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.

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 #2

1) *Apparently, this study builds up on Miyasaki et al. (2012a). It is not clear whether you use the same data assimilation experiment than in Miyasaki et al. (2012a) or whether you had done some new developments compared to Miyasaki et al. (2012a) or performed a new data assimilation experiment. I think this has to be clarified in the Introduction section.*

The following sentence has been added to the Introduction section.
"Compared to the system described in Miyazaki et al. (2012a), several updates have been applied to the data assimilation settings on the a priori emissions and the assimilated measurements."

2) *The goals of the sections 5.2 and 5.3 are not very clear. Please clarify. In addition, in Results, you can maybe first present the validation of the data assimilation and after present the LNOX source estimation.*

The following sentences have been added to describe the purpose of these sections:
"Lightning strongly influences the O3 production and chemistry, especially in the tropical troposphere, as discussed in Sect. 3.2.2 and suggested by Sauvage et al. (2007a). Lightning activity and surface NOx sources differ considerably among the tropical regions, reflecting variations in the meteorological conditions including cumulus convection activity. This section demonstrates the ability of CHASER-DAS to analyse the LNOx sources and O3 distributions in several tropical regions."

We believe the structure of Section 5, first showing the general performance of data assimilation and then demonstrating detailed analysis results, is reasonable.

3) *Section 5.1: Why do you perform the validation at only 4 Shadoz websites ? I think it will be more rigorous to have a comparison for all the sites otherwise one can think that you chose to show the sites for which it works well. You could show some of the comparisons and present the results of all*

the comparisons in term of bias, correlation and rms for the LT and UT in a table. You can also refer to the extensive validation of the CHASER-DAS system presented in Miyasaki et al. (2012a), if this is relevant (see my question 1).

The following sentence has been added:

"Ozonesonde observations from 39 locations (9 locations in the tropics) have been used to validate the global ozone profiles (see Sect. 6.3). In the tropics, the data assimilation reduces the mean ozone concentration bias: by 11 % in the lower troposphere (750-450 hPa), by 63 % in the middle troposphere (450-200 hPa), and by 79 % in the upper troposphere (200-90 hPa) in January. Similar improvements were reported before by Miyazaki et al. (2012a)."

4) *Section 3.2.1: The parameterization of Price and Rind (1992) should be only applied to convective clouds. I wonder whether you apply it to every cloud. Indeed, LNOx over oceans in figure 7 is maximum in the lower troposphere below 900hPa. This seems unrealistic. Please clarify this point.*

The parameterization was applied to convective clouds only. This is clearly described in the revised manuscript as follow:

"The global distribution of the flash rate is calculated in CHASER for convective clouds on the basis of the observed relation between the lightning activity and the cloud top height (Price and Rind, 1992) in the AGCM at each forecast step."

The AGCM tended to produce low convective clouds over the oceans, and thus lower tropospheric source maxima are produced. Even when high convective clouds are simulated, lower tropospheric LNOx source maxima could be present because of the averages of individual LNOx profiles with different cloud top height (i.e., the lower maxima are always present in the lower troposphere but the upper maxima occur at various altitudes). The following explanation is provided 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."

The following sentence has been added in Sections 4.3 and 6.1.1 to discuss the reality of the analyzed LNOx sources over the oceans:

"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."

5) *In Pickering et al. (1998), 3 vertical profiles of LNOx are provided depending on the environment (land/ocean, tropical/midlatitudes). It is not clear if you used these 3 profiles or only one of them. Can you be more precise on this point?*

The following sentences have been added:
"The three profiles provided by Pickering et al. (1998) is averaged and applied in the parameterization."

6) *Section 4.4 : I do not understand your explanation for the negative analysis increment in the upper tropospheric LNOx obtained from the assimilation of TES (figure 8) due to the negative bias of TES in the UT. I thought TES had a general small positive bias in the upper troposphere according to Worden et al. (2006) and Nassar et al. (2008). In this last paper, the only systematic negative bias occur in southern subtropics. In figure 8, the negative analysis increment due to TES is for the southern tropics and also for the northern midlatitudes.*

Although the reason for the negative increment is not very clear from the analysis, the positive bias in the simulated O3 concentrations in the UTLS region (Fig. 11 in Miyazaki et al. (2012a)) could be partly responsible for the negative increment. The sentence has been rewritten as:
"The negative analysis increments in the upper tropospheric LNOx sources obtained from the assimilation of TES O3 data likely arises from the TES negative bias (up to 20 %) from the upper troposphere to the lower stratosphere in the southern subtropics, see e.g. Nassar et al. (2008), whereas those in the northern mid-latitudes may be associated with the positive bias in the simulated O3 (Miyazaki et al., 2012a)."

7) *section 6.1.4 : Could you explain the latest step in the calculation of the total error (p 29230 l l2-18) ?*

The explanation has been expanded.

8) *When speaking about lightning activity over the ocean in section 6.2.1 you can refer to Boccippio, Dennis J., 2002: Lightning Scaling Relations Revisited. J. Atmos. Sci., 59, 1086–1104. It is shown in this paper that the lightning parameterization of Price and Rind (1992) over the oceans is not*

consistent with observations.

The following sentence has been added:
"Boccippio (2002) also pointed out inconsistencies between the scheme of Price and Rind (1992) and satellite observations over the ocean."

Minor comments:
Page 29206 line 10: 'etc' to be removed

Removed.

Page 29207 line 14: could you put the expression of the observation operator in section 3.1.2 ? Please also clarify the explanation of the expression. In particular, please better define the operators S and A and explain the utility of H.

The sentences have been written as:
"The observation operator (H) is constructed on the basis of the spatial interpolation operator (S), the a priori profile (xa) and the averaging kernel (A), which maps the model fields (x: N- (the system dimension) dimensional state vector) into observation space (y: p- (the number of observation) dimensional observational vector) while taking into account the vertical averaging implicit in the observations as follows:"
The following sentences have been added:
"The spatial interpolation operator (S) is first applied to the model fields x in order to interpolate to the horizontal location of each observation and the height of each of the vertical layers. The averaging kernel (A) is then applied to define the sensitivity of the satellite retrieved state to changes to the true state. For weak absorbers, the a priori profile (xapriori) does not, or only weakly, influence the relative model-observation difference (Eskes and Boersma, 2003). The averaging kernel (A) and the a priori profile (xapriori) information provided for each retrieval is used in the data assimilation."

Page 29212 line 14: the ensemble mean analysis is then Page 29213, line 13: typo

Corrected.

Page29221 l 20-24, could you put the influence of the length of the assimilation cycle in the discussion of the errors in section 6.1.3?

The following sentence has been added to Section 6.1.3:

“The choice of the length of the data assimilation cycle could also influence the data assimilation result associated with distinct diurnal variations in tropospheric chemistry.”

Page 29224, line 4: 153 S -> 15S Page 29229, line 2: typo

Corrected.

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 #3

1. *The a posteriori lightning NOx product will reflect corrections to convolved errors in the model representation of both flash activity and NOx yields per flash. The lightning flash rate was not assimilated (satellite coverage is poor; global ground networks have low detection efficiencies). However, the flash rate parameterization was also not adjusted to match the satellite climatology from LIS/OTD, as is done for most global models. This is surprising, because the global lightning flash rate distribution is the best-known aspect of the lightning NOx source. If the authors wish to maintain discussion of the assimilated LNOx emissions in the individual context of the unconstrained flash rate (Section 6.2.1) versus NOx yields per flash (Section 6.2.2) –Âa ˘both of which have very large uncertainties in models – then the flash rate distribution of the model should be shown and quantitatively evaluated against the spatial and seasonal distribution from LIS/OTD. The authors seem to suggest that the lightning flash rate parameterization performs very well when unconstrained, which would be a very surprising result in the context of the literature (e.g., Tost et al., 2007), and therefore should be documented.*

Evaluation results using the LIS/OTD measurements are shown in Tables 1, 7, 8, and in Figs. 2, 14 in the revised manuscript. To discuss the results, the following sentences have been added 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 parameterisation. 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 relevant sentence in Section 7 has been rewritten as follow:

"First, errors in flash rates can explain only a small fraction of the uncertainty in LNOx estimates, as the main observed features of the annual global flash rate are generally reproduced by the model, except for the large low bias over central Africa."

The evaluation results for the flash rate and the NO production efficiency using the LIS/OTD observations are discussed in Section 6.2.1 and Section 6.2.2, respectively, in the revised manuscript. Please also see my reply below.

2. *The technique used here should not be able to distinguish between co-located NOx emission sources in a grid cell (e.g., surface lightning and anthropogenic sources, free tropospheric aircraft and lightning), and assumably depends on the a priori fraction of emissions for source attribution. If this is the case, some discussion should be included as it pertains to the results presented here. E.g., if the Ott et al. (2010) vertical probability distributions for lightning emissions were used instead of Pickering et al. (1998), which had a much smaller fraction emitted in the boundary layer, then the assimilation would attribute more of its surface NOx corrections to anthropogenic sources than lightning, which would influence the total lightning NOx value. Corrections of biases in surface sources in strongly polluted but lightning-prone regions (e.g., Gulf Coast, Congo) may be erroneously ascribed to lightning. Similarly, it is unclear to me how this technique could be used to differentiate between IC and CG flash yields, unless they have very separate spatiotemporal signatures from one another.*

The combined use of the multiple datasets with different vertical sensitivities will provide some information on the vertical LNOx profile (see section 6.2.3) and allows the assimilation to distinguish between the surface NOx emissions and LNOx sources as described in Section 3.2.2. Furthermore, transport from the source region is different for sources at different altitudes. In the boundary layer, the LNOx fraction can in principle be constrained in a meaningful way if it is the dominant source. In case of strong, simultaneous surface sources the emission adjustments will be distributed according to the assumed errors in the surface and lightning sources, and the a priori fraction of the NOx emissions (see Section 6.1.3). As suggested by the referee, use of a different a priori LNOx profile may indeed affect the LNOx source analysis.

The relevant sentence in Section 6.1.3 has been written as:
"Therefore, the estimated LNOx sources could have large uncertainties, especially where the surface emissions are large and variable."

The importance of separately estimating for IC and CG flashes is discussed in Section 6.2.2. The following sentence has been added in Section 3.2.2:
"The data assimilation optimizes the multiplication factors for the total LNOx sources, and does not separately optimise for IG and CG flashes.

The following sentence has been added in Section 6.2.3:
"When the observational constraints are insufficient to adjust the vertical profiles, changes in a priori LNOx source profiles (e.g., from the profiles of Pickering et al. (1998) to those of Ott et al. (2010)) or changes in the vertical structure of the covariance matrix will affect the analysed profiles."

Specific Comments
p29206 l25-27 - Does it not also have the potential to introduce larger errors if uncertainties are large in the additional constraint? e.g., the bias in TES UT ozone as shown in Fig. 8?

Additional error sources can be introduced by simultaneous data assimilation. This point is discussed in Section 6.1.1.

p29206 l29 - I suspect the "while" is erroneous?

Replaced by "whole".

p29207 l13-17 - Equation 1 would be better placed in Section 3.

Moved to Section 3.1.2.

p29207 l23 - remove subjective term "strong," perhaps replace with "useful"

Replaced.

p29208 l23-25 - There appears to be a missing word after "halfway"?

Replaced by "in the middle of the clouds".

p29210 l6-10 - Version and access date should be given for the OMI/MLS product, which has changed over time.

Added.

p29211 l25 – p29212 l4 - What is meant by "based on"?

Replaced by “obtained from”.

p29212 l3-4 - The authors should compare the aircraft emissions used here in the context other estimates from the literature (e.g., Wilkerson et al., 2010, http://doi.dx.org/10.5194/acp-10-6391-2010). The interpretation of the assimilated LNOx results will be sensitive to uncertainty in aircraft emissions, which should be acknowledged.

The sentence has been rewritten as:
“The total NOx emission by aircraft is obtained from EDGAR as 0.55 TgN yr-1, which is similar to a more recent estimate of 0.49 TgN yr-1 for 2004 (Wilkerson et al., 2010).”

The following sentence has been added in Section 6.1.2:
“Although the aircraft NOx emissions likely have relatively small uncertainties (e.g., Wilkerson et al., 2010), the LNOx source estimates might be influenced by errors in the aircraft emissions especially along the major flight routes in the northern extratropics.”

Section 3.1.2. This section could use clarification, particularly for readers not familiar with data assimilation and/or EnKF. It would be helpful to include a sentence or two that qualitatively describe how the EnKF works. Does the error covariance matrix take into account errors in the observations (e.g., those discussed in Section 6.1.1), or does EnKF blindly treat all the satellite products as truth, even in instances where we know the observations to be poor or highly uncertain? What averaging kernels are used in H(x), assumably those from each satellite product? What is the value of k?

Section 3.1.2 has been expanded and reformulated. The following sentences have been added:
“The EnKF uses an ensemble forecast to estimate the background error covariance matrix. The advantage of the EnKF over 4D-VAR is its easy implementation for complicated systems; i.e., it does not require the development of an adjoint code. The EnKF data assimilation technique employed is local ensemble transform Kalman filter (LETKF, Hunt et al., 2007). The LETKF scheme, which is based on the ensemble square root filter (SRF) method (e.g., Whitaker and Hamill, 2002), generates an analysis ensemble mean and covariance that satisfy the Kalman filter equations for linear models. The LETKF has conceptual and computational advantages over the original EnKF. The analysis performed locally in space and time reduces sampling errors caused by limited ensemble size, which also enable us to perform parallel computation. The computational advantages are important for this study because of the large state vector size.”

“The spatial interpolation operator (S) is first applied to the model fields x in order to interpolate to the horizontal location of each observation and the height of each of the vertical layers. The averaging kernel (A) is then applied to define the sensitivity of the estimated state to changes to the true state. Because of the operator, the a priori profile (xapriori) does not, or only weakly, influence the model-observation difference in the data assimilation. The averaging kernel (A) and the a priori profile (xapriori) information provided for each retrieval is used in the data assimilation.”
“In conclusion, the data assimilation updates model variables (the concentrations and the emission multiplication factors) for every grid point. This analysis is based on the observational information (i.e., the satellite retrievals) and the background error covariance estimated from the ensemble forecast with 48 members in our case. The estimated concentrations and emissions are used as initial conditions in the next step of ensemble model simulations and updated at every analysis step (i.e., 100 min.).”

p29213 l22-23, p29231 l9-10 neglect to acknowledge the existence of ground-based networks with global coverage, e.g., the World Wide Lightning Location Network (WWLLN; Abarca et al., 2010, http://dx.doi.org/10.1029/2009JD013411) or Vaisala’s GLD360.

The sentence has been rewritten as:
“The ground-based operational lightning detection networks (e.g., the World Wide Lightning Location Network (WWLLN)) provide lightning maps but they have low detection efficiencies (Abarca et al., 2010), whereas satellite instruments provide limited coverage on a daily basis.”

p29214 l6-10 - why was a global scaling factor chosen to give 41.2 flashes s-1, rather than one to match the climatological value from satellites? Also, the more recent climatology using the combined LIS and OTD instruments (46 flashes s-1; Cecil et al., 2012, http://doi.dx.org/10.1016/j.atmosres.2012.06.028) should be referenced, instead of the old OTD-only reference.

The scaling factor was used to obtain a realistic flash estimate based on a comparison with an older observation data. The sentences have 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.”
“The simulated average global flash rate for 2007 is 41.2 flashes s-1, which is comparable to the climatological estimates of 44+-5 flashes s-1 derived from the Optical Transient Detector (OTD) measurements (Christian et al., 2003) and 46 flashes s-1 derived from the Lightning Imaging Sensor (LIS) and OTD measurements (Cecil et al., 2014). The difference between the model simulation and the

observations is partly attributed to interannual variations in flash activity; the annual total flash rate for the latitude band 35S-35N in 2007 observed from the LIS measurement is about 3 % lower than those from the climatology. Because only LIS measurements are available in 2007 and because the global coverage was not obtained, this study uses the climatological observations obtained from a combination of LIS and OTD measurements to validate the global flash rate."

p29214 l20-24 - z is not the IC/CG ratio as stated by the authors, but the CG proportion of total flashes. (Otherwise, setting z to zero makes no sense). Also, the coefficients for z given here are those from Price and Rind (1993, http://dx.doi.org/10.1029/93GL00226), not those in Price et al. (1997).

Corrected.

p29214 l26 – p29215 l1, p29232 l18-19 - The difference in yields between IC and CG flashes is still very uncertain. Comparison of what is used here with the literature should be given. Most recent work suggests the CG/IC production ratio should be closer to unity, cf. Table 19 of Schumann and Huntrieser (2007), although not all (e.g., Koshak et al., 2013; http://dx.doi.org/10.1016/j.atmosres.2012.12.015).

The sentences have been rewritten as:
"Second, following Price et al. (1997), the LNOx source amounts are calculated on the basis of a lightning NO production of 1100 moles per CG flash and 110 moles per IC flash, with a mean energy per CG flash of 6.7x10^9 J flash-1."
"However, it has been suggested that the ratio should be closer to 1 than to 10 (Gallardo and Cooray, 1996; Fehr et al., 2004; DeCaria et al., 2005), although a more recent estimate by Koshaz et al (2014) showed the ratio to be closer to 10."

p29215 l1-5 - Were the Pickering et al. (1998) profiles scaled to local cloud top height, or were fixed altitudes used? Why were the Pickering et al. (1998) profiles used instead of the Ott et al. (2010) profiles?

The Pickering et al. (1998) profiles were scaled to local cloud top height. CHASER uses the Pickering et al. (2008) profiles because the Ott et al. (2010) profiles were not available when CHASER was developed (and the model has not yet been updated).

p29216 l4 - "lighting" should be "lightning."

Corrected.

p29216 l5-6 - The Cooper et al. (2007) and Hudman et al. (2007) studies examined North America, not the tropical upper troposphere. Better references for comparison would be Sauvage et al. (2007, http://dx.doi.org/10.5194/acp-7-815-2007) or Murray et al. (2012), who examined the influence of lightning in the tropics.

Corrected.

p29221 l3-7 - The assimilated changes in mean OH could be independently evaluated by comparison to the methyl chloroform and methane lifetimes, available from observational constraints (cf. John et al., 2012, and references therein; http://dx.doi.org/10.5194/acp-12-12021-2012). In addition to OH, I would also expect a major benefit of the multiple-species to be in its ability to constrain ozone production efficiencies (OPE, which may be approximated as PO3/PHNO3, cf. Cooper et al., 2010, http://dx.doi.org/10.1029/2010JD015056), which are non-linearly dependent on NOx, and would be important for inversely determining LNOx emissions from ozone observations.

Evaluations of the analysed OH fields and the OPE are very interesting topics. However, these are beyond the main scope of this study and should be discussed in a separate paper. To note the importance, the following sentence has been added in Section 4.2:
"The simultaneous assimilation also has the ability to constrain ozone production efficiencies (OPE) through the NOx-CO-OH-O3 set of chemical reactions, which may improve the LNOx source estimation with the assimilation of TES O3 data. Detailed analyses are required to measure the impact of the simultaneous assimilation on OPE."

p29222 Section 5.1 - The authors might consider showing Ascension instead of Irene, given the expected strong influence of lightning on the South Atlantic ozone maximum, the dominant mode of seasonal variability in tropical ozone (e.g., Sauvage et al., JGR, 2007, http://dx.doi.org/10.1029/2006JD008008).

Added.

p29223 l4-5 - Convection and lightning are heavily parameterized everywhere in the model. Please cut, or give an objective argument as to why tropical W Pacific is expected to have worse convection or lightning than elsewhere in the model.

The sentence has been rewritten as:
“Large uncertainties in the LNOx sources are expected over the tropical western Pacific because of errors in the tropical Pacific ITCZ cumulus clouds simulated by the AGCM (Emori et al., 2005).”

p29223 l15-17 - Please clarify what is being compared in these sentences.

The following sentence has been added:
“The analysed NO2 and O3 concentrations show better agreements with the observations (Fig. 11) because of the simultaneous data assimilation.”

p29224 l26-27 - Please justify why large uncertainties in cumulus cloud and biomass burning activity are "expected" in this region

The sentence has been rewritten as follow:
“In this region, large uncertainties in the simulated cumulus cloud and biomass burning activity are expected, as suggested by Emori et al. (2005) and Stroppiana et al. (2010), respectively.”

p29228 l18 - “tests is” should be “tests are”

Corrected.

p29228 l25-28 - Please clarify what is meant by the phrases “mean analysis spread” and “spin-up period for the assimilation” (I thought Kalman filters only require the previous state?). Also, “week” should be plural.

The sentences have been rewritten as
”The mean analysis spread, as estimated by transforming the background ensemble in the data assimilation (c.f., Eq. (4)), is about 0.9 TgNyr-1 for the annual global source strength”
“The LNOx analysis is obtained from information of roughly two weeks of measurements, as demonstrated by the spin-up period of the assimilation (i.e., the spin-up period was required to obtain a converged solution in the analysis)”

p29231 l16-19, p29237 l2-3 - I find this conclusion weak unless more is done to objectively evaluate the flash rate distribution in the model. It could easily be due to a systematic low bias in the a priori NOx production per flash over the ocean. Whether or not this is primarily due to underestimation of (1) the

flash rate, or (2) NOx yields per flash over marine regions could be determined by comparison of the simulated flash rates with the LIS/OTD climatology.

Table 7 and the following discussions have been added in Section 6.2.1:

"We note that comparisons against the LIS/OTD observations consistently reveal a larger underestimation in the parameterised global flash rate over the oceans (about 27 %) than over land (about 5 %). On the other hand, over the tropical oceans (Pacific, Atlantic, and Indian Oceans), the difference between the observed and the parameterised flash rate is relatively small, as summarised in Table 7. This suggests that errors in the NOx production efficiency rather than those in the flash rate could be responsible for the large increase in the LNOx sources over the tropical oceans. This will be further discussed in Sect. 6.2.2."

The relevant discussions in Section 6.2.2 have been rewritten as follow:

"The annual global LNOx source from our estimates corresponds to a mean NO production of about 350 mol flash-1 based on the parameterized flash rate, as summarized in Table 7. Because errors in the parameterized flash rate influence this estimation, we also use the LIS/OTD climatological observations; a global mean NO production of about 310 mol flash-1 is estimated using the flash observations. Both these values are within the range of most other recent estimates."

"Our analysis for July consistently reveals a large production per flash of 430 and 350 mol of NO in the NH (20-90N) compared to 360 and 240 mol of NO in the tropics (20S-20N) based on the parameterised flash rate and the LIS/OTD observations, respectively. There are also obvious regional differences; e.g., a large production per flash of about 440 and 570 are estimated for the northern Eurasia continent based on the parameterised flash rate and the LIS/OTD observations, respectively, as summarised in Table 7 and shown in Fig. 14. The detailed spatial structures in the production efficiencies estimated from the analysed LNOx sources and the observed and the parameterised flash rates (Fig. 14) may reflect not only variations in flash characteristics but also noises and errors in the assimilated and flash measurements (c.f., Section 6.1.2). Note that the production efficiency estimated using the observed flash rate becomes unrealistically large locally where the observed flash rate is much smaller than the model flash rate."

The following paragraph has been also added in Section 6.2.2:

"The NO production efficiencies estimated using the simulated total LNOx sources and the simulated flash rate by the model parameterization (without any assimilation) are about 20 % lower over land and about 11 % lower over the oceans, compared with those estimated using the analysed LNOx sources and the LIS/OTD observations. The obtained results imply general underestimations in the NOx production efficiency simulated by the model, although there are obvious regional differences in the estimates (Table

7). The underestimation could be attributed to errors either in the parameterised IC/CG flash ratio (c.f., Eq. (5)) or in the assumptions on the production efficiency of IC and CG flashes."

The following sentence has added in Section 7:
"It is also suggested that the model parameterisation may underestimate the annual and global mean NO production efficiency by about 10 % over land and 20 % over the oceans."

p29231 l23-24 - 6.3 Tg N yr-1 using a global mean flash rate of 46 flashes s-1 from the LIS/OTD climatology corresponds to 310 mol per flash. Do you expect your 41.2 flashes s-1 for 2006 could be explained by interannual variability in the global mean lightning flash rate?

Please see my reply above.

p29232 l10-14 - An extremely useful figure for the community would be a map of the average NOx yield per flash, calculated by using the assimilated LNOx emissions divided by the flash rate distribution from (1) the model parameterization, and (2) the LIS/OTD climatology. This would be helpful for informing CTMs/CCMs as how to implement differential LNOx yields per flash, which are typically done in arbitrary manner, but necessary for matching global ozone distributions. To me, this is the most useful and unique scientific contribution enabled by this work. The greatest uncertainty global models face at present in reproducing the lightning NOx source is in NOx yields per flash, since most constrain the flash rate magnitude and distribution to the LIS/OTD climatology.

Thank you for your advice. In the revised paper, Table 7 and Fig. 14 demonstrate the NO production efficiencies estimated from the model flash rate and the LIS/OTD observations. Table 8 compares our estimates with other estimates.

The following sentence has been added in Section 6.3:
"The annual global LNOx source from our estimates corresponds to a mean NO production of about 310 mol flash-1 based on the LIS/OTD climatological observations. This value is also within most of the recent estimates (c.f., Table 8)."

Please also see my reply above.

p29237 l7-8 - Please rephrase to make it clear that this is because errors in simulated flash rates are small in this study. Many CTM studies find it necessary to constrain the lightning flash rates for their

ozone simulations (e.g., Martin et al., 2007; Sauvage et al., 2007; Jourdain et al., 2010; Allen et al., 2010; Murray et al., 2012).

The sentence has been rewritten as:
"First, errors in flash rates can explain only a small fraction of the uncertainty in LNOx estimates, as the main observed features of the annual global flash rate are generally reproduced by the model, except for the large low bias over central Africa."

Fig. 9 - superfluous axis labels and titles could be removed to increase panel box sizes

Removed.

Fig. 12 caption should clearly state which difference is taken (I assume with minus without the cloud-covered observations)?

Corrected.

%% Copernicus Publications Manuscript Preparation Template for LaTeX Submissions
%% ---------------------------------
%% This template should be used for the following class files: copernicus.cls, copernicus2.cls, copernicus_discussions.cls
%% The class files, the Copernicus LaTeX Manual with detailed explanations regarding the comments
%% and some style files are bundled in the Copernicus Latex Package which can be downloaded from the different journal webpages.
%% For further assistance please contact the Publication Production Office (production@copernicus.org).
%% http://publications.copernicus.org

%% Differing commands regarding the specific class files are highlighted.
~~dsada~~

%% copernicus.cls
\documentclass[acp]{copernicus}

%% copernicus_discussions.cls
%\documentclass[acpd, hvmath, online]{copernicus_discussions}

\begin{document}

\linenumbers

\title{Global lightning \chem{NO_x} production estimated by an assimilation of multiple satellite datasets}

~~\Author[1]{K.}{Miyazaki}~~

~~\Author[2]{H.~J.}{Eskes}~~

~~\Author[3]{K.}{Sudo}~~

~~\Author[4]{C.}{Zhang}~~\author[1]{K. Miyazaki}

\author[2]{H.~J. Eskes}
\author[3]{K. Sudo}
\author[4]{C. Zhang}

\affil[1]{Japan Agency for Marine-Earth Science and Technology, Yokohama 236-0001, Japan}

\affil[2]{Royal Netherlands Meteorological Institute (KNMI), Wilhelminalaan 10, 3732 GK, De Bilt, the Netherlands}
\affil[3]{Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan}
\affil[4]{International Pacific Research Center, University of Hawaii at Manoa, Honolulu, Hawaii, USA}

%% The [] brackets identify the author to the corresponding affiliation, 1, 2, 3, etc. should be inserted.

\runningtitle{Lightning \chem{NO_x} production estimation}

\runningauthor{Miyazaki et al.}

~~\correspondence{K. Miyazaki\\ (kmiyazaki@jamstec.go.jp)}~~

\runningauthor{K.~Miyazaki et~al.}

\correspondence{K.~Miyazaki (kmiyazaki@jamstec.go.jp)}

\received{28~October 2013}
\accepted{29~October 2013}
\published{}

%% These dates will be inserted by the Publication Production Office during the typesetting process.

\firstpage{1}

\maketitle

\begin{abstract}
The global source of lightning-produced \chem{NO_x} (\chem{LNO_x}) is estimated by assimilating observations of \chem{NO_2}, \chem{O_3}, \chem{HNO_3}, and \chem{CO} measured by multiple satellite measurements into a chemical transport model. Included are observations from the Ozone Monitoring

Instrument (OMI), Microwave Limb Sounder (MLS), Tropospheric Emission Spectrometer (TES), and Measurements of Pollution in the Troposphere (MOPITT) instruments. The assimilation of multiple chemical datasets with different vertical sensitivity profiles provides comprehensive constraints on the global \chem{LNO_x} source while improving the representations of the entire chemical system affecting atmospheric \chem{NO_x}, including surface emissions and inflows from the stratosphere. The annual global \chem{LNO_x} source amount and NO production efficiency are estimated at 6.3\,\unit{Tg\,N\,yr^{-1}} and 3~~5~~10\,\unit{mol\,NO\,flash^{-1}}, respectively. Sensitivity studies with perturbed satellite datasets, model and data assimilation settings leads to an error estimate of about 1.4\,\unit{Tg\,N\,yr^{-1}} on this global \chem{LNO_x} source. These estimates are significantly different from those estimated from a parameter inversion that optimises only the \chem{LNO_x} source from ~~those derived from~~ \chem{NO_2} observations alone, which may lead to an overestimate of the source adjustment. The total \chem{LNO_x} source is predominantly corrected by the assimilation of OMI \chem{NO_2} observations, while TES and MLS observations add important constraints on the vertical source profile. The results indicate that the widely used lightning parameterization based on the C-shape assumption underestimates the source in the upper troposphere and overestimates the peak source height by up to about 1\,\unit{km} over land and the tropical western Pacific. Adjustments are larger over ocean than over land, suggesting that the cloud height dependence is too weak over the ocean in the Price and Rind (1992) approach. The significantly improved agreement between the analysed ozone fields and independent observations gives confidence in the performance of the \chem{LNO_x} source estimation.
\end{abstract}

\introduction
Lightning-produced \chem{NO_x} (\chem{LNO_x}) plays an important role in tropospheric chemistry through influences on ozone formation and oxidation capacity (e.g. Schumann and Huntrieser, 2007, and references therein). \chem{LNO_x} accounts for only about 10--20\,{\%} of the global \chem{NO_x} sources, but is the most dominant source in the upper troposphere (e.g. Galloway et~al., 2004). A~small fraction of \chem{LNO_x} can lead to significant ozone production in the upper troposphere (Thompson et~al., 1994), because the \chem{O_3} production efficiency per \chem{NO_x} molecule typically increases with height owing to the

longer lifetime of \chem{NO_x} and the highly non-linear dependence of ozone production on \chem{NO_x} (Pickering et~al., 1998; Martin et~al., 2000; Jenkins and Ryu, 2004). Therefore, accurate estimates of \chem{LNO_x} source strength and its global distribution are important for understanding tropospheric chemical systems and for improving chemical transport models (CTMs).

The lightning and subsequent \chem{NO_x} formation are estimated with the aid of parameterizations in CTMs. Various schemes have been developed for determining the global distribution of flashes and \chem{LNO_x} sources on the basis of assumptions regarding the \chem{NO_x} production efficiency per flash, energy ratio of cloud-to-ground (CG) flashes to intra-cloud (IC) flashes, and vertical source profiles (Schumann and Huntrieser, 2007). The parameterizations are generally too simplified and have large uncertainties. First, ~~the~~any lightning parameterization cannot fully represent the regional variability of lightning activity (e.g. Boccippio, 2002; Jourdain et~al., 2010). Second, most studies have assumed that the energy ratio of CG flashes and IC flashes equals 10, but this ratio is likely to have a~much smaller value (e.g. DeCaria et~al., 2000; Fehr et~al., 2004; DeCaria et~al., 2005; Ott et~al., 2007, 2010). Third, assumption of a~C-shaped vertical \chem{LNO_x} profile, with a~first maximum in the upper troposphere and a~second maximum in the boundary layer as proposed by Pickering et~al.~(1998), may place too much \chem{NO_x} near the surface and too little in the middle and upper troposphere (e.g. Ott et~al., 2010).

The \chem{LNO_x} sources can be optimized through a~top-down approach, in which estimates of the \chem{LNO_x} sources are obtained by finding the best match between model and observations. Tropospheric \chem{NO_2} column observations from satellite instruments have been used to constrain the global \chem{LNO_x} source (e.g. Boersma et~al., 2005; Beirle et~al., 2006; Martin et~al., 2007; Lin, 2012). In these estimates, however, the mismatches between observed and simulated \chem{NO_2} concentrations are influenced by not only the \chem{LNO_x} sources but also by other processes such as surface emissions, stratospheric inflows, and chemical production and loss processes. Errors in these processes other than those in the \chem{LNO_x} sources could cause large uncertainties in the \chem{LNO_x} source estimates when observations are used to constrain

only the \chem{LNO_x} sources.

Satellite measurements of chemical species other than \chem{NO_2} provide important constraints on the \chem{LNO_x} source by constraining the chemical interactions with \chem{NO_x}~~ etc~~ and by reducing errors in other chemical species that influence the \chem{NO_x} chemistry. Martin et~al. (2007) demonstrated the ability of satellite \chem{NO_2}, \chem{O_3}, and \chem{HNO_3} measurements to constrain the \chem{LNO_x} source. Advanced data assimilation techniques, such as four-dimensional variation (4~~-~~D-Var) and ensemble Kalman filtering (EnKF), are powerful tools to combine multiple observations with models to obtain comprehensive constraints on \chem{LNO_x} sources. 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.

The EnKF differs from 4-D-Var by allowing us to take advantage of the detailed chemical processes in a~CTM without developing an adjoint code.

Based on the EnKF approach, Miyazaki et~al.~(2012a) developed a~data assimilation system (CHASER-DAS) using a~global CTM CHASER (chemical atmospheric general circulation model for study of the atmospheric environment and radiative forcing). CHASER-DAS simultaneously optimizes the \chem{LNO_x} sources and the surface emissions of \chem{NO_x} and \chem{CO} as well as the concentrations of 35 chemical species including the assimilated species (\chem{NO_2}, \chem{O_3}, \chem{HNO_3}, and \chem{CO}), while taking into account the chemical interactions through error covariance. The simultaneous optimization of ~~several~~various chemical fields improves the representation of the whole chemical system and thus reduces the model--observation mismatch arising from non-\chem{LNO_x} sources and chemical processes. Therefore, this approach has the potential to improve global estimates of \chem{LNO_x} sources when compared to previous top-down approaches that optimize \chem{LNO_x} sources only. In this study, CHASER-DAS is utilized to assimilate multiple satellite datasets in order to analy~~z~~se the global \chem{LNO_x} sources, including the vertical profiles, for the wh~~i~~ole year 2007. Compared to the system described in Miyazaki et al. (2012a), several updates have been applied to the data assimilation settings on the a priori emissions and the assimilated measurements.

The rest of this paper is structured as follows: Sect.ion~2 describes the observations used for assimilation and validation. Section~3 introduces the data assimilation system. Section~4 presents the results of the global \chem{LNO_x} source estimation. Section~5 presents the regional \chem{LNO_x} structure over the Pacific and central Africa. Section~6 discusses the possible errors in the source estimation and the implications for the lightning parameterizations. Section~7 summarizes this study.

\section{Data}

\subsection{Assimilated data}

As depicted in Fig.~1, the \chem{LNO_x} source is estimated from a~simultaneous assimilation of \chem{NO_2}, \chem{O_3}, CO, and \chem{HNO_3} retrievals from satellite measurements by the Ozone Monitoring Instrument (OMI), Tropospheric Emission Spectrometer (TES), Microwave Limb Sounder (MLS), and Measurement of Pollution in the Troposphere (MOPITT) instruments. The observation operator,

\begin{equation}
y= H(x) = x_{\mathrm{a}} + \mathbf{A} (S(x) x_{\mathrm{a}}),
\end{equation}

is constructed on the basis of the spatial interpolation operator (S), the a~priori profile (x_{a}) and the averaging kernel ($\mathbf{A}$), which maps the model fields (x) into observation space (y) while taking into account the vertical averaging implicit in the observations.

This section describes the observations, with a~focus on the application of \chem{LNO_x} source estimation. More extended descriptions of the observations, their quality, and the filtering method used in CHASER-DAS are provided in Miyazaki et~al.~(2012a). Figure~1 shows how the individual satellite datasets provide information on different aspects of the chemical system in the middle and upper troposphere. Combined, these instruments provide strong useful constraints on \chem{LNO_x}. The contributions from the individual satellite sensors are highlighted below.

\subsubsection{OMI \chem{NO_2}}

Tropospheric \chem{NO_2} column retrievals obtained from the version-2 OMI DOMINO data product (Boersma et~al., 2011) are used to constrain the \chem{LNO_x} sources, the surface emissions of \chem{NO_x}, and the concentrations of \chem{NO_y} species. The overpass time of OMI (13:30) is more suitable for \chem{LNO_x} source estimation than the morning time observation by other satellite instruments (GOME, GOME-2, and SCIAMACHY), because lightning activity over land is strongest in the late afternoon and very weak in the morning (e.g. Lay et~al., 2007). We employ the super observation approach to produce representative data with a~horizontal resolution of $2.5{\degree} \times 2.5{\degree}$ for OMI \chem{NO_2} and MOPITT \chem{CO} (c.f., Sect. 2.1.4) observations, following Miyazaki et~al. (2012b). 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. This approach avoids complications caused by the small (13--24\,\unit{km}) footprint of OMI at nadir. Therefore, the data assimilation adjusts the \chem{LNO_x} sources at grid scale rather than individually at the OMI footprint scale. Further details are described in Miyazaki et~al.~(2012b).

Boersma et~al.~(2005, 2011) summarized the general error characteristics of tropospheric \chem{NO_2} retrievals. For retrievals with small values, as over the oceans, the uncertainty is dominated by the combined error from spectral fitting and stratospheric column estimation. For columns exceeding 0.5×10^{15}\,\unit{molec\,cm^{-2}}, as over most continents, the uncertainty grows due to increasing errors related to cloud fraction, albedo, and profile shape. Clouds have a~large influence on the errors and sensitivity in the measurements of the retrieved columns. Clouds below an \chem{NO_2} layer increase the effective albedo of the scene and increase the detected slant column, whereas high clouds partly screen the \chem{NO_2} column below (Boersma et~al., 2005). We employ both clear-sky data and cloud-scene data in the \chem{LNO_x} estimation because both are sensitive to \chem{NO_2} produced by lightning higher up in the atmosphere. For the cloud-covered observations the averaging kernel shows a~sharp drop roughly ~~halfway~~in the middle of –

the cloud, and very small sensitivities below. The location and magnitude of the drop is based on the cloud fraction and effective top height retrieved from the observations.

\subsubsection{TES \chem{O_3}}

The TES \chem{O_3} retrievals used are the version-4 level-2 nadir data obtained in the global survey mode (Bowman et~al., 2006). This product represents 16 orbits daily, with a~horizontal resolution of 5--8\,\unit{km}. Its vertical resolution is typically 6\,\unit{km}, with sensitivity to both the lower and upper troposphere (Worden et~al., 2004; Bowman et~al., 2006; Jourdain et~al., 2007). Jourdain et~al.~(2010) argued that the TES provides direct observations of ozone-enhanced layers downwind of convective events and thus is a~valuable dataset for estimating the vertical \chem{LNO_x} profiles. This can be attributed to its high sensitivity to \chem{LNO_x} relevant altitude layers, typically with more than one degree of freedom (DOF) for the middle and upper troposphere (from 500\,\unit{hPa} to the tropopause).

\subsubsection{MLS \chem{O_3}, \chem{HNO_3}}

The version-3.3 level-2 MLS products for \chem{O_3} and \chem{HNO_3} (Livesey et~al., 2011) are used to constrain the \chem{LNO_x} sources in the upper troposphere and the chemical concentrations in the upper troposphere and the lower stratosphere. We use data on \chem{O_3} and \chem{HNO_3} only for pressures lower than 215\,\unit{hPa} and 150\,\unit{hPa}, respectively, owing to data quality problems for higher pressures. Martin et~al.~(2007) demonstrated that \chem{O_3} and \chem{HNO_3} measurements by limb viewing spaceborne sounders have a~great potential to constrain the \chem{LNO_x} sources in the upper troposphere, based on Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) measurements.

\subsubsection{MOPITT CO}

The MOPITT \chem{CO} retrievals employed are the version-5 level-2 thermal-infrared (TIR) data (Deeter et~al., 2011, 2013). These observations are used for optimizing the surface \chem{CO} emissions and the concentrations of \chem{CO} and non-methane hydrocarbons

(NMHCs). However, the covariances between the \chem{CO} observations and the \chem{NO_x} sources are neglected in the analysis, since the error correlations are not expected to contain meaningful information, and the limited ensemble size creates spurious correlations between non- or weakly-related variables (see Sect.~3.1.2). Even so, the \chem{CO} observations indirectly affect the \chem{LNO_x} source estimation through their influence on the oxidation capacity and the \chem{NO_x} chemistry.

\subsection{Validation data}

Independent ozone observations are used to validate the performance of the data assimilation. The spatial distribution of tropospheric \chem{O_3} in the tropics is validated against the monthly mean tropospheric ozone column (TOC) derived using the OMI total columns and the MLS profiles from Ziemke et~al.~(2006) with a~horizontal resolution of $1{\degree} \times 1.25{\degree}$ (\url{http://acd-ext.gsfc.nasa.gov/Data_services/cloud_slice/new_data.html}). Ozonesonde observations taken from the database of the the Southern Hemisphere Additional Ozonesondes (SHADOZ) project (Thompson et~al., 2007) are used to validate the vertical profile of \chem{O_3} in the troposphere and the lower stratosphere. The validation is performed at ~~four~~five sites in different regions of the tropics: Costa Rica in central America (10{\degree}\,N, 84{\degree}\,W), Irene in South Africa (25.9{\degree}\,S, 28.2{\degree}\,E), Pago Pago in American Samoa (14.4{\degree}\,S, 170.6{\degree}\,W), ~~and~~ San Cristobal in Ecuador (0.9{\degree}\,S, 89.6{\degree}\,W), and Ascension in the tropical Atlantic (8.0{\degree}\,S, 14.4{\degree}\,W). -We also use the global ozonesonde observations from 39 locations taken from the World Ozone and Ultraviolet Data Center (WOUDC) database, as in Miyazaki et al. (2012a). For the purpose of comparison, all ozonesonde profiles are interpolated to a~common vertical pressure grid with a~cell size of 25\,\unit{hPa}. The model profiles are linearly interpolated to the location and time of each observation point.

\section{Methodology}

\subsection{Data assimilation system}

CHASER-DAS has been developed for the analysis of chemical compounds

in the troposphere (Miyazaki et~al., 2012a, b; Miyazaki and Eskes, 2013). This system simultaneously optimizes the \chem{LNO_x} sources and surface emissions of \chem{NO_x} and \chem{CO} as well as the predicted concentrations of 35 chemical species. With the assimilation of data on multiple species, an improved description of the chemical interactions can be obtained, especially in relation to the \chem{NO_x}-\chem{CO}-\chem{OH}-\chem{O_3} set of chemical reactions.

Miyazaki and Eskes (2013) demonstrated that multi-species data assimilation improves the analysis of surface \chem{NO_x} emissions, in comparison with an inversion derived from \chem{NO_2} measurements alone. They showed that the assimilation of measurements for species other than \chem{NO_2} changes the regional estimates of monthly mean surface \chem{NO_x} by up to $-$58\,{\%} to +32\,{\%}. These large changes emphasize that uncertainties in the model chemistry affect the quality of the emission estimates. Similar benefits may be expected from the multi-species data assimilation to improve the \chem{LNO_x} source estimation through corrections made to the concentrations of various chemical species. This is especially true for \chem{LNO_x}, because all satellite sensors are sensitive in the altitude range where \chem{LNO_x} and the ozone produced by \chem{LNO_x} resides, see Fig.~1.

\subsubsection{A~global chemical transport model CHASER}

The forecast model used is the global CTM CHASER (Sudo et~al., 2002), which describes chemical and transport processes in the troposphere. The model has a~so-called T42 horizontal resolution (2.8{\degree}) and 32 vertical levels from the surface to 4\,\unit{hPa}. CHASER is coupled to the atmospheric general circulation model (AGCM) version 5.7b of the Center for Climate System Research and Japanese National Institute for Environmental Studies (CCSR/NIES). At each time step of the model, the AGCM fields are nudged toward the reanalysis (Kanamitsu et~al., 2002) by the Atmospheric Model Intercomparison Project II of the National Centers for Environmental Prediction and US Department of Energy (NCEP-DOE/AMIP-II). Hence, the model realistically reproduces large-scale circulation while simulating sub-grid-scale convection using the cumulus convection parameterization (Arakawa and Schubert, 1974; Pan and Randall, 1998).

Anthropogenic \chem{NO_x} emissions are ~~based on~~obtained from the Emission Database for Global Atmospheric Research (EDGAR) version 4.2. Emissions from biomass burning are based on the Global Fire Emissions Data base (GFED) version 3.1 (van der Werf et~al., 2010). Emissions from soils are based on monthly mean Global Emissions Inventory Activity (GEIA) (Graedel et~al., 1993). A~diurnal variability scheme is implemented for the surface \chem{NO_x} emissions, depending on the dominant category for each emission category (Miyazaki et~al., 2012b). The total \chem{NO_x} emission by aircraft is obtained from EDGAR as 0.55\,\unit{Tg\,N\,yr^{-1}}~~.~~, which is similar to a more recent estimate of 0.49 \,\unit{Tg\,N\,yr^{-1}} for 2004 (Wilkerson et al., 2010). 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.

\subsubsection{Local ensemble transform Kalman filter}

The EnKF uses an ensemble forecast to estimate the background error covariance matrix. The advantage of the EnKF over 4D-VAR is its easy implementation for complicated systems; i.e., it does not require the development of an adjoint code. The EnKF data assimilation technique employed is local ensemble transform Kalman filter (LETKF, Hunt et~al., 2007). The LETKF scheme, which is based on the ensemble square root filter (SRF) method (e.g., Whitaker and Hamill, 2002), generates an analysis ensemble mean and covariance that satisfy the Kalman filter equations for linear models. The LETKF has conceptual and computational advantages over the original EnKF. The analysis performed locally in space and time reduces sampling errors caused by limited ensemble size, which also enable us to perform parallel computation. The computational advantages are important for this study because of the large state vector size.

In the forecast step, a~background ensemble, $\vec{x}^b_i (i=1,...,k)$, is globally obtained from the evolution of each ensemble model realisation, where $\vec{x}$ represents the model variable; b the background state; and k the ensemble size (i.e., 48 in this study). An ensemble of background observation vectors in the observation space, $\vec{y}^b_i=H(\vec{x}^b_i)$, is then estimated using the observation operator H. The observation operator (H) is constructed on the basis of the spatial interpolation operator (S), the a~priori profile ($\vec{x}_{apriori}$) and the averaging

kernel ($\vec{A}$), which maps the model fields ($\vec{x}$: N- (the system dimension) dimensional state vector) into
observation space ($\vec{y}$: p- (the number of observation) dimensional observational vector) while taking into account the vertical averaging implicit in the observations as follows:
\begin{equation}
\vec{y}= H(\vec{x}) = \vec{x}_{apriori} + \vec{A} (S(x)-\vec{x}_{apriori}),
\end{equation}

The spatial interpolation operator (S) is first applied to the model fields $\vec{x}$ in order to interpolate to the horizontal location of each observation and the height of each of the vertical layers. The averaging kernel ($\vec{A}$) is then applied to define the sensitivity of the satellite retrieved state to changes to the true state. For weak absorbers, the a priori profile ($\vec{x}_{apriori}$) does not, or only weakly, influence the relative model-observation difference (Eskes and Boersma, 2003). The averaging kernel ($\vec{A}$) and the a~priori profile ($\vec{x}_{apriori}$) information provided for each retrieval is used in the data assimilation.

A~background ensemble mean in the observation space, $\overline{\vec{y}^b}=\frac{1}{k} \sum\nolimits^k_{i=1} \vec{y}_i^b$, or in the model space, $\overline{\vec{x}^b}=\frac{1}{k} \sum\nolimits^k_{i=1} \vec{x}_i^b$, and an ensemble of background perturbations in the observation space, $\mathbf{Y}^b=\vec{y}^b_i-\overline{\vec{y}^b}$, or in the model space, $\mathbf{X}^b=\vec{x}^b_i-\overline{\vec{x}^b}$, are also computed.

~~In the analysis step, an ensemble of background observation vectors in the observation space, $\vec{y}^b_i=H(\vec{x}^b_i)$, is estimated using the observation operator H.~~

~~A~background ensemble mean in the observation space, $\overline{\vec{y}^b}=\frac{1}{k} \sum\nolimits^k_{i=1} \vec{y}_i^b$, or in the model space, $\overline{\vec{x}^b}=\frac{1}{k} \sum\nolimits^k_{i=1} \vec{x}_i^b$, and an ensemble of background perturbations in the observation space, $\mathbf{Y}^b=\vec{y}^b_i-\overline{\vec{y}^b}$, or in the model space, $\mathbf{X}^b=\vec{x}^b_i-\overline{\vec{x}^b}$, are also computed.~~

The ensemble mean analysis is then updated by
\begin{equation}

\overline{\vec{x}^{\mathrm{a}}}=\overline{\vec{x}^b}+ \mathbf{X}^b \tilde{\mathbf{P}}
^{\mathrm{a}} \big(\mathbf{Y}^b\big)^{\mathrm{T}} \mathbf{R}^{-1} \big(\vec{y}
^{\mathrm{o}} - \overline{\vec{y}^b}\big),
\end{equation}
where $\vec{y}^{\mathrm{o}}$ is the observation vector, and $\mathbf{R}$
is the $p \times p$ observation error covariance ~~(p is the number of observation)~~. The observation error information is obtained for each retrieval, which includes the smoothing error, the model parameter error, the forward model error, the geophysical noise, the instrument error, and the representativeness error (see Miyazaki et al. 2012a, for details).

$\tilde{\mathbf{P}}^{\mathrm{a}}$ is the local analysis
error covariance in the ensemble space,
\begin{equation}
\tilde{\mathbf{P}}^{\mathrm{a}} =
\left[(k-1) I + \big(\mathbf{Y}^b\big)^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}^b \right]^{-1}.
\end{equation}

The new analysis ensemble perturbation matrix in the model space
$\mathbf{X}^{\mathrm{a}}$ is obtained by transforming the background ensemble
$\mathbf{X}^b$,
\begin{equation}
\mathbf{X}^{\mathrm{a}} =
\mathbf{X}^b \left[(k-1) \tilde{\mathbf{P}}^{\mathrm{a}} \right]^{1/2}.
\end{equation}

–

The new background error covariance used in the next forecast step is
obtained from an ensemble simulation with the analysis ensemble.

In –
the analysis, a~covariance localization was applied to neglect the
covariance among non- or weakly-related variables. 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. For the
optimization of \chem{LNO_x} sources, the covariances with TES
\chem{O_3}, OMI \chem{NO_2}, MLS \chem{O_3}, and MLS \chem{HNO_3}
observations are considered, while those with MOPITT \chem{CO} data
are not. MOPITT \chem{CO} data affect the concentrations of CO,
hydrocarbons, and formaldehyde only. Surface emissions of \chem{NO_x}
and \chem{CO} are optimized using OMI \chem{NO_2} data and MOPITT

\chem{CO} data, respectively (see Miyazaki et~al.,~2012a, for details). The covariance localization helps to avoid sampling errors resulting from the limited ensemble size and to improve the \chem{LNO_x} analysis.

The localization is also applied to avoid the influence of remote observations for improving the filter performance. The influence of an observation is cut off when the distance between the observation and an analysis point is larger than $2L\times\sqrt{10/3}$ based on the formulation of Gaspari and Cohn (1999). The localization scale L is 600\,\unit{km} in our setting. As a~result, the analysis is solved at every grid point by choosing nearby observations. The emission and concentration fields are updated at an analysis interval of every 100\,\unit{min}~~, with the ensemble size of 48~~.

In conclusion, the data assimilation updates model variables (the concentrations and the emission multiplication factors) for every grid point. This analysis is based on the observational information (i.e., the satellite retrievals) and the background error covariance estimated from the ensemble forecast with 48 members in our case. The estimated concentrations and emissions are used as initial conditions in the next step of ensemble model simulations and updated at every analysis step (i.e., 100 min.). Further details are described in Miyazaki et~al.~(2012a).

\subsection{\chem{LNO_x} estimation}

\subsubsection{Parameterization}

Lightning is routinely monitored from ground-based networks and detected from satellite instruments. Nevertheless, these data cannot be directly used in CTM simulations. The ground-based operational lightning detection networks (e.g., the World Wide Lightning Location Network (WWLLN)) provide lightning maps but they have low detection efficiencies (Abarca et al., 2010), whereas satellite instruments provide~~, because of the very small amount of global~~ limited coverage on a~daily basis. Thus, a~parameterization is required in order to estimate the lightning flash frequency in CTM simulations. The global distribution of the flash rate is calculated in CHASER for convective clouds on the basis of the observed relation between the lightning activity and the cloud top height (Price and Rind, 1992) in the AGCM at each forecast step. In this approach, high clouds are expected to exhibit strong lightning activity. The frequencies of lightning over

the continents and the oceans are estimated separately as follows (Price et~al., 1997): $F_{\mathrm{c}}= 3.44 \times 10^{-5} \times H^{4.92}$ (\unit{flashes\,min^{-1}}) over continents; $F_{\mathrm{m}}=6.40 \times 10^{-4} \times H^{1.73}$ (\unit{flashes\,min^{-1}}) over ocean, where F is the total flash rate (\unit{flashes\,min^{-1}}) and H is the cloud top height (\unit{km}), in which a. 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.

The simulated average global flash rate for 2007 is 41.2\,\unit{flashes\,s^{-1}}, which is comparable to the climatological estimates of 44-\pm$-5$\,\unit{flashes\,s^{-1}} derived from the Optical Transient Detector (OTD) measurements (Christian et~al., 2003) and 46\,\unit{flashes\,s^{-1}} derived from the Lightning Imaging Sensor (LIS) and OTD measurements (Cecil et al., 2014).- The difference between the model simulation and the observations is partly attributed to interannual variations in flash activity; the annual total flash rate for the latitude band 35 {\degree}\,S-35 {\degree}\,N in 2007 observed from the LIS measurement is about 3 \% lower than those from the climatology. Because only LIS measurements are available in 2007 and because the global coverage was not obtained, this study uses the climatological observations obtained from a combination of LIS and OTD measurements to validate the global flash rate.

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 parameterisation. Compared with the observations, the global distribution of the total flash rate is generally reproduced by the model. The simulated global flash rate shows a~maximum in boreal summer and a~minimum in boreal winter, with frequent occurrences over Central Africa, South America, and the maritime continent. These features are commonly found in the climatological observations. Conversely, in comparison with the observed flash activity, the simulated flash activity is stronger over Northern South America but weaker over central Africa and most of the oceanic ITCZ. These systematic differences found in studies using the scheme of Price and Rind (1992) have been reported before (e.g. Allen and Pickering, 2002; Labrador et~al., 2005; Martin et~al., 2007). Mainly because of the low bias over central Africa, the model – underestimates the annual flash rate in the tropics (20{\degree}\,S--20{\degree}\,N) by about 27 \%, leading to about 13 \% underestimation in the global total flash rate.

The \chem{LNO_x} source is estimated at each grid point of CHASER by using the simulated lightning activity and making several assumptions. First, the ~~ratio~~CG proportion of total flashes, z, ~~between the IC flashes and CG flashes~~ is estimated as a~function of the cold cloud thickness (ΔH for < 0\,{\degree}C) on the basis of the following relationship (Price ~~et~al.~~and Rind, 199~~7~~3):

\begin{equation}
z = 0.021\times\Delta H^4 - 0.648\times\Delta H^3 + 7.493\times\Delta H^2 - 36.54\times\Delta H + 63.09.
\end{equation}

This relationship is applied for clouds with $\Delta H>5.5$\,\unit{km}, while z is set to zero for clouds with $\Delta H<5.5$\,\unit{km} based on the observation that low clouds almost exclusively have IC flashes during the growth stage. Second, following Price et~al. (1997), the \chem{LNO_x} source amounts are calculated on the basis of a~lightning NO production of 1100 moles per CG flash and 110 moles per IC flash, with a~mean energy per CG flash of 6.7×10^9\,\unit{J\,flash^{-1}}~~(Price et~al., 1997)~~. Third, the vertical profiles of the \chem{LNO_x} sources are determined on the basis of the C-shaped profile given by Pickering et~al.~(1998), with a~first maximum in the upper troposphere and a~second maximum in the lower troposphere. The three profiles provided by Pickering et al. (1998) is averaged and applied in the parameterization. The global annual total amount of the \chem{LNO_x} source for 2007 was estimated at 4.7\,\unit{Tg\,N\,yr^{-1}} in the CHASER simulations.

\subsubsection{Optimization by data assimilation}

The multiplication factors for the \chem{LNO_x} production rate (\unit{mol\,s^{-1}}) and the \chem{NO_x} surface emissions are optimized in the assimilation analysis step, by adding them to the state vector together with the forecast variables (i.e. concentrations). In this approach, the background error covariance matrix estimated from the ensemble simulations is used to obtain best estimates of the source factors at each grid point of the model. Figure~~2~~3 shows the mean background error covariance structure between the \chem{LNO_x} source and the concentrations of various species at different altitudes over central Africa in July. The concentrations of chemical species such as \chem{O_x}, \chem{NO_x}, \chem{N_2O_5}, \chem{HNO_3}, \chem{HNO_4}

exhibit high positive correlations with the \chem{LNO_x} sources in the middle and upper troposphere. The high correlations indicate the utility of measurements of these species to constrain the \chem{LNO_x} sources. The background error covariance varies in time and space in the EnKF approach, reflecting variations in the lightning activity and the chemical concentrations. The data assimilation optimizes the multiplication factors for the total \chem{LNO_x} sources, and does not separately optimise for IG and CG flashes.

In order to provide meaningful constraints on \chem{LNO_x}, the observation error of each retrieval must be sufficiently small compared to lightning signal. As shown in Fig.~~3~~4, the CHASER simulations with and without the lightning signal in the tropospheric \chem{NO_2} columns gives the magnitude of the boreal summer (June--August) mean lightning signal as roughly $1\sim5\times 10^{14}$\,\unit{molec\,cm^{-2}} over the tropical Atlantic, Africa, and India but roughly 1×10^{14}\,\unit{molec\,cm^{-2}} over remote land sites. The large enhancements are over, and downwind of, active convective regions (e.g. over the tropical Atlantic), as found by Martin et~al.~(2007). These signals are large compared to the local super-observation error of the OMI retrievals, in which the total super observation error is computed as a~combination of the measurement error and the representativeness error as in Miyazaki et~al.~(2012b). The super observation error is relatively small because of the large number of OMI observations per grid cell. Large enhancements in \chem{O_3} due to lightning are observed in the tropical upper troposphere, with contributions as large as 13\,\unit{ppbv} to the mean concentration, consistent with the analyses of Sauvage ~~Cooper~~ et~al.~(200~~6~~7) ~~and Hudman et~al.~(2007).~~. These signals are slightly less than or nearly equal to the mean TES observation error. The mean MLS observation errors are generally much larger than the lightning signals, especially for \chem{HNO_3}. Although the mean ratio of the lightning signals to the retrieval error is small, a~large number of observations can still provide constraints on \chem{LNO_x} estimates.

The observed concentrations reflect contributions not only from \chem{LNO_x} but also from other sources such as surface emissions and inflows from the stratosphere. Any model errors in the other sources will produce model--observation mismatches that will negatively affect the accuracy of the estimated \chem{LNO_x} source in the top down

framework. To avoid difficulties related to contributions from surface sources, for instance, Boersma et~al.~(2005) focused on situations downwind of storm systems over areas relatively free of pollution. In contrast, our analysis simultaneously corrects the various model error sources, which benefits the \chem{LNO_x} source analysis even over polluted regions. Moreover, the combined use of the multiple datasets with different vertical sensitivities is expected to facilitate the estimation of the vertical \chem{LNO_x} profile and to distinguish between the surface \chem{NO_x} emissions and \chem{LNO_x} sources. The MLS and TES data not only provides information on the \chem{NO_x} source amount in the middle and upper troposphere, but also constrain the the lower tropospheric source when combined with the OMI tropospheric column \chem{NO_2} retrievals.

The a~priori error is set to 40\,{\%} for the surface emissions of \chem{NO_x} and \chem{CO} and 60\,{\%} for the \chem{LNO_x} sources, considering the large uncertainties in the lightning parameterization. Since no model error term is implemented for the source factors during the forecast step, the background error covariance may continuously decrease and become underestimated through the data assimilation cycle. Thus, we apply covariance inflation to the source factors to prevent covariance underestimation in the analysis step, as was done for the surface emission in Miyazaki et~al.~(2012a). The standard deviation is artificially inflated to a~minimum predefined value (i.e., 30\,{\%} of the initial standard deviation) at each analysis step for both the surface and lightning sources. The sensitivity of the analysis to the choice of these parameters is discussed in Sect.~6.~~2~~1.3.

\section{Results}

\subsection{Global \chem{LNO_x} source distributions}

Data assimilation increases the global annual \chem{LNO_x} source from 4.7 to 6.3\,\unit{Tg\,N\,yr^{-1}}, as summarized in Table~~1~~2. The instantaneous uncertainty estimated from the mean analysis spread for the global source is $\pm$0.9\,\unit{Tg\,N\,yr^{-1}}, while the chi-square (χ^2) diagnostic gives confidence in the estimated analysis spread as a~measure of the analysis uncertainty (c.f., Sect.~6.1.3). This spread of the ensemble is relatively small, showing

that \chem{LNO_x} is well constrained by the assimilation. The Northern Hemisphere (NH, 20--90{\degree}\,N), the tropics (20{\degree}\,S--20{\degree}\,N), and the Southern Hemisphere (SH, 90--20{\degree}\,S) contribute 31\,{\%}, 56\,{\%}, and 13\,{\%} to the analyzsed global source, respectively. These relative contributions are slightly modified from the a~priori sources. The analyzsed global source is maximal in July, primarily resulting from the seasonal variation in the NH, as depicted in Fig.~45.

Figure~56 compares the simulation with the assimilation in terms of the global distribution of annual total \chem{LNO_x} sources. Data assimilation substantially increases the annual sources over central Africa, the central and eastern United States, northern Eurasia, South America, south and southeast Asia, the maritime continent, and over the tropical oceans around the ITCZ (left panels). The seasonal amplitudes are also enhanced by 10--40\,{\%} over most of these regions (middle panels). Data assimilation introduces distinct seasonal variations in sources over the oceans, especially along the ITCZ. The maximum \chem{LNO_x} sources mostly appear in June or July over the NH temperate regions and in December or January over the SH continents (right panels). Compared to those over most other NH regions, the peak sources over northern Africa and India occur 1--2 months later (i.e., in July--August), reflecting the local convective activity during the African and Asian monsoons, respectively. The timing of the peak sources is very similar between the simulation and the assimilation, because the seasonal variation of the lightning flash frequency is generally well predicted by the model as compared with the OTD climatology (figure not shownc.f., Table 1). However, data assimilation changes the peak time by 1--2 months over central Africa and South America.

\subsection{Regional \chem{LNO_x} source distributions}

Figure~67 shows the seasonal variations of the \chem{LNO_x} source over the 11 regions shown in the top panel. Over the NH continents (e.g. North America, Europe, and northern Eurasia), the broad summertime source peak is predicted by the model, where the monthly peak source strength is further increased by 30--60\,{\%} due to assimilation. Over northern Eurasia, the large increase in the summer

dominates the approximately 40\,{\%} increase in the annual total source. The relative increase is more constant with season over Europe.

The seasonal variations of the sources differ significantly among the regions in the tropics, reflecting the locality of convection and monsoon activity. The predicted sources exhibit broad maxima in the rainy seasons over the tropical continental regions; that is, from October to April over South America, from April to September over Northern Africa, and from May to September over Southeast Asia. Data assimilation generally enhances the seasonality of the tropical sources, e.g. producing two source maxima in May and July over northern Africa. The sources over South America are increased by about 40\,{\%} from October to February, which is primarily attributed to enhanced sources over the Amazon during the South American monsoon.

Over the oceans, the retrieval uncertainties of OMI, TES, and MOPITT measurements are generally larger compared to over land. Nevertheless, substantial changes in the regional \chem{LNO_x} sources are obtained by assimilation over the Pacific, the Indian Ocean, and the Atlantic in the tropics, especially along the ITCZ, because of large lightning signals in the chemical concentrations. Data assimilation produces obvious source maxima in boreal late winter over the Atlantic and in boreal autumn--early winter over the Indian Ocean. The west African and south Asian monsoons, respectively, may be responsible for these enhancements where the model failed to predict the distinct seasonality. 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 \chem{NO_x} production efficiency over the oceans in the model simulation. The \chem{LNO_x} sources over the tropical Pacific are largely increased throughout the year, with a~mean factor of about two. As a~result, the data assimilation increases the annual global source by about 56\,{\%} over the oceans. The relative increase of the total source is smaller over land (i.e., 32\,{\%}) than over the oceans.

\subsection{Vertical distribution of the \chem{LNO_x}source}

Figure~~7~~8 shows the changes in the vertical profiles of the \chem{LNO_x} sources. Data assimilation increases the \chem{LNO_x} sources over most land regions by 20--50\,{\%} in the upper troposphere, with a~maximum increase at 240\,\unit{hPa} in the global and annual mean, which is attributed to the source increase in the tropical upper troposphere~~.~~. The corrections below the middle troposphere are much smaller. Compared to other land regions, the source increase in the upper troposphere is much smaller over Australia. 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. A~roughly twofold increase in the upper tropospheric sources occurs over the Pacific. In the annual and global mean, the \chem{LNO_x} sources over the oceans are increased by 40--50\,{\%} in the lower troposphere and by up to 65\,{\%} in the upper troposphere. We note that errors in the OMI tropospheric \chem{NO_2} column retrievals could cause large uncertainties in the analyzed \chem{LNO_x} sources over the oceans, as will be discussed in Section 6.1.1.

Table~~2~~3 summarizes the altitude (in km) with maximum \chem{LNO_x} emission (i.e., source peak height). The predicted source peak heights of the regional sources over land mostly range from 5\,\unit{km} to 12\,\unit{km} but exceed 13\,\unit{km} locally over central Africa and central America. Data assimilation generally lowers the regional mean peak source heights, with large decreases at low latitudes over land. Substantial decreases in the peak heights of the annual sources occur over Southeast Asia ($-$0.74\,\unit{km}), Australia ($-$0.41\,\unit{km}), and southern Africa ($-$0.38\,\unit{km}), whereas the changes are small over North America and northern Eurasia. Over the oceans, and in particular over the tropical western Pacific, the peak height is substantially lowered by about 1.2\,\unit{km}.

\subsection{Relative contributions of individual assimilated datasets}

The effects of individual assimilated datasets on the estimated \chem{LNO_x} sources were evaluated through observing system experiments (OSEs) by separately assimilating each dataset with CHASER-DAS (Fig.~~8~~9). The assimilation of OMI \chem{NO_2} measurements dominates the overall structure of the total analysis increment. The

spatial distribution in the monthly mean analysis increment of the source column reveals a~high correlation between the full assimilation and the OSE with OMI measurements ($r=0.55$--0.60). This demonstrates the dominant role of the OMI \chem{NO_2} retrievals in determining the \chem{LNO_x} source distribution in the analysis.

The measurements of species other than \chem{NO_2} also provide important constraints on the global source estimation. The assimilated datasets all have an impact on the global source analysis, but with different contributions from individual datasets, as summarized in Table~~3~~4. The vertical and latitudinal structure of the analysis increments obtained from the assimilation of MLS \chem{O_3} and \chem{HNO_3} data are generally similar (Fig.~~8~~9), revealing consistent constraints from datasets gathered for different species but with the same instrument. In contrast, the respective corrections from MLS \chem{O_3} and TES \chem{O_3} measurements show differences in magnitude and distribution. This arises from differences between the two sets of measurements in the coverage, vertical sensitivity, and systematic error. The negative analysis increments in the upper tropospheric \chem{LNO_x} sources obtained from the assimilation of TES \chem{O_3} data likely arises from the TES negative bias (up to 20\,{\%}) from the upper troposphere to the lower stratosphere in the southern subtropics, see – e.g. Nassar et~al.~(2008),~~.~~ whereas those in the northern mid-latitudes may be associated with the positive bias in the simulated \chem{O_3} (Miyazaki et al., 2012a).

The influences measured by the OSEs mostly reflect the effect of direct source optimization through the background error covariance. In addition, each retrieval indirectly affects the source estimation through adjustments made to the various concentration fields. For instance, the assimilation of the MLS retrievals corrects the concentrations of \chem{O_3} and \chem{NO_y} species in the stratosphere, which has the potential to improve the modelled impact of stratospheric air on tropospheric concentrations and benefits the source estimate derived from tropospheric column. Furthermore, the assimilation of MOPITT \chem{CO} data changes the free tropospheric OH concentration by up to 5\,{\%}, corresponding to an increase of about 25\,{\%} in the annual NH total \chem{CO} emissions (Miyazaki et~al., 2012a), which changed the monthly global \chem{LNO_x} source by up to

5\,{\%} in combination with the other satellite datasets. The simultaneous assimilation also has the ability to constrain ozone production efficiencies (OPE) through the \chem{NO_x}-\chem{CO}-\chem{OH}-\chem{O_3} set of chemical reactions, which may improve the \chem{LNO_x} source estimation with the assimilation of TES \chem{O_3} data. Detailed analyses are required to measure the impact of the simultaneous assimilation on OPE.

In most previous studies, \chem{NO_2} measurements were used to optimise only the \chem{LNO_x} production. In this case, the accuracy of the \chem{LNO_x} source estimates is negatively affected by various sources of error, including the surface \chem{NO_x} and \chem{CO} emissions and inflows from the stratosphere. We found differences reaching about 50\,{\%} on the regional scale when comparing the multi-species assimilation and a~simpler \chem{LNO_x} source (i.e. single-parameter) inversion derived from \chem{NO_2} measurements only. The single-parameter approach tends to have larger sources (e.g., the estimated global \chem{LNO_x} source was 8.89\,\unit{Tg\,N\,yr^{-1}} for January and 12.8\,\unit{Tg\,N\,yr^{-1}} for July), because of the analysis increment introduced to compensate for other sources of error. Miyazaki and Eskes (2013) showed that the a~priori surface \chem{NO_x} emissions are possibly underestimated by up to 67\,{\%} on the regional scale. The simulated \chem{O_3} concentrations in the stratosphere are also biased (Miyazaki et~al., 2012a). These sources of error leads to an excessive model--observation mismatch in the \chem{LNO_x} source inversion. Note that, because of the short assimilation cycle (i.e., 100\,\unit{min}.), the \chem{LNO_x} source can be frequently and continuously increased to compensate for persistent model errors. This situation may cause larger \chem{LNO_x} sources in our estimates compared to those with a~longer (e.g., one month) assimilation cycle especially when only \chem{LNO_x} sources are optimized.

\section{Tropical regions}

Lightning strongly influences the \chem{O_3} production and chemistry, especially in the tropical troposphere, as discussed in Sect. 3.2.2 and suggested by Sauvage et al. (2007a). Lightning activity and surface \chem{NO_x} sources differ considerably among the tropical regions, reflecting variations in the meteorological conditions including cumulus convection activity. This section demonstrates the ability of CHASER-DAS to analyse the \chem{LNO_x} sources and \chem{O_3} distributions in several tropical regions.

\subsection{Validation of tropospheric ozone}

The validation of ozone profiles provides useful information on the performance of the \chem{LNO_x} source estimation, because of the strong chemical links between the \chem{LNO_x} sources and the \chem{O_3} concentration, which influence the simultaneous optimization in the data assimilation. Figure~~9~~10 shows a~comparison of modeled and analysed vertical \chem{O_3} profiles with the ozonesonde observations taken at f~~our~~ive tropical SHADOZ sites. The assimilation removes most of the \chem{O_3} bias in the upper troposphere when the predicted lightning activity is maximal; namely, from June to August over Costa Rica, from December to February over Irene, and from September to November over American Samoa. Data assimilation also removed most of the free tropospheric negative bias over San Cristobal throughout the year and over Ascension from December to February and from September to November. Sauvage et al. (2007b) suggested great lightning contributions to \chem{O_3} concentrations over the tropical Atlantic. Conversely, the assimilation does not obviously improve the lower tropospheric \chem{O_3}. Ozonesonde observations from 39 locations (9 locations in the tropics) have been used to validate the global ozone profiles (see Sect.~6.3). In the tropics, the data assimilation reduces the mean ozone concentration bias: by 11 \% in the lower troposphere (750--450 hPa), by 63 \% in the middle troposphere (450--200 hPa), and by 79 \% in the upper troposphere (200--90 hPa) in January. Similar improvements were reported before by Miyazaki et al. (2012a).

Table~45 summarizes the validation results of the monthly mean tropospheric ozone column (TOC) against the MLS and OMI measurements in the tropics. The general spatial structures observed are well captured by both the simulation and the assimilation, as confirmed by the high spatial correlation ($r=0.85\sim0.92$). The CTM simulation overestimates the TOC in the northern subtropics, especially over northern Africa, India, and eastern Asia (figure not shown). Large model errors are also found over the western – Atlantic, where the simulated ozone is too low in January and too high in July. Data assimilation mitigates these errors, removing most of the bias and reducing the global RMSE by 20--30\,{\%}. The improved agreement with TOC data is mainly attributed to the assimilation of TES \chem{O_3} (Miyazaki et al., 2012a). Because lightning substantially influences the amount of \chem{O_3} in the

tropics, and because the data assimilation simultaneously optimizes the \chem{O_3} and the \chem{LNO_x} source, significantly improved agreement with ~~the~~ independent ozone observations gives confidence in the performance of the \chem{LNO_x} estimates.

\subsection{The tropical western Pacific}

The tropical western Pacific is a~region with active cumulus convection which substantially influences the vertical profile of chemical compounds. The warm sea surface and high convective available potential energy (CAPE) activate vertical uplifting and lightning especially over the maritime continent. The parameterization for the flash rate estimation assumes that the meteorological fields in the model represent the deep convection that generates lightning. Large uncertainties in the ~~simulated clouds and the~~ \chem{LNO_x} sources are expected over the tropical western Pacific because of errors in the tropical Pacific ITCZ cumulus clouds simulated by the AGCM (Emori et al., 2005). ~~, since these are heavily parameterized in the model.~~

Figure~1~~0~~1 shows the regional distribution of total cloud fraction and tropospheric concentrations of \chem{NO_2} and \chem{O_3} over the tropical western Pacific in mid-August. Compared to the OMI cloud data, the AGCM shows systematic errors in the location and total cloud fraction of the ITCZ. Although the meteorological conditions in the AGCM are nudged toward the NCEP-II reanalysis, the use of the cumulus parameterization causes a~large uncertainty in the simulated cloud and \chem{LNO_x} source structures~~.~~. Accurate simulations of the cloud position are important to properly distribute the \chem{LNO_x} sources, while errors in the simulated cloud top height lead to uncertainties in the total source strength. The regional mean strength of the \chem{LNO_x} sources is increased by a~factor of two due to data assimilation, but the spatial distribution is only slightly modified because of the sparse coverage of the observations and large uncertainty in individual data.

The analysed \chem{NO_2} and \chem{O_3} concentrations show better agreements with the observations (Fig. 11) because of the simultaneous data assimilation.

The simulation underestimates the regional mean \chem{NO_2} concentration in the upper troposphere by 65\,{\%} compared to the OMI \chem{NO_2} retrieval, whereas data assimilation removes most of the

bias. Agreements with TES \chem{O_3} fields are also greatly improved by data assimilation, despite the sparse coverage of the TES measurements. Because of the simultaneous optimization of the \chem{LNO_x} sources and these chemical concentrations, the improvements suggest that the regional total \chem{LNO_x} sources are reasonably estimated by data assimilation. In contrast, to better represent the observed fine structure associated with individual occurrences of cloud, a~high-resolution model is required. This is discussed further in Sect.~6.1.2.

\subsection{Central Africa}

Africa is the region with the strongest lightning activity and the largest source of biomass burning over the globe (e.g. Christian et~al., 2003). In this region, large uncertainties in the simulated cumulus cloud and biomass burning activity are expected, as suggested by Emori et al. (2005) and Stroppiana et al. (2010), respectively, which will cause errors in the predicted \chem{LNO_x} sources and chemical concentrations. \chem{NO_2} concentrations exhibit distinct vertical and latitudinal variations over western Africa (Fig.~1~~1~~2). These variations are produced by various factors such as convective uplifting, stratospheric inflows, surface sources, and lightning sources. The lower tropospheric \chem{NO_2} concentrations are maximal over northern (around 5--20{\degree}\,N) and central (15~~3~~{\degree}\,S--Equator) Africa, owing to in situ emissions from biomass burning. In the upper troposphere, strong \chem{LNO_x} sources cause a~maximum \chem{NO_2} concentration over northern Africa (5--15{\degree}\,N). Data assimilation increases the surface \chem{NO_x} emissions over northern and central Africa by up to 90\,{\%} and the \chem{LNO_x} source over northern Africa by about 50\,{\%}, which acts to strengthen the local maximum in \chem{NO_2} concentrations.

The tropospheric \chem{O_3} distributions also show distinct variations over central Africa. The surface \chem{NO_x} and \chem{CO} emissions and \chem{LNO_x} sources lead to \chem{O_3} production, whereas the inflows from the stratosphere along the subtropical jet stream predominantly determine the latitudinal \chem{O_3} variations in the upper troposphere. The multi-species data assimilation provides

comprehensive constraints on these processes and the \chem{O_3} variations. The assimilation of TES data modifies the \chem{O_3} distribution around the African monsoon circulation, whereas the assimilation of MLS observations improves the stratospheric inflow (figure not shown). Assimilation of MOPITT and OMI measurements provides important constraints on the OH fields and chemical \chem{O_3} production. These case studies demonstrate once more the utility of the simultaneous data assimilation for regional process studies.

\section{Discussion}

\subsection{Uncertainty of the \chem{LNO_x} source estimate}

\subsubsection{Systematic satellite observation errors}

The quality of the assimilated measurements largely influences the \chem{LNO_x} source uncertainty. Boersma et~al.~(2011) showed that the different OMI \chem{NO_2} retrievals have biases up to 40\,{\%} because of errors in the retrieval processes. Although this study uses the latest improved retrievals, systematic errors will still be present in the retrievals. We performed a~sensitivity experiment by adding an artificial positive bias of 15\,{\%} to the OMI \chem{NO_2} retrievals. This increased the monthly regional and global \chem{LNO_x} sources by up to 14\,{\%} and 3\,{\%}, respectively, as summarized in Table~~5~~6. It is emphasized that low \chem{NO_2} concentrations over the oceans are mostly smaller than the OMI noise level. Errors related to the separation of stratospheric and tropospheric \chem{NO_2} could also cause errors in the OMI tropospheric \chem{NO_2} column retrievals (Lamsal et al., 2010; Boersma et al., 2011). These may cause large uncertainties in the analyzed \chem{LNO_x} sources, especially over the oceans.

Any bias in the measurements of species other than \chem{NO_2} also affects the quality of the sources estimated in the simultaneous assimilation framework. TES \chem{O_3} data are known to have positive biases in the upper troposphere compared to ozonesonde observations (Worden et~al., 2007; Nassar et~al., 2008). A~sensitivity experiment was performed in which a~bias correction for TES \chem{O_3} data was applied. This consisted of a~uniform 3.3\,\unit{ppbv} above 500\,\unit{hPa} and 6.5\,\unit{ppbv} below 500\,\unit{hPa}, as recommended by Worden et~al.~(2009). The result of this experiment suggests that such a~bias in TES \chem{O_3}

increases the estimated monthly global \chem{LNO_x} source by up to 14\,{\%} (Table~~5~~6). Implementing a~reasonable bias correction scheme for individual retrievals is therefore clearly important to obtain unbiased source estimates.

The presence of clouds influences both the quality and sensitivity of the satellite retrievals. The retrieval uncertainty is increased by errors related to cloud fraction, while the retrieval sensitivity at cloud top or above is enhanced by multiple scattering and the high albedo (Boersma et~al., 2005). In situations of high cloud with strong lightning activity, a~large fraction of the \chem{LNO_x} reaches the top and anvil of the cloud (Ridley et~al., 1996). The \chem{LNO_x} produced can reside above the cloud for several days, because of the long lifetime of \chem{NO_x} in the upper troposphere, and the enhanced concentrations may be detected by remote observations. Furthermore, an instrument like OMI is sensitive to \chem{NO_2} in the upper part of thunderstorm clouds. In most previous studies, however, cloud-covered observations were simply ignored to avoid anticipated retrieval complications. A~sensitivity experiment was performed in which cloud-covered OMI \chem{NO_2} observations were removed when the cloud radiance fraction is larger than 50\,{\%}. This experiment confirmed that the cloud-covered data have an impact on the \chem{LNO_x} source estimation, as summarized in Table~~5~~6 and depicted in Fig.~1~~2~~3. The large changes associated with the cloud-covered data are found above roughly 450\,\unit{hPa} in the tropics and in the NH, with source increases up to about 40\,{\%} in the NH subtropical upper troposphere in July. The global \chem{LNO_x} source is increased by 12\,{\%} in July when cloud-covered observations are used. These results imply that cloud-covered OMI \chem{NO_2} data contain important information on the \chem{LNO_x} source amounts above and inside clouds. However, errors in the simulated cloud profiles (c.f., Fig.~1~~0~~1), and in the retrieved cloud-altitude dependent averaging kernels, may cause biases in the cloudy scene analysis.

The lifetimes of \chem{NO_x}, \chem{HNO_3}, and \chem{O_3} are much longer in the upper troposphere than in the lower troposphere, and also the satellite observations (e.g. TES, MLS, and cloud-covered OMI observations) show a~higher sensitivity in the middle to upper troposphere. This situation also helps to provide constraints on the

\chem{LNO_x} source far from the convective clouds that produced the \chem{LNO_x}. The large spread of the different estimates indicates that estimates of the \chem{LNO_x} source distribution and the global \chem{LNO_x} source amount are highly sensitive to the satellite data used. Subsequent use of new measurements is expected to influence the source estimation to a~considerable degree.

\subsubsection{Systematic model errors}

The mismatch between the simulated and observed concentrations is partly caused by various sources of error in the model. For instance, errors in the lifetime of \chem{NO_x}, the emissions of ozone precursors, the inflows from the stratosphere, and the parameterization of convective transport will dominate the mismatch and thus may have an impact on the quality of the \chem{LNO_x} source estimation (e.g., Jourdain et~al., 2010). As suggested in Sect.~4.4, the use of multiple datasets causes the analysis increment of the \chem{LNO_x} sources to become smaller because several sources of error are corrected simultaneously. In a~parameter inversion that optimizes the \chem{LNO_x} sources only, the analysis increments will compensate for errors that occur not only in the \chem{LNO_x} sources but also in the other sources and species concentrations. This will result in overcorrections in the \chem{LNO_x} source estimation. Despite this advantage of the simultaneous assimilation of multiple satellite datasets, sources of error will remain because the satellite data do not fully constrain the model. Possible systematic error sources are discussed below.

Cumulus convection plays an important role in determining the vertical profile of chemical concentrations. Because the \chem{NO_x} chemical lifetime in the troposphere depends on altitude (i.e., longer at higher altitude), cumulus convection affects the total amount of \chem{NO_x} in the troposphere and, accordingly, \chem{NO_x} source inversion. However, the cumulus parameterization is highly uncertain and leads to systematic errors in the simulated clouds (e.g., Emori et~al., 2005). A~sensitivity experiment in which the convective mass flux from the cumulus parameterization was increased by 20\,{\%} resulted in a~variation of up to 6\,{\%} in the tropical \chem{LNO_x} sources (Table~~~56~~). The high sensitivity of the \chem{LNO_x} source estimation to model convection is commonly reported by Lin (2012).

The model also shows systematic errors in the location of cumulus clouds (c.f., Fig.~1~~0~~1). An accurate determination of cloud positions is important for analysing the \chem{LNO_x} source patterns. The location of the ITCZ clouds is sensitive to sea surface temperature (SST) in the GCM simulation. A~sensitivity experiment using the SST data set obtained for 1997 (a~strong El Ni\~{n}o year) showed a~variation of up to 4\,{\%} in the global \chem{LNO_x} estimates (Table~~~5~~6). The impact of changing the SST data was different for different regions; e.g., ~~and increased~~ the \chem{LNO_x} sources over the Pacific increased by 14\,{\%} in January. This indicates that the uncertainty in the \chem{LNO_x} estimate due to errors in the simulated clouds is significant.

The satellite measurements used in our analysis contain limited information to constrain the influences of model errors due to fast chemistry (e.g. which determines the $\chem{NO_2}/\chem{NO}$ ratio) and transient transport processes (e.g. due to convection and boundary layer mixing). On the basis of a~comparison with aircraft measurements, Miyazaki et~al.~(2012b) implied that the $\chem{NO_2}/\chem{NO}$ ratio in the free troposphere over Mexico is not realistically represented even when observations of multiple species are assimilated. Accordingly, changes in the chemical scheme are expected to affect the estimated sources. For instance, Stavrakou et~al.~(2013) demonstrated the strong effect of \chem{NO_x} loss uncertainties on top-down \chem{NO_x} source inversion by using several different chemical schemes. In addition, the diurnal variations of lightning activity are determined with the aid of the parameterization in the model, since the assimilated measurements provide constraints mainly around noon. Although the departures (observation minus forecast) reflect model errors accumulated over a~period of several days, because of the long lifetime of the chemical constituents in the upper troposphere, errors in the simulated diurnal variability in the chemical concentrations and in the lightning activity will lead to large uncertainties in the source estimate.

The actual spatiotemporal scale of lightning activity is typically much smaller than the model resolution. A~finer resolution model is required to capture the influences of lightning and atmospheric

transport on the scale of individual clouds and efficiently assimilate fine-scale retrievals. We confirmed that in comparison with the GCM simulation the high-resolution Weather Research and Forecasting model (WRF-ARW) version 3.4.1 (Skamarock et~al., 2008) with 3\,\unit{km} horizontal grid spacing and without any cumulus convection parameterization provides a~better agreement with the satellite observations in terms of the cloud distribution over the western Pacific, by better representing the small-scale cloud and wind structures (figure not shown). Data assimilation with high-resolution models are expected to improve the \chem{LNO_x} source estimate.

Although the aircraft \chem{NO_x} emissions likely have relatively small uncertainties (e.g., Wilkerson et al., 2010), the \chem{LNO_x} source estimates might be influenced by errors in the aircraft emissions especially along the major flight routes in the northern extratropics.

\subsubsection{Data assimilation error modelling}

χ^2 diagnostic tests (e.g. Menard and Chang, 2000) ~~is~~are used to measure the data assimilation statistics. The χ^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 χ^2 ratio becomes 1 if the background error covariance matches the model-observation differences.

The annual mean χ^2 was about 1.2 for OMI \chem{NO_2}, 0.8 for TES \chem{O_3}, 0.7 for MOPITT CO, 2.1 for MLS \chem{O_3}, and 1.6 for MLS \chem{HNO_3}, demonstrating that the overall magnitude of the background error covariances are reasonably modelled in the data assimilation. The too large χ^2 for the MLS data imply an overconfidence in the model, and is largely attributed to the prescribed concentrations for \chem{O_3}, \chem{NO_x}, \chem{HNO_3}, and \chem{N_2O_5} above 20\,\unit{km} altitude in CHASER.

The mean analysis spread, as estimated by transforming the background ensemble in the data assimilation (c.f., Eq. (4)),~~for the annual global source strength~~ is about 0.9\,\unit{Tg\,N\,yr^{-1}} for the annual global source strength. This can be translated into an error estimate for the total yearly source. The \chem{LNO_x} analysis is obtained from information of roughly two weeks of measurements, as demonstrated by the spin-up period of the assimilation (i.e., the spin-up period was required to obtain a converged solution in the analysis). If we assume that individual

two-weekly \chem{LNO_x} estimations are uncorrelated, then the impact of the analysis spread on the uncertainty in the total yearly \chem{LNO_x} can be estimated as $0.9/\sqrt{24}$\,\unit{Tg\,N\,yr^{-1}}, or about 0.2\,\unit{Tg\,N\,yr^{-1}}. This contribution is insignificant in comparison to other error terms discussed in this section.

Although both the analysis mean and spread fields are updated by the data assimilation, the a~priori error assumption may influence the analysis results. In our study, the a~priori error was set to 60\,{\%} for the lightning source and 40\,{\%} for the surface emission. We confirmed that increasing the a~priori error of one of the surface or lightning sources of \chem{NO_x} by 20\,{\%} changes the global estimate of \chem{LNO_x} sources by not more than a~few percent, demonstrating the robustness of the source analysis (Table~~5~~6). On the regional scale, however, the impact is no longer negligible.~~-~~ For instance, over the tropical biomass burning regions and the mid-latitude polluted regions, the change in the a~priori error influences the monthly regional source by up to 7\,{\%}. Therefore, the estimated \chem{LNO_x} sources could have large uncertainties, especially where the surface emissions are large and variable. The a~priori error dependence of the surface and \chem{LNO_x} sources in the analysis is related to the vertical sensitivity of the assimilated measurements. Retrievals with a~strongly varying vertical sensitivity will help to separate the surface and \chem{LNO_x} sources more efficiently. A~sensitivity experiment in which the a~priori global total \chem{LNO_x} source is increased by 15\,{\%} demonstrates that the estimated \chem{LNO_x} source amount is influenced by the a~priori source setting (Table~6); the global a-posteriori \chem{LNO_x} sources are increased by 4 \% in January and 10 \% in July~~to some extent (Table~5)~~. Further constraints from additional measurements or longer assimilation cycles may be required to fully remove the a~priori setting dependence.

The data assimilation employs a~localization technique to avoid sampling errors caused by the limited ensemble size. The horizontal localization scale (L) of 600\,\unit{km} applied in this study was optimised based on sensitivity experiments on the basis of comparisons with independent observations (Miyazaki et~al., 2012a). In the upper troposphere, the lifetimes of NO, \chem{HNO_3}, and \chem{O_3} are generally longer than a~day, and long-range transport of \chem{LNO_x}

can occur. A~larger localization length may be useful when satellite measurements detect aged \chem{LNO_x} air. We confirmed that doubling the localization scale (i.e. to 1200\,\unit{km}) changes the monthly global source by up to 12\,{\%}, but this generally degrades the agreement with the independent observations, especially the ozonesonde observations made in the lower and middle troposphere. Introducing an adaptive localization technique that considers the structures of the chemical lifetime and atmospheric circulation, or increasing the ensemble size may be useful to improve the efficiency with which remote observations are used. The choice of the length of the data assimilation cycle could also influence the data assimilation result associated with distinct diurnal variations in tropospheric chemistry.

\subsubsection{Total systematic error}

The total error on the estimate of the \chem{LNO_x} source will be dominated by systematic errors such as the ones discussed above. The uncertainty caused by the ensemble spread was found to be small compared to the numbers given in Table~~5~~6. From the systematic satellite and model uncertainties listed in Table~~5~~6 we obtain an error estimate of about 1\,\unit{Tg\,N\,yr^{-1}} (0.7 in January, 1.6 in July), if we assume that the individual systematic error sources are uncorrelated. Apart from these sensitivity experiments (summarised in Table~~5~~6) several other sources of error may be introduced. In particular, Stavrakou et~al.~(2013) claimed that the uncertainty in the loss processes is very large for \chem{NO_x} resulting in a~factor of 1.8 difference (3.3--5.9 \,\unit{Tg\,N\,yr^{-1}}) between upper and lower estimates of the \chem{LNO_x} source. This would translate into an error bar of about 1\,\unit{Tg\,N\,yr^{-1}}. If this is added as an independent source of error on top of the ones listed in Table 6, ~~If such an effect is added to the results in Table~5,~~ the total error bar would increase to about 1.4 ($=\sqrt{1^2+1^2}$)\,\unit{Tg\,N\,yr^{-1}}.

\subsection{Implication for the parameterization of \chem{LNO_x}}

\subsubsection{Flash activity}

With respect to the climatology based on satellite observations, the scheme of Price and Rind (1992) overestimated the flash activity over South America but underestimated that over central Africa and over

most of the tropical convergence zones, as reported in previous studies (e.g. Allen and Pickering, 2002; Labrador et~al., 2005; Martin et~al., 2007). Boccippio (2002) also pointed out inconsistencies between the scheme of Price and Rind (1992) and satellite observations over the oceans. The modelling of the flash rate may be improved by using a~more advanced parameterisation. However, Tost et~al.~(2007) concluded that the observed lightning distributions were not even approximately reproduced with any of the lightning parameterisations based on either cloud-top height, updraft velocity, updraft mass flux, or convective precipitation.

Murray et~al.~(2012) demonstrated that applying monthly scaling factors obtained from the LIS/OTD satellite instruments improves the tropical ozone simulation. However, we found that the simulated concentrations are only slightly changed by scaling the global lightning flash count to the climatological observations from the LIS/OTD. The satellite observations cannot be used to adjust the detailed spatial structure of the flash frequency because of the small amount of coverage on a~daily basis. On the other hand, an increase in the annual \chem{LNO_x} amount from 4.7 to 6.3\,\unit{Tg\,N\,yr^{-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 ~~LIS/OTD~~ climatological observations~~y~~ (41.2 \,\unit{flashes\,s^{-1}} v.s. 44 or 46 \,\unit{flashes\,s^{-1}}) and considering about 3 \% lower flash frequency in 2007 compared to the climatology (c.f., Sect. 3.2.1). ~~V~~Various factors, such as for instance the \chem{NO_x} production per flash, are responsible for uncertainty in the \chem{LNO_x} source parameterisation. Meanwhile, the relative positive adjustment of the lightning source is larger over the oceans compared to land (58\,{\%} vs. 30\,{\%}). This finding may indicate that the power 1.73 in the modelling of the height dependence of the lightning activity over ocean is underestimated in the Price and Rind (1992) approach. We note that comparisons against the LIS/OTD observations consistently reveal a larger underestimation in the parameterised global flash rate over the oceans (about 27 \%) than over land (about 5 \%). On the other hand, over the tropical oceans (Pacific, Atlantic, and Indian Oceans), the difference between the observed and the parameterised flash rate is relatively small, as summarised in Table 7. This suggests that errors in the \chem{NO_x} production efficiency rather than those in the flash rate could be responsible for the large increase in the \chem{LNO_x} sources over the tropical oceans. This will be further discussed in Sect. 6.2.2. –

\subsubsection{\chem{NO_x} production per flash and CG/IC ratio}

Even if the flash frequency could be predicted accurately by the model, an uncertainty in the amount of \chem{NO_x} produced per flash would lead to an error in the \chem{LNO_x} source estimates. The annual global \chem{LNO_x} source from our estimates corresponds to a~mean NO production of about 350\,\unit{mol\,flash^{-1}}~~.~~ based on the parameterized flash rate, as summarized in Table 7. Because errors in the parameterized flash rate influence this estimation, we also use the LIS/OTD climatological observations; a global mean NO production of about 310\,\unit{mol\,flash^{-1}} is estimated using the flash observations. ~~This~~Both these values ~~is~~are within the range of most other recent estimates, as summarised in Table 8. For instance, Schumann and Huntrieser (2007) suggested a~best estimate of 250\,\unit{mol\,NO\,flash^{-1}}. Ott et~al.~(2010) reported a~mean value of 500\,\unit{mol\,flash^{-1}} from a~cloud-resolved modelling analysis of mid-latitude and sub-tropical storms. Hudman et~al.~(2007) and Jourdain et~al.~(2010) showed, respectively, that a~continental production rate of 500 or 520\,\unit{mol\,NO\,flash^{-1}} gives reasonable performance in a~chemical simulation over the United States. Boersma et~al.~(2005) used the Global Ozone Monitoring Experiment (GOME) \chem{NO_2} data and estimated a~global \chem{LNO_x} source strength of 1.~~6~~1--6.4\,\unit{Tg\,N\,yr^{-1}}, implying a~production rate of 82--328\,\unit{mol\,NO\,flash^{-1}} based on the LIS/OTD climatology.

The amount of \chem{NO_x} produced per flash may not be constant over the globe. It varies with flash strength, extension, type, branching, and other factors. Huntrieser et~al.~(2008) suggested that tropical thunderstorms are less effective than mid-latitude storms in \chem{LNO_x} production per flash due to lower wind shears and smaller stroke lengths. Our analysis for July consistently reveals a~large production per flash of 430 and 350 \,\unit{mol} of NO in the NH (20--90{\degree}\,N) compared to 360 and 240\,\unit{mol} of NO in the tropics (20{\degree}\,S--20{\degree}\,N) based on the parameterised flash rate and the LIS/OTD observations, respectively. There are also obvious regional differences; e.g., a~large production per flash of about 440 and 570 are estimated for the northern Eurasia continent based on the parameterised flash rate and the LIS/OTD observations, respectively, as summarised in Table 7 and shown in Fig. 14. The detailed spatial structures in the production efficiencies estimated from the analysed \chem{LNO_x} sources and the observed and the parameterised flash rates (Fig. 14) may reflect not only variations in flash

characteristics but also noises and errors in the assimilated and flash measurements (c.f., Section 6.1.2). Note that the local production efficiency estimated using the observed flash rate becomes unrealistically large where the observed flash rate is much smaller than the model flash rate (e.g., over most of the remote oceans).

The NO production efficiencies estimated using the simulated total \chem{LNO_x} sources and the simulated flash rate by the model parameterization (without any assimilation) are about 20 \% lower over land and about 11 \% lower over the oceans, compared with those estimated using the analysed \chem{LNO_x} sources and the flash observations. The obtained results imply general underestimations in the \chem{NO_x} production efficiency simulated by the model, although there are obvious regional differences in the estimates (Table 7). The underestimation could be attributed to errors either in the parameterised IC/CG flash ratio (c.f., Eq. (5)) or in the assumptions on the production efficiency of IC and CG flashes (c.f., Section 3.2.1). Detailed analyses of individual storms with a~high-resolution model are required to provide further insights into the NO production efficiency for individual cases.

A~large uncertainty ~~also~~ remains regarding the ratio of \chem{NO_x} production per flash by IC and CG flashes. Following Price et~al.~(1997), a~lightning NO production of 1100\,\unit{mol\,(CG\,flash)^{-1}} and 110\,\unit{mol\,(IC\,flash)^{-1}} was assumed in the parameterisation. However, it has been suggested that the ratio should be closer to 1 than to 10 (Gallardo and Cooray, 1996; Fehr et~al., 2004; DeCaria et~al., 2005)~~.~~, although a more recent estimate by Koshaz et al (2014) showed the ratio to be closer to 10. ~~We attempted to optimize these parameters from the multi-species data assimilation but could not find any significant differences between the two parameters in the analysis.~~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. The observational constraints still seem insufficient for optimizing such detailed parameters. Further insights may be obtained with observations that are higher in accuracy, density, and vertical resolution.

\subsubsection{The C-shaped vertical profile}

The assumption of a~C-shaped vertical profile as proposed by Pickering et~al.~(1998) implies that a~majority of the \chem{LNO_x} is present

in the upper troposphere, while a~secondary maximum occurs in the boundary layer as a~result of downdrafts. Our data assimilation analysis suggests that the C-shape assumption underestimates the source strength in the middle and upper troposphere over land. Ott et~al.~(2010) reported a~consistent result from analyses of a~cloud-scale chemical transport model. They also suggested that the upper tropospheric maximum in \chem{LNO_x} mass may be located too high because of the C-shape assumption at mid-latitudes. Our analysis also revealed that the peak source height is overestimated by up to about 1\,\unit{km} over land and the tropical oceans. ~~in the tropics.~~ Ott et~al.~(2010) suggested that the simplified treatment of \chem{LNO_x} and wind fields by Pickering et~al.~(1998) will cause errors in the vertical \chem{LNO_x} source profile.

When the observational constraints are insufficient to adjust the vertical profiles, changes in a priori LNOx source profiles (e.g., from the profiles of Pickering et al. (1998) to those of Ott et al. (2010)) or changes in the vertical structure of the covariance matrix will affect the analysed profiles. The robustness of the analysed vertical profile of the \chem{LNO_x} source in the assimilation was evaluated with an assimilation sensitivity experiment that optimized height-independent source scaling factors for each grid point. Compared to the height-independent analysis, the height-dependent analysis (i.e. the standard data assimilation) produces sources larger by 24\,{\%} at 300\,\unit{hPa} in January, and sources smaller by 14\,{\%} at 200\,\unit{hPa} in July in the tropics (20{\degree}\,S--20{\degree}\,N) for grid points with the \chem{LNO_x} source column greater than 5×10^{-15}\,\unit{kg\,m^{-2}\,s^{-1}} (figure not shown). In both seasons, the height-dependent analysis generally produces a~lower peak height for the source in the upper troposphere. The global sources also exhibit systematic differences between the assimilations with height-independent and height-dependent source factors (Table~~3~~4). The two estimations show a~large discrepancy especially when TES \chem{O_3} data are assimilated. In the height-dependent analysis, it was estimated that the TES data assimilation mostly decreases the \chem{LNO_x} sources in the upper troposphere (c.f., Fig.~~8~~9). In contrast, in the height-independent analysis, the \chem{LNO_x} sources are increased throughout the troposphere, because positive increments obtained by the constraints in the middle troposphere mostly dominate the total adjustment. These

results demonstrate the capability of the simultaneous assimilation of multiple datasets to modify the vertical source shape.

\subsection{Validation using forward CTM simulations}

The \chem{O_3} concentrations simulated using the estimated lightning and surface sources in CHASER are used to indirectly validate the performance of the estimated sources, as summarized in Table~~69~. In the validation, the multiplication factors for the \chem{LNO_x} sources and the surface emissions estimated from the assimilation are used as inputs to forward CHASER simulations without adjusting the chemical concentrations by assimilation. This validation demonstrates the importance of correcting the \chem{NO_x} sources for reproducing the \chem{O_3} fields.

The validation is made when lightning is most active; e.g. for July in the NH and for January in the tropics and the SH. The ozonesonde observations from 39 locations were taken from the ~~World Ozone and Ultraviolet Data Center (~~WOUDC~~)~~/~~Southern Hemisphere Additional Ozonesondes (~~SHADOZ~~)~~ database, as in Miyazaki et~al.~(2012a). By using the estimated \chem{LNO_x} sources instead of the sources predicted by the model parameterization, CHASER simulations showed improved agreement with independent global ozonesonde observations. The improved agreement includes 13\,{\%} reductions in the negative bias in the middle/upper troposphere for the NH, 17\,{\%} reductions in the positive bias in the upper troposphere for the tropics, and about 25--50\,{\%} reductions in the positive bias in the middle/upper troposphere for the SH. The CHASER simulation showed further improved agreement with the ozonesonde observations, by using the surface \chem{NO_x} emission data from the multiple data assimilation instead of the emission inventories, together with the estimated \chem{LNO_x} sources. This reduced the ozone bias in the NH and the tropics throughout the troposphere. These results demonstrate the improved consistency of the concentrations and emissions through the multiple datasets assimilation and confirm the quality of the estimated sources as inputs to CTM simulations. We note that the concentration adjustment by the simultaneous data assimilation play an important role in further improving the ozone fields especially in the upper troposphere and the lower stratosphere.

\subsection{Comparisons with previous estimates}

Based on various estimation results, Schumann and Huntrieser (2007) have provided a~best estimate of 5 ± 3\,\unit{Tg\,N\,yr^{-1}} for the annual global \chem{LNO_x} source. Our estimate of 6.3 ± 1.4\,\unit{Tg\,N\,yr^{-1}} is well within the range of the best estimate. This is small compared to recent estimates of the uncertainty in the lightning sources (Schumann and Huntrieser, 2007). More recently, Murray et~al.~(2012) and Stavrakou et~al.~(2013) estimated a~global annual \chem{LNO_x} source of 6 ± 0.5\,\unit{Tg\,N\,yr^{-1}} and 3.3--5.9\,\unit{Tg\,N\,yr^{-1}}, respectively. These estimates are also close to our estimate. The annual global \chem{LNO_x} source from our estimates corresponds to a mean \chem{NO} production of about 310\,\unit{mol\,flash^{-1}} based on the LIS/OTD climatological observations. This value is also within most of the recent estimates (c.f., Table 8).

In spite of the good agreement in the estimates of the annual global source and the \chem{NO} production efficiency, the lightning activity and the \chem{LNO_x} 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 \chem{NO_x} 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.

\conclusions

The global source of lightning-produced \chem{NO_x} (\chem{LNO_x}) is estimated from an assimilation of multiple chemical species based on an ensemble Kalman filter approach. \chem{NO_2}, \chem{O_3}, \chem{HNO_3}, and \chem{CO} measurements obtained from multiple satellite instruments (OMI, MLS, TES, and MOPITT) provide comprehensive constraints on estimates of the global \chem{LNO_x} source. This approach has the potential to reduce the influence of model errors on the \chem{LNO_x} source estimation by simultaneously optimizing various aspects of the chemical system, including the surface emissions of \chem{NO_x} and \chem{CO} as well as the concentrations of 35 chemical species. Errors in these model fields other than the \chem{LNO_x} sources introduce additional model--observation mismatches into the inversion and degrade the \chem{LNO_x} source estimation. In most previous top-down estimates,

only \chem{LNO_x} sources were optimized from \chem{NO_2} measurements. In such cases, the \chem{LNO_x} sources may be overcorrected since analysis increments are introduced to compensate for various sources of model error. Substantial differences in the estimated \chem{LNO_x} sources are obtained between the single-parameter (\chem{LNO_x}) inversion and the combined optimization of sources and concentrations, which emphasizes the ability of the assimilation system presented in this paper to improve the \chem{LNO_x} source estimation.

The assimilation provides substantial adjustments to the \chem{NO_x} sources both at the surface and in the middle--upper troposphere, because of the use of multiple satellite data sets with different vertical sensitivities, see Fig.~1. The relative importance of the individual assimilated datasets varies with height and season, reflecting the measurement sensitivity and its relation to lightning activity. The cloud-covered OMI \chem{NO_2} retrievals provide important constraints on the estimation of the \chem{LNO_x} above, and inside the upper part of clouds, because of the enhanced measurement sensitivity of air masses transported upward by the deep convection. TES and MLS measurements add important constraints on the vertical profiles of the \chem{LNO_x} sources, especially in the upper troposphere. Regional studies of the atmosphere over Africa and the western Pacific demonstrated that the optimization of multiple chemical aspects is a~powerful approach for correcting various processes controlling variations in \chem{O_3} and \chem{NO_2}.

The annual global \chem{LNO_x} source amount and NO production efficiency based on the LIS/OTD observations ~~for 2007~~ are estimated by the assimilation system to be 6.3\,\unit{Tg\,N\,yr^{-1}} and 3~~5~~10\,\unit{mol\,NO\,flash^{-1}} for 2007, respectively, which are within the ranges of recent values from top-down estimations and cloud-resolving simulations. The total error on the mean global \chem{LNO_x} source due to uncertainties in the observation, the model, and the assimilation settings have been studied with a~series of sensitivity experiments and is estimated as 1.4\,\unit{Tg\,N\,yr^{-1}}.

The annual \chem{LNO_x} source columns are increased over most parts of the land by up to about 40\,{\%} compared to the a~priori emissions

predicted using the \chem{LNO_x} parameterization. The analyzsed \chem{LNO_x} sources exhibit obvious regional differences in the tropics, reflecting the regionality of cumulus convection and monsoon circulation. The analysis increments significantly differ between land and the oceans, with annual global source increases of about 56\,{\%} over the oceans and by about 32\,{\%} over land. These increases are largely attributed to the positive increments in the lower troposphere over the oceans and in the upper troposphere over land. We find that the relative positive adjustment of the lightning source is significantly larger over the oceans compared to land. This finding may indicate that the power 1.73 in the modelling of the height dependence of the source over ocean is underestimated. The significantly improved agreement with independent ozone observations from ozonesondes and TOC retrievals gives confidence in the performance of the data assimilation.

The analysed \chem{LNO_x} sources have important implications for improving \chem{LNO_x} parameterisations. First, errors in flash rates can explain only a~small fraction of the uncertainty in \chem{LNO_x} estimates, as the main observed features of the annual global flash rate are generally reproduced by the model, except for the large low bias over central Africa.. The remaining uncertainty in estimates from the bottom-up approach arises from the NO production efficiency that can be very different for individual storms. Our analysis suggests that tropical thunderstorms are less effective than mid-latitude storms in generating NO with each flash in boreal summer, as commonly suggested by previous studies. It is also suggested that the model parameterisation may underestimate the annual and global mean NO production efficiency by about 10 \% over land and 20 \% over the oceans.

Second, the widely used C-shape assumption underestimates the source strength in the upper troposphere and overestimates the peak source height over land and the tropical oceans, especially along the ITCZ. Finally, as the two types of discharges (IC and CG) behave differently, these should be considered separately. Although parameters related to these different types are hardly discriminated with the currently available observations, the approach of combining all available satellite datasets is expected to provide further insights into such detailed processes in future studies with measurements that are more advanced (i.e. higher in accuracy, density, and vertical resolution).

~~\begin{acknowledgements}~~
~~We would like to thank Masayuki Takigawa, Hisashi Yashiro, Kevin Hamilton, Yuqing Wang, and Folkert Boersma for their helpful comments on this study. This work was supported by the JSPS Grant-in-Aid for Yong Scientists (B) 24740327.~~
~~\end{acknowledgements}~~

\begin{acknowledgements}
We would like to thank Masayuki Takigawa, Hisashi Yashiro, Kevin Hamilton, Yuqing Wang, and Folkert Boersma for their helpful comments on this study. We also would like to thank the three anonymous reviewers and the editor for their valuable comments. This work was supported by the JSPS Grant-in-Aid for Yong Scientists (B) 24740327.
\end{acknowledgements}

\begin{thebibliography}{~~40~~69}

\bibitem{1}
Abarca, S. F., K. L. Corbosiero, and T. J. Galarneau Jr.: An evaluation of the Worldwide Lightning Location Network (WWLLN) using the National Lightning Detection Network (NLDN) as ground truth, J. Geophys. Res., 115, D18206, doi:10.1029/2009JD013411., 2010.

\bibitem{1}
Allen,~D.~J. and Pickering,~K.~E.: Evaluation of lightning flash rate parameterizations for use in a~global chemical transport model,~J. Geophys. Res., 107, 4711, \doi{10.1029/2002JD002066}, 2002.

\bibitem{2}
Arakawa,~A. and Schubert,~W.~H.: The interaction of a~cumulus cloud ensemble with the large-scale environment,~J. Atmos. Sci., 31, 674--701, 1974.

\bibitem{3}
Beirle,~S., Spichtinger,~N., Stohl,~A., Cummins,~K.~L., Turner,~T., Boccippio,~D., Cooper,~O.~R., Wenig,~M., Grzegorski,~M., Platt,~U., and Wagner,~T.: Estimating the \chem{NO_x} produced by lightning from GOME

and NLDN data: a case study in the Gulf of Mexico,
Atmos. Chem. Phys., 6, 1075--1089, \doi{10.5194/acp-6-1075-2006},
2006. % ### SELF-REFERENCE ###

\bibitem{4}
Boccippio,~D.~J.: Lightning scaling relations revisited,~J. Atmos. Sci., 59, 1086--1104, 2002.

\bibitem{5}
Boersma,~K.~F., Eskes,~H.~J., Meijer,~E.~W., and Kelder,~H.~M.:
Estimates of lightning \chem{NO_x} production from GOME satellite
observations, Atmos. Chem. Phys., 5, 2311--2331,
\doi{10.5194/acp-5-2311-2005}, 2005. % ### SELF-REFERENCE ###

\bibitem{6}
Boersma,~K.~F., Eskes,~H.~J., Dirksen,~R.~J., van~der~A,~R.~J.,
Veefkind,~J.~P., Stammes,~P., Huijnen,~V., Kleipool,~Q.~L.,
Sneep,~M., Claas,~J., Leit\~{a}o,~J., Richter,~A., Zhou,~Y., and
Brunner,~D.: An improved tropospheric \chem{NO_2} column retrieval
algorithm for the Ozone Monitoring Instrument, Atmos. Meas. Tech., 4,
1905--1928, \doi{10.5194/amt-4-1905-2011}, 2011. % ### SELF-REFERENCE
% ###

\bibitem{7}
Bowman,~K.~W., Rodgers,~C.~D., Sund-Kulawik,~S., Worden,~J.,
Sarkissian,~E., Osterman,~G., Steck,~T., Luo,~M., Eldering,~A.,
Shephard,~M.~W., Worden,~H., Clough,~S.~A., Brown,~P.~D.,
Rinsland,~C.~P., Lampel,~M., Gunson,~M., and Beer,~R.: Tropospheric
emission spectrometer: retrieval method and error analysis,
IEEE~T. Geosci. Remote, 44, 1297--1307, 2006.

\bibitem{8}
Bowman,~K.~W., Rodgers,~C.~D., Sund-Kulawik,~S., Worden,~J.,
Sarkissian,~E., Osterman,~G., Steck,~T., Luo,~M., Eldering,~A.,
Shephard,~M.~W., Worden,~H., Clough,~S.~A., Brown,~P.~D.,
Rinsland,~C.~P., Lampel,~M., Gunson,~M., and Beer,~R.: Tropospheric
emission spectrometer: retrieval method and error analysis,
IEEE~T. Geosci. Remote, 44, 1297--1307, 2006.

\bibitem{9}

Cecil, D. J., Buechler, D. E., Blakeslee, R. J., Gridded lightning climatology from TRMM-LIS and OTD: Dataset description, Atmospheric Research, 135--136, 404-414, http://dx.doi.org/10.1016/j.atmosres.2012.06.028, 2014.

~~\bibitem{8}~~
~~Cooper,~O.~R., Stohl,~A., Trainer,~M., Thompson, A. M., Witte, J. C., Oltmans, S., Morris, G. A., Pickering, K. E., Crawford, J., Chen, G., Cohen, R. C., Bertram, T. H., Wooldridge, P. J., Perring, A., Brune, W. H., Merrill, J., Moody, J., Tarasick, D., N\'ed\'elec, P., Forbes, G., Newchurch, M., Schmidlin, F., Johnson, B. J., Turquety, S., Baughcum, S. L., Ren, X., Fehsenfeld, F., Meagher, J. F., Spichtinger, N., Brown, C., McKeen, S. A., McDermid, I., and Leblanc, T.: Large upper tropospheric ozone enhancements above mid-latitude North America during summer: in situ evidence from the IONS and MOZAIC ozone measurement network,~J. Geophys. Res., 111, D24S05, \doi{10.1029/2006JD007306}, 2006.~~

\bibitem{~~9~~10}
Christian,~H.~J., Blakeslee,~R.~J., Boccippio,~D.~J., Boeck,~W.~L., Buechler,~D.~E., Driscoll,~K.~T., Goodman,~S.~J., Hall,~J.~M., Koshak,~W.~J., Mach,~D.~M., and Stewart,~M.~F.: Global frequency and distribution of lightning as observed from space by the Optical Transient Detector,~J. Geophys. Res., 108, 4005, \doi{10.1029/2002JD002347}, 2003.

\bibitem{11~~0~~}
DeCaria,~A.~J., Pickering,~K.~E., Stenchikov,~G.~L., Scala,~J.~R., Stith,~J.~L., Dye,~J.~E., Ridley,~B.~A., and Laroche,~P.: A~cloud-scale model study of lightning-generated \chem{NO_x} in an individual thunderstorm during STERAO-A,~J. Geophys. Res., 105, 11601--11616, 2000.

\bibitem{11~~1~~2}
DeCaria,~A.~J., Pickering,~K.~E., Stenchikov,~G.~L., and Ott,~L.~E.: Lightning-generated \chem{NO_x} and its impact on tropospheric ozone production: a~three-dimensional modeling study of a~Stratosphere-Troposphere Experiment: Radiation, Aerosols and Ozone (STERAO-A) thunderstorm,~J. Geophys. Res., 110, 1--13, \doi{10.1029/2004JD005556}, 2005.

\bibitem{13~~2~~}
Deeter,~M.~N., Martinez-Alonso,~S., Edwards,~D.~P., Emmons,~L.~K., Gille,~J.~C., Worden,~H.~M., Pittman,~J.~V., Daube,~B.~C., and

Wofsy,~S.~C.: Validation of MOPITT Version 5 thermal-infrared,
near-infrared, and multispectral carbon monoxide profile retrievals
for 2000--2011,~J. Geophys. Res. Atmos., 118, 6710--6725,
\doi{10.1002/jgrd.50272}, 2003.

\bibitem{143}
Deeter,~M.~N., Worden,~H.~M., Edwards,~D.~P., Gille,~J.~C., Mao,~D., and
Drummond,~J.~R., MOPITT multispectral CO retrievals: origins and effects of
geophysical radiance errors,~J. Geophys. Res., 116, D15303,
\doi{10.1029/2011JD015703}, 2011.

\bibitem{154}
Deeter,~M.~N., Martinez-Alonso,~S., Edwards,~D.~P., Emmons,~L.~K.,
Gille,~J.~C., Worden,~H.~M., Pittman,~J.~V., Daube,~B.~C., and
Wofsy,~S.~C.: Validation of MOPITT Version 5 thermal-infrared,
near-infrared, and multispectral carbon monoxide profile retrievals
for 2000--2011,~J. Geophys. Res. Atmos., 118, 6710--6725,
\doi{10.1002/jgrd.50272}, 2013.

\bibitem{165}
Emori,~S., Hasegawa,~A., Suzuki,~T., and Dairaku,~K.: Validation,
parameterization dependence, and future projection of daily
precipitation simulated with a~high-resolution atmospheric GCM,
Geophys. Res. Lett., 32, L06708, \doi{10.1029/2004GL022306}, 2005.

\bibitem{17}
Eskes, H. J. and Boersma, K. F.: Averaging kernels for DOAS total-column satellite retrievals,
Atmos. Chem. Phys., 3, 1285-1291, doi:10.5194/acp-3-1285-2003, 2003.

\bibitem{186}
Fehr,~T., H{\"o}ller,~H., and Huntrieser,~H.: Model study on
production and transport of lightning-produced \chem{NO_x} in a~EULINOX
supercell storm,~J. Geophys. Res., 109, 1--17,
\doi{10.1029/2003JD003935}, 2004.

\bibitem{197}
Gallardo,~L. and Cooray,~V.: Could cloud-to-cloud discharges be as
effective as cloud-to-ground discharges in producing \chem{NO_x}?, Tellus B,
48, 641--651, 1996.

\bibitem{20}
Galloway, J. N., F. J. Dentener, D. G. Capone, E. W. Boyer, R. W. Howarth, S. P. Seitzinger, G. P. Asner, C. Cleveland, P. Green, E. Holland, D. M. Karl, A. F. Michaels, J. H. Porter, A. Townsend, and C. V\"{o}r\"{o}smarty. 2004. Nitrogen Cycles: Past, Present and Future. Biogeochemistry 70: 153-226.

\bibitem{2118}
Gaspari,~G. and Cohn,~S.~E.: Construction of correlation functions in two and three dimensions, Q.~J. Roy. Meteor. Soc., 125, 723--757, 1999.

\bibitem{2219}
Graedel,~T.~E., Bates,~T.~S., Bouwman,~A.~F., Cunnold,~D., Dignon,~J., Fung,~I., Jacob,~D.~J., Lamb,~B.~K., Logan,~J.~A., Marland,~G., Middleton,~P., Pacyna,~J.~M., Placet,~M., and Veldt,~C.: A~compilation of inventories of emissions to the atmosphere, Global Biogeochem. Cy., 7, 1--26, 1993.

\bibitem{230}
Hudman,~R.~C., Jacob,~D.~J., Turquety,~S., Leibensperger, E. M., Murray, L. T., Wu, S., Gilliland, A. B., Avery, M., Bertram, T. H., Brune, W., Cohen, R. C., Dibb, J. E., Flocke, F. M., Fried, A., Holloway, J., Neuman, J. A., Orville, R., Perring, A., Ren, X., Ryerson, T. B., Sachse, G. W., Singh, H. B., Swanson, A., and Wooldridge, P. J.: Surface and lightning sources of nitrogen oxides over the United States: magnitudes, chemical evolution, and outflow,~J. Geophys. Res., 112, D12S05, \doi{10.1029/2006JD007912}, 2007.

\bibitem{241}
Hunt,~B.~R., Kostelich,~E.~J., and Szunyogh,~I.: Efficient data assimilation for spatiotemporal chaos: a~local ensemble transform Kalman filter, Physica D, 230, 112--126, 2007.

\bibitem{252}
Huntrieser,~H., Schumann,~U., Schlager,~H., H\"{o}ller,~H., Giez,~A., Betz,~H.-D., Brunner,~D., Forster,~C., Pinto~Jr.,~O., and Calheiros,~R.: Lightning activity in Brazilian thunderstorms during TROCCINOX: implications for \chem{NO_x} production, Atmos. Chem. Phys., 8, 921--953, \doi{10.5194/acp-8-921-2008}, 2008. % ### SELF-REFERENCE % ###

\bibitem{263}

Jenkins,~G.~S. and Ryu,~J.-H.: Space-borne observations link the tropical atlantic ozone maximum and paradox to lightning, Atmos. Chem. Phys., 4, 361--375, \doi{10.5194/acp-4-361-2004}, 2004. % ### SELF-REFERENCE ###

\bibitem{275}
Jourdain,~L., Worden,~H.~M., Worden,~J.~R., Bowman,~K., Li,~Q.~B., Eldering,~A., Kulawik,~S.~S., Osterman,~G., Boersma,~F., Fisher,~B., Rinsland,~C.~P., Beer,~R., and Gunson,~M.: Tropospheric vertical distribution of tropical Atlantic ozone observed by TES during the Northern African biomass burning season, Geophys. Res. Lett., 34, L04810, \doi{10.1029/2006GL028284}, 2007.

\bibitem{284}
Jourdain,~L., Kulawik,~S.~S., Worden,~H.~M., Pickering,~K.~E., Worden,~J., and Thompson,~A.~M.: Lightning \chem{NO_x} emissions over the USA constrained by TES ozone observations and the GEOS-Chem model, Atmos. Chem. Phys., 10, 107--119, \doi{10.5194/acp-10-107-2010}, 2010. % ### SELF-REFERENCE ###

\bibitem{296}
Kanamitsu,~M., Ebisuzaki,~W., Woollen,~J., Yang,~S.~K., Hnilo,~J.~J., Fiorino,~M., and Potter,~G.~L.: NCEP-DOE AMIP-II reanalysis (R-2),~B. Am. Meteorol. Soc., 83, 1631--1643, \doi{10.1175/BAMS-83-11-1631}, 2002.

\bibitem{30}
Kanamitsu,~M., Ebisuzaki,~W., Woollen,~J., Yang,~S.~K., Hnilo,~J.~J., Fiorino,~M., and Potter,~G.~L.: NCEP-DOE AMIP-II reanalysis (R-2),~B. Am. Meteorol. Soc., 83, 1631--1643, \doi{10.1175/BAMS-83-11-1631}, 2002.

\bibitem{31}
Koshak, W., Peterson, H., Biazar, A., Khan, M., and Wang, L.: The NASA Lightning Nitrogen Oxides Model (LNOM): Application to air quality modeling, Atmospheric Research, Volumes 135--136, 363-369, http://dx.doi.org/10.1016/j.atmosres.2012.12.015, 2014.

\bibitem{3227}
Labrador,~L.~J., von~Kuhlmann,~R., and Lawrence,~M.~G.: The effects of lightning-produced \chem{NO_x} and its vertical distribution on atmospheric chemistry: sensitivity simulations with MATCH-MPIC, Atmos. Chem. Phys., 5, 1815--1834, \doi{10.5194/acp-5-1815-2005},

2005. % ### SELF-REFERENCE ###

\bibitem{33}
Lamsal, L. N., Martin, R. V., van Donkelaar, A., Celarier, E. a., Buc- sela, E. J., Boersma, K. F., Dirksen, C. L., and Wang, Y.: Indirect validation of tropospheric nitrogen dioxide retrieved from the OMI satellite instrument: insight into the seasonal variation of nitrogen oxides at northern midlatitudes, J. Geophys. Res., 115, 1–15, doi:10.1029/2009JD013351, 2010.

\bibitem{3428}
Lay,~E.~H., Jacobson,~A.~R., Holzworth,~R.~H., Rodger,~C.~J., and Dowden,~R.~L.: Local time variation in land/ocean lightning flash density as measured by the World Wide Lightning Location Network,~J. Geophys. Res., 112, D13111, \doi{10.1029/2006JD007944}, 2007.

\bibitem{3529}
Lin,~J.-T.: Satellite constraint for emissions of nitrogen oxides from anthropogenic, lightning and soil sources over East China on a high-resolution grid, Atmos. Chem. Phys., 12, 2881--2898, \doi{10.5194/acp-12-2881-2012}, 2012. % ### SELF-REFERENCE ###

\bibitem{360}
Livesey,~N.~J., Read,~W.~G., Froidevaux,~L., Lambert,~A., Manney,~G.~L., Pumphrey,~H.~C., Santee,~M.~L., Schwartz,~M.~J., Wang,~S., Cofield,~R.~E., Cuddy,~D.~T., Fuller,~R.~A., Jarnot,~R.~F., Jiang,~J.~H., Knosp,~B.~W., Stek,~P.~C., Wagner,~P.~A., and Wu,~D.~L.: Aura Microwave Limb Sounder (MLS), Version 3.3 Level 2 data quality and description document, Tech. Rep. JPL D-33509, Jet Propul. Lab., Pasadena, CA, 2011.

\bibitem{371}
Martin,~R.~V., Jacob,~D.~J., Logan,~J.~A., Ziemke,~J.~M., and Washington,~R.: Detection of a~lightning influence on tropical tropospheric ozone, Geophys. Res. Lett., 27, 1639--1642, 2000.

\bibitem{382}
Martin,~R.~V., Sauvage,~B., Folkins,~I., Sioris,~C.~E., Boone,~C., Bernath,~P., and Ziemke,~J.~R.: Space-based constraints on the production of nitric oxide by lightning,~J. Geophys. Res., 112, D09309, \doi{10.1029/2006JD007831}, 2007.

\bibitem{393}
M\"{e}nard,~R. and Chang,~L.-P.: Assimilation of stratospheric
chemical tracer observations using a~Kalman filter, Part 2:
2-validated results and analysis of variance and correlation dynamics,
Mon. Weather Rev., 128, 2672--2686, 2000.

\bibitem{3406}
Miyazaki,~K. and Eskes,~H.: Constraints on surface \chem{NO_x} emissions by
assimilating satellite observations of multiple species, Geophys. Res. Lett.,
40, 4745--4750, \doi{10.1002/grl.50894}, 2013.

\bibitem{4134}
Miyazaki,~K., Eskes,~H.~J., Sudo,~K., Takigawa,~M., van~Weele,~M., and
Boersma,~K.~F.: Simultaneous assimilation of satellite \chem{NO_2}, \chem{O_3},
CO, and \chem{HNO_3} data for the analysis of tropospheric chemical
composition and emissions, Atmos. Chem. Phys., 12, 9545--9579,
\doi{10.5194/acp-12-9545-2012}, 2012a. % ### SELF-REFERENCE ###

\bibitem{4235}
Miyazaki,~K., Eskes,~H.~J., and Sudo,~K.: Global \chem{NO_x} emission
estimates derived from an assimilation of OMI tropospheric \chem{NO_2}
columns, Atmos. Chem. Phys., 12, 2263--2288,
\doi{10.5194/acp-12-2263-2012}, 2012b. % ### SELF-REFERENCE ###

\bibitem{4337}
Murray,~L.~T., Jacob,~D.~J., Logan,~J.~A., Hudman,~R.~C., and
Koshak,~W.~J.: Optimized regional and interannual variability of
lightning in a~global chemical transport model constrained by LIS/OTD
satellite data,~J. Geophys. Res., 117, D20307,
\doi{10.1029/2012JD017934}, 2012.

\bibitem{4438}
Nassar,~R., Logan,~J.~A., Worden,~H.~M., Megretskaia,~I.~A.,
Bowman,~K.~W., Osterman,~G.~B., Thompson,~A.~M., Tara- sick,~D.~W.,
Austin,~S., Claude,~H., Dubey,~M.~K., Hocking,~W.~K., Johnson,~B.~J.,
Joseph,~E., Merrill,~J., Morris,~G.~A., Newchurch,~M., Oltmans,~S.~J.,
Posny,~F., Schmidlin,~F.~J., Vomel,~H., Whiteman,~D.~N., and
Witte,~J.~C.: Validation of Tropospheric Emission Spectrometer (TES)
nadir ozone profiles using ozonesonde measurements,~J. Geophys. Res.,

113, D15S17, \doi{10.1029/2007JD008819}, 2008.

\bibitem{45}
Ott, L. E., K. E. Pickering, G. L. Stenchikov, H. Huntrieser, and U. Schumann: Effects of lightning NOx production during the 21 July European Lightning Nitrogen Oxides Project storm studied with a three-dimensional cloud-scale chemical transport model, J. Geophys. Res., 112, D05307, doi:10.1029/2006JD007365, 2007.

\bibitem{4639}
Ott,~L.~E., Pickering,~K.~E., Stenchikov,~G.~L., Allen,~D.~J., DeCaria,~A.~J., Ridley,~B., Lin,~R.-F., Lang,~S., and Tao,~W.-K.: Production of lightning \chem{NO_x} and its vertical distribution calculated from three-dimensional cloud-scale chemical transport model simulations,~J. Geophys. Res., 115, D04301, \doi{10.1029/2009JD011880}, 2010.

\bibitem{470}
Pan,~D.-M., and Randall,~D.~A.: A~cumulus parameterization with a~prognostic closure, Q.~J. Roy. Meteor. Soc., 124, 949--981, 1998.

\bibitem{48}
Peterson, H. S. and Beasley, W. H.: Possible catalytic effects of ice particles on the production of NOx by lightning discharges, Atmos. Chem. Phys., 11, 10259-10268, doi:10.5194/acp-11-10259-2011, 2011.

\bibitem{491}
Pickering,~K.~E., Wang,~Y., Tao,~W.~K., Price,~C., and Muller,~J.~F.: Vertical distributions of lightning \chem{NO_x} for use in regional and global chemical transport models,~J. Geophys. Res., 103, 31203--31216, \doi{10.1029/98JD02651}, 1998.

\bibitem{5042}
Price,~C. and Rind,~D.: A~simple lightning parameterization for calculating global lightning distributions,~J. Geophys. Res., 97, 9919--9933, \doi{10.1029/92JD00719}, 1992.

\bibitem{51}
Price,~C. and Rind,~D.: What determines the cloud-to-ground lightning! fraction in thunderstorms?, Geophys. Res. Lett., 20, 463--466, doi:10.1029/93GL00226, 1993.

\bibitem{4523}
Price,~C., Penner,~J., and Prather,~M.: \chem{NO_x} from lightning, 1. Global distribution based on lightning physics,~J. Geophys. Res., 102, 5929--5941, \doi{10.1029/96JD03504}, 1997.

\bibitem{4535}
Ridley,~B.~A., Dye,~J.~E., Walega,~J.~G., Zheng,~J., Grahek,~F.~E., and Rison,~W.: On the production of active nitrogen by thunderstorms over New Mexico,~J. Geophys. Res., 101, 20985--21005, \doi{10.1029/96JD01706}, 1996.

\bibitem{4754}
Schumann,~U. and Huntrieser,~H.: The global lightning-induced nitrogen oxides source, Atmos. Chem. Phys., 7, 3823--3907, \doi{10.5194/acp-7-3823-2007}, 2007. % ### SELF-REFERENCE ###

\bibitem{4855}
Skamarock,~W.~C., Klemp,~J.~B., Dudhia,~J., Gill,~D.~O., Barker,~D.~M., Wang,~W., and Powers,~J.~G.: A~description of the Advanced Research WRF Version 3, NCAR Tech Notes-475+STR, 113 pp., 2008.

\bibitem{56}
Sauvage, B., Martin, R. V., {van Donkelaar}, A., Liu, X., Chance, K., Jaegl\'{u}, L., Palmer, P. I., Wu, S., and Fu, T.-M.: Remote sensed and in situ constraints on processes affecting tropical tropospheric ozone, Atmos. Chem. Phys., 7, 815-838, doi:10.5194/acp-7-815-2007, 2007a.

\bibitem{57}
Sauvage, B., R. V. Martin, A. {van Donkelaar}, and J. R. Ziemke: Quantification of the factors controlling tropical tropospheric ozone and the South Atlantic maximum, J. Geophys. Res., 112, D11309, doi:10.1029/2006JD008008, 2007b.

\bibitem{5846}
Stavrakou,~T., M\"{u}ller,~J.-F., Boersma,~K.~F., van~der~A,~R.~J., Kurokawa,~J., Ohara,~T., and Zhang,~Q.: Key chemical \chem{NO_x} sink uncertainties and how they influence top-down emissions of nitrogen oxides, Atmos. Chem. Phys., 13, 9057--9082, \doi{10.5194/acp-13-9057-2013}, 2013. % ### SELF-REFERENCE ###

\bibitem{59}

Stroppiana, D., Brivio, P. A., Gr\'{e}goire, J.-M., Liousse, C., Guillaume, B., Granier, C., Mieville, A., Chin, M., and P\'{e}tron, G.: Comparison of global inventories of CO emissions from biomass burning derived from remotely sensed data, Atmos. Chem. Phys., 10, 12173--12189, doi:10.5194/acp-10-12173-2010, 2010.

\bibitem{6049}
Sudo,~K., Takahashi,~M., and Akimoto,~H.: CHASER: a~global chemical model of the troposphere, 2. Model results and evaluation,~J. Geophys. Res., 107, 4586, \doi{10.1029/2001JD001114}, 2002.

\bibitem{5610}
Thompson,~A.~M., Pickering,~K.~E., Dickerson,~R.~R., Ellis Jr.,~W.~G., Jacob,~D.~J., Scala,~J.~R., Tao,~W.-K., McNamara,~D.~P., and Simpson,~J.: Convective transport over the central United States and its role in regional CO and ozone budgets,~J. Geophys. Res., 99, 18703--18711, \doi{10.1029/94JD01244}, 1994.

\bibitem{5621}
Thompson,~A.~M., Witte,~J.~C., Smit,~H.~G.~J., Oltmans,~S.~J., Johnson,~B.~J., Kirchhoff,~V.~W.~J.~H., and Schmidlin,~F.~J.: Southern Hemisphere Additional Ozonesondes (SHADOZ) 1998--2004 tropical ozone climatology: 3. Instrumentation, station-to-station variability, and evaluation with simulated flight profiles,~J. Geophys. Res., 112, D03304, \doi{10.1029/2005JD007042}, 2007.

\bibitem{6352}
Tost,~H., J\"{o}ckel,~P., and Lelieveld,~J.: Lightning and convection parameterisations -- uncertainties in global modelling, Atmos. Chem. Phys., 7, 4553--4568, \doi{10.5194/acp-7-4553-2007}, 2007. % ### SELF-REFERENCE ###

\bibitem{5364}
van~der~Werf,~G.~R., Randerson,~J.~T., Giglio,~L., Collatz,~G.~J., Mu,~M., Kasibhatla,~P.~S., Morton,~D.~C., DeFries,~R.~S., Jin,~Y., and van~Leeuwen,~T.~T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997--2009), Atmos. Chem. Phys., 10, 11707--11735, \doi{10.5194/acp-10-11707-2010}, 2010. % ### SELF-REFERENCE ###

\bibitem{65}
Whitaker, J. and Hamill, T.: Ensemble data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913--1924, 2002.

\bibitem{66}
Wilkerson, J. T., Jacobson, M. Z., Malwitz, A., Balasubramanian, S., Wayson, R., Fleming, G., Naiman, A. D., and Lele, S. K.: Analysis of emission data from global commercial aviation: 2004 and 2006, Atmos. Chem. Phys., 10, 6391--6408, doi:10.5194/acp-10-6391-2010, 2010.

\bibitem{~~5~~467}
Worden,~J., Kulawik,~S.~S., Shephard,~M.~W., Clough,~S.~A., Worden,~H., Bowman,~K., and Goldman,~A.: Predicted errors of tropospheric emission spectrometer nadir retrievals from spectral window selection,~J. Geophys. Res., 109, D09308, \doi{10.1029/2004JD004522}, 2004.

\bibitem{~~5~~568}
Worden,~J., Liu,~X., Bowman,~K., Chance,~K., Beer,~R., Eldering,~A., Gunson,~M., and Worden,~H.: Improved tropospheric ozone profile retrievals using OMI and TES radiances,~J. Geophys. Res., 34, L01809, \doi{10.1029/2006GL027806}, 2007.

\bibitem{~~5~~669}
Worden,~J., Jones,~D.~B.~A., Liu,~J., Parrington,~M., Bowman,~K., Stajner,~I., Beer,~R., Jiang,~J., Thouret,~V., Kulawik,~S., Li,~J.~L.~F., Verma,~S., and Worden,~H.: Observed vertical distribution of tropospheric ozone during the Asian summertime monsoon,~J. Geophys. Res., 114, D13304, \doi{10.1029/2008JD010560}, 2009.

\bibitem{~~5~~770}
Ziemke,~J.~R., Chandra,~S., Duncan,~B.~N., Froidevaux,~L., Bhartia,~P.~K., Levelt,~P.~F., and Waters,~J.~W.: Tropospheric ozone determined from Aura OMI and MLS: evaluation of measurements and comparison with the Global Modeling Initiative's Chemical Transport Model,~J. Geophys. Res., 111, D19303, \doi{10.1029/2006JD007089}, 2006.

~~\end{thebibliography}~~

```
\end{thebibliography}

\clearpage

\begin{table*}[t]
  \caption
{The annual and seasonal total flash rate (in \unit{flashes\,s^{-1}}) estimated from the LIS/OTD high--resolution monthly climatology (HRMC) data (Cecil et al., 2014) and the model simulation for the Northern Hemisphere (NH, 20--90{\degree}\,N), the tropics (TR, 20{\degree}\,S--20{\degree}\,N), the Southern Hemisphere (SH, 90--20{\degree}\,S), and the globe (GL, 90{\degree}\,S--90{\degree}\,N) for 2007 and for four seasons of the year: December--February (DJF), March--May (MAM), June--August (JJA), and September--November (SON).}
%\scalebox{.85}[.85]
{\begin{tabular}{lccccccccccc}
\tophline
&\multicolumn{5}{c}{ LIS/OTD} & &\multicolumn{5}{c}{Model}\\
\cline{2-6}\cline{8-12}
& Annual & DJF & MAM & JJA & SON & & Annual & DJF & MAM & JJA & SON    \\
\middlehline
NH & 12.7 & 2.5 & 11.6 & 28.4 & 8.0 & &12.7 & 6.6 & 12.6 & 22.2 & 9.5 \\
TR & 27.1 & 22.9 & 27.4 & 25.7 & 32.6 & & 21.3 & 23.5 & 22.1 & 18.9 & 20.7 \\
SH & 6.8 & 11.0 & 6.1 & 3.1 & 7.0 & & 7.2 & 8.6 & 7.1 & 5.5 & 7.5 \\
GL & 46.5 & 36.3 & 45.0 & 57.1 & 47.7 & & 41.2 & 38.7 & 41.8 & 46.7 & 37.6  \\
\bottomhline
\end{tabular}}
%\hack{
%\setlength\tabularwidth{0.9\tabularwidth}
%}
%\scalebox{.7}[.7]{
\belowtable{%
%\hack{\vspace*{2mm}}
}
%}
\end{table*}



\begin{table*}[t]
  \caption{The annual total \chem{LNO_x} sources (in
    \unit{Tg\,N\,yr^{-1}}) obtained from the CTM simulation and the
```

```
   data assimilation for the Northern Hemisphere (NH,
   20--90{\degree}\,N), the tropics (TR,
   20{\degree}\,S--20{\degree}\,N), the Southern Hemisphere (SH,
   90--20{\degree}\,S), and the globe (GL,
   90{\degree}\,S--90{\degree}\,N). The analysis uncertainty
   estimated from the mean analysis spread is shown in brackets.}
%\scalebox{.85}[.85]
{\begin{tabular}{lcccc}
\tophline
&NH &TR &SH &GL   \\
\middlehline
CTM &1.4 &2.7 &0.6 &4.7  \\
Assimilation &2.0 ($\pm$0.3) &3.5 ($\pm$0.49) &0.8 ($\pm$0.10) &6.3 ($\pm$0.9) \\
\bottomhline
\end{tabular}}
%\hack{
%\setlength\tabularwidth{0.9\tabularwidth}
%}
%\scalebox{.7}[.7]{
\belowtable{%
%\hack{\vspace*{2mm}}
}
%}
\end{table*}


\begin{table*}[t]
\caption{
The regional averages of the mean altitude (in km) with maximum annual
\chem{LNO_x} emission (i.e., source peak height) estimated from the
CTM simulation and the data assimilation and ~~. The~~the corresponding analysis increments
(the
data assimilation minus the simulation)~~ are shown in brackets.~~. The
definitions of the regions are same as for Fig.~~6~~7, except for the
tropical western Pacific (130--165{\degree}\,E,
1--18{\degree}\,N). Grid points without any \chem{LNO_x} sources or
with peak height for pressures higher than 850\,\unit{hPa} are removed
from the average, in order to measure the upper tropospheric peak
height.}
```

```
%\scalebox{.85}[.85]
{\begin{tabular}{lccc}
\tophline
&CTM &Assim &Increment    \\
\middlehline
North America          & 7.82  & 7.85  & +0.03   \\
Europe                 & 5.72  & 5.58  & $-$0.14 \\
Northern Eurasia       & 9.39  & 9.39  & $\pm$0  \\
Tropical Western Pacific & 11.15 & 9.96  & $-$1.19 \\
South America          & 9.97  & 9.83  & $-$0.14 \\
Northern Africa        & 11.00 & 10.83 & $-$0.17 \\
Southern Africa        & 9.36  & 8.98  & $-$0.38 \\
Southeast Asia         & 11.51 & 10.77 & $-$0.74 \\
Australia              & 8.01  & 7.60  & $-$0.41 \\
\bottomhline
\end{tabular}}
%\hack{
%\setlength\tabularwidth{0.9\tabularwidth}
%}
%\scalebox{.7}[.7]{
\belowtable{%
%\hack{\vspace*{2mm}}
}
%}
\end{table*}

%\clearpage

\begin{table*}[t]
\caption{The monthly \chem{LNO_x} sources (in Tg\,N) for the Northern Hemisphere
 (NH, 20--90{\degree}\,N), the tropics (TR,
 20{\degree}\,S--20{\degree}\,N), the Southern Hemisphere (SH,
 90--20{\degree}\,S), and the globe (GL,
 90{\degree}\,S---90{\degree}\,N) as obtained from the CTM simulation,
 the OSEs with TES \chem{O_3}, OMI \chem{NO_2}, MLS \chem{O_3}, and
 MLS \chem{HNO_3} observations, and from the assimilation of all the
 datasets. Also shown are results for data assimilation that optimizes
 a~height-independent \chem{LNO_x} source scale factor (2-D). Standard
 deviations obtained from all the data assimilation estimates are also
 listed.}
```

```
%\scalebox{.85}[.85]
{\begin{tabular}{lccccccccc}
\tophline
&\multicolumn{4}{c}{Jan} &&\multicolumn{4}{c}{Jul}\\
\cline{2-5}\cline{7-10}
&NH &TR &SH &GL &&NH &TR &SH &GL   \\
\middlehline
CTM    &0.58 &2.82 &0.95 &4.35 &&3.17 &2.41 &0.31 &5.89 \\
\cline{1-10}
TES  \chem{O_3} &0.64 &2.34 &1.01 &4.00 &&3.21 &3.21 &0.34 &6.76 \\
OMI  \chem{NO_2}  &0.78 &3.25 &1.20 &5.22 &&3.99 &3.51 &0.51 &8.01 \\
MLS \chem{O_3}   &0.74 &4.89 &1.21 &6.84 &&4.82 &4.69 &0.31 &9.83 \\
MLS \chem{HNO_3}   &0.82 &4.89 &1.89 &7.56 &&4.33 &3.66 &0.31 &8.30 \\
ALL   &0.78 &3.99 &1.39 &6.15 &&4.69 &2.99 &0.50 &8.18 \\
\cline{1-10}
TES  \chem{O_3} (2-D) &0.72 &3.35 &1.67 &5.75 &&4.94 &3.74 &0.37 &9.05 \\
OMI  \chem{NO_2} (2-D) &0.78 &3.29 &1.12 &5.18  &&2.86 &3.54 &0.49 &6.89 \\
MLS \chem{O_3} (2-D)  &0.85 &3.92 &1.07 &5.85 &&5.46 &4.10 &0.37 &9.92 \\
MLS \chem{HNO_3} (2-D)  &0.74 &3.66 &1.34 &5.75 &&3.93 &2.80 &0.33 &7.07 \\
ALL (2-D)  &0.84 &3.56 &1.23 &5.63 &&3.00 &3.14 &0.51 &6.65 \\
\cline{1-10}
Standard dev.   &0.07 &0.77 &0.28 &0.96 &&0.89 &0.56 &0.09 &1.23 \\
\bottomhline
\end{tabular}}
%\hack{
%\setlength\tabularwidth{0.9\tabularwidth}
%}
%\scalebox{.7}[.7]{
\belowtable{%
%\hack{\vspace*{2mm}}
}
%}
\end{table*}

%\clearpage

\begin{table*}[t]
\caption{The ~~global~~ spatial correlation (Corr), ~~global~~ mean difference (Bias),
and ~~global~~ mean root-mean-square error (RMSE) of the three-monthly
```

```
mean tropospheric \chem{O_3} columns (TOCs) for the OMI/MLS data of
December--February (DJF) and June--August (JJA) in 2007. The results
of the CTM simulation and data assimilation for
are shown for the latitude band 30{\degree}\,S--30{\degree}\,N for the globe (180{\degree}
\,W--180{\degree}\,E), Africa (15{\degree}\,W--50{\degree}\,E), India (50{\degree}
\,E--90{\degree}\,E), the maritime continent (90{\degree}\,E--140{\degree}\,E), Pacific
(140{\degree}\,E--80{\degree}\,W), South America (80{\degree}\,W--40{\degree}\,W), and
Atlantic (40{\degree}\,W--15{\degree}\,W)are shown.}
%\scalebox{.85}[.85]
{\begin{tabular}{llccccccc}
\tophline
& &\multicolumn{3}{c}{DJF} &&\multicolumn{3}{c}{JJA}\\
\cline{23-45}\cline{67-89}

& &Corr &Bias &RMSE &&Corr &Bias &RMSE   \\
\middlehline
Globe & CTM    &0.85 &1.92 &4.16 &&0.92 &1.41 &3.26\\
& Assim.     &0.86 &$-$0.55 &2.85 &&0.92 &0.19 &2.59\\
\bottomhline
\middlehline
Africa& CTM    &0.65 &3.26 &22.22 &&0.88 &3.54 &18.91\\
& Assim.     &0.75 &0.78 &6.96 &&0.84 &2.92 &15.64\\
\middlehline
India & CTM    &0.83 &2.15 &24.36 &&0.96 &2.91 &16.28\\
& Assim.     &0.84 &0.25 &8.67 &&0.95 &1.03 &4.67\\
\middlehline
Maritime & CTM    &0.88 &1.23 &14.26 &&0.96 &1.49 &8.38\\
continent & Assim.     &0.87 &0.00 &5.78 &&0.94 &$-$1.03 &5.52\\
\middlehline
Pacific & CTM    &0.89 &0.27 &9.83 &&0.96 &$-$0.15 &8.41\\
& Assim.     &0.89 &$-$1.35 &8.24 &&0.97 &$-$1.67 &4.92\\
\middlehline
South & CTM    &0.80 &3.36 &15.83 &&0.61 &1.18 &13.98\\
America & Assim.     &0.75 &$-$0.13 &5.32 &&0.81 &0.12 &6.46\\
\middlehline
Atlantic & CTM    &0.01 &4.10 &25.64 &&0.74 &1.91 &8.18\\
& Assim.     &0.32 &0.62 &5.42 &&0.83 &1.56 &5.84\\
\bottomhline
\end{tabular}}
%\hack{
%\setlength\tabularwidth{0.9\tabularwidth}
```

```
%}
%\scalebox{.7}[.7]{
\belowtable{%
%\hack{\vspace*{2mm}}
}
%}
\end{table*}

%\clearpage

\begin{table*}[t]
\caption{Similar to Table~34, but lists the \chem{LNO_x} sources
  obtained from the control data assimilation run (Control), with
  a~15\,{\%} addition of artificial OMI \chem{NO_2} bias (w/ OMI bias),
  with the TES \chem{O_3} bias correction (TES bias correction), without
  the OMI cloud-covered observations (w/o OMI cloud), with the SST
  data for 1997 (year 1997 SST), with 20\,{\%} increases in the
  convective mass flux (+20\,{\%} convection), with 20\,{\%} increases
  in the a~priori errors of the \chem{LNO_x} source and the surface
  \chem{NO_x} emissions (+20\,{\%} \chem{LNO_x} a priori error and +20\,{\%}
  \chem{SNO_x} a priori error), and with 15\,{\%} increases in the a~priori values of
  the \chem{LNO_x} sources (+15\,{\%} \chem{LNO_x} a priori sourceprior). The total
  bias due to all terms is computed as a~random addition of the
  individual biases. See the text for details.}
%\scalebox{.85}[.85]
{\begin{tabular}{lccccccccc}
\tophline
&\multicolumn{4}{c}{Jan} &&\multicolumn{4}{c}{Jul}\\
\cline{2-5}\cline{7-10}
&NH &TR &SH &GL &&NH &TR &SH &GL   \\
\middlehline
Control    &0.78 &3.99 &1.39 &6.15 &&4.69 &2.99 &0.50 &8.18 \\
\cline{1-10}
w/ OMI bias   &0.87 &3.97 &1.46 &6.31 &&4.61 &3.08 &0.50 &8.18 \\
TES bias correction   &0.68 &3.79 &1.36 &5.83 &&4.19 &2.74 &0.29 &7.21 \\
w/o cloud OMI   &0.76 &4.04 &1.31 &6.09 &&4.13 &2.89 &0.29 &7.33 \\
year 1997 SST   &0.76 &3.89 &1.37 &6.03 &&4.71 &3.06 &0.51 &8.26 \\
+20\,{\%} convection   &0.80 &3.76 &1.37 &5.89 &&4.27 &2.99 &0.50 &8.09 \\
+20\,{\%} \chem{LNO_x} a priori error   &0.83 &3.75 &1.32 &5.90 &&4.59 &2.93 &0.51 &8.03
\\
```

```
+20\,{\%} \chem{SNO_x} a priori err~~or~~  &0.81 &3.77 &1.27 &5.85 &&4.58 &2.83 &0.50 &7.90
\\
+15\,{\%} \chem{LNO_x} a priori ~~prior~~source   &0.83 &4.10 &1.48 &6.41 &&5.29 &3.16 &0.57
&9.02 \\
\cline{1-10}
Total bias  &0.16 &0.47 &0.20 &0.66 &&1.06 &0.38 &0.31 &1.58 \\
\bottomhline
\end{tabular}}
%\hack{
%\setlength\tabularwidth{0.9\tabularwidth}
%}
%\scalebox{.7}[.7]{
\belowtable{%
%\hack{\vspace*{2mm}}
}
%}
\end{table*}

%\clearpage

\begin{table*}[t]
\caption{
The global and regional total flash rate (FR, \unit{flashes\,s^{-1}}) estimated from the model parameterization (1st row) and the LIS/OTD climatological observations (2nd row), and the NO production efficiency (NO prod., \unit{mol\,NO\,flash^{-1}}) estimated from the total \chem{LNO_x} sources analysed from data assimilation with the model flash rate (3rd row) and with the LIS/OTD observations (4th row). The NO production efficiency predicted by the model, as estimated from the simulated \chem{LNO_x} sources and the model flash rate, is also shown (5th row). The definitions of the regions are same as for Fig.~7 and Table 1. The analysis results as measured from the LIS measurements for 2007 for regions within latitudes between 35{\degree}\,S and 35{\degree}\,N are shown in brackets.}
%\scalebox{.85}[.85]
{\begin{tabular}{lcccccc}
\tophline
&\multicolumn{2}{c}{FR [\unit{flashes\,s^{-1}}]} & &\multicolumn{3}{c}{NO prod. [\unit{mol
\,NO\,flash^{-1}}]}\\
\cline{2-3}\cline{5-7}
& Model & LIS/OTD & & Assim. w/ model FR & Assim. w/  LIS/OTD FR & Model   \\
\middlehline
NH          & 12.7  & 12.6 &  & 351  & 353 & 256 \\
```

```
TR          & 21.3 & 27.1 (26.2) & & 377  & 296 (306) & 285\\
SH          & 7.2  & 6.8 &  & 246 & 261 & 179 \\
GL          & 41.2  & 46.5 &  & 347 & 308 & 258\\
\middlehline
Land          & 32.2  & 33.9  & & 388  & 368 & 294 \\
Ocean         & 9.1  & 12.5  & & 201 & 145 & 128 \\
\middlehline
North America          & 2.4  & 4.9 &  & 385  & 191 & 282 \\
Europe                 & 0.7  & 0.9  & & 383  & 268 & 244\\
Northern Eurasia        & 2.8  & 2.2 &  & 443  & 574 & 311\\
Pacific           & 1.6 & 0.9 (0.9) & & 268  & 460 (487) & 143 \\
South America          & 7.9 & 8.1 (7.8) &  & 389  & 379 (394) & 304 \\
Atlantic Ocean          & 0.3  & 0.3 (0.3) &  & 194  & 215 (194) & 117 \\
Northern Africa         & 5.0 & 6.2 (6.0) & & 364 & 393 (304) & 288 \\
Southern Africa         & 4.3  & 7.7 (7.5) & & 399 & 224 (228) & 306 \\
Indian Ocean          & 0.5 & 0.1 &&  196 & 883 &  121 \\
Southeast Asia          & 3.0 & 4.5 (3.9) & & 363 & 224 (280) & 280 \\
Australia             & 1.5  & 2.0 &  & 270 &  202 & 226 \\
\bottomhline
\end{tabular}}
%\hack{
%\setlength\tabularwidth{0.9\tabularwidth}
%}
%\scalebox{.7}[.7]{
\belowtable{%
%\hack{\vspace*{2mm}}
}
%}
\end{table*}

%\clearpage

\begin{table*}[t]
\caption{
A comparison of estimates of \chem{NO_x} amount produced per flash (\unit{mol\,NO
\,flash^{-1}}), adapted from Peterson and Beasley (2011) and Koshak et al. (2014).}
%\scalebox{.85}[.85]
{\begin{tabular}{lllc}
\tophline
First author & Year & Methodology  & Moles/flash     \\
```

```
\middlehline
Levine  & 1981  & Laboratory  & 8.30  \\
Kumar  & 1995  & Field study  & 8.30  \\
Dawson  & 1980  & Theoretical  & 13.28  \\
Beirle  & 2010  & Satellite  & 16.61  \\
Tuck  & 1976  & Theoretical  & 18.27  \\
Hill  & 1980  & Theoretical  & 19.93  \\
Koshak  & 2010  & Theoretical  & 23.40  \\
Cooray  & 2009  & Theoretical  & 33.21  \\
Lawrence  & 1995  & Review  & 38.19  \\
Nesbitt  & 2000  & Field study  & 44.25  \\
Huntrieser  & 2002  & Field study  & 44.84  \\
Wang  & 1998  & Laboratory  & 51.48  \\
Peyrous  & 1982  & Laboratory  & 53.14  \\
Ridley  & 2004  & Field study  & 53.14  \\
Beirle  & 2006  & Satellite  & 89.67  \\
Koshak  & 2014  & Theoretical  & 101.17  \\
Sisterson  & 1990  & Theoretical  & 136.17  \\
Noxon  & 1976  & Field study  & 166.06  \\
Chameides  & 1977  & Theoretical  & 166.06  \\
Kowalczyk  & 1982  & Theoretical & 166.06  \\
Bucsela  & 2010  & Satellite  & 174.36  \\
Schumann  & 2007  & Review  & 249.09  \\
Huntrieser  & 2011  & Field study  & 250.00  \\
DeCaria  & 2000  & Theoretical  & 258.39  \\
Miyazaki & (This study)  & Satellite  & 307.55  \\
Fehr  & 2004  & Field study & 348.72  \\
Rahman  & 2007  & Field study  & 398.54  \\
Chameides  & 1979  & Theoretical  & 415.14  \\
DeCaria  & 2005  & Theoretical  & 460.00  \\
Martini  & 2011  & Theoretical  & 480.88  \\
Hudman  & 2007  & Theoretical & 500.00  \\
Ott  & 2010  & Theoretical & 500.00  \\
Jourdain  & 2010  & Theoretical  & 520.00  \\
Drapcho  & 1983  & Field study  & 664.23  \\
Franzblau  & 1989  & Field study  & 4981.73  \\
\bottomhline
\end{tabular}}
%\hack{
%\setlength\tabularwidth{0.9\tabularwidth}
```

```
%}
%\scalebox{.7}[.7]{
\belowtable{%
%\hack{\vspace*{2mm}}
}
%}
\end{table*}

\clearpage

\begin{table*}[t]
\caption{The mean ozone concentration bias (in ppbv) between the
  CHASER simulations and the global ozonesonde observations for
  January 2007 in the NH (25{\degree}\,N--90{\degree}\,N) and for July
  2007 in the tropics (TR, 25{\degree}\,S--25{\degree}\,N) and the SH
  (90{\degree}\,S--25{\degree}\,S). The CHASER simulation results
  using the a~priori emissions sources (A~priori), the \chem{LNO_x}
  sources (\chem{LNO_x}), and the \chem{LNO_x} sources and surface
  \chem{NO_x} emissions (L\,+\,\chem{SNO_x}) are shown. The results from the
  CHASER-DAS simultaneous assimilation are also listed (DAS).}
\scalebox{.688}[.688]
{\begin{tabular}{lccccccccccccc}
\tophline
&\multicolumn{4}{c}{NH in Jul} &&\multicolumn{4}{c}{TR in Jan} &&\multicolumn{4}{c}{SH in
Jan}\\
\cline{2-5}\cline{7-10}\cline{12-15}
&A~priori &\chem{LNO_x} &L\,+\,\chem{SNO_x} &DAS &&A~priori &\chem{LNO_x} &L\,+\,
\chem{SNO_x} &DAS &&A~priori &\chem{LNO_x} &L\,+\,\chem{SNO_x} &DAS \\
\middlehline
750--450\,\unit{hPa}   &$-$12.3 &$-$11.7 &$-$0.2 &$-$1.8 &&18.5 &20.2 &16.6 &16.4 &&$-
$4.1 &$-$2.0 &$-$2.8 &$-$4.9 \\
450--200\,\unit{hPa}  &$-$6.8 &$-$5.9 &0.7 &1.3 &&8.9 &9.5 &3.3 &3.3 &&9.9 &7.4 &3.4 &$-
$1.0 \\
200--90\,\unit{hPa}   &19.8 &19.7 &4.8 &4.5 &&42.2 &34.9 &21.7 &10.4 &&219.5 &136.2
&149.5 &45.3 \\
\bottomhline
\end{tabular}}
%\hack{
%\setlength\tabularwidth{0.9\tabularwidth}
%}
```

```
%\scalebox{.7}[.7]{
\belowtable{%
%\hack{\vspace*{2mm}}
}
%}
\end{table*}

\clearpage

\begin{figure}
\includegraphics[width=1200mm]{acpd-2013-0838-f01}
\caption{Schematic diagram of the constraints on \chem{LNO_x} brought
  by the different satellite retrieval products. The vertical bars
  indicate the vertical sensitivity range for the species
  observed. Through these different sensitivities the assimilation
  system extracts information about the total \chem{LNO_x} source and
  its profile, the surface emissions, inflows from the stratosphere,
  and the chemical interactions in the troposphere through the
  observation of multiple species. Because these sensitivity ranges
  cover a~large part of the troposphere, it is important that the
  analysis simultaneously optimises the \chem{LNO_x} source strength,
  surface emissions as well as concentrations of the reactive gases
  involved.}
\end{figure}

\clearpage

\begin{figure}
\includegraphics[width=90mm]{acpd-2013-0838-f02}
\caption{Global distributions of the mean annual flash rate (in flashes \unit{km^{-2}}
\unit{yr^{-1}}) estimated from (a) the LIS/OTD high--resolution monthly climatology (HRMC)
data (Cecil et al., 2014) and (b) the model simulation for 2007.
}
\end{figure}

\clearpage

\begin{figure}
\includegraphics[width=120mm]{acpd-2013-0838-f023}
\caption{Vertical profiles of correlations between the \chem{LNO_x}
```

```
  sources and the concentrations of various chemical species as
  estimated from the background error covariance matrix based on
  CHASER ensemble simulations, averaged over central Africa for July
  2007. The regional monthly mean of the covariance estimated for each
  grid point is plotted. The correlation is shown in red or blue where
  positive or negative, respectively.}
\end{figure}

\clearpage

\begin{figure*}
\includegraphics[width=~~8~~130mm]{acpd-2013-0838-f0~~3~~4}
\caption{Global maps of (left) the concentration differences between
  the CHASER simulations with and without lightning sources and
  (right) the mean ratio of the lightning signals to the measurement
  errors as estimated along each satellite track by applying the
  averaging kernels of OMI \chem{NO_2} (in
  $10^{14}$\,\unit{molec\,cm^{-2}}), TES \chem{O_3} (ppbv) at
  300\,\unit{hPa}, MLS \chem{O_3} (ppbv) at 215\,\unit{hPa}, and MLS
  \chem{HNO_3} (pptv) at 150\,\unit{hPa} for June, July, and August in
  2007. A~super observation approach is employed to the OMI
  measurements, whereas individual observations are used in the
  analysis of the others.}
\end{figure*}

\clearpage

\begin{figure}
\includegraphics[width=1~~2~~00mm]{acpd-2013-0838-f0~~4~~5}
\caption{Seasonal variations of the total \chem{LNO_x} sources (in
  TgN) analysed from the data assimilation (solid lines) and estimated
  from the model simulation (dashed lines) over the globe
  (90{\degree}\,S--90{\degree}\,N), the Northern Hemisphere (NH,
  20--90{\degree}\,N), the tropics (TR,
  20{\degree}\,S--20{\degree}\,N), and the Southern Hemisphere (SH,
  90--20{\degree}\,S) for 2007.}
\end{figure}

\clearpage
```

\begin{figure*}
\includegraphics[width=1~~2~~80mm]{acpd-2013-0838-f0~~5~~6}

\caption{Global distributions of the annual \chem{LNO_x} source
(left, 10^{-12}\,\unit{kg\,m^{-2}\,s^{-1}}), its seasonal amplitude
(centre, in 10^{-12}\,\unit{kg\,m^{-2}\,s^{-1}}), and the timing of peak
sources (right, in months) for 2007. Shown are the a~priori sources
estimated from the CTM parameterisation (upper), the a~posteriori
sources from the data assimilation (middle), and the analysis
increment (lower). The analysis increment equals the a~posteriori
sources minus the a~priori sources. The peak timing is estimated for
regions with the analysed annual sources of \chem{LNO_x} greater than
0.7×10^{-13}\,\unit{kg\,m^{-2}\,s^{-1}}.}
\end{figure*}

\clearpage

\begin{figure*}
\includegraphics[width=1~~2~~70mm]{acpd-2013-0838-f0~~6~~7}

\caption{Seasonal variations of the regional \chem{LNO_x} sources (in
\unit{Tg\,N}) for (1) North America (120--65{\degree}\,W,
20--60{\degree}\,N), (2) Europe (10{\degree}\,W--30{\degree}\,E,
35--60{\degree}\,N), (3) northern Eurasia (60--130{\degree}\,E,
30--68{\degree}\,N), (4) the Pacific
(154--180{\degree}\,E, 35{\degree}\,S--20{\degree}\,N
and 180{\degree}\,E--88{\degree}\,W,
35{\degree}\,S--12{\degree}\,N), (5) South America
(77--39{\degree}\,W, 35{\degree}\,S--10{\degree}\,N), (6) the
Atlantic ocean (35{\degree}\,W--8{\degree}\,E,
30{\degree}\,S--3{\degree}\,N), (7) northern Africa
(15{\degree}\,W--48{\degree}\,E, 3--25{\degree}\,N), (8) southern
Africa (10--48{\degree}\,E, 30{\degree}\,S--3{\degree}\,N), (9) the
Indian ocean (52--108{\degree}\,E, 40--9{\degree}\,S), (10)
Southeast Asia (95--146{\degree}\,E, 9{\degree}\,S--26{\degree}\,N),
and (11) Australia (112--154{\degree}\,E,
40--12{\degree}\,S) analysed from the data assimilation (black) and estimated from the model simulation (red). The total annual values (in \unit{Tg\,N\,yr^{-1}}) are displayed in each panel. -Results for all land areas and for all the

oceans are also plotted.}

```
\end{figure*}

\clearpage

\begin{figure*}
\includegraphics[width=1~~270mm]{acpd-2013-0838-f0~~78}
\caption{Similar to Fig.~~67, but shows for the vertical profile of the
  annual mean a~priori (black) and a~posteriori (red) \chem{LNO_x}
  sources (left panels) and the seasonal variation of the monthly mean
  vertical profile of the a~posteriori \chem{LNO_x} source (right
  panels) in \unit{pptv\,day^{-1}}.}
\end{figure*}

\clearpage

\begin{figure*}
\includegraphics[width=1~~270mm]{acpd-2013-0838-f0~~89}
\caption{Latitude-pressure cross-sections of the monthly mean
  analysis increment (the assimilation minus the CTM simulation) for
  the \chem{LNO_x} source (in \unit{pptv\,day^{-1}}) obtained from assimilation of
  (left) all the data, (2nd from left) TES \chem{O_3} data, (centre)
  OMI \chem{NO_2} data, (2nd from right) MLS \chem{O_3} data, and
  (right) MLS \chem{HNO_3} data in (top) January 2007 and (bottom)
  July 2007. The interval of the contour lines is 10\,\unit{pptv\,day^{-1}}.}
\end{figure*}

\clearpage

\begin{figure*}
\includegraphics[width=1~~170mm]{acpd-2013-0838-f~~0910}
\caption{Vertical \chem{O_3} profiles (in ppbv) obtained from
  ozonesondes (black), the CTM simulation (blue), and the data
  assimilation (red) for Costa Rica (left), Irene in South Africa (2nd
  from left), American Samoa (~~2nd from right~~center), ~~and~~ San Cristobal in
  Ecuador (2nd from right), and Ascension in
  the tropical Atlantic (right) during December--February (DJF, top), March--May
  (MAM, 2nd from top), June--August (JJA, 2nd from bottom), and
  September--November (SON, bottom) in 2007. The error bars represent
  the standard deviation of all the data within one bin.}
```

\end{figure*}

\clearpage

\begin{figure*}
\includegraphics[width=1~~2~~80mm]{acpd-2013-0838-f10~~1~~}

\caption{Spatial distributions of tropospheric \chem{NO_2} column concentration (upper, 10^{15}\,\unit{molec\,cm^{-2}}), \chem{O_3} concentration at 300\,\unit{hPa} (middle, in ppbv), and cloud fraction (lower) over the western Pacific. Each is averaged over 14--21 August 2007. For the \chem{NO_2} and \chem{O_3} concentrations, the results obtained from the OMI and TES satellite retrievals (left), the CTM simulation (centre), and the data assimilation (right) are shown. For the cloud fraction, the results obtained from the OMI retrievals (left) and the GCM simulation (centre) are shown. The numbers in brackets represent the regional mean value for each plot.}
\end{figure*}

\clearpage

\begin{figure*}
\includegraphics[width=~~9~~170mm]{acpd-2013-0838-f1~~1~~2}

\caption{Latitude-pressure cross-section of the longitudinal-mean (6{\degree}\,W--30{\degree}\,E) of the \chem{LNO_x} source (upper, in \unit{pptv\,day^{-1}}), \chem{NO_2} concentration (middle, in \unit{pptv}), and \chem{O_3} concentration (lower, ppbv) over Africa. The results obtained from the CTM simulation (centre) and analysed from the data assimilation (right) are presented. The lower left panel shows the result obtained from the TES measurement for \chem{O_3} concentration. Also shown are the latitudinal distributions of (upper left panel) the longitudinal-mean surface emissions of \chem{NO_x} (in 10^{-11}\,\unit{kg\,m^{-2}\,s^{-1}}) as obtained from the a priori emissions constructed based on the EDGAR 4.2, the GFED 3.1, and the GEIA inventories (see text in Sect. 3.1.1) (blue line) and analysed from the data assimilation (red line) ~~as obtained from the emission inventories (blue line) and the assimilation (red line)~~ and (centre left) the longitudinal-mean tropospheric \chem{NO_2} columns (in 10^{15}\,\unit{molec\,cm^{-2}}) as obtained from the

```
  OMI measurements (black line), the CTM simulation (blue line), and
  analysed from the data assimilation (red line). The vectors represent
  meridional-vertical winds. Each is averaged over 10--20 July 2007.}
\end{figure*}

\clearpage

\begin{figure*}
\includegraphics[width=1~~230mm]{acpd-2013-0838-f1~~23}
\caption{Latitude-pressure cross-section of the \chem{LNO_x} source
  differences between the data assimilations with and without (with minus without) the
  cloud-covered OMI \chem{NO_2} observations (in \unit{pptv\,day^{-1}}) for January
  and July in 2007. The increases and decreases in the source due to
  assimilation of the cloud-covered observations correspond to
  positive and negative values represented by red and blue,
  respectively.}
\end{figure*}

\clearpage

\begin{figure}
\includegraphics[width=90mm]{acpd-2013-0838-f14}
\caption{Global distributions of the \chem{NO} production efficiency (\unit{mol\,NO
\,flash^{-1}}) estimated from the annual total \chem{LNO_x} sources analysed from data
assimilation for 2007 with (a) the LIS/OTD HRMC flash climatology data (Cecil et al., 2014), and
with (b) the model flash rate for 2007. The results in (a) are shown for the region with the
observed annual flash rates of greater than 0.44 flashes \unit{km^{-2}} \unit{yr^{-1}} to avoid
unrealistically large estimates.
}
\end{figure}

%\addtocounter{figure}{-1}\renewcommand{\thefigure}{\arabic{figure}a}


\end{document}
\endpoint
```