Author comments in reply to two anonymous referees on "Global NOx emission estimates derived from an assimilation of OMI tropospheric NO_2 columns" by K. Miyazaki et al.

We want to thank both referees for their helpful comments and suggestions. We have revised the paper according to the reviewer's comments, and hope that the revised manuscript is now suitable for publication. Below in italics are the referee comments with our replies in normal font.

Reply to Referee #1

the readability of the article is seriously hampered by its excessive length. Discussion could be more synthetic and insightful. At many instances, it is either limited to a simple reporting of the content of the figures with too much detail (e.g. Sect. 3.2, 3.3, 5, 6) or it consists of commonplace statements which do not bring essential information. I strongly recommend that authors make an effort to increase the concise- ness of their text and remove speculative/unproved arguments.

Based on the reviewer's comment, the length of the manuscript has been shortened. Several qualitative arguments (see below) have been removed from the manuscript. We believe that the revised manuscript is more exact and provides more insight.

List of deleted sentences/paragraphs:
P. 31542 line 10
The first paragraph of Section 3.3 (half)
  P. 31543 line 22-24
  P. 31543 line 24-27
  P. 31543 line 27- P.31544 line 1
The first paragraph of Section 3.3 (all)
P. 31547 line 24-25
P. 31548 line 23-24
P. 31549 line 27-28
P. 31551 line 14-15
P. 31551 line 25-27
P. 31553 line 20-22
P. 31536 line 3-4

Comments:

1. *Could the emission increases over several regions (e.g. China, Eastern US, S. Africa) inferred by the assimilation (cf. Table 3) be driven by the general underestimation of tropospheric NO2 columns in the CHASER model as reported in the intercomparison paper of van Noije et al. 2006?*

Yes, the positive increments correspond to the general underestimation of tropospheric NO_2 columns in CHASER which was commonly revealed by van Noije et al. (2006). The following sentence has thus been added:

"These positive increments are consistent with the general underestimation of tropospheric NO_2 columns in CHASER, consistent with the results by van Noije et al. (2006)."

2. *To assess the validity of the assimilation results the a posteriori regional/global emissions should be compared with previously reported top-down estimates.*

The following sentences have been added to Section 5.1 to compare our estimates with previously reported estimates.

"In the regional scale, the a priori emissions estimated from this study generally show agreement with other top-down studies. The 0.465 TgN estimated over the Eastern United States (102-64◦W, 22- 50◦N) from the OMI observations for March 2006 (Boersma et al., 2008a) is comparable to our estimate of 0.500 TgN for the same period. The 0.73 TgN estimated over the United States (130-70◦ W, 25-50◦ N) from ICARTT observations for July 1-August 15, 2004 (Hudman et al., 2007) is slightly smaller than our estimates of 0.98 TgN for July 2005. The 7.72 TgN (8.0 TgN) estimated for July 2008 (January 2009) over east China (103.75-123.75◦E, 19- 45◦N) from OMI and GOME-2 observations (Lin and McElroy, 2010) are comparable to our estimates of 7.8 TgN (6.5 TgN) for July 2005 (January 2005). The 11.0 TgN/yr estimated over East Asia (80-150◦ E, 10-50◦ N) for July 2007 from OMI observations (Zhao and Wang, 2009) is comparable to our estimates of 11.8 TgN/yr for July 2005. Differences in analysis years, together with those in retrieval data and models used in the analysis, will contribute to the difference in NOx emission estimates (e.g., Jaegle´ et al., 2005). This will be further discussed in Section 6."

3. *Besides introducing the basic LETKF formulation, the article does not go deeply enough in the analysis of the individual steps needed to obtain the emission updates. Furthermore, more details are needed on the advantages of the LETKF method over EnKF, on the localization strategy and the covariance inflation. In addition, please explain also how the perturbations Xb (p. 31534, l. 14) are defined. In Equation 4, the covariance inflation parameter Δ as well as the particular type of inflation assumed here should be defined. In the same formula, I is the identity matrix, so*

it should be in bold roman style.

The first paragraph of Section 2.3.1 has been rewritten to describe the advantages of the LETKF over EnKF as follows:

"The data assimilation technique employed is a local ensemble transform Kalman filter (LETKF). There are two types in EnKF approaches, the perturbed observation (PO) method and the ensemble square root filter (SRF) method (e.g., Whitaker and Hamill, 2002). SRF methods generate an analysis ensemble mean and covariance that satisfy the Kalman filter equations for linear models (e.g., Ott et al., 2004), whereas PO methods introduce an additional source of sampling errors. The LETKF is related to the SRF method (e.g., Whitaker and Hamill, 2002), and it has conceptual and computational advantages over the original EnKF (e.g., Ott et al., 2004; Hunt et al., 2007; Kalnay, 2010). One of the advantages is that the LETKF performs the analysis locally in space and time, and reduces sampling errors caused by a limited ensemble size. In addition, the analyses at different grid points are performed independently, which reduces the computational cost because most calculations are performed in parallel in the LETKF (e.g., Miyoshi and Yamane, 2007)."

The following sentence has been also added to describe the advantages of the LETKF.

"calculations of large vectors or matrices with N dimension are not necessary to obtain the T matrix in the LETKF different from the original EnKF."

The following sentences have been added to Section 2.3.1 to explain the LETKF update processes.

"In summary, the LETKF analyzes variables (i.e., NOx emissions) for every grid point by choosing observations (i.e., OMI retrievals) that determine the observational space. Then, the analysis is solved independently at every grid point located at the local volume center using the observational information and background error covariance estimated from the ensemble forecast. The new global analysis ensemble of the variables (i.e., NOx emissions) is then obtained by combining the local analysis. The estimated emissions are used in the next step ensemble model simulations (after the forecast process) and updated at every analysis step. The forecast and analysis processes for NOx emissions are further described in Sections 2.3.2."

Several descriptions have been also added to explain the inflation more precisely. Eq. (8) has been added to define the X_b used in this study. "I" has been corrected.

4. *Section 3.2 on the seasonal variation of the NO2 columns is too long and can be easily shortened. Some of the sentences (e.g. p. 31543, l. 22-24, l. 27-28) are commonplace and should be omitted.*

Further, it is not true that there is an obvious underestimation in winter in Europe (as stated in page 31544, l. 3), at least when we compare with DOMINOv2 data. Please correct the text (p. 31544, l. 4).

Several sentences (see above list) have been removed from the original paper. The discussion of the underestimation in Europe has been corrected.

5. *The first paragraph of Section 3.3 could be shortened or even omitted. In p. 31544, second paragraph, it is stated that the use of diurnal profiles in CHASER improves the model-data agreement globally. However, the agreement worsens over Central Africa (Fig. 4b) after application of the assumed biomass burning profile. The authors should first give more details about this profile and make a point on this in the main text to explain the model behaviour in Fig. 4b.*

The first paragraph of Section 3.3 has been deleted. The second paragraph of Section 3.3 has been rewritten to become the first; it describes the diurnal profile and then explains the model behavior as follows:

"To improve the simulation, we applied pre-defined functions for the diurnal variations of the surface NOx emissions. As described in Section 2.2, we applied different diurnal variation profiles for different sources: maxima in the morning and evening for anthropogenic sources; a rapid increase in the morning and maximal emissions at mid-day for biomass burning sources; and maximal emissions in the afternoon for soil sources. By applying the diurnal variability scheme, CHASER generally shows better agreements with the satellite retrievals, with a global mean RMSE reduction of about 10-15 (30-40) % compared to the OMI (SCIAMACHY) retrievals. Similar results were demonstrated with other CTMs (van Noije et al., 2006; Boersma et al., 2008b). The diurnal variability scheme generally decreases the NO_2 concentration in the morning, but increases it in the afternoon in the industry and biomass burning areas (Fig. 4). It improves the agreement with DOMINO v2 data over Europe (Fig. 4a), whereas the increased biomass burning emission during daytime caused the NO_2 columns over Central Africa to be too high compared to DOMINO v2 data (Fig. 4b). The diurnal variability for the biomass burning source is highly variable and uncertain."

6. *Section 5.1 is too long and contains much self-explained information. The authors do not make any point on the difference between the seasonality of emissions over Europe and the Eastern US (Fig 7, top panels) and on the much higher NOx emissions derived by the assimilation over the Eastern US. This latter result seems to be at odds with what we know from EPA trends of NOx*

emissions over the US. Some discussion is needed.

Several sentences (self-explained information) have been removed from Section 5.1 (see above list).

The following sentences have been added to describe the seasonality of emissions over Europe:
"Both the a priori and a posteriori emissions reveal maximum emissions in summer, but the seasonal amplitude is about 15 % higher for the a posteriori emissions over Europe."

The second paragraph of Section 5.1 has been rewritten to describe the emissions over the US more carefully:
"Over the eastern United States, both the a priori and newer emission inventories are significantly lower than the estimated emissions. The EPA 2005 National Emission Inventory (NEI-05) (U.S. EPA, 2009) showed a larger decrease in anthropogenic NOx emissions for the United States in 2000-2005 (-15.3 %) than in 1995-2000 (-9.0 %); this appears to be inconsistent with the positive increment obtained for 2005 in this study from the a priori emissions created based on 1995- 2000 trends. However, we found that although the trend in 1995-2000 is similar between the a priori emissions (-8.5 %) and the EPA NEI-05 (-9.0 %), the absolute value is 20-30 % lower in the a prior emissions. As a result, the a priori emissions obtained for 2005 over the United States (5.32 TgN/yr) are lower than the EPA NEI-05 (6.23 TgN/yr, for this estimates we used the emission ratio between anthropogenic, soil, and biomass burning emissions over the United States estimated from Zhang et al. (2012)), whereas the a posterior emissions of 6.90 TgN/yr are even higher than the EPA NEI-05. The a posteriori emissions reveal higher emissions from autumn to spring than during summer, which differs from the seasonal variation of the a priori emissions. In contrast, the a posteriori NOx emissions over the contiguous United States are maximized in summer with a July/January ratio of 1.2, consistent with the analysis of NOx emissions inventories, including the EPA NEI-05, soil emissions from Yienger and Levy (1999), and GFED ver. 2, performed by Zhang et al. (2012). Within the eastern United States domain (lower middle panels in Fig. 8), the annual mean a posteriori emissions show higher values than the a priori emissions around large cities in the eastern United States; e.g., around Chicago, Indianapolis, Atlanta, and Florida peninsula. In contrast, the a posteriori emissions are smaller in the northern part of North America (e.g., around Montreal and Toronto), as well as around Houston, with factors of less than 0.6 being observed."

7. *In Section 5.1 (p. 31547, l. 20) the authors state that the REAS inventory largely underestimates the NOx emissions over Eastern China in 2005, as commonly revealed by van Noije et al. (2006). But van Noije et al. (2006) do not make an evaluation of the REAS inventory. This should be corrected. On the other hand, it should be specified how the extrapolation of emissions is*

performed. In the case of the REAS inventory, an extrapolation is not necessary since 2005 and 2006 REAS emissions are already available online. In addition, the GEIA emission inventory for soils cannot be considered as a "latest" inventory, as mentioned in p. 31547, l.26. Also in p. 31548, l. 11, "the a priori and the latest ones" should be changed.

van Noije et al., (2006) has been removed from the sentence.

For all emissions categories, emission data for the simulation years 2005-2006 were obtained by extrapolating the emission inventories from the years 1995 and 2000, even though REAS is available for the years 2005-2006. To describe this more clearly, the related sentence (in Section 2.2) has been modified as follows:
"For all emission categories, the emission values for the simulation years 2005-2006 are obtained by extrapolating the emission inventories from the years 1995 and 2000."

"latest" has been replaced with "newer" throughout the paper.

8. *In Section 5.1 (p. 31550, first paragraph) the statement about oceanic NOx data is misleading since the Boersma et al. (2008) OMI data are not the same as the DOMINOv2 product used in the assimilation. To make a meaningful statement, you must compare DOMINOv1 and DOMINOv2 over the oceans.*

We confirmed that the difference in tropospheric NO2 columns between DOMINO v1 and v2 data over the oceans was very small, which is consistent with the results of Boersma et al. (2011). To state this, the related sentences have been modified as follows:
"Boersma et al. (2008b) found that DOMINO v1 data generally have lower columns with a mean bias of 0.6 × 1015 molec. cm−2 over the ocean when compared to aircraft measurements during the INTEX-B campaign. The difference in tropospheric NO2 columns between DOMINO v1 and v2 data are generally very small over the ocean (Boersma et al., 2011), which suggests a similar bias for v2 compared to INTEX-B data."

9. *In Section 5.3, p. 31553, l. 9-10, it is stated that "above the PBL, the NO2 concentrations decrease with height, primarily because of the NO2/NO ratio, which decreases with temperature". The main reason for the decrease is the relatively short lifetime of the NOx family, due to the NO2+OH reaction. Change in the NO2/NO ratio play only a minor role.*

Based on the comment, the statement has been rewritten as follows:

"Above the PBL, the NO2 concentrations decrease with height, mainly due to the relatively short lifetime of the NOx family."

p. 31531, l. 32 : correct "metrological" in this sentence

Corrected.

p. 31533, l. 10 : correct "maxima emissions"

Replaced with "maximal emissions".

p. 31534, l.1 : missing reference to Whitaker and Hamill

Added.

p. 31537, l. 11 : correct "in out setting"

Replaced with "in our setting"

p. 31541, l. 21 : "occurs in relatively fewer observations", do you mean "leads to relatively fewer observations"?

Yes, it has been replaced.

read names "van Noije" and "van der A" throughout

Corrected.

p. 31542, l. 11 : "the CHASER" remove "the"

Removed.

read "molec./cm2" instead of "mol/cm2" throughout

Corrected.

p. 31549, l. 18 : "of greater" remove "of"

Removed.

p. 31552, l. 27 : "below 950 hPa", read "at pressures higher than 950 hPa", similar correction in l. 28.

Corrected.

p. 31553, l. 6 and 8 : read "increase by a factor" 21.

Corrected.

p. 31553, l. 17 : "Cabauw...100 km", poor phrasing

The phrase has been replaced with "Cabauw is surrounded by major populated areas within a distance of a few 100 km,".

Reply to Referee #2

Major Comments

1) *The limited validation conducted with in situ aircraft data does suggest that the inversion is improving the modeled NOx distribution. However, I think that in the absence of more extensive validation with in situ data, it would be extremely important for the authors to put their estimated NOx emissions in context with what is in the literature. For example, Hudman et al. (JGR, 2007) suggested that power plant and industrial emissions of NOx in North America decreased by 50% between 1999 and 2004. How consistent are the OMI derived NOx estimates with those from other studies, such as the Hudman et al. analysis?*

Several sentences have been added to compare with previously estimated emissions. Please also see our reply to Referee #1 (comment #2 & #6).

2) *The vertical profile comparison (Figure 11) shows that the model is significantly underestimating free tropospheric ozone. My guess is that this is linked to an underestimate of lightning NOx emissions (LNOx), which could bias the surface source estimates of NOx. In particular, I wonder to what extent could this account for the in- ferred increase in NOx emissions in the eastern USA in the inversion? Indeed, when the authors reduced the LNOx source in the model, the a posteriori NOx emissions were much larger. I believe that it would be helpful for the authors to compare their regional LNOx estimates with those from other studies such as Hudman et al. (JGR, 2007), Sauvage et al. (ACP, 2007), and Jourdain et al. (ACP, 2010).*

Thank you very much for the valuable comment. We have compared LNOx with other studies and added the following sentences to discuss the comparison results:

"The free tropospheric ozone is too low by about 10 pptv in both the simulation and the assimilation. An underestimation of nearly 10 ppvt was commonly observed for a GEOS-Chem model simulation over the United States during the International Consortium on Atmospheric Transport and Transformation (ICARTT) aircraft campaign (Hudman et al., 2007). Hudman demonstrated that an enhanced lightning NOx source (0.27 TgN over the United States from 1 July to 15 August 2004) removed most of the upper tropospheric ozone bias in their standard simulation (which had only 0.068 TgN from lightning). The similar magnitude of the ozone underestimation and lightning source (0.061 TgN) in our CHASER simulation shows that although the global total lightning source is similar for our simulation (7.5 TgN/yr) and estimates from chemical observations (mostly 6-8 TgN/yr) (e.g., Martin et al., 2007; Sauvage et al., 2007), CHASER may underestimate lightning NOx sources and their-induced ozone production in the free troposphere over Mexico and North America. Errors in the stratospheric ozone transport into the troposphere may also contribute to the ozone underestimation."

3) *The authors invested significant effort in evaluating the performance of the forward model and the LETFK assimilation system, which I appreciate, since this is the first application of the LETKF for inverse modeling of NOx emissions. However, there are a few key parameters in the assimilation system for which the values were specified in an ad hoc manner without explanation. For example, on lines 16-18 on page 31537, the authors state that the analysis errors were inflated to 30% of the initial standard deviation, based on sensitivity tests. What tests? What metric was used in these test to determine that 30% was the best value? Similarly, on line 1, page 31540, there is no explanation as to why a 15% error correlation was used in generating the super-obs. It is clear that there was quite a bit of tuning of the assimilation system and it would be helpful if the authors gave the reader a better sense as to why the different parameter values were selected for the standard inversion case.*

The following sentences have been added to explain the criteria used to obtain these values.

"The minimum predefined ensemble spread of 30 % and the random noise magnitude of 4 % used in data assimilation were obtained from sensitivity experiments by changing the predefined magnitude to 15, 30, 45, and 60 %, and the noise magnitude to 0, 4, 8, and 12 %, respectively. Quantitative criteria for the selection of these values are the daily Observation-minus-Forecast (OmF) check and the chi-square (χ^2) test (see Section 4.2 for details). The optimal values of 30 % and 4 % were obtained by minimizing the global mean and root mean square of the OmF and by requiring that the χ^2 value approaches 1."

Please see below about the 15% error correlation.

Minor Comments

1) *Line 10, page 31525: Please change "formations of HNO3" to "formation of HNO3"*

Corrected.

2) *Lines 2 – 5, page 31526: I don't understand what the authors are saying here. How is the total column more closely related to the area average emissions than surface data? With the column data you have the contribution from the stratosphere and upper tropospheric sources, such as lightning.*

The sentence has been replaced with:

"Because the satellite measures an area-averaged amount of NO_2, the satellite observations are more representative for the global model grid scale emissions than surface in situ point observations which depend strongly on (finer scale) local sources and local removal processes."

3) *Line 28, page 31526: Please add "us" after "allow"*

Added.

4) *Line 17, page 31528: The averaging kernel is not important in the retrieval. In fact, it is not used in the retrieval. It is a by-product of the retrieval that enables one to assess the sensitivity of the retrieval.*

I agree with the reviewer. The sentence has been replaced with "…averaging kernel (AK) information are important for the use of the observations in data assimilation."

5) *Line 11, page 31537: Please change "out setting" to "our setting".*

Corrected.

6) *Line 27, page 21537: Why did the authors add noise with a magnitude of 4% of the initial spread? This 4% seems ad hoc. Can they justify this value?*

The optimal value is obtained by minimizing the OmF and having the chi-square value approach 1 from sensitivity experiments. This is noted in the revised manuscript. Please also see above (reply to Major comment #3)

7) *Lines 1, page 31540: On what analysis is the 15% correlation based?*

There is no evidence for this number, because this number is very difficult to estimate. This is explained more carefully in the revised manuscript.

8) *Equation (11), page 31540: I don't quite understand the role of alpha here. What does alpha = 0 imply physically?*

When alpha = 0, no data are used for the estimate, and it cannot occur. Thus, the alpha must be larger than 0. This is noted in the revised manuscript.

9) *Figure 2: Are the panels with the difference maps showing the differences between the data and the model with the averaging kernels? It is not clear from the caption. Using the untransformed model in the difference maps would be less useful.*

Yes, we applied the averaging kernel of each satellite retrieval to the model profile to make each plot.

The sentence has been rewritten as follows:

"Global distributions of annual mean tropospheric NO_2 columns (in 10^{15} molec. cm^{-2}) obtained from the satellite retrievals (left columns): DOMINO v2 (upper rows), DOMINO v1 (middle rows), and SCIAMACHY (lower rows), and from the CHASER simulation estimated using the AK of each retrieval to be compared with the simulation (middle columns) for 2005".

10) Line 4, page 31544: Europe looks good, contrary to what is stated in the text.

Yes, the statement has been removed.

11) Figure 4: What are the error bars for the satellite data? Since this is an average over a large region, what is the variability around the mean value?

The plot shows regional- and monthly-averaged results. The error bar in this case is dominated by systematic errors (biased). A rough guess of these errors was used in the construction of the super-observation, but the uncertainty is very high. Also one should note that systematic offsets may be similar for SCIAMACHY and OMI. Because of this we did not add any error bar of the satellite data in this figure. Day to day changes in the area-averaged measured amount are caused in particular by changes in the cloud cover. Such changes are for a large part reproduced by the model. Adding a (grey-shading) range of values is therefore not very meaningful.

12) Lines 21-25, page 31544: The improvement with the diurnal variability scheme is not obvious to me.

We agree with the reviewer that, in particular for central Africa, the improvement is not obvious. The sentences were rewritten to describe the results more precisely. Please also see our reply to Referee #1 (Major comment #5)

13) Lines 10-14, page 31546: Is the better performance with the super-obs just due to the fact that you are averaging a large number of retrievals in each gridbox, and as a result the retrieval noise is significantly reduced? Since the precision on an individual retrieval is not very high, one could image that the noise from each retrieval could cause the inversion to produce noisy increments.

Yes, large noises in individual retrieval sometime cause very noisy increments during the data assimilation. Assuming that the retrieval error is random, the super-observation error becomes much less noisy (or high precision). The sentence has been rewritten to describe this more carefully as follows:

"The super-observation approach generally provides more representative data with a recued random error (e.g., than the individual observation) and results in systematic and smaller analysis increments. Furthermore, the super-observation approach reduces the computational cost of the data assimilation, by reducing the number of data processed in the analysis step"

14) *Lines 1-6, page 315550: I am surprised that the inversion has sufficient sensitivity to optimize the ship emissions. Over the oceans the NO2 column is low, and the stratospheric contribution to the column will be higher than over a continental source region. It would be helpful to see the results of an OSSE showing that the inversion is indeed sensitive to these emissions, given the specified measurement and model errors.*

Because we applied the super-observation error approach, the observation error of each data used for data assimilation appeared to be small enough for optimizing ship emissions in our analysis. We agree with the reviewer that OSSE would be interesting. However, this is a study by itself and is outside of the scope of this paper.

15) *Figure 12: It is not clear what the different lines represent in the scatter plots in panel (b) and (c). Is each line a linear fit to the data points of the same color as the line? Given the significant scatter in each panel, I wonder how meaningful is the linear fit?*

Yes, each line represents a linear fit to the data points of the same color. The following sentence has been added to explain this more clearly.

"Each line represents a linear fit to the points of the same colour, and the colours represent the altitude level. The black line shows a linear fit to all of the data."

Although the distribution is scattered, there are systematic differences in the distributions between (b) and (c). Thus we believe this is useful information.

A note from the authors: Since the quality (resolution, color setting (related to the EPS file conversion), etc...) of several figures were not very good, they will be replaced with improved ones in the revised manuscript. No change will be made to the results and discussions.