Response to referee comments

We thank the reviewers for the valuable feedback. Responding to the general comments, we have introduced the following modifications to the revised paper:

- We have extended the discussion on analysis biases and their dependence on assimilation setup
- We have included a discussion about the interference effects in NO2 measurements with Molybdenum converters
- We have the clarified the system description regarding the separate assimilation of NO2 and O3.

We address the specific comments below (the reviewers' comments are shown in italic).

Responses to referee #1

First concerning O3 assimilation results, i do not well understand how you explain the increase of the bias when using the tuned B matrix.

The increase of bias is explained by the differences in background and observation error standard deviations. For the first-guess B and R matrices, σ_b is assumed to be about twice σ_{obs}. With the adjusted covariance matrices, the ratio is time-dependent but the diagnosed observation errors are generally larger than the background errors.

Due to the higher ratio between background/observation errors, the first-guess assimilation setup results in much stronger increments in the model fields than the final setup. Because of a negative bias in the free-running model, the increments are on average positive. In the final setup, the larger diagnosed observation errors result in more damped increments, and therefore smaller reduction of bias. However, the final setup results in analyses with better temporal correlation and RMSE, which in our view indicates that the analyses are more robust against observational (both instrument and representativeness) errors.

Reducing the negative bias would certainly be desirable. However, we believe that this is better addressed by improving the forecast model, and possibly, with a dedicated bias correction scheme.

The corresponding additional discussion is included in the paper (Discussion section).

Also, could you explain why results for EMEP stations are less good than for MACC stations.

For ozone, the free-running model performs better on the MACC stations, while for NO2 the situation is reversed. The most important difference between the MACC and EMEP station sets is the spatial coverage.

The MACC validation stations are concentrated in the Central Europe, while the EMEP network covers the computational domain more uniformly.

The densest coverage in the MACC dataset (and Airbase as whole) thus coincides with the area where regional air quality models often perform best; the SILAM forecast scores follow a similar spatial distribution as reported by eg. Vautard et al. (2009) for an ensemble of CTMs. Consequently, the average scores for the MACC stations are better than for the EMEP stations even for the free-running model.

In the assimilation runs, the performance difference in favour of the MACC validation stations is even larger, because the assimilation stations have similar spatial coverage as the MACC validation stations.

This discussion is also included in the paper.

Concerning NO2 assimilation results, it seems also that biases are not always reduced. How do you explain this? I understand that you select background rural stations for NO2 for representativeness issues but for these stations you can have measurements problems, indeed with most common devices (using molybden converters) your are measuring NO2 only but also other nitrogen oxides such like PAN, HNO3, HONO. This issues have been raised by Dunlea et al (2007) and Steinbacher et al (2007). I think that you have to mention these aspects.

As indicated in Table 4, the free-running model has a bias of -1.18 μg/m3 for hourly values measured at MACC stations but +0.47 μg/m3 for EMEP stations. For the MACC stations, the assimilation reduces the bias to -0.38 μg/m3, while the bias at EMEP stations is increased to 0.99 μg/m3. Therefore, the positive bias on EMEP stations appears to be primarily a feature of the forward model.

We calculated the contribution of NOz (CB4 species PAN, HNO3, NO3, HONO and PNA) to NOy (NOx + NOz) from the free-running SILAM simulation. On a few stations in the EMEP subset, the contribution can be up about 50% on yearly mean level, however more typical range was 10-20%. We agree that the ambiguity of NO2 measurements introduces uncertainty to the analysis fields and complicates the model evaluation, and this is discussed in the revised manuscript.

The discussion is added to the paper.

Also have you checked the impact of assimilating NO2 on O3 ? You do not mention it.

We cannot assess the impact of NO2 assimilation on O3 with the current setup, because the NO2 and O3 are assimilated into separate runs with different chemical schemes. This is mentioned in Section 4.2, but we have clarified the issue in Section 2.3 in the revised version.

You are showing that the results of the assimilation on forecast do not last more than 24hours. Do you think that you could do better ? Do you need more data, different data (I'm thinking to satellite data for example) ?

Assimilation of satellite O3 data has been investigated by other modellers (Coman et al., 2012). Although they do not assess the impact on forecast performance, we could expect such data to be useful for improving the forecasts on areas not well covered by the in-situ networks.

However, on areas strongly affected by local emission forcing or chemical processes, any assimilation scheme based on adjusting the forecast initial condition is likely to become ineffective as the forecast length increases. One way to overcome this limitation is to extend the state vector with additional parameters, like emission fluxes, which we investigated in an earlier paper (Vira and Sofiev, 2012). This approach can improve forecasts on longer range, but this requires that the obtained a posteriori emission rates can be extrapolated to the forecast window, and that the assimilation scheme is able to correctly attribute the observed discrepancies to the uncertain parameters.

The discussion above has been introduced in the manuscript**.**

You show that assimilation improves the simulation of daily maxima for ozone and it is of importance for AQ control but do you have checked if you were improving the highest values of the distribution or values exceeding the regulation thresholds ?

Following the suggestion, we computed the hit rates (the number of correctly predicted exceedances divided by the number of observed exceedances) for the 180 μg/m3 threshold with and without assimilation. It turns out that assimilation (with the final B and R matrices) improves the hit rate, albeit only slightly: from 0.25 to 0.26 on average for rural MACC validation stations, and from 0.13 to 0.15 for EMEP stations. If the averaging is restricted to the stations with more than 10 exceedances during 2012, the values change from 0.32 to 0.36 for MACC and from 0.21 to 0.43 for EMEP stations.

The hit rates suffer from the negative bias present in the daily maximum values. Consequently the "first guess" assimilation setup has somewhat higher average hit rates (0.32 for MACC and 0.22 for EMEP stations).

We have added a short discussion on this topic**.**

To finish, just a short remark on the form. You are referring to figure 6 before figure 5, it is only a detail but you should invert the order of these two figures.

Done**.**

Responses to referee #2

Is the separate assimilation of ozone and NO2 an issue, technical and/or scientific?

There is no technical reason for not assimilating NO2 together with O3 into the CB4 simulation. However, our experience is that assimilation of NO2 can have negative impact on O3 predictions. The issue is scientifically interesting, but would require efforts beyond the scope of the current paper.

P5595 L7-8: Could you elaborate on the representativeness issue?

Urban and suburban areas typically have significant NOx sources whose variability is not resolved by the 0.25 degree grid. For this reason we expect the rural stations to represent more reliably the pollutant levels in the scales that are resolved by our model. The corresponding statement is included in the paper**.**

When is the iteration stopped? What criterion is followed?

We stopped the iteration when the RMSE at validation stations stopped improving. This does not imply actual convergence of the adjusted parameters (σ_{obs} and σ_{b}), however, we prefer this criterion in order to avoid overfitting the parameters to the calibration periods**.**

Maybe I am missing something, but it is not immediately clear to me what is the link with summertime. Could you please elaborate?

Wang et al. (2011) reported somewhat longer-lasting forecast improvement in winter conditions due to lower photochemical activity. Since our forecast experiment was set in July-August, we cannot comment on the forecast in winter conditions. We have rephrased the sentence as “at least under the photochemically active summertime conditions” to emphasize this**.**

Could you provide examples of these previous studies?

We have added references to the NO2 studies, and also added an emphasis that our results are for an episode and not the whole year 2012.

Fig. 3 caption: Identify the end points of the colour scale. For example, red/blue indicate relatively high/low concerntrations of ozone and NO2.

We have redrawn the colour bars. The colour scale of Fig. 3a has also been changed to better match the range of values**.**

Finally, we have introduced the technical and editorial remarks, and thank the reviewer especially for advice on English language**.**

References

Coman, A., Foret, G., Beekmann, M., Eremenko, M., Dufour, G., Gaubert, B., Ung, A., Schmechtig, C., Flaud, J.-M., Bergametti, G., 2012. Assimilation of IASI partial tropospheric columns with an Ensemble Kalman Filter over Europe. Atmos. Chem. Phys. 12, 2513–2532. doi:10.5194/acp-12-2513-2012

Wang, X., Mallet, V., Berroir, J., Herlin, I., 2011. Assimilation of OMI NO2 retrievals into a regional chemistry-transport model for improving air quality forecasts over Europe. Atmos. Environ. 45, 485–492. doi:10.1016/j.atmosenv.2010.09.028

Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P.J.H., Christensen, J.H., Cuvelier, C., Foltescu, V., Graff, a., 2009. Skill and uncertainty of a regional air quality model ensemble. Atmos. Environ. 43, 4822–4832. doi:10.1016/j.atmosenv.2008.09.083

Vira, J., Sofiev, M., 2012. On variational data assimilation for estimating the model initial conditions and emission fluxes for short-term forecasting of SOx concentrations. Atmos. Environ. 46, 318–328. doi:10.1016/j.atmosenv.2011.09.066

Changes introduced to the revised manuscript

Abstract

This paper describes assimilation of trace gas observations into the chemistry transport model SILAM (System for Integrated modeLling of Atmospheric coMposition) using the 3D-Var method. Assimilation results for the year 2012 are presented for the prominent photochemical pollutants ozone (O3) and nitrogen dioxide (NO2). Both species are covered by the Airbase observation database, which provides the observational dataset used in this study.

Attention is paid to the background and observation error covariance matrices which are obtained primarily by iterative application of a posteriori diagnostics. The diagnostics are computed separately for two months representing summer and winter conditions, and further disaggregated by time of day. This allows deriving background and observation error covariance definitions, which include both seasonal and diurnal variation. The consistency of the obtained covariance matrices is verified using χ^2 diagnostics.

The analysis scores are computed for a control set of observation stations withheld from assimilation. Compared to a free-running model simulation, the correlation coefficient for daily maximum values is improved from 0.8 to 0.9 for O3 and from 0.53 to 0.63 for NO2.

1 Introduction[JV1]

During the last 10-15 years, assimilating observations into atmospheric chemistry transport models has been studied with a range of computational methods and observational datasets. The interest has been driven by the success of advanced data assimilation methods in numerical weather prediction (Rabier, 2005), as well as by development of operational forecast systems for regional air quality (Kukkonen et al., 2012). Furthermore, the availability of remote sensing data on atmospheric composition has permitted construction of global analysis and forecasting systems such as those described by Benedetti et al. (2009) and Zhang et al. (2008). Assimilation of satellite observations into stratospheric chemistry models has been demonstrated eg. by Errera et al.(2008).

Data assimilation is ~~classically~~ defined (eg. Kalnay, 2003) as the numerical process of using model fields and observations to produce a physically and statistically consistent representation of the atmospheric state - often in order to initialize the subsequent forecast. The main techniques used in atmospheric models include the optimal interpolation (OI, Gandin 1963), variational methods (3D-Var and 4D-Var, Le Dimet and Talagrand, 1986; Lorenc, 1986), and the stochastic methods based on the Ensemble Kalman Filter (EnKF, Evensen, 2003, 1994). Each of the methods has been applied in air quality modelling. Statistical interpolation methods were used by Blond and Vautard (2004) for surface ozone analyses and by Tombette et al. (2009) for particulate matter. The EnKF method has been utilized by several authors (Constantinescu et al., 2007; Curier et al., 2012; Gaubert et al., 2014) especially for ozone modelling. The 3D-Var method has been applied in regional air quality models by Jaumouillé et al. (2012) and Schwartz et al. (2012), while the computationally more demanding 4D-Var method has been demonstrated by Elbern & Schmidt (2001) and Chai et al. (2007). Partly due to its significance in relation to health effects, the most commonly assimilated chemical component has been ozone

Performance of most data assimilation methods depends on correctly prescribed background error covariance matrices (BECM). This is particularly important for 3D-Va r, where the BECM is prescribed and fixed throughout the whole procedure, in contrast to the EnKF based assimilation methods, where the BECM is described by the ensemble of states, and to the 4D-Var method, where the BECM is prescribed but evolves implicitly within the assimilation window.

A range of methods of varying complexity have been employed to estimate the BECM in previous studies on chemical data assimilation. The “National Meteorological Centre” (NMC) method introduced by Parrish & Derber (1992) is based on using differences between forecasts with differing lead times as a proxy for the background error. Kahnert (2008), as well as Schwartz et al. (2012), applied the NMC method for estimating the BECM for assimilation of aerosol observations. Chai et al. (2007) based the BECM on a combination of

NMC method and the observational method of Hollingsworth & Lönnberg (1986). The observational method was used in assimilation of NO2 and O3 observations also by Kumar et al. (2012).

The BECM can also be estimated using ensemble modelling; this approach was taken by Massart et al. (2012) for global and by Jaumouillé et al. (2012) for regional ozone analyses. Finally, Desroziers et al. (2005) presented a set of diagnostics which can be used to adjust the background and observation error covariances. This method has been previously applied in chemical data assimilation for example by Schwinger and Elbern (2010) and Gaubert et al. (2014).

In contrast to short and medium range weather prediction, the influence of initial condition on an air quality forecast has been found to diminish as the forecast length increases. For ozone, Blond and Vautard, (2004) and Wu et al. (2008) found that the effect of the adjusted initial condition extended for up to 24 hours. Among other reactive gases, NO2 has been a subject for studies of Silver et al. (2013) and Wang et al. (2011). However, the shorter lifetime of NO2 limits the timescale for forecast improvements especially in summer conditions.

An approach for improving effectiveness of data assimilation for short-lived species is to extend the adjusted state vector with model parameters. Among the possible choices are emission and deposition rates (Bocquet, 2012; Curier et al., 2012; Elbern et al., 2007; Vira and Sofiev, 2012).

The aim of the current paper is to describe and evaluate a regional air quality analysis system based on assimilating hourly near-surface observations of NO2 and O3 into the SILAM chemistry transport model. The assimilation scheme was initially presented by Vira and Sofiev (2012); in the current study, the scheme is applied to photochemical pollutants and moreover, we discuss how its performance can be improved by introducing statistically consistent background and observation error matrices. The analysis fields are produced for the assimilated species at hourly frequency using the standard 3D-Var assimilation method (Lorenc, 1986). The diagnostics of Desroziers et al. (2005) are applied in this work for estimating the background and observation error standard deviations, in particular resolving their seasonal and diurnal variations. The evaluation is performed for year 2012 using stations withheld from assimilation. In addition to assessing the analysis quality, the effectiveness of assimilation for initializing the model forecasts is evaluated.

The following Section 2 presents the model setup and briefly reviews the 3D-Var assimilation method. The procedure for estimating the background and observation error covariance matrices is discussed in Section 3. The assimilation results for O3 and NO2 for the year 2012 are discussed in Section 4. Section 5 concludes the paper.

2 Materials and methods

This section presents the SILAM dispersion model, the observation datasets used, and describes the assimilation procedure.

2.1 The SILAM dispersion model and experiment setup

This study employs the SILAM chemistry transport model (CTM) version 5.3. The model utilizes the semi-Lagrangian advection scheme of Galperin (2000) combined with the vertical discretization described by Sofiev (2002) and the boundary layer scheme of Sofiev et al. (2010). Wet and dry deposition are parameterized as described in Sofiev et al. (2006).

Chemistry of ozone and related reactive pollutants is simulated using the Carbon Bond 4 chemical mechanism (CB4, Gery et al., 1989). However, the NO2 analyses are produced with separate simulations employing the DMAT chemical scheme of Sofiev (2000). This follows the setup used in operational air quality forecasts with the SILAM model, where the two model runs are necessary since the primary and secondary inorganic aerosols are only included in the DMAT scheme. The SILAM model has been previously applied in simulating regional ozone and NO2 concentrations (Huijnen et al., 2010; Langner et al., 2012; Solazzo et al., 2012), for global-scale aerosol simulations (Sofiev et al., 2011) as well as for simulating emission and dispersion of allergenic pollen (Siljamo et al., 2012). The daily, European-scale air quality forecasts contributing to the MACC-II project are publicly available at http://macc-raq.gmes-atmosphere.eu.

In this study, the model is configured for a European domain covering the area between 35.2° and 70.0° N and -14.5° and 35.0° E with a regular lon-lat grid. The vertical discretization consists of eight terrain-following levels reaching up to about 6.8 km. The vertical coordinate is geometric height. The model is driven by operational ECMWF IFS forecast fields, which are initially extracted in a 0.125 degree lon-lat grid and further interpolated to the CTM resolution. Chemical boundary conditions are provided by the MACC reanalysis (Inness et al., 2013), which uses the MOZART global chemistry-transport model.

The emissions of anthropogenic pollutants are provided by the MACC-II European emission inventory (Kuenen et al., 2014) for the reference year 2009. The biogenic isoprene emissions, required by the CB4 run, are simulated by the BEM emission model (Poupkou et al., 2010).

Three sets of SILAM simulations are carried out in this study. First, the background and observation error covariance matrices are calibrated using one-month simulations for June and December 2011. The results of calibration are used in reanalysis simulations covering year 2012. Finally, a set of 72 hour hindcasts is generated for the period between 16 July and 5 August, 2012, to evaluate the forecast impact of assimilation. The hindcasts are initialized from the 00 UTC analysis fields. The timespan includes an ozone

episode affecting parts of Southern and Western Europe (EEA, 2013). The reanalysis and hindcasts use identical meteorological and boundary input data, and hence, the hindcasts only assess the effect of chemical data assimilation.

The analysis and forecast runs are performed at a horizontal resolution of 0.2 degrees. The setup for calibrations runs (June and December 2011) is identical except that a coarser horizontal resolution of 0.5° is chosen in order to reduce the computational burden. The model timestep is 15 minutes for both setups.

2.2 Observations

This study ~~utilises~~uses the hourly observations of NO2 and O3 at background stations available in the Airbase database (http://acm.eionet.europa.eu/databases/airbase/) maintained by the European Environmental Agency. Separate subsets are employed for assimilation and evaluation.

Two sets of stations are withheld for evaluation. The first set, referred here as the MACC set, has been used in the regional air quality assessments within the MACC and MACC-II projects (Rouïl, 2013, also Curier et al., 2012). The second set consists of the stations reported as EMEP stations in the database. The MACC validation stations include about a third of the available background stations for each species, and are chosen with the requirement to cover the same area as the assimilation stations. ~~However, due to uncertainties in representativeness of suburban or urban stations, only rural stations are considered for the evaluation of the 2012 reanalyses.~~ The EMEP network is sparser and has no particular relation to the assimilation stations. It can be noted that the EMEP stations included in Airbase do not comprise the full EMEP monitoring network.

~~All other stations are available for assimilation. However, to reduce the effect of representativeness errors, data from urban stations are not assimilated, and for NO2, also suburban stations are excluded.~~The in-situ data are used for assimilation and evaluation under the assumption that they represent the pollutant levels in spatial scales resolved by the model. We expect this assumption to be violated especially at many urban and suburban stations due to local variations in emission fluxes. For this reason, only rural stations are used for evaluation of the 2012 reanalysis. The NO2 assimilation set also excludes both urban and suburban stations. For ozone, the data from suburban stations are assimilated, however, the observation errors are assessed separately for suburban and rural stations, as outlined in Section 3. The station sets are presented on a map in Figure 1.

The statistical indicators used for model evaluation are correlation, mean bias and root mean squared error (RMSE). Since air quality models are frequently used to evaluate daily maximum concentrations, the indicators are evaluated separately for the daily maximum values.

2.3 The 3D-Var assimilation

In the 3D-Var method, the analysis $\mathbf{x}_a$ minimises the cost function

(1) $$J(\mathbf{x}) = \frac{1}{2}\left(\mathbf{y} - \mathcal{H}(\mathbf{x})\right)^T \mathbf{R}^{-1}\left(\mathbf{y} - \mathcal{H}(\mathbf{x})\right) + \frac{1}{2}\left(\mathbf{x} - \mathbf{x}_b\right)^T \mathbf{B}^{-1}\left(\mathbf{x} - \mathbf{x}_b\right),$$

where $\mathbf{x}_b$ is the background state, $\mathbf{y}$ is the vector of observations, and $\mathcal{H}$ is the possibly nonlinear observation operator. The uncertainties of the background state $\mathbf{x}_b$ and the observations $\mathbf{y}$ are described by the background and observation error covariance matrices $\mathbf{B}$ and $\mathbf{R}$, respectively. In this study, the control variable $\mathbf{x}$ consists of the three-dimensional airborne concentration ~~field~~ for either NO2 or ozone. The `m1qn3` minimization code (Gilbert and Lemaréchal, 1989) is used for solving the optimisation problem (1).

For the surface measurements, the operator $\mathcal{H}$ is linear and consists of horizontal interpolation only, since the surface concentrations are considered to be represented by the lowest model level. Following the hourly observation frequency, the analysis is performed every hour followed by a one-hour forecast. The forecast provides the background field for the subsequent analysis.

In the current study, only single chemical component is assimilated in each run. Since O3 is not a prognostic variable in the DMAT scheme, it cannot be assimilated into the NO2 simulation. Assimilating NO2 observations into the CB4 simulation would be technically feasible; however, simultaneous assimilation of NO2 and O3 would require care due to the strong chemical coupling between the species. The background and observation error covariance matrices would also need to be estimated jointly.

3 Background and observation error covariance matrices

The numerical formulation of the BECM in the current work follows the assumptions made by Vira and Sofiev (2012). We assume that the background error correlation is homogeneous in space, and its horizontal component is described by a Gaussian function of distance between the grid points. Furthermore, we assume that the background error standard deviation σ_b is independent of location. This allows writing the BECM as $\mathbf{B} = \sigma_b^2 \mathbf{C}$, where $\mathbf{C}$ is the correlation matrix and σ_b is the background error standard deviation.

For estimation of the parameters for the covariance matrices $\mathbf{B}$ and $\mathbf{R}$, we combined the NMC method, which is used for determining the correlation matrix $\mathbf{C}$, and the approach of Desroziers et al. (2005), which is used for diagnosing the observation and background error standard deviations.

In the NMC method, the difference between two forecasts valid at a given time is taken as a proxy of the forecast error. In this work, the proxy dataset is extracted from 24 and 48 hour regional air quality forecasts for year 2010. The forecasts are generated with the SILAM model in a configuration similar to the one used in this study. Since no chemical data assimilation is used in the forecasts, the differences are due to changes in ~~forecasted~~forecast meteorology and boundary conditions only. The lead times are chosen to allow sufficient spread to develop between the forecasts. The forecast data are segregated by hour resulting in separate sets for hours 00, 06, 12 and 18 UTC, and the correlations are interpolated for all other times of day.

The horizontal and vertical components of the correlation matrix $\mathbf{C}$ are estimated separately. The horizontal correlation is determined by the length scale *L*, which is obtained by fitting a Gaussian correlation function to the dataset. First, the sample correlation matrix $\mathbf{C}$ of the forecast differences is calculated. Then, the Gaussian correlation function is fitted to the empirical correlations $\tilde{\mathbf{C}}_{ij}$ by minimizing

(2) $$f(L) = \sum_{|r_i - r_j| < d} \left| \tilde{\mathbf{C}}_{ij} - \mathbf{C}_{ij}(L) \right|^2,$$

where the fitted correlation function is $\mathbf{C}_{ij}(L) = \exp(-(|x_i - x_j|^2 + |y_i - y_j|^2)/L^2)$ and x and y are the Cartesian coordinates for each grid point. To reduce the effect of spurious long-distance correlations due to the limited sample size, the fitting is restricted to grid points r_i closer than *d*=1000 km to each other. The distances, shown in Table 1, are computed for the lowest model layer.

The vertical correlation function is obtained directly as the sample correlation across all vertical columns for each time of day. As an example, the correlation matrix obtained for NO2 at 12 UTC is shown in Figure 2.

Since the NMC dataset includes only meteorological perturbations, it is expected to underestimate the total uncertainty of the CTM simulations. Hence, the standard deviations are not diagnosed from the NMC dataset, but instead, and approach based on a posteriori diagnostics is taken. The approach, devised by (Desroziers et al., 2005), is based on a set of identities which relate the BECM and OECM to expressions which can be estimated statistically from a set of analysis and corresponding background fields.

First, the standard deviation $\sigma_{obs}^{(i)}$ of the i th observation component is equal to

(3) $$E[(\mathbf{y}^{(i)} - \mathbf{y}_a^{(i)})(\mathbf{y}^{(i)} - \mathbf{y}_b^{(i)})] = \sigma_{obs}^{(i)^2},$$

where E denotes the expectation, $\mathbf{y}$ is the observation vector and $\mathbf{y}_a = \mathcal{H}(\mathbf{x}_a)$ and $\mathbf{y}_b = \mathcal{H}(\mathbf{x}_b)$ are evaluated from the analysis and background fields, respectively.

The background error covariance matrix cannot be uniquely expressed in observation space. However, assuming that each observation only depends (linearly) on a single model grid cell (ie. horizontal interpolation is neglected), then

(4) $$E[(\mathbf{y}_{\mathbf{a}}^{(i)} - \mathbf{y}_{\mathbf{b}}^{(i)})(\mathbf{y}^{(i)} - \mathbf{y}_{\mathbf{b}}^{(i)})] = \sigma_b^{(i)^2}.$$

The identities (3) and (4) hold for an ideally defined analysis system, provided that the background and observation errors are normally distributed and assuming the observation operator is not strongly nonlinear.

Furthermore, Equations (3) and (4) can be used to tune the parameters σ_{obs} and σ_b by means of fixed point iteration. First, a set of analyses is produced using initial parameter values. Then, the left-hand sides of (3) and (4) are evaluated as averages over the analyses, resulting in new parameter values. The procedure is then repeated using the updated σ_b and σ_{obs} to produce a new set of analyses. In this work, we stopped the iteration when the RMSE at validation stations was no longer improving. We chose this criterion to avoid overfitting the parameters to the calibration data.

In this work, the observation error covariance matrix R is assumed diagonal. The initial values for σ_{obs} and σ_b were set to 11.2 and 20.6 µg m^{-3} for O3, and 4.0 and 8.0 µg m^{-3} for NO2. The values correspond to typical mean-squared errors for a free-running model, which are attributed to the model and observation error variances in the ratio of 80/20, respectively. The standard deviations, together with the correlation matrices obtained with the NMC procedure, are then employed in the iterations to calculate a set of hourly analyses for the two calibration periods spanning June and December 2011.

The choice of calibration periods representing both winter and summer conditions is motivated by the strong seasonal variations in both O3 and NO2. Both σ_{obs} and σ_b are segregated by hour, while for O3 σ_{obs} is also evaluated separately for suburban stations. For the reanalysis of year 2012, the standard deviations, obtained separately for June and December, are interpolated linearly for all other months.

Finally, the overall consistency can be evaluated by checking the identity (Ménard et al., 2000)

(5) $$E(\chi^2) = N,$$

where $\chi^2 = 2J(\mathbf{x}_a)$ is twice the value of cost function (1) at the minimum, and N is dimension of the observation vector $\mathbf{y}$. The identity (5) tests the overall consistency of the analysis and is affected by both $\mathbf{B}$ and $\mathbf{R}$.

4 Results and discussion

The SILAM model was run for year 2012 with and without assimilation. Since the 3D-Var analyses require no additional model integrations in form of iterations or ensemble simulations, the hourly analyses increase the simulation runtime by only 10-15%.

The effect of assimilation to the yearly-mean concentrations on the lowest model level is shown in Figure 3. On average, the ozone concentrations are increased by the assimilation especially around the Mediterranean Sea, which indicates corresponding low bias in the free model run. The main changes in NO2 levels are confined to somewhat more limited areas; in particular areas near major mountain ranges (Alps and Pyrenees) show enhanced NO2 levels in the analysis run.

4.1 Background and observation error covariance matrices

Refining the background and observation standard deviations iteratively both improves the consistency of the assimilation setup as measured by the χ^2 indicator (Eq. (5)), and improves the model-measurement comparison on the validation stations over the calibration period. However, after five iterations (for both June and December), the changes in χ^2 become slow and the validation scores no longer improve. Hence, the values for σ_{obs} and σ_b in fifth iterations were taken as the final values for 2012 reanalysis. The changes in χ^2 and model-measurement RMSE are summarized in Table 2.

The diagnosed observation and background error standard deviations for O3 and NO2 are shown in Figure 4. For June, the standard deviations for ~~for~~ ozone range between 11 and 21 µg/m^{-3} for rural stations. For December, the diurnal variation is flatter, but the absolute values ~~essentially~~are generally not reduced, in contrast to the ~~general~~overall seasonality of O3.

Especially for summertime night conditions, the values are higher than the values adopted in most of the earlier studies (Chai et al., 2007; Curier et al., 2012; Jaumouillé et al., 2012). However, the errors are comparable to the observation errors diagnosed using the CHIMERE model by Gaubert et al. (2014). The main error component is likely to be due to lack of representativeness: using the AIRNOW observation network, Chai et al. (2007) found standard deviations between 5 and 13 ppb for observations inside a grid cell with 60 km resolution. The maximum values occurred during night time.

The diagnosed observation and background error parameters are subject to uncertainty, since they are not uniquely determined (Schwinger and Elbern, 2010). Also, the parameters depend on the assumptions made regarding the correlation function. Nevertheless, the relative magnitude of observation errors during night is interesting for interpreting the model-to-measurement comparisons.

The diagnosed background errors for ozone are between 5 and 9 µg/m^{-3} depending on month and time of day. For June, the diagnosed errors are largest between 9-10 and 21-22 UTC, which coincides with transitions between stable and convective boundary layers in summertime conditions. For December, only minor diurnal variation is observed.

The observation error standard deviation for NO2 varies between 2.8 and 5.2 µg/m^{-3} for rural stations. Suburban or urban stations were not assimilated for NO2. Contrary to ozone, the diurnal variation of background and observation errors both positively correlate with the diurnal variation of the pollutant.

The BECM and OECM were adjusted to optimize self-consistency for two months in 2011. To assess the robustness of the obtained formulations, the χ^2 indicator was computed also for all analysis steps for the 2012 reanalysis simulation.

As seen in **Error! Reference source not found.**, the analyses using the adjusted BECM and OECM generally satisfy the consistency relation better throughout the year, when compared to the first-guess values. The yearly-mean values for χ^2 are 1.05 and 0.97 for ozone and NO2, respectively.

Overall, the assimilation system is based on rather simplistic assumptions regarding the background and observation error statistics. In addition to computational efficiency, this approach benefits ~~of~~from having few tuning parameters, and the remaining parameters (σ_{obs}, σ_b and L) can be estimated using an automated procedure. As shown in the following section, the refined background and observation error definitions provide a clear improvement on analysis scores at the control stations, despite the rather limited training datasets.

4.2 Evaluation against independent observations

Tables 3 and 4 present the analysis skill scores for runs with both first guess and final BECM and OECM, and for the free-running model with no assimilation.

In terms of correlation and RMSE, both analysis and free model runs show better performance for predicting the daily maximum than hourly values. This applies to both O3 and NO2, although the difference is more marked for ozone. The opposite holds for bias, which tends to be higher when calculated for daily maxima.

The comparison reveals a number of contrasts between the “MACC” and “EMEP” validation stations. First, the free-running model shows better performance for NO2 on the EMEP stations, while for ozone, the performance is better on the MACC stations. On the other hand, the data assimilation has stronger impact on the scores for the MACC validation stations. This is especially visible the case for NO2, a result which is

consistent with the shorter lifetime of NO2 compared to O3~~. The shorter lifetime would make the MACC validation stations, which are generally located closer to the assimilation stations, more sensitive to the assimilation~~.

The differences largely originate from the different representativeness and coverage of the MACC and EMEP station sets. As seen in Figure 1, the EMEP network covers the computational domain more evenly than the MACC validation stations, which are concentrated in Central Europe. Since the coverage of assimilation and MACC validation stations is similar, the average impact of assimilation is stronger on the MACC than EMEP stations.

For the free-running simulations, the better performance for O3 at the MACC stations is consistent with the geographical variations in the model skill: the densest coverage of the MACC validation stations coincides with the parts of Europe where many regional air quality models perform best for ozone (eg. Vautard et al., 2009). The scores for NO2 also vary by region, however, due to the shorter chemical lifetime, the forecasts of NO2 are more sensitive to unresolved variations in local emissions. This probably explains the better scores for NO2 on the EMEP stations, since the EMEP network is specifically aimed at monitoring the background levels of pollutants.

For ozone, the assimilation had ~~uneven~~a variable effect on the model bias. While the correlation and RMSE were always improved by assimilation, the analyses have slightly larger negative mean bias (-4.6 vs -4.0 µg m^3 on MACC stations) than the free model. This is confirmed by the average diurnal profile shown in Figure 6. Diurnal variation of model bias (µg m^{-3}). The first guess assimilation setup is shown in red and the final setup in blue. The reference run with no assimilation is drawn in green. The values are shown for the rural MACC validation stations and averaged over each day of year 2012 and over the stations.. However, the diurnal variation of analysis errors is flatter, and the strongest bias no longer coincides with the afternoon hours, when the highest O3 concentrations are typically observed.

~~The analysis biases for O3 are not surprising given the similar bias in the free-running simulation, since the analysis scheme assumes an unbiased model. This also explains why~~For NO2, the analyses have only slight negative bias (-0.38 µg/m^3) on the MACC stations, which turns positive (about 1 µg/m^3) for the more remote EMEP sites. As seen in Table 4, the difference between the station sets is similar to that of the free-running model. Given the differences between the MACC and EMEP station sets, this suggests that the model overestimates the lifetime of NO2, which in turn results in the positive bias in the analyses. The long lifetime of NO2 in the SILAM DMAT chemistry scheme was also noticed by Huijnen et al. (2010).

The analysis scheme assumes an unbiased model, and hence, the negative bias present in the free-running simulations is reduced but not removed in the analysis fields. The assimilation setup including tuned OECM

and BECM produces more biased analyses compared to the first-guess setup, as seen in Figure 6. Diurnal variation of model bias (μg m^{-3}). The first guess assimilation setup is shown in red and the final setup in blue. The reference run with no assimilation is drawn in green. The values are shown for the rural MACC validation stations and averaged over each day of year 2012 and over the stations.. ~~As shown by Dee (2005), such issues can in principle be addressed by the assimilation system.~~This is a consequence of the differences between the diagnosed and first-guess background and observation error standard deviations. Contrary to the tuned setup, the first-guess attributes most of the model-observation discrepancy to the background error, which results in stronger increments towards the observed values. Consequently, the analysis bias is smaller. However, the tuned assimilation setup has consistently better RMSE and correlation than the first guess assimilation setup.

Since the analysis bias is mainly a consequence of a bias in the forecast model, the bias should be addressed primarily by improving the model. As shown by Dee (2005), model biases can in principle be addressed also by the assimilation system. However, a possible bias correction scheme should be implemented with care, since also observational biases could arise due to representativeness errors.

~~The assimilation setup obtained by iterative tuning of observation and background error parameters has consistently better RMSE and correlation than the first guess assimilation setup.~~

In addition to computing the regular statistical indicators for daily maxima, we evaluated the hit rates (the number of correctly predicted exceedances divided by the number of observed exceedances) for the 180 μg/m^3 threshold for O3, with and without assimilation. Assimilation turns out to improve also the hit rate, albeit only slightly: from 0.25 to 0.26 on average for rural MACC validation stations, and from 0.13 to 0.15 for EMEP stations. If the averaging is restricted to the stations with more than 10 exceedances during 2012, the values change from 0.32 to 0.36 for MACC and from 0.21 to 0.43 for the EMEP stations. Obviously, the hit rates are sensitive to the low bias in the daily maxima.

For NO2, a specific source of observational errors is due to the molybdenum converters used in the chemiluminescence technique, which is the most common measurement technique for monitoring NO2. As discussed by Dunlea et al. (2007) and Steinbacher et al. (2007), this technique is subject to positive interference by the NOz species such as PAN, HNO3 and HONO.

The interference can lead to overestimation of NO2 by up to factor of two, however, the error varies by location and time, and may depend on features of the instrument (Steinbacher et al., 2007). We estimated the magnitude of this effect from the free-running CB4 simulation. On most continental EMEP sites, the contribution of the NOz species to the total NOz + NO2 was about 10-20% of the simulated yearly mean. However, for a few sites the contribution could reach 50%.

The O3 and NO2 observations were assimilated into separate model runs. Assimilation of O3 had only a minor influence on NO2 in the CB4 simulation; however, the mean bias was reduced by about 5% on average for the MACC validation stations. Because the DMAT simulation does not include ozone as a tracer, the impact of NO2 assimilation on ozone fields was not evaluated in this study.

4.3 Forecast experiments

In order to quantify the usefulness of data assimilation forecast applications, a set of simulations without data assimilation were generated using the analysis fields at 00 UTC as initial conditions. The forecast experiment covered time between 16 July and 5 August, 2012.

The effect of chemical data assimilation on forecast performance was assessed as a function of the forecast lead time. Figures 7 and 8 present the correlation and bias for the O3 and NO2 forecasts, respectively, and compare them to the corresponding indicators for the analyses and the control run.

For ozone, the forecast improvements due to data assimilation were largely limited to the first 24 hours of forecast. Also, the forecast initialized at 00:00 UTC from the analysis shows a larger negative bias for the daytime than the free model run. This is a result of the corresponding night time positive bias of the free model run. The bias is effectively removed in the 00 analysis; however, the subsequent forecast is unable to recover the level observed during daytime. The correlation coefficient during daytime is nevertheless improved slightly (from 0.75 to 0.78) by initializing from the analysis. While the forecast shows somewhat reduced positive bias for hours between 18 and 30 (ie. the following night), the subsequent daytime scores are already almost unchanged by assimilation. The results in Figure 7 are computed for the MACC station network; similar impact is observed ~~on~~at the EMEP stations.

Due to the shorter chemical lifetime, the effect of initial condition on forecasts of NO2 can be expected to ~~vanish~~fall away more quickly than for ozone. This has been confirmed in the previous works based on assimilation of data from the OMI instrument. Under summer conditions, Wang et al. (2011) found assimilation to provide no improvement in RMSE with regard to surface observations, while Silver et al. (2013) reported the NO2 concentration to relax to its background values within 3-4 hours.

In the forecast experiments performed within this study, the effect of assimilation on NO2 forecast scores was limited to the first 6 forecast hours, which coincides with the night in most of the domain. Hence, at least under the photochemically active summertime conditions, the analyses are only marginally useful for improving forecasts of NO2.

The forecast for short-lived pollutants like NO2 is poorly constrained by the initial condition, because the boundary layer concentrations become driven mainly by local emissions, chemical transformations and

deposition. This limits effectiveness of any assimilation scheme based updating only the initial condition. A possible way to extend the forecast impact is to include more persistent parameters, such as emission rates, into the state vector. This has been demonstrated by Elbern et al. (2007) for forecasting an ozone episode. In general, such an approach requires that the obtained a posteriori emission rates can be extrapolated to the forecast window, and that the assimilation scheme is able to correctly attribute the observed discrepancies to the uncertain parameters.

5 Conclusions

An assimilation system coupled to the SILAM chemistry transport model has been described along with its application in reanalysis of ozone and NO2 concentrations for year 2012. Furthermore, the impact of using the O3 and NO2 analyses to initialize forecasts has been assessed for an ozone episode occurring in July 2012.

The assimilation consistently improves the model-measurement comparison for stations not included in the assimilation. For daily maximum values, the correlation coefficient is improved over the free running model from 0.8 to 0.9 for O3 and from 0.53 to 0.63 for NO2 on rural validation stations. The respective biases are also decreased, however, a bias of -7.4 µg m^{-3} remains in the O3 analyses due to a negative bias in the free-running model.

~~Initializing the forecasts from the analysis fields provided an improvement in ozone forecast skill for maximum of 24 hours. For NO2, the improvement was limited to a window of 6 hours. These findings are similar to the results published in previous studies.~~

During a three-week forecast experiment, initializing the forecasts from the analysis fields provided an improvement in ozone forecast skill for a maximum of 24 hours. For NO2, the improvement was limited to a window of 6 hours. The findings for NO2 are similar to the results published in previous studies (Silver et al., 2013; Wang et al., 2011).

The diagnosed observation error standard deviations for ozone have a strong diurnal variation, and reach up to about 21 µg m^{-3} during night. These values are higher than usually assumed in chemical data assimilation, but ~~corroborate~~agree well with the results obtained by Gaubert et al. (2014) with similar diagnostics.

The 3D-Var based assimilation has a low computational overhead. This makes it especially suitable for reanalyses in yearly or longer time scales, as well as for high-resolution forecasting under operational time constraints. Future work will include more accurate characterization of station representativeness as well as further investigation of model biases for O3.

Acknowledgements

This work has been supported by the FP7 projects MACC and MACC-II and the NordForsk project EmblA.

The authors thank Marje Prank for constructive comments on the manuscript.

References

Benedetti, A., Morcrette, J.-J., Boucher, O., Dethof, A., Engelen, R.J., Fisher, M., Flentje, H., Huneeus, N., Jones, L., Kaiser, J.W., Kinne, S., Mangold, A., Razinger, M., Simmons, A.J., Suttie, M., 2009. Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation. J. Geophys. Res. 114, D13205. doi:10.1029/2008JD011115

Blond, N., Vautard, R., 2004. Three-dimensional ozone analyses and their use for short-term ozone forecasts. J. Geophys. Res. 109, 1–14. doi:10.1029/2004JD004515

Bocquet, M., 2012. Parameter-field estimation for atmospheric dispersion: application to the Chernobyl accident using 4D-Var. Q. J. R. Meteorol. Soc. 138, 664–681. doi:10.1002/qj.961

Chai, T., Carmichael, G.R., Tang, Y., Sandu, A., Hardesty, M., Pilewskie, P., Whitlow, S., Browell, E. V., Avery, M.A., Nédélec, P., Merrill, J.T., Thompson, A.M., Williams, E., 2007. Four-dimensional data assimilation experiments with International Consortium for Atmospheric Research on Transport and Transformation ozone measurements. J. Geophys. Res. 112, 1–18. doi:10.1029/2006JD007763

Constantinescu, E.M., Sandu, A., Chai, T., Carmichael, G.R., 2007. Assessment of ensemble-based chemical data assimilation in an idealized setting. Atmos. Environ. 41, 18–36. doi:10.1016/j.atmosenv.2006.08.006

Curier, R.L., Timmermans, R., Calabretta-Jongen, S., Eskes, H., Segers, a., Swart, D., Schaap, M., 2012. Improving ozone forecasts over Europe by synergistic use of the LOTOS-EUROS chemical transport model and in-situ measurements. Atmos. Environ. 60, 217–226. doi:10.1016/j.atmosenv.2012.06.017

Dee, D.P., 2005. Bias and data assimilation. Q. J. R. Meteorol. Soc. 131, 3323–3343. doi:10.1256/qj.05.137

Desroziers, G., Berre, L., Chapnik, B., Poli, P., 2005. Diagnosis of observation, background and analysis-error statistics in observation space. Q. J. R. Meteorol. Soc. 131, 3385–3396. doi:10.1256/qj.05.108

Dunlea, E.J., Herndon, S.C., Nelson, D.D., Volkamer, R.M., Martini, F.S., Sheehy, P.M., Zahniser, M.S., 2007. Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment. Atmos. Chem. Phys. 7, 2691–2704.

EEA, 2013. Air pollution by ozone across Europe during summer 2012, EEA Technical report.

Elbern, H., Schmidt, H., 2001. Ozone episode analysis by four-dimensional variational chemistry data assimilation. J. Geophys. Res. 106, 3569–3590.

Elbern, H., Strunk, ~~a~~A., Schmidt, H., Talagrand, O., 2007. Emission rate and chemical state estimation by 4-dimensional variational inversion. Atmos. Chem. Phys. ~~Discuss.~~ 7, ~~1725–1783~~3749–3769. doi:10.5194/acpd-7-1725-2007

Errera, Q., Daerden, F., Chabrillat, S., Lambert, J.C., Lahoz, W.A., Viscardy, S., Bonjean, S., Fonteyn, D., 2008. 4D-Var assimilation of MIPAS chemical observations : ozone and nitrogen dioxide analyses. Atmos. Chem. Phys. 8, 6169–6187.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99, 10143–10162.

Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dyn. 53, 343–367. doi:10.1007/s10236-003-0036-9

Galperin, M., 2000. The approaches to correct computation of airborne pollution advection, in: Problems of Ecological Monitoring and Ecosystem Modelling. XVII (in Russian). Gidrometeoizdat, pp. 54–68.

Gandin, L.S., 1963. Objective analysis of meteorological fields, Gidrometeorologischeskoe Izdatel'stvo. Translated (1965) by Israel Programme for Scientific Translation, Jerusalem., Leningrad.

Gaubert, B., Coman, A., Foret, G., Meleux, F., Ung, A., Rouil, L., Ionescu, A., Candau, Y., Beekmann, M., 2014. Regional scale ozone data assimilation using an ensemble Kalman filter and the CHIMERE chemical transport model. Geosci. Model Dev. 7, 283–302. doi:10.5194/gmd-7-283-2014

Gery, M.W., Whitten, G.Z., Killus, J.P., Dodge, M.C., 1989. A photochemical kinetics mechanism for urban and regional scale computer modeling. J. Geophys. Res. 94, 12925–12956.

Gilbert, J.C., Lemaréchal, C., 1989. Some numerical experiments with variable-storage quasi-Newton algorithms. Math. Program. 45, 407–435.

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

Huijnen, V., Eskes, H.J., Poupkou, A., Elbern, H., Boersma, K.F., Foret, G., Sofiev, M., Valdebenito, A., Flemming, J., Stein, O., Gross, A., Robertson, L., D'Isidoro, M., Kioutsioukis, I., Friese, E., Amstrup, B., Bergstrom, R., Strunk, A., Vira, J., Zyryanov, D., Maurizi, A., Melas, D., Peuch, V.-H., Zerefos, C., 2010. Comparison of OMI NO2 tropospheric columns with an ensemble of global and European regional air quality models. Atmos. Chem. Phys. 10, 3273–3296. doi:10.5194/acp-10-3273-2010

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R.J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J.W., Kapsomenakis, J., Lefever, K., Leitão, J., Razinger, M., Richter, A., Schultz, M.G., Simmons, a. J., Suttie, M., Stein, O., Thépaut, J.-N., Thouret, V., Vrekoussis, M., Zerefos, C., 2013. The MACC reanalysis: an 8 yr data set of atmospheric composition. Atmos. Chem. Phys. 13, 4073–4109. doi:10.5194/acp-13-4073-2013

Jaumouillé, E., Massart, S., Piacentini, a., Cariolle, D., Peuch, V.-H., 2012. Impact of a time-dependent background error covariance matrix on air quality analysis. Geosci. Model Dev. 5, 1075–1090. doi:10.5194/gmd-5-1075-2012

Kahnert, M., 2008. Variational data analysis of aerosol species in a regional CTM: background error covariance constraint and aerosol optical observation operators. Tellus B 60, 753–770. doi:10.1111/j.1600-0889.2008.00377.x

Kalnay, E., 2003. Atmospheric modeling, data assimilation and predicability. Cambridge University Press.

Kuenen, J.J.P., Visschedijk, a. J.H., Jozwicka, M., Denier van der Gon, H. a. C., 2014. TNO-MACC_II emission inventory: a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling. Atmos. Chem. Phys. Discuss. 14, 5837–5869. doi:10.5194/acpd-14-5837-2014

Kukkonen, J., Olsson, T., Schultz, D.M., Baklanov, a., Klein, T., Miranda, a. I., Monteiro, a., Hirtl, M., Tarvainen, V., Boy, M., Peuch, V.-H., Poupkou, a., Kioutsioukis, I., Finardi, S., Sofiev, M., Sokhi, R., Lehtinen, K.E.J., Karatzas, K., San José, R., Astitha, M., Kallos, G., Schaap, M., Reimer, E., Jakobs, H., Eben, K., 2012. A review of operational, regional-scale, chemical weather forecasting models in Europe. Atmos. Chem. Phys. 12, 1–87. doi:10.5194/acp-12-1-2012

Kumar, U., Ridder, K. De, Lefebvre, W., Janssen, S., 2012. Data assimilation of surface air pollutants (O3 and NO2) in the regional-scale air quality model AURORA. Atmos. Environ. 60, 99–108. doi:10.1016/j.atmosenv.2012.06.005

Langner, J., Engardt, M., Baklanov, a., Christensen, J.H., Gauss, M., Geels, C., Hedegaard, G.B., Nuterman, R., Simpson, D., Soares, J., Sofiev, M., Wind, P., Zakey, a., 2012. A multi-model study of impacts of climate change on surface ozone in Europe. Atmos. Chem. Phys. 12, 10423–10440. doi:10.5194/acp-12-10423-2012

Le Dimet, F.-X., Talagrand, O., 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. Tellus A 38A, 97–110. doi:10.1111/j.1600-0870.1986.tb00459.x

Lorenc, A.C., 1986. Analysis methods for numerical weather prediction. Q. J. R. Meteorol. Soc. 112, 1177–1194. doi:10.1002/qj.49711247414

Massart, S., Piacentini, a., Pannekoucke, O., 2012. Importance of using ensemble estimated background error covariances for the quality of atmospheric ozone analyses. Q. J. R. Meteorol. Soc. 138, 889–905. doi:10.1002/qj.971

Ménard, R., Cohn, S.E., Chang, L.-P., Lyster, P.M., 2000. Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part I: Formulation. Mon. Weather Rev. 128, 2654–2671.

Parrish, D.F., Derber, J.C., 1992. The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. Mon. Weather Rev. 120, 1747–1763.

Poupkou, A., Giannaros, T., Markakis, K., Kioutsioukis, I., Curci, G., Melas, D., Zerefos, C., 2010. A model for European Biogenic Volatile Organic Compound emissions: Software development and first validation. Environ. Model. Softw. 25, 1845–1856. doi:10.1016/j.envsoft.2010.05.004

Rabier, F., 2005. Overview of global data assimilation developments in numerical weather-prediction centres. Q. J. R. Meteorol. Soc. 131, 3215–3233. doi:10.1256/qj.05.129

Rouïl, L. (Ed.), 2013. Validation report for the 2010 Air Quality Assessment Report.

Schwartz, C.S., Liu, Z., Lin, H.-C., McKeen, S. a., 2012. Simultaneous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth. J. Geophys. Res. 117. doi:10.1029/2011JD017383

Schwinger, J., Elbern, H., 2010. Chemical state estimation for the middle atmosphere by four-dimensional variational data assimilation: A posteriori validation of error statistics in observation space. J. Geophys. Res. 115. doi:10.1029/2009JD013115

Siljamo, P., Sofiev, M., Filatova, E., Grewling, L., Jäger, S., Khoreva, E., Linkosalo, T., Ortega Jimenez, S., Ranta, H., Rantio-Lehtimäki, A., Svetlov, A., Veriankaite, L., Yakovleva, E., Kukkonen, J., 2013. A numerical model of birch pollen emission and dispersion in the atmosphere. Model evaluation and sensitivity analysis. Int. J. Biometeorol. 57, 125–136. doi:10.1007/s00484-012-0539-5

Silver, J.D., Brandt, J., Hvidberg, M., Frydendall, J., Christensen, J.H., 2013. Assimilation of OMI NO_2 retrievals into the limited-area chemistry-transport model DEHM (V2009.0) with a 3-D OI algorithm. Geosci. Model Dev. 6, 1–16. doi:10.5194/gmd-6-1-2013

Sofiev, M., 2000. A model for the evaluation of long-term airborne pollution transport at regional and continental scales. Atmos. Environ. 34, 2481–2493. doi:10.1016/S1352-2310(99)00415-X

Sofiev, M., 2002. Extended resistance analogy for construction of the vertical diffusion scheme for dispersion models. J. Geophys. Res. 107. doi:10.1029/2001JD001233

Sofiev, M., Genikhovich, E., Keronen, P., Vesala, T., 2010. Diagnosing the Surface Layer Parameters for Dispersion Models within the Meteorological-to-Dispersion Modeling Interface. J. Appl. Meteorol. Climatol. 49, 221–233. doi:10.1175/2009JAMC2210.1

Sofiev, M., Siljamo, P., Valkama, I., Ilvonen, M., Kukkonen, J., 2006. A dispersion modelling system SILAM and its evaluation against ETEX data. Atmos. Environ. 40, 674–685. doi:10.1016/j.atmosenv.2005.09.069

Sofiev, M., Soares, J., Prank, M., de Leeuw, G., Kukkonen, J., 2011. A regional-to-global model of emission and transport of sea salt particles in the atmosphere. J. Geophys. Res. 116. doi:10.1029/2010JD014713

Solazzo, E., Bianconi, R., Vautard, R., Appel, K.W., Moran, M.D., Hogrefe, C., Bessagnet, B., Brandt, J., Christensen, J.H., Chemel, C., Coll, I., Denier van der Gon, H., Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A.B., Jeričević, A., Kraljević, L., Miranda, A.I., Nopmongcol, U., Pirovano, G., Prank, M., Riccio, A., Sartelet, K.N., Schaap, M., Silver, J.D., Sokhi, R.S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S.T., Galmarini, S., 2012. Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII. Atmos. Environ. 53, 60–74. doi:10.1016/j.atmosenv.2012.01.003

Steinbacher, M., Zellweger, C., Schwarzenbach, B., Bugmann, S., Buchmann, B., Ordóñez, C., Prevot, a. S.H., Hueglin, C., 2007. Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques. J. Geophys. Res. 112, D11307. doi:10.1029/2006JD007971

Tombette, M., Mallet, V., Sportisse, B., 2009. PM 10 data assimilation over Europe with the optimal interpolation method. Atmos. Chem. Phys. 9, 57–70.

Wang, X., Mallet, V., Berroir, J., Herlin, I., 2011. Assimilation of OMI NO2 retrievals into a regional chemistry-transport model for improving air quality forecasts over Europe. Atmos. Environ. 45, 485–492. doi:10.1016/j.atmosenv.2010.09.028

Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P.J.H., Christensen, J.H., Cuvelier, C., Foltescu, V., Graff, a., 2009. Skill and uncertainty of a regional air quality model ensemble. Atmos. Environ. 43, 4822–4832. doi:10.1016/j.atmosenv.2008.09.083

Vira, J., Sofiev, M., 2012. On variational data assimilation for estimating the model initial conditions and emission fluxes for short-term forecasting of SOx concentrations. Atmos. Environ. 46, 318–328. doi:10.1016/j.atmosenv.2011.09.066

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

Zhang, J., Reid, J.S., Westphal, D.L., Baker, N.L., Hyer, E.J., 2008. A system for operational aerosol optical depth data assimilation over global oceans. J. Geophys. Res. 113. doi:10.1029/2007JD009065

Table 1. Correlation length scales *L* (km) diagnosed from the NMC dataset.

	UTC hour			
Species	00	06	12	18
O3	45.5	51.0	57.6	59.5
NO2	35.8	39.0	41.1	42.3

Table 2. The χ^2 / N consistency indicator and RMSE on rural MACC validation stations during the first and fifth iteration for tuning the observation and background error standard deviations.

		O3		NO2	
		χ^2 / N	RMSE	χ^2 / N	RMSE
June	First guess	0.86	20.94	0.39	6.14
	5th iteration	1.05	18.93	1.16	5.80
December	First guess	0.74	17.39	1.20	9.91
	5th iteration	1.05	16.89	1.14	9.54

Table 3. Comparison of performance indicators for ozone in the 2012 reanalysis. The scores are given for station sets "MACC" and "EMEP" as defined in Section 2.2. For the analysis runs, scores are shown for the different background error covariance matrices discussed in Section 3.

		Hourly			Daily maximum		
		Corr	Bias	RMSE	Corr	Bias	RMSE
MACC	No assimilation	0.67	-4.00	24.91	0.80	-11.39	22.09
	Assimilation, first guess B	0.77	-4.62	21.35	0.86	-2.71	15.51
	Assimilation, final B	0.8	-4.64	19.2	0.9	-7.4	14.52
EMEP	No assimilation	0.58	-6.32	24.06	0.71	-12.11	22.00
	Assimilation, first guess B	0.66	-5.79	21.83	0.77	-5.32	17.96
	Assimilation, final B	0.68	-6.00	20.22	0.8	-9.57	17.15

Table 4. Comparison of performance indicators for NO2 in the 2012 reanalysis. ~~Station~~The station sets MACC and EMEP and assimilation options are as in Table 3.

		Hourly			Daily maximum		
		Corr	Bias	RMSE	Corr	Bias	RMSE
MACC	No assimilation	0.50	-1.18	9.01	0.53	-3.41	13.58
	Assimilation, first guess B	0.58	-0.25	8.6	0.61	-0.96	12.78
	Assimilation, final B	0.6	-0.38	8.04	0.63	-2.35	12.01
EMEP	No assimilation	0.52	0.47	6.19	0.55	-0.02	9.17
	Assimilation, first guess B	0.55	1.17	6.45	0.59	1.75	9.63
	Assimilation, final B	0.57	0.99	5.92	0.6	0.74	8.66

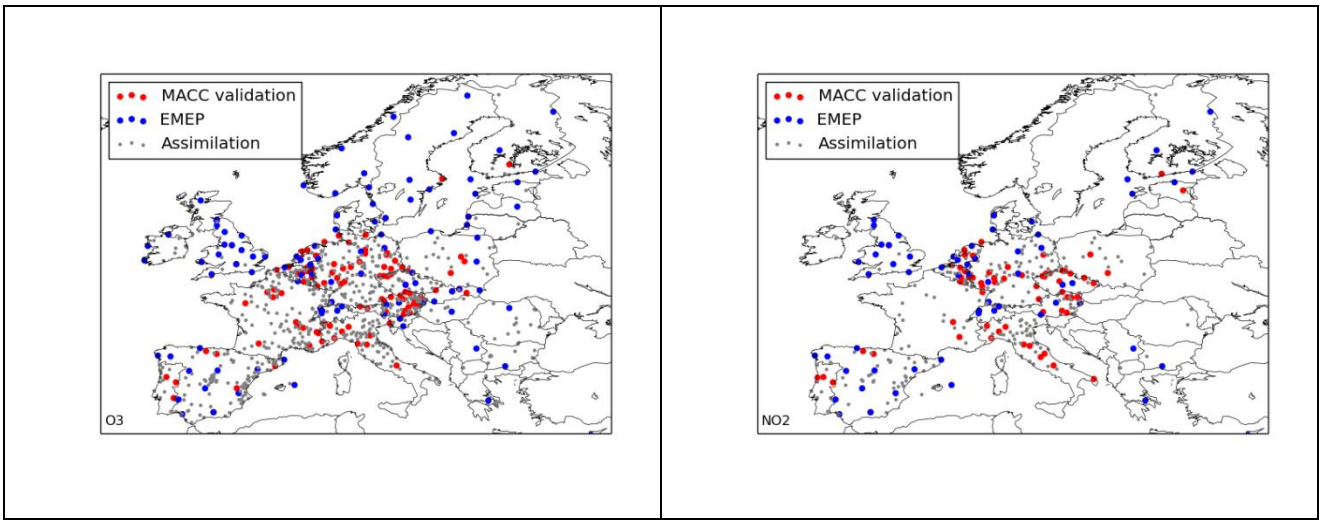


Figure 1. The stations networks used for assimilation and validation for O3 (left) and NO2 (right). The assimilation stations for O3 include rural and suburban stations, for NO2 only rural stations. For validation, only rural stations are shown. The red and blue colours refer to the MACC validation and EMEP stations subsets.

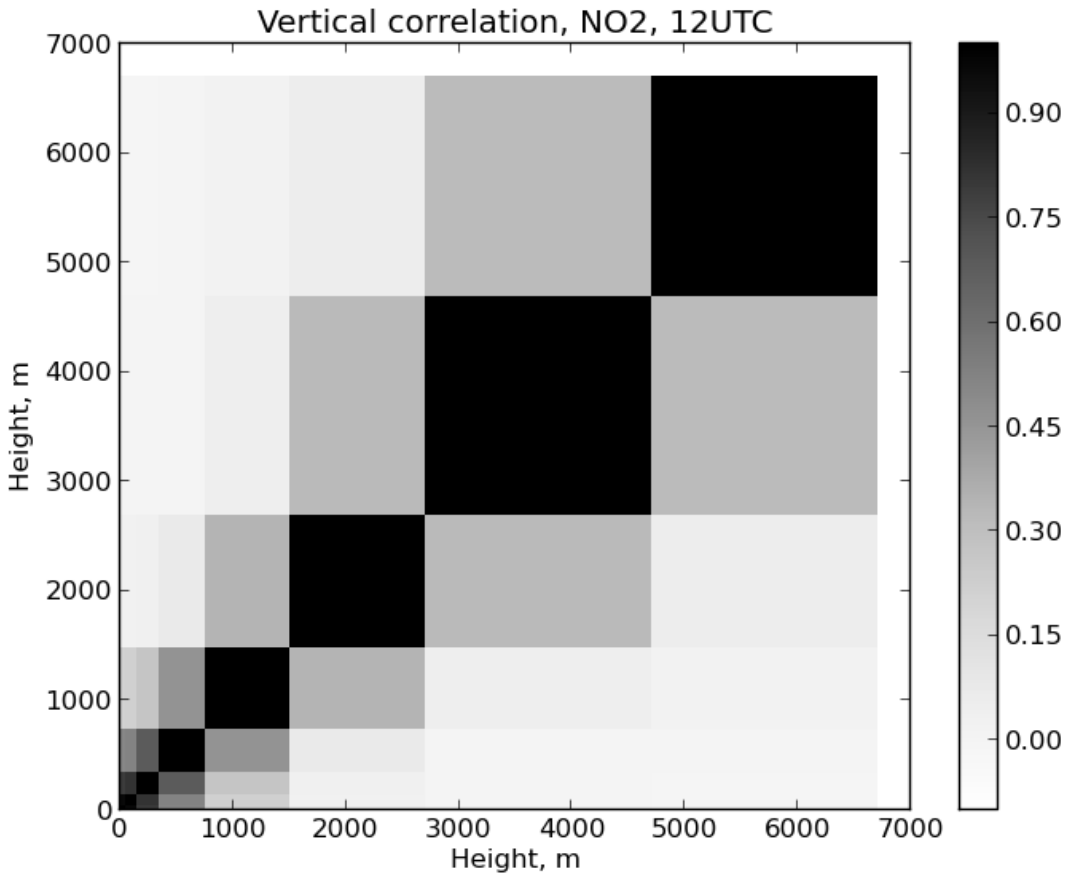
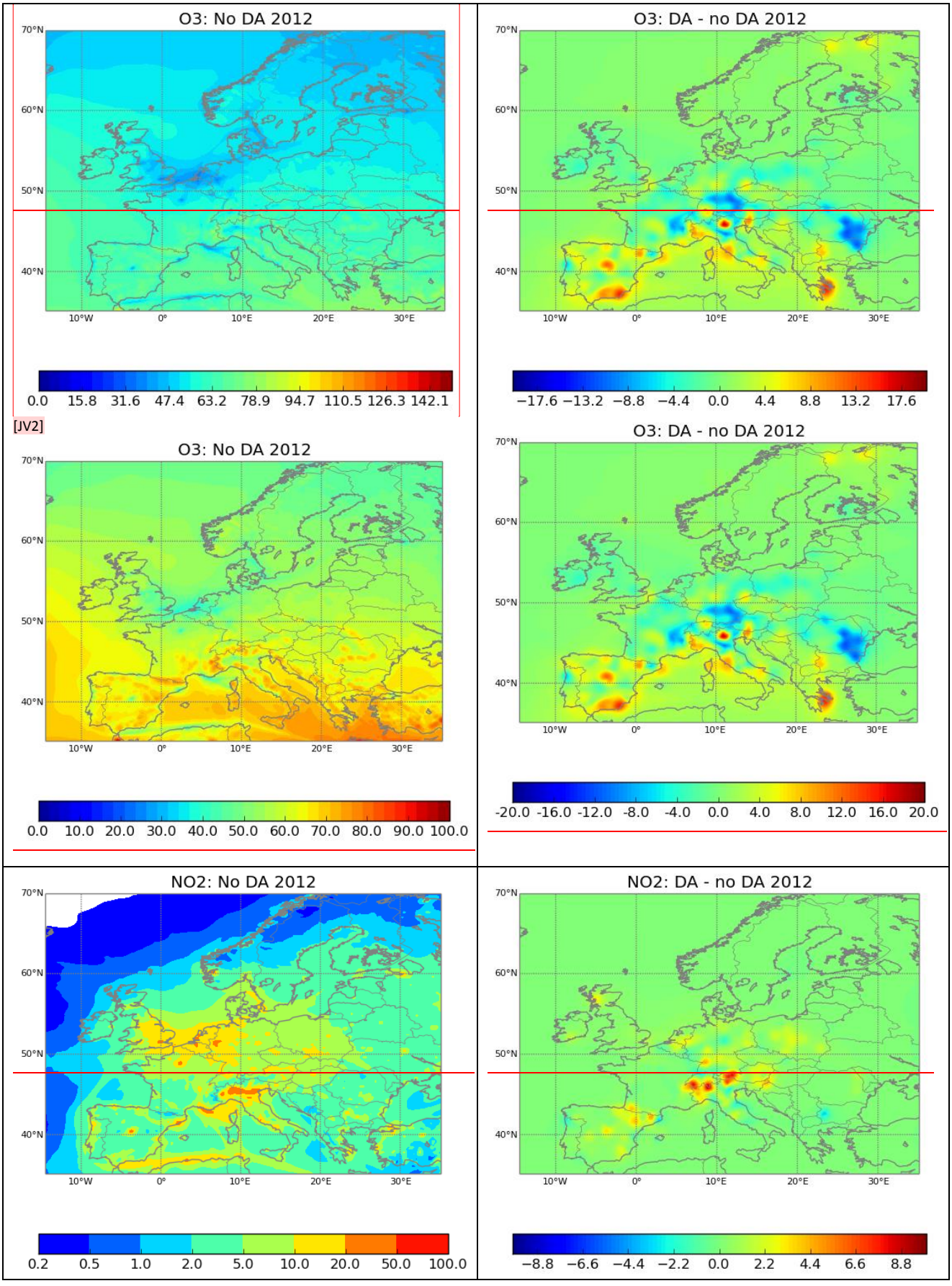


Figure 2. Vertical correlation function for NO2 at 12UTC.


O3: No DA 2012
70°N
60°N
50°N
40°N
10°W
0°
10°E
20°E
30°E
0.0 15.8 31.6 47.4 63.2 78.9 94.7 110.5 126.3 142.1
O3: DA - no DA 2012
−17.6 −13.2 −8.8 −4.4 0.0 4.4 8.8 13.2 17.6
[JV2]
O3: No DA 2012
0.0 10.0 20.0 30.0 40.0 50.0 60.0 70.0 80.0 90.0 100.0
O3: DA - no DA 2012
-20.0 -16.0 -12.0 -8.0 -4.0 0.0 4.0 8.0 12.0 16.0 20.0
NO2: No DA 2012
0.2 0.5 1.0 2.0 5.0 10.0 20.0 50.0 100.0
NO2: DA - no DA 2012
−8.8 −6.6 −4.4 −2.2 0.0 2.2 4.4 6.6 8.8

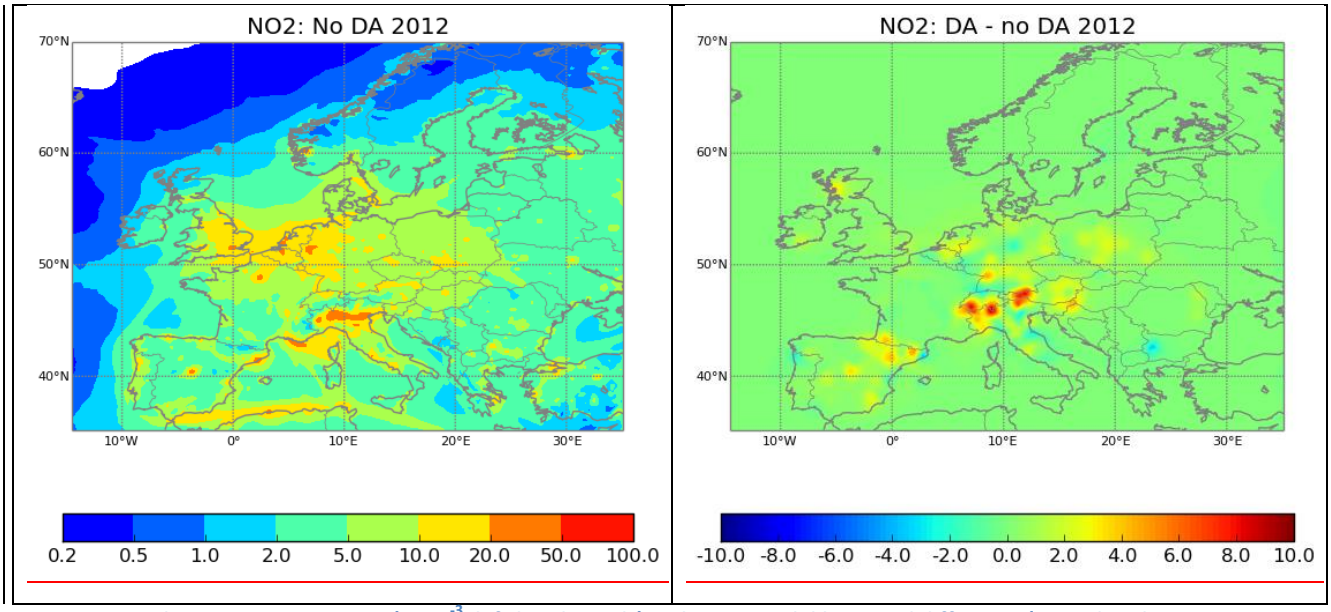


Figure 3. Yearly mean concentration (µg m^{-3}, left-hand panels) on lowest model layer and difference (assimilated – not assimilated, right-hand panels) due to assimilation of O3 (top panels) and NO2 (bottom panels).

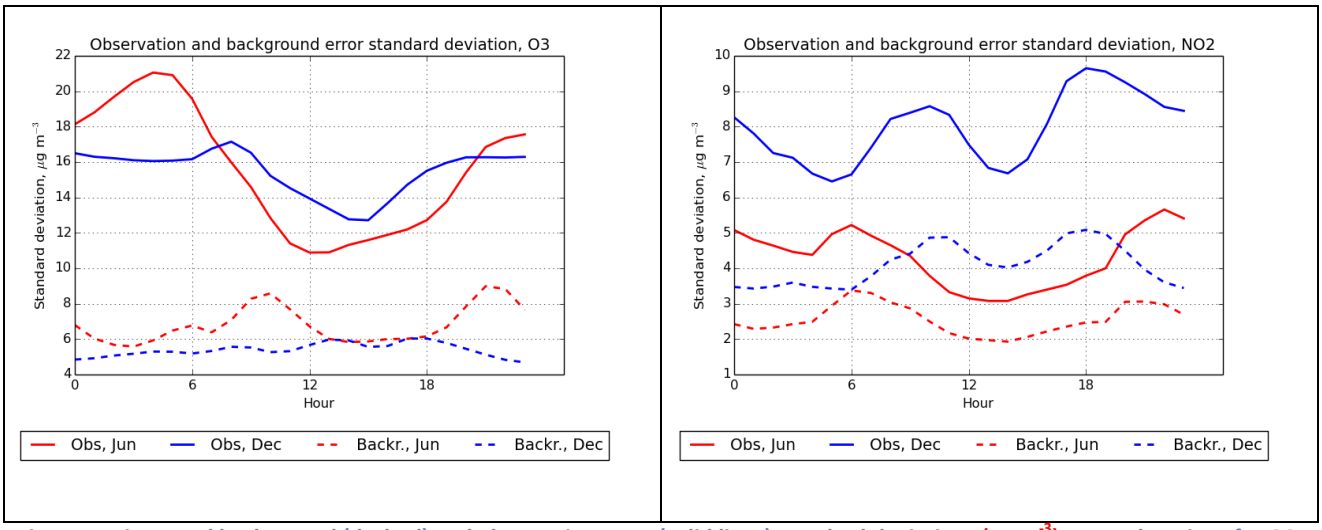


Figure 4. Diagnosed background (dashed) and observation error (solid lines) standard deviations (µg m^{-3}) on rural stations for O3 (left) and NO2 (right). Red lines correspond to the calibration made for June 2011, blue lines correspond to calibration for December 2011.

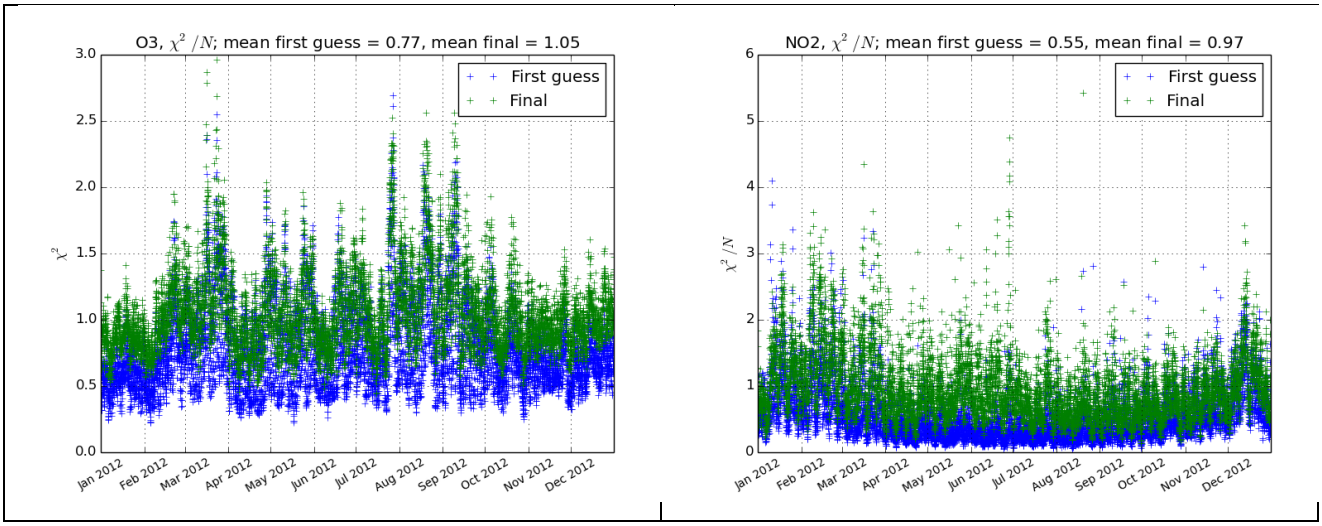


Figure 5. The χ^2 / N_{obs} consistency indicator for hourly analyses of O3 (left) and NO2 (right). The values in blue and green are shown for the first-guess and final assimilation setups, respectively. Note the different scales for O3 and NO2.

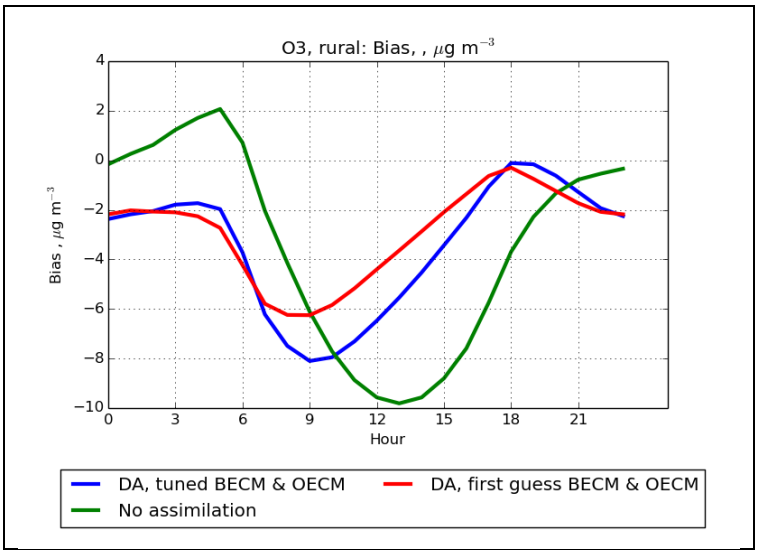


Figure 6. Diurnal variation of model bias (µg m^{-3}). The first guess assimilation setup is shown in red and the final setup in blue. The reference run with no assimilation is drawn in green. The values are shown for the rural MACC validation stations and averaged over each day of year 2012 and over the stations.

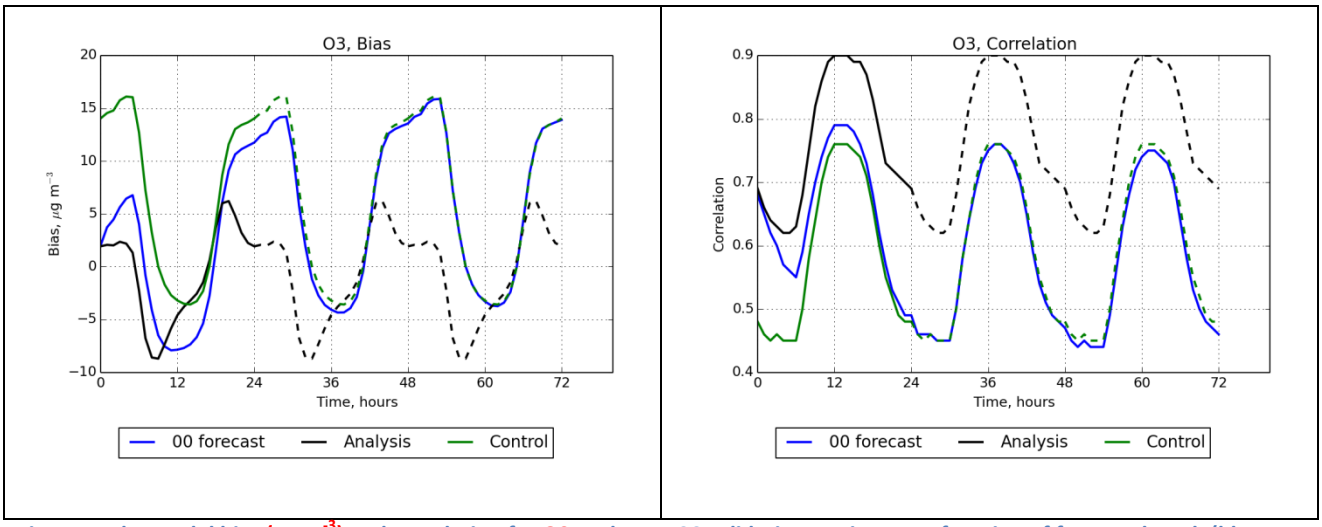


Figure 7. The model bias (μg m^{-3}) and correlation for O3 at the MACC validation stations as a function of forecast length (blue lines). The corresponding indicators the analyses (black) and control run (green) are shown averaged by time of day and replicated over the forecast window.

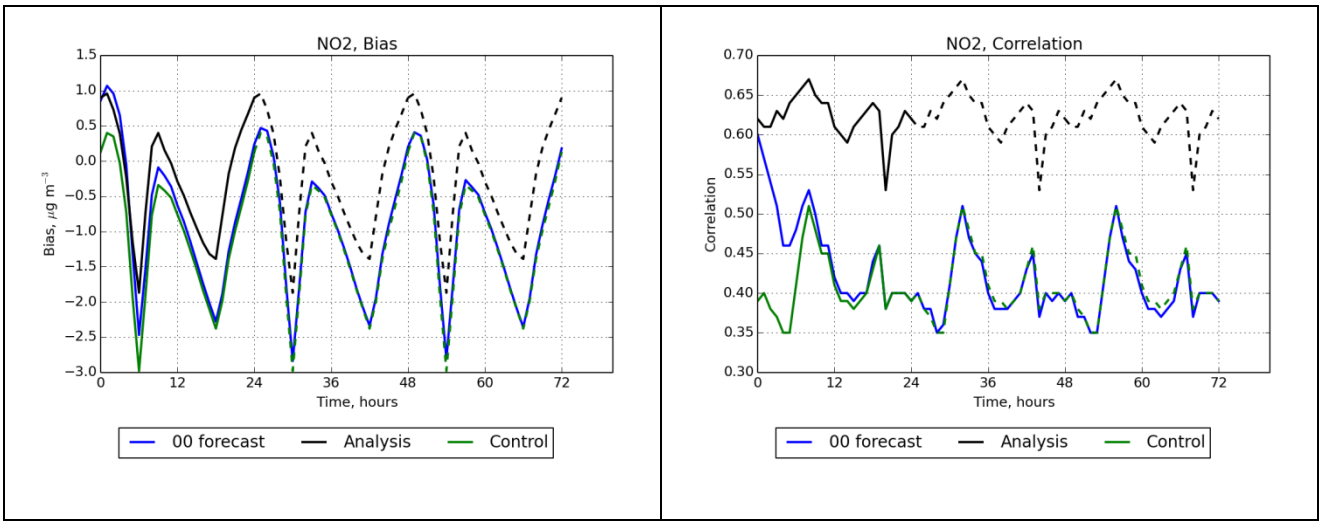


Figure 8. As, but for NO2.