Geoscientific Model Development Discussions

Open Access

Assimilation of surface NO$_2$ and O$_3$ observations into the SILAM chemistry transport model

J. Vira and M. Sofiev

Finnish Meteorological Institute, Helsinki, Finland

Correspondence to: J. Vira (julius.vira@fmi.fi)

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

GMDD

7, 5589–5621, 2014

Assimilation of surface NO$_2$ and O$_3$ observations into the SILAM chemistry transport model

J. Vira and M. Sofiev

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀ | ▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

This paper describes assimilation of trace gas observations into the chemistry transport model SILAM using the 3D-Var method. Assimilation results for year 2012 are presented for the prominent photochemical pollutants ozone (O_3) and nitrogen dioxide

5 (NO_2). 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 condi-

10 tions, 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

15 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 O_3 and from 0.53 to 0.63 for NO_2.

1 Introduction

During the last 10–15 years, assimilating observations into atmospheric chemistry

20 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 compo-

25 sition has permitted construction of global analysis and forecasting systems such as those described by Benedetti et al. (2009) and Zhang et al. (2008).

Data assimilation is classically (e.g. Kalnay, 2003) defined 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 and 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-Var, 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 in previous studies on chemical data assimilation. The "National Meteorological Centre" (NMC) method introduced by Parrish and 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 and Lönnberg (1986).

Assimilation of surface NO$_2$ and O$_3$ observations into the SILAM chemistry transport model

J. Vira and M. Sofiev

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

The observational method was used in assimilation of NO_2 and O_3 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 the effect of adjusted initial condition to extend for up to 24 h. Among other reactive gases, NO_2 has been a subject for studies of Silver et al. (2013) and Wang et al. (2011). However, the shorter lifetime of NO_2 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 NO_2 and O_3 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 definitions. 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.

2 Materials and methods

This section presents the SILAM dispersion model, the utilized observation datasets, 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 (Gery et al., 1989). However, the NO_2 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 NO_2 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

J. Vira and M. Sofiev

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° 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 h 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 00:00 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°. 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 min for both setups.

2.2 Observations

This study utilises the hourly observations of NO$_2$ and O$_3$ 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.

J. Vira and M. Sofiev

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 sta- tions include about third of the available background stations for each species, and are chosen with the requirement to cover the same area as the assimilation stations. How- ever, 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 rep- resentativeness errors, data from urban stations are not assimilated, and for NO$_2$, also suburban stations are excluded. For ozone, the data from suburban stations are assim- ilated, however, the observation errors are assessed separately for suburban and rural stations, as outlined in Sect. 3. The station sets are presented on a map in Fig. 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 evalu- ate 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 x_a minimises the cost function

$$J(x) = \frac{1}{2}(y - \mathcal{H}(x))^\top R^{-1}(y - \mathcal{H}(x)) + \frac{1}{2}(x - x_b)^\top B^{-1}(x - x_b), \tag{1}$$

where x_b is the background state, y is the vector of observations, and $\mathcal{H}$ is the ob- servation operator. The uncertainties of the background state x_b and the observations y are described by the background and observation error covariance matrices B and

R, respectively. In this study, the control variable x consists of the three-dimensional airborne concentration field for either NO_2 or ozone. The m1qn3 minimization code (Gilbert and Lemaréchal, 1989) is used for solving the optimisation problem (Eq. 1).

For the surface measurements, the operator $\mathcal{H}$ is linear and consists of horizontal
5 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.

3 Background and observation error covariance matrices

10 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
15 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
20 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 h 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.
25 Since no chemical data assimilation is used in the forecasts, the differences are due to changes in forecasted 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 00:00, 06:00, 12:00 and 18:00 UTC, and the correlations are interpolated for all other times of day.

The horizontal and vertical components of the correlation matrix **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 $\tilde{\mathbf{C}}$ of the forecast differences is calculated. Then, the Gaussian correlation function is fitted to the empirical correlations $\tilde{\mathbf{C}}_{ij}$ by minimizing

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

where the fitted correlation function is $\mathbf{C}_{ij}(L) = \exp(-(|x_i - x_j|^2 + |y_i - y_j|^2)/L^2)$. 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 NO_2 at 12:00 UTC is shown in Fig. 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 $\sigma_{\text{obs}}^{(i)}$ of the ith observation component is equal to

$$E[(\boldsymbol{y}^{(i)} - \boldsymbol{y}_{\text{a}}^{(i)})(\boldsymbol{y}^{(i)} - \boldsymbol{y}_{\text{b}}^{(i)})] = \sigma_{\text{obs}}^{(i)^2}, \tag{3}$$

where E denotes the expectation, $\boldsymbol{y}$ is the observation vector and $\boldsymbol{y}_{\text{a}} = \mathcal{H}(\boldsymbol{x}_{\text{a}})$ and $\boldsymbol{y}_{\text{b}} = \mathcal{H}(\boldsymbol{x}_{\text{b}})$ are evaluated from the analysis and background fields, respectively.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

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 (i.e. horizontal interpolation is neglected), then

$$E[(\boldsymbol{y}_a^{(i)} - \boldsymbol{y}_b^{(i)})(\boldsymbol{y}^{(i)} - \boldsymbol{y}_b^{(i)})] = {\sigma_b^{(i)}}^2. \tag{4}$$

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, Eqs. (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 Eqs. (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, the observation error covariance matrix $\mathbf{R}$ is assumed diagonal. The initial values for σ_{obs} and σ_b were set to 11.2 and 20.6 $\mu g\,m^{-3}$ for O$_3$, and 4.0 and 8.0 $\mu g\,m^{-3}$ for NO$_2$. 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 O$_3$ and NO$_2$. Both σ_{obs} and σ_b are segregated by hour, while for O$_3$ 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)

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

where $\chi^2 = 2J(x_a)$ is twice the value of cost function (Eq. 1) at the minimum, and N is dimension of the observation vector y. The identity (Eq. 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 lowest model level is shown in Fig. 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 NO_2 levels are confined to somewhat more limited areas; in particular areas near major mountain ranges (Alps and Pyrenees) show enhanced NO_2 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 O$_3$ and NO$_2$ are shown in Fig. 4. For June, the standard deviations 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 not reduced, in contrast to the general seasonality of O$_3$.

Especially for summertime night conditions, the values are higher than the values adapted 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 09:00–10:00 and 21:00–22:00 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 NO$_2$ varies between 2.8 and 5.2 µg m^3 for rural stations. Suburban or urban stations were not assimilated for NO$_2$. Contrary to ozone, the diurnal variation of background and observation errors both 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 Fig. 6, 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 NO$_2$, 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 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 O$_3$ and NO$_2$, 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 NO$_2$ 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 for NO$_2$, which is consistent with the shorter lifetime of NO$_2$ compared to O$_3$. The shorter lifetime would make the MACC validation stations, which are generally located closer to the assimilation stations, more sensitive to the assimilation.

For ozone, the assimilation had uneven effect on the model bias. While the correlation and RMSE were always improved by assimilation, the analyses have slightly larger

J. Vira and M. Sofiev

negative mean bias (-4.6 vs. $-4.0\,\mu g\,m^3$ on MACC stations) than the free model. This is confirmed by the average diurnal profile shown in Fig. 5. However, the diurnal variation of analysis errors is flatter, and the strongest bias no longer coincides with the afternoon hours, when the highest O$_3$ concentrations are typically observed.

5   The analysis biases for O$_3$ are not surprising given the similar bias in the free-running simulation, since the analysis scheme assumes an unbiased model. This also explains why assimilation setup including tuned OECM and BECM produces more biased analyses compared to the first-guess setup, as seen in Fig. 5. As shown by Dee (2005), such issues can in principle be addressed by the assimilation system. However, a possible

10   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.

15   The O$_3$ and NO$_2$ observations were assimilated into separate model runs. Assimilation of O$_3$ had only minor influence on NO$_2$ 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 NO$_2$ assimilation on ozone fields was not evaluated in this study.

20   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 00:00 UTC as initial conditions. The forecast experiment covered time between 16 July and 5 August 2012.

25   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 O_3 and NO_2 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 h of forecast. Also, the forecast initialized at 00:00 UTC from 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 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 Fig. 7 are computed for the MACC station network; similar impact is observed on the EMEP stations.

Due to the shorter chemical lifetime, the effect of initial condition on forecasts of NO_2 can be expected to vanish 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 NO_2 concentration to relax to its background values within 3–4 h.

In the forecast experiments performed within this study, the effect of assimilation on NO_2 forecast scores was limited to the first 6 forecast hours, which coincides with the night in most of the domain. Hence, under summertime conditions, the analyses are only marginally useful for improving forecasts of NO_2.

5 Conclusions

An assimilation system coupled to the SILAM chemistry transport model has been described along with its application in reanalysis of ozone and NO_2 concentrations

for year 2012. Furthermore, the impact of using the O$_3$ and NO$_2$ 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 assimilation. For daily maximum values, the correlation coefficient is improved over the free running model from 0.8 to 0.9 for O$_3$ and from 0.53 to 0.63 for NO$_2$ on rural validation stations. The respective biases are also decreased, however, a bias of $-7.4\,\mu g\,m^{-3}$ remains in the O$_3$ 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 h. For NO$_2$, the improvement was limited to a window of 6 h. These findings are similar to the results published in previous studies.

The diagnosed observation error standard deviations for ozone have a strong diurnal variation, and reach up to about $21\,\mu g\,m^{-3}$ during night. These values are higher than usually assumed in chemical data assimilation, but corroborate 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.

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., and Suttie, M.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J. Geophys. Res., 114, D13205, 2009.

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

Bocquet, M.: Parameter-field estimation for atmospheric dispersion: application to the Chernobyl accident using 4D-Var, Q. J. Roy. Meteor. Soc., 138, 664–681, 2012.

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., and Williams, E.: Four-dimensional data assimilation experiments with International Consortium for Atmospheric Research on transport and transformation ozone measurements, J. Geophys. Res., 112, 1–18, 2007.

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

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

Dee, D. P.: Bias and data assimilation, Q. J. Roy. Meteor. Soc., 131, 3323–3343, 2005.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Q. J. Roy. Meteor. Soc., 131, 3385–3396, 2005.

EEA: Air Pollution by Ozone Across Europe During Summer 2012, EEA Technical Report, 2013.

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

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

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

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynam., 53, 343–367, 2003.

Galperin, M.: The approaches to correct computation of airborne pollution advection, Gidrometeoizdat, XVII, 54–68, 2000.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Gandin, L. S.: Objective Analysis of Meteorological Fields, Gidrometeorologischeskoe Iz-datel'stvo, Translated (1965) by Israel Programme for Scientific Translation, Jerusalem, Leningrad, 1963.

Gaubert, B., Coman, A., Foret, G., Meleux, F., Ung, A., Rouil, L., Ionescu, A., Candau, Y., and Beekmann, M.: 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, 2014.

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

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

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

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., Kiout-sioukis, I., Friese, E., Amstrup, B., Bergstrom, R., Strunk, A., Vira, J., Zyryanov, D., Mau-rizi, A., Melas, D., Peuch, V.-H., and Zerefos, C.: Comparison of OMI NO_2 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, 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., Huij-nen, 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., and the MACC team: The MACC reanalysis: an 8 yr data set of atmospheric composition, Atmos. Chem. Phys., 13, 4073–4109, doi:10.5194/acp-13-4073-2013, 2013.

Jaumouillé, E., Massart, S., Piacentini, A., Cariolle, D., and Peuch, V.-H.: 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, 2012.

Kahnert, M.: Variational data analysis of aerosol species in a regional CTM: background er-ror covariance constraint and aerosol optical observation operators, Tellus B, 60, 753–770, 2008.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predicability, Cambridge University Press, 2003.

Kuenen, J. J. P., Visschedijk, A. J. H., Jozwicka, M., and Denier van der Gon, H. A. C.: 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, 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., and Eben, K.: 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, 2012.

Kumar, U., Ridder, K. De, Lefebvre, W., and Janssen, S.: Data assimilation of surface air pollutants (O_3 and NO_2) in the regional-scale air quality model AURORA, Atmos. Environ., 60, 99–108, 2012.

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., and Zakey, A.: 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, 2012.

Le Dimet, F.-X. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, Tellus A, 38, 97–110, 1986.

Lorenc, A. C.: Analysis methods for numerical weather prediction, Q. J. Roy. Meteor. Soc., 112, 1177–1194, 1986.

Massart, S., Piacentini, A., and Pannekoucke, O.: Importance of using ensemble estimated background error covariances for the quality of atmospheric ozone analyses, Q. J. Roy. Meteor. Soc., 138, 889–905, 2012.

Ménard, R., Cohn, S. E., Chang, L.-P., and Lyster, P. M.: Assimilation of stratospheric chemical tracer observations using a Kalman Filter, Part I: Formulation, Mon. Weather Rev., 128, 2654–2671, 2000.

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

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Poupkou, A., Giannaros, T., Markakis, K., Kioutsioukis, I., Curci, G., Melas, D., and Zerefos, C.: A model for European Biogenic Volatile Organic Compound emissions: software development and first validation, Environ. Modell. Softw., 25, 1845–1856, 2010.

Rabier, F.: Overview of global data assimilation developments in numerical weather-prediction centres, Q. J. Roy. Meteor. Soc., 131, 3215–3233, 2005.

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

Schwartz, C. S., Liu, Z., Lin, H.-C., and McKeen, S. A.: Simultaneous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth. J. Geophys. Res., 117, D13202, doi:10.1029/2011JD017383, 2012.

Schwinger, J. and Elbern, H.: 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, 1–19, 2010.

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., and Kukkonen, J.: 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, 2013.

Silver, J. D., Brandt, J., Hvidberg, M., Frydendall, J., and Christensen, J. H.: 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, 2013.

Sofiev, M.: A model for the evaluation of long-term airborne pollution transport at regional and continental scales, Atmos. Environ., 34, 2481–2493, 2000.

Sofiev, M.: Extended resistance analogy for construction of the vertical diffusion scheme for dispersion models, J. Geophys. Res., 107, ACH10.1–ACH10.8, doi:10.1029/2001JD001233, 2002.

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

Sofiev, M., Genikhovich, E., Keronen, P., and Vesala, T.: Diagnosing the surface layer parameters for dispersion models within the meteorological-to-dispersion modeling interface, J. Appl. Meteorol. Clim., 49, 221–233, 2010.

Sofiev, M., Soares, J., Prank, M., de Leeuw, G., and Kukkonen, J.: A regional-to-global model of emission and transport of sea salt particles in the atmosphere, J. Geophys. Res., 116, D21302, doi:10.1029/2010JD014713, 2011.

Assimilation of surface NO$_2$ and O$_3$ observations into the SILAM chemistry transport model

J. Vira and M. Sofiev

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

◄◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

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., Jerièeviæ, A., Kraljeviæ, 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., and Galmarini, S.: Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII, Atmos. Environ., 53, 60–74, 2012.

Tombette, M., Mallet, V., and Sportisse, B.: PM_{10} data assimilation over Europe with the optimal interpolation method, Atmos. Chem. Phys., 9, 57–70, doi:10.5194/acp-9-57-2009, 2009.

Wang, X., Mallet, V., Berroir, J., and Herlin, I.: Assimilation of OMI NO_2 retrievals into a regional chemistry-transport model for improving air quality forecasts over Europe, Atmos. Environ., 45, 485–492, 2011.

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

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

Zhang, J., Reid, J. S., Westphal, D. L., Baker, N. L., and Hyer, E. J.: A system for operational aerosol optical depth data assimilation over global oceans, J. Geophys. Res., 113, D10208, doi:10.1029/2007JD009065, 2008.

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

Species	UTC hour			
	00:00	06:00	12:00	18:00
O$_3$	45.5	51.0	57.6	59.5
NO$_2$	35.8	39.0	41.1	42.3

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

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.

		O$_3$		NO$_2$	
		χ^2/N	RMSE	χ^2/N	RMSE
Jun	First guess	0.86	20.94	0.39	6.14
	5th iteration	1.05	18.93	1.16	5.80
Dec	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 Sect. 2.2. For the analysis runs, scores are shown for the different background error covariance matrices discussed in Sect. 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 NO$_2$ in the 2012 reanalysis. 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
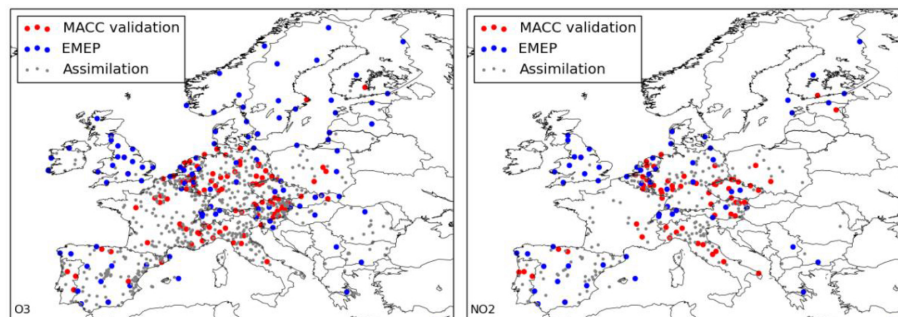
Figure 1. The stations networks used for assimilation and validation for O_3 (left) and NO_2 (right). The assimilation stations for O_3 include rural and suburban stations, for NO_2 only rural stations. For validation, only rural stations are shown.
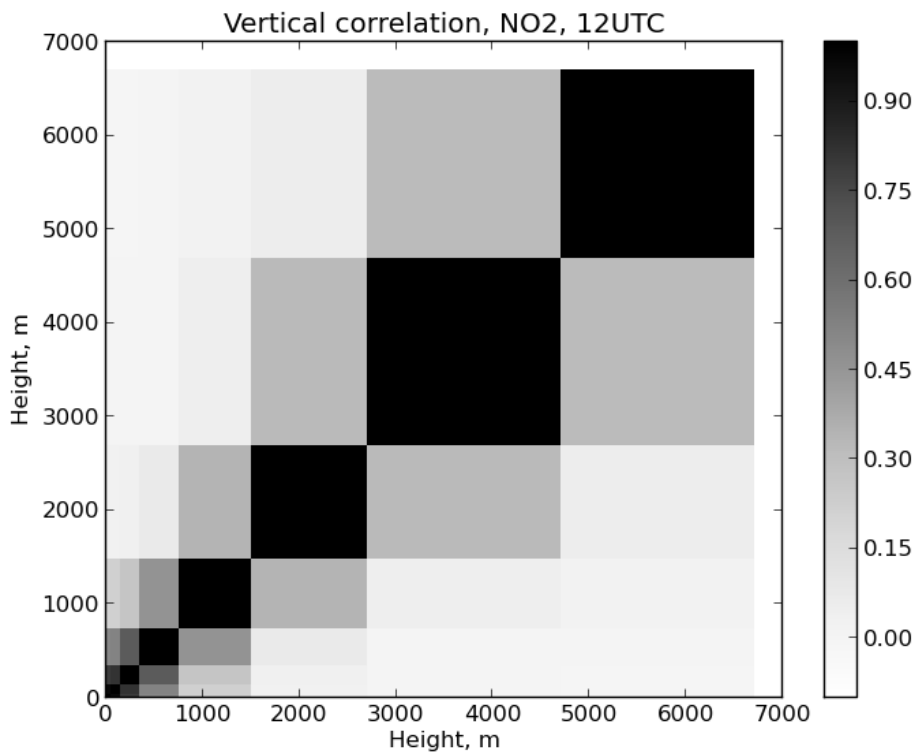
Figure 2. Vertical correlation function for NO_2 at 12:00 UTC.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
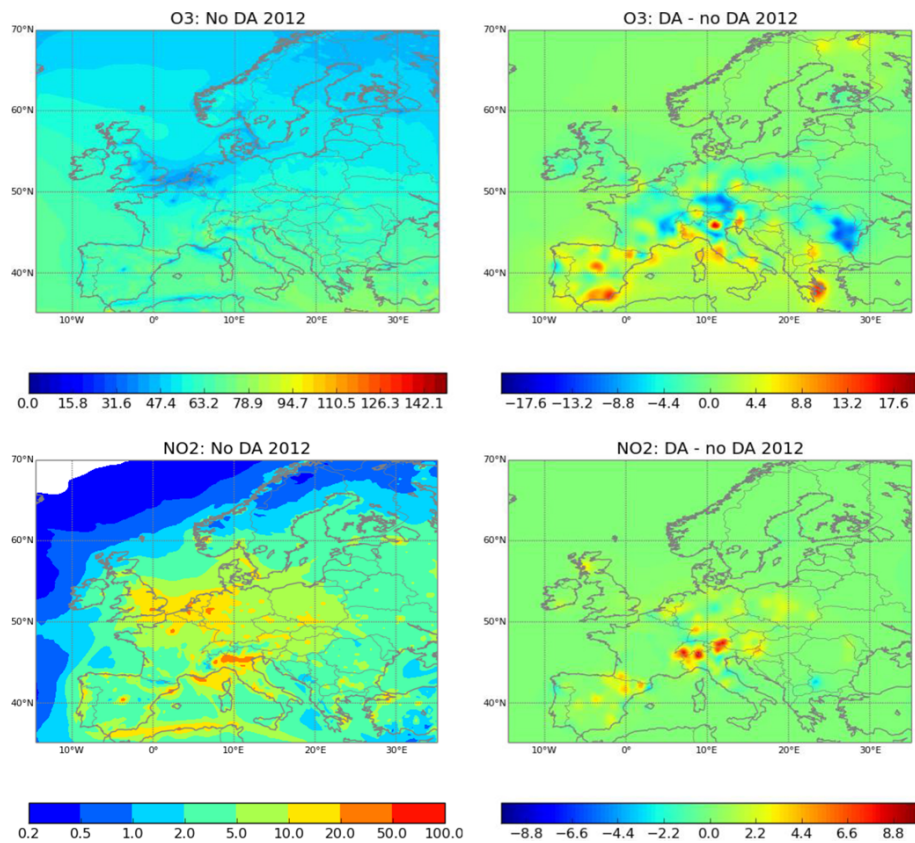
Figure 3. Yearly mean concentration ($\mu g\,m^{-3}$, left-hand panels) on lowest model layer and difference (assimilated – not assimilated, right-hand panels) due to assimilation of O_3 (top panels) and NO_2 (bottom panels).

Assimilation of surface NO$_2$ and O$_3$ observations into the SILAM chemistry transport model

J. Vira and M. Sofiev
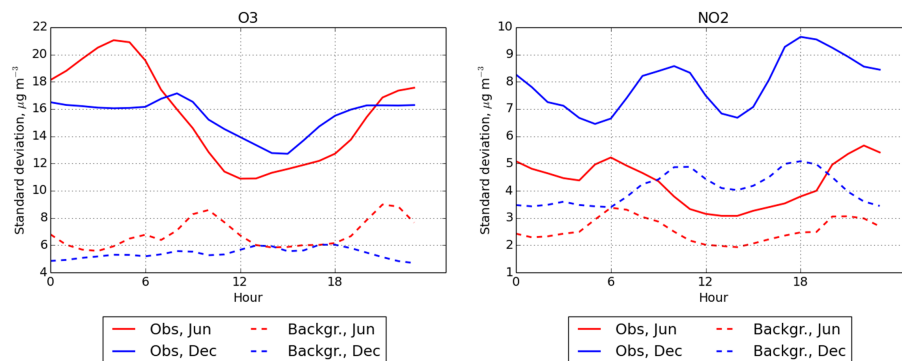
Figure 4. Diagnosed background (dashed) and observation error (solid lines) standard deviations on rural stations for O$_3$ (left) and NO$_2$ (right). Red lines correspond to the calibration made for June 2011, blue lines correspond to calibration for December 2011.

Printer-friendly Version

Interactive Discussion
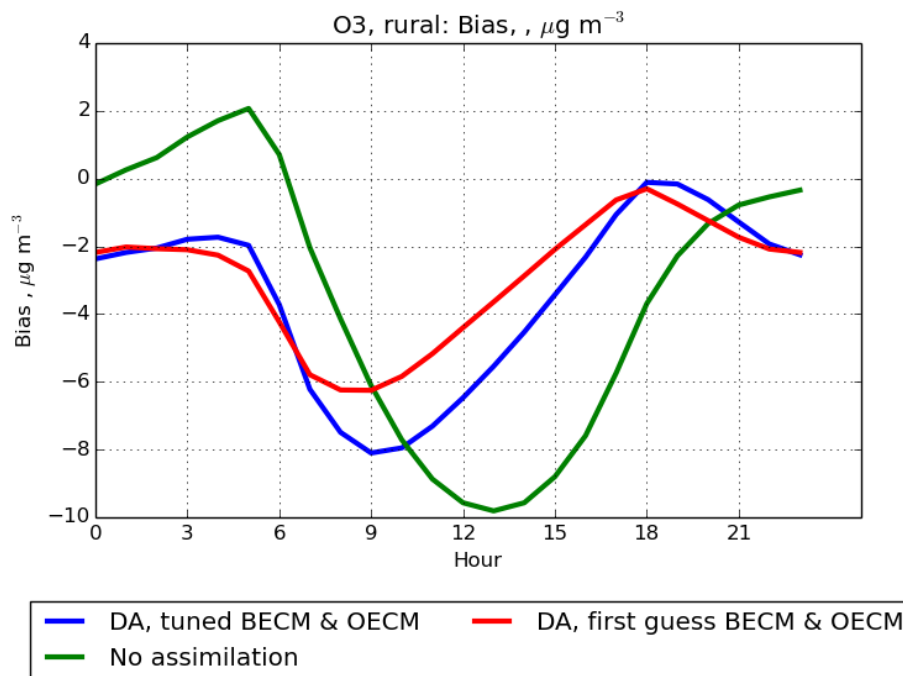
Figure 5. Diurnal variation of model bias. The first guess assimilation setup is shown in red and the final setup in blue. The reference run with no assimilation is drawn in green. The values are shown for the rural MACC validation stations and averaged over each day of year 2012 and over the stations.
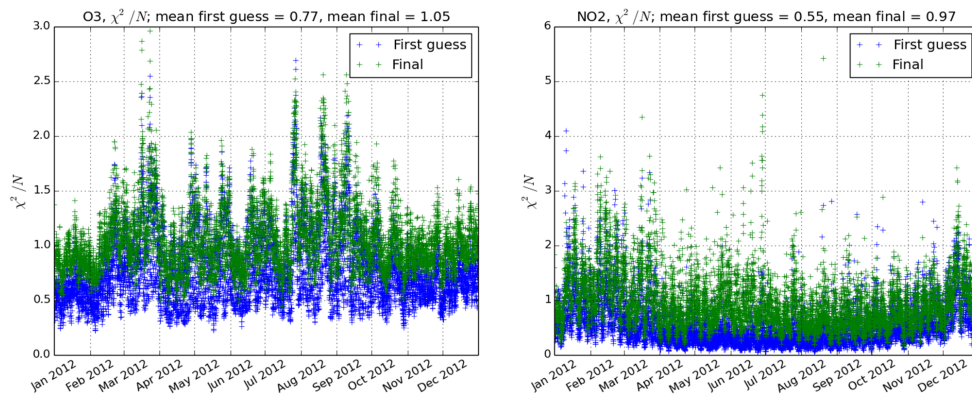
Figure 6. The χ^2/N_{obs} consistency indicator for hourly analyses of O_3 (left) and NO_2 (right). The values in blue and green are shown for the first-guess and final assimilation setups, respectively. Note the different scales for O_3 and NO_2.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀◀ | ▶▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
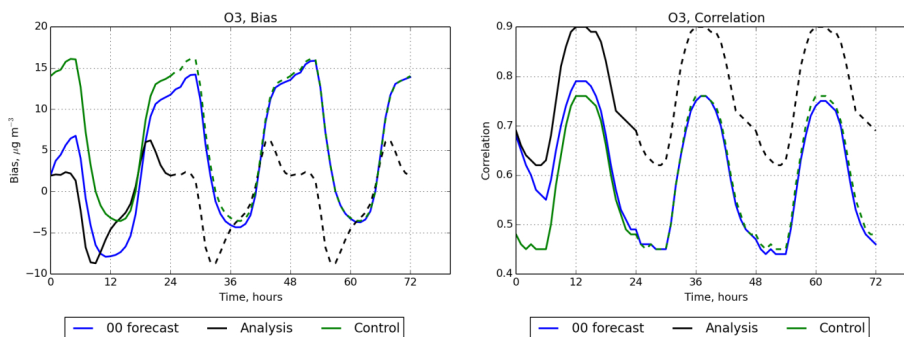
Figure 7. The model bias and correlation for the MACC validation stations as a function of forecast length (blue lines). The corresponding indicators the analyses (black) and control run (green) are shown averaged by time of day and replicated over the forecast window.
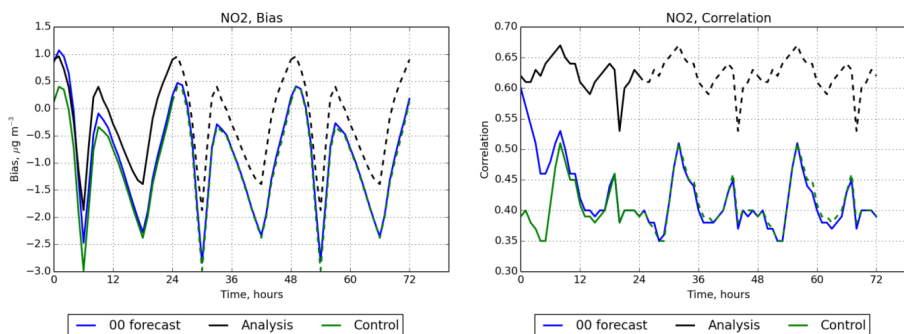
Figure 8. As Fig. 7, but for NO$_2$.