Author comments in reply to the anonymous referee on "Global lightning NOx production estimated by an assimilation of multiple satellite datasets" by K. Miyazaki et al.

We want to thank the referee for the helpful comments and suggestions. We have revised the manuscript according to the comments, and hope that the revised version of the manuscript is now suitable for publication. Below are the referee comments in italics with our replies in normal font.

Reply to Referee #1

1) *The approach considers data assimilation of multiple species to constrain LNOx and surface emissions sources as well as species concentrations. For Ozone (O3) the results from the assimilated system are compared to observations in Fig 9. It would be useful to be able to quantify how important the correct simulation of the LNOx source is in itself for O3, since much of the improvement will emanate from assimilation of O3 itself.*

The O3 concentrations simulated using the LNOx sources are discussed in Section 6.3 and shown in Table 9. To clarify the purpose of the validation, the following sentence has been added in Section 6.3.
"This validation demonstrates the importance of correcting the NOx sources for reproducing the O3 fields."

2) *In fact "LNOx –only" optimisation is discussed in section 4.4, but this text is confusing where it is currently placed since this section refers to Table 3 which shows the relative contribution of assimilation of each of the different satellite datasets on simulated O3 chemistry including surface and LNOx sources.*

The result is presented in Table 9 in the revised manuscript. Please also see my reply above.

3) *It is not totally clear, but it seems year 2007 was chosen for both model simulations and for assimilation with measurements? Have any other years been examined to see how well this approach performs in other years?*

Both the model simulation and the assimilation were performed for year 2007. The following sentence has been added to the manuscript in Section 3.1.1:
"Both the model simulation and the data assimilation are conducted for the entire year 2007, because a large amount of satellite data is available for this year."

The inter-annual variability of the LNOx source will be investigated in a future study. The last sentences in Section 6.4 have been rewritten as follow:

"In spite of the good agreement in the estimates of the annual global source and the NO production efficiency, the lightning activity and the LNOx source varies significantly with season and year (e.g., Cecil et al., 2014), and differences will be more pronounced when comparisons are made regionally. The amount of NOx produced per flash may also vary considerably with season and region (c.f., Table 7). Detailed comparisons on monthly and regional scales including those seasonal and inter-annual variations remain an important topic for future studies."

4) *There is no validation with LIS/OTD lightning flash rates though this is discussed briefly. In particular, it would be useful to see if there is any seasonality in flash rates over the oceans in line with those found in Figure 6 (when data assimilation is included)*

Table 1 and Figure 2 have been added to compare with the LIS/OTD measurements. The following sentences have been added to discuss the comparison result in Section 3.2.1:

"Table 1 and Figure 2 compare the global flash rate between the LIS/OTD high resolution monthly climatology (HRMC) data (Cecil et al., 2014) and the model parameterization. Compared with the observations, the global distribution of the total flash rate is generally reproduced by the model."

"Mainly because of the low bias over central Africa, the model underestimates the annual flash rate in the tropics (20S-20N) by about 27 %, leading to about 13 % underestimation in the global total flash rate."

The following sentence has been added to discuss the seasonality in the flash rate over the oceans in Section 4.2:

"Because the predicted flash rate does not show such distinct seasonality over the oceans, and because the seasonal amplitude of the flash rate is generally smaller in the model simulation than in the LIS/OTD measurements over the oceans (figure not shown), these changes imply errors in the seasonal variation of either the flash rate or the NOx production efficiency over the oceans in the model simulation."

5) *Figure 7 shows low clouds over oceans producing maximum amounts of LNOx. Is this signal really due to low clouds or is it that the re-distribution of the LNOx source towards the surface is greater with assimilation?*

By analysing the simulated and analysed LNOx source profiles, we confirmed that the LNOx source maxima in the lower troposphere are closely associated with the occurrence of low convective clouds.

However, the analysis result may have uncertainties associated with errors in the assimilated retrievals over the oceans. The following explanations are provided in Section 4.3 and 6.1.1 in the manuscript:

"Over the oceans, persistent strong sources associated with the simulated low clouds and the occurrence of IC flashes are predicted in the lower troposphere. Data assimilation further increases the lower tropospheric sources by a factor of up to two."

"We note that errors in the OMI tropospheric NO2 column retrievals could cause large uncertainties in the analyzed LNOx sources over the oceans, as will be discussed in Section 6.1.1."

"It is emphasized that low NO2 concentrations over the oceans are mostly smaller than the OMI noise level. Errors related to the separation of stratospheric and tropospheric NO2 could also cause errors in the OMI tropospheric NO2 column retrievals (Lamsal et al., 2010; Boersma et al., 2011). These may cause large uncertainties in the analyzed LNOx sources, especially over the oceans."

6) *The conclusion regarding IC/CG ratios in the discussion (section 6.2.2) is rather confusing but may be insightful. Was it possible to obtain cold cloud thickness and hence ranges of z values from the satellite measurements, in order to comment on whether ratios of 1 or 10 were more likely?*

To more clearly describe the implication obtained from the result, the relevant sentence in Section 6.2.2 has been rewritten as:

"We attempted to optimize the production per flash parameters separately for IG and CG flashes from the multi-species data assimilation but could not find any significant differences between the two parameters in the analysis."

We agree that estimating the relationships between the cloud information from satellite measurements and the analysed LNOx sources provide useful information. However, the treatment of cloud information from satellite measurements needs special cautions (e.g., spatial representativeness, error estimation), and this point remains an important topic for future studies.

Specific comments:

P29204, Line 1 "assimilating observations" add "into a chemistry transport model".

Added

P29204, line 14, "These estimates . . ." This sentence is confusing as Table 3 shows a lower value for the global source when using OMI NO2 alone. It is likely referring to results not shown trying to optimise

LNOx production alone.

The sentence has been rewritten as follow:
"These estimates are significantly different from those estimated from a parameter inversion that optimises the LNOx source only from NO2 observations alone, which may lead to an overestimate of the source adjustment."

P29205, line 4, provide reference for 10-20% is it from Grewe et al. ?

The following paper is cited in the revised manuscript:
Galloway, J. M., Dentener, F. J., Capone, D. G., et al.: Nitrogen Cycles: Past, Present and Future, Biogeochemistry, 70, 153–226, 2004.

P29205, Line 19: explain "the lightning parametrization" – either state which one or rephrase as "any lighting parametrization"

Replaced by "any lightning parameterisation".

P 29205, Line 21: GC to ID flashes equals 10- there is more recent litera-ture on this e.g. DeCaria, et al (2005), J. Geophys. Res., 110, D14303, 860 doi:10.1029/2004JD005556. Ott et al. (2007), J. Geophys. Res., 112, D05307, doi:10.1029/2006JD007365 Ott, et al. (2010), J. Geophys. Res., 115, D04301, doi:10.1029/2009JD011880, 2010

The papers are cited in the revised manuscript. Thank you for the information.

P29206, line 7-8, "errors in these processes ..." – rephrase this text for clarity and provide references.

The sentence has been rewritten as:
"Errors in these processes other than those in the LNOx sources could cause large uncertainties in the LNOx source estimates when observations are used to constrain only the LNOx sources."

P29206 line 10 remove or rephrase "etc". This is a key point so it would be helpful to add an example to reinforce the text.

Removed. The following sentence has been added:

"Martin et al. (2007) demonstrated the ability of satellite NO2, O3, and HNO3 measurements to constrain the LNOx source."

P29206, line 15, it is not obvious that the 4-D var method goes hand in hand with an adjoint approach rather than a forward running model being re-ran. Can this text be explained in more detail, as again it is a key point of the methodology?

The following sentence has been added:
"The 4D-Var requires minimization algorithms to compute gradient information with adjoint models, in which the necessity of the development and maintenance of the adjoint model is the main disadvantage of 4D-Var."

P29206, line 18, define "CTM".

CTM is already defined before.

P29206, line 21, when discussing the 35 chemical species, it would be useful to relate those to the species that are directly measured: O3, NO2, NHO3 and CO.

The sentence is rewritten as:
"…as well as the concentrations of 35 chemical species including the assimilated species (NO2, O3, HNO3, and CO), while taking into account the chemical interactions…"

P29206, line 23, "several "? 35 species are referred to in the line above?

Replaced by "various".

P29206, line 29 "the while year 2007"?

Replaced by "the whole year 2007".

P29207, line 14, Define all terms in the equation here and provide reference/s. Relate this equation more clearly to the sub-sections that follow describing different satellite observations- or move this equation and text to 3.1.2 where this information is used and re-name the section?

The equation has been moved to Section 3.1.2 and the definitions are provided in the revised manuscript.

P29208, line 8, briefly explain here what is meant by "the super observation approach". It is not clear how all the observations are considered together. Are all the datasets re-gridded onto a 2.5 by 2.5 degree grid or is it only for NO2?

The sentence has been rewritten as:
"We employ the super observation approach to produce representative data with a horizontal resolution of 2.5x2.5 for OMI NO2 and MOPITT CO (c.f., Sect. 2.1.4) observations, following Miyazaki et al. (2012b)."

To provide more information, the following sentences have been added:
"A super observation is generated by averaging all data located within a super observation grid cell. The measurement error for the super observation is estimated by considering an error correlation of 15% among data. A representativeness error is introduced when the super-observation grid is not fully covered by OMI pixels."

P29208, line 12, rephrase "OMI scale" for clarity.

Replaced by "at the OMI footprint scale".

P 29208, line 24, rephrase "halfway the cloud"

Replaced by "in the middle of the cloud".

P29209, line 20, are there any issues with MOPITT being on a different satellite to the other 3 instruments on AURA? The MOPITT CO contribution is not shown in Figure 8. Is this because the differences that feed through to the LNOx source from the CO corrections are too small?

We did not find any problem with the use of MOPITT observations. Because the covariance between the LNOx source and CO concentrations are neglected in the analysis, the CO observations do not directly influence the LNOx source. Thus the result is not presented in the figure.

P29212, line 11, H is the observation operator. In section 2.1 y was defined as the observation operator,

please clarify.

The sentences have been rewritten.

P29212, line 18, change to "observations".

Corrected.

P29213, line 2, explain what the term "covariance localization" means.

The following sentences have been added:
“This technique allows us to neglect the correlations among variables that may suffer significantly from spurious correlations, by setting the covariance among non- or weakly related variables to zero.”

P29214, line 6, What is the tuning factor and what is it based on? How does this scaling factor affect the LNOx error?

The tuning factor is applied to obtain a realistic estimate of the global total lightning frequency based on a comparison with an older satellite flash observation data. This tuning factor does not affect the spatial distribution of the lightning frequency. The sentence has been rewritten as:
“A globally and annually constant tuning factor is applied for the total flash frequency in CHASER simulations to obtain a realistic estimate of the global total flash occurrence, whereas the spatial distribution of the flash frequency is determined by the model parameterization.”

P 29216. Line 2, "super observation"

Corrected.

P29216, line 22, "provides".

Corrected.

P29218, line 7, it would be useful to show this figure.

Figure 2 has been added.

P29218, line 8, in fig 5 right hand panels it is difficult to see any coherent differences over Africa, can the description be more precise and include the sign of change.

Since the difference is unclear, the sentence has been removed.

P29218, line 10, it isn't clear which are the model results "with and without assimilation".

The following words have been added to the caption of Fig. 7 in the revised manuscript:
"analysed from the data assimilation (black) and estimated from the model simulation (red)"

P29219, line 13, add where at "240 hpa" (since this could be in the stratosphere at mid-latitudes).

The sentence has been rewritten as:
"Data assimilation increases the LNOx sources over most land regions by 20-50 % in the upper troposphere, with a maximum increase at 240 hPa in the global and annual mean, which is attributed to the source increase in the tropical upper troposphere."

P29222, line 5, It would be helpful to split Table 4 into different regions to accompany the text in this paragraph. It would also be helpful to remind the reader that the assimilation process influences the O3 distribution through the assimilated O3 as well as LNOx. It would be useful to comment if the improved O3 is wholly due to the assimilation of O3.

The table (Table 5 in the revised manuscript) has been expanded to include comparisons for several tropical regions. The last sentence in Section 5.1 has been rewritten as follow:
"Because lightning substantially influences the amount of O3 in the tropics, and because the data assimilation simultaneously optimizes the O3 and the LNOx source, significantly improved agreement with independent ozone observations gives confidence in the performance of the LNOx estimates."

The following sentence has been added in Section 5.1:
"The improved agreement with TOC data is mainly attributed to the assimilation of TES O3 (Miyazaki et al., 2012a)."

P29222, line 8-12, although the O3 bias in the upper troposphere is improved there seems to be a greater bias in the lower troposphere?

The following sentence has been added:
“Conversely, the assimilation does not obviously improve the lower tropospheric O3.”

P 29222, line13 define TOC. This table caption discusses "global" but the text here discusses "in the tropics".

Defined. The table caption has been corrected.

P29223, line 15, it would be useful to note that the LNOx parametrization is not based on cloud fraction but cloud top height although clearly in the GCM cloud top height must be related to cloud existence. It would also be helpful to remind the reader that this region encompasses the maritime continent where significant lightning activity occurs.

The following sentence has been added:
“Accurate simulations of the cloud position are important to properly distribute the LNOx sources, while errors in the simulated cloud top height lead to uncertainties in the total source strength.”

The relevant sentence has been rewritten as:
“The warm sea surface and high convective available potential energy (CAPE) activate vertical uplifting and lightning especially over the maritime continent.”

P 29224, line 15, it is hard to see the improvements discussed in Fig 11 from assimilation of TES and MLS O3.

Additional figures are required to show these improvements more clearly. However, we believe these figures are not really necessary in the manuscript. Therefore, “(figure not shown)” has been added.

P 29227, line 17, the text discusses an increase using SSTs for 1997, but Table 5 shows a decrease for year 1997 compared to the control. This experiment with SSTs for 1997 will have a number of differences besides cloud location and so should be interpreted with caution.

The relevant sentence has been rewritten as follow:
“The impact of changing the SST data was different for different regions; e.g., the LNOx sources over the Pacific increased by 14 % in January.”

P 22928, line 18, change to "are" used. State what the chi squared test results given are actually measuring.

Corrected. The following sentences have been added:
"The chi^2 is estimated from the ratio of the differences between the model forecast and observations to the estimated background covariances. This measures whether the background covariance matrix producing realistic errors. The chi^2 ratio becomes 1 if the background error covariance matches the model-observation differences."

P22929, line 5, the text discussing LNOX a priori errors and a priori source estimates could be made clearer so the reader knows which rows in Table 5 to look at. P22929, line 7, correct to "a priori"

The table has been revised to clarify the meanings.

P22929, line 19, explain "to some extent" more precisely, the value for GL for July is 10%.

The following sentence has been rewritten as:
"A sensitivity experiment in which the a priori global total LNOx source is increased by 15 % demonstrates that the estimated LNOx source amount is influenced by the a priori source setting (Table 6); the global a-posteriori LNOx sources are increased by 4 % in January and 10 % in July."

P22931, line 12, add appropriate reference for 7% underestimation – Murray et al. 2012?

The sentence has been rewritten as:
"On the other hand, an increase in the annual LNOx amount from 4.7 to 6.3 TgNyr-1 is obtained from assimilation but cannot simply be explained by a roughly 4-9 % (=7-12 % minus 3 %) underestimation of the global lightning flash frequency as compared to the climatological observations (41.2 flashes s-1 v.s. 44 or 46 flashes s-1) and considering about 3 % lower flash frequency in 2007 compared to the climatology (c.f., Sect. 3.2.1)."

P29233, line 7, "overestimated by 1km in the tropics"- did Ott et al (2010) find any difference in the tropics?

Ott et al. (2010) showed results for the subtropics and the northern mid-latitudes. To more clearly

describe our result, the sentence has been written as:
“Our analysis also revealed that the peak source height is overestimated by up to about 1 km over land and the tropical oceans.”

P29234, line7, change to "most active".

Corrected.

P29235, line 6, other papers discussed earlier in the paper provide estimates of global LNOx constrained from satellite- Boersma et al. 2005, Bierle et al. 2006, Martin et al. 2007, Lin et al. 2012. It would be useful to add these ranges here.

The results of Boersma et al. (2005), Beirle et al. (2006), and Martin et al. (2007) are already included in the estimate of Schumann and Huntrieser (2007), whereas Lin et al. (2012) estimated the LNOx sources for China only.

Table 2: remove "are shown in brackets".

The sentence has been rewritten as:
“The regional averages of the mean altitude (in km) with maximum annual LNOx emission (i.e., source peak height) estimated from the CTM simulation and the data assimilation and the corresponding analysis increments (the data assimilation minus the simulation).”

Fig 3. Why does the panel for TES O3 in Fig 3 (difference with and without lightning) show a large difference in northern polar latitudes?

The large differences in the northern high latitudes seem to reflect the large simulated LNOx sources over the northern Eurasian continent and North America and also the fact that meridional air transport in the northern extratropics is relatively suppressed during summer.

Fig 5. "analysed sources" add "of LNOx".

Added.

Fig 6. The black and red lines and numerical values need to be explained.

The following sentences have been added:

"…analysed from the data assimilation (black) and estimated from the model simulation (red). The total annual values (in TgNyr-1) are displayed in each panel."

Fig 11. Some of the caption is rather unclear. Explain what "inventories" mean. Are these the datasets used in the CTM? Re-phrase more clearly and give references. Rephrase "the data assimilation" to "the CTM simulation using data assimilation" or such like.

The figure caption has been rewritten to clarify the descriptions.