**Atmospheric
Chemistry
and Physics**

Assimilation of IASI partial tropospheric columns with an Ensemble Kalman Filter over Europe

A. Coman[1], G. Foret[1], M. Beekmann[1], M. Eremenko[1], G. Dufour[1], B. Gaubert[1], A. Ung[2], C. Schmechtig[1], J.-M. Flaud[1], and G. Bergametti[1]

[1]Laboratoire Inter-universitaire des Systèmes Atmosphériques (LISA), UMR7583, Universités Paris Est Créteil et Paris Diderot, CNRS, Créteil, France
[2]Institut National de l'Environnement Industriel et des Risques (INERIS), Verneuil-en-Halatte, Oise, France

Correspondence to: A. Coman (coman@lisa.u-pec.fr)

Abstract. Partial lower tropospheric ozone columns provided by the IASI (Infrared Atmospheric Sounding Interferometer) instrument have been assimilated into a chemistry-transport model at continental scale (CHIMERE) using an Ensemble Square Root Kalman Filter (EnSRF). Analyses are made for the month of July 2007 over the European domain. Launched in 2006, aboard the MetOp-A satellite, IASI shows high sensitivity for ozone in the free troposphere and low sensitivity at the ground; therefore it is important to evaluate if assimilation of these observations can improve free tropospheric ozone, and possibly surface ozone. The analyses are validated against independent ozone observations from sondes, MOZAIC[1] aircraft and ground based stations (AIRBASE – the European Air quality dataBase) and compared with respect to the free run of CHIMERE. These comparisons show a decrease in error of 6 parts-per-billion (ppb) in the free troposphere over the Frankfurt area, and also a reduction of the root mean square error (respectively bias) at the surface of 19 % (33 %) for more than 90 % of existing ground stations. This provides evidence of the potential of data assimilation of tropospheric IASI columns to better describe the tropospheric ozone distribution, including surface ozone, despite the lower sensitivity.

The changes in concentration resulting from the observational constraints were quantified and several geophysical explanations for the findings of this study were drawn. The corrections were most pronounced over Italy and the Mediter-
ranean region, we noted an average reduction of 8–9 ppb in the free troposphere with respect to the free run, and still a reduction of 5.5 ppb at ground, likely due to a longer residence time of air masses in this part associated to the general circulation pattern (i.e. dominant western circulation) and to persistent anticyclonic conditions over the Mediterranean basin. This is an important geophysical result, since the ozone burden is large over this area, with impact on the radiative balance and air quality.

1 Introduction

In the last decade, there has been an increased need in developing air quality standards, guidelines and strategies for air quality management. In particular, special attention has been attached to ozone. Repetitive exposure to enhanced ozone levels can cause various health problems, mostly respiratory illnesses like bronchitis and asthma (WHO, 2003). To comply with these needs, a large variety of observational platforms has been developed: a network of ground based in situ or satellite data, ozone sondes, lidar, and commercial or research aircraft; in parallel, deterministic models (e.g. regional chemical transport model) have been developed, which take into account a large number of physical and chemical interactions between predictor variables as well as required input data (emissions, meteorology and landcover). Lately, European projects such as GEMS[2] (Global and regional Earth System (Atmosphere) Monitoring using

---

[1]Measurements of OZone, water vapour, carbon monoxide and nitrogen oxides by in-service AIrbus airCraft (http://mozaic.aero.obs-mip.fr/web/).

[2]http://gems.ecmwf.int/

Satellite and in situ data) and MACC[3] (Monitoring Atmospheric Composition and Climate) have attempted to integrate enhanced observing systems and modelling tools to develop a comprehensive atmospheric composition monitoring and forecasting system. Combining the two pieces of information (observations and modeling) provides us a framework named data assimilation (DA), where the knowledge about the two components is used in order to obtain an optimal estimate for the state of the system under study: in our case the tropospheric ozone 3-dimensional (3-D) distribution.

The problem addressed by data assimilation is to obtain accurate analyses and/or forecasts of an evolving system, given an imperfect dynamical model and observations. There are two classical ways to deal with this problem: one is sequential data assimilation that only considers observations made in the past until the time of analysis (which is the case of real-time assimilation systems) and the other one non-sequential, or retrospective, where observations from the future can be used, for instance in a reanalysis exercise (Bouttier and Courtier, 1999). Examples of the two types of algorithms identified can be found in the above cited work (in Fig. 2, Sect. 1): Optimal Interpolation, 3D-Var, Extended Kalman Filter for the first category and 4D-Var, fixed-lag Kalman Smoother and Kalman Smoother for the second one, roughly classified according to their complexity and applicability to real-time problems. Compromises between these approaches are possible (for example the use of the 3D-Var in the ERA-40 reanalysis following Uppala et al., 2005). A rigorous approach to solve the problem formulated above is Bayesian estimation (Maybeck, 1979), but its full-scale implementation is, most of the time, not possible due to the size of the problem. In many practical applications, like for air quality it is necessary to make simplifying assumptions. Talagrand (2003) emphasized that there are two main lines in such assumptions: statistical linear estimation and ensemble assimilation. Statistical linear estimation achieves Bayesian estimation when the system is linear and the errors are Gaussian. In particular, it provides a way of estimating the Best Linear Unbiased Estimate (BLUE) (Talagrand, 2003). Ensemble assimilation is a form of Monte-Carlo approximation which attempts to estimate probability distribution functions (pdf) from the spread of the ensemble (Lahoz et al., 2007). The first category is represented by algorithms such as: Optimal Interpolation, the variational methods or various suboptimal versions of the sequential Kalman Filter. The second category is represented by the Ensemble Kalman Filter (EnKF) or the Particle filters (PF). Initiated by Evensen (1994), the EnKF method uses the Monte Carlo approach to sample the error statistics.

There are two important objectives when we apply data assimilation. First, described above, is the state estimation problem which consists in finding the best estimate of the model state which best fits the model equations and the observed data. The second one is the parameter estimation. In this case, we want to improve estimates of a set of poorly known model parameters, the errors in the model being associated with uncertainties in the selected parameters. There exists also combined state and parameter estimation, where the two problems are addressed simultaneously. This issue can be solved very efficiently using ensemble or variational methods. In this study only the state estimation problem using ensemble methods is discussed.

Data assimilation is a relatively young research application in atmospheric chemistry and air quality, but has been widely applied in meteorology, where the principal objective was to improve the initial conditions and forecasts. Starting in the nineties of the last century, assimilation of satellite observation of chemical constituents was begun with a strong focus on stratospheric ozone and related constituents (Khattatov et al., 2000; Lahoz et al., 2007). Recently, the production of total ozone forecasts has become routine in a number of operational centres such as ECMWF (European Centre for Medium-Range Weather Forecasts) and KNMI (Koninklijk Nederlands Meteorologisch Instituut). At ECMWF, TOMS (Total Ozone Mapping Spectrometer) and SBUV (Solar Backscatter Ultraviolet Instrument) data were assimilated during the 40-yr reanalysis (Dethof and Holm, 2002), and more recently MIPAS (Michelson Interferometer for Passive Atmospheric Sounding) and SCIAMACHY (SCanning Imaging Absorption spectroMeter for Atmospheric CartrograpHY) measurements, while the KNMI provides daily global ozone forecasts based on assimilation of total ozone from GOME (Global Ozone Monitoring Experiment) (Eskes et al., 2003) and nowadays from SCIAMACHY. Lately, particular attention has been paid to assimilation of ozone in the free troposphere, because tropospheric ozone is also known as a strong greenhouse gas (Forster et al., 2007). Lamarque et al. (2002) assimilated tropospheric ozone columns from TOMS in the global chemistry-transport model MOZART: assimilation consistently improved the model results when compared with the available independent ozonesonde observations. Segers et al. (2004) assimilated GOME ozone profiles into the global chemistry-transport model TM3 using a Kalman Filter with anisotropic covariance. A case study showed that the assimilation of GOME profiles is able to improve the simulation of the vertical ozone distribution even in the case of strong vertical gradients.

Concerning air quality, there are several studies presenting assimilation of ozone data. In the case of ground based ozone measurements, the first objective was the improvement of the ozone field together with a bias/root mean square error (rmse) reduction (with respect to independent data sets). This type of result was presented by Hanea et al. (2004) using hourly ozone concentrations from European ground stations in a sequential framework and Chai et al. (2007) using various platforms (aircraft, surface and ozone sondes) and a variational approach. A second objective, equally important, especially in air quality, is the improvement in the ozone

---

[3]http://www.gmes-atmosphere.eu/

forecast. We know that, at continental scale, data assimilation can partially correct model deficiencies (Blond and Vautard, 2004), but the information injected in the system decays quickly in time; nevertheless, this permits significant improvements in up to one-day ozone forecasts as emphasized by Elbern and Schmidt (2001), Blond and Vautard (2004) and Wu et al. (2009). In order to avoid this rapid decay, a reasonable alternative could be the use of satellite data, because these data would add information into the system upwind of the considered target areas. To complement the existing, but sparse, observations from sondes and commercial aircraft, the new generation of space borne infrared sounders appears promising. This is the case of the TES (Tropospheric Emission Spectrometer) instrument, launched in July 2004 aboard the NASA EOS Aura satellite, with a routine operating procedure on a 1-day-on, 1-day-off cycle. In a first study, Parrington et al. (2008) examined the potential of TES observations (ozone and CO retrievals) for July and August 2006 to constrain the tropospheric ozone distribution. In a second paper, Parrington et al. (2009) tested the impact of assimilating ozone observations from TES on North American surface ozone abundances in the GEOS-Chem model in August 2006. They found significant improvements in the bias in the modeled free tropospheric ozone, as well as a large reduction of the model bias with respect to surface ozone observations in the western USA.

Recently, a new instrument called IASI (Infrared Atmospheric Sounding Interferometer) was launched aboard the European Metop-A satellite (in October 2006). The instrument design resulted from compromises between the meteorology requirements (high spatial coverage) and the atmospheric chemistry needs (accuracy and vertical information for trace gases) (Clerbaux et al., 2009). In comparison with TES, IASI has a better spatial coverage due to its 200 km swath along the track that allows a twice daily overpass frequency. Its good horizontal resolution and strong sensitivity to free tropospheric ozone make IASI well suited for measurements of tropospheric ozone with focus on air quality (Eremenko et al., 2008; Dufour et al., 2010).

The assimilation of IASI data has already been studied. Massart et al. (2009) studied the quality of IASI derived total ozone columns measurements by comparing them with an ozone field obtained by assimilation of the ozone profiles from the MLS (Microwave Limb Sounder) instrument and of total ozone columns from the SCIAMACHY nadir instrument. The comparison shows that on average, the IASI data tend to overestimate the total ozone columns by 2 % to 8 %; additionally the random observation error of the same data is estimated to be about 7 % except over polar and deserts areas where it is higher due to high surface emissivity.

In view of the elements presented above, the aim of the present study was to show the potential of the IASI data to constrain the 3-D tropospheric ozone distribution and to quantify the gain obtained by assimilating these data in a regional chemistry-transport model, CHIMERE. IASI obser-

vations are especially useful for this purpose, because of their high horizontal resolution (about 12 km under the satellite track), and because of their strong sensitivity to tropospheric ozone (Clerbaux et al., 2009). To our knowledge, this is the first study, trying to assimilate IASI data into a regional chemistry-transport model over Europe, in order to improve the estimate of free troposheric ozone fields and surface ozone fields. In this region, a large number of surface stations and vertical profiles from sounding sites and commercial aircraft can be used as independent data for evaluation/validation purposes.

Section 2 describes the satellite data used in the assimilation exercise and in Sect. 3 a description of the chemistry-transport model is given. Section 4 contains a detailed description of the selected assimilation algorithm and the set-up of the experiments performed in this work. Results of the application described above and its evaluation are presented and discussed in Sect. 5, followed by the conclusions in Sect. 6.

2 The satellite data

In this work we use the tropospheric O_3 retrievals obtained using measurements performed with the Infrared Atmospheric Sounding Interferometer (IASI). The retrieval of ozone profiles from IASI spectra was performed with the radiative transfer model KOPRA (Karlsruhe Optimised and Precise Radiative transfer Algorithm, Stiller et al., 2000) and its numerical inversion module KOPRAFIT. Both were developed for limb retrievals (MIPAS mission, Fisher, 2008) and adapted later for the Nadir geometry. The inversion method for ozone was developed to achieve maximal information content in the lower troposphere and it was set-up and first applied by Eremenko et al. (2008). A validation exercise performed over the first one-year-and-a-half of IASI operation for the northern midlatitudes shows no significant bias (less than 5 %) in the retrieved ozone and that the errors derived from the error budget calculation (about 18 % for 0–6 km partial columns for mid-latitudes) are consistent with the standard deviation of the differences between sonde measurements and IASI observations (Keim et al., 2009).

IASI observations have a good spatial resolution (the vertical Nadir field of view for one IASI pixel has the diameter of 12 km at the surface), comparable to the usual continental range resolution of the models (0.2–0.5°); however, a major limitation of the IASI data is linked to the temporal resolution, with two overpasses per day: first in the morning and a second one in the evening. Note although, that a twice daily measurement frequency is a major achievement of the IASI instrument (compared for example with GOME2, OMI-Ozone Monitoring Instrument or TES instrument).

The principal diagnostic of the inversion of IASI ozone partial columns is the Averaging Kernel Matrix (AVK), which expresses the sensitivity of the retrieved profile to the
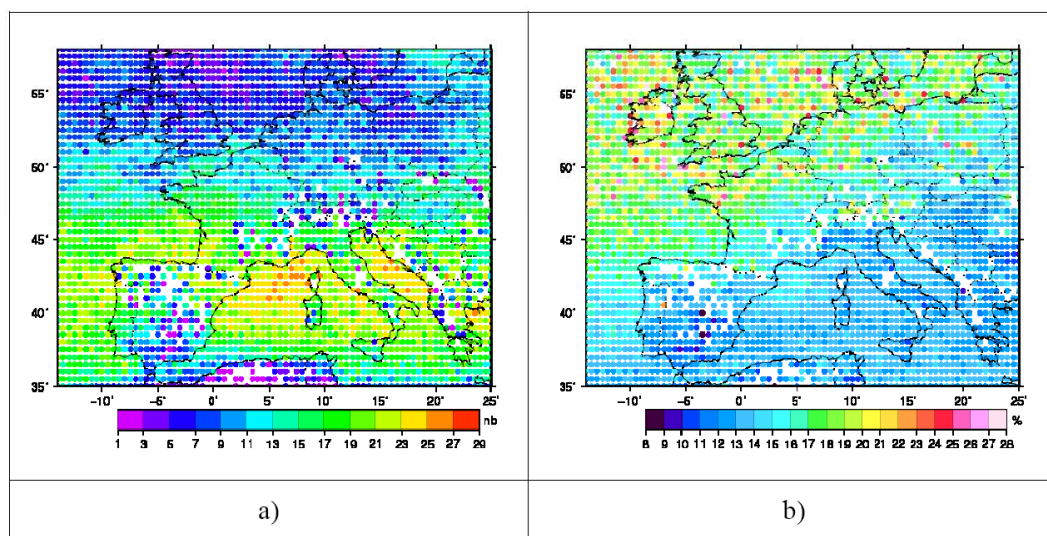
Fig. 1. (a) The spatio-temporal distribution of days when at least one pixel was available for assimilation for a given grid cell (maximum of 30 days i.e. number of days included in the assimilation experiment, color scale varying from violet to red: violet represents relatively low values and red relatively high values) and **(b)** the mean relative error (expressed in percent %) of the IASI inversions (error calculated on the ozone partial columns 0–6 km) averaged over the month of July 2007 (black represents relatively low values and pink relatively high values).

true profile. The averaging kernel matrix describes both the vertical resolution of each level in the retrieval and the relative weighting in the solution between the true profile and a priori profile (in this case the climatology from McPeters et al., 2007). Thus, the averaging kernel matrix provides the information which, if properly applied to a particular in situ profile data, transforms that measurement in order to have the same resolution and a priori dependence as the IASI retrievals (see the equation below from Rodgers, 2000):

$$\hat{x} = x_{\mathrm{a}} + \mathbf{A}(x_{\mathrm{m}} - x_{\mathrm{a}}) = \mathbf{A} x_{\mathrm{m}} + (\mathbf{I} - \mathbf{A}) x_{\mathrm{a}} \qquad (1)$$

where x_{m} represents the model simulated profile, x_{a} is the a priori profile (here issued from the McPeters climatology), $\mathbf{I}$ is the identity matrix and $\mathbf{A}$ is the averaging kernel; in this way we transform the model profile into a pseudo retrieved profile. This matrix is calculated during the retrieval process for each individual retrieved profile. Examples of typical averaging kernels can be found in Dufour et al. (2010). Due to the nature of the measurement, note that (1) it is impossible to separate information originating from nearby vertical levels; (2) the sensitivity to the lower levels of the ozone profile (below 3 km) is relatively small. However, the method developed by Eremenko et al. (2008) allows to discriminate the lower and the upper troposphere when thermal conditions are favorable, typically during summer (Dufour et al., 2010).

We use the 0–6 km (partial) columns calculated from the morning pass of the satellite over the European domain. Use of these columns is a compromise between using observations as sensitive as possible towards surface ozone (values of the AVK matrix for the lower levels greater than 0) and representing as much as possible one independent piece of

information (the sum of the AVK diagonal values greater than 1). The latter requirement would mean to use ozone columns integrated from 0 to 8–9 km to increase DOF (degrees of freedom), but which would be less sensitive to free tropospheric ozone than the 0–6 km columns. Only morning overpasses are used, because they are more sensitive in the lowermost atmospheric layers, due to a higher contrast between the temperature at the surface and temperature in the first atmospheric layers. In this study, we consider only the pixel with the hottest surface temperature in the grid cell i.e., with the potentially largest sensitivity to the lowermost troposphere.

For a more visual approach, in Fig. 1a (left panel) we present the spatio-temporal distribution of the IASI data (pixels) available for July 2007 on the European domain under study. As we can see, the monthly spatial coverage is not homogeneous. In terms of number of pixels available for each day, the Mediterranean Basin is the most represented (more than 20 observations). On the contrary, north of 50° N, the number of pixels per month is in general below 10. However, even in the case of sparse observations, the assimilation system can significantly spread information in space and time, through local assimilation or advection from information-rich to information-poor regions.

The error budget, calculated for each individual retrieved profile accounts for errors due to the measurement noise, to the uncertainty in the temperature profile, to smoothing and to other sources (Eremenko et al., 2008). The corresponding total error in the 0–6 km columns used in this study ranges from 10 to 20 %, corresponding to less than 3 Dobson Units (DU). A map with the monthly averaged (July 2007) error

calculated for the European domain is presented in Fig. 1b (right panel). As we can see, errors are often larger over the western part of the domain (over the ocean) because of less favorable thermal conditions (lower surface temperature and thermal contrast than over the continent) and cloudy conditions. It is worth noting that the errors described above represent an essential piece of information to be used in the assimilation scheme as we will see in Sect. 4.

For assimilation purposes, the DOF is an important parameter that gives a proxy of the information injected in the system. In Rodgers (2000), "the degrees of freedom for signal describe the number of useful independent pieces of information in the retrieved quantity". It can be evaluated by taking the trace of the averaging kernel matrix. As DOFs for lower tropospheric ozone columns (0–6 km), averaged over July 2007, reach a maximum value of 0.7 (Fig. 2), we do not have one independent piece of information. There is generally more independent information in the southern half of the domain than in the northern part and the same difference can be observed between the continental part and the sea or ocean. This is likely due to the lower thermal contrast over northern and oceanic areas than over southern and continental ones. We expect that these differences will have some influence on the behavior of the assimilation system. This is one of the reasons for splitting the study domain into four quadrants for evaluation of the assimilation system (see Fig. 2).

To close the section dedicated to the observation set, we present in Fig. 3a the temporal evolution of the altitude at which the maximum AVK is reached for the four zones shown in Fig. 2. As we can see, the maximum of the AVK for different days in July 2007 occurs in general between 3.5 and 5.5 km altitude with more variability during the first ten days. We remark that, for the zone D (south-eastern part), maxima are lower than for the other three zones for almost the entire period under study; this is a priori a more suitable situation for obtaining more information about surface and boundary layer ozone levels. In Fig. 3b, we show the daily coverage (available pixels) for the four regions considered during July 2007. Most pixels are available in zone D and C (north-eastern part). Over the south-eastern part of Europe (zone D), the presence of persistent anticyclonic conditions during summer limits the cloud formation. The analysis of meteorological conditions of July 2007 also shows a heat wave event located over the north-eastern part of the domain during the second decade of the month that was also associated with cloud cover. We will see in the results section the impact of this on the assimilation experiment.

3 Regional Chemistry-Transport Model (RCTM)

In order to simulate ozone concentrations, we use the CHIMERE RCTM version described in Bessagnet et al. (2008) and documented at www.lmd.polytechnique.fr/
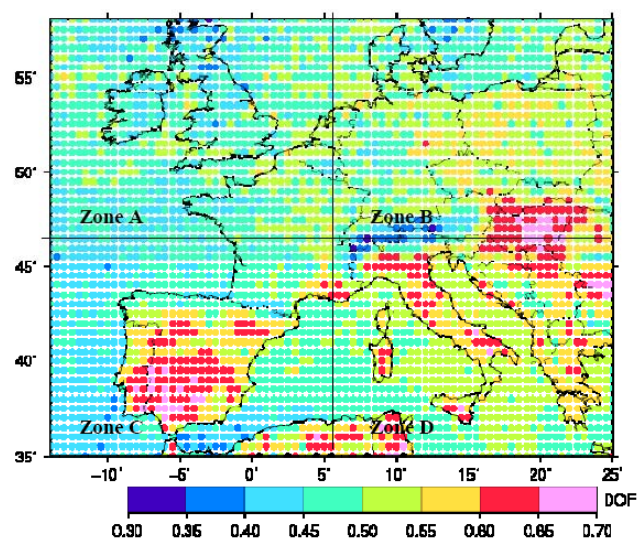

Fig. 2. Monthly mean of Degrees of Freedom (DOF) for the IASI data (0–6 km columns) used in the study (blue color represents relatively low values; pink color represents relatively high values). The domain was split in four zones: zone A for North-West, zone B for North-East, zone C for South-West and zone D for South-East.

chimere. This model simulates a wide variety of gaseous pollutants (O_3, NO_x, SO_2, CO, Volatile Organic Compounds . . .) as well as airborne particulate matter (this latter option is switched off for the current simulations). It has been used for numerous air quality studies dealing with gaseous and/or particulate pollution (e.g. Vautard et al., 2005; Coll et al., 2005; Hodzic et al., 2006; Deguillaume, 2008). It works operationally on the national French PREV'AIR platform (www.prevair.org, Rouil et al., 2009) to produce ozone, NO_x and particulate concentrations short-term forecasts and analyses at the continental scale. In this framework, ozone concentrations simulated with the CHIMERE model have been compared to ground-based measurements for three spring/summer periods between 2004 and 2006. Honoré et al. (2008) showed that the mean model bias of daily ozone maxima was mostly under $5\,\mu g\,m^{-3}$ (micrograms per cubic meter), RMSE (root mean square error) was generally less than $20\,\mu g\,m^{-3}$ and temporal correlation was more than 0.8 on average over Western Europe. Also in a European framework, CHIMERE is currently used for several operational applications, with recent examples in ensemble forecasting in the GEMS project (Hollingsworth, 2008) and data assimilation within the MACC project (http://www. gmes-atmosphere.eu/).

For this work, the simulation is set up over a large western European domain (ranging from $14°$ W to $25°$ E and from $35°$ N to $58°$ N). To cover this domain, we used 3713 (79×47) horizontal grid points with $0.5° \times 0.5°$ horizontal resolution. 17 vertical levels are defined following a hybrid (σ, p) scheme; their thickness varies from 50 m in
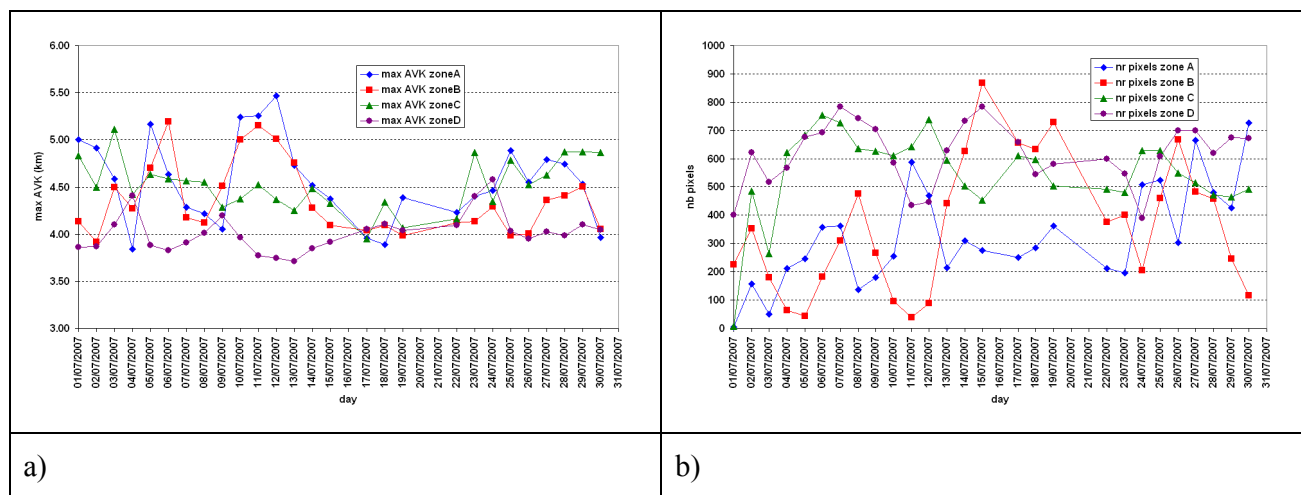
Fig. 3. (a) Day to day variation of the altitude of the maximum AVK (averaging kernel) averaged over the four quadrants of the domain, for July 2007; **(b)** day to day variation of the number of pixels used in assimilation for the four zones (blue for zone A, red for zone B, green for zone C and violet for zone D).

the surface layer to a maximum value of about 1 km in the free troposphere. We use a monthly climatology obtained from simulations with the LMDz-INCA model (Interaction of Chemistry and Aerosol model coupled to the Laboratoire de Météorologie Dynamique General Circulation model-LMDz) (Hauglustaine et al., 2004) to impose gaseous concentrations at the domain boundaries (top and lateral conditions). Primary pollutant emissions are based on the EMEP (European Monitoring and Evaluation Programme) 2003 inventory (Vestreng et al., 2005). Meteorological fields (pressure, temperature, wind components, relative humidity, liquid water content and precipitation) are calculated off-line by the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) at a $0.5° \times 0.5°$ horizontal resolution.

4 Assimilation method

4.1 The Ensemble Kalman filter method

In this study, an advanced sequential data assimilation method (EnKF) has been set-up for the purpose of 4-D data assimilation. We examine the possibility of using ensembles, generated by using Monte Carlo methods, to calculate spatially and temporally varying forecast-error covariances for the purpose of performing data assimilation.

There is a long list of successful EnKF applications to a wide range of problems including meteorology (Houtekamer and Mitchell, 2001; Buehner, 2005), oceanography (Keppenne and Rienecker, 2002; Evensen, 2007), oil reservoir modelling (Evensen et al., 2006), air quality (Hanea et al., 2004), stratospheric chemistry (Milewski and Bourqui,

2011) or land modelling (Reichle et al., 2002; Clark et al., 2008).

In mathematical terms, the general data assimilation problem is defined by the computation of the probability density function (PDF) of the model solution, conditioned on the measured observations, (i.e. following the Bayes theorem, we have to estimate a posterior PDF). This PDF is usually represented using statistical moments or an ensemble of model states and searching for estimators like mean, mode or maximum likelihood. In the case of the EnKF, since the size of the ensemble is limited, it is difficult to obtain a very accurate representation of the PDF in high dimensional problems (Evensen, 2007). We restrict ourselves to finding a good estimate for the mean of the PDF. In the case of the EnKF, the ensemble mean and covariance are presumed to fully describe the PDF of both the prior and assimilated fields which are assumed to be Gaussian; thus the solution becomes computationally feasible.

The analysis equation which allows us to update each ensemble member is written as:

$$\mathbf{\Psi}_i^a = \mathbf{\Psi}_i^f + \mathbf{P}_e^f \mathbf{H}^T \left(\mathbf{H} \mathbf{P}_e^f \mathbf{H}^T + \mathbf{R} \right)^{-1} \left(\boldsymbol{d} - \mathbf{H} \mathbf{\Psi}_i^f \right) \tag{2}$$

where $\mathbf{\Psi}_i^f$ represents an ensemble member i (model state) ("f" stands for forecast, "a" for analysis), $\boldsymbol{d}$ is the vector of observations available at the time of analysis, $\mathbf{H}$ represents the linear version of the observation operator which permits the projection from the model space onto the observation space, $\mathbf{P}_e^f$ is the forecast covariance error matrix, $\mathbf{R}$ is the observation covariance error matrix and

$$\mathbf{K}_e = \mathbf{P}_e^f \mathbf{H}^T \left(\mathbf{H} \mathbf{P}_e^f \mathbf{H}^T + \mathbf{R} \right)^{-1} \tag{3}$$

is known as the Kalman gain matrix, where "T" refers to the transpose of a matrix. The "best estimate" is calculated as a mean over the ensemble members using the formula (with N the ensemble size):

$$\overline{\Psi^a} = \frac{1}{N} \sum_{i=1}^{N} \Psi_i^a \qquad (4)$$

And the analysed covariance error matrix $\mathbf{P}^a$ as the covariance over the ensemble:

$$\mathbf{P}_e^a = \frac{1}{N-1} \sum_{i=1}^{N} \left(\Psi_i^a - \overline{\Psi^a}\right)\left(\Psi_i^a - \overline{\Psi^a}\right)^T \qquad (5)$$

The same formula is used for the forecast covariance error matrix $\left(\mathbf{P}_e^f\right)$ using Ψ_i^f instead Ψ_i^a (and $\overline{\Psi^f}$ instead $\overline{\Psi^a}$).

The ensemble $\left(\Psi_i^f\right)_{i=1,N}$, whose mean is the current state estimate, is updated in the analysis step (Eq. 2) taking into account all knowledge about the error statistics (in the model, $\mathbf{P}_e^f$, and in the measurements, $\mathbf{R}$). The aim here is to transform the forecast ensemble into an analysis ensemble $\left(\Psi_i^a\right)_{i=1,N}$ with appropriate statistics. The key of this method is the use of an ensemble to get the forecast/analysis covariance matrices (see Eq. 5).

They are two ways to treat observations in the EnKF: one consists in adding perturbations to them according to the observational error in order to obtain N vectors of measurements $(d_i)_{i=1,N}$ (N is the ensemble size). In this case, in Eq. (1), we will use d_i instead of d to update each member of the ensemble (details in Burgers et al., 1998). In this manner, we avoid the loss of the ensemble spread after assimilation. An alternative is to use a Square Root Filter formulation (Maybeck, 1979). This formulation avoids the loss of positive definiteness of the error covariance matrices. It was demonstrated that the elimination of the sampling error associated with the perturbed observations makes the EnSRF (Ensemble Square Root Filter) more accurate than the EnKF for the same ensemble size (Whitaker and Hamill, 2002; Sakov and Oke, 2008). This is the reason for selecting the square root formulation in our study (we use the same formulas and notations as in Evensen, 2004).

4.2 Set-up of the assimilation experiments

In this work, we selected the Ensemble Square Root Kalman Filter in order to implement an assimilation system in the RCTM CHIMERE at a continental scale. We allow the model to contain errors, the model error being represented only by the perturbations added to the ozone model field, and use the information both from observations and the model to improve the actual model state. In the configuration applied in this study, the state vector contains the concentrations for all the chemical species in each grid cell. The focus here is on the state estimation rather than on parameter estimation.

There are several methods to set-up an ensemble, but no unified theory has been developed yet, at least for chemistry-transport simulations (Galmarini et al., 2004). Ensembles can be derived from a single model while perturbing model parameters (Beekmann and Derognat, 2003) or numerical and physical parameterisations (Mallet and Sportisse, 2006). An alternative could be to select the ensemble from a long model simulation or just to perturb a set of initial conditions. In this study, the initial ensemble was created by applying 3-dimensional "pseudo-random" perturbations to a reference run. These perturbations were taken from a Gaussian distribution with zero mean, unitary variance and a Gaussian spatial covariance with a fixed decorrelation length (Evensen, 1994) in order to obtain a 2-dimensional field. Given two such pseudo-random fields for two distinct layers, a new couple of fields vertically correlated with a specific covariance between layers can be generated (Eqs. A13–A14 from Evensen, 1994). This procedure is applied for the 17 correlated perturbations fields corresponding to the model layers. The 3-dimensional perturbation field obtained after this procedure is characterised by a decorrelation length fixed at 200 km in the horizontal and at 1 km in the vertical, following the comparisons between a model reference run and ground-based/MOZAIC observations, presented in Boynard et al. (2010). The amplitude of perturbations was fixed at 10 % of the simulated ozone concentrations in each grid cell. Perturbations were applied each 3 h during the spin-up period of 24 h, and then during the whole assimilation period. During the one day forecast periods, between two analyses, these perturbations accumulated to give a dispersion of the ensemble from the mean varying between 17 % and 25 %. This is consistent with the model error statistics established by comparison with surface and free tropospheric ozone observations (Honoré et al., 2008). No temporal correlation was used in this configuration (i.e. white noise was assumed).

The errors in the observations were grouped (representativeness and instrumental error) in the $\mathbf{R}$ matrix, whose diagonal is filled with the results obtained during the inversion procedure (Eremenko et al., 2008). We consider that there is no error correlation between different satellite observations used simultaneously in the assimilation ($\mathbf{R}$ is diagonal). This is certainly a simplification, but the degree of horizontal error correlation is unknown. However, only a limited set of satellite observations is used within the adopted localisation procedure (see below). We see that in our case, the observation error of the 0–6 km ozone column is about 16 % on the average over the pixels available for the whole month (see Fig. 2b with the monthly average relative errors). In the retrieval procedure the diagnosed error is not temporally correlated, thus we do not consider such a correlation in our system.

An ensemble with a limited number of members cannot estimate accurately the forecast error across the entire state space due to spurious error correlations; therefore it is better to restrict the new information provided by the measurement to a local neighbourhood. In a local region, the ensemble size may be sufficient to represent a large portion of the state uncertainty (Szunyogh et al., 2005). "Localisation" has become a very widely used technique to filter out the spurious

long-range correlations, and increase the rank of the forecast covariance matrix. In particular, the most used technique in assimilation is distance based localisation. This method requires the existence of some concept of distance between a state vector element and an observation (Sakov and Bertino, 2010). In complex multivariate models, "even the appropriate distance between spatially and temporally collocated observations and state variables becomes unclear when the observation and state are of different types" (Anderson, 2007). In our case, this problem was simplified because we use the concept of distance only in the Euclidean sense (as a function of geographical coordinates) between an analysed grid point and an observation location.

A first method of localisation was described in Houtekamer and Mitchell (1998). They limit the influence of an observation on the analysis to a surrounding local region using a cutoff radius beyond which covariances between variables are assumed to be zero, also called "covariance localisation". Another localisation method, scheme-independent, is "local analysis" or localisation in model grid space which uses "local approximation of the state error covariance for each updated state vector element by building a virtual local spatial window around this element" (Sakov and Bertino, 2010).

We apply here a "local analysis" in order to avoid spurious correlations in the forecast ensemble, which are introduced by the perturbation method for finite ensemble sizes, and which do not have any geophysical reality. The basic idea of this method is to perform the analysis on a given grid point using the observations within a local region centred at that point and this analysis is performed grid point by grid point. The radius of this region was fixed at 200 km, corresponding to the decorrelation length in the horizontal perturbations applied (following Boynard et al., 2010). The maximum number of observations to be assimilated was limited to 30 pixels. This parameter was subject to sensitivity tests (see later). No vertical localisation was applied. In certain cases, this method can lead to discontinuities in the analysis when an observation is taken into account into a local window and not in the next one, when we move from updating one state vector element to another, but this problem is beyond the scope of this study. Note however that this unsuitable occurrence has been addressed in Hunt et al. (2007).

To assimilate satellite data, we have to be able to project the vectors from the model space onto the observation space, calculate the innovations (differences between the observations and the simulated fields projected in the observational space) and reproject this information onto the forecast model space. All these operations are achieved by constructing $\mathbf{H}$, the observation operator. The formula used is:

$$\mathbf{H}\left(\boldsymbol{\psi}_i^{\mathrm{f}}\right) = S \cdot \mathbf{A} \cdot L\left(\boldsymbol{\psi}_i^{\mathrm{f}}\right) \tag{6}$$

Making observations and simulated fields comparable first implies performing a vertical interpolation L (in order to

have the same number of vertical layers for the model and for the IASI retrieval, one layer for each km up to 12 km). The second operation is a convolution by the averaging kernel $\mathbf{A}$. As already mentioned, the averaging kernel matrix provides the information which, if properly applied to a particular in situ profile data, transforms that profile in order to have the same resolution and a priori dependence as the IASI retrievals. Note that in the assimilation case, adding the a priori profile is not needed (Rodgers, 2000) because the a priori was removed from the IASI columns before, therefore only the first term in Eq. (1) ($A\boldsymbol{x}_{\mathrm{m}}$) is required. The last step for constructing H is the integration on the vertical (S), up to 6 km, in order to obtain a scalar value corresponding to the column value.

There are two remaining parameters to discuss: the ensemble size and the local patch size (maximum number of pixels used in the assimilation step for each grid cell); the sensitivity tests described in the next subsection will clarify the choices made for these parameters in the present study.

4.3 Sensitivity tests

In this subsection, we briefly describe the sensitivity experiments performed for the two parameters considered as the most influential: the ensemble size, which controls the accuracy used for the evaluation of the model prediction covariance error and the local patch size which represents the maximum number of pixels used in the assimilation step and located within 200 km of the analysed grid cell. Closely related to the last parameter is the horizontal decorrelation length. As this parameter was tested and fixed following the results obtained by Boynard et al. (2010) we choose to keep it constant in order to avoid higher CPU requirements. The tests were conducted for three days (17, 18 and 19 July 2007) when the CHIMERE model simulated high ozone concentrations over Central and Eastern Europe. The strong gradient of the simulated ozone field together with a large coverage from the IASI instrument for these three days are two important reasons for choosing this time period.

The ensemble sizes tested were 10, 20, 40 and 80. The increase of sample numbers slightly improves the forecast scores (see Table 1) but we can see that results are very similar. We calculated the RMSE between the simulated columns and the IASI data for each test and we compared also with the RMSE of a reference run (without assimilation); the modest improvements shown in Table 1 are likely due to the saturation of the errors in the assimilation system. It appears that increasing the ensemble size will not improve the assimilation system accuracy: about 40 members (therefore 40 CHIMERE simulations) were found to be sufficient for a good assimilation; therefore this is the size retained in this study. As for the local patch size, its increase from 10 to 20 pixels slightly improved the quality of our analysis and then the increase from 20 to 30 pixels leads to a small loss of precision (values are very close, differences at second digit).

Table 1. Root mean square error (expressed in DU) calculated for the sensitivity tests performed on 17, 18 and 19 July 2007. Tests of ensemble sizes are presented in different rows, maximum local patch sizes (10, 20 and 30 pixels) in different columns. The first row of the table shows the root mean square error calculated for a reference run "without assimilation" (REF). The lowest value of the RMSE is highlighted in bold.

RMSE (DU)	10 pixels	20 pixels	30 pixels
REF	5.40	5.40	5.40
10ens	3.01	2.97	2.98
20ens	2.96	2.94	2.96
40ens	2.95	2.94	2.94
80ens	2.95	**2.92**	2.93

Finally, the largest value tested (30) was used in the simulations presented here. As an example of assimilation for this testing period, in Fig. 4, we plot a comparison between the mean ozone columns before assimilation (left panel), the IASI data (central panel) and the mean ozone columns after assimilation (right panel) for the 17 July 2007 using an ensemble of 40 members and 30 assimilated pixels. Assimilation increases the spatial correlation coefficient (between forecast field and IASI data compared to analysis field versus IASI data, always for 0–6 km columns) from 0.06 to 0.23 (following a linear fit for the scatter plots, not shown). Assimilation results show a large negative correction especially in the Eastern part of the domain. The situation chosen here was special, due to the presence over Eastern Europe of a large anticyclonic system with enhanced temperatures and ozone levels for this day.

5 Results

5.1 General results

Using the assimilation system described above we perform a one month simulation, for July 2007, and we assimilate the tropospheric ozone columns (0–6 km) daily at 09:00 a.m. We start the results section, with some general statistics which help us to evaluate quantitatively the corrections performed on the ozone field using the IASI data (in terms of ozone columns, therefore expressed in Dobson Units).

In Fig. 5 we compare the mean ozone columns before assimilation (left panel) with IASI columns (central panel) and the mean ozone columns after assimilation (right panel) over the entire domain and over the whole month. For these comparisons (as already for those presented in Fig. 4), averaging kernels are applied to the collocated model profiles (in order to remove the dependency of the comparison on the a priori ozone profile information used in the retrieval) and the same a priori was added for the three data sets (details in the

set-up Sect. 2, Eq. 1). From this figure, we can see that the model largely overestimates partial columns in the northern part of the domain. As a consequence the mean European north to south gradient observed by IASI is not captured by the model. This feature is consistent with the results obtained with an ensemble of models by Zyryanov et al. (2011). As a result of assimilation, the large-scale structures in the ozone distribution are as expected more consistent with the IASI data. Except for the large negative correction in the northern half of the domain, the corrections are more pronounced in the southern half of the domain over Spain, Italy and Eastern Europe where the CHIMERE model overestimates the ozone columns.

In a second step, we performed self-consistency diagnostic tests. OmA (observations minus analysis) and OmF (observations minus forecast) differences were computed for every day of the month (for each morning overpass). Technically, this evaluation consists in a comparison between the innovations (defined as observations minus projected forecasts in the observational space – last parenthesis in Eq. 2), called OmF and the difference between observations and projected analysis in the same space, called OmA. Daily, OmA and OmF values for individual grid cells were processed over the whole model domain to calculate the bias (mean bias error – MBE) and the RMSE/RMSED (root mean square error and/or debiased). The formulas used for these performance indices are:

$$MBE = \frac{1}{n} \sum_{i=1}^{n} (O_i - P_i) \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (O_i - P_i)^2} \tag{8}$$

$$RMSED = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(O_i - P_i - (\bar{O} - \bar{P})\right)^2} \tag{9}$$

where P_i and O_i represent the simulated (predicted) and measured concentrations and $\bar{P}$, $\bar{O}$ the means calculated for all n model grid cells with pixels available at time t.

The temporal evolution of these indices is plotted in Fig. 6. Except for some gaps due to the lack of the IASI data, we can notice the same linear shape of the RMSE/bias of the mean of the ensemble columns after analysis (step) versus the IASI columns, which shows rather constant values over the whole period. Note that the large bias/RMSE reduction for 1 July is characteristic for the first assimilation time step performed, since the model was not constrained prior to this date. In terms of gain with respect to forecast we notice a reduction of 1.5 DU for the RMSE, 1 DU for RMSED and 2 DU for the bias, nearly removing the bias. Thus, the assimilation system is very efficient in removing bias and also partly improving the variability of the quadratic error (reduction of the RMSED when comparing green and red curves).
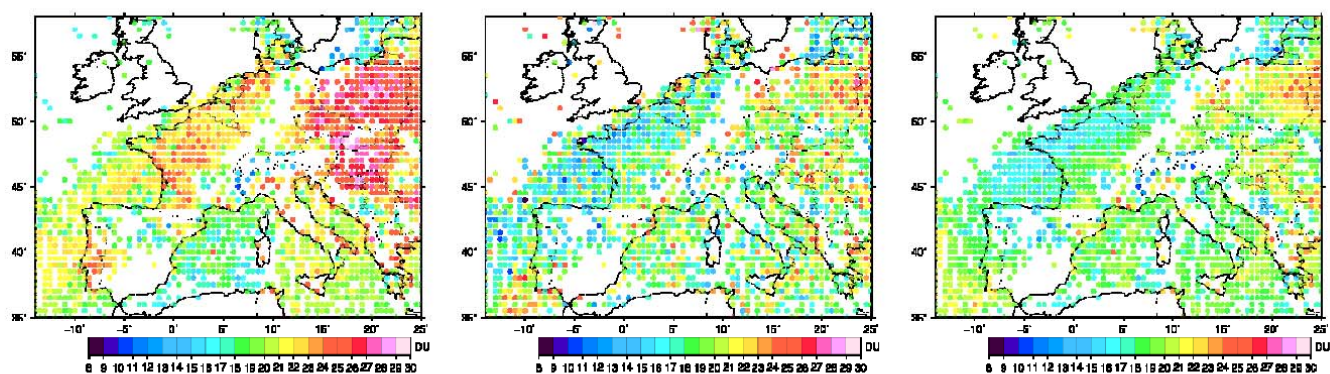
Fig. 4. 0–6 km ozone partial columns (Dobson Units) from the mean of the forecast ensemble (i.e. before assimilation; left panel), IASI (central panel) and the mean of analysis ensemble (i.e. the ensemble after assimilation; right panel) for the 17 July 2007 (at 09:00 a.m.). The white points on the maps correspond to the missing pixels (clouds or bad inversion) (color scale varies from black, representing relatively low values, to pink representing relatively high values).
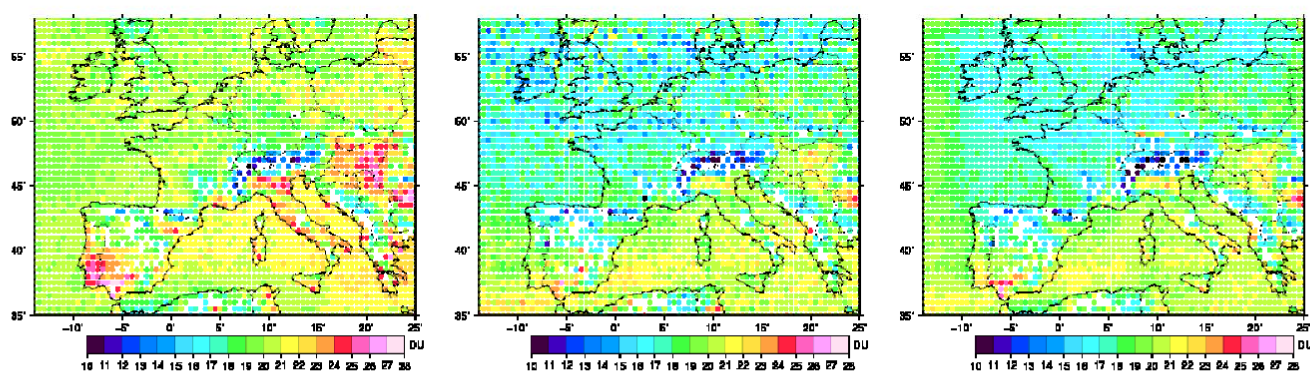

Fig. 5. 0–6 km ozone partial columns (Dobson Units) from the mean of the forecast ensemble (i.e. before assimilation; left panel), from IASI retrievals (central panel) and the mean of analysis ensemble (i.e. the ensemble after assimilation; right panel). Values are averaged over the whole month at 09:00 a.m. The white points on the maps correspond to the missing pixels (clouds or bad inversion) (color scale varies from black, representing relatively low values, to pink representing relatively high values).

The remaining RMSE of analysis is consistent with the IASI error shown in Fig. 1 (about 2 to 2.5 DU). Thus, these first error statistics show a satisfying behaviour of the assimilation system. We can note that the mean bias observed for the monthly mean appears to be a systematic feature. It is likely that chemical boundary conditions (BC) are mainly responsible for this since it largely controls simulated free tropospheric concentrations. The bias build up (of about 2 DU) between two assimilation steps (i.e. 24 h) may be partly due to the advection of biased BC inside the domain.

5.2 Evaluation of the data assimilation system

In a second step, we quantified how much the assimilation procedure improved the forecast of IASI observations for the next day after assimilation, with respect to a free model run without assimilation (called reference model run). IASI observations used for this evaluation have not been employed

prior for assimilation. However, it is difficult to prove that data used for assimilation and data used for evaluation are strictly speaking independent, because the temporal correlation of the errors is not known. This time, the evaluation consists in a comparison between the innovations (OmF) and the difference between observations and the reference run. We called the latter quantity OmRef. The performance indices are presented in Fig. 7 as an integrated quantity column 0–6 km (so the unit is DU) for the four quadrants of the domain (A for north-west, B for north-east, C for south-west and D for south-east). They show that, at least for the second ten days, when the number of available pixels was higher (see Fig. 3b), the filter allows an improvement in ozone columns, especially on the eastern half of the domain; thus we obtained an averaged (over the entire month) gain in bias of 1 DU for the zone B and 1.5 DU for the zone D for the whole month. The better improvement in zone D in the south-eastern part of the domain can be explained by two facts: first, more pixels
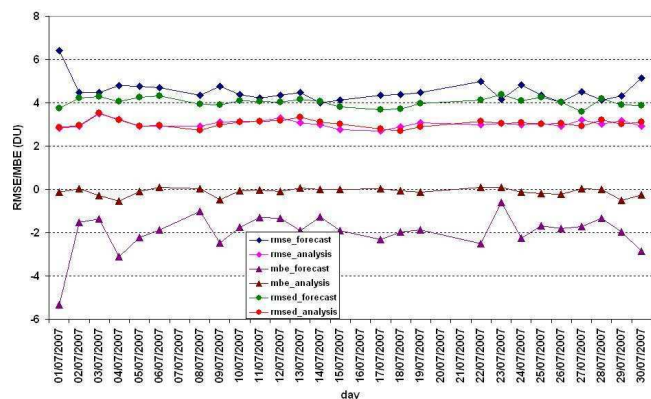
Fig. 6. Temporal evolution of the general statistics RMSE, RMSED and MBE calculated for the data described in Fig. 5: monthly averaged mean of the ensemble ozone columns (forecast and analysis) versus the IASI partial columns for the whole month. The symbols are as follows: diamonds for the RMSE "forecast" or "analysis" (dark blue or magenta respectively); circles for the debiased RMSE "forecast" or "analysis" (green or red respectively); and triangles for the MBE (violet or brown respectively). The unit measure is DU.

were available for the southern part of the domain because of more cloud free conditions (see Fig. 3b) and second, most air masses present in this area have undergone westward transport and spent several days in the simulation domain (cf. back trajectories presented in Foret et al., 2009). It is likely that the initial gain in bias is lost, especially for the quadrant A (north-west), because air masses which have not benefited from assimilation (coming from outside the domain), are advected over the domain within the one day time span between assimilation and evaluation. We present also the RMSED (the red and blue curves); these results show no improvement in forecast except for the zone D where a slight reduction of 0.3 DU was found. Again, for the quadrants A and B where also the bias was not improved much, this can be explained by advection of fresh air masses into the domain during the one day time span between assimilation and evaluation. For quadrant D, it is likely that assimilated information is not transported faithfully to the correct model grid cell one day later, thus RMSED does not improve much, unlike the bias. To test this hypothesis, a scale analysis of the model improvement due to assimilation would need to be conducted, but this is beyond the scope of our paper.

We next evaluate the effect of the assimilation on the vertical ozone distribution. For this, we first analyze the vertical distribution of corrections (with respect to the free run) in terms of 2-dimensional fields. Figure 8 presents the correction maps (ozone concentrations expressed in parts-per-billion i.e. ppb) performed at several altitudes (surface, 1, 5 and 9 km) for the same date (here for 17 July 2007) in order to verify the consistency of our corrections (with the Fig. 4). The comparison reveals the existence of a large-scale correc-

tion over the Eastern part of the domain extending from the Baltic Sea to Southern Italy, as already mentioned in Sect. 4.3 when we compared the 0–6 km columns for the same date (Fig. 4). We notice that the corrections performed at the surface are generally small and that the same structure, a little more pronounced, is observed at 1 km altitude. At 5 km, the negative corrections occur again over the eastern part of the model domain, but in addition also over a region reaching from Denmark to Spain. These zones are consistent with the differences between the ensemble mean and the IASI partial columns (see Fig. 4). When we reach the 9 km altitude, the corrections are smaller than at 5 km, but the structure of the field remains consistent with differences observed at other altitudes. As expected, this altitude dependence of the corrections is consistent with the vertical structure of the averaging kernel. As shown in Fig. 3a, the maximum sensitivity of IASI 0–6 km partial columns occurs between 4 and 5 km height. This information is used in the Kalman gain matrix (via the observation operator **H**, Eq. 2) which projects the innovations from the observation space into the model space. The negative innovations (for partial columns) lead to negative corrections at surface in the eastern part of the domain, but not for the western part. This is likely due to a larger sensitivity of IASI observations to ground for the region over Eastern Europe where surface temperatures (and thus the thermal contrast with the atmosphere) were high during this day.

Keeping these results in mind, we analyze the impact of the ozone correction on the vertical ozone profile. For this, we averaged for each quadrant of the domain and over the whole month of July the vertical profile for a run with assimilation, called EnSRF and a reference run without assimilation called Ref (shown in Fig. 9). First, we can notice a large difference between the vertical profiles simulated by the model for the four zones (Ref – blue curves on Fig. 9). Simulated ozone levels in the reference simulation are larger in the south-eastern quadrant of the domain especially in the lower free troposphere due to persistent anticyclonic conditions and transport of polluted air masses coming from Western Europe (Roelofs, 2003). We remark also the absence of a vertical gradient in the first two kilometres for the same region and more variability in the Eastern part of the domain. Second, as expected from the discussion of the corrections of the 0–6 km columns, corrections are always negative (Fig. 9) and they are strongest in the 2–6 km region, which corresponds to the largest IASI sensitivity (for the four regions the maximum AVK is reached at approximately 4–4.5 km altitude shown in Fig. 3a). Above about 10 km, the corrections become negligible. We can also see that the correction at the surface remains significant. This is an important result, because it implies that IASI observations can constrain surface ozone levels. In addition to direct correction due to analysis, vertical transport of assimilated information affects these profiles, in particular downward transport increases the differences at the surface level. The magnitude of the corrections, in the free troposhere and at the surface, is spatially
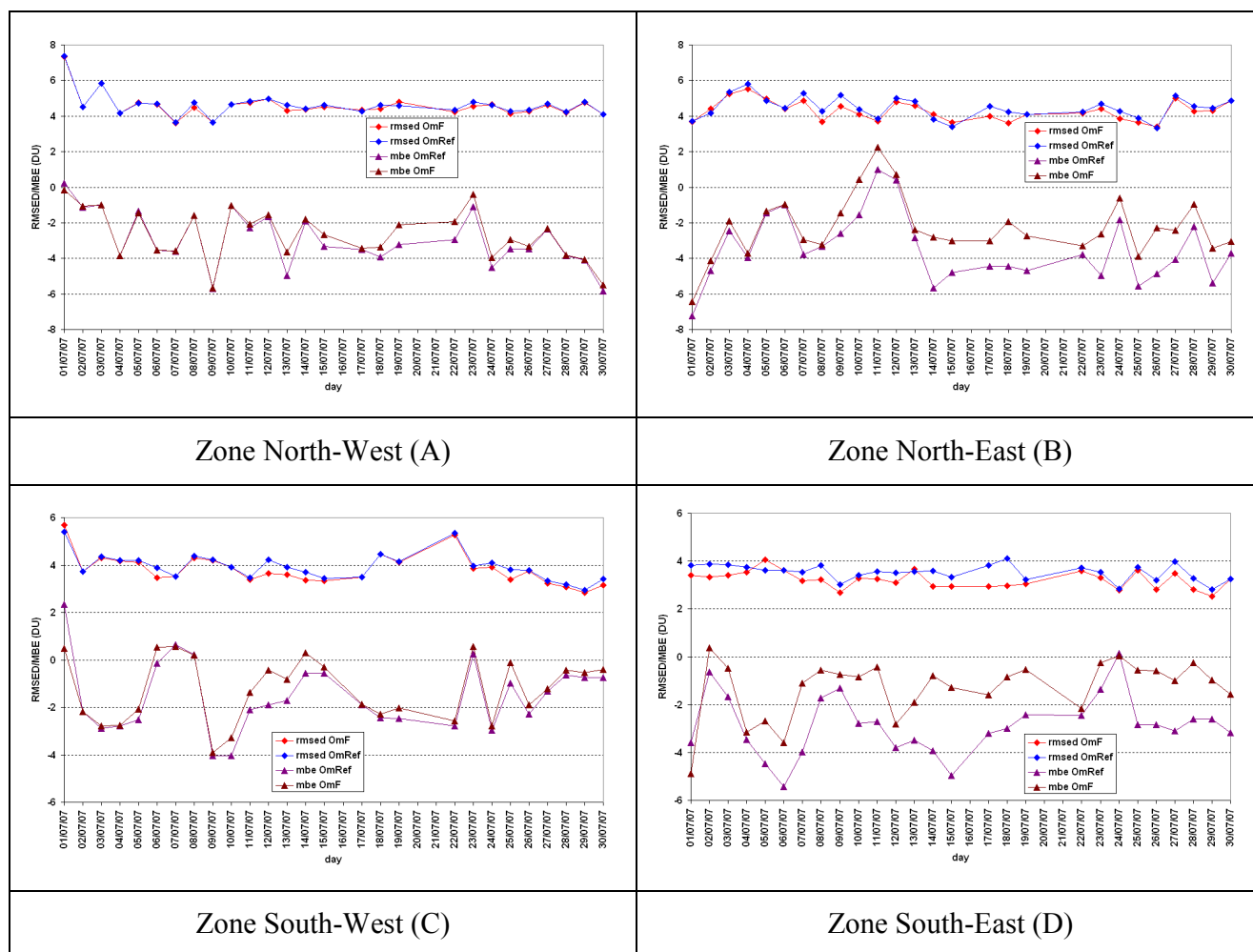
Fig. 7. Temporal evolution of the general statistics RMSED and MBE calculated for the differences (observations minus model/assimilation forecasts) for the four quadrants of the domain (defined in Fig. 2). The statistics were calculated for two types of differences called OmF (i.e. innovations) and OmRef (corresponding to the model reference run). The symbols are as follows: diamonds for the debiased RMSE "forecast" or "reference" (red or blue respectively); and triangles for the MBE "forecast" or "reference" (brown or violet respectively). The unit measure is DU.

modulated: in the western part of the domain, the corrections are smaller than in the eastern part (as was already noted for the columns). At 4 km, the mean decrease was of 7.4 ppb for zone A, 11.4 ppb for zone B, 6.6 ppb for zone C and 12.4 ppb for zone D. At surface level, these values were still 2.5 ppb for zone A, 6 ppb for zone B, 4.6 ppb for zone C and 7 ppb for zone D. Thus, for the Eastern part of the domain (zones B and D), surface corrections are roughly about half of those at 4 km. This important impact of the assimilation of IASI observations on surface ozone will be further analysed in Sect. 5.4 (evaluation with surface ozone data).

5.3 Improvement of vertical ozone distribution

In order to evaluate the impact of the assimilation procedure on the vertical ozone distribution, we use vertical ozone profiles from the MOZAIC project (Thouret et al., 1998) and vertical ozone soundings as independent validation data sets. The first data set used for validation is MOZAIC data from aircraft taking off or landing in Frankfurt (50° N, 8.7° E), where 83 profiles were available over July 2007. Ozone levels simulated by the CHIMERE model (with or without assimilation) considered along the MOZAIC flight track were calculated horizontally by bilinear interpolation (using the neighbouring grid cells) and vertically by linear interpolation every 100 m. As shown in Fig. 10, the analysis (red curve) produces a decrease in bias of about 6 ppb in the free
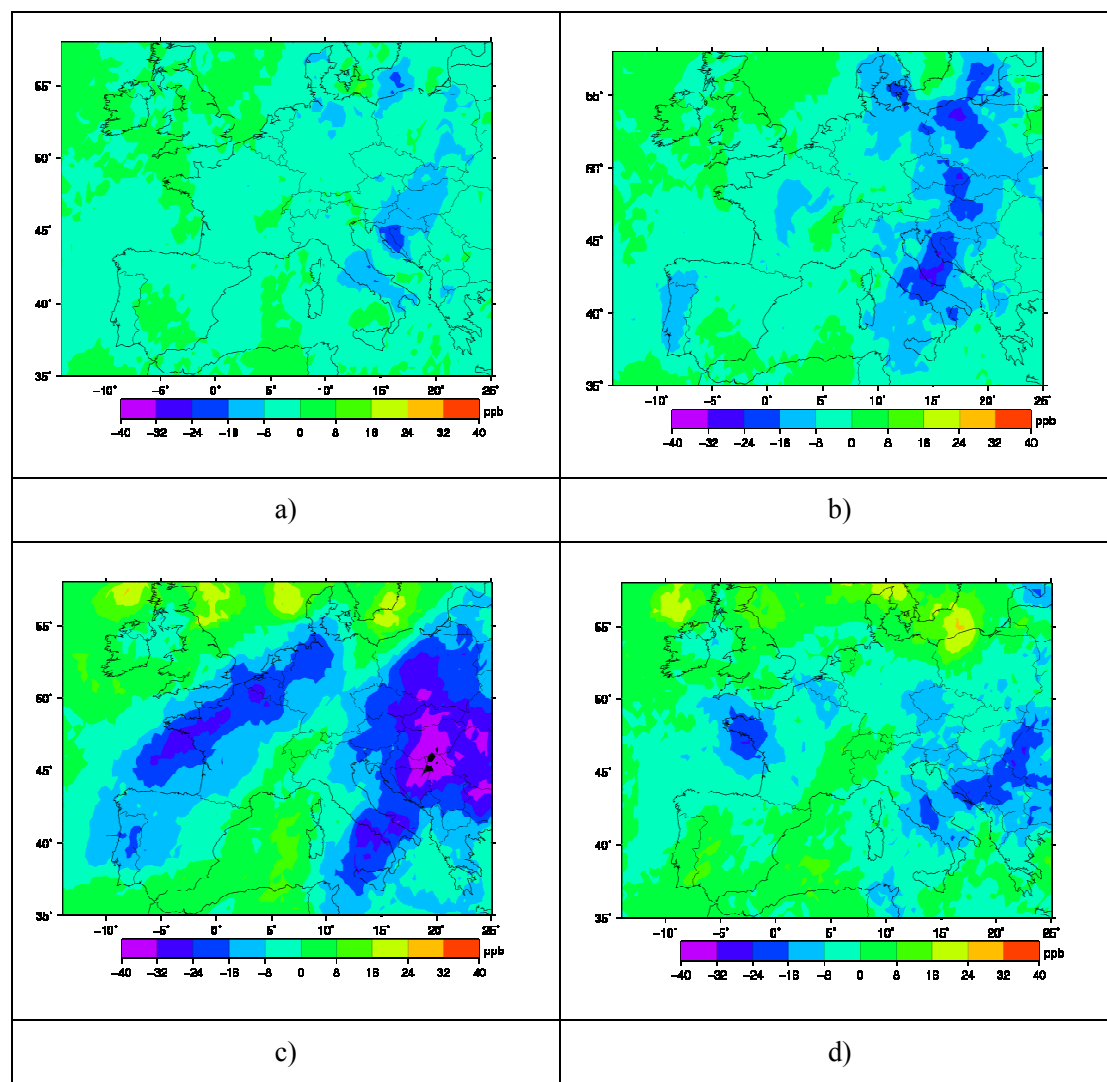
Fig. 8. 2-dimensional maps with corrections performed at various altitudes for 17 July 2007. The four contour plots are the corrections (concentration after assimilation minus concentration before assimilation) at **(a)** surface, **(b)** 1 km, **(c)** 5 km and **(d)** 9 km altitude. The ozone concentrations are expressed in ppb (color scale varies from violet representing relatively low values, to red representing relatively high values and the min/max values are −40/40 ppb).

troposphere when compared with observations (blue curve). Within a 2 to 5 km height range nearly all bias is removed. This altitude range corresponds to the maximum sensitivity of the IASI instrument. However, even at the surface, the bias is reduced, both due to a larger than zero sensitivity to surface ozone, and due to downward transport of assimilated information (as described in Foret et al., 2009). Therefore, we can conclude that, for this location, the constraints induced by IASI data (0–6 km columns) on the model outputs lead to a significant bias reduction in ozone estimate.

We selected also for validation the sonde measurements provided by the WOUDC (World Ozone and Ultraviolet radiation Data Centre). For comparison purposes, we interpo-

lated the soundings every 100 m and then we compared them to the model estimated profiles (reference run or analysed one) for which we performed, like for the MOZAIC flights, a bilinear interpolation in the horizontal plane and linear on the vertical every 100 m. For the five locations available: De Bilt, Lindenberg, Barajas, Payerne and Uccle, we analysed 46 profiles. The coordinates and the number of profiles available for each location are shown in the Table 2. The first remark is that the model profiles are smoother than the observed ones and this is due to the limited model vertical resolution. Second, at two stations, Barajas and Payerne (not shown), which are situated on the southern half of the domain, and where the model initially overestimates
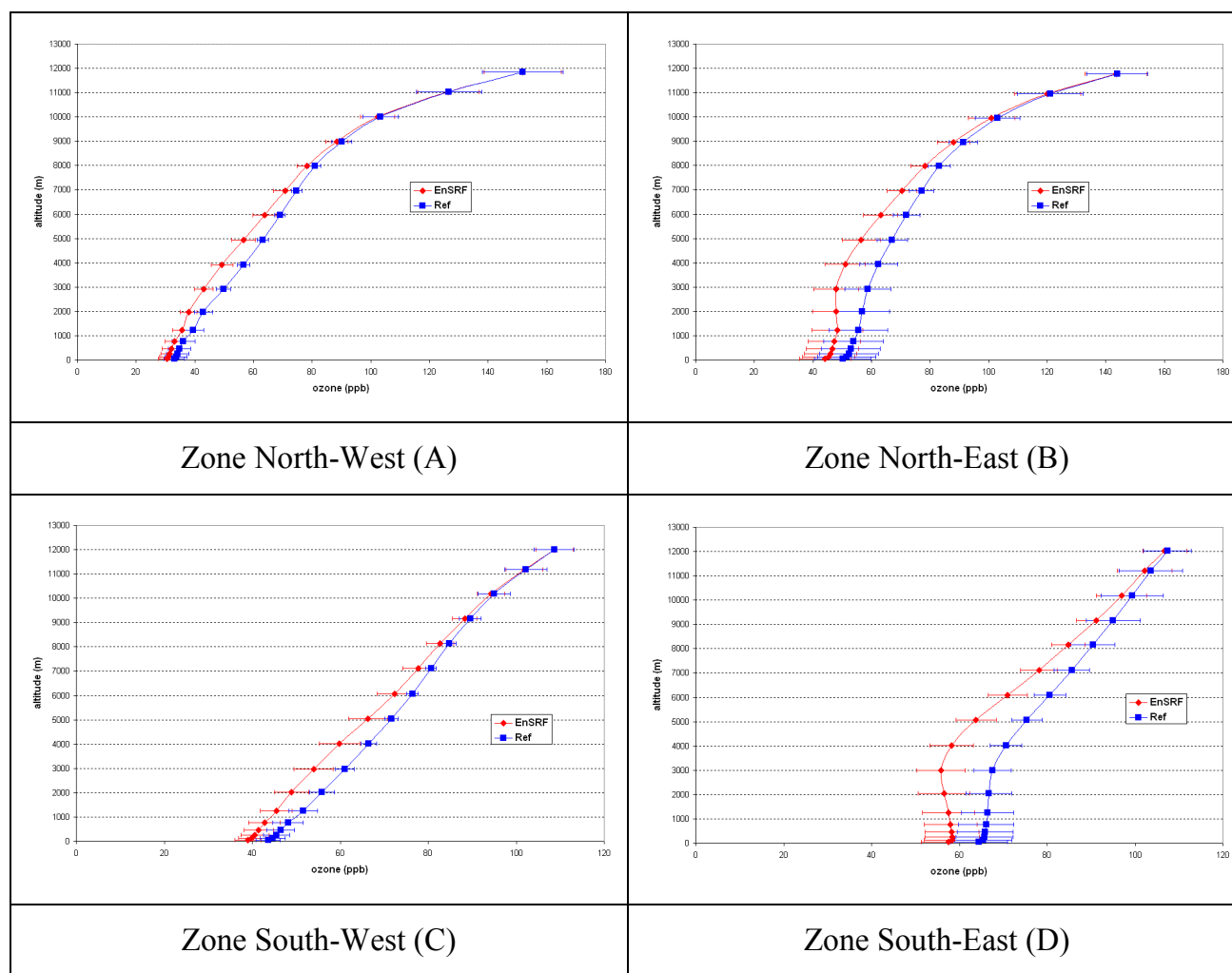
Fig. 9. Comparison between monthly mean ozone vertical profiles (surface to 12 km) for the four quadrants of the domain. The profiles were drawn from two distinctive runs: one with assimilation, called EnSRF (red curve) and the other one, free run "without assimilation" called Ref (blue curve); the averages were calculated first over the corresponding grid cells and then over the month. The error bars correspond to the monthly standard deviation of the vertical profiles. The unit measure for the vertical profiles is ppb.

Table 2. Geographical characteristics of the sounding sites as well as the number of profiles available for the study. The sondes used in this paper are taken from the World Ozone and Ultraviolet Data Centre (WOUDC) (http://www.woudc.org).

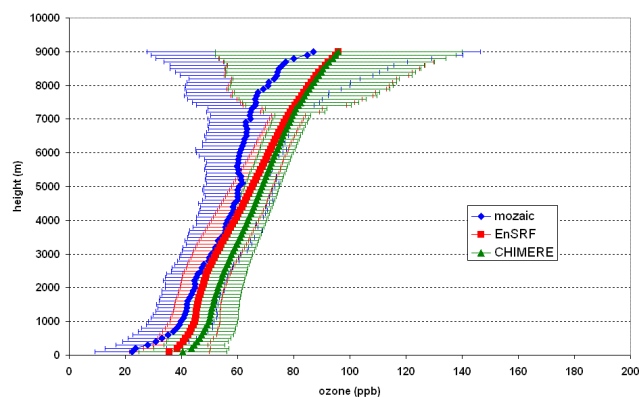| Name	Country	Time	Coordinates	Height (m)	# of profiles
Barajas	Spain	11:00 UT	40.47° N, 3.58° W	631	4
De Bilt	The Netherlands	11:00 UT	52.1° N, 5.18° E	4	4
Lindenberg	Germany	11:00 UT	52.21° N, 14.12° E	112	12
Payerne	Switzerland	11:00 UT	46.8° N, 6.95° E	491	14
Uccle	Belgium	11:00 UT	50.8° N, 4.35° E	85	12

Fig. 10. Comparison between the averaged vertical profile of MOZAIC aircraft (average calculated over 83 profiles) (blue curve) with the mean profile calculated over the analysed ensemble, called EnSRF, (red curve) and a mean profile drawn from a reference run "without assimilation", called CHIMERE (green curve). The error bars correspond to the standard deviation of the mean.

Fig. 11. Same comparison as in Fig. 10 using an average profile calculated for the 46 soundings available (for information about the five locations, see Table 2).

the tropospheric ozone, bias is improved, while for the three other sites, situated on the northern half of the domain, the analysis underestimates the ozone concentrations. However, for these latter soundings the reference run already exhibits a good agreement (not shown). Nevertheless, the lack of data at several sites does not allow for firm conclusions on the geographical distribution of the assimilation performance. If we average over the 46 profiles available, we can see in Fig. 11 the lack of improvement (compare red curve with the blue one) especially in the free troposphere.

To conclude, the independent data sets used for validating the assimilation system show different behaviour. We saw that, when comparing with soundings, a slight deterioration of bias is observed. The reference being already close to the measurements, the analysis performed with the EnSRF does not provide any gain (in the average) in the free troposphere. On the contrary, the comparison with MOZAIC data shows a strong improvement when assimilating IASI ozone partial columns and a bias reduction of about 6 ppb (12 %) in the free troposphere.

5.4 Improvement of surface ozone

In spite of a theoretical low sensitivity of IASI measurement to the surface (cf. AVK on Fig. 3a), it is important for air quality concerns to quantify the impact of IASI assimilation on the improvement of the surface ozone distribution, and this is the case of Europe, where it has been shown that downward transport of air masses could transfer a part of information carried by the IASI measurements to the surface (Foret et al., 2009). The potential improvement of surface ozone due to the assimilation of IASI data was tested against the observations provided by the Airbase (the European Air quality dataBase) network. For 2007, we have 703 ground-based
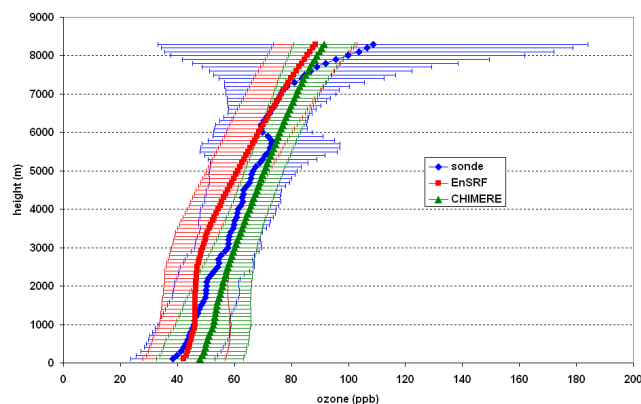
stations available. These stations were classified following Flemming et al. (2005) in four groups. The classification is based on the medians of daily average concentration, relative daily variation and altitude. We identified four classes; their sizes and characteristics are listed in Table 3. The most represented class contains stations which exhibit high diurnal variability (called peri-urban, P) and the second class in order of size is represented by the rural stations (R). The third one in order of size groups the stations which exhibit low variability (M) and finally, the fourth one is represented by the ground stations situated at altitude (above 900 m). In Table 4 we present the monthly morning and afternoon (09:00 a.m. and 03:00 p.m. UT) averages of the RMSE/bias (of assimilation run and the reference run against measurements) at the surface in absolute values (ppb) for all the stations. 09:00 a.m. corresponds to the hour of IASI overpass, 03:00 p.m. is close to the afternoon surface ozone maximum. As we can see, for all the groups we obtain a slight decrease of the RMSE/bias of the analysis compared to reference at 09:00 a.m., except for the bias at 03:00 p.m., where the decrease is substantial. Thus information introduced into the system in the morning affects surface ozone especially in the afternoon, likely due to rapid mixing within the convective boundary layer, and therefore significantly reduces the bias in the afternoon. We also evaluated the number of stations which improve their RMSE (more than 95 %, at 09:00 a.m. and 86 % at 03:00 p.m. see Table 5) and the number of stations which improve their bias (98 % in the morning and 89 % in the afternoon as shown in Table 6). As for the absolute values, we obtained an averaged reduction of the root mean square error of about 3 ppb at 09:00 a.m. and 2 ppb at 03:00 p.m., while for the bias the mean reduction is 3.9 ppb at 09:00 a.m. and 4.15 ppb at 03:00 p.m. To conclude, for the two selected times, the filter shows RMSE/bias reductions, a sign that the IASI tropospheric columns constitute an efficient direct or indirect constraint to correct the simulated surface ozone concentrations.

Table 3. Classification of ground-based stations (function of diurnal variability): 4 classes with the number of sites available (following Flemming, 2005). The fourth class (ALT) is special: it groups together all the ground stations situated at altitude (above 900 m).

TYPE	VARIABILITY	#
M	low	135
P	high	266
R	medium	222
ALT	–	80

Table 4. Monthly averaged statistics RMSE and bias (in ppb) when comparing a run with assimilation (EnSRF) and a reference one (Ref) against the measurements provided by the ground stations, differentiated for each type of station (as indicated by Table 3) at 09:00 a.m. and 03:00 p.m.

PPB	RMSE EnSRF	RMSE Ref	bias EnSRF	bias Ref
M 09:00 a.m.	10.60	13.11	−6.59	−10.46
M 03:00 p.m.	10.02	12.01	−3.70	−8.66
P 09:00 a.m.	13.12	16.12	−9.42	−13.70
P 03:00 p.m.	10.55	13.00	−3.74	−9.51
R 09:00 a.m.	11.43	14.11	−7.75	−11.50
R 03:00 p.m.	10.29	12.42	−3.58	−8.73
ALT 09:00 a.m.	15.89	19.23	−11.89	−16.70
ALT 03:00 p.m.	11.08	12.86	−1.36	−7.51

In these spatially averaged results, large spatial differences occur. The difference between RMSE after and before assimilation for all surface stations (in ppb) is presented in Fig. 12. The spatial pattern of corrections shows a clear west-east gradient, better performances being observed in the east, and especially over Italy, with RMSE reductions of more than 5 ppb. It should be noted that the type of stations (represented by different symbols cf. Fig. 13) does not impact the result. This may be related to the fact that assimilated information is more efficiently transported from the lower free troposphere to the surface over the south-eastern domain part where subsiding anticyclonic conditions are generally observed (Foret et al., 2009). It may also be related to the fact that the bias in the reference model is especially high over these regions and the mean DOF for the same region (independent information) is also higher than elsewhere (see the Fig. 2).

To conclude, at the surface, almost all the AIRBASE stations tested exhibit significant improvements in terms of root mean square error and bias, a sign that the IASI data can constitute an important indirect constraint to correct simulated surface ozone concentrations even if the instrument shows low sensitivity at the ground.

Table 5. Number of stations (in %) for which the root mean square error (RMSE) after analysis is reduced, the RMSE calculated as in Table 4. The total number of sites for each type is indicated in the first line.

nr %	M (135)	P (266)	R (222)	ALT (80)
09:00 a.m.	94 %	98 %	95 %	94 %
03:00 p.m.	86 %	90 %	90 %	76 %

Table 6. Number of stations (in %) for which the mean bias after analysis is reduced, the mean bias calculated as in Table 4. The total number of sites for each type is indicated in the first line.

nr %	M (135)	P (266)	R (222)	ALT (80)
09:00 a.m.	97 %	99 %	99 %	96 %
03:00 p.m.	90 %	93 %	94 %	81 %

6 Conclusions

In this study, an assimilation experiment based on the EnSRF has been performed in order to investigate and evaluate the potential of the IASI ozone partial columns to better represent the 3-dimensional distribution of the ozone fields over Europe. To our knowledge, this is the first assimilation experiment conducted using IASI data in order to improve the estimate of the ozone fields, especially in the free and lowermost troposphere.

We have quantified the impact of assimilation using self-consistency tests and evaluation against all available independent data for the period of the study. It has been shown that DA of 0–6 km ozone columns from IASI is improving simulated ozone fields in the middle and lower troposphere. These corrections with respect to the free run were more pronounced in the South-Eastern part of the domain, up to 8–9 ppb in the free troposphere, likely due to (1) the transport of corrections from the Western part of the domain associated to the general circulation pattern (i.e. dominant western circulation) and (2) a longer residence time of air masses in this area due to persistent anticyclonic conditions over the Mediterranean basin (Foret et al., 2009). It may also be related to the fact that the bias in the reference model was especially high over these regions and the mean DOF for the same region (independent information) is also higher than elsewhere. Surprisingly, the impact of IASI columns assimilation on surface ozone is quite large which is not obvious a priori considering the low sensitivity of TIR (Thermal InfraRed) measurements at the surface. The reduction of the errors at the surface is more important in the afternoon than in the morning, pointing out the fact that the ozone information introduced into the system needs some time to be
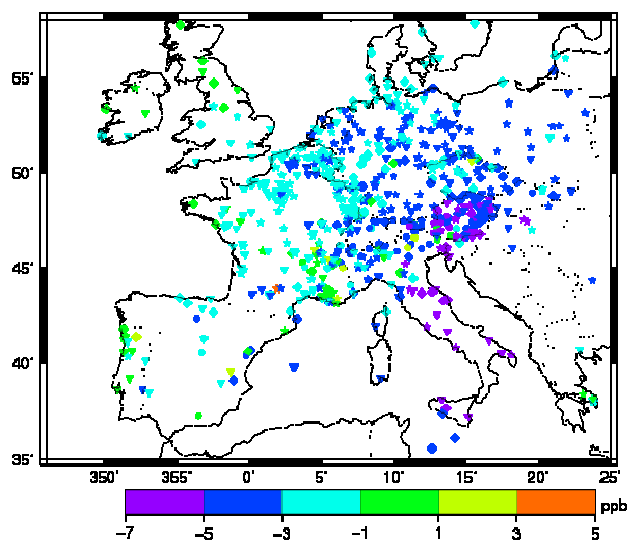
Fig. 12. Difference RMSE_EnSRF minus RMSE_Ref calculated using two simulations: one with assimilation, called EnSRF and the second one without assimilation, called Ref (against measurements) at 09:00 a.m. calculated over the month July 2007, for each of 703 ground-based stations from the AIRBASE network. Different symbols were used for each class type: circles for the ALT stations, diamonds for the M, stars for the P and triangles for the R (see the four types of stations in Table 3). The color scale varies from violet representing relatively low values, to red representing relatively high values and min/max values from −7 ppb to 5 ppb.

transported downward. Corrections are stronger again in the south-eastern part of the European domain. Even if the major aim of this paper was not a geophysical analysis of processes determining the tropospheric ozone distribution over Europe, our results confirm the importance of the downward transport (by subsidence, and mixing in the boundary layer) of tropospheric ozone over the Eastern part of the Mediterranean basin.

Both error reductions by IASI data assimilation in free tropospheric and surface ozone over Europe and in particular the Mediterranean are important results in the perspective of monitoring/understanding of trends and variability of Mediterranean tropospheric ozone concentrations as well as for air quality concerns. Especially, in the context of GMES (Global Monitoring for Environment and Security) and the set up of the future monitoring systems dedicated to AQ, these results show the importance of using the IASI measurements to obtain more accurate ozone analysis even at the surface. Improvement of forecast has still to be demonstrated. The use of these data for operational and assimilation/ forecast context shall be investigated in the framework of the GMES/MACC project.

To further explore the potential of IASI data, many improvements can be imagined and will be pursued in future studies. First, an expected improvement would be the construction and use of a physically sound ensemble by perturbing various parameters of the model (emissions, boundary conditions, meteorology, etc.). Second, the coupling of IASI data assimilation with that of in situ observations, i.e. combining complementary information sources, is expected to improve the system performance. Finally, longer assimilation periods would be beneficial in order to sample major ozone pollution episodes.

The publication of this article is financed by CNRS-INSU.

References

Anderson, J. L.: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter, Physica D, 230, 99–111, 2007.

Beekmann M. and Derognat C.: Monte Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements from the atmospheric pollution over the Paris area (ESQUIF) campaign, J. Geophys. Res., 108, 8559, doi:10.1029/2003JD003391, 2003.

Bessagnet, B., Menut, L., Aymoz, G., Chepfer, H., and Vautard, R.: Modeling dust emissions and transport within Europe: the Ukraine March 2007 event, J. Geophys. Res., 113, D15202, doi:10.1029/2007JD009541, 2008.

Blond, N. and Vautard R.: Three-dimensional ozone analyses and their use for short-term ozone forecasts, J. Geophys. Res., 109, D17303, doi:10.1029/2004JD004515, 2004.

Boynard, A., Beekmann, M., Foret, G., Ung, A., Szopa, S., Schmechtig, C., and Coman, A.: Assessment of regional ozone model uncertainty with a modelling ensemble using an explicit error representation, Atmos. Environ., 45, 784–793, 2011.

Bouttier, F. and Courtier, P.: Data assimilation concepts and methods, Meteorological Training Course Lecture Series, ECMWF, 1999.

Buehner, M.: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting, Q. J. Roy. Meteor. Soc., 131, 1013–1043, 2005.

Burgers, G., Van Leeuwen, P. J., and Evensen, G.: Analysis Scheme in the Ensemble Kalman Filter, Mon. Weather Rev., 126, 1719–1724, 1998.

Chai, T., Carmichael, G. R., Tang, Y., Sandu, A., Hardesty, M., Pilewskie, P., Whitlow, S., Browell, E. V., Avery, M. A., Nédélec, P., Mererill, J. T., Thompson, A. M., and Williams, E.: Four Dimensional Data Assimilation Experiments with ICARTT (International Consortium for Atmospheric Research on transport and Transformation), J. Geophys. Res., 112, D12S15, doi:10.1029/2006JD007763, 2007.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbit, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M.: Hydrological data assimilation with the Ensemble Kalman Filter; Use of streamflow data to update the states in a distributed hydrological model, Adv. Water Res., 31, 1309–1324, 2008.

Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., Hurtmans, D., Pommier, M., Razavi, A., Turquety, S., Wespes, C., and Coheur, P.-F.: Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder, Atmos. Chem. Phys., 9, 6041–6054, doi:10.5194/acp-9-6041-2009, 2009.

Coll, I., Pinceloup, S., Perros, P. E., Laverdet, G., and Le Bras, G.: 3D analysis of high ozone production rates observed during the ESCOMPTE campaign, Atmos. Res., 74, 477–505, 2005.

Constantinescu, E. M., Chai, T., Sandu, A., and Carmichael, G. R.: Autoregressive Models of Background Errors for Chemical Data Assimilation, J. Geophys. Res., 112, D12309, doi:10.1029/2006JD008103, 2007.

Deguillaume, L., Beekmann, M., and Derognat, C.: Uncertainty evaluation of ozone production and its sensitivity to emission changes over the Ile-de-France region during summer periods, J. Geophys. Res., 113, D02304, doi:10.1029/2007JD009081, 2008.

Dethof, A. and Hólm, E.: Ozone in era40: 1991–1996, Technical Memorandum 377, ECMWF, Reading, UK, 2002.

Dufour, G., Eremenko, M., Orphal, J., and Flaud, J.-M.: IASI observations of seasonal and day-to-day variations of tropospheric ozone over three highly populated areas of China: Beijing, Shanghai, and Hong Kong, Atmos. Chem. Phys., 10, 3787–3801, doi:10.5194/acp-10-3787-2010, 2010.

Elbern, H. and Schmidt, H.: Ozone episode analysis by fourdimensional variational chemistry data assimilation, J. Geophys. Res., 106, 3569–3590, 2001.

Eremenko, M., Dufour, G., Foret, G., Keim, C., Orphal, J., Beekmann, M., Bergametti, G., and Flaud, J.-M.: Tropospheric ozone distributions over Europe during the heat wave in July 2007 observed from infrared nadir spectra recorded by IASI, Geophys. Res. Lett., 35, L18805, doi:10.1029/2008GL034803, 2008.

Eskes, H. J., Van Velthoven, P. F. J., Valks, P., and Kelder, H. M.: Assimilation of GOME total ozone satellite observations in a three-dimensional tracer transport model, Q. J. Roy. Meteorol. Soc., 129, 1663–1681, 2003.

Evensen G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics, J. Geophys. Res., 99, 143–162, 1994.

Evensen, G.: The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation, Ocean Dynam., 53, 343–367, 2003.

Evensen, G.: Sampling strategies and square root analysis schemes for the EnKF, Ocean Dynam., 54, 539–560, doi:10.1007/s10236-004-0099-2, 2004.

Evensen, G.: Data assimilation: The Ensemble Kalman Filter, Springer-Verlag Berlin Heidelberg, 2007.

Fiore, A. M., Jacob, D. J., Bey, I., Yantosca, M. R., Field B. D., and Fusco, A. C.: Background ozone over the United States in summer: Origin, trend, and contribution to pollution episodes, J. Geophys. Res., 107, 4275, doi:10.1029/2001JD000982, 2002.

Fischer, H., Birk, M., Blom, C., Carli, B., Carlotti, M., von Clarmann, T., Delbouille, L., Dudhia, A., Ehhalt, D., Endemann, M., Flaud, J. M., Gessner, R., Kleinert, A., Koopman, R., Langen, J., López-Puertas, M., Mosner, P., Nett, H., Oelhaf, H., Perron, G., Remedios, J., Ridolfi, M., Stiller, G., and Zander, R.: MIPAS: an instrument for atmospheric and climate research, Atmos. Chem. Phys., 8, 2151–2188, doi:10.5194/acp-8-2151-2008, 2008.

Flemming, J., Stern, R., and Yamartino, R. J.: A new air quality regime classification scheme for O_3, NO_2, SO_2 and PM_{10} observation sites, Atmos. Environ., 39, 6121–6129, 2005.

Foret, G., Hamaoui, L., Schmechtig, C., Eremenko, M., Keim, C., Dufour, G., Boynard, A., Coman, A., Ung, A., and Beekmann, M.: Evaluating the potential of IASI ozone observations to constrain simulated surface ozone concentrations, Atmos. Chem. Phys., 9, 8479–8491, doi:10.5194/acp-9-8479-2009, 2009.

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D. W., Haywood, J., Lean, J., Lowe, D. C., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., and Van Dorland, R.: Changes in Atmospheric Constituents and in Radiative Forcing. In: Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Galmarini, S., Bianconi, R., Klug, W., Mikkelsen, T., Addis, R., Andronopoulos, S., Astrup, P., Baklanov, A., Bartniki, J., Bartzis, J. C., Bellasio, R., Bompay, F., Buckley, R., Bouzom, M., Champion, H., D'Amours, R., Davakis, E., Eleveld, H., Geertsema, G. T., Glaab, H., Kollaxo, M., Ilvonenu, M., Manningm, A., Pechingerp, U., Perssono, C., Polreichp, E., Potemskiq, S., Prodanovar, M., Saltbonesh, J., Slaperj, H., Sofievt, M. A., Syrakovr, D., Sørenseng, J. H., Van der Auweras, L., Valkamat, I., and Zelaznyq, R.: Ensemble dispersion forecasting – part I: Concept,

approach and indicators, Atmos. Environ., 38, 4607–4617, 2004.

Hanea, G. R., Velders, G. J. M., and Heemink, A.: Data assimilation of ground-level ozone in Europe with a Kalman filter and chemistry transport model, J. Geophys. Res., 109, D10302, doi:10.1029/2003JD004283, 2004.

Hauglustaine, D. A., Hourdin, F., Walters, S., Jourdain, L., Filiberti, M.-A., Larmarque, J.-F., and Holland, E. A.: Interactive chemistry in the Laboratoire de Météorologie Dynamique general circulation model: description and background tropospheric chemistry evaluation, J. Geophys. Res., 109, D04314, doi:10.1029/2003JD003957, 2004.

Hodzic, A., Vautard, R., Chepfer, H., Goloub, P., Menut, L., Chazette, P., Deuzé, J. L., Apituley, A., and Couvert, P.: Evolution of aerosol optical thickness over Europe during the August 2003 heat wave as seen from CHIMERE model simulations and POLDER data, Atmos. Chem. Phys., 6, 1853–1864, doi:10.5194/acp-6-1853-2006, 2006.

Honoré, C., Rouil, L., Vautard, R., Beekmann, M., Bessagnet, B., Dufour, A., Elichegaray, C., Flaud, J.-M., Malherbe, L., Meleux, F., Menut, L., Martin, D., Peuch, A., Peuch, V.-H., and Poisson, N.: Predictability of European air quality: Assessment of 3 years of operational forecasts and analyses by the PREV'AIR system, J. Geophys. Res., 113, D04301, doi:10.1029/2007JD008761, 2008.

Hollingsworth, A., Engelen, R. J., Textor, C., Benedetti, A., Boucher, O., Chevallier, F., Dethof, A., Elbern, H., Eskes, H., Flemming, J., Granier, C., Kaiser, J. W., Morcrette, J.-J., Rayner, P., Peuch, V.-H., Rouil, L., Schultz, M. G., Simmons, A. J., and The GEMS Consortium: Toward a Monitoring and Forecasting System For Atmospheric Composition: The GEMS Project, B. Am. Meteorol. Soc., 89, 1147–1164, 2008.

Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an Ensemble Kalman Filter technique, Mon. Weather Rev., 126, 796–811, 1998.

Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev., 129, 123–137, 2001.

Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L. and Hansen, B.: Atmospheric data assimilation with the ensemble Kalman filter: Results with real observations, Mon. Weather Rev., 133, 604–620, 2005.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D 230, 112–126, 2007.

Keim, C., Eremenko, M., Orphal, J., Dufour, G., Flaud, J.-M., Höpfner, M., Boynard, A., Clerbaux, C., Payan, S., Coheur, P.-F., Hurtmans, D., Claude, H., Dier, H., Johnson, B., Kelder, H., Kivi, R., Koide, T., López Bartolomé, M., Lambkin, K., Moore, D., Schmidlin, F. J., and Stübi, R.: Tropospheric ozone from IASI: comparison of different inversion algorithms and validation with ozone sondes in the northern middle latitudes, Atmos. Chem. Phys., 9, 9329–9347, doi:10.5194/acp-9-9329-2009, 2009.

Keppenne, C. and Rienecker, M.: Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter, J. Mar. Syst., 40–41, 363–380, 2003.

Khattatov, B., Lamarque, J.-F., Lyjak, L., Menard, R., Levelt, P., Tie, X., Brasseur, G., and Gille, J. C.: Assimilation of satellite observation of long-lived chemical species in global chemistry transport models, J. Geophys. Res., 105, 29135–29144, 2000.

Lahoz, W. A., Errera, Q., Swinbank, R., and Fonteyn, D.: Data assimilation of stratospheric constituents: a review, Atmos. Chem. Phys., 7, 5745–5773, doi:10.5194/acp-7-5745-2007, 2007.

Lamarque, J.-F., Kattatov, B. V., and Gille, J. C.: Constraining tropospheric ozone column through data assimilation, J. Geophys. Res, 107, 4651, doi:10.1029/2001JD001249, 2002.

Mallet, V. and Sportisse, B.: Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling, J. Geophys. Res., 111, D01302, doi:10.1029/2005JD006149, 2006.

Massart, S., Clerbaux, C., Cariolle, D., Piacentini, A., Turquety, S., and Hadji-Lazaro, J.: First steps towards the assimilation of IASI ozone data into the MOCAGE-PALM system, Atmos. Chem. Phys., 9, 5073–5091, doi:10.5194/acp-9-5073-2009, 2009.

Maybeck, P.: Stochastic models, estimation, and control, Academic Press, London, 1979.

McPeters, R. D., Labow, G. J., and Logan, J. A.: Ozone climatological profiles for satellite retrieval algorithms, J. Geophys. Res., 112, D05308, doi:10.1029/2005JD006823, 2007.

Milewski, T. and Bourqui, M. S.: Assimilation of stratospheric temperature and ozone with an ensemble Kalman filter in a chemistry-climate model, Mon. Weather Rev., 139, 3389–3404, doi:10.1175/2011MWR3540.1, 2011.

Parrington, M., Jones, D. B. A., Bowman, K. W., Horowitz, L. W., Thompson, A. M., Tarasick, D. W., and Witte, J. C.: Estimating the summertime tropospheric ozone distribution over North America through assimilation of observations from the Tropospheric Emission Spectrometer, J. Geophys. Res., 113, D18307, doi:10.1029/2007JD009341, 2008.

Parrington, M., D. B. A. Jones, K. W. Bowman, A. M. Thompson, D. W. Tarasick, J. Merrill, S. J. Oltmans, T. Leblanc, J. C. Witte, and D. B. Millet, Impact of the assimilation of ozone from the Tropospheric Emission Spectrometer on surface ozone across North America, Geosphys. Res. Lett., 36, L04802, doi:10.1029/2008GL036935, 2009.

Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: Hydrologic Data Assimilation with the Ensemble Kalman Filter, Mon. Weather Rev., 130, 103–114, 2002.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific, Series on Atmospheric, Ocean. Planet. Phys., 2, Hackensack, NJ, 2000.

Roelofs, G. J., Scheeren, H. A., Heland, J., Ziereis, H., and Lelieveld, J.: A model study of ozone in the eastern Mediterranean free troposphere during MINOS (August 2001), Atmos. Chem. Phys., 3, 1199–1210, doi:10.5194/acp-3-1199-2003, 2003.

Rouil L., Honoré C., Vautard R., Beekmann M., Bessagnet B., Malherbe L., Meleux F., Dufour A., Elichegaray C., Flaud J.-M., Menut L., Martin D., Peuch A., Peuch V.-H., and Poisson N.: PREV'AIR: an operational forecasting and mapping system for air quality in Europe, B. Am. Meteor. Soc., 90, 73–83, doi:10.1175/2008BAMS2390.1, 2009.

Sakov, P. and Bertino, L.: Relation between two common localization methods for the EnKF, Comput. Geosci., 15, 225–236, doi:10.1007/s10596-010-9202-6, 2010.

Sakov, P. and Oke, P. R.: Implications of the form of the ensemble transformation in the ensemble square root filters, Mon. Weather Rev., 136, 1042–1053, 2008.

Segers, A. J., Eskes, H. J., Van der A, R. J., Van Oss, R. F., and Van Velthoven, P. F. J.: Assimilation of GOME ozone profiles and a global chmistry-transport model, using a Kalman Filter with anisotropic covariance, Q. J. Roy. Meteorol. Soc., 130, 477–502, 2004.

Stiller, G. P. (Ed.) with contributions from v. Clarmann, T., Dudhia, A., Echle, G., Funke, B., Glatthor, N., Hase, F., Höpfner, M., Kellmann, S., Kemnitzer, H., Kuntz, M., Linden, A., Linder, M., Stiller, G. P., and Zorn, S.: The Karlsruhe Optimized and Precise Radiative Transfer Algorithm (KOPRA), vol. FZKA 6487 of Wissenschaftlich Berichte, Forschungszentrum Karlsruhe, Germany, 672 pp., available online at: http://www-imk.fzk.de/asf/ame/publications/kopra_docu/, 2000.

Szunyogh, I., Kostelich, E. J., Gyarmati, G., Patil, D. J., Hunt, B. R., Kalnay, E., Ott, E., and Yorke, J. A.: Assessing a local ensemble Kalman filter: perfect model experiments with the National Centers for Environmental Prediction global model, Tellus, 57A, 528–545, 2005.

Talagrand, O.: Bayesian estimation. Optimal interpolation. Statistical linear estimation, in: Data assimilation for the Earth System, NATO ASI Series, edited by: Swinbank, R., Shutyaev, V., and Lahoz, W. A., Kluwer, 21–35, 2003.

Thouret, V., Marenco, A., Logan, J. A., Nédélec, P., and Grouhel, C.: Comparisons of ozone measurements from the MOZAIC airborne program and the ozone sounding network at eight locations, J. Geophys. Res., 103, 25695–25720, 1998.

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteorol. Soc., 131, 2961–3012, 2005.

Vautard, R., Honore, C., Beekmann, M., and Rouil, L.: Simulation of ozone during the August 2003 heat wave and emission control scenarios, Atmos. Environ., 39, 2957–2967, 2005.

Vestreng, V., Breivik, K., Adams, M., Wagener, A., Goodwin, J., Rozovskkaya, O., and Pacyna, J. M.: Inventory Review 2005, Emission Data reported to LRTAP Convention and NEC Directive, Initial review of HMs and POPs, Technical report MSC-W 1/2005, ISSN 0804-2446, 2005.

Whitaker, J. S. and Hamill, T. M.: Ensemble Data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913–1924, 2002.

WHO: Health aspects of air pollution with Particulate Matter, Ozone and Nitrogen Dioxide, Report on a WHO Working Group, Bonn, Germany, 30–45, 2003.

Wu, L., Mallet, V., Bocquet, M., and Sportisse, B.: A comparison study of data assimilation algorithms for ozone forecast, J. Geophys. Res., 113, D20310, doi:10.1029/2008JD009991, 2008.

Zyryanov, D., Foret, G., Eremenko, M., Beekmann, M., Cammas, J.-P., D'Isidoro, M., Elbern, H., Flemming, J., Friese, E., Kioutsioutkis, I., Maurizi, A., Melas, D., Meleux, F., Menut, L., Moinat, P., Peuch, V.-H., Poupkou, A., Razinger, M., Schultz, M., Stein, O., Suttie, A. M., Valdebenito, A., Zerefos, C., Dufour, G., Bergametti, G., and Flaud, J.-M.: 3-D evaluation of tropospheric ozone simulations by an ensemble of regional Chemistry Transport Model, Atmos. Chem. Phys. Discuss., 11, 28797–28849, doi:10.5194/acpd-11-28797-2011, 2011.