Manuscript prepared for Atmos. Chem. Phys.
with version 5.0 of the LaTeX class copernicus.cls.
Date: 10 February 2014

Global lightning NO_x production estimated by an assimilation of multiple satellite datasets

K. Miyazaki[1], H. J. Eskes[2], K. Sudo[3], and C. Zhang[4]

[1]Japan Agency for Marine-Earth Science and Technology, Yokohama 236-0001, Japan
[2]Royal Netherlands Meteorological Institute (KNMI), Wilhelminalaan 10, 3732 GK, De Bilt, the Netherlands
[3]Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan
[4]International Pacific Research Center, University of Hawaii at Manoa, Honolulu, Hawaii, USA

Correspondence to: K. Miyazaki (kmiyazaki@jamstec.go.jp)

Abstract. The global source of lightning-produced NO_x (LNO_x) is estimated by assimilating observations of NO_2, O_3, HNO_3, and 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 LNO_x source while improving the representations of the entire chemical system affecting atmospheric NO_x, including surface emissions and inflows from the stratosphere. The annual global LNO_x source amount and NO production efficiency are estimated at $6.3\,\mathrm{Tg\,N\,yr^{-1}}$ and $310\,\mathrm{mol\,NO\,flash^{-1}}$, respectively. Sensitivity studies with perturbed satellite datasets, model and data assimilation settings leads to an error estimate of about $1.4\,\mathrm{Tg\,N\,yr^{-1}}$ on this global LNO_x source. These estimates are significantly different from those estimated from a parameter inversion that optimises only the LNO_x source from NO_2 observations alone, which may lead to an overestimate of the source adjustment. The total LNO_x source is predominantly corrected by the assimilation of OMI 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 $1\,\mathrm{km}$ 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 LNO_x source estimation.

1 Introduction

Lightning-produced NO_x (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). LNO_x accounts for only about 10–20 % of the global NO_x sources, but is the most dominant source in the upper troposphere (e.g. Galloway et al., 2004). A small fraction of LNO_x can lead to significant ozone production in the upper troposphere (Thompson et al., 1994), because the O_3 production efficiency per NO_x molecule typically increases with height owing to the longer lifetime of NO_x and the highly non-linear dependence of ozone production on NO_x (Pickering et al., 1998; Martin et al., 2000; Jenkins and Ryu, 2004). Therefore, accurate estimates of 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 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 LNO_x sources on the basis of assumptions regarding the 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, any lightning parameterization cannot fully represent the regional variabil-

ity 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 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 NO_x near the surface and too little in the middle and upper troposphere (e.g. Ott et al., 2010).

The LNO_x sources can be optimized through a top-down approach, in which estimates of the LNO_x sources are obtained by finding the best match between model and observations. Tropospheric NO_2 column observations from satellite instruments have been used to constrain the global 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 NO_2 concentrations are influenced by not only the 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 LNO_x sources could cause large uncertainties in the LNO_x source estimates when observations are used to constrain only the LNO_x sources.

Satellite measurements of chemical species other than NO_2 provide important constraints on the LNO_x source by constraining the chemical interactions with NO_x and by reducing errors in other chemical species that influence the NO_x chemistry. Martin et al. (2007) demonstrated the ability of satellite NO_2, O_3, and HNO_3 measurements to constrain the LNO_x source. Advanced data assimilation techniques, such as four-dimensional variation (4D-Var) and ensemble Kalman filtering (EnKF), are powerful tools to combine multiple observations with models to obtain comprehensive constraints on 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 LNO_x sources and the surface emissions of NO_x and CO as well as the concentrations of 35 chemical species including the assimilated species (NO_2, O_3, HNO_3, and CO), while taking into account the chemical interactions through error covariance. The simultaneous optimization of various chemical fields improves the representation of the whole chemical system and thus reduces the model–observation mismatch arising from non-LNO_x sources and chemical processes. Therefore, this approach has the potential to improve global estimates of LNO_x sources when compared to previous top-down approaches that optimize LNO_x sources only. In this study, CHASER-DAS is utilized to assimilate multiple satellite datasets in order to analyse the global LNO_x sources, including the vertical profiles, for the whole 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: Section 2 describes the observations used for assimilation and validation. Section 3 introduces the data assimilation system. Section 4 presents the results of the global LNO_x source estimation. Section 5 presents the regional 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.

2 Data

2.1 Assimilated data

As depicted in Fig. 1, the LNO_x source is estimated from a simultaneous assimilation of NO_2, O_3, CO, and 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.

This section describes the observations, with a focus on the application of 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 useful constraints on LNO_x. The contributions from the individual satellite sensors are highlighted below.

2.1.1 OMI NO_2

Tropospheric NO_2 column retrievals obtained from the version-2 OMI DOMINO data product (Boersma et al., 2011) are used to constrain the LNO_x sources, the surface emissions of NO_x, and the concentrations of NO_y species. The overpass time of OMI (13:30) is more suitable for 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

$2.5^\circ \times 2.5^\circ$ for OMI NO_2 and MOPITT 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 km) footprint of OMI at nadir. Therefore, the data assimilation adjusts the 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 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} 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 NO_2 layer increase the effective albedo of the scene and increase the detected slant column, whereas high clouds partly screen the NO_2 column below (Boersma et al., 2005). We employ both clear-sky data and cloud-scene data in the LNO_x estimation because both are sensitive to NO_2 produced by lightning higher up in the atmosphere. For the cloud-covered observations the averaging kernel shows a sharp drop roughly 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.

2.1.2 TES O_3

The TES 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 km. Its vertical resolution is typically 6 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 LNO_x profiles. This can be attributed to its high sensitivity to LNO_x relevant altitude layers, typically with more than one degree of freedom (DOF) for the middle and upper troposphere (from 500 hPa to the tropopause).

2.1.3 MLS O_3, HNO_3

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

2.1.4 MOPITT CO

The MOPITT 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 CO emissions and the concentrations of CO and non-methane hydrocarbons (NMHCs). However, the covariances between the CO observations and the 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 CO observations indirectly affect the LNO_x source estimation through their influence on the oxidation capacity and the NO_x chemistry.

2.2 Validation data

Independent ozone observations are used to validate the performance of the data assimilation. The spatial distribution of tropospheric 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 $1^\circ \times 1.25^\circ$ (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 O_3 in the troposphere and the lower stratosphere. The validation is performed at five sites in different regions of the tropics: Costa Rica in central America (10° N, 84° W), Irene in South Africa (25.9° S, 28.2° E), Pago Pago in American Samoa (14.4° S, 170.6° W), San Cristobal in Ecuador (0.9° S, 89.6° W), and Ascension in the tropical Atlantic (8.0° S, 14.4° 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 hPa. The model profiles are linearly interpolated to the location and time of each observation point.

3 Methodology

3.1 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 LNO_x sources and surface emissions of NO_x and 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 NO_x-CO-OH-O_3 set of chemical reactions.

Miyazaki and Eskes (2013) demonstrated that multi-species data assimilation improves the analysis of surface NO_x emissions, in comparison with an inversion derived from NO_2 measurements alone. They showed that the assimilation of measurements for species other than NO_2 changes the regional estimates of monthly mean surface 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 LNO_x source estimation through corrections made to the concentrations of various chemical species. This is especially true for LNO_x, because all satellite sensors are sensitive in the altitude range where LNO_x and the ozone produced by LNO_x resides, see Fig. 1.

3.1.1 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°) and 32 vertical levels from the surface to 4 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 NO_x emissions are 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 NO_x emissions, depending on the dominant category for each emission category (Miyazaki et al., 2012b). The total NO_x emission by aircraft is obtained from EDGAR as 0.55 $\mathrm{Tg\,N\,yr^{-1}}$, which is similar to a more recent estimate of 0.49 $\mathrm{Tg\,N\,yr^{-1}}$ 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.

3.1.2 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, $\boldsymbol{x}_i^b (i = 1, ..., k)$, is globally obtained from the evolution of each ensemble model realisation, where $\boldsymbol{x}$ 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, $\boldsymbol{y}_i^b = H(\boldsymbol{x}_i^b)$, 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 ($\boldsymbol{x}_{apriori}$) and the averaging kernel ($\boldsymbol{A}$), which maps the model fields ($\boldsymbol{x}$: N- (the system dimension) dimensional state vector) into observation space ($\boldsymbol{y}$: p- (the number of observation) dimensional observational vector) while taking into account the vertical averaging implicit in the observations as follows:

$$\boldsymbol{y} = H(\boldsymbol{x}) = \boldsymbol{x}_{apriori} + \boldsymbol{A}(S(x) - \boldsymbol{x}_{apriori}), \quad (1)$$

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

A background ensemble mean in the observation space, $\overline{\boldsymbol{y}^b} = \frac{1}{k}\sum_{i=1}^{k} \boldsymbol{y}_i^b$, or in the model space, $\overline{\boldsymbol{x}^b} = \frac{1}{k}\sum_{i=1}^{k} \boldsymbol{x}_i^b$, and an ensemble of background perturbations in the obser-

vation space, $\mathbf{Y}^b = \boldsymbol{y}_i^b - \overline{\boldsymbol{y}^b}$, or in the model space, $\mathbf{X}^b = \boldsymbol{x}_i^b - \overline{\boldsymbol{x}^b}$, are also computed.

The ensemble mean analysis is then updated by

$$\overline{\boldsymbol{x}^{\mathrm{a}}} = \overline{\boldsymbol{x}^b} + \mathbf{X}^b \tilde{\mathbf{P}}^{\mathrm{a}} \left(\mathbf{Y}^b\right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\boldsymbol{y}^{\mathrm{o}} - \overline{\boldsymbol{y}^b}\right), \quad (2)$$

where $\boldsymbol{y}^{\mathrm{o}}$ is the observation vector, and $\mathbf{R}$ is the $p \times p$ observation error covariance. 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,

$$\tilde{\mathbf{P}}^{\mathrm{a}} = \left[(k-1)I + \left(\mathbf{Y}^b\right)^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}^b\right]^{-1}. \quad (3)$$

The new analysis ensemble perturbation matrix in the model space $\mathbf{X}^{\mathrm{a}}$ is obtained by transforming the background ensemble $\mathbf{X}^b$,

$$\mathbf{X}^{\mathrm{a}} = \mathbf{X}^b \left[(k-1)\tilde{\mathbf{P}}^{\mathrm{a}}\right]^{1/2}. \quad (4)$$

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 $\mathrm{LNO_x}$ sources, the covariances with TES $\mathrm{O_3}$, OMI $\mathrm{NO_2}$, MLS $\mathrm{O_3}$, and MLS $\mathrm{HNO_3}$ observations are considered, while those with MOPITT CO data are not. MOPITT CO data affect the concentrations of CO, hydrocarbons, and formaldehyde only. Surface emissions of $\mathrm{NO_x}$ and CO are optimized using OMI $\mathrm{NO_2}$ data and MOPITT 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 $\mathrm{LNO_x}$ 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 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 min.

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).

3.2 $\mathrm{LNO_x}$ estimation

3.2.1 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 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): $F_{\mathrm{c}} = 3.44 \times 10^{-5} \times H^{4.92}$ (flashes min^{-1}) over continents; $F_{\mathrm{m}} = 6.40 \times 10^{-4} \times H^{1.73}$ (flashes min^{-1}) over ocean, where F is the total flash rate (flashes min^{-1}) and H is the cloud top height (km). 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 35 ° S-35 ° 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° S–20° N) by about 27 %, leading to about 13 % underestimation in the global total flash rate.

The LNO_x source is estimated at each grid point of CHASER by using the simulated lightning activity and making several assumptions. First, the CG proportion of total flashes, z, is estimated as a function of the cold cloud thickness (ΔH for $< 0\,^{\circ}\mathrm{C}$) on the basis of the following relationship (Price and Rind, 1993):

$$z = 0.021\times\Delta H^4 - 0.648\times\Delta H^3 + 7.493\times\Delta H^2 - 36.54\times\Delta H + 63.09 \quad (5)$$

This relationship is applied for clouds with $\Delta H > 5.5\,\mathrm{km}$, while z is set to zero for clouds with $\Delta H < 5.5\,\mathrm{km}$ based on the observation that low clouds almost exclusively have IC flashes during the growth stage. Second, following Price et al. (1997), the 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 $6.7\times10^9\,\mathrm{J\,flash^{-1}}$. Third, the vertical profiles of the 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 LNO_x source for 2007 was estimated at $4.7\,\mathrm{Tg\,N\,yr^{-1}}$ in the CHASER simulations.

3.2.2 Optimization by data assimilation

The multiplication factors for the LNO_x production rate ($\mathrm{mol\,s^{-1}}$) and the 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 3 shows the mean background error covariance structure between the LNO_x source and the concentrations of various species at different altitudes over central Africa in July. The concentrations of chemical species such as O_x, NO_x, N_2O_5, HNO_3, HNO_4 exhibit high positive correlations with the LNO_x sources in the middle and upper troposphere. The high correlations indicate the utility of measurements of these species to constrain the 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 LNO_x sources, and does not separately optimise for IG and CG flashes.

In order to provide meaningful constraints on LNO_x, the observation error of each retrieval must be sufficiently small compared to lightning signal. As shown in Fig. 4, the CHASER simulations with and without the lightning signal in the tropospheric NO_2 columns gives the magnitude of the boreal summer (June–August) mean lightning signal as roughly $1 \sim 5\times10^{14}\,\mathrm{molec\,cm^{-2}}$ over the tropical Atlantic, Africa, and India but roughly $1\times10^{14}\,\mathrm{molec\,cm^{-2}}$ 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 O_3 due to lightning are observed in the tropical upper troposphere, with contributions as large as 13 ppbv to the mean concentration, consistent with the analyses of Sauvage 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 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 LNO_x estimates.

The observed concentrations reflect contributions not only from 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 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 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 LNO_x profile and to distinguish between the surface NO_x emissions and LNO_x sources. The MLS and TES data not only provides information on the 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 NO_2 retrievals.

The a priori error is set to 40 % for the surface emissions of NO_x and CO and 60 % for the 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. 6.1.3.

4 Results

4.1 Global LNO_x source distributions

Data assimilation increases the global annual LNO_x source from 4.7 to 6.3 $TgN\,yr^{-1}$, as summarized in Table 2. The instantaneous uncertainty estimated from the mean analysis spread for the global source is $\pm 0.9\,TgN\,yr^{-1}$, 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 LNO_x is well constrained by the assimilation. The Northern Hemisphere (NH, 20–90° N), the tropics (20° S–20° N), and the Southern Hemisphere (SH, 90–20° S) contribute 31 %, 56 %, and 13 % to the analysed global source, respectively. These relative contributions are slightly modified from the a priori sources. The analysed global source is maximal in July, primarily resulting from the seasonal variation in the NH, as depicted in Fig. 5.

Figure 6 compares the simulation with the assimilation in terms of the global distribution of annual total 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 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 (c.f., Table 1).

4.2 Regional LNO_x source distributions

Figure 7 shows the seasonal variations of the 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 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 NO_x production efficiency over the oceans in the model simulation. The 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.

4.3 Vertical distribution of the LNO_x source

Figure 8 shows the changes in the vertical profiles of the LNO_x sources. Data assimilation increases the LNO_x 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. 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 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 NO_2 column retrievals could cause large uncertainties in the analyzed LNO_x sources over the oceans, as will be discussed in Section 6.1.1.

Table 3 summarizes the altitude (in km) with maximum LNO_x emission (i.e., source peak height). The predicted source peak heights of the regional sources over land mostly range from 5 km to 12 km but exceed 13 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 km), Australia (-0.41 km), and southern Africa (-0.38 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 km.

4.4 Relative contributions of individual assimilated datasets

The effects of individual assimilated datasets on the estimated LNO_x sources were evaluated through observing system experiments (OSEs) by separately assimilating each dataset with CHASER-DAS (Fig. 9). The assimilation of OMI 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 NO_2 retrievals in determining the LNO_x source distribution in the analysis.

The measurements of species other than 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 4. The vertical and latitudinal structure of the analysis increments obtained from the assimilation of MLS O_3 and HNO_3 data are generally similar (Fig. 9), revealing consistent constraints from datasets gathered for different species but with the same instrument. In contrast, the respective corrections from MLS O_3 and TES 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 LNO_x sources obtained from the assimilation of TES 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 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 O_3 and 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 CO data changes the free tropospheric OH concentration by up to 5 %, corresponding to an increase of about 25 % in the annual NH total CO emissions (Miyazaki et al., 2012a), which changed the monthly global 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 NO_x-CO-OH-O_3 set of chemical reactions, which may improve the LNO_x source estimation with the assimilation of TES O_3 data. Detailed analyses are required to measure the impact of the simultaneous assimilation on OPE.

In most previous studies, NO_2 measurements were used to optimise only the LNO_x production. In this case, the accuracy of the LNO_x source estimates is negatively affected by various sources of error, including the surface NO_x and 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 LNO_x source (i.e. single-parameter) inversion derived from NO_2 measurements only. The single-parameter approach tends to have larger sources (e.g., the estimated global LNO_x source was 8.89 $Tg\,N\,yr^{-1}$ for January and 12.8 $Tg\,N\,yr^{-1}$ 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 NO_x emissions are possibly underestimated by up to 67 % on the regional scale. The simulated 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 LNO_x

source inversion. Note that, because of the short assimilation cycle (i.e., 100 min.), the LNO_x source can be frequently and continuously increased to compensate for persistent model errors. This situation may cause larger LNO_x sources in our estimates compared to those with a longer (e.g., one month) assimilation cycle especially when only LNO_x sources are optimized.

5 Tropical regions

Lightning strongly influences the 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 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 LNO_x sources and O_3 distributions in several tropical regions.

5.1 Validation of tropospheric ozone

The validation of ozone profiles provides useful information on the performance of the LNO_x source estimation, because of the strong chemical links between the LNO_x sources and the O_3 concentration, which influence the simultaneous optimization in the data assimilation. Figure 10 shows a comparison of modeled and analysed vertical O_3 profiles with the ozonesonde observations taken at five tropical SHADOZ sites. The assimilation removes most of the 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 O_3 concentrations over the tropical Atlantic. Conversely, the assimilation does not obviously improve the lower tropospheric 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 5 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 ($r = 0.85 \sim 0.92$). 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 O_3 (Miyazaki et al., 2012a). Because lightning substantially influences the amount of O_3 in the tropics, and because the data assimilation simultaneously optimizes the O_3 and the LNO_x source, significantly improved agreement with independent ozone observations gives confidence in the performance of the LNO_x estimates.

5.2 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 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).

Figure 11 shows the regional distribution of total cloud fraction and tropospheric concentrations of NO_2 and 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 LNO_x source structures. Accurate simulations of the cloud position are important to properly distribute the 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 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 NO_2 and O_3 concentrations show better agreements with the observations (Fig. 11) because of the simultaneous data assimilation. The simulation underestimates the regional mean NO_2 concentration in the upper troposphere by 65 % compared to the OMI NO_2 retrieval, whereas data assimilation removes most of the bias. Agreements with TES 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 LNO_x sources and these chemical concentrations, the improvements suggest that the regional total 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.

5.3 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 LNO_x sources and chemical concentrations. NO_2 concentrations exhibit distinct vertical and latitudinal variations over western Africa (Fig. 12). These variations are produced by various factors such as convective uplifting, stratospheric inflows, surface sources, and lightning sources. The lower tropospheric NO_2 concentrations are maximal over northern (around 5–20° N) and central (15° S–Equator) Africa, owing to in situ emissions from biomass burning. In the upper troposphere, strong LNO_x sources cause a maximum NO_2 concentration over northern Africa (5–15° N). Data assimilation increases the surface NO_x emissions over northern and central Africa by up to 90 % and the LNO_x source over northern Africa by about 50 %, which acts to strengthen the local maximum in NO_2 concentrations.

The tropospheric O_3 distributions also show distinct variations over central Africa. The surface NO_x and CO emissions and LNO_x sources lead to O_3 production, whereas the inflows from the stratosphere along the subtropical jet stream predominantly determine the latitudinal O_3 variations in the upper troposphere. The multi-species data assimilation provides comprehensive constraints on these processes and the O_3 variations. The assimilation of TES data modifies the 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 O_3 production. These case studies demonstrate once more the utility of the simultaneous data assimilation for regional process studies.

6 Discussion

6.1 Uncertainty of the LNO_x source estimate

6.1.1 Systematic satellite observation errors

The quality of the assimilated measurements largely influences the LNO_x source uncertainty. Boersma et al. (2011) showed that the different OMI 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 NO_2 retrievals. This increased the monthly regional and global LNO_x sources by up to 14 % and 3 %, respectively, as summarized in Table 6. It is emphasized that low NO_2 concentrations over the oceans are mostly smaller than the OMI noise level. Errors related to the separation of stratospheric and tropospheric NO_2 could also cause errors in the OMI tropospheric NO_2 column retrievals (Lamsal et al., 2010; Boersma et al., 2011). These may cause large uncertainties in the analyzed LNO_x sources, especially over the oceans.

Any bias in the measurements of species other than NO_2 also affects the quality of the sources estimated in the simultaneous assimilation framework. TES 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 O_3 data was applied. This consisted of a uniform 3.3 ppbv above 500 hPa and 6.5 ppbv below 500 hPa, as recommended by Worden et al. (2009). The result of this experiment suggests that such a bias in TES O_3 increases the estimated monthly global LNO_x source by up to 14 % (Table 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 LNO_x reaches the top and anvil of the cloud (Ridley et al., 1996). The LNO_x produced can reside above the cloud for several days, because of the long lifetime of NO_x in the upper troposphere, and the enhanced concentrations may be detected by remote observations. Furthermore, an instrument like OMI is sensitive to 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 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 LNO_x source estimation, as summarized in Table 6 and depicted in Fig. 13. The large changes associated with the cloud-covered data are found above roughly 450 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 LNO_x source is increased by 12 % in July when cloud-covered observations are used. These results imply that cloud-covered OMI NO_2 data contain important information on the LNO_x source amounts above and inside clouds. However, errors in the simulated cloud profiles (c.f., Fig. 11), and in the retrieved cloud-altitude dependent averaging kernels, may cause biases in the cloudy scene analysis.

The lifetimes of NO_x, HNO_3, and 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 LNO_x source far from the convective clouds that produced the LNO_x. The large spread of the different estimates indicates that estimates of the LNO_x source distribution and the global 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.

6.1.2 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 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 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 LNO_x sources to become smaller because several sources of error are corrected simultaneously. In a parameter inversion that optimizes the LNO_x sources only, the analysis increments will compensate for errors that occur not only in the LNO_x sources but also in the other sources and species concentrations. This will result in overcorrections in the 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 NO_x chemical lifetime in the troposphere depends on altitude (i.e., longer at higher altitude), cumulus convection affects the total amount of NO_x in the troposphere and, accordingly, 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 LNO_x sources (Table 6). The high sensitivity of the 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. 11). An accurate determination of cloud positions is important for analysing the 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 Niño year) showed a variation of up to 4 % in the global LNO_x estimates (Table 6). The impact of changing the SST data was different for different regions; e.g., the LNO_x sources over the Pacific increased by 14 % in January. This indicates that the uncertainty in the 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 NO_2/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 NO_2/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 NO_x loss uncertainties on top-down 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 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 LNO_x source estimate. Although the aircraft NO_x emissions likely have relatively small uncertainties (e.g., Wilkerson et al., 2010), the LNO_x source estimates might be influenced by errors in the aircraft emissions especially along the major flight routes in the northern extratropics.

6.1.3 Data assimilation error modelling

χ^2 diagnostic tests (e.g. Menard and Chang, 2000) 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 covari-

ance 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 NO_2, 0.8 for TES O_3, 0.7 for MOPITT CO, 2.1 for MLS O_3, and 1.6 for MLS 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 O_3, NO_x, HNO_3, and N_2O_5 above 20 km altitude in CHASER.

The mean analysis spread, as estimated by transforming the background ensemble in the data assimilation (c.f., Eq. (4)), is about 0.9 $\mathrm{Tg\,N\,yr^{-1}}$ for the annual global source strength. This can be translated into an error estimate for the total yearly source. The 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 LNO_x estimations are uncorrelated, then the impact of the analysis spread on the uncertainty in the total yearly LNO_x can be estimated as $0.9/\sqrt{24}\,\mathrm{Tg\,N\,yr^{-1}}$, or about 0.2 $\mathrm{Tg\,N\,yr^{-1}}$. 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 NO_x by 20 % changes the global estimate of LNO_x sources by not more than a few percent, demonstrating the robustness of the source analysis (Table 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 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 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 LNO_x sources more efficiently. A sensitivity experiment in which the a priori global total LNO_x source is increased by 15 % demonstrates that the estimated LNO_x source amount is influenced by the a priori source setting (Table 6); the global a-posteriori LNO_x sources are increased by 4 % in January and 10 % in July. 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 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, HNO_3, and O_3 are generally longer than a day, and long-range transport of LNO_x can occur. A larger localization length may be useful when satellite measurements detect aged LNO_x air. We confirmed that doubling the localization scale (i.e. to 1200 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.

6.1.4 Total systematic error

The total error on the estimate of the 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 6. From the systematic satellite and model uncertainties listed in Table 6 we obtain an error estimate of about 1 $\mathrm{Tg\,N\,yr^{-1}}$ (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 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 NO_x resulting in a factor of 1.8 difference (3.3–5.9 $\mathrm{Tg\,N\,yr^{-1}}$) between upper and lower estimates of the LNO_x source. This would translate into an error bar of about 1 $\mathrm{Tg\,N\,yr^{-1}}$. If this is added as an independent source of error on top of the ones listed in Table 6, the total error bar would increase to about 1.4 ($=\sqrt{1^2+1^2}$) $\mathrm{Tg\,N\,yr^{-1}}$.

6.2 Implication for the parameterization of LNO_x

6.2.1 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 parameterization. However, Tost et al. (2007) concluded that the observed lightning

distributions were not even approximately reproduced with any of the lightning parameterizations 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 LNO_x amount from 4.7 to 6.3 $TgN\,yr^{-1}$ 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 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). Various factors, such as for instance the NO_x production per flash, are responsible for uncertainty in the LNO_x source parameterization. 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 NO_x production efficiency rather than those in the flash rate could be responsible for the large increase in the LNO_x sources over the tropical oceans. This will be further discussed in Sect. 6.2.2.

6.2.2 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 NO_x produced per flash would lead to an error in the LNO_x source estimates. The annual global LNO_x 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, as summarised in Table 8. For instance, Schumann and Huntrieser (2007) suggested a best estimate of 250 $mol\,NO\,flash^{-1}$. Ott et al. (2010) reported a mean value of 500 $mol\,flash^{-1}$ 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 $mol\,NO\,flash^{-1}$ gives reasonable performance in a chemical simulation over the United States. Boersma et al. (2005) used the Global Ozone Monitoring Experiment (GOME) NO_2 data and estimated a global LNO_x source strength of 1.1–6.4 $TgN\,yr^{-1}$, implying a production rate of 82–328 $mol\,NO\,flash^{-1}$ based on the LIS/OTD climatology.

The amount of 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 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 mol of NO in the NH (20–90° N) compared to 360 and 240 mol of NO in the tropics (20° S–20° 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 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 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 LNO_x sources and the flash observations. The obtained results imply general underestimations in the 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 remains regarding the ratio of NO_x production per flash by IC and CG flashes. Following Price et al. (1997), a lightning NO production of 1100 $mol\,(CG\,flash)^{-1}$ and 110 $mol\,(IC\,flash)^{-1}$ 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 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.

6.2.3 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 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 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 $1\,\mathrm{km}$ over land and the tropical oceans. Ott et al. (2010) suggested that the simplified treatment of LNO_x and wind fields by Pickering et al. (1998) will cause errors in the vertical 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 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 $300\,\mathrm{hPa}$ in January, and sources smaller by 14 % at $200\,\mathrm{hPa}$ in July in the tropics (20° S–20° N) for grid points with the LNO_x source column greater than $5 \times 10^{-15}\,\mathrm{kg\,m^{-2}\,s^{-1}}$ (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 4). The two estimations show a large discrepancy especially when TES O_3 data are assimilated. In the height-dependent analysis, it was estimated that the TES data assimilation mostly decreases the LNO_x sources in the upper troposphere (c.f., Fig. 9). In contrast, in the height-independent analysis, the 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.

6.3 Validation using forward CTM simulations

The 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 9. In the validation, the multiplication factors for the 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 NO_x sources for reproducing the 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 WOUDC/SHADOZ database, as in Miyazaki et al. (2012a). By using the estimated 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 NO_x emission data from the multiple data assimilation instead of the emission inventories, together with the estimated 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.

6.4 Comparisons with previous estimates

Based on various estimation results, Schumann and Huntrieser (2007) have provided a best estimate of $5 \pm 3\,\mathrm{Tg\,N\,yr^{-1}}$ for the annual global LNO_x source. Our estimate of $6.3 \pm 1.4\,\mathrm{Tg\,N\,yr^{-1}}$ 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 LNO_x source of $6 \pm 0.5\,\mathrm{Tg\,N\,yr^{-1}}$ and 3.3–5.9 $\mathrm{Tg\,N\,yr^{-1}}$, respectively. These estimates are also close to our estimate. The

annual global LNO_x source from our estimates corresponds to a mean NO production of about $310\,\mathrm{mol\,flash^{-1}}$ 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 NO production efficiency, the lightning activity and the 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 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.

7 Conclusions

The global source of lightning-produced NO_x (LNO_x) is estimated from an assimilation of multiple chemical species based on an ensemble Kalman filter approach. NO_2, O_3, HNO_3, and CO measurements obtained from multiple satellite instruments (OMI, MLS, TES, and MOPITT) provide comprehensive constraints on estimates of the global LNO_x source. This approach has the potential to reduce the influence of model errors on the LNO_x source estimation by simultaneously optimizing various aspects of the chemical system, including the surface emissions of NO_x and CO as well as the concentrations of 35 chemical species. Errors in these model fields other than the LNO_x sources introduce additional model–observation mismatches into the inversion and degrade the LNO_x source estimation. In most previous top-down estimates, only LNO_x sources were optimized from NO_2 measurements. In such cases, the LNO_x sources may be overcorrected since analysis increments are introduced to compensate for various sources of model error. Substantial differences in the estimated LNO_x sources are obtained between the single-parameter (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 LNO_x source estimation.

The assimilation provides substantial adjustments to the 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 NO_2 retrievals provide important constraints on the estimation of the 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 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 O_3 and NO_2.

The annual global LNO_x source amount and NO production efficiency based on the LIS/OTD observations are estimated by the assimilation system to be $6.3\,\mathrm{Tg\,N\,yr^{-1}}$ and $310\,\mathrm{mol\,NO\,flash^{-1}}$ 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 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 $1.4\,\mathrm{Tg\,N\,yr^{-1}}$.

The annual 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 LNO_x parameterization. The analysed 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 LNO_x sources have important implications for improving LNO_x parameterisations. First, errors in flash rates can explain only a small fraction of the uncertainty in 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 ex-

pected 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).

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.

References

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.

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:http://dx.doi.org/10.1029/2002JD00206610.1029/2002JD002066, 2002.

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.

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 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:http://dx.doi.org/10.5194/acp-6-1075-200610.5194/acp-6-1075-2006, 2006.

Boccippio, D. J.: Lightning scaling relations revisited, J. Atmos. Sci., 59, 1086–1104, 2002.

Boersma, K. F., Eskes, H. J., Meijer, E. W., and Kelder, H. M.: Estimates of lightning NO_x production from GOME satellite observations, Atmos. Chem. Phys., 5, 2311–2331, doi:http://dx.doi.org/10.5194/acp-5-2311-200510.5194/acp-5-2311-2005, 2005.

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., Leitão, J., Richter, A., Zhou, Y., and Brunner, D.: An improved tropospheric NO_2 column retrieval algorithm for the Ozone Monitoring Instrument, Atmos. Meas. Tech., 4, 1905–1928, doi:http://dx.doi.org/10.5194/amt-4-1905-201110.5194/amt-4-1905-2011, 2011.

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.

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.

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.

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:http://dx.doi.org/10.1029/2002JD00234710.1029/2002JD002347, 2003.

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 NO_x in an individual thunderstorm during STERAO-A, J. Geophys. Res., 105, 11601–11616, 2000.

DeCaria, A. J., Pickering, K. E., Stenchikov, G. L., and Ott, L. E.: Lightning-generated 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:http://dx.doi.org/10.1029/2004JD00555610.1029/2004JD005556, 2005.

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:http://dx.doi.org/10.1002/jgrd.5027210.1002/jgrd.50272, 2003.

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:http://dx.doi.org/10.1029/2011JD01570310.1029/2011JD015703, 2011.

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:http://dx.doi.org/10.1002/jgrd.5027210.1002/jgrd.50272, 2013.

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:http://dx.doi.org/10.1029/2004GL02230610.1029/2004GL022306, 2005.

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.

Fehr, T., Höller, H., and Huntrieser, H.: Model study on production and transport of lightning-produced NO_x in

a EULINOX supercell storm, J. Geophys. Res., 109, 1–17, doi:http://dx.doi.org/10.1029/2003JD00393510.1029/2003JD003935, 2004.

Gallardo, L. and Cooray, V.: Could cloud-to-cloud discharges be as effective as cloud-to-ground discharges in producing NO_x?, Tellus B, 48, 641–651, 1996.

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. Vörösmarty. 2004. Nitrogen Cycles: Past, Present and Future. Biogeochemistry 70: 153-226.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Q. J. Roy. Meteor. Soc., 125, 723–757, 1999.

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.

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:http://dx.doi.org/10.1029/2006JD00791210.1029/2006JD007912, 2007.

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.

Huntrieser, H., Schumann, U., Schlager, H., Höller, 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 NO_x production, Atmos. Chem. Phys., 8, 921–953, doi:http://dx.doi.org/10.5194/acp-8-921-200810.5194/acp-8-921-2008, 2008.

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:http://dx.doi.org/10.5194/acp-4-361-200410.5194/acp-4-361-2004, 2004.

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:http://dx.doi.org/10.1029/2006GL02828410.1029/2006GL028284, 2007.

Jourdain, L., Kulawik, S. S., Worden, H. M., Pickering, K. E., Worden, J., and Thompson, A. M.: Lightning NO_x emissions over the USA constrained by TES ozone observations and the GEOS-Chem model, Atmos. Chem. Phys., 10, 107–119, doi:http://dx.doi.org/10.5194/acp-10-107-201010.5194/acp-10-107-2010, 2010.

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:http://dx.doi.org/10.1175/BAMS-83-11-163110.1175/BAMS-83-11-1631, 2002.

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:http://dx.doi.org/10.1175/BAMS-83-11-163110.1175/BAMS-83-11-1631, 2002.

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.

Labrador, L. J., von Kuhlmann, R., and Lawrence, M. G.: The effects of lightning-produced NO_x and its vertical distribution on atmospheric chemistry: sensitivity simulations with MATCH-MPIC, Atmos. Chem. Phys., 5, 1815–1834, doi:http://dx.doi.org/10.5194/acp-5-1815-200510.5194/acp-5-1815-2005, 2005.

Lamsal, L. N., Martin, R. V., van Donkelaar, A., Celarier, E. a., Bucsela, 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, 15, doi:10.1029/2009JD013351, 2010.

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:http://dx.doi.org/10.1029/2006JD00794410.1029/2006JD007944, 2007.

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:http://dx.doi.org/10.5194/acp-12-2881-201210.5194/acp-12-2881-2012, 2012.

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.

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.

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:http://dx.doi.org/10.1029/2006JD00783110.1029/2006JD007831, 2007.

Mënard, 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.

Miyazaki, K. and Eskes, H.: Constraints on surface NO_x emissions by assimilating satellite observations of multiple species, Geophys. Res. Lett., 40, 4745–4750,

doi:http://dx.doi.org/10.1002/grl.5089410.1002/grl.50894, 2013.

Miyazaki, K., Eskes, H. J., Sudo, K., Takigawa, M., van Weele, M., and Boersma, K. F.: Simultaneous assimilation of satellite NO_2, O_3, CO, and HNO_3 data for the analysis of tropospheric chemical composition and emissions, Atmos. Chem. Phys., 12, 9545–9579, doi:http://dx.doi.org/10.5194/acp-12-9545-201210.5194/acp-12-9545-2012, 2012a.

Miyazaki, K., Eskes, H. J., and Sudo, K.: Global NO_x emission estimates derived from an assimilation of OMI tropospheric NO_2 columns, Atmos. Chem. Phys., 12, 2263–2288, doi:http://dx.doi.org/10.5194/acp-12-2263-201210.5194/acp-12-2263-2012, 2012b.

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:http://dx.doi.org/10.1029/2012JD01793410.1029/2012JD017934, 2012.

Nassar, R., Logan, J. A., Worden, H. M., Megretskaia, I. A., Bowman, K. W., Osterman, G. B., Thompson, A. M., Tarasick, 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:http://dx.doi.org/10.1029/2007JD00881910.1029/2007JD008819, 2008.

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.

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 NO_x and its vertical distribution calculated from three-dimensional cloud-scale chemical transport model simulations, J. Geophys. Res., 115, D04301, doi:http://dx.doi.org/10.1029/2009JD01188010.1029/2009JD011880, 2010.

Pan, D.-M., and Randall, D. A.: A cumulus parameterization with a prognostic closure, Q. J. Roy. Meteor. Soc., 124, 949–981, 1998.

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.

Pickering, K. E., Wang, Y., Tao, W. K., Price, C., and Muller, J. F.: Vertical distributions of lightning NO_x for use in regional and global chemical transport models, J. Geophys. Res., 103, 31203–31216, doi:http://dx.doi.org/10.1029/98JD0265110.1029/98JD02651, 1998.

Price, C. and Rind, D.: A simple lightning parameterization for calculating global lightning distributions, J. Geophys. Res., 97, 9919–9933, doi:http://dx.doi.org/10.1029/92JD0071910.1029/92JD00719, 1992.

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.

Price, C., Penner, J., and Prather, M.: NO_x from lightning, 1. Global distribution based on lightning physics, J. Geophys. Res., 102, 5929–5941, doi:http://dx.doi.org/10.1029/96JD0350410.1029/96JD03504, 1997.

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:http://dx.doi.org/10.1029/96JD0170610.1029/96JD01706, 1996.

Schumann, U. and Huntrieser, H.: The global lightning-induced nitrogen oxides source, Atmos. Chem. Phys., 7, 3823–3907, doi:http://dx.doi.org/10.5194/acp-7-3823-200710.5194/acp-7-3823-2007, 2007.

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.

Sauvage, B., Martin, R. V., van Donkelaar, A., Liu, X., Chance, K., Jaeglú, 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.

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.

Stavrakou, T., Müller, J.-F., Boersma, K. F., van der A, R. J., Kurokawa, J., Ohara, T., and Zhang, Q.: Key chemical NO_x sink uncertainties and how they influence top-down emissions of nitrogen oxides, Atmos. Chem. Phys., 13, 9057–9082, doi:http://dx.doi.org/10.5194/acp-13-9057-201310.5194/acp-13-9057-2013, 2013.

Stroppiana, D., Brivio, P. A., Gregoire, J.-M., Liousse, C., Guillaume, B., Granier, C., Mieville, A., Chin, M., and Petron, 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.

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:http://dx.doi.org/10.1029/2001JD00111410.1029/2001JD001114, 2002.

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:http://dx.doi.org/10.1029/94JD0124410.1029/94JD01244, 1994.

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:http://dx.doi.org/10.1029/2005JD00704210.1029/2005JD007042, 2007.

Tost, H., Jöckel, P., and Lelieveld, J.: Lightning and convection parameterisations – uncertainties in global modelling, Atmos. Chem. Phys., 7, 4553–4568, doi:http://dx.doi.org/10.5194/acp-7-4553-200710.5194/acp-7-4553-2007, 2007.

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:http://dx.doi.org/10.5194/acp-10-11707-201010.5194/acp-10-11707-2010, 2010.

Whitaker, J. and Hamill, T.: Ensemble data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913–1924, 2002.

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.

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:http://dx.doi.org/10.1029/2004JD00452210.1029/2004JD004522, 2004.

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:http://dx.doi.org/10.1029/2006GL02780610.1029/2006GL027806, 2007.

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:http://dx.doi.org/10.1029/2008JD01056010.1029/2008JD010560, 2009.

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:http://dx.doi.org/10.1029/2006JD00708910.1029/2006JD007089, 2006.

Table 1. The annual and seasonal total flash rate (in $\mathrm{flashes\,s^{-1}}$) 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° N), the tropics (TR, 20° S–20° N), the Southern Hemisphere (SH, 90–20° S), and the globe (GL, 90° S–90° N) for 2007 and for four seasons of the year: December–February (DJF), March–May (MAM), June–August (JJA), and September–November (SON).

	LIS/OTD					Model				
	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON
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

Table 2. The annual total $\mathrm{LNO_x}$ sources (in $\mathrm{Tg\,N\,yr^{-1}}$) obtained from the CTM simulation and the data assimilation for the Northern Hemisphere (NH, 20–90° N), the tropics (TR, 20° S–20° N), the Southern Hemisphere (SH, 90–20° S), and the globe (GL, 90° S–90° N). The analysis uncertainty estimated from the mean analysis spread is shown in brackets.

	NH	TR	SH	GL
CTM	1.4	2.7	0.6	4.7
Assimilation	2.0 (±0.3)	3.5 (±0.49)	0.8 (±0.10)	6.3 (±0.9)

Table 3. The regional averages of the mean altitude (in km) with maximum annual $\mathrm{LNO_x}$ 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). The definitions of the regions are same as for Fig. 7, except for the tropical western Pacific (130–165° E, 1–18° N). Grid points without any $\mathrm{LNO_x}$ sources or with peak height for pressures higher than $850\,\mathrm{hPa}$ are removed from the average, in order to measure the upper tropospheric peak height.

	CTM	Assim	Increment
North America	7.82	7.85	+0.03
Europe	5.72	5.58	−0.14
Northern Eurasia	9.39	9.39	±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

Table 4. The monthly LNO_x sources (in Tg N) for the Northern Hemisphere (NH, 20–90° N), the tropics (TR, 20° S–20° N), the Southern Hemisphere (SH, 90–20° S), and the globe (GL, 90° S—90° N) as obtained from the CTM simulation, the OSEs with TES O_3, OMI NO_2, MLS O_3, and MLS HNO_3 observations, and from the assimilation of all the datasets. Also shown are results for data assimilation that optimizes a height-independent LNO_x source scale factor (2-D). Standard deviations obtained from all the data assimilation estimates are also listed.

	Jan				Jul			
	NH	TR	SH	GL	NH	TR	SH	GL
CTM	0.58	2.82	0.95	4.35	3.17	2.41	0.31	5.89
TES O_3	0.64	2.34	1.01	4.00	3.21	3.21	0.34	6.76
OMI NO_2	0.78	3.25	1.20	5.22	3.99	3.51	0.51	8.01
MLS O_3	0.74	4.89	1.21	6.84	4.82	4.69	0.31	9.83
MLS 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
TES O_3 (2-D)	0.72	3.35	1.67	5.75	4.94	3.74	0.37	9.05
OMI NO_2 (2-D)	0.78	3.29	1.12	5.18	2.86	3.54	0.49	6.89
MLS O_3 (2-D)	0.85	3.92	1.07	5.85	5.46	4.10	0.37	9.92
MLS 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
Standard dev.	0.07	0.77	0.28	0.96	0.89	0.56	0.09	1.23

Table 5. The spatial correlation (Corr), mean difference (Bias), and mean root-mean-square error (RMSE) of the three-monthly mean tropospheric 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 are shown for the latitude band 30° S–30° N for the globe (180° W–180° E), Africa (15° W–50° E), India (50° E–90° E), the maritime continent (90° E–140° E), Pacific (140° E–80° W), South America (80° W–40° W), and Atlantic (40° W–15° W).

		DJF			JJA		
		Corr	Bias	RMSE	Corr	Bias	RMSE
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
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
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
Maritime continent	CTM	0.88	1.23	14.26	0.96	1.49	8.38
	Assim.	0.87	0.00	5.78	0.94	−1.03	5.52
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
South America	CTM	0.80	3.36	15.83	0.61	1.18	13.98
	Assim.	0.75	−0.13	5.32	0.81	0.12	6.46
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

Table 6. Similar to Table 4, but lists the LNO_x sources obtained from the control data assimilation run (Control), with a 15 % addition of artificial OMI NO_2 bias (w/ OMI bias), with the TES 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 LNO_x source and the surface NO_x emissions (+20 % LNO_x a priori error and +20 % SNO_x a priori error), and with 15 % increases in the a priori values of the LNO_x sources (+15 % LNO_x a priori source). The total bias due to all terms is computed as a random addition of the individual biases. See the text for details.

	Jan				Jul			
	NH	TR	SH	GL	NH	TR	SH	GL
Control	0.78	3.99	1.39	6.15	4.69	2.99	0.50	8.18
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 % LNO_x a priori error	0.83	3.75	1.32	5.90	4.59	2.93	0.51	8.03
+20 % SNO_x a priori error	0.81	3.77	1.27	5.85	4.58	2.83	0.50	7.90
+15 % LNO_x a priori source	0.83	4.10	1.48	6.41	5.29	3.16	0.57	9.02
Total bias	0.16	0.47	0.20	0.66	1.06	0.38	0.31	1.58

Table 7. The global and regional total flash rate (FR, $\mathrm{flashes\,s^{-1}}$) estimated from the model parameterization (1st row) and the LIS/OTD climatological observations (2nd row), and the NO production efficiency (NO prod., $\mathrm{mol\,NO\,flash^{-1}}$) estimated from the total 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 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° S and 35° N are shown in brackets.

	FR [$\mathrm{flashes\,s^{-1}}$]		NO prod. [$\mathrm{mol\,NO\,flash^{-1}}$]		
	Model	LIS/OTD	Assim. w/ model FR	Assim. w/ LIS/OTD FR	Model
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
Land	32.2	33.9	388	368	294
Ocean	9.1	12.5	201	145	128
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

Table 8. A comparison of estimates of NO_x amount produced per flash ($mol\,NO\,flash^{-1}$), adapted from Peterson and Beasley (2011) and Koshak et al. (2014).

First author	Year	Methodology	Moles/flash
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

Table 9. The mean ozone concentration bias (in ppbv) between the CHASER simulations and the global ozonesonde observations for January 2007 in the NH (25° N–90° N) and for July 2007 in the tropics (TR, 25° S–25° N) and the SH (90° S–25° S). The CHASER simulation results using the a priori emissions sources (A priori), the LNO_x sources (LNO_x), and the LNO_x sources and surface NO_x emissions ($L+SNO_x$) are shown. The results from the CHASER-DAS simultaneous assimilation are also listed (DAS).

	NH in Jul				TR in Jan				SH in Jan			
	A priori	LNO_x	$L+SNO_x$	DAS	A priori	LNO_x	$L+SNO_x$	DAS	A priori	LNO_x	$L+SNO_x$	DAS
750–450 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 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 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

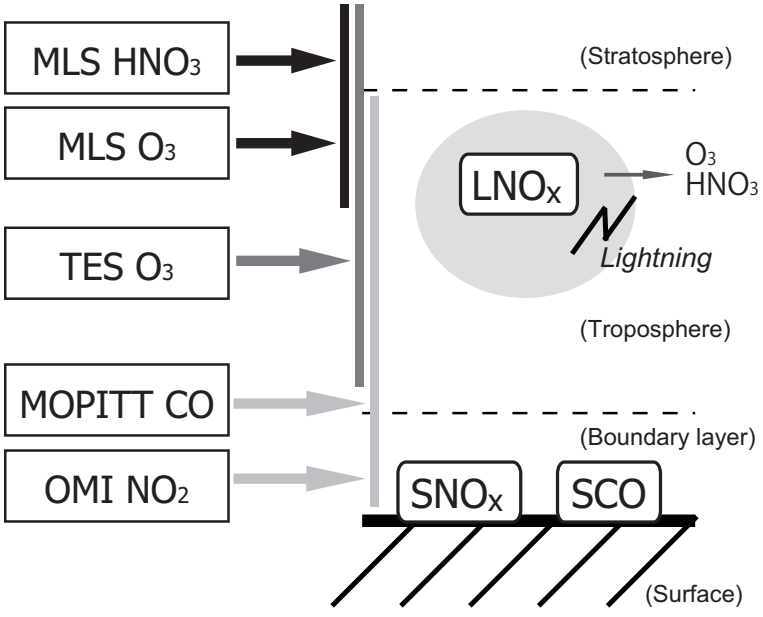

Fig. 1. Schematic diagram of the constraints on LNO_x brought by the different satellite retrieval products. The vertical bars indicate the vertical sensitivity range for the species observed. Through these different sensitivities the assimilation system extracts information about the total LNO_x source 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 LNO_x source strength, surface emissions as well as concentrations of the reactive gases involved.

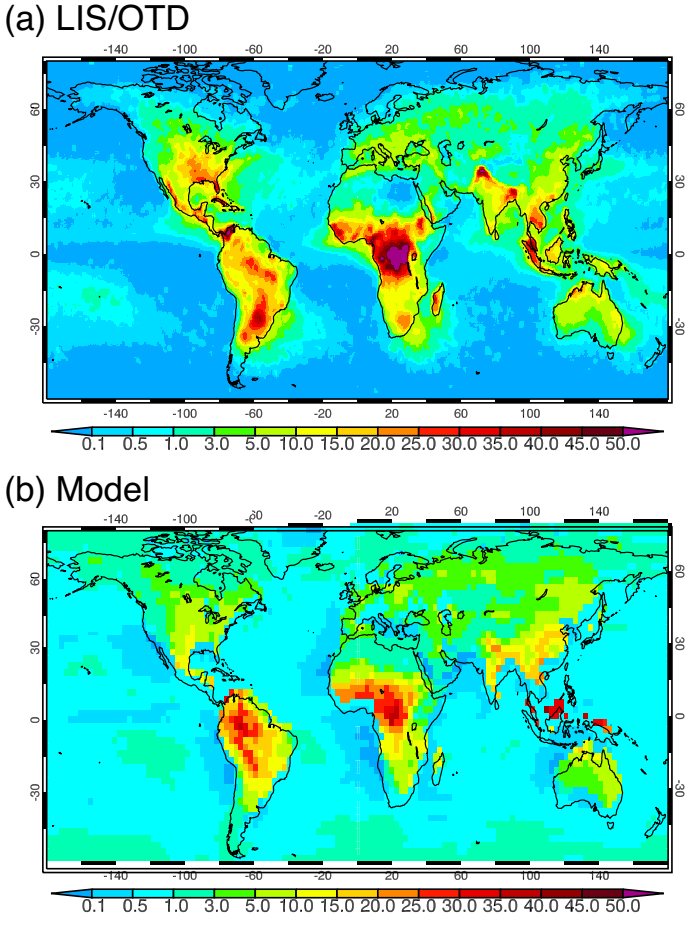

Fig. 2. Global distributions of the mean annual flash rate (in flashes km^{-2} yr^{-1}) estimated from (a) the LIS/OTD high–resolution monthly climatology (HRMC) data (Cecil et al., 2014) and (b) the model simulation for 2007.

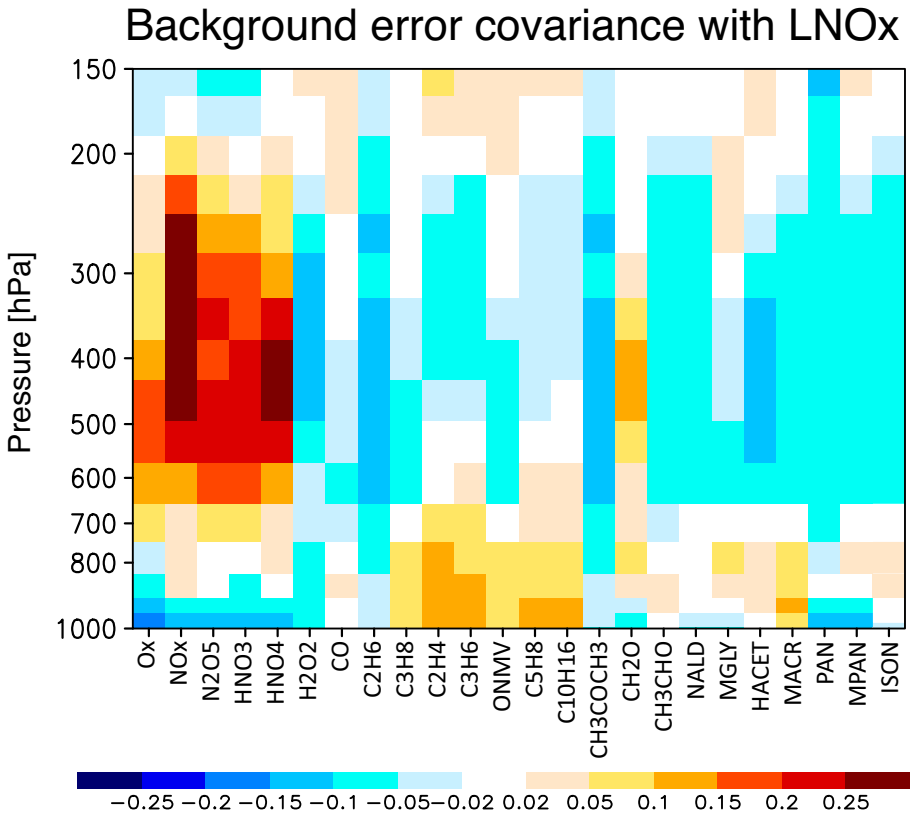

Fig. 3. Vertical profiles of correlations between the LNO_x sources and the concentrations of various chemical species as estimated from the background error covariance matrix based on CHASER ensemble simulations, averaged over central Africa for July 2007. The regional monthly mean of the covariance estimated for each grid point is plotted. The correlation is shown in red or blue where positive or negative, respectively.

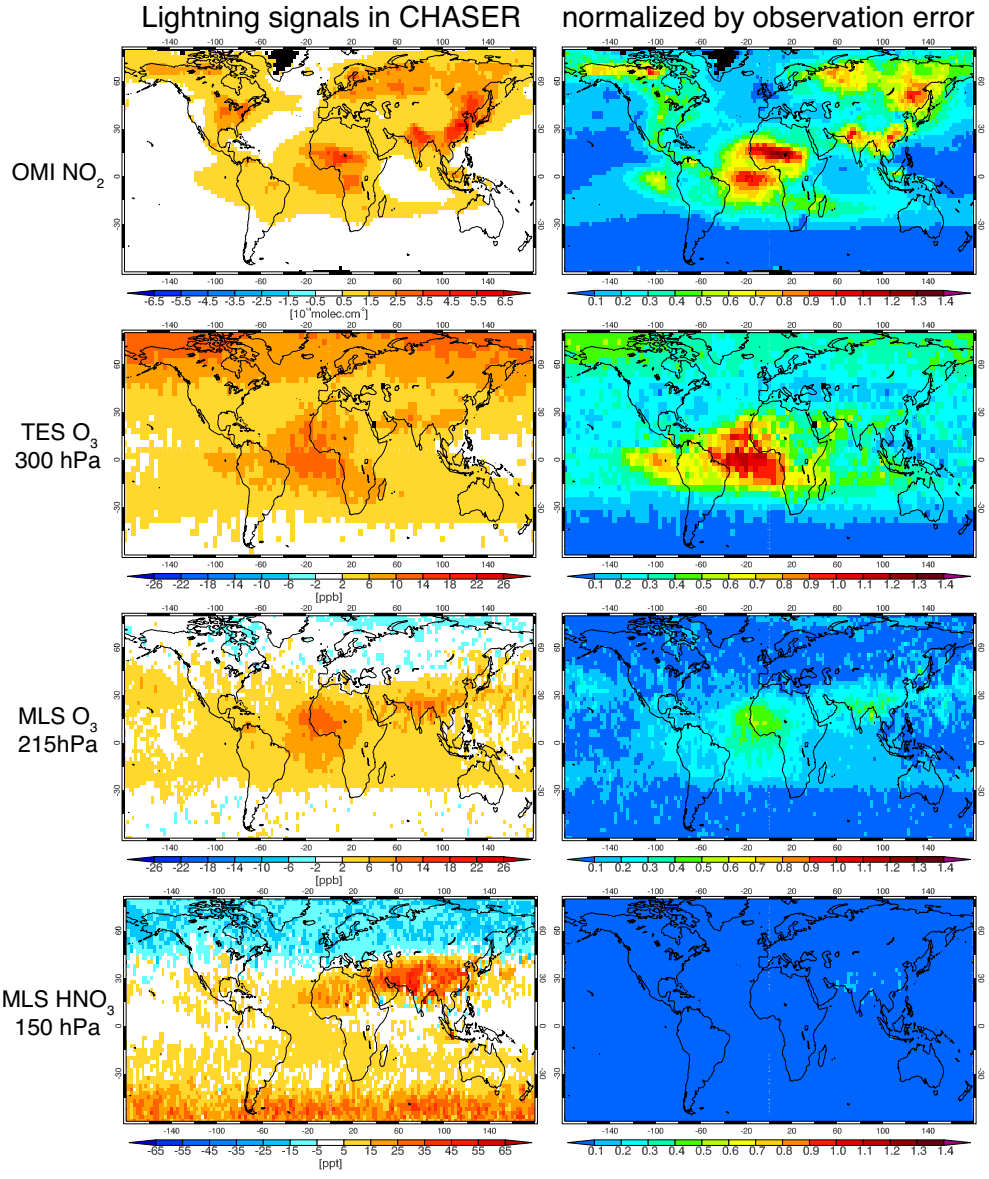

Fig. 4. Global maps of (left) the concentration differences between the CHASER simulations with and without lightning sources and (right) the mean ratio of the lightning signals to the measurement errors as estimated along each satellite track by applying the averaging kernels of OMI NO_2 (in 10^{14} molec cm^{-2}), TES O_3 (ppbv) at 300 hPa, MLS O_3 (ppbv) at 215 hPa, and MLS HNO_3 (pptv) at 150 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.

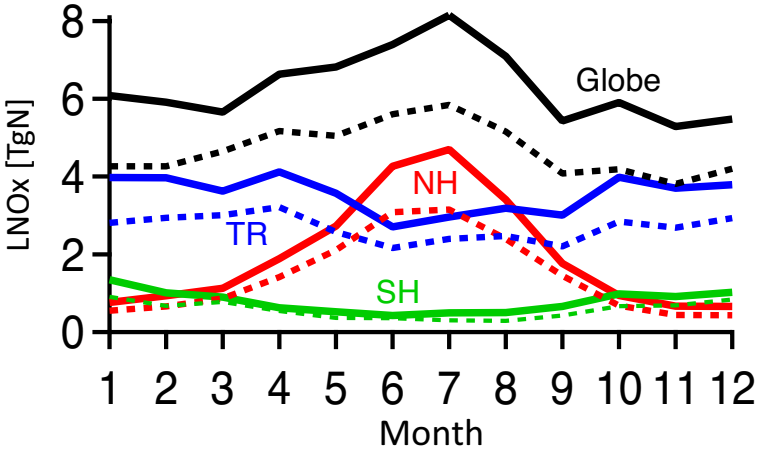

Fig. 5. Seasonal variations of the total LNO_x sources (in TgN) analysed from the data assimilation (solid lines) and estimated from the model simulation (dashed lines) over the globe (90° S–90° N), the Northern Hemisphere (NH, 20–90° N), the tropics (TR, 20° S–20° N), and the Southern Hemisphere (SH, 90–20° S) for 2007.

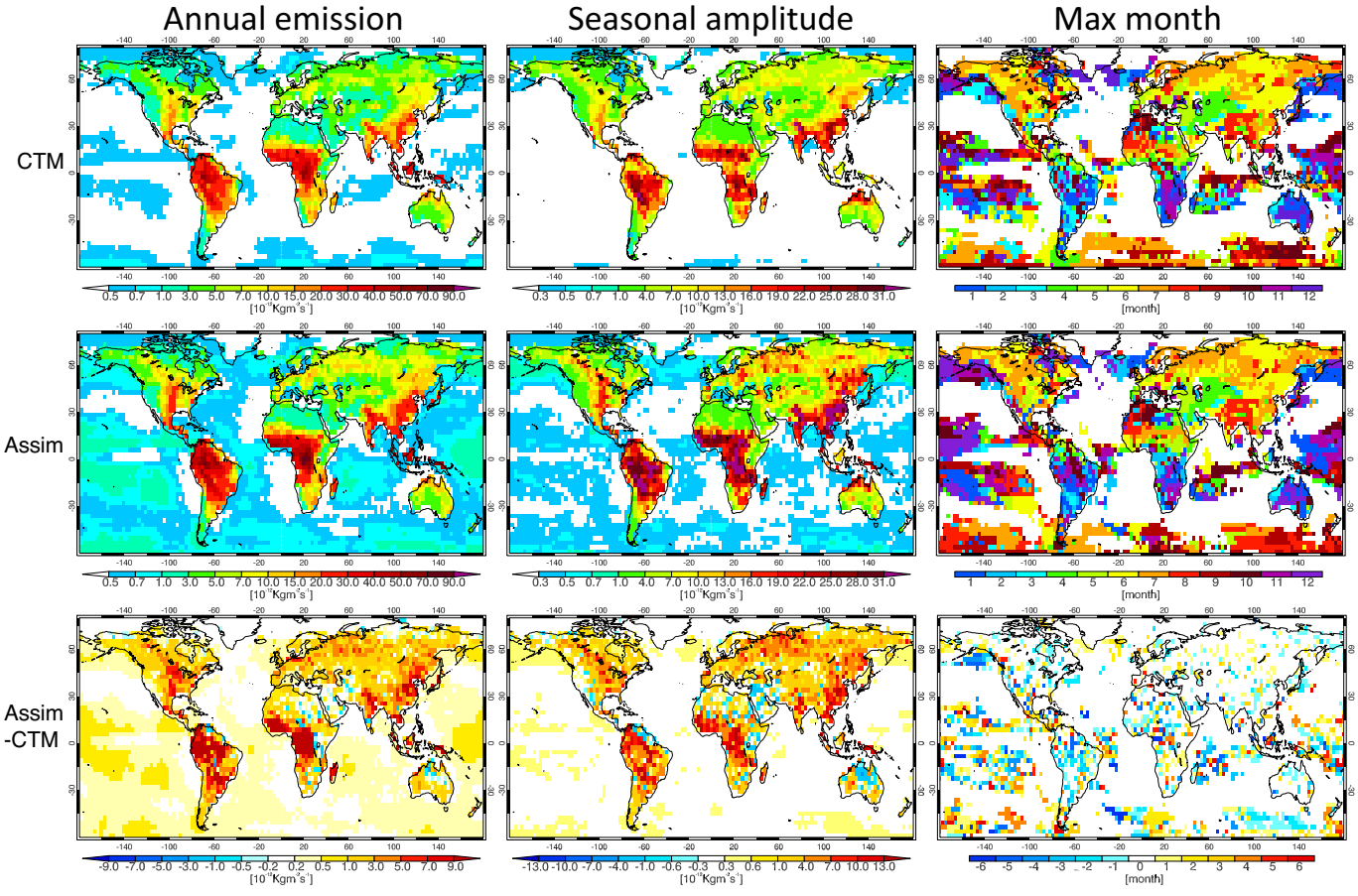

Fig. 6. Global distributions of the annual LNO_x source (left, $10^{-12}\,\mathrm{kg\,m^{-2}\,s^{-1}}$), its seasonal amplitude (centre, in $10^{-12}\,\mathrm{kg\,m^{-2}\,s^{-1}}$), and the timing of peak sources (right, in months) for 2007. Shown are the a priori sources estimated from the CTM parameterisation (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 LNO_x greater than $0.7 \times 10^{-13}\,\mathrm{kg\,m^{-2}\,s^{-1}}$.

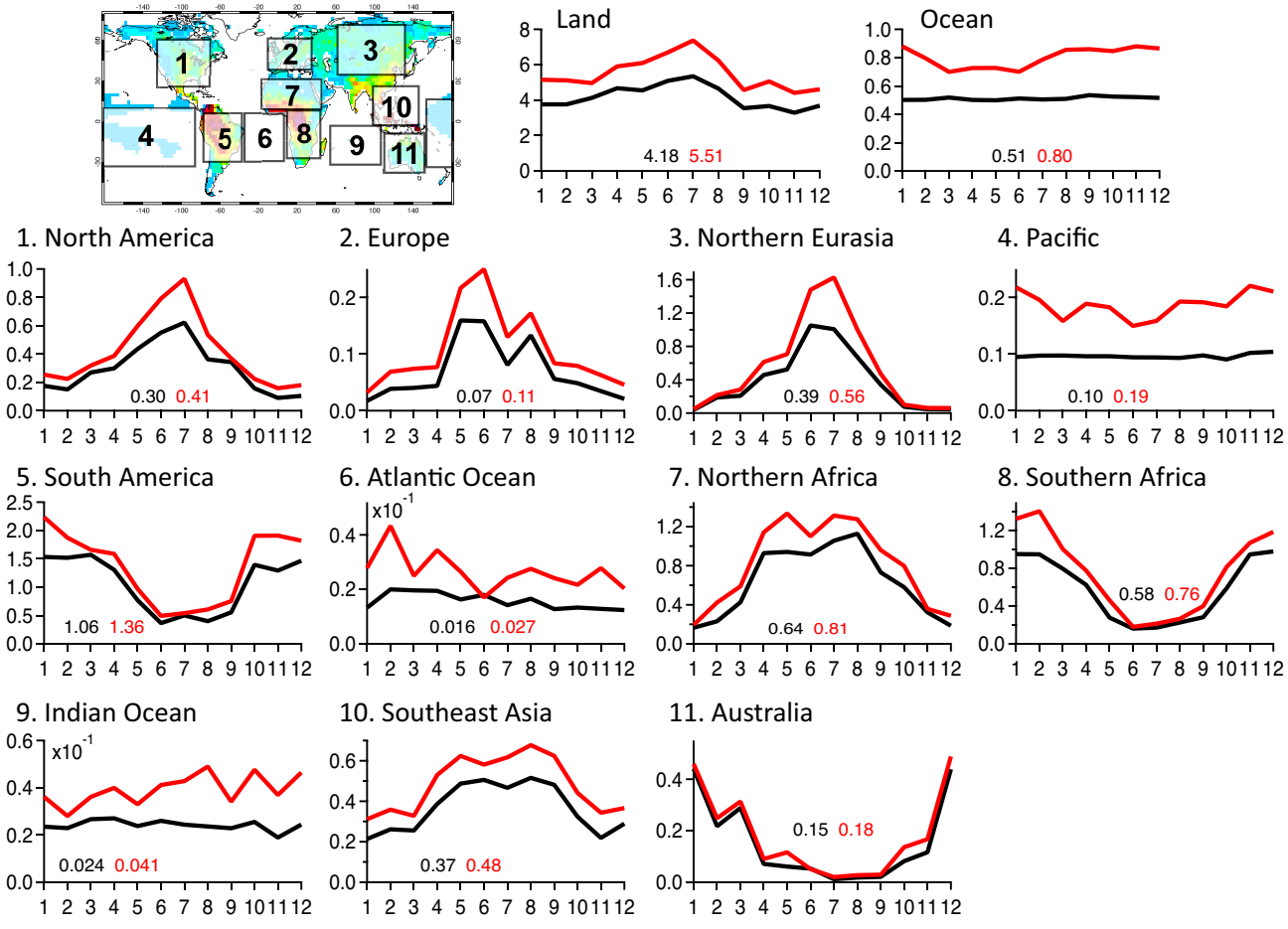

Fig. 7. Seasonal variations of the regional LNO_x sources (in $Tg\,N$) for (1) North America (120–65° W, 20–60° N), (2) Europe (10° W–30° E, 35–60° N), (3) northern Eurasia (60–130° E, 30–68° N), (4) the Pacific (154–180° E, 35° S–20° N and 180° E–88° W, 35° S–12° N), (5) South America (77–39° W, 35°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

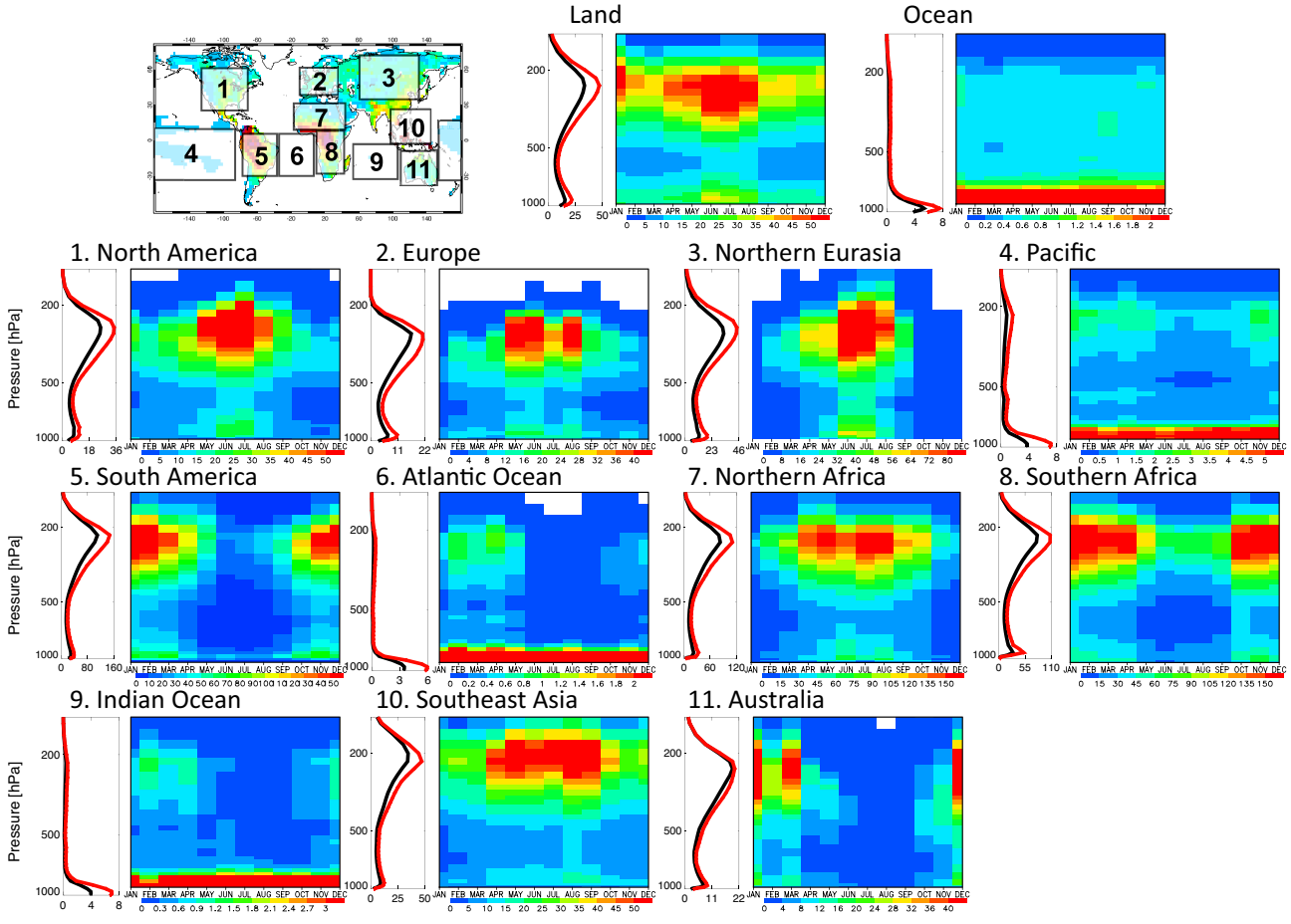

Fig. 8. Similar to Fig. 7, but shows for the vertical profile of the annual mean a priori (black) and a posteriori (red) LNO_x sources (left panels) and the seasonal variation of the monthly mean vertical profile of the a posteriori LNO_x source (right panels) in $\mathrm{pptv\,day^{-1}}$.

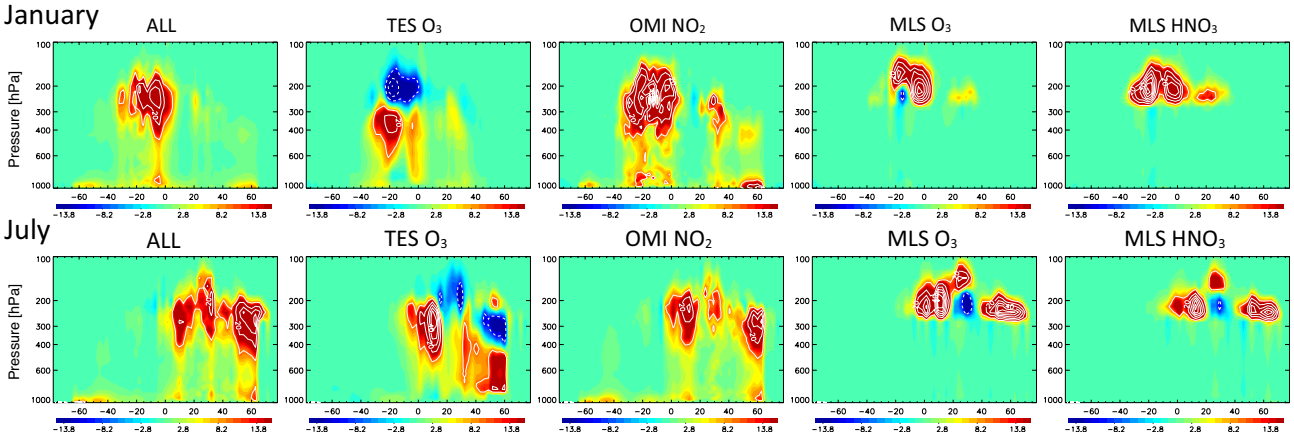

Fig. 9. Latitude-pressure cross-sections of the monthly mean analysis increment (the assimilation minus the CTM simulation) for the LNO_x source (in $\mathrm{pptv\,day^{-1}}$) obtained from assimilation of (left) all the data, (2nd from left) TES O_3 data, (centre) OMI NO_2 data, (2nd from right) MLS O_3 data, and (right) MLS HNO_3 data in (top) January 2007 and (bottom) July 2007. The interval of the contour lines is $10\,\mathrm{pptv\,day^{-1}}$.

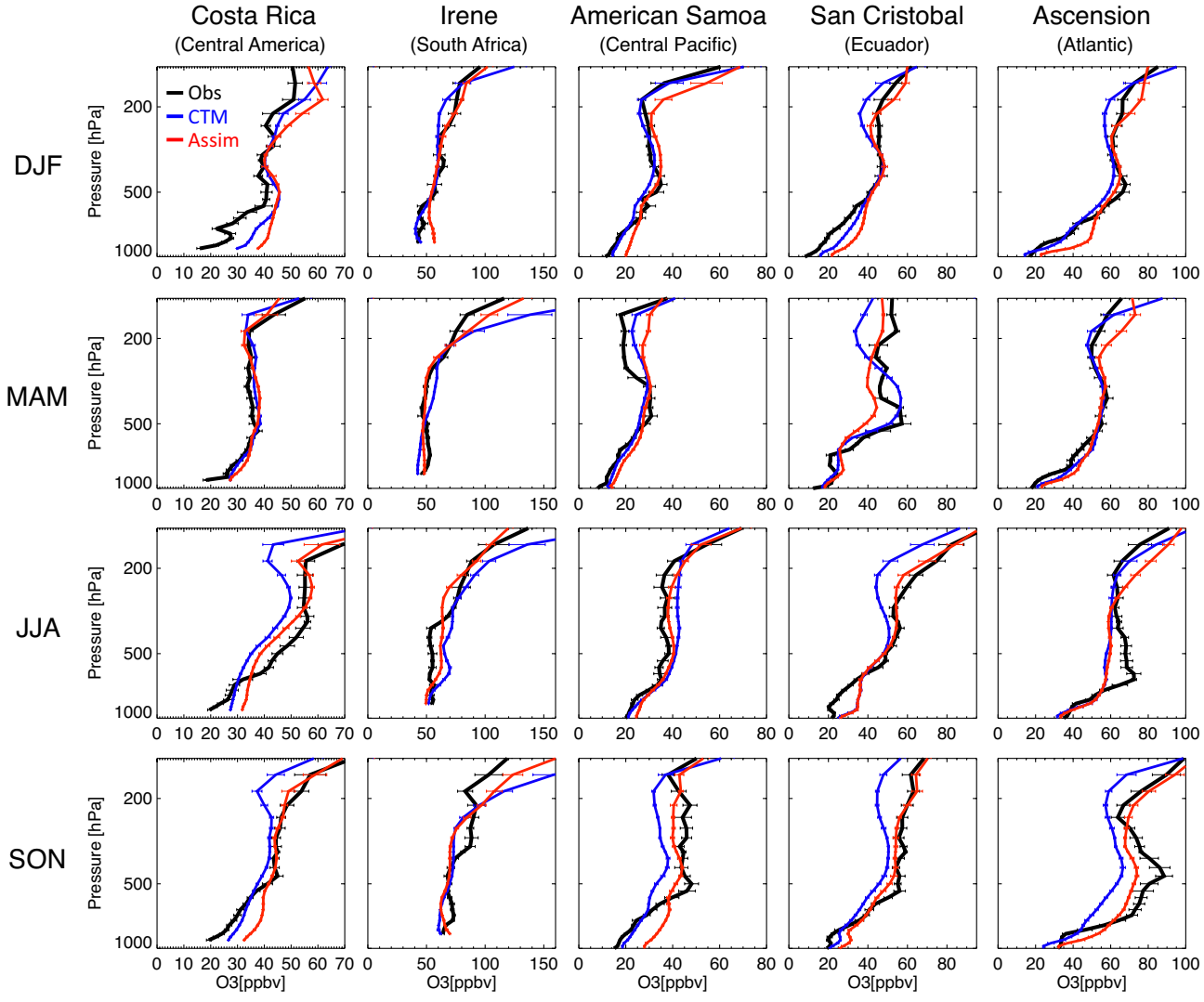

Fig. 10. Vertical O_3 profiles (in ppbv) obtained from ozonesondes (black), the CTM simulation (blue), and the data assimilation (red) for Costa Rica (left), Irene in South Africa (2nd from left), American Samoa (center), San Cristobal in Ecuador (2nd from right), and Ascension in the tropical Atlantic (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.

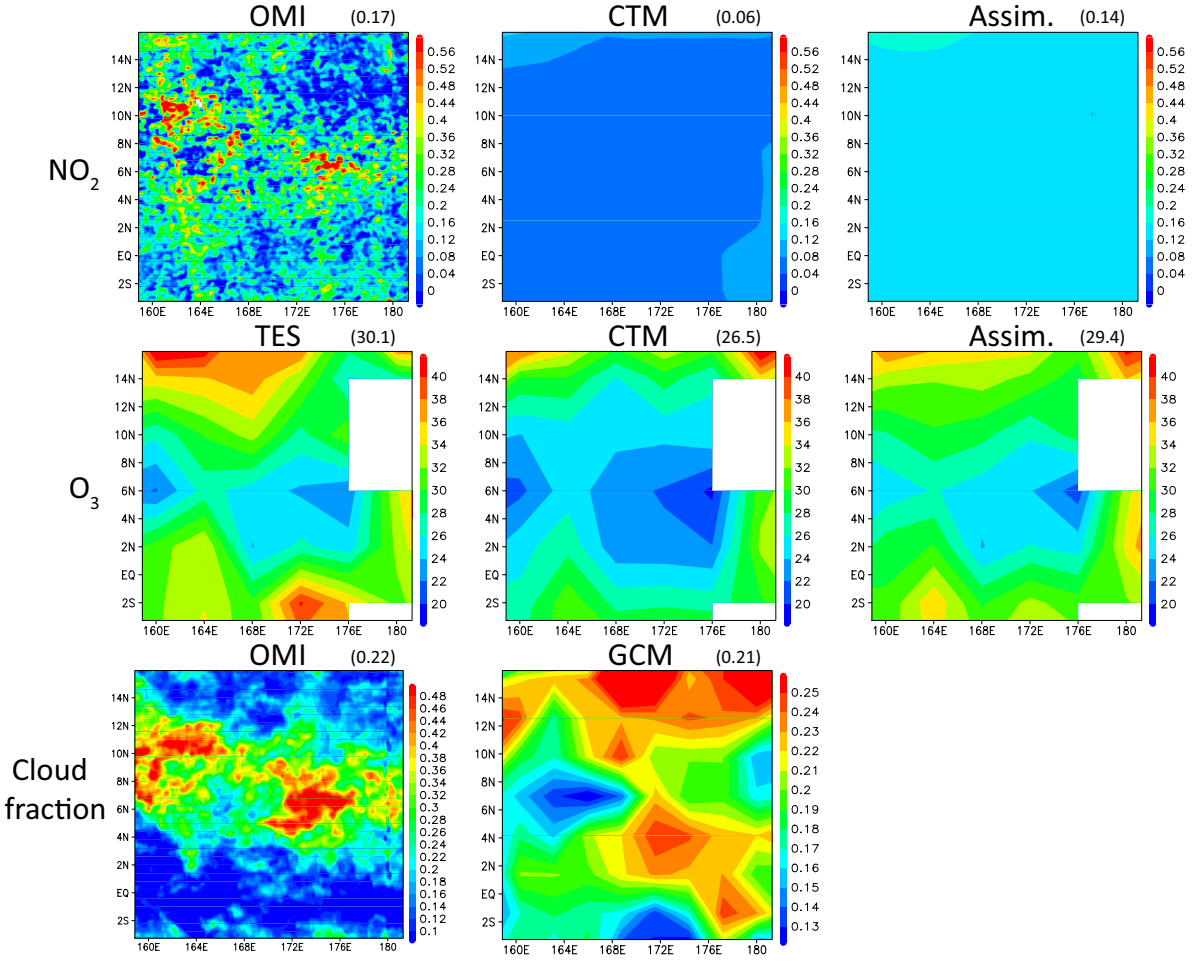

Fig. 11. Spatial distributions of tropospheric NO_2 column concentration (upper, 10^{15} molec cm^{-2}), O_3 concentration at 300 hPa (middle, in ppbv), and cloud fraction (lower) over the western Pacific. Each is averaged over 14–21 August 2007. For the NO_2 and O_3 concentrations, the results obtained 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.

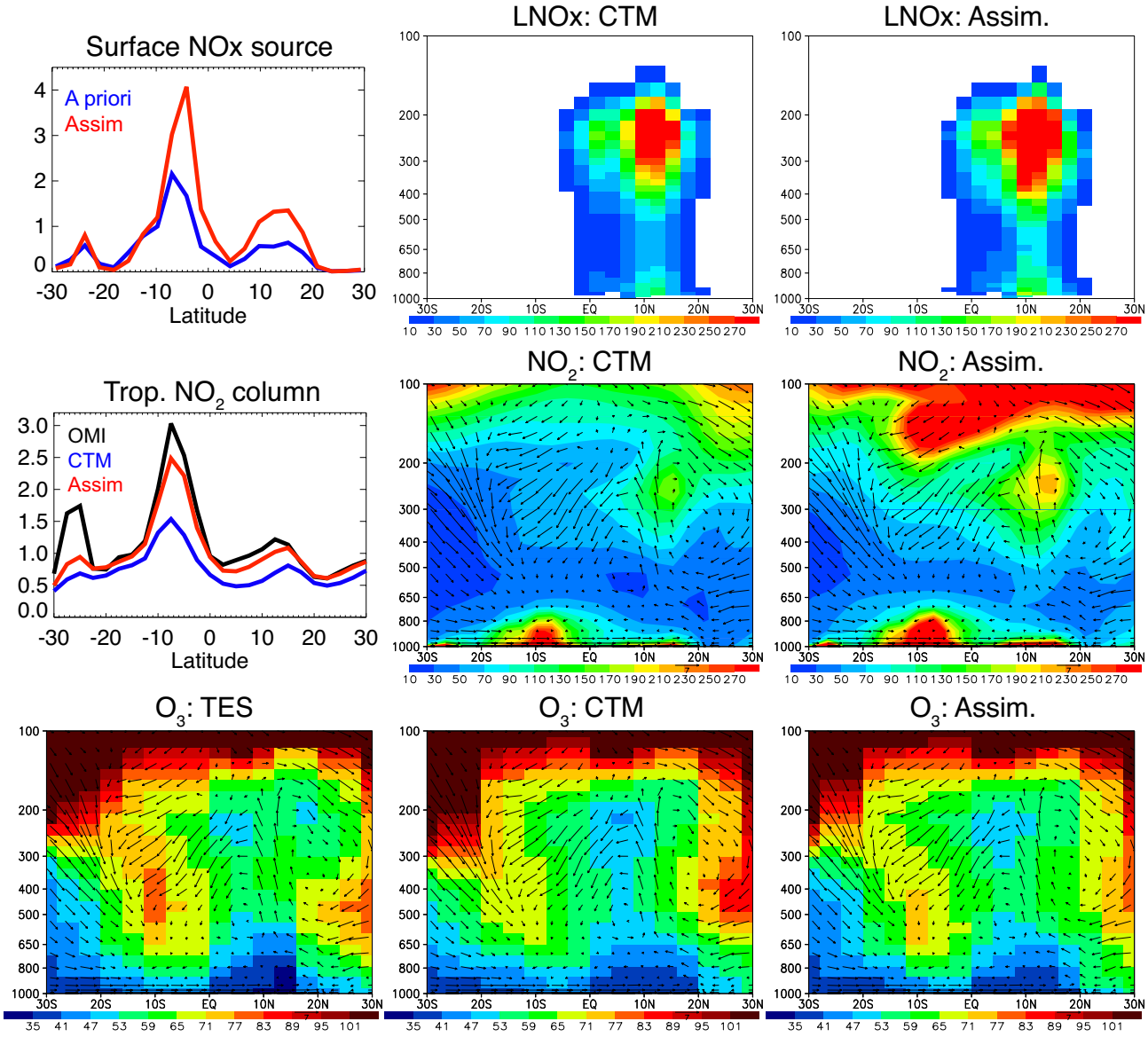

Fig. 12. Latitude-pressure cross-section of the longitudinal-mean (6° W–30° E) of the LNO_x source (upper, in $\mathrm{pptv\,day^{-1}}$), NO_2 concentration (middle, in pptv), and O_3 concentration (lower, ppbv) over Africa. The results obtained from the CTM simulation (centre) and analysed from the data assimilation (right) are presented. The lower left panel shows the result obtained from the TES measurement for O_3 concentration. Also shown are the latitudinal distributions of (upper left panel) the longitudinal-mean surface emissions of NO_x (in $10^{-11}\,\mathrm{kg\,m^{-2}\,s^{-1}}$) 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) and (centre left) the longitudinal-mean tropospheric NO_2 columns (in $10^{15}\,\mathrm{molec\,cm^{-2}}$) 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.

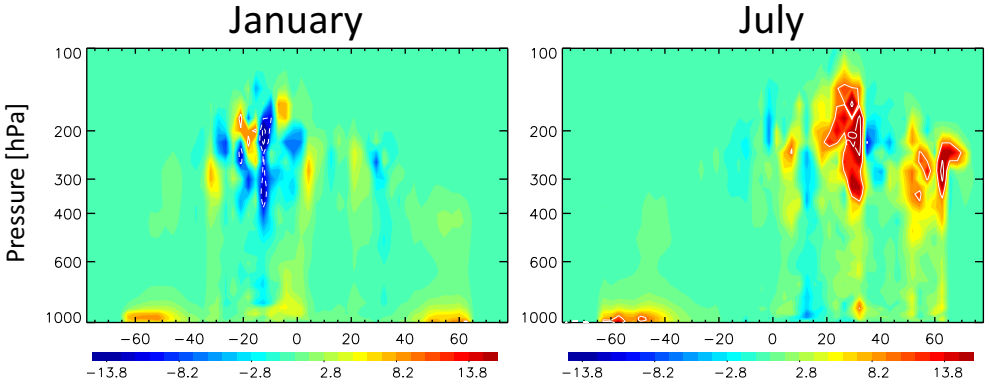

Fig. 13. Latitude-pressure cross-section of the LNO_x source differences between the data assimilations with and without (with minus without) the cloud-covered OMI NO_2 observations (in $\mathrm{pptv\,day^{-1}}$) for January and July in 2007. The increases and decreases in the source due to assimilation of the cloud-covered observations correspond to positive and negative values represented by red and blue, respectively.

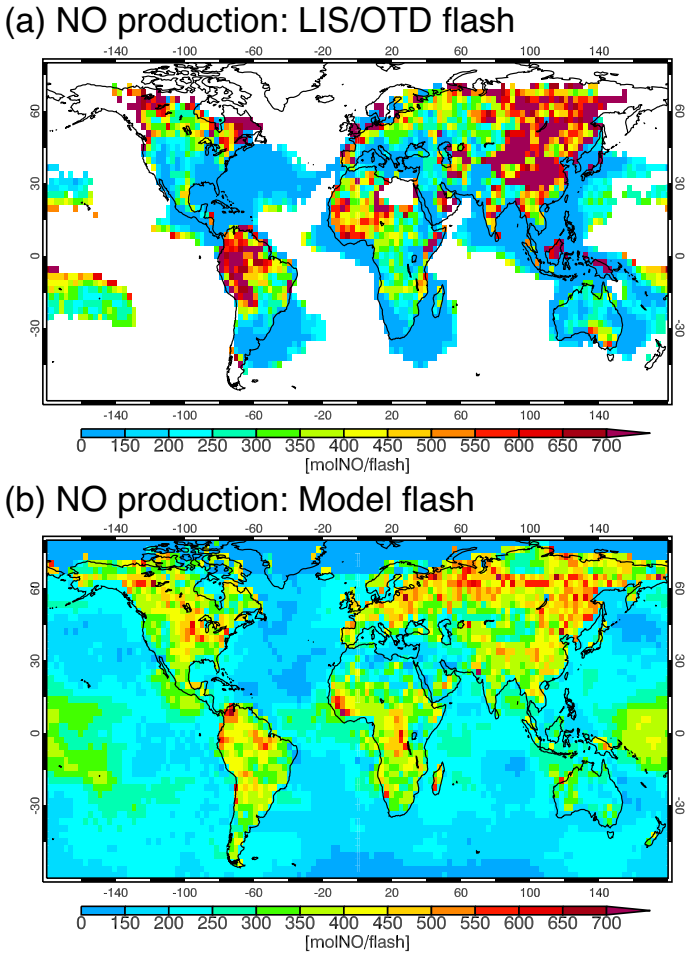

Fig. 14. Global distributions of the NO production efficiency ($\mathrm{mol\,NO\,flash^{-1}}$) estimated from the annual total $\mathrm{LNO_x}$ sources analysed from data assimilation for 2007 with (a) the LIS/OTD HRMC flash climatology data (Cecil et al., 2014), and with (b) the model flash rate for 2007. The results in (a) are shown for the region with the observed annual flash rates of greater than 0.44 flashes $\mathrm{km^{-2}\ yr^{-1}}$ to avoid unrealistically large estimates.