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/m^3 for hourly values measured at MACC stations but +0.47 μg/m^3 for EMEP stations. For the MACC stations, the assimilation reduces the bias to -0.38 μg/m^3, while the bias at EMEP stations is increased to 0.99 μg/m^3. 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/m^3 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