Global lightning NO\textsubscript{x} production estimated by an assimilation of multiple satellite data sets

K. Miyazaki\textsuperscript{1}, H. J. Eskes\textsuperscript{2}, K. Sudo\textsuperscript{3}, and C. Zhang\textsuperscript{4}

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

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

Received: 28 October 2013 – Published in Atmos. Chem. Phys. Discuss.: 7 November 2013
Revised: 10 February 2014 – Accepted: 12 February 2014 – Published: 2 April 2014

Abstract. The global source of lightning-produced NO\textsubscript{x} (LNO\textsubscript{x}) is estimated by assimilating observations of NO\textsubscript{2}, O\textsubscript{3}, HNO\textsubscript{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 data sets with different vertical sensitivity profiles provides comprehensive constraints on the global LNO\textsubscript{x} source while improving the representations of the entire chemical system affecting atmospheric NO\textsubscript{x}, including surface emissions and inflows from the stratosphere. The annual global LNO\textsubscript{x} source amount and NO production efficiency are estimated at 6.3 Tg N yr\textsuperscript{-1} and 310 mol NO flash\textsuperscript{-1}, respectively. Sensitivity studies with perturbed satellite data sets, model and data assimilation settings lead to an error estimate of about 1.4 Tg N yr\textsuperscript{-1} on this global LNO\textsubscript{x} source. These estimates are significantly different from those estimated from a parameter inversion that optimizes only the LNO\textsubscript{x} source from NO\textsubscript{2} observations alone, which may lead to an overestimate of the source adjustment. The total LNO\textsubscript{x} source is predominantly corrected by the assimilation of OMI NO\textsubscript{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 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 analyzed ozone fields and independent observations gives confidence in the performance of the LNO\textsubscript{x} source estimation.

1 Introduction

Lightning-produced NO\textsubscript{x} (LNO\textsubscript{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\textsubscript{x} accounts for only about 10–20 % of the global NO\textsubscript{x} sources, but is the most dominant source in the upper troposphere (e.g., Galloway et al., 2004). A small fraction of LNO\textsubscript{x} can lead to significant ozone production in the upper troposphere (Thompson et al., 1994) because the O\textsubscript{3} production efficiency per NO\textsubscript{x} molecule typically increases with height, owing to the longer lifetime of NO\textsubscript{x} and the highly non-linear dependence of ozone production on NO\textsubscript{x} (Pickering et al., 1998; Martin et al., 2000; Jenkins and Ryu, 2004). Therefore, accurate estimates of LNO\textsubscript{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\textsubscript{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\textsubscript{x} sources on the basis of...
assumptions regarding the NO$_x$ production efficiency per flash, energy ratio of cloud-to-ground (CG) flashes to intracloud (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 variability of lightning activity (e.g., Boccioppio, 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, 2005; Fehr et al., 2004; 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 4D-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 data sets in order to analyze 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: Sect. 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 data sets 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 UT) 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 (cf., 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 $0.5 \times 10^{15}$ 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 data set 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 LNO₃ source estimation through their influence on the oxidation capacity and the NO₃ chemistry.

2.2 Validation data

Independent ozone observations are used to validate the performance of the data assimilation. The spatial distribution of tropospheric O₃ 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° × 1.25° (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 O₃ 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 LNO₃ sources and surface emissions of NOₓ 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 NOₓ-CO-OH-O₃ set of chemical reactions.

Miyazaki and Eskes (2013) demonstrated that multi-species data assimilation improves the analysis of surface NOₓ emissions, in comparison with an inversion derived from NO₂ measurements alone. They showed that the assimilation of measurements for species other than NO₂ changes the regional estimates of monthly mean surface NOₓ by up to −58 % to +32 %. These large changes emphasize the 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 LNO₃ source estimation through corrections made to the concentrations of various chemical species. This is especially true for LNO₃ because all satellite sensors are sensitive in the altitude range where LNO₃ and the ozone produced by LNO₃ 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 NOₓ 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 Database (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 NOₓ emissions, depending on the dominant category for each emission category (Miyazaki et al., 2012b). The total NOₓ emission by aircraft is obtained from EDGAR as 0.55 Tg N yr⁻¹, which is similar to a more recent estimate of 0.49 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-V AR 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, \( x^b_i (i = 1, \ldots, k) \), is globally obtained from the evolution of each ensemble model realization, where \( x \) 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, \( y^b_i = H(x^b_i) \), is then estimated using the observation operator \( H \). The observation operator (\( H \)) is constructed on the basis of the spatial interpolation operator (\( S \)), the a priori profile (\( x_{\text{apriori}} \)) and the averaging kernel (\( A \)), which maps the model fields (\( x: N - (\text{the system dimension}) \) dimensional state vector) into observation space (\( y: p - (\text{the number of observation}) \) dimensional observational vector) while taking into account the vertical averaging implicit in the observations as follows:

\[
y = H(x) = x_{\text{apriori}} + A(S(x) - x_{\text{apriori}}). \tag{1}
\]

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

A background ensemble mean in the observation space, \( \bar{y}^b = \frac{1}{k} \sum_{i=1}^{k} y^b_i \), or in the model space, \( \bar{x}^b = \frac{1}{k} \sum_{i=1}^{k} x^b_i \), and an ensemble of background perturbations in the observation space, \( y^b = y^b - \bar{y}^b \), or in the model space, \( x^b = x^b - \bar{x}^b \), are also computed.

The ensemble mean analysis is then updated by

\[
\bar{x} = \bar{x}^b + X^b \bar{P}^a (Y^b) \bar{R}^{-1} (y^o - \bar{y}^b), \tag{2}
\]

where \( y^o \) is the observation vector, and \( R \) 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). \( \bar{P}^a \) is the local analysis error covariance in the ensemble space,

\[
\bar{P}^a = \left[ (k-1) I + (Y^b)^{T} \bar{R}^{-1} Y^b \right]^{-1}. \tag{3}
\]

The new analysis ensemble perturbation matrix in the model space \( X^a \) is obtained by transforming the background ensemble \( X^b \),

\[
X^a = X^b \left[ (k-1) \bar{P}^a \right]^{1/2}. \tag{4}
\]

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 LNOx sources, the covariances with TES O3, OMI NO2, MLS O3, and MLS HNO3 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 NOx and CO are optimized using OMI NO2 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 LNOx 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 LNOx 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): \( F = 3.44 \times 10^{-5} \times H^{1.92} \) (flashes min\(^{-1}\)) over continents; and \( F_{\text{in}} = 6.40 \times 10^{-4} \times H^{1.73} \) (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 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 Fig. 2 compare the global flash rate between the LIS/OTD high-resolution monthly climatology (HRMC) data (Cecil et al., 2014) and the model parameterization. Compared with the observations, the global distribution of the total flash rate is generally reproduced by the model. 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 (\( \Delta H \) for < 0 °C) on the basis of the following relation:

\[
z = 0.021 \times \Delta H^4 - 0.648 \times \Delta H^3 + 7.493 \times \Delta H^2 - 36.54 \times \Delta H + 63.09.
\]

This relationship is applied for clouds with \( \Delta H > 5.5 \) km, while \( z \) is set to zero for clouds with \( \Delta H < 5.5 \) 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 × 10\(^3\) 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) are averaged and applied in the parameterization. The global annual total amount of the LNO\(_x\) source for 2007 was estimated at 4.7 Tg N yr\(^{-1}\) in the CHASER simulations.
Table 1. The annual and seasonal total flash rate (in 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 the four seasons of the year: December–February (DJF), March–May (MAM), June–August (JJA), and September–November (SON).

<table>
<thead>
<tr>
<th>LIS/OTD</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Annual</td>
</tr>
<tr>
<td>NH</td>
<td>12.7</td>
</tr>
<tr>
<td>TR</td>
<td>27.1</td>
</tr>
<tr>
<td>SH</td>
<td>6.8</td>
</tr>
<tr>
<td>GL</td>
<td>46.5</td>
</tr>
</tbody>
</table>

3.2.2 Optimization by data assimilation

The multiplication factors for the LNOₓ production rate (mol s⁻¹) and the NOₓ 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 LNOₓ source and the concentrations of various species at different altitudes over central Africa in July. The concentrations of chemical species such as O₅, NOₓ, N₂O₅, HNO₃, and HNO₄ exhibit high positive correlations with the LNOₓ sources in the middle and upper troposphere. The high correlations indicate the utility of measurements of these species to constrain the LNOₓ 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 chemical concentrations optimize the multiplication factors for the total LNOₓ sources, and does not separately optimize for IG and CG flashes.

In order to provide meaningful constraints on LNOₓ, 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 NO₂ columns give the magnitude of the boreal summer (June–August) mean lightning signal as roughly 1–5 × 10¹⁴ molec cm⁻² over the tropical Atlantic, Africa, and India but roughly 1 × 10¹⁴ molec cm⁻² over remote land sites. The large enhancements are over, and downwind of, active convective regions (e.g., over the tropical Atlantic). 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. 3. Vertical profiles of correlations between the LNOₓ 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 (left panels) of the concentration differences between the CHASER simulations with and without lightning sources and (right panels) 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\textsubscript{2} (in 10\textsuperscript{14} molec cm\textsuperscript{-2}), TES O\textsubscript{3} (ppbv) at 300 hPa, MLS O\textsubscript{3} (ppbv) at 215 hPa, and MLS HNO\textsubscript{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.

estimated LNO\textsubscript{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\textsubscript{x} source analysis even over polluted regions. Moreover, the combined use of the multiple data sets with different vertical sensitivities is expected to facilitate the estimation of the vertical LNO\textsubscript{x} profile and to distinguish between the surface NO\textsubscript{x} emissions and LNO\textsubscript{x} sources. The MLS and TES data not only provide information on the NO\textsubscript{x} source amount in the middle and upper troposphere, but also constrain the lower tropospheric source when combined with the OMI tropospheric column NO\textsubscript{2} retrievals.

The a priori error is set to 40\% for the surface emissions of NO\textsubscript{x} and CO and 60\% for the LNO\textsubscript{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
4 Results

4.1 Global LNOx source distributions

Data assimilation increases the global annual LNOx source from 4.7 to 6.3 Tg N yr\(^{-1}\), as summarized in Table 2. The instantaneous uncertainty estimated from the mean analysis spread for the global source is \(\pm 0.9\) Tg N yr\(^{-1}\), while the chi-square \((\chi^2)\) diagnostic gives confidence in the estimated analysis spread as a measure of the analysis uncertainty (cf., Sect. 6.1.3). This spread of the ensemble is relatively small, showing that LNOx 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° S–90° S) contribute 31%, 56%, and 13% to the analyzed global source, respectively. These relative contributions are slightly modified from the a priori sources. The analyzed 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 LNOx 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 LNOx 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 (cf., Table 1).

4.2 Regional LNOx source distributions

Figure 7 shows the seasonal variations of the LNOx 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.

Table 2. The annual total LNOx sources (in 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° S–90° S), and the globe (GL, 90° S–90° N). The analysis uncertainty estimated from the mean analysis spread is shown in brackets.

<table>
<thead>
<tr>
<th>Region</th>
<th>CTM</th>
<th>Assimilation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td>1.4</td>
<td>2.0 (±0.3)</td>
</tr>
<tr>
<td>TR</td>
<td>2.7</td>
<td>3.5 (±0.49)</td>
</tr>
<tr>
<td>SH</td>
<td>0.6</td>
<td>0.8 (±0.10)</td>
</tr>
<tr>
<td>GL</td>
<td>4.7</td>
<td>6.3 (±0.9)</td>
</tr>
</tbody>
</table>

Fig. 5. Seasonal variations of the total LNOx sources (in Tg N) analyzed 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° S–90° S) for 2007.
Fig. 6. Global distributions of the annual LNO$_x$ source (left panels, $10^{-12}$ kg m$^{-2}$ s$^{-1}$), its seasonal amplitude (center panels, in $10^{-12}$ kg m$^{-2}$ s$^{-1}$), and the timing of peak sources (right panels, in months) for 2007. Shown are the a priori sources estimated from the CTM parameterization (upper panels), the a posteriori sources from the data assimilation (middle panels), and the analysis increment (lower panels). The analysis increment equals the a posteriori sources minus the a priori sources. The peak timing is estimated for regions with the analyzed annual sources of LNO$_x$ greater than $0.7 \times 10^{-13}$ kg m$^{-2}$ s$^{-1}$.

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 Sect. 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 South-east 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 data sets

The effects of individual assimilated data sets on the estimated LNO$_x$ sources were evaluated through observing system experiments (OSEs) by separately assimilating each data set 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
Table 3. The regional averages of the mean altitude (in km) with maximum annual 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 the same as for Fig. 7, except for the tropical western Pacific (130°–165° E, 1–18° N). Grid points without any LNO\(_x\) sources or with peak height for pressures higher than 850 hPa are removed from the average in order to measure the upper tropospheric peak height.

<table>
<thead>
<tr>
<th>Region</th>
<th>CTM</th>
<th>Assim</th>
<th>Increment</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>7.82</td>
<td>7.85</td>
<td>+0.03</td>
</tr>
<tr>
<td>Europe</td>
<td>5.72</td>
<td>5.58</td>
<td>−0.14</td>
</tr>
<tr>
<td>Northern Eurasia</td>
<td>9.39</td>
<td>9.39</td>
<td>±0</td>
</tr>
<tr>
<td>Tropical western Pacific</td>
<td>11.15</td>
<td>9.96</td>
<td>−1.19</td>
</tr>
<tr>
<td>South America</td>
<td>9.97</td>
<td>9.83</td>
<td>−0.14</td>
</tr>
<tr>
<td>Northern Africa</td>
<td>11.00</td>
<td>10.83</td>
<td>−0.17</td>
</tr>
<tr>
<td>Southern Africa</td>
<td>9.36</td>
<td>8.98</td>
<td>−0.38</td>
</tr>
<tr>
<td>Southeast Asia</td>
<td>11.51</td>
<td>10.77</td>
<td>−0.74</td>
</tr>
<tr>
<td>Australia</td>
<td>8.01</td>
<td>7.60</td>
<td>−0.41</td>
</tr>
</tbody>
</table>

The measurements of species other than NO\(_2\) also provide important constraints on the global source estimation. The assimilated data sets all have an impact on the global source analysis, but with different contributions from individual data sets, 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 data sets 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 arise 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\(_x\) species in the stratosphere, which has the potential to improve the modeled impact of stratospheric air on tropospheric concentrations and benefits the source estimate derived from tropospheric column. Furthermore, the assimilation of MO-

Table 4. The table provides regional mean concentrations (ppbv) of LNO\(_x\) observed by OSEs from January 2007 to July 2010, along with the PITT 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 data sets. 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 optimize 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 lead 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 analyze 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
Fig. 8. Similar to Fig. 7, but shows for the vertical profile of the annual mean a priori (black) and a posteriori (red) LNOx sources (left panels) and the seasonal variation of the monthly mean vertical profile of the a posteriori LNOx source (right panels) in pptvday⁻¹.

Fig. 9. Latitude-pressure cross-sections of the monthly mean analysis increment (the assimilation minus the CTM simulation) for the LNOx source (in pptv day⁻¹) obtained from assimilation of (left panels) all the data, (2nd from left panels) TES O₃ data, (center panels) OMI NO₂ data, (2nd from right panels) MLS O₃ data, and (right panels) MLS HNO₃ data in (top row) January 2007 and (bottom row) July 2007. The interval of the contour lines is 10 pptv day⁻¹.
the O₃ concentration, which influence the simultaneous optimization in the data assimilation. Figure 10 shows a comparison of modeled and analyzed vertical O₃ profiles with the ozonesonde observations taken at five tropical SHADOZ sites. The assimilation removes most of the O₃ 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 O₃ concentrations over the tropical Atlantic. Conversely, the assimilation does not obviously improve the lower tropospheric O₃. 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 = 0.85–0.92). The CTM simulation overestimates the TOC in the northern sub tropics, 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 O₃ (Miyazaki et al., 2012a). Because lightning substantially influences the amount of O₃ in the tropics, and because the data assimilation simultaneously optimizes the O₃ and the LNOₓ source, significantly improved agreement with independent ozone observations gives confidence in the performance of the LNOₓ estimates.

### 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 LNOₓ sources are expected over the tropical western Pacific because of errors in the tropical Pacific ITCZ cumulus clouds simulated by the AGCM (Emori et al., 2005).

Figure 11 shows the regional distribution of total cloud fraction and tropospheric concentrations of NO₂ and O₃ 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
Fig. 10. Vertical O₃ profiles (in ppbv) obtained from ozonesondes (black), the CTM simulation (blue), and the data assimilation (red) for Costa Rica (left panels), Irene in South Africa (2nd from left panels), American Samoa (center panels), San Cristobal in Ecuador (2nd from right panels), and Ascension in the tropical Atlantic (right panels) during December–February (DJF, top row), March–May (MAM, 2nd from top row), June–August (JJA, 2nd from bottom row), and September–November (SON, bottom row) in 2007. The error bars represent the standard deviation of all the data within one bin.

Parameterization causes a large uncertainty in the simulated cloud and LNOₓ source structures. Accurate simulations of the cloud position are important to properly distribute the LNOₓ sources, while errors in the simulated cloud top height lead to uncertainties in the total source strength. The regional mean strength of the LNOₓ 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 analyzed NO₂ and O₃ concentrations show better agreements with the observations (Fig. 11) because of the simultaneous data assimilation. The simulation underestimates the regional mean NO₂ concentration in the upper troposphere by 65% compared to the OMI NO₂ retrieval, whereas data assimilation removes most of the bias. Agreements with TES O₃ fields are also greatly improved by data assimilation, despite the sparse coverage of the TES measurements. Because of the simultaneous optimization of the LNOₓ sources and these chemical concentrations, the improvements suggest that the regional total LNOₓ 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.
Table 5. The spatial correlation (corr), mean difference (bias), and mean root-mean-square error (RMSE) of the three-monthly mean tropospheric O\textsubscript{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\degree S–30\degree N for the globe (180\degree W–180\degree E), Africa (15\degree W–50\degree E), India (50–90\degree E), the maritime continent (90–140\degree E), Pacific (140\degree E–80\degree W), South America (80–40\degree W), and Atlantic (40–15\degree W).

<table>
<thead>
<tr>
<th></th>
<th>DJF</th>
<th></th>
<th></th>
<th>JJA</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Corr</td>
<td>Bias</td>
<td>RMSE</td>
<td>Corr</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>Globe</td>
<td>CTM</td>
<td>0.85</td>
<td>1.92</td>
<td>4.16</td>
<td>0.92</td>
<td>1.41</td>
</tr>
<tr>
<td></td>
<td>Assim.</td>
<td>0.86</td>
<td>−0.55</td>
<td>2.85</td>
<td>0.92</td>
<td>0.19</td>
</tr>
<tr>
<td>Africa</td>
<td>CTM</td>
<td>0.65</td>
<td>3.26</td>
<td>22.22</td>
<td>0.88</td>
<td>3.54</td>
</tr>
<tr>
<td></td>
<td>Assim.</td>
<td>0.75</td>
<td>0.78</td>
<td>6.96</td>
<td>0.84</td>
<td>2.92</td>
</tr>
<tr>
<td>India</td>
<td>CTM</td>
<td>0.83</td>
<td>2.15</td>
<td>24.36</td>
<td>0.96</td>
<td>2.91</td>
</tr>
<tr>
<td></td>
<td>Assim.</td>
<td>0.84</td>
<td>0.25</td>
<td>8.67</td>
<td>0.95</td>
<td>1.03</td>
</tr>
<tr>
<td>Maritime</td>
<td>CTM</td>
<td>0.88</td>
<td>1.23</td>
<td>14.26</td>
<td>0.96</td>
<td>1.49</td>
</tr>
<tr>
<td></td>
<td>Assim.</td>
<td>0.87</td>
<td>0.00</td>
<td>5.78</td>
<td>0.94</td>
<td>−1.03</td>
</tr>
<tr>
<td>Pacific</td>
<td>CTM</td>
<td>0.89</td>
<td>0.27</td>
<td>9.83</td>
<td>0.96</td>
<td>−0.15</td>
</tr>
<tr>
<td></td>
<td>Assim.</td>
<td>0.89</td>
<td>−1.35</td>
<td>8.24</td>
<td>0.97</td>
<td>−1.67</td>
</tr>
<tr>
<td>South America</td>
<td>CTM</td>
<td>0.80</td>
<td>3.36</td>
<td>15.83</td>
<td>0.61</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>Assim.</td>
<td>0.75</td>
<td>−0.13</td>
<td>5.32</td>
<td>0.81</td>
<td>0.12</td>
</tr>
<tr>
<td>Atlantic</td>
<td>CTM</td>
<td>0.01</td>
<td>4.10</td>
<td>25.64</td>
<td>0.74</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>Assim.</td>
<td>0.32</td>
<td>0.62</td>
<td>5.42</td>
<td>0.83</td>
<td>1.56</td>
</tr>
</tbody>
</table>

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\textsubscript{x} sources and chemical concentrations. NO\textsubscript{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\textsubscript{2} concentrations are maximal over northern (around 5–20\degree N) and central (15\degree S–Equator) Africa, owing to in situ emissions from biomass burning. In the upper troposphere, strong LNO\textsubscript{x} sources cause a maximum NO\textsubscript{2} concentration over northern Africa (5–15\degree N). Data assimilation increases the surface NO\textsubscript{x} emissions over northern and central Africa by up to 90\% and the LNO\textsubscript{x} source over northern Africa by about 50\%, which acts to strengthen the local maximum in NO\textsubscript{2} concentrations.

The tropospheric O\textsubscript{3} distributions also show distinct variations over central Africa. The surface NO\textsubscript{2} and CO emissions and LNO\textsubscript{x} sources lead to O\textsubscript{3} production, whereas the inflows from the stratosphere along the subtropical jet stream predominantly determine the latitudinal O\textsubscript{3} variations in the upper troposphere. The multi-species data assimilation provides comprehensive constraints on these processes and the O\textsubscript{3} variations. The assimilation of TES data modifies the O\textsubscript{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\textsubscript{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\textsubscript{x} source estimate

6.1.1 Systematic satellite observation errors

The quality of the assimilated measurements largely influences the LNO\textsubscript{x} source uncertainty. Boersma et al. (2011) showed that the different OMI NO\textsubscript{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\textsubscript{2} retrievals. This increased the monthly regional and global LNO\textsubscript{x} sources by up to 14\% and 3\%, respectively, as summarized in Table 6. It is emphasized that low NO\textsubscript{2} concentrations over the oceans are mostly smaller than the OMI noise level. Errors related to the separation of stratospheric and tropospheric NO\textsubscript{2} could also cause errors in the OMI tropospheric NO\textsubscript{2} column retrievals (Lamsal et
Fig. 11. Spatial distributions of tropospheric NO\textsubscript{2} column concentration (upper panels, 10\textsuperscript{15} molec cm\textsuperscript{-2}), O\textsubscript{3} concentration at 300 hPa (middle panels, in ppbv), and cloud fraction (lower panels) over the western Pacific. Each is averaged over 14–21 August 2007. For the NO\textsubscript{2} and O\textsubscript{3} concentrations, the results obtained from the OMI and TES satellite retrievals (left panels), the CTM simulation (center panels), and the data assimilation (right panels) are shown. For the cloud fraction, the results obtained from the OMI retrievals (left) and the GCM simulation (right) are shown. The numbers in brackets represent the regional mean value for each plot.

K. Miyazaki et al.: Lightning NO\textsubscript{x} production estimation

(2010; Boersma et al., 2011). These may cause large uncertainties in the analyzed LNO\textsubscript{x} sources, especially over the oceans.

Any bias in the measurements of species other than NO\textsubscript{2} also affects the quality of the sources estimated in the simultaneous assimilation framework. TES O\textsubscript{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\textsubscript{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\textsubscript{3} increases the estimated monthly global LNO\textsubscript{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\textsubscript{x} reaches the top and anvil of...
Surface NOx source

LNOx: CTM

LNOx: Assim.

Trop. NOx column

NO2: CTM

NO2: Assim.

O3: TES

O3: CTM

O3: Assim.

Fig. 12. Latitude-pressure cross-section of the longitudinal-mean (6°W–30°E) of the LNOx source (upper, in pptv day\(^{-1}\)), NO\(_2\) concentration (middle, in pptv), and O\(_3\) concentration (lower, ppbv) over Africa. The results obtained from the CTM simulation (center) and analyzed 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\(^{-11}\) 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 analyzed from the data assimilation (red line) and (center left panel) the longitudinal-mean tropospheric NO\(_2\) columns (in 10\(^{15}\) molec cm\(^{-2}\)) as obtained from the OMI measurements (black line), the CTM simulation (blue line), and analyzed from the data assimilation (red line). The vectors represent meridional-vertical winds. Each is averaged over 10–20 July 2007.

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\textsubscript{x} source is increased by 12 % in July when cloud-covered observations are used. These results imply that cloud-covered OMI NO\textsubscript{2} data contain important information on the LNO\textsubscript{x} source amounts above and inside clouds. However, errors in the simulated cloud profiles (cf., Fig. 11) and in the retrieved cloud-altitude dependent averaging kernels may cause biases in the cloudy scene analysis.

The lifetimes of NO\textsubscript{x}, HNO\textsubscript{3}, and O\textsubscript{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\textsubscript{x} source far from the convective clouds that produced the LNO\textsubscript{x}. The large spread of the different estimates indicates that estimates of the LNO\textsubscript{x} source distribution and the global LNO\textsubscript{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\textsubscript{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\textsubscript{x} source estimation (e.g., Jourdain et al., 2010). As suggested in Sect. 4.4, the use of multiple data sets causes the analysis increment of the LNO\textsubscript{x} sources to become smaller because several sources of error are corrected simultaneously. In a parameter inversion that optimizes the LNO\textsubscript{x} sources only, the analysis increments will compensate for errors that occur not only in the LNO\textsubscript{x} sources but also in the other sources and species concentrations. This will result in overcorrections in the LNO\textsubscript{x} source estimation. Despite this advantage of the simultaneous assimilation of multiple satellite data sets, 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\textsubscript{x} chemical lifetime in the troposphere depends on altitude (i.e., longer at higher altitude), cumulus convection affects the total amount of NO\textsubscript{x} in the troposphere and, accordingly, NO\textsubscript{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\textsubscript{x} sources (Table 6). The high sensitivity of the LNO\textsubscript{x} source estimation to model convection is commonly reported by Lin (2012).

The model also shows systematic errors in the location of cumulus clouds (cf., Fig. 11). An accurate determination of cloud positions is important for analyzing the LNO\textsubscript{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\textsubscript{x} estimates (Table 6). The impact of changing the SST data was different for different regions; e.g., the LNO\textsubscript{x} sources over the Pacific increased by 14 % in January.

---

Table 6. Similar to Table 4, but lists the LNO\textsubscript{x} sources obtained from the control data assimilation run (control), with a 15 % addition of artificial OMI NO\textsubscript{2} bias (w/ OMI bias), with the TES O\textsubscript{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\textsubscript{x} source and the surface NO\textsubscript{x} emissions (+20 % LNO\textsubscript{x} a priori error and +20 % SNO\textsubscript{x} a priori error), and with 15 % increases in the a priori values of the LNO\textsubscript{x} sources (+15 % LNO\textsubscript{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.

<table>
<thead>
<tr>
<th>Jan</th>
<th>Jul</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NH</td>
</tr>
<tr>
<td>Control</td>
<td>0.78</td>
</tr>
<tr>
<td>w/ OMI bias</td>
<td>0.87</td>
</tr>
<tr>
<td>TES bias correction</td>
<td>0.68</td>
</tr>
<tr>
<td>w/o cloud OMI</td>
<td>0.76</td>
</tr>
<tr>
<td>year 1997 SST</td>
<td>0.76</td>
</tr>
<tr>
<td>+20 % convection</td>
<td>0.80</td>
</tr>
<tr>
<td>+20 % LNO\textsubscript{x} a priori error</td>
<td>0.83</td>
</tr>
<tr>
<td>+20 % SNO\textsubscript{x} a priori error</td>
<td>0.81</td>
</tr>
<tr>
<td>+15 % LNO\textsubscript{x} a priori source</td>
<td>0.83</td>
</tr>
<tr>
<td>Total bias</td>
<td>0.16</td>
</tr>
</tbody>
</table>
This indicates that the uncertainty in the LNO\textsubscript{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., determines the NO\textsubscript{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\textsubscript{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, Stavrou et al. (2013) demonstrated the strong effect of NO\textsubscript{x} loss uncertainties on top-down NO\textsubscript{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\textsubscript{x} source estimate. Although the aircraft NO\textsubscript{x} emissions likely have relatively small uncertainties (e.g., Wilkerson et al., 2010), the LNO\textsubscript{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 modeling

$\chi^2$ diagnostic tests (e.g., Menard and Chang, 2000) are used to measure the data assimilation statistics. The $\chi^2$ is estimated from the ratio of the differences between the model forecast and observations to the estimated background covariances. This measures whether the background covariance matrix producing realistic errors. The $\chi^2$ ratio becomes 1 if the background error covariance matches the model–observation differences. The annual mean $\chi^2$ was about 1.2 for OMI NO\textsubscript{2}, 0.8 for TES O\textsubscript{3}, 0.7 for MOPITT CO, 2.1 for MLS O\textsubscript{3}, and 1.6 for MLS HNO\textsubscript{3}, demonstrating that the overall magnitude of the background error covariances are reasonably modeled in the data assimilation. The too large $\chi^2$ for the MLS data imply an overconfidence in the model, and is largely attributed to the prescribed concentrations for O\textsubscript{3}, NO\textsubscript{x}, HNO\textsubscript{3}, and N\textsubscript{2}O\textsubscript{5} above 20 km altitude in CHASER.

The mean analysis spread, as estimated by transforming the background ensemble in the data assimilation (cf., Eq. 4), is about 0.9 TgN yr\textsuperscript{-1} for the annual global source strength. This can be translated into an error estimate for the total yearly source. The LNO\textsubscript{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\textsubscript{x} estimations are uncorrelated, then the impact of the analysis spread on the uncertainty in the total yearly LNO\textsubscript{x} can be estimated as $0.9/\sqrt{24}$ TgN yr\textsuperscript{-1}, or about 0.2 TgN yr\textsuperscript{-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\textsubscript{x} by 20% changes the global estimate of LNO\textsubscript{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\textsubscript{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\textsubscript{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\textsubscript{x} sources more efficiently. A sensitivity experiment in which the a priori global total LNO\textsubscript{x} source is increased by 15% demonstrates that the estimated LNO\textsubscript{x} source amount is influenced by the a priori source setting (Table 6); the global a posteriori LNO\textsubscript{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 optimized 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\textsubscript{3}, and O\textsubscript{3} are generally longer than a day, and long-range transport of LNO\textsubscript{x} can occur. A larger localization length may be useful when satellite measurements detect aged LNO\textsubscript{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\textsubscript{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 TgN yr\textsuperscript{-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 (summarized in Table 6), several other sources of error may be introduced. In particular, Stavroukou et al. (2013) claimed that the uncertainty in the loss processes is very large for NO\textsubscript{x}, resulting in a factor of 1.8 difference (3.3–5.9 TgN yr\textsuperscript{-1}) between upper and lower estimates of the LNO\textsubscript{x} source. This would translate into an error bar of about 1 TgN yr\textsuperscript{-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 (= $\sqrt{1^2+1^2}$) TgN yr\textsuperscript{-1}. 

www.atmos-chem-phys.net/14/3277/2014/ Atmos. Chem. Phys., 14, 3277–3305, 2014
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). Boccioppio (2002) also pointed out inconsistencies between the scheme of Price and Rind (1992) and satellite observations over the oceans. The modeling 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 Tg N yr$^{-1}$ is obtained from assimilation but cannot simply be explained by a roughly 4–9 % (7–12 % minus 3 %) underestimation of the global lightning flash frequency as compared to the 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 (cf., 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 modeling 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 parameterized 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 parameterized flash rate is relatively small, as summarized 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 summarized 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 modeling 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$_x$ data and estimated a global LNO$_x$ source strength of 1.1–6.4 Tg N 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 parameterized 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 parameterized flash rate and the LIS/OTD observations, respectively, as summarized in Table 7 and shown in Fig. 14. The detailed spatial structures in the production efficiencies estimated from the analyzed LNO$_x$ sources and the observed and the parameterized flash rates (Fig. 14) may reflect not only variations in flash characteristics but also noises and errors in the assimilated and flash measurements (cf., Sect. 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 analyzed...
Table 7. The global and regional total flash rate (FR, 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., mol NO flash$^{-1}$) estimated from the total LNO$_x$ sources analyzed 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.

<table>
<thead>
<tr>
<th>Region</th>
<th>FR [flashes s$^{-1}$]</th>
<th>NO prod. [mol NO flash$^{-1}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model LIS/OTD</td>
<td>Assim. w/ model FR LIS/OTD FR</td>
</tr>
<tr>
<td>NH</td>
<td>12.7</td>
<td>12.6</td>
</tr>
<tr>
<td>TR</td>
<td>21.3</td>
<td>27.1 (26.2)</td>
</tr>
<tr>
<td>SH</td>
<td>7.2</td>
<td>6.8</td>
</tr>
<tr>
<td>GL</td>
<td>41.2</td>
<td>46.5</td>
</tr>
<tr>
<td></td>
<td>Land</td>
<td>32.2</td>
</tr>
<tr>
<td></td>
<td>Ocean</td>
<td>9.1</td>
</tr>
<tr>
<td>North America</td>
<td>2.4</td>
<td>4.9</td>
</tr>
<tr>
<td>Europe</td>
<td>0.7</td>
<td>0.9</td>
</tr>
<tr>
<td>Northern Eurasia</td>
<td>2.8</td>
<td>2.2</td>
</tr>
<tr>
<td>Pacific</td>
<td>1.6</td>
<td>0.9 (0.9)</td>
</tr>
<tr>
<td>South America</td>
<td>7.9</td>
<td>8.1 (7.8)</td>
</tr>
<tr>
<td>Atlantic Ocean</td>
<td>0.3</td>
<td>0.3 (0.3)</td>
</tr>
<tr>
<td>Northern Africa</td>
<td>5.0</td>
<td>6.2 (6.0)</td>
</tr>
<tr>
<td>Southern Africa</td>
<td>4.3</td>
<td>7.7 (7.5)</td>
</tr>
<tr>
<td>Indian Ocean</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Southeast Asia</td>
<td>3.0</td>
<td>4.5 (3.9)</td>
</tr>
<tr>
<td>Australia</td>
<td>1.5</td>
<td>2.0</td>
</tr>
</tbody>
</table>

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 parameterized IC/CG flash ratio (cf., Eq. 5) or in the assumptions on the production efficiency of IC and CG flashes (cf., Sect. 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 parameterization. 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 Koshak 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 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 LNO$_x$ 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 analyzed profiles. The robustness of the analyzed 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 hPa in January, and sources smaller by 14% at 200 hPa in July in the tropics (20°S–20°N) for grid points with the LNO$_x$ source column greater than $5 \times 10^{-15}$ 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 (cf., 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 data sets 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 data sets 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 ± 3 Tg N yr$^{-1}$ for the annual global LNO$_x$ source. Our estimate of 6.3 ± 1.4 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 ± 0.5 Tg N yr$^{-1}$ and 3.3–5.9 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 mol flash$^{-1}$ based on the LIS/OTD climatological observations. This value is also within most of the recent estimates (cf., 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 (cf., Table 7). Detailed comparisons on monthly and regional scales including those seasonal and interannual 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 data sets 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 Tg N yr$^{-1}$ and 310 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 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 analyzed 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 modeling 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 analyzed LNO$_x$ sources have important implications for improving LNO$_x$ parameterizations. 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 parameterization 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 data sets is expected to provide further insights into such detailed processes in future studies with measurements that are more advanced (i.e., higher in accuracy, density, and vertical resolution).

Acknowledgements. We would like to thank M. Takigawa, H. Yashiro, K. Hamilton, Y. Wang, and F. 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.

Edited by: D. Shindell

References


Table 9. The mean ozone concentration bias (in ppbv) between the CHASER simulations and the global ozonesonde observations for January 2007 in the NH (25–90°N) and for July 2007 in the tropics (TR, 25–25°N) and the SH (90–25°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).

<table>
<thead>
<tr>
<th>A priori</th>
<th>LNO$_x$</th>
<th>L + SNO$_x$</th>
<th>DAS</th>
<th>A priori</th>
<th>LNO$_x$</th>
<th>L + SNO$_x$</th>
<th>DAS</th>
<th>A priori</th>
<th>LNO$_x$</th>
<th>L + SNO$_x$</th>
<th>DAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH in Jul</td>
<td>750–450 hPa</td>
<td>−12.3</td>
<td>−11.7</td>
<td>−0.2</td>
<td>−1.8</td>
<td>18.5</td>
<td>20.2</td>
<td>16.6</td>
<td>16.4</td>
<td>−4.1</td>
<td>−2.0</td>
</tr>
<tr>
<td></td>
<td>450–200 hPa</td>
<td>−6.8</td>
<td>−5.9</td>
<td>0.7</td>
<td>1.3</td>
<td>8.9</td>
<td>9.5</td>
<td>3.3</td>
<td>3.3</td>
<td>9.9</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>200–90 hPa</td>
<td>19.8</td>
<td>19.7</td>
<td>4.8</td>
<td>4.5</td>
<td>42.2</td>
<td>34.9</td>
<td>21.7</td>
<td>10.4</td>
<td>21.9</td>
<td>136.2</td>
</tr>
<tr>
<td>TR in Jan</td>
<td>750–450 hPa</td>
<td>−12.3</td>
<td>−11.7</td>
<td>−0.2</td>
<td>−1.8</td>
<td>18.5</td>
<td>20.2</td>
<td>16.6</td>
<td>16.4</td>
<td>−4.1</td>
<td>−2.0</td>
</tr>
<tr>
<td></td>
<td>450–200 hPa</td>
<td>−6.8</td>
<td>−5.9</td>
<td>0.7</td>
<td>1.3</td>
<td>8.9</td>
<td>9.5</td>
<td>3.3</td>
<td>3.3</td>
<td>9.9</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>200–90 hPa</td>
<td>19.8</td>
<td>19.7</td>
<td>4.8</td>
<td>4.5</td>
<td>42.2</td>
<td>34.9</td>
<td>21.7</td>
<td>10.4</td>
<td>21.9</td>
<td>136.2</td>
</tr>
<tr>
<td>SH in Jan</td>
<td>750–450 hPa</td>
<td>−12.3</td>
<td>−11.7</td>
<td>−0.2</td>
<td>−1.8</td>
<td>18.5</td>
<td>20.2</td>
<td>16.6</td>
<td>16.4</td>
<td>−4.1</td>
<td>−2.0</td>
</tr>
<tr>
<td></td>
<td>450–200 hPa</td>
<td>−6.8</td>
<td>−5.9</td>
<td>0.7</td>
<td>1.3</td>
<td>8.9</td>
<td>9.5</td>
<td>3.3</td>
<td>3.3</td>
<td>9.9</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>200–90 hPa</td>
<td>19.8</td>
<td>19.7</td>
<td>4.8</td>
<td>4.5</td>
<td>42.2</td>
<td>34.9</td>
<td>21.7</td>
<td>10.4</td>
<td>21.9</td>
<td>136.2</td>
</tr>
</tbody>
</table>


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


