interpolation method (OI) is used (Daley, 1993), since its performance should be similar to that of more complex methods (Wu et al., 2008). A brief description of OI method is presented below.

Let $\mathbf{x_b}$ be the a priori state estimation (*background*) with error covariance matrix $\mathbf{B}$, $\mathbf{y}$ be the observation with error covariance matrix $\mathbf{R}$, and $\mathbf{x_a}$ be the posterior state estimation (*analysis*). The innovation $\mathbf{d}$ represents the difference between the observation and the background state vector, i.e., $\mathbf{d} = \mathbf{y} - \mathbf{H}\mathbf{x_b}$, where $\mathbf{H}$ is a linear observation operator that maps the state vector to the observation space. NO_2 column is the state in the study. Our objective is to find the analysis state vector $\mathbf{x_a}$ according to best linear unbiased estimator theory as follows:

$$\mathbf{x_a} = \mathbf{x_b} + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x_b}) \tag{2}$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{3}$$

In practice, choosing the error covariance matrices for observations ($\mathbf{R}$) and background ($\mathbf{B}$) remains problematic. In this study $\mathbf{R}$ is set to be diagonal, with the value of 1.6 $10^6 \mu g^2\, m^{-4}$, which is the square of mean satellite data errors (section 3). $\mathbf{B}$ can be in the diagonal or Balgovind forms. If $\mathbf{B}$ is diagonal, the combination of formulae (2) and (3) leads to a similar expression as formulae

(1) in section 4. The ratio between errors of observation and background is then similar to the value of parameter A.

In the following study **B** is set to be Balgovind form and the error covariance between two points is given by

$$f(r) = \left(1 + \frac{r}{L}\right) \exp\left(-\frac{r}{L}\right) v \tag{4}$$

where L is a characteristic length, r is the distance between two points and v is a variance (Balgovind et al., 1983). The value of v is set to 1.14 $10^6 \mu g^2\, m^{-4}$, which is derived from the satellite data error divided by 1.4, the optimal value of A in section 4.

In order to use satellite data for assimilation, an observation operator **H** has been developed to link the CTM model and satellite data. The model makes forward simulations every Δt (10 minutes in this study). At a simulation time t, the model forecasts NO_2 and a vertical integration of the forecasted values produces the background state vector $\mathbf{x_b}$. $\mathbf{x_b}$ is then mapped to the same locations as satellite data through operator **H**. Using the OI formulae (2) and (3), $\mathbf{x_b}$ is updated to analysis state $\mathbf{x_a}$. The new NO_2 concentrations (i.e., analyzed concentrations) of all levels are then computed by multiplying the background concentrations with $\mathbf{x_a}/\mathbf{x_b}$. These analyzed concentrations can be used as the initial conditions for the next-step simulation ($t + \Delta t$).

Two simulations are normally carried out: one without data assimilation (i.e., reference, the same as that discussed in section 2), and the other with assimilation of observations. For the latter, the process can consist of two steps: assimilation and prediction. During the assimilation period, e.g., [$\mathbf{t_0}$,$\mathbf{t_N}$], observations are assimilated when they are available and the period usually covers more than one day. The outputs are model forecasts at various time intervals, *i.e.,* 10 minutes to several hours, since each day the available satellite data in Europe are within a few hours (usually from 10:00 UT to 14:00 UT). In the subsequent prediction period [$\mathbf{t_{N+1}}$,$\mathbf{t_T}$], simulation is made without using observations. The results from all these simulations (reference, assimilation and prediction) can be compared with ground measurements to understand the effects of assimilation.

5.1 Assimilation process

In the following the assimilation process is presented which covers the whole studied period. Data assimilation with different values of Balgovind characteristic length L is carried out since L describes the spatial structure of the error and affects the results.

The model results (with and without assimilation) are compared with ground observations and the comparison statistics are shown in Table 2. It is found that the assimilation is sensitive to L value. When L increases from 0.1° to 4° (recall that the model grid interval is 0.5°), the RMSE against ground observations is reduced significantly from 6.9 to 5.9, and the correlation changes slightly from 0.61 to 0.58. In order to choose a proper value of L, some practical conditions, e.g., meteorology and emissions, need be considered. The lifetime of NOx emissions is about 13 hours in Europe in winter (Schaub et al., 2007); the mean surface wind speed is 5.0 $m\,s^{-1}$ during this season (computed from ECMWF). Therefore, the NOx emissions can affect locations about 234 km

Table 2: Comparisons of model reference (without assimilation) and assimilation results with EMEP ground observations. The characteristics length L_h of Balgovind background covariance matrix varies from 0.1 to 4.

statistics	Reference	$L_h = 0.1$	$L_h = 0.25$	$L_h = 0.5$	$L_h = 1.5$	$L_h = 2$	$L_h = 3$	$L_h = 4$
Model mean	12.4	12	11.7	11.3	10.9	10.9	10.8	10.8
Ground mean	8.1	8.1	8.1	8.1	8.1	8.1	8.1	8.1
Correlation	0.62	0.61	0.60	0.60	0.59	0.59	0.58	0.58
RMSE	6.9	6.6	6.4	6.2	6.0	6.0	5.9	5.9

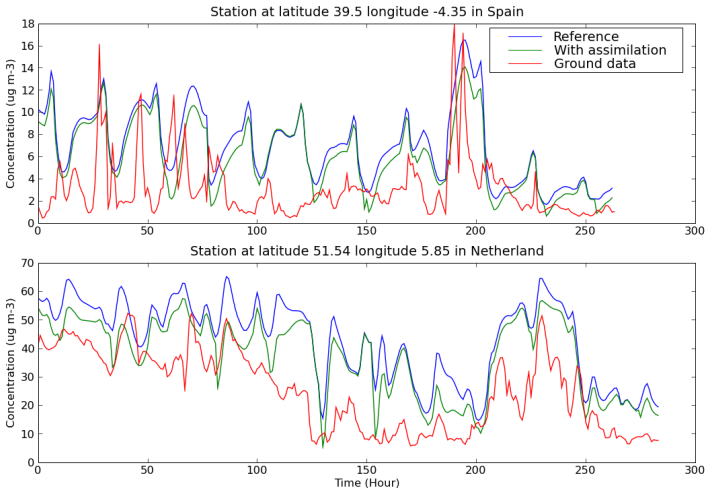

Figure 2: The model results(with and without assimilation) and ground observations at two example stations during 14–31 December 2005. Top: at a station in Spain. Bottom: at a station in Netherland. The model simulation with assimilation shows better performance compared to ground observations.

away from their original source. Since L is a characteristic decorrelation length in background errors from one location to the other, it should be related to the concentration transportation. Therefore, a rough estimation for L may be 3°. In the following study we use L equal to 1.5°.

When using $L = 1.5°$, the comparisons between model (with and without assimilation) and ground observations are shown at two example stations in Figure 2. Clear changes are seen in the ground concentrations with the assimilation applied. The mean RMSE between ground observations and the assimilation is 6.0 and the correlation is 0.59; the RMSE is therefore decreased by 13% (from 6.9 to 6.0) compared to that using the reference. The improvement of RMSE is similar, even slightly better than that (10%) from the inverse modeling of NOx emissions using GOME data (Konovalov et al., 2006). Therefore, data assim-

Table 3: Comparisons of model reference (without assimilation), the assimilation and predition processes with EMEP ground observations of NO_2.

statistics	Reference	Assimilation	Prediction
Model mean	12.4	10.9	11.6
Ground mean	8.1	8.1	8.1
Correlation	0.62	0.59	0.58
RMSE	6.9	6.0	6.5

ilation of OMI observations has a significant potential to improve the surface NO_2 forecast.

5.2 Operational conditions

It is highly desirable to obtain the next-day forecast of air quality for practical applications. For this purpose, the assimilation/prediction processes constitute a cycle which is applied to every two days within the studied period. For the first cycle, the assimilation process ranges from 1 November 2005 at 0100 UT to 2 November 2005 at 0000 UT, followed by the prediction from 2 November at 0100 UT to 3 November 0000 UT. The second cycle performs the assimilation from 2 November at 0100 UT to 3 November 0000 UT, followed by the prediction from 3 November at 0100 UT to 4 November 0000 UT. The cycling continues until the end of the considered period, *i.e.,* 31 December 2005. The predictions are then compared with ground observations. Their statistics are displayed in Table 3. While the correlation between the predictions and ground observations changes slightly, the RMSE is reduced from 6.9 (reference simulation) to 6.5 (prediction), hence by 7%. This indicates that a better next-day forecast of surface NO_2 can be achieved using assimilation of OMI data. The correlation between the predicted and satellite columns is also calculated, with a value of 0.6, similar to that (0.57) using the reference simulations.

6 Discussion

It is known that NO_2 variations differ from the warm to cold season (Kramer et al., 2008). We applied the same study to NO_2 in summer (June, July and August 2005) in order to understand the seasonal differences. The correlation between model and satellite data has a value of 0.51, slightly less than that (0.57) found in November and December. Shown in Figure 3 is the temporal mean over summer from model and satellite data respectively. Good consistency is seen in England, Spain, Italy, *etc.*, and less agreements in some parts of central Europe. Satellite data are on average higher than the model columns, which is the reverse from the cold season.

The assimilation process is also applied to the summer period using Balgovind error covariance matrix. It is found that the model NO_2 columns and ground concentrations are both increased with data assimilation. A comparison between OMI data and model columns shows a better correlation (0.59) with assimilation than without it (0.51), indicating an improved forecast of

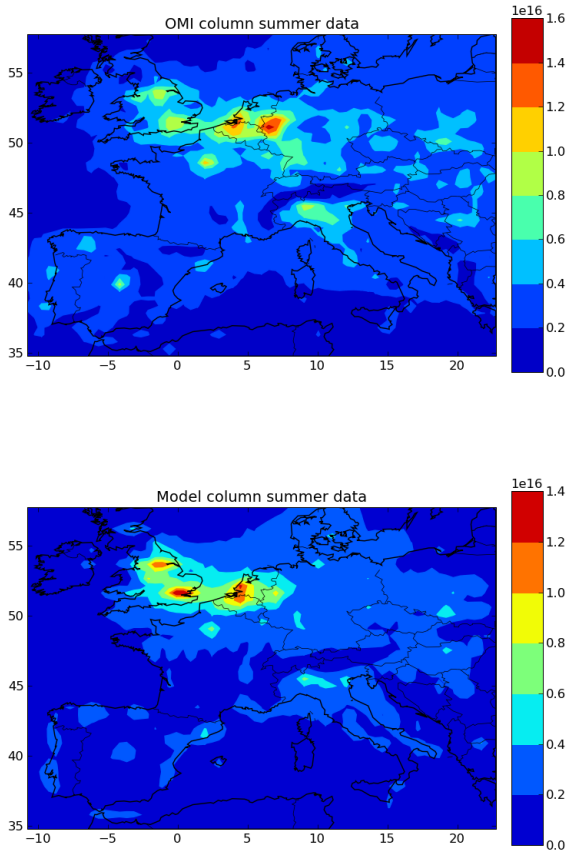

Figure 3: The temporal average of data in Europe over June, July and August of 2005. Top: OMI column data. Bottom: model simulation of columns. Good consistency is seen in Spain, Italy, Great Britain, North Europe, and the less in east Europe. Model simulations have lower values than satellite observations.

columns. The ground NO_2 results are then compared against EMEP observations; however, similar values are seen with or without assimilation. For better understanding the assimilation effects on the ground concentrations, we examine the NO_2 diurnal variations using the model. Shown in Figure 4 is an example of the spatial averages of NO_2 (for both columns and surface concentrations) with/without assimilation from 17 to 20 July. There are significant diurnal differences of NO_2 and the daily minimum is near noon time. When satellite observes NO_2 in Europe, it is usually between 1000 - 1400 UT, the same period when NO_2 concentrations are low. With the large diurnal variations and shorter life time of NOx in summer (Schaub et al., 2007), the assimilation effects decrease quickly with time and do not change the daily NO_2 profile as show in

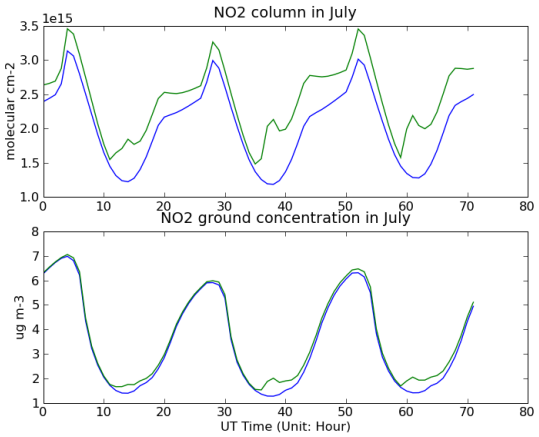

Figure 4: The spatial-averaged model results with and without data assimilation in 17-20 July 2005 (assimilation process). Top: Columns. Bottom: Ground concentration.

Figure 4. This probably explains why the statistics of ground comparisons are similar with or without assimilation.

The prediction process is also carried out in summer. Its results show almost the same statistics as those without assimilation, either evaluating model columns or ground concentrations. This indicates that the assimilation of OMI columns has no clear impact on the next day forecast of NO_2 in summer, consistent with NOx having a shorter life time in the warm season.

It is known that NO_2 is an important precursor of O_3. In the following we examine if the assimilation of satellite NO_2 data can affect O_3 forecast. The assimilation process is discussed here since the prediction has the similar or even less effects. The model surface O_3 concentrations are compared with EMEP observations with and without assimilating NO_2, and their correlation and RMSE are computed with the daily peak values of O_3. The daily maximum is usually seen around 1300-1500 UT from the model simulation, which is near or a little later than the satellite measuring time (around 1000-1400 UT) in Europe. It is found that with assimilation the daily peak O_3 concentrations are increased in summer; the RMSE against ground O_3 observations becomes even larger, 24.9 without assimilation and 28.7 with assimilation. Since the mean daily peak concentration is 106.5 from the reference simulation and 97.8 from ground observations, the model already overestimating the O_3 without assimilation. This may explain the increase of RMSE with assimilation. The correlation with ground data changes slightly with and without assimilation (0.61 versus 0.63). Therefore, the data assimilation of NO_2 columns may not necessarily improve the forecast of ground O_3 in summer. In November or December, the changes of O_3 due to NO_2 assimilation are not clear and the statistics are almost the same as those without assimilation. This may be explained by O_3 concentrations being generally lower and less sensitive to NO_2 concentrations in cold seasons than in warm ones.

7 Conclusions

In this study we have performed the data assimilation of OMI NO_2 columns in the Polyphemus air quality system. Good consistency is seen in the comparisons of model simulation, satellite data and ground observations before assimilation. The optimal interpolation method is then applied to produce analysis of NO_2. The assimilation is carried out with available satellite data in the studied period, and the next-day prediction is also performed. The model results (from the reference simulation, assimilation and prediction) are compared with the EMEP ground obsevations for evaluation.

It is found that in winter the RMSE between model and ground observations is less with assimilation than that without assimilation, reduced by 13% on average. The next-day predictions also show a better forecast of NO_2 with RMSE decreased by 7%. Therefore, OMI data assimilation has the potential to improve the forecast of surface NO_2 concentrations in the cold season.

On the other hand OMI data assimilation shows no clear impact on the surface NO_2 forecast in summer, probably due to the shorter life time and higher diurnal variability of NO_2 in the warm season.

The assimilation of NO_2 also has an impact on O_3 and the effects are more clearer in summer than in winter. This however does not improve the O_3 forecast due to model already overestimating O_3.

References

Austin, J. (1992). Toward the four dimensional assimilation of stratospheric chemical constituents. *J. Geophys. Res.*, 97(D2):2569–2588.

Balgovind, R., Dalcher, A., Ghil, M., and Kalnay, E. (1983). A stochastic-dynamic model for the spatial structure of forecast error statistics. *Mon. Wea. Rev.*, 111(4):701–722.

Boersma, K. F., Eskes, H. J., Veefkind, J. P., Brinksma, E. J., van der A, R. J., Sneep, M., van der Oord, G. H. J., Levelt, P. F., Stammes, P., Gleason, J. F., and Bucsela, E. J. (2006). Near-real time retrieval of tropospheric NO_2 from OMI. *Atmospheric Chemistry and Physics Discussions*, 6(6):12301–12345.

Boutahar, J., Lacour, S., Mallet, V., Quélo, D., Roustan, Y., and Sportisse, B. (2004). Development and validation of a fully modular platform for numerical modelling of air pollution: POLAIR. *Int. J. Env. and Pollution*, 22(1/2):17–28.

Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V.-V., V., K. V. C. K., and Goede, A. P. H. (1999). Scimachy: Mission Objectives and Measurement Modes. *J. Atmos. Sci.*, 56(127–150).

Bucsela, E. J., Celarier, E. A., Wenig, M. O., Gleason, J. F., Veefkind, J. P., Boersma, K. F., and Brinsma, E. J. (2006). Algorithm for NO_2 vertical column retrieval from the Ozone Monitoring Instrument. *IEEE Trans. Geosci. Remote Sens*, 44(5).

Burnett, R. T., Steib, D., Brook, J. R., Cakmak, S., Dales, R., Raizenne, M., Vincent, R., and Dann, T. (2004). The short-term effects of nitrogen dioxide on mortality in Canadian cities. *Arch. Environ. Health*, 59(228–237).

Burrows, J. P., Webber, M., and et al. (1999). The Global Ozone Monitoring Experiment (GOME): Mission concept and first scientific results. *J. Atmos. Sci.*, 56(151–175).

Daley, R. (1993). *Atmospheric Data Analysis*. Cambridge University Press.

Dunlea, E. J., Herndon, S. C., Nelson, D. D., Volkamer, R. M., Martini, F. S., Sheehy, P. M., Zahniser, M. S., Shorter, J. H., Wormhoudt, J. C., Lamb, B. K., Allwine, E. J., Gaffney, J. S., Marley, N. A., Grutter, M., Marquez, C., Blanco, S., Cardenas, B., Retama, A., Villegas, C. R. R., Kolb, C. E., Molina, L. T., and Molina, M. J. (2007). Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment. *Atmos. Chem. Phys. Discuss.*, 7:569—-604.

Fisher, M. and Lary, D. J. (1995). Lagrangian four dimensional variational data assimilation of chemical species. *Quart. J. Roy. Meteor. Soc.*, 121(527):1681–1704 Part A.

Jeuken, A., Eskes, H., Velthoven, P. V., Kelder, H., and Holm, E. (1999). Assimilation of total ozone satellite measurements in a three-dimensional tracer transport model. *J. Geophys. Res.*, 104(5551—5563).

Kalnay, E. (2003). *Atmospheric Modeling, Data assimilation and Prediction*. Cambridge.

Konovalov, I. B., Beekmann, M., Richter, A., and Burrows, J. P. (2006). Inverse modelling of the spatial distribution of NO_x emissions on a continental scale using satellite data. *Atmos. Chem. Phys.*, 6(1747-1770).

Korsakissok, I. and Mallet, V. (2008). Comparative study of Gaussian dispersion formulae within the Polyphemus platform: evaluation with Prairie Grass and Kincaid experiments. *J. Applied Meteor.* Submitted.

Kramer, L., Leigh, R. J., Remedios, J. J., and Monks, P. S. (2008). Comparison of OMI and ground-based in situ and MAX-DOAS measurements of tropospheric nitrogen dioxide in an urban area. *J. Geophys. Res.*, 113(D16S39).

Lamasal, L. N., Martin, R., Donkelaar, A. V., Steibacher, M., Celarier, E. A., Bucsela, E., Dunlea, E. J., and Pinto, J. P. (2008). Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instruments. *J. Geophys. Res.*, 113(D16308).

Leue, C., Wenig, M., Wagner, T., Klimm, O., Platt, U., and Jahne, B. (2001). Quantitative analysis of NO_x emissions from gome satellite image sequences. *J. Geophys. Res.*, 106(5493—5505).

Levelt, P., Khattatov, B., Gille, J., Brasseur, G., Tie, X., and Waters, J. (1999). Assimilation of MLS ozone measurements in the global 3D CTM ROSE. *J. Geophys. Res.*, 25(4493—4496).

Levelt, P. F., van den Oord, B., and et al. (2006). The Ozone Monitoring Instrument. *IEEE Trans. Geosci. Remote Sens*, 44(1093–1101).

Logan, J. A. (1983). Nitrogen oxides in the troposphere: Global and regional budgets. *J. Geophys. Res.*, 88(10).

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. (2007). Technical Note: The air quality modeling system Polyphemus. *Atmos. Chem. Phys.*, 7(20):5,479–5,487.

Middleton, P., Stockwell, W. R., and Carter, W. P. L. (1990). Aggregation and analysis of volatile organic compound emissions for regional modeling. *Atmos. Env.*, 24A(5):1,107–1,133.

Murphy, D. M., Fahey, D. W., Proffitt, M. H., Liu, S. C., Chan, K. R., Eubank, C. S., Kawa, S. R., and Kelly, K. K. (1993). Reactive nitrogen and its correlation with ozone in the lower stratosphere and upper troposphere. *J. Geophys. Res.*, 98(8751–8773).

Napelenok, S. L., Pinder, R. W., Gilliland, A. B., and Martin, R. V. (2008). A method for evaluating spatially-resolved NO_x emissions using Kalman filter inversion, direct sensitivities, and space-based NO_2 observations. *Atmos. Chem. Phys. Discuss.*, 8:6469–6499.

Panella, M., Tommasini, V., Binotti, M., Palin, L., , and Bona, G. (2000). Monitoring nitrogen dioxide and its effects on asthmatic patients: Two different strategies compared. *Environ. Monit. Assess*, 63(447–458).

Quélo, D., Krysta, M., Bocquet, M., Isnard, O., Minier, Y., and Sportisse, B. (2007). Validation of the Polyphemus platform on the ETEX, Chernobyl and Algeciras cases. *Atmos. Env.*, 41(26):5,300–5,315.

Richter, A. and Burrows, J. P. (2002). Tropospheric NO_2 from GOME measurements. *Appl. Sci. Res.*, 29(1673–1683).

Richter, A., Burrows, J. P., Nüß, H., Granier, C., and Niemeier, U. (2005). Increase in Tropospheric Nitrogen Dioxide Over China Observed from Space. *Nature*, 437(129—132).

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

Roustan, Y., Sartelet, K. N., Tombette, M., Debry, E., and Sportisse, B. (2008). Simulation of aerosols and gas-phase species over Europe with the Polyphemus system. part II: model sensitivity. *Atmos. Env.* Submitted.

Sartelet, K., Debry, E., Fahey, K., Roustan, Y., Tombette, M., and Sportisse, B. (2007). Simulation of aerosols and gas-phase species over Europe with the Polyphemus system: Part I, model-to-data comparison for 2001. *Atmos. Env.*, 41(6116-–6131).

Sartlet, K. N., Hayami, H., and Sportisse, B. (2008). MICS Asia Phase II—-Sensitivity to the aerosol module. *Atmos. Env.*, 42(15).

Schaub, D., Brunner, D., Boersma, K. F., Keller, J., Folini, D., Buchmann, B., Berresheim, H., and Staehelin, J. (2007). Sciamachy tropospheric NO_2 over Switzerland: Estimates of NO_x lifetimes and impact of the complex Alpine topography on the retrieval. *Atmos. Chem. Phys.*, 7(5971-–5987).

Simpson, D., Winiwarter, W., Börjesson, G., Cinderby, S., Ferreiro, A., Guenther, A., Hewitt, C. N., Janson, R., Khalil, M. A. K., Owen, S., Pierce, T. E., Puxbaum, H., Shearer, M., Skiba, U., Steinbrecher, R., Tarrasón, L., and Öquist, M. G. (1999). Inventorying emissions from nature in Europe. *J. Geophys. Res.*, 104(D7):8,113–8,152.

Smith, B. J., Nitschke, M., Pilotto, L. S., Ruffin, R. E., Pisaniello, D. L., , and Willson, K. J. (2000). Health effects of daily indoor nitrogen dioxide exposure in people with asthma. *Eur. Respir. J.*, 16(879–885).

Steib, D., Judek, S., , and Brunett, R. T. (2003). Meta-analysis of time-series studies of air pollution and mortality: Update in relation to the use of generalized additive models. *J. Air Waste Manage. Assoc.*, 53(258-–261).

Steinbacher, M., Zellweger, C., Schwarzenbach, B., Bugmann, S., Buchmann, B., Ordóñez, C., Prevot, A. S. H., and Hueglin, C. (2007). Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques. *J. Geophys. Res.*, 112(D11307).

Wu, L., Mallet, V., Bocquet, M., and Sportisse, B. (2008). A comparison study of data assimilation algorithms for ozone forecasts. *J. Geophys. Res.*, 113(D20310).

Zhang, L., Brook, J. R., and Vet, R. (2003). A revised parameterization for gaseous dry deposition in air-quality models. *Atmospheric Chemistry and Physics*, 3(6):2067–2082.

Centre de recherche INRIA Paris – Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399