



Simultaneous assimilation of satellite NO₂, O₃, CO, and HNO₃ data for the analysis of tropospheric chemical composition and emissions

K. Miyazaki^{1,2}, H. J. Eskes¹, K. Sudo^{2,3}, M. Takigawa², M. van Weele¹, and K. F. Boersma^{1,4}

¹Royal Netherlands Meteorological Institute (KNMI), Wilhelminalaan 10, 3732 GK, De Bilt, The Netherlands

²Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

³Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan

⁴Eindhoven University of Technology, Fluid Dynamics Lab, Eindhoven, The Netherlands

Correspondence to: K. Miyazaki (miyazaki@knmi.nl)

Received: 22 May 2012 – Published in Atmos. Chem. Phys. Discuss.: 2 July 2012

Revised: 3 October 2012 – Accepted: 4 October 2012 – Published: 22 October 2012

Abstract. We have developed an advanced chemical data assimilation system to combine observations of chemical compounds from multiple satellites. NO₂, O₃, CO, and HNO₃ measurements from the Ozone Monitoring Instrument (OMI), Tropospheric Emission Spectrometer (TES), Measurement of Pollution in the Troposphere (MOPITT), and Microwave Limb Sounder (MLS) satellite instruments are assimilated into the global chemical transport model CHASER for the years 2006–2007. The CHASER data assimilation system (CHASER-DAS), based on the local ensemble transform Kalman filter technique, simultaneously optimizes the chemical species, as well as the emissions of O₃ precursors, while taking their chemical feedbacks into account. With the available datasets, an improved description of the chemical feedbacks can be obtained, especially related to the NO_x-CO-OH-O₃ set of chemical reactions. Comparisons against independent satellite, aircraft, and ozonesonde data show that the data assimilation results in substantial improvements for various chemical compounds. These improvements include a reduced negative tropospheric NO₂ column bias (by 40–85%), a reduced negative CO bias in the Northern Hemisphere (by 40–90%), and a reduced positive O₃ bias in the middle and upper troposphere (from 30–40% to within 10%). These changes are related to increased tropospheric OH concentrations by 5–15% in the tropics and the Southern Hemisphere in July. Observing System Experiments (OSEs) have been conducted to quantify the relative importance of each data set on constraining the emissions and concentrations. The OSEs confirm that the assimilation of individual data sets results in a strong influence on both assimilated and

non-assimilated species through the inter-species error correlation and the chemical coupling described by the model. The simultaneous adjustment of the emissions and concentrations is a powerful approach to correcting the tropospheric ozone budget and profile analyses.

1 Introduction

Tropospheric ozone (O₃) is an important chemical species for air quality and climate (IPCC, 2007). It is an atmospheric pollutant in the lower troposphere and an effective greenhouse gas in the upper troposphere. Surface emissions of carbon monoxide (CO) and nitrogen oxides (NO_x) play an important role in determining tropospheric O₃ abundances. CO is an important precursor of tropospheric O₃ under high NO_x conditions. The concentration of CO is strongly related to the oxidising capacity of the atmosphere since it reacts primarily with OH (e.g. Logan et al., 1981; Daniel and Solomon, 1998; Thompson, 1992). In the middle and upper troposphere, O₃ can be generated efficiently through lightning NO_x sources (e.g. Pickering et al., 1998; Jenkins and Ryu, 2004; Martin et al., 2007). The abundance of tropospheric CO and NO₂ influences the atmospheric lifetime of the important greenhouse gases, methane (CH₄), O₃ (Shindell et al., 2009), and also CO₂ (Folberth et al., 2005).

NO_x and CO have both anthropogenic and natural sources. Anthropogenic sources include fossil fuel and biofuel combustion. Natural sources include biomass burning, soil, and also lightning emissions for NO_x. CO is produced from

the oxidation of hydrocarbons by the incomplete combustion of fossil fuels and biofuels, and during biomass burning events (Holloway et al., 2000). Knowledge about variations in surface emissions is important, but currently available bottom-up emissions inventories have large uncertainties (e.g., Jaeglé et al., 2005; Zhao et al., 2011). These inventories use statistical data, which generally have coarse resolution and large uncertainties. The extent of emission-related activities and emission factors are sources of error. For instance, Zhao et al. (2011) estimated the uncertainties of a bottom-up inventory of Chinese anthropogenic NO_x emissions to be $-13\% \sim 37\%$. In addition, temporal (e.g. diurnal, weekly, seasonal, inter-annual) variations in emissions are generally poorly represented in the inventories. For instance, rapid economic growth in industrialized Asia has led to a rapid increase in the concentrations of O_3 precursors, such as NO_2 , CO (Richter et al., 2005; Stavrou and Müller, 2006; van der A et al., 2008), and Volatile Organic Compounds (VOCs) (Fu et al., 2007), but these may not be captured well by most of the inventories (Lamsal et al., 2011).

In the past decade, top-down inverse modelling approaches have been proposed to estimate emission variations in CO (e.g. Kasibhatla et al., 2002; Arellano et al., 2004; Stavrou and Müller, 2006; Kopacz et al., 2009; Hooghiemstra et al., 2011) and in NO_x (e.g. Martin et al., 2003; Boersma et al., 2008b; Zhao and Wang, 2009; Lamsal et al., 2010; Miyazaki et al., 2012). The inversion adjusts the emissions in order to minimize the discrepancy between the model predictions and observations, while taking the observation errors into account. The estimated regional emissions show large discrepancies among different estimates, reflecting differences in inversion frameworks, atmospheric models (e.g. Arellano and Hess, 2006), and datasets (e.g. Miyazaki et al., 2012) employed in the analyses. Since the relationship between surface emissions and atmospheric abundances is assumed to be predicted well by the model in the inversions, it is important to represent the chemical processes in a realistic way when estimating the emissions. The CO-OH- NO_x -non-methane VOC (NMVOC) chemical interactions may have large impacts on the inversion of NO_x and CO emissions (Müller and Stavrou, 2005). For instance, neglecting the chemical feedback of changes in surface emissions on the abundance of OH could introduce biases in the a posteriori estimates of the CO sources (Jones et al., 2009).

Data from satellite sensors can provide strong constraints on tropospheric composition in inversions. Tropospheric NO_2 column concentrations have been retrieved by Ozone Monitoring Instrument (OMI) (Levelt et al., 2006), Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) (Bovensmann et al., 1999), and Global Ozone Monitoring Experiment (GOME) and GOME-2 Callies et al. (2000). Tropospheric CO has been retrieved from Tropospheric Emission Spectrometer (TES) (Lopez et al., 2008), Infrared Atmospheric Sounding Interferometer (IASI) (Turquety et al., 2004), Measurement of Pollution

in the Troposphere (MOPITT) (Deeter et al., 2003; Edwards et al., 2006), and SCIAMACHY (Gloude-mans et al., 2009). In recent years, tropospheric O_3 retrievals in the thermal infrared (TIR) have been made from TES (Beer, 2006; Parrington et al., 2008), the Atmospheric Infrared Sounder (AIRS), and IASI (Coman et al., 2012). Limb-viewing satellite instruments, like the Microwave Limb Sounder (MLS), are capable of providing valuable information on the upper troposphere and the lower stratosphere (UTLS) (Waters et al., 2006). Because of the differences between the instruments in terms of sensitivity, retrieval techniques, and observing schedules, the combined use of different satellite instruments is challenging (e.g. Kopacz et al., 2010).

Data assimilation is a technique to combine different observational data sets with a model (e.g., Kalnay, 2003). Data assimilation systems for tropospheric chemistry have been developed in the past decade for mapping the global distribution of chemical species, including O_3 and its precursors. In the past decade, advanced techniques involving the variational approach (Elbern and Schmidt, 2001; Errera et al., 2008; Flemming et al., 2009; Elguindi et al., 2010) and Kalman filters (Khattatov et al., 2000; Eskes and Boersma, 2003; Grassi et al., 2004; Hanea et al., 2004; Segers et al., 2005; Parrington et al., 2008) have been applied to atmospheric chemistry. Recently, the ensemble Kalman filter (EnKF) technique has been applied for tropospheric chemical data assimilation (van Loon et al., 2000; Arellano et al., 2007; Constantinescu et al., 2007; Coman et al., 2012). The EnKF uses an ensemble forecast to estimate the background error covariance matrix. The advantage of the EnKF is its easy implementation for complicated systems; i.e. it does not require the development of an adjoint code.

The use of data assimilation for atmospheric chemistry, especially for short-lived chemical species, is still challenging, as discussed by Lahoz et al. (2007) and Sandu and Chai (2011). Short-lived species concentrations vary on timescales from less than a minute to one day, and detailed treatment of various chemical processes is required to simulate the variability. A large part of the atmospheric chemical system is not sensitive to the initial conditions because of the chemical equilibrium, which is different from the chaotic system involved in the numerical weather prediction (Constantinescu et al., 2007; Lahoz et al., 2007), but is sensitive to the model parameters (e.g. emission, chemical reaction rate, and deposition velocity) and processes (e.g. chemical reaction equation, wet and dry deposition, and atmospheric transport). Although the errors in simulated tropospheric composition are caused by many factors, they are largely affected by highly uncertain emissions (e.g. Mallet and Sportisse, 2005). Thus, the simultaneous adjustment of emissions and concentrations is a powerful framework in tropospheric chemical data assimilation. However, most recent satellite data assimilation systems optimize either the concentration of a very limited number of chemical species (e.g. Mallet and Sportisse, 2005; Parrington et al., 2008; Flemming et al., 2009; Elguindi et al.,

Table 1. List of satellite observations used for the data assimilation.

Sensor	Platform	Species	Resolution	Version	Reference
OMI	AURA	NO ₂	13 × 25 km	DOMINO ver. 2	Boersma et al. (2011)
TES	AURA	O ₃	5.3 × 8.3 km	Nadir ver. 4	Beer (2006)
MOPITT	TERRA	CO	22 × 22 km	TIR ver. 5	Deeter et al. (2010)
MLS	AURA	O ₃ , HNO ₃		ver. 3.3	Livesey et al. (2011)

Table 2. List of observations used for the validation.

Sensor	Platform	Species	Resolution	Version	Reference
GOME-2	MetOp	NO ₂	80 × 40 km	TEMIS ver. 2	Boersma et al. (2004, 2011)
SCIAMACHY	Envisat	NO ₂	60 × 30 km	TEMIS ver. 2	Boersma et al. (2004, 2011)
TES	AURA	CO	5.3 × 8.3 km	Nadir ver. 4	Parrington et al. (2008)
MLS/OMI	AURA	O ₃	1° × 1.25°	NASA/GSFC	Ziemke et al. (2006)
Ozonesonde	Sonde	O ₃		WOUDC/SHADOZ	
INTEX-B	Aircraft (DC-8)	Several			Singh et al. (2009)

2010) or emissions (e.g. Müller and Stavrou, 2005; Kopacz et al., 2010; Hooghiemstra et al., 2011). Only a few advanced studies (Hanea et al., 2004; Elbern et al., 2007) have demonstrated that the simultaneous optimization of multiple chemical states including emissions is an effective way to improve air quality near the surface using surface in-situ observations.

In this study, an advanced EnKF data assimilation system is presented to simultaneously optimize the chemical concentrations and emissions in the troposphere. Satellite observations of O₃, CO, NO₂, and HNO₃ obtained from TES, MOPITT, OMI, and MLS are assimilated into the global chemical transport model (CTM) “Chemical AGCM for study of atmospheric environment and radiative forcing” (CHASER). TES has the potential to efficiently constrain tropospheric O₃ profiles (Foret et al., 2009). MOPITT is suitable for global CO emission estimates because of its good global coverage. MLS is expected to provide important constraints on the background concentrations of O₃, HNO₃, and other O₃ precursors in the UTLS together with lightning NO_x sources. The high temporal and spatial resolutions of the OMI are useful to optimize NO_x emissions on a daily basis. The assimilation results are validated against independent data, obtained from five satellite instruments, MLS/OMI (tropospheric O₃ column, TOC), TES (CO), and GOME-2 and SCIAMACHY (tropospheric NO₂ column). Global ozonesonde data and aircraft observations obtained during the INTEX-B campaign (Singh et al., 2009) are also used for the validation of the vertical profiles. To the authors best knowledge, this is the first advanced data assimilation system that simultaneously optimizes the concentrations and emissions of multiple tropospheric trace gases, based on multiple satellite sensor/species data sets. The structure of this paper is as follows. Section 2 describes the data. Section 3 introduces the data assimilation system. Section 4 presents Observing System Experiment (OSE) results to identify the relative contribution of each as-

simulated data set. Section 5 presents the data assimilation results including the estimated emissions, the validation, and the properties of the assimilated fields. Section 6 concludes this study. Section 7 discusses future challenges.

2 Observations

This section introduces the observations used for the data assimilation (Sect. 2.1 and Table 1) and validation (Sect. 2.2 and Table 2). The data assimilation requires a non-linear observation operator, H , for each satellite retrieval. The model fields, x , are first interpolated to the horizontal location of each observation and the height of each of the vertical layers using the spatial interpolation operator, S . Then the averaging kernel, \mathbf{A} , and the a priori profile, x_a , of each observation are applied to obtain the model fields in the observation space, y^b ,

$$y^b = H(x) = x_a + \mathbf{A}(S(x) - x_a). \quad (1)$$

The averaging kernel matrix is used to define the sensitivity of the estimated state to changes to the true state, while the trace of the averaging kernel matrix gives a measure of the number of independent pieces of information, i.e. the Degree of Freedom for Signals (DOFs) (Rodgers, 2000). In this approach, the satellite-model difference ($y^o - y^b$) is not, or only weakly, biased by the a priori profile x_a (Eskes and Boersma, 2003; Rodgers and Connor, 2003),

$$y^o - y^b = \mathbf{A}(x_{\text{true}} - S(x)) + \epsilon, \quad (2)$$

where the observational error ϵ is the sum of the measurement error and the representativeness error (both random and systematic), and x_{true} represents the true atmosphere profile. The same observation operator has been also applied for validating the model profile against retrievals in order to remove

the influence of the smoothing error and the retrieval error arising from the a priori profile. For plotting the global distribution, both the retrieved and simulated concentrations are mapped onto a same resolution of $2.5 \times 2.5^\circ$ ($1.25 \times 1^\circ$ for MLS/OMI TOC only).

2.1 Measurements used in the assimilation

2.1.1 OMI tropospheric NO₂ column

The Dutch-Finnish OMI instrument, which was launched aboard the Aura satellite in July 2004, is a nadir-viewing imaging spectrograph (Levelt et al., 2006). Aura traces a sun-synchronous, polar orbit with a period of 100 min. OMI provides measurements of both direct and atmosphere-backscattered sunlight in the ultraviolet visible range from 270 to 500 nm. OMI pixels are 13×24 km at nadir, increasing in size to 24×135 km for the largest viewing angles. OMI tropospheric NO₂ column retrievals, with their daily global coverage, are effective to constrain global NO_x emissions on a daily basis, unlike GOME-2 and SCIAMACHY retrievals which have poorer spatial and temporal resolutions and less global coverage (Richter and Burrows, 2002; Boersma et al., 2008b). The overpass time of OMI (about 13:40 LT) is more suitable for the estimation of lightning NO_x sources than that of GOME-2 and SCIAMACHY (both in the morning). The Dutch OMI tropospheric NO₂ data product DOMINO version 2 (Boersma et al., 2011) is used in this study. The error in OMI NO₂ retrievals for individual pixels can be approximated as 1.0×10^{15} molec cm⁻² + 25 % (Boersma et al., 2011). Details of the retrieval and error estimates are described by Boersma et al. (2004, 2007, 2011). Only observations with a radiance reflectance from clouds of less than 50 % (i.e. cloud fraction less than about 20 %) and surface albedo of less than 0.3 with quality flag = 0 (meaningful tropospheric retrievals) are used, as recommended by the product specification document (Boersma et al., 2011).

The averaging kernel is used to create modeled tropospheric NO₂ columns from the observation operator, which removes the contribution of the retrieval error due to the a priori profile assumed (Eskes and Boersma, 2003), as described by Miyazaki et al. (2012). The spatial resolution of the OMI data is much finer than that of the model used in this study (2.8° , about 300 km in the equator). Thus, there are large representativeness errors in the model because of unresolved small-scale variations. To fill the spatial scale gaps and to obtain more representative data, a super-observation approach has been developed and applied to the OMI data, as described by Miyazaki et al. (2012). The super-observation error covariance matrix includes contributions from the measurement error and the representation error.

2.1.2 TES O₃ profile

TES onboard the Aura satellite was designed to measure the global, vertical distribution of tropospheric O₃ and its precursors (Beer, 2006; Bowman et al., 2009). TES is an infrared Fourier transform spectrometer (FTS) with high spectral resolution (0.1 cm⁻¹) and a wide spectral range from 650 to 3250 cm⁻¹. The version 4 level 2 nadir data obtained from the global survey mode are used in this study. This product consists of 16 daily orbits of nadir-viewing measurements with a spatial resolution of 5×8 km spaced 1.6° apart along the orbit track every other day. The TES algorithm is described by Bowman et al. (2002), Worden et al. (2004), and Bowman et al. (2006). The vertical resolution of TES O₃ profile retrievals is typically 6 km in the tropics and in the summer hemisphere for cloud free conditions (Worden et al., 2004). The peaks of the TES O₃ averaging kernel matrix are generally in the middle troposphere, while its sensitivity is reduced greatly in the lower troposphere. On average, there are less than 2 DOFs for the tropospheric profile in the tropics (Jourdain et al., 2007).

The observation operator is applied to account for the vertical smoothing of the retrievals as reflected by the averaging kernel and for the TES a priori profile. This removes the influence of the a priori profile in the data assimilation, as performed by Jones et al. (2003). The observation error includes the smoothing error, the systematic error, and the measurement error. Vertical correlations due to the smoothing influence of the TES retrievals are accounted for in the forecast error covariance matrix through the influence of the averaging kernel. The TES data used in the data assimilation are filtered following the TES L2 Data Users Guide (Osterman et al., 2009). The C-Curve flag and the emission layer quality flag were used to exclude low-quality data. We excluded data poleward of 70° , where satellite sensitivities are low because of the low brightness temperature. TES O₃ profiles are positively biased by less than 15 % from the surface to the upper troposphere (to 100 hPa) and negatively biased by less than 20 % from the upper troposphere to the lower stratosphere (100 to 30 hPa) compared to ozonesonde data (Worden et al., 2007; Nassar et al., 2008; Boxe et al., 2010). We will investigate the effect of the bias in TES O₃ data on the data assimilation in Sect. 5.1.3.

2.1.3 MOPITT CO profile

The MOPITT instrument was launched onboard EOS Terra in December 1999. MOPITT measures thermal emission in the 4.7 μ m and 2.2 – 2.4 μ m absorption band. The equator crossing time is 10:30 LT/22:30 LT with global coverage every 3 days. The data employed are the version 5 level 2 TIR data (Deeter, 2011). The MOPITT instrument is mainly sensitive to free tropospheric CO, especially in the middle troposphere, but it also provides boundary layer information (Deeter et al., 2003, 2007, 2010). DOF is typically much

larger than 0.5, indicating that most of the information comes from the measurement as opposed to the a priori (Kopacz et al., 2010). Maximum zonal mean DOF values of approximately 1.5 occur in daytime overpasses over land in the tropics.

The retrieved error represents the cumulative error from the smoothing error, model parameter error, forward model error, geophysical noise, and instrument error. These are accounted for in the observation error covariance. We exclude MOPITT data in polar regions ($>65^\circ$ latitude), where the quality deteriorates because of potential problems related to cloud detection and icy surfaces. Also, the retrievals in these regions have low information content related to poor thermal contrast conditions. Daytime conditions typically provide better thermal contrast conditions for TIR-based retrievals than nighttime conditions over land, whereas nighttime observations have not been validated and appear subject to larger bias (Heald et al., 2004). We thus exclude the nighttime MOPITT data using a filter based on solar zenith angle. The super-observation approach is applied to the MOPITT data in the same manner as for the OMI data. The representativeness error for the MOPITT super-observations derived from the variability of the observed concentrations in a super-observation grid-cell is typically much smaller (less than 5 %) than that for OMI tropospheric NO_2 columns. Validation results based on in situ profiles exhibit a bias of less than 1 % at the surface, 700 hPa, and 100 hPa, and nearly -6% at 400 hPa for version 4 data (Deeter et al., 2010). No bias correction is applied to MOPITT data in this study, which may lead to slight bias in the estimated CO emissions. The MOPITT data on the 9 pressure levels (900, 800, 700, 600, 500, 400, 300, 200, and 100 hPa) and at the surface are used in the data assimilation, while the data only at 700 hPa is used for the CO emission optimization.

2.1.4 MLS O_3 and HNO_3 profile

The MLS instrument was launched in August 2004 onboard the Aura satellite. Vertical profiles of several atmospheric parameters are retrieved from the millimeter and sub-millimeter thermal emissions measured in the atmospheric limb (Waters et al., 2006). The vertical resolution for the standard O_3 product is up to 2.5 km in the uppermost troposphere and stratosphere. We use the version 3.3 level 2 MLS O_3 and HNO_3 products. A detailed validation and comparison with other data sets is available in Livesey et al. (2011).

We used data with good quality flags, with quality fields greater than 0.6 (1.0), odd status fields, and convergence fields less than 1.18 (1.6) for O_3 (HNO_3), following the recommendations in Livesey et al. (2011). In the UTLS, the MLS version 3.3 retrieval provides data at 6 levels, 316, 261, 215, 150, 100, and 68 hPa. Since further evaluations are still required for data for pressures higher than 261 hPa, we use only data for pressures lower than 215 hPa. For HNO_3 , data for pressures less than 150 hPa are used because of large

systematic uncertainties at 215 hPa ($\pm 30\%$). Detailed instructions for screening tropical-cloud-induced outliers in the HNO_3 and O_3 products given in the version 3.3 data quality document (Livesey et al., 2011) were applied before data assimilation. Because the instrument's vertical resolution is reasonably comparable to the model grid, the averaging kernel is neglected. The measurement error is used as the diagonal element of the observation error covariance matrix, while the vertical correlation is neglected.

2.2 Measurements used for validation

2.2.1 SCIAMACHY tropospheric NO_2 column

SCIAMACHY, which was launched in March 2002 on board ENVISAT (Bovensmann et al., 1999), is a passive remote sensing spectrometer observing backscattered, reflected, transmitted and emitted radiation from the atmosphere and the Earth's surface, in the wavelength range between 240 nm and 2380 nm and with a spectral resolution of 0.25 nm in the UV and 0.4 nm in the visible. We use the version 2 tropospheric NO_2 data from the KNMI retrieval algorithm (Boersma et al., 2004, 2011). The ground pixel of the nadir mode is generally 60×30 km, but depends on the solar zenith angle, with global coverage approximately once every six days. The local overpass time is 10:00 LT. The approach adopted to calculate the AMF is the same as that for DOMINO version 2 data. Errors in the slant column fitting, the stratospheric corrections, and in the AMFs lead to an overall error in the SCIAMACHY retrieval, as described in Boersma et al. (2004). The error for individual pixels can be approximated as $0.7 \times 10^{15} \text{ molec cm}^{-2} + 25\%$ (Boersma et al., 2011). Cloud radiance fraction of less than 50 % with quality flag = 0 is used for the comparison.

2.2.2 GOME-2 tropospheric NO_2 column

GOME-2, which is an improved version of the GOME instrument, is a nadir UV-visible spectrometer (Callies et al., 2000). GOME-2 covers the spectral range between 240 nm and 790 nm and has a spectral resolution between 0.25 nm and 0.5 nm. The ground pixel size of GOME-2 tropospheric NO_2 retrievals is 80×40 km, with a global coverage within 1.5 day. The equatorial overpass time is at 09:30 LT in the descending node. This study employs the version 2 tropospheric NO_2 data from the KNMI retrieval algorithm (Boersma et al., 2004, 2011). The error for individual pixels can be approximated as $0.7 \times 10^{15} \text{ molec cm}^{-2} + 25\%$ (Boersma et al., 2011). Only observations with a radiance reflectance of less than 50 % from clouds with quality flag = 0 were used.

2.2.3 TES CO

Version 4 CO profiles retrieved from TES measurements are used for the validation. The TES CO retrievals are sensitive

primarily to CO in the troposphere, with a DOF between 1 and 1.5 for the tropospheric profile. The maximum sensitivity appears in the lower troposphere, below 500 hPa (Parrington et al., 2008).

Global patterns of CO as measured by TES are in good qualitative agreement with those seen by MOPITT. The mean difference between column abundances of CO from TES and MOPITT was less than 5%. TES CO agrees within the estimated uncertainty of the aircraft instruments, including both errors and the variability of CO itself (Luo et al., 2007; Ho et al., 2009). The TES and MOPITT retrievals both have a maximum sensitivity mainly from 300 to 800 hPa.

2.2.4 OMI/MLS tropospheric O₃ column (TOC)

Several approaches have been developed to derive global TOC from satellite measurements that involve subtracting the stratospheric O₃ column measured in the limb from the total O₃ column measured independently in the nadir (Ziemke et al., 2006; Schoeberl et al., 2007). The monthly mean TOC data derived using the OMI total columns and the MLS profiles from Ziemke et al. (2006) with a horizontal resolution of $1 \times 1.25^\circ$ are used for the validation. Ziemke et al. (2006) produced TOCs at the MLS measurement locations in daylight, where OMI retrievals are available, and where it is not excessively cloudy. Note that the quality of the derived TOC can be very sensitive to the choice of the tropopause definition in this approach (Stajner et al., 2008). Outside the tropics, the large and rapid tropospheric O₃ variability complicates determining tropospheric O₃, as it requires individual observations to be of sufficient accuracy.

2.2.5 Ozone sonde

Ozone sonde observations are taken from the World Ozone and Ultraviolet radiation Data Center (WOUDC) and the Southern Hemisphere Additional Ozone sondes project (SHADOZ) database. The accuracy of the ozone sonde measurement is about $\pm 5\%$ in the troposphere (Smit and Kley, 1998). The observation sites considered for the validation are listed in Table 3. We use data from 39 locations for a total number of 99 (89) observations in January (July) 2007.

To compare ozone sonde measurements with the simulation and the data assimilation, all ozone sonde profiles have been interpolated to a common vertical pressure grid, with a bin of 25 hPa. Then, for each interpolated observed profile, the co-located model profile is computed using the nearest neighbor grid point data for the linear space/time interpolation. The averaged profile is computed globally and for three latitudinal bands, the Northern Hemisphere (30–90° N), the tropics (30° S–30° N), and the Southern Hemisphere (90–30° S). The standard deviations of the normalized differences are computed over these regions.

Table 3. List of ozone sonde stations used for the validation.

Station no.	Name	Latitude	Longitude
STN021	Stonyplain	53.5	−114.1
STN221	Legionowo	52.4	21.0
STN174	Lindendeberg	52.2	14.1
STN316	De Bilt	52.1	5.2
STN053	Uccle	50.8	4.3
STN099	Hohenpeissenberg	47.8	11.0
STN156	Payerine	46.5	6.6
STN458	Yarmouth	43.9	−66.1
STN012	Sapporo	43.1	141.3
STN308	Barajas	40.5	−3.6
STN107	Wallops Island	37.9	−75.5
STN014	Tsukuba	36.1	140.1
STN418	Huntsville	34.7	−86.6
STN190	Naha	26.2	127.7
STN344	Hong Kong	22.3	114.2
STN109	Hilo	19.7	−155.1
STN187	Poona	18.6	73.9
STN494	Alajuela	10.0	−84.2
STN435	Paramaribo	5.8	−55.2
STN434	San cristobal	−0.9	−89.6
STN175	Nairobi	−1.3	36.8
STN466	Maxaranguape	−5.5	−35.3
STN437	Watukosek	−7.5	112.6
STN328	Ascension Island	−8.0	−14.4
STN191	Samoa	−14.2	−170.6
STN394	Broadmeadows	−37.7	144.9
STN256	Lauder	−45.0	169.7
STN029	Macquarie Island	−54.5	158.9

2.2.6 INTEX-B aircraft measurements

Aircraft vertical trace gas (CO, SO₂, O₃, OH, NO₂, NO, CO, HNO₃, PAN, HO₂, CH₂O, and H₂O) profiles were obtained using the UC Berkeley Laser-Induced Fluorescence (TD-LIF) instrument on a DC-8 during the INTEX-B campaign over the Gulf of Mexico (Singh et al., 2009). Thornton et al. (2003), Bucselo et al. (2008), Hains et al. (2010) provide a detailed description and discuss the performance of the measurements. In the comparison between model and assimilation results, the data were binned on a pressure grid with an interval of 30 hPa, while the model output was interpolated to the time and space of each sample. Data collected over highly polluted areas (over Mexico City and Houston) have been removed from the comparison, since they can cause a serious representativeness error in the comparison (Hains et al., 2010). The comparisons were made for March 2006.

3 Data assimilation system

The CHASER data assimilation system (CHASER-DAS) is developed based on an ensemble Kalman filter approach.

This section introduces the forecast model, the data assimilation approach, and the experimental settings.

3.1 The global chemical transport model CHASER

The forecast model used in the data assimilation system is the global CTM CHASER (Sudo et al., 2002). CHASER includes detailed chemical and transport processes in the troposphere, including 88 chemical and 25 photolytic reactions with 47 chemical species, and has a horizontal resolution of T42 (2.8°) and 32 vertical levels from the surface to 4 hPa. CHASER is coupled to the atmospheric general circulation model, Center for Climate System Research/National Institute for Environmental Studies (CCSR/NIES) atmospheric general circulation model (AGCM) ver. 5.7b. The AGCM fields are nudged toward National Centers for Environmental Prediction/Department of Energy Atmospheric Model Intercomparison Project II (NCEP-DOE/AMIP-II) reanalysis (Kanamitsu et al., 2002) at each time step (i.e. every 20 min) in order to reproduce past meteorological conditions.

As described by Miyazaki et al. (2012), the anthropogenic emissions are based on a yearly mean inventory of national emissions obtained from the Emission Database for Global Atmospheric Research (EDGAR) version 3.2 (Olivier et al., 2005). The Global Fire Emissions Data base (GFED) version 2.1 (Randerson et al., 2007), estimated on a monthly basis, is employed for emissions from biomass burning. The monthly biogenic emissions from vegetation, obtained via the GEIA inventory (Guenther et al., 1995), are considered for isoprene, terpenes, and other non-methane VOCs. NO_x emissions from soils are based on monthly mean Global Emissions Inventory Activity (GEIA) (Graedel et al., 1993). The emissions over Asia were obtained from Regional Emission inventory in Asia (REAS) version 1.1 (Ohara et al., 2007). The emissions for the simulation years 2006–2007 are obtained by extrapolating the emissions inventories from the years 1995 and 2000. Emissions of lightning NO_x are linked to convective cloud top height following the parameterization of Price and Rind (1992). The lightning NO_x production is calculated at each time step of CHASER using the convection scheme in the AGCM. The total aircraft NO_x emission is 0.55 Tg N yr⁻¹, which is obtained from the EDGAR inventory. We apply a diurnal variability scheme to the surface NO_x emissions depending on the dominant category for each area: anthropogenic, biogenic, and soil emissions, as in Miyazaki et al. (2012).

3.2 Ensemble Kalman filter data assimilation

The data assimilation technique used in this study is a local ensemble transform Kalman filter (LETKF) (Hunt et al., 2007). The implementation is the same as in Miyazaki et al. (2012). The LETKF has conceptual and computational advantages over the original EnKF (e.g. Ott et al., 2004; Hunt et al., 2007). The LETKF performs the analysis locally in

space and time, reducing sampling errors caused by limited ensemble size. It also reduces the computational cost by performing most calculations in parallel (Miyoshi and Yamane, 2007). Because of the large state vector size and the large number of grid cells in a global CTM, the computational advantages of the LETKF over the original EnKF is important for global tropospheric chemistry data assimilation.

The LETKF transforms a background ensemble ($\mathbf{x}_i^b; i = 1, \dots, k$) into an analysis ensemble ($\mathbf{x}_i^a; i = 1, \dots, k$) and updates the analysis mean, where \mathbf{x} represents the model variable; b the background state; a the analysis state; and k the ensemble size. In the forecast step, a background ensemble, \mathbf{x}_i^b , is globally obtained from the evolution of each ensemble model simulation. The background ensemble mean, $\bar{\mathbf{x}}^b$, and its perturbations (spread), \mathbf{X}^b , are thus estimated from the ensemble forecast,

$$\bar{\mathbf{x}}^b = \frac{1}{k} \sum_{i=1}^k \mathbf{x}_i^b; \quad \mathbf{X}_i^b = \mathbf{x}_i^b - \bar{\mathbf{x}}^b. \quad (3)$$

These are $N \times k$ matrices, where N indicates the system dimension and k indicates the ensemble size. The background error covariance ($\mathbf{P}^b = \mathbf{X}^b (\mathbf{X}^b)^T$) tends to underestimate the true background error covariance because of model errors and sampling errors (Houtekamer and Mitchell, 1998). To prevent the covariance underestimation, the covariance inflation technique (with a covariance inflation parameter of 5 %) is applied at each analysis step, as in Miyazaki et al. (2012).

In the analysis step, an ensemble of background observation vectors in the observation space, $\mathbf{y}_i^b = H(\mathbf{x}_i^b)$, is estimated using the non-linear observational operator H . An ensemble of background perturbations $\mathbf{Y}^b = \mathbf{y}_i^b - \bar{\mathbf{y}}^b$ is also computed. The ensemble mean is then updated by

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{X}^b \tilde{\mathbf{P}}^a (\mathbf{Y}^b)^T \mathbf{R}^{-1} (\mathbf{y}^o - \bar{\mathbf{y}}^b), \quad (4)$$

where \mathbf{y}^o is the observation vector, \mathbf{R} is the $p \times p$ observation error covariance, $\tilde{\mathbf{P}}^a$ is the local analysis error covariance in the ensemble space. The new analysis ensemble perturbation matrix in the model space \mathbf{X}^a is simultaneously obtained by transforming the background ensemble \mathbf{X}^b . An ensemble simulation with the new analysis ensemble is then used to predict the new background error covariance \mathbf{X}^b in the next forecast step. Further details are described in Hunt et al. (2007) and Miyazaki et al. (2012).

EnKF approaches always have a spurious long distance correlation problem because of imperfect sampling of the probability distribution due to limited ensembles (Houtekamer and Mitchell, 2001). In complex chemical data assimilation systems, a realistic estimation of the background error distribution is very important (Singh et al., 2011; Massart et al., 2012). Boynard et al. (2011) demonstrated that the spatial correlations estimated from ensemble simulations are overestimated in the chemical model error covariance fields, and suggested the need for special attention to avoid

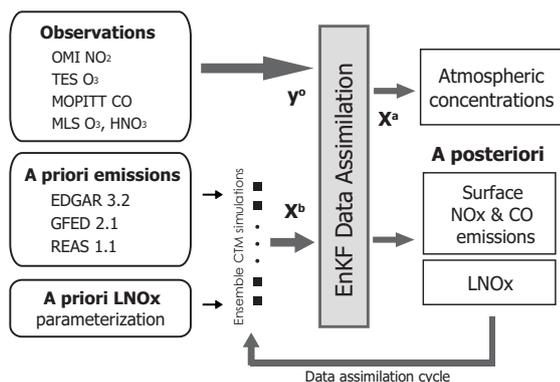


Fig. 1. Schematic diagram of the data assimilation system. The ensemble model simulation with a priori emissions is used to provide the background error covariance information (\mathbf{X}^b). The data assimilation is performed using the background error information and the observation information (\mathbf{y}^o). Then the data assimilation provides a posteriori estimates of surface NO_x emissions, surface CO emissions, lightning NO_x emissions, and 3-D distributions of the chemical species (\mathbf{X}^a). Assimilation of MLS O_3 and HNO_3 data affects the concentrations only above 260 and 220 hPa, respectively. O_x is the sum of O_3 and $\text{O}(^1\text{D})$, and NO_x is the sum of NO , NO_2 , and NO_3 . See Sect. 3 for details.

too large correlation of fields distant from the location of the observation. A covariance localization technique is used to avoid possible degradation because of under sampling. We assumed that observations located far from the analysis point have larger errors and that those observations have less effect on the analysis (Miyoshi and Yamane, 2007). A correct choice of ensemble size and correlation lengths is important to improve the data assimilation performance, as will be discussed in Sect. 3.3.4

3.3 Experimental setting

Three series of one-month data assimilation experiments have been conducted, starting from the 1 March 2006, 1 January 2007, and 1 July 2007. The March 2006 experiment was used to validate against the INTEX-B air craft data, while the January and July 2007 experiments were used to compare the seasonal difference in the data assimilation performance. The data assimilation cycle is 100 min; e.g. each orbit cycle of polar-orbit satellites. This setting is useful to reduce the time discrepancy (sampling errors) between the observations and the model in the data assimilation, given distinct diurnal variation in tropospheric chemistry (Miyazaki et al., 2012). Figure 1 shows a schematic diagram of the data assimilation process.

In addition to the full assimilation run (with all the data), we have also conducted a control run (without any assimilation), five observing system experiments (see Sect. 4), an emission inversion run, and a fixed-flux assimilation run. In the emission inversion run, only surface emissions are opti-

mized (i.e., without concentrations in the state vector). In the fixed-emission assimilation run, only concentrations are updated from the data assimilation (i.e., without emissions in the state vector). The emission inversion run and the fixed-emission assimilation run have been compared with the full assimilation run in which both the concentration and emission are updated. The comparison allows us to understand the relative importance of the emission optimization and the direct concentration adjustment in the simultaneous assimilation (see Sect. 5.3). Further, we have conducted an idealized data assimilation experiment in which synthetic observations are derived from a perturbed model run. The results obtained from the idealized experiment confirmed that the data assimilation system is properly implemented, and the simultaneous optimization for O_3 concentration and its precursors emissions is a powerful framework for the tropospheric chemistry analysis (see Appendix A).

3.3.1 State vector

The state vector is chosen to include uncertain model aspects that most effectively optimize the tropospheric chemical system. First, emissions are a major source of uncertainty in CTM simulations. The solution of a tropospheric chemical model is only weakly influenced by the initial conditions, because of the strong stiffness of tropospheric chemical processes (Constantinescu et al., 2007; Lahoz et al., 2007). An improvement could be achieved by an ensemble obtained by perturbing various parameters of the model (emissions, reaction rates, etc.). The EnKF can be extended to include such parameters in the data assimilation process. A state vector which includes both the concentrations and the emissions makes it possible to find the optimal values for the emissions, which are linked to the concentrations by the CTM. In the EnKF system, the background error covariance, estimated from the ensemble CTM simulations, varies with time and space, reflecting dominant atmospheric processes. The local analysis increment for emissions thus reflects the complex indirect relationship between concentrations and emissions of related species.

The surface emissions of NO_x , $e(\text{NO}_x)$, the surface emissions of CO, $e(\text{CO})$, the lightning sources of NO_x , $e(\text{LNO}_x)$, and the concentrations of all the predicted (i.e., transported, total 35) chemical species, c , are optimized at all the models grid cells for each data assimilation cycle. The concentrations of radical and members of family species are not included in the state vector. The data assimilation influences their concentrations through the chemical coupling during the forecast. The background ensemble can be represented as follows,

$$\mathbf{x}_i^b = \begin{bmatrix} c_i^b \\ e(\text{NO}_x)_i^b \\ e(\text{CO})_i^b \\ e(\text{LNO}_x)_i^b \end{bmatrix}. \quad (5)$$

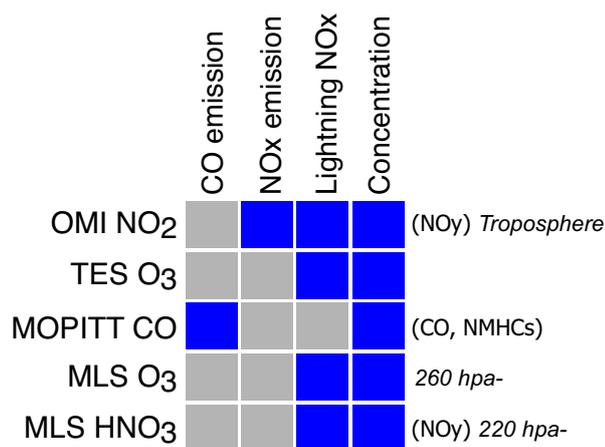


Fig. 2. Schematic diagram of the correlation matrix between observations and the state variables. Satellite data used for the data assimilation are listed in the left column. The model variables updated during the data assimilation are listed in the top row. The blue (gray) colour indicates that correlations between the observed variables and the model variables are considered (neglected using the variable localization technique). See Sect. 3.3 for details.

Although the data assimilation system simultaneously updates emissions of NO_x and CO, we treat the data independently and do not include NO_x-CO emissions covariance in the background error matrix. This is to avoid the effects of spurious multi-variate correlations in the background error covariance, possibly developed because of limited ensembles, and errors in both model and observations. However, the forecasted atmospheric concentrations of NO₂ and CO are coupled chemically through their effect on the tropospheric chemistry.

Based on sensitivity experiment results (see Sect. 4), we have also applied the variable localization to improve the analysis. This means the covariance among non- or weakly-related variables is set to zero. This technique allows us to neglect the correlations among variables that may suffer significantly from spurious correlations. The optimization of the variable localization was based on a comparison against satellite data. If the data assimilation significantly deteriorated the agreement with at least one of the data used for the data assimilation and the validation, variable localization was applied to reduce the deterioration by considering dominant chemical processes, as will be further described in Sect. 4.2. The state vector structure used is summarized in Fig. 2. With the technique, lightning NO_x sources are optimized using TES O₃, OMI NO₂, and MLS O₃ and HNO₃ observations, whereas the covariance between CO concentration and lightning NO_x sources was set to be zero, since their error correlation are not expected to contain meaningful information. Similarly, OMI tropospheric NO₂ column data are used to update the concentrations of NO_y (= NO_x + HNO₃ + HNO₄ + PAN + MPAN + N₂O₅)

species only, since the ensemble may not contain meaningful information on the profile of other chemical species. For the same reason, and related to their poor quality, MLS HNO₃ data are only allowed to influence the NO_y species in the analysis. Similarly, MOPITT CO data affect the concentration of CO, hydrocarbons, and formaldehyde only. CO emissions are optimized using MOPITT CO data only. The variable localization is found to significantly improve the analysis (see Sect. 4.2).

3.3.2 Parameter estimation

A diurnal variability is implemented for the NO_x emissions as in Miyazaki et al. (2012), depending on the dominant source category for each area. The lightning NO_x sources vary in time and space, reflecting the variability in meteorological fields. However, because a model error term is not implemented during the forecast step, the background error covariance can be continuously deflated and underestimated during the data assimilation. To prevent covariance underestimation during the data assimilation, we have applied a covariance inflation to the analyzed emission as in Miyazaki et al. (2012). The analyzed standard deviation (i.e., background error) is artificially inflated to a minimum predefined value at each analysis step. This minimum value is chosen as 30 % of the initial standard deviation, based on sensitivity experiments. Because of the absence of any forecast model (i.e., model bias) to the emissions, and of the use of the background covariance inflation, initial bias in the a priori emissions can be reduced gradually through the data assimilation cycle using the state-augmentation approach, as discussed by Lin et al. (2008).

The initial error is set to 40 % of the a priori emissions for surface emissions of NO_x and CO. For lightning NO_x sources, the initial error is set to 60 %, considering large discrepancies among different estimates (Schumann and Huntrieser, 2007). For the concentrations, it is set to 10 %. Although the optimized emissions (i.e., the analysis mean) and the uncertainty (i.e., the analysis spread) are not strongly sensitive to the choice of the initial error after some assimilation cycles (e.g. several weeks) because of the analysis applied for both the mean and spread fields and the use of the inflation, convergence is generally attained faster in the case for larger initial uncertainties.

3.3.3 Observation error

The observation error covariance matrix contains the measurement error provided by each retrieval. The representativeness error is also considered for the OMI NO₂ and MOPITT CO super-observations as in Miyazaki et al. (2012). The off-diagonal components are neglected for MLS data; the observation error of one measurement is assumed to be independent of the observation error of other measurements. For TES O₃ and MOPITT CO data, the full error covariance

is used, including correlations between vertical layers. We also account for the influence of the averaging kernels from the instruments, which captures the vertical sensitivity profiles of the retrievals. The horizontal correlation in the observation error covariance matrix is neglected. We do not attempt to remove possible biases from the observations before assimilation, mainly because of the difficulty in estimating the true bias structure; this will be further discussed in Sect. 5.1.2.

3.3.4 Assimilation parameters

Since the DOF of the state vector employed in this study is large ($\sim O(10^6)$), a large ensemble size is essential to capture the proper background error covariance structure, but at the expense of an increased computational cost. We have optimized the data assimilation parameters based on sensitivity experiments. The observation-minus-forecast (OmF) analysis (see Sect. 5.1.1) was used to choose the best value of the ensemble size and localization length, as summarized in Table 4. The sensitivity experiment showed that the analysis is improved significantly by increasing the ensemble size from 16 to 32 and is further somewhat improved by increasing it from 32 to 48, as seen in the OmF reduction in the comparison, for instance, with MOPITT CO, MLS O₃, and MLS HNO₃ data. In contrast, the impact was much less significant by increasing it from 48 to 64. The ensemble size is accordingly set to 48. The sensitivity experiments also show that the analysis results are sensitive to the horizontal localization length. The inclusion of spatial correlations with appropriately chosen correlation lengths leads to improvements. From the sensitivity experiments, the horizontal localization length was set to 450 km for NO_x emissions and 600 km for CO emissions, lightning NO_x, and the concentrations. Too short localization length (i.e. half size) increases the OmF error, for instance, for MOPITT and MLS data, because of the neglected influence of remote observation information. Although the larger localization length (i.e. double size) somewhat reduces the OmF for some cases, we use the above-mentioned setting to avoid possible serious spurious correlations. The physical vertical localization length was set to $\ln P$ [hPa] = 0.2 based on sensitivity experiments (results not shown). The optimal length, however, may depend on the location, season, species, and model resolution (Pajot et al., 2011), reflecting the chemical lifetime of the species and atmospheric wind patterns.

4 Observing system experiments

Observation system experiments (OSEs) are used to study how each individual observational data set improves the overall performance. We have conducted five OSEs by separately assimilating OMI NO₂, TES O₃, MOPITT CO, MLS O₃, and MLS HNO₃ data, and the results are compared with

the control run (without any assimilation) and the full assimilation run (with all the data).

4.1 Background error covariance

The background error covariance estimated from ensemble simulations allow unobserved species to be constrained by observed species. Inter-species adjustment can be expected when observed and unobserved species chemically interact on a time scale of the order of the assimilation cycle. The background error covariance follows from the assumption that the background ensemble perturbations \mathbf{X}^b sample the forecast errors,

$$\mathbf{P}^b = \mathbf{X}^b (\mathbf{X}^b)^T. \quad (6)$$

Figure 3 shows the simulated global mean background error covariance structure, \mathbf{P}^b . The covariance analysis shows tight correlations between variations in surface emissions and low-level concentrations of chemically-related species. NO_x emissions show strong positive correlations with low-level (950 hPa) concentrations of NO_x ($r = 0.66$), O_x ($r = 0.60$), N₂O₅ ($r = 0.69$), HNO₃ ($r = 0.62$), and HNO₄ ($r = 0.59$), whereas its relation to upper-level (500 hPa) concentrations is much less significant. CO emissions have a significant correlation with the lower tropospheric CO concentration ($r = 0.74$), but does not relate to other species obviously. Because of the time delays associated with vertical mixing, the middle tropospheric CO is generally delayed in phase, with less variability associated with the CO emission variability. Positive correlations are found between lightning NO_x sources and concentrations of O_x ($r = 0.18$) and NO_y species (e.g. $r = 0.30$ for NO_x) in the middle troposphere, demonstrating the potential to constrain lightning NO_x sources from those observations. Note that correlations with lightning NO_x sources are more robust in the tropics ($r = 0.30$ for O_x and $r = 0.36$ for NO_x between 25° S and 25° N) than the global mean. Negative correlations are also found between reactive species. For instance, $r = -0.63$ between O_x and C₂H₄ (ethene) at the surface results from the removal of C₂H₄ as a result of the fast reaction with OH and O₃ (Sawada and Totsuka, 1986).

The background error covariance shows significant correlations among the concentrations of related chemical species, reflecting the complex tropospheric chemical processes. For instance, O_x shows large correlations with NO_y species, CO, CH₂O, SO₄, and PAN at low levels (with $r > 0.30$). NO_x shows a similar covariance structure, reflecting strong chemical links between O_x and NO_x (with $r = 0.41$) both in the lower and middle troposphere. There are large correlations among the hydrocarbons throughout the troposphere.

The background error structure strongly depends on the model characteristics, and it may have a critical effect on the data assimilation performance. In complex chemical data assimilation systems, a realistic estimation of the background error distribution is very important, given the noisy

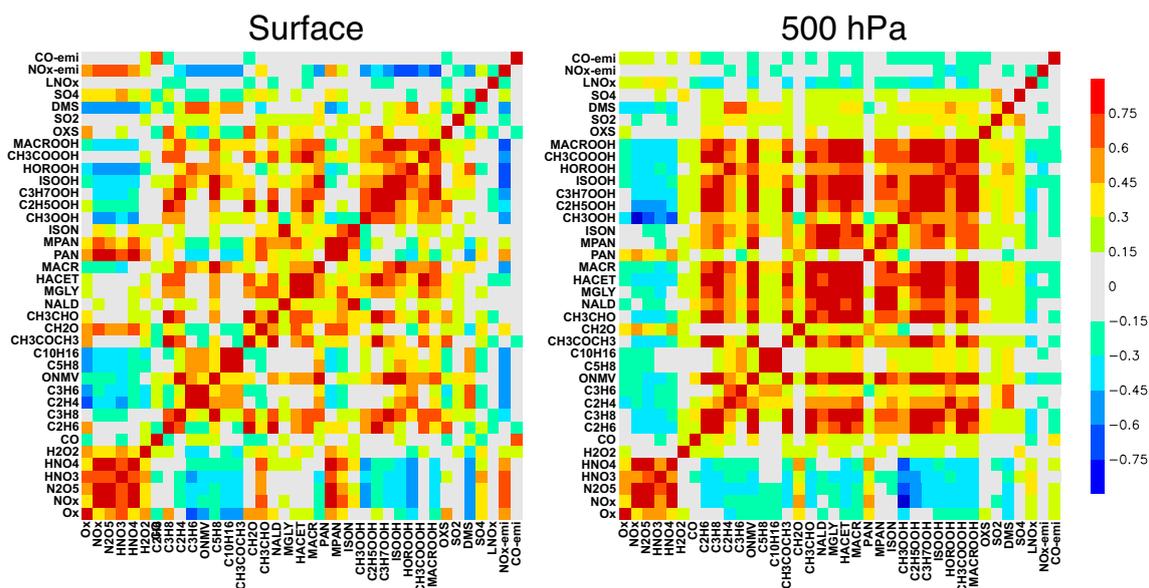


Fig. 3. Correlations between species in the background error covariance matrix, estimated from the LETKF ensemble at 950 hPa (left) and 500 hPa (right) averaged over 15–20 July 2007. The global mean of the covariance estimated for each grid point is plotted. The matrix includes concentrations of all the predicted species, surface NO_x emission ($\text{NO}_x\text{-emi.}$), surface CO emissions (CO-emi.), and lightning NO_x sources (LNO_x). O_x is the sum of O_3 and $\text{O}(^1\text{D})$, and NO_x is the sum of NO , NO_2 , and NO_3 . The red (blue) colour represents positive (negative) correlations.

Table 4. The performance of the data assimilation for different parameters: the horizontal localization length (loc) and the ensemble number (ens). Ten-day mean (averaged over 20–30 January 2007) global mean RMS innovation of the OmF for each assimilated data are shown. The control (CTL) simulation was conducted with $\text{loc} = 450$ km for NO_x emissions and with 600 km for CO emissions, lightning NO_x , and the concentrations, and $\text{ens} = 48$. The simulations with different loc values were conducted with $\text{ens} = 48$. The smallest RMS innovation for each comparison is shown in bold.

	OMI NO_2 (10^{15} molec cm^{-2})	MOPPIT CO (ppbv)	TES O_3 (ppbv)	MLS O_3 (ppbv)	MLS HNO_3 (ppbv)
CTL	1.10	9.05	11.3	81.2	0.64
$\text{loc} \times 0.5$	1.13	9.44	11.3	89.1	0.75
$\text{loc} \times 2.0$	1.15	8.90	10.8	81.1	0.69
$\text{ens} = 16$	1.11	9.09	11.4	84.4	0.70
$\text{ens} = 32$	1.12	9.06	11.3	82.7	0.66
$\text{ens} = 64$	1.10	9.05	11.3	80.9	0.63

observations along with imperfect model predictions, as suggested by Singh et al. (2011).

4.2 Results

The OSEs confirm that the assimilation of each species data set has a strong influence on both assimilated and non-assimilated species through the use of the inter-species error correlation and through the chemical coupling provided by the model forecast. The assimilation of OMI NO_2 data provides some changes in O_3 and CO concentrations, whereas the assimilation of TES O_3 data has some effects on NO_2 fields, as will be shown in Sect. 5.1.2.. These changes are tightly associated with the changes in OH because of the

chemical interactions in the CO-OH-NO_x system, as depicted in Fig. 4. The assimilation of OMI NO_2 data generally increases OH concentrations in the tropical troposphere by 5–15 % and decreases it in the extratropics by 10–20 %. These changes correspond to the increased (decreased) NO_2 concentration in the tropics (in the extratropics) through $\text{NO}_x\text{-OH-O}_3$ chemical reactions in the NO_x -sensitive regime. The assimilation of TES O_3 data also significantly changes OH concentrations. The obtained O_3 increment results in a 10–20 % increase in OH concentration in the extratropics. As a result of the combined assimilation of all satellite data sets, zonal mean OH concentration is increased by 5–15 % in the tropics and the Southern Hemisphere, and the north-to-south

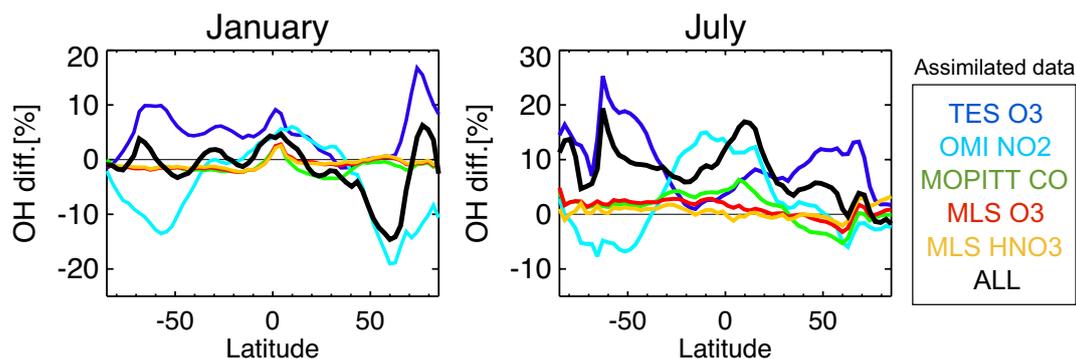


Fig. 4. Latitudinal distributions of the effect of data assimilation on the mean concentration of OH, averaged between 800 and 550 hPa for 16–30 January 2007 (left) and 16–30 July 2007 (right). The percentage difference of the zonal mean concentration, averaged between 800 and 550 hPa, between the assimilation runs and the control run is shown for six different assimilation runs; the full assimilation run (with all the data, black) and the five OSEs with TES O₃ data (marble), OMI NO₂ data (light blue), MOPITT CO data (green), MLS O₃ data (red), and MLS HNO₃ data (yellow). A positive (negative) value indicates that the assimilation run has a higher (lower) concentration than the control run.

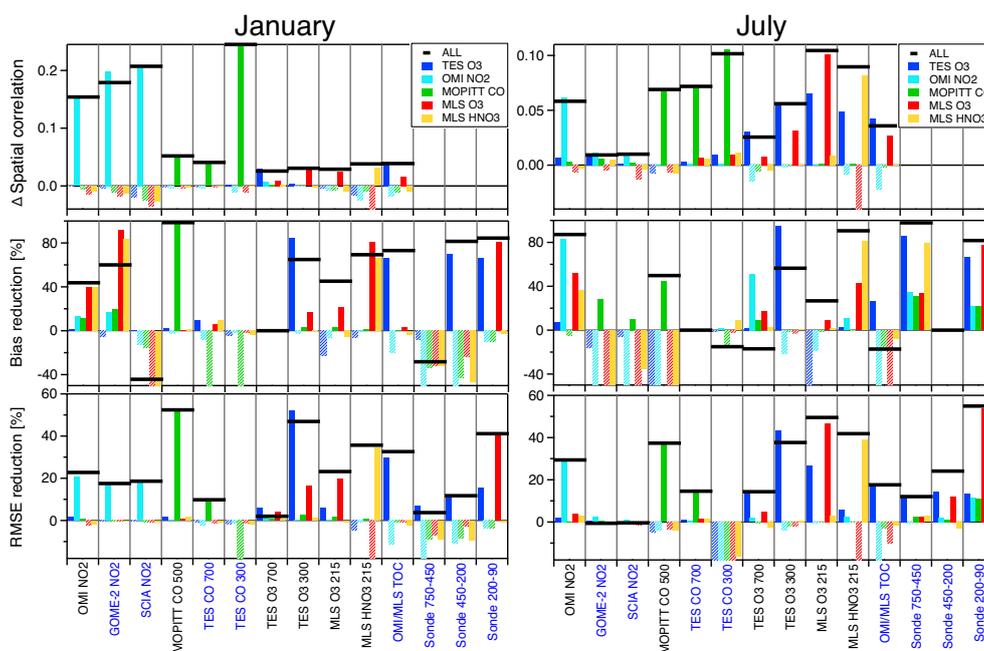


Fig. 5. The differences in the global spatial correlation, the global mean bias, and the global RMSE between the data assimilation runs and the control run for the 16–30 (from the 7–30 only for the ozonesondes) of January (left) and July (right) in 2007. These scores are first estimated from the comparison against observations listed at the bottom (assimilated data in black and independent data in blue), and then compared with the control run. For the spatial correlation, the difference (the data assimilation runs minus the control run) is positive (negative) when the spatial correlation is higher (lower) in the data assimilation runs than in the control run. For the bias and the RMSE, the error reduction rate defined by Eq. (7) is plotted; the positive (negative) value represents that the error is smaller (larger) in the data assimilation runs than in the control run. A reduction rate of 100 % indicates that the error in the model is completely removed by the data assimilation. The results are shown for six different data assimilation runs (the full assimilation run and the five OSEs). The number shown in the bottom list represents the approximate altitude level in hPa.

gradient in OH concentration in the free troposphere is reduced in July.

The OSEs quantify the improvement due to the assimilation of each individual species data set in comparison with the assimilation of all data sets and the control run with-

out assimilation. The global spatial correlation, root-mean-square-error (RMSE), and mean bias for 15-days (from the 16th to the 30th of each month) mean fields were estimated for the control run and the OSEs. The improvement rate due to each data set was then estimated by comparing scores

between the control run and the OSE, as shown in Fig. 5. For the RMSE and bias, the error reduction rate is estimated by comparing these statistics between the control run (E_{cntl}) and the OSE (E_{OSE}) as follows,

$$\frac{|E_{\text{cntl}}| - |E_{\text{OSE}}|}{|E_{\text{cntl}}|} \times 100. \quad (7)$$

When the global mean model bias of the control run is nearly zero, the error reduction rate is not meaningful and is set to zero. This is done for the comparisons with TES O₃ data at 700 hPa in January, TES CO data at 700 hPa in July, and ozonesonde data between 450 and 200 hPa in July. The nearly zero bias compared to TES O₃ and CO data at 700 hPa can be largely attributed to the very small sensitivity of the retrievals at these levels, and does not reflect the true model bias which may be large.

The comparison demonstrates significant improvements of the scores obtained by the assimilation. Improvements in the non-assimilated chemical species show that the ensemble simulation is capable of correctly representing inter-species error correlations and propagating observation information through assimilation cycle. For instance, the assimilation of MLS O₃ and HNO₃ data leads to an improved agreement with OMI NO₂, as shown by the large reduction of the bias. Furthermore, all the assimilated data sets improve the agreement with O₃ profiles obtained from ozonesondes in July, as will be further discussed in Sect. 5.1.2.

Note that the effect of the data assimilation on non-observed species is not always positive. Consideration of inter-species error correlations sometime causes the error to grow. Optimization of the state vector structure is thus conducted to minimize the error based on the OSEs, by neglecting the inter-species correlations that result in serious error growth; the optimized state vector is depicted in Fig. 2. Serious negative effects arose, for example, from the TES O₃ data assimilation on CO fields, the MLS HNO₃ data assimilation on CO fields, the OMI NO₂ data assimilation on O₃ fields, and the MOPITT CO data assimilation on O₃ fields (figure not shown). This is primarily because limited ensembles can cause spurious error correlation among chemical species, especially for species having insignificant chemical links. For instance, because of its relatively long chemical lifetime (~2 months), CO may not have significant correlations with chemically active species such as O₃ in the lower troposphere with a time scale on the order of the data assimilation cycle. Similarly, the OMI NO₂ tropospheric columns may not have enough information to directly constrain the vertical profile of O₃ because of its smooth averaging kernel profile and large observation error. Since we applied the variable localization to avoid these negative influences (see Sect. 3.3.1), the full assimilation run provides the best performance among the individual data assimilation in most cases, see Fig. 5. Note that all the observation data can affect all the chemical fields throughout the forecast.

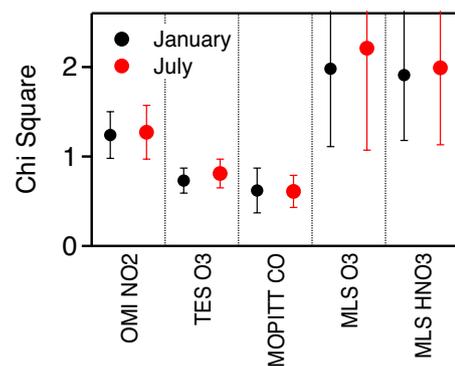


Fig. 6. χ^2 value (in vertical axis) estimated for each assimilated data set (in horizontal axis) averaged over the 10–30 January (black) and July (red) in 2007.

5 Data assimilation results

5.1 Validation

5.1.1 Self-consistency tests

An important test for the quality of data assimilation is whether the differences between the model forecast and observations (the innovations) are consistent with the covariance matrices for the model forecast and observations (e.g. Segers et al., 2005; Lahoz et al., 2007). The background covariance matrix is important in reaching an appropriate balance between the background and the observations. A quantitative criterion for the choice of the background error is a chi-square (χ^2) test, the χ^2 diagnostics (e.g. Ménard and Chang, 2000). χ^2 should approach 1 if the background error covariances are properly specified, while a value higher (lower) than 1 indicates an underestimation (overestimation) of the background error covariance matrices. The χ^2 determined for each assimilated data set is shown in Fig. 6. The χ^2 is greater than 1 for the MLS O₃ and HNO₃ data assimilation, indicating too much confidence in the model. The model overconfidence is associated with the limited lower stratospheric variations in the ensemble, which are strongly constrained by the fixed upper boundary conditions in CHASER. For MOPITT CO and TES O₃ data assimilation, χ^2 lower than 1, which indicates a possible overestimation of the background errors, may result in too much correction of the model fields.

Figure 7 shows the latitude dependence of the bias and root-mean-square (RMS) innovation of the OmF computed in the observation space. The innovation between forecast and assimilated data is a sum of three contributions; the observation error, the forecast error, and the representativeness error caused by mismatches between the satellite ground pixel and the model grid cell (Eskes et al., 2003; Lahoz et al., 2007). A persistent model bias is found in the underestimation of tropospheric NO₂ columns compared to OMI NO₂

Table 5. Comparisons between the data assimilation run and the satellite retrievals. The results are obtained from 15-day averages (from the 16–30 of each month) for January and July in 2007. Shown are the global spatial correlation (Corr), the global mean difference (Bias), and the global root-mean-square error (RMSE). The model simulation results (without data assimilation) are shown in brackets.

	January			July		
	Corr.	Bias	RMSE	Corr.	Bias	RMSE
OMI NO ₂ (10 ¹⁵ molec cm ⁻²)	0.92 (0.77)	-0.05 (-0.09)	0.80 (1.04)	0.93 (0.87)	0.01 (-0.10)	0.25 (0.35)
GOME-2 NO ₂ (10 ¹⁵ molec cm ⁻²)	0.87 (0.69)	-0.02 (-0.06)	1.09 (1.32)	0.80 (0.79)	0.19 (0.04)	0.33 (0.33)
SCIAMACHY NO ₂ (10 ¹⁵ molec cm ⁻²)	0.90 (0.69)	0.11 (0.08)	1.25 (1.53)	0.79 (0.78)	0.23 (0.11)	0.66 (0.66)
MOPITT CO 500 hPa (ppbv)	0.97 (0.92)	0.1 (-7.1)	6.0 (12.6)	0.92 (0.86)	1.1 (-2.2)	7.0 (11.2)
TES CO 700 hPa (ppbv)	0.90 (0.86)	3.3 (-1.5)	15.7 (17.5)	0.83 (0.76)	0.97 (0.03)	13.0 (15.3)
TES CO 300 hPa (ppbv)	0.77 (0.53)	22.2 (13.3)	26.2 (22.1)	0.77 (0.67)	26.1 (17.1)	31.8 (21.7)
TES O ₃ 700 hPa (ppbv)	0.92 (0.89)	-2.7 (-0.4)	5.9 (6.0)	0.91 (0.88)	-3.1 (-2.7)	7.0 (8.0)
TES O ₃ 300 hPa (ppbv)	0.96 (0.93)	5.1 (14.7)	11.0 (20.7)	0.95 (0.90)	3.2 (7.5)	11.7 (18.8)
MLS O ₃ 215 hPa (ppbv)	0.96 (0.93)	10.7 (19.4)	36.8 (47.9)	0.95 (0.90)	10.3 (14.0)	31.9 (63.3)
MLS HNO ₃ 215 hPa (ppbv)	0.81 (0.77)	-0.13 (-0.43)	0.37 (0.57)	0.84 (0.75)	-0.04 (-0.37)	0.36 (0.62)
OMI/MLS TOC (DU)	0.77 (0.73)	-1.1 (3.9)	4.2 (6.3)	0.89 (0.85)	1.3 (1.1)	3.4 (4.2)
Sonde 800–450 hPa (ppbv)		3.4 (4.8)	11.1 (11.6)		-0.15 (-7.14)	19.9 (22.6)
Sonde 450–200 hPa (ppbv)		-0.50 (11.9)	30.7 (34.8)		0.13 (0.98)	30.6 (40.3)
Sonde 200–90 hPa (ppbv)		13.5 (107.0)	122 (208)		4.3 (23.5)	20.6 (45.7)

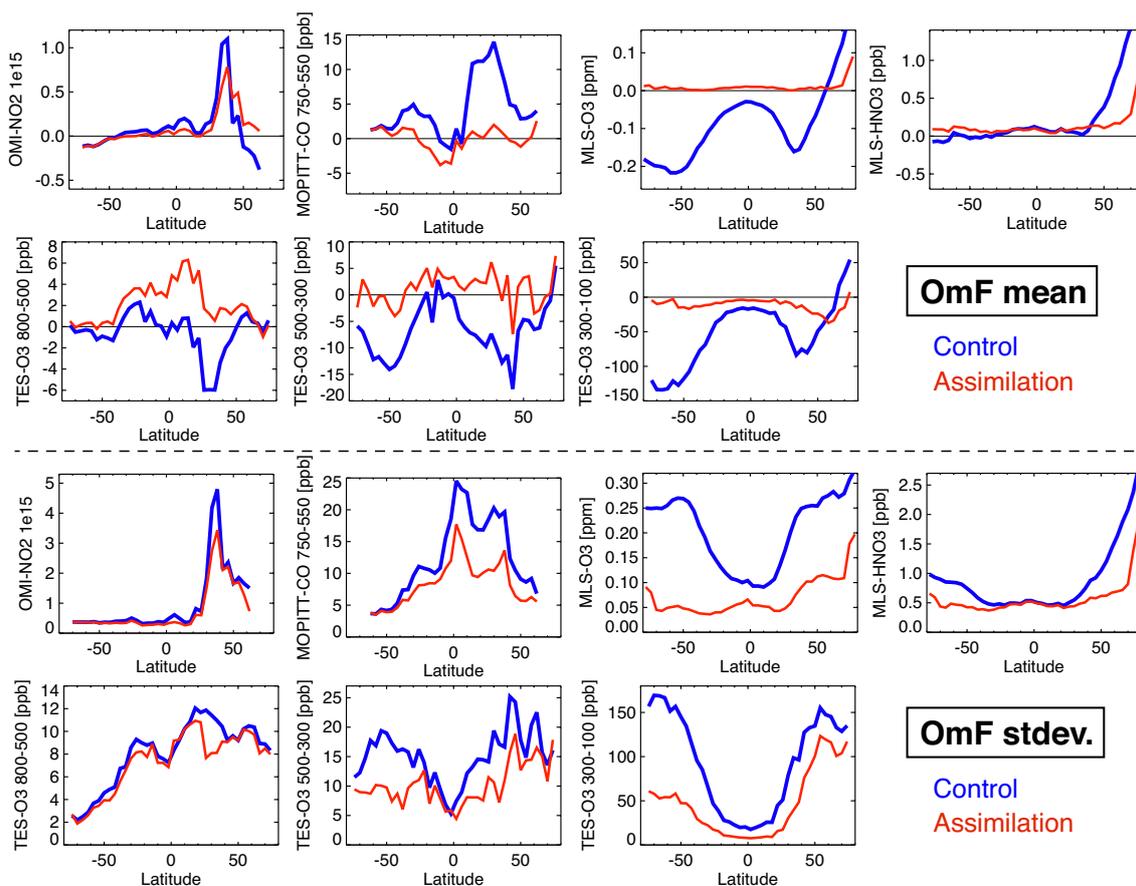


Fig. 7. Latitudinal distributions of the mean OmF (upper panels) and its standard deviation around the mean (lower panels) estimated in the observation space for each assimilated data set, averaged over the period 16–30 January 2007. The results are shown for the data assimilation run (red) and the control run (blue).

data, the overestimation of the middle tropospheric CO in the extratropics compared to MOPITT CO data, and the overestimation in the middle and upper tropospheric O₃ compared to TES and MLS O₃ data. The data assimilation removes most of the OmF bias. The large reduction of the O₃ OmF bias for TES O₃ data in the middle and upper troposphere, which reflected the reduction of the OmA bias, implies that TES O₃ has meaningful information for constraining the O₃ fields at these altitudes, as similarly reported by Parrington et al. (2008). In contrast, the bias reduction is not obvious in the lower troposphere (800–500 hPa), especially at high latitudes. This is because the DOFs of the TES retrieval in the troposphere are generally smaller than 1 poleward of 45° and TES has little sensitivity to the lower tropospheric O₃ (Worden et al., 2004; Osterman et al., 2008). The near zero OmF bias for MLS O₃ in the data assimilation reflects a good coverage and high quality of MLS O₃ data. A long lifetime of O₃ in the UTLS also helps to accumulate the observation information. The observation-minus-analysis (OmA) histogram shows a more pronounced peak than that for OmF (closer to a Gaussian curve, figure not shown) in many cases, as the analysis is closer to the assimilated observations than to the forecast, as shown by Miyazaki et al. (2012).

The standard deviation about the mean of the OmF was found to be mostly equal to the observation error, indicating that the data assimilation captures the observed variability well and satisfies the data assimilation assumptions. A substantial part of the RMS of the OmF has been removed by the data assimilation for MOPITT CO, MLS O₃, and TES O₃ in the middle and upper troposphere. A reduction of the RMS is less pronounced for OMI NO₂, TES O₃ in the lower troposphere, and for MLS HNO₃. These are, respectively, related to rapid spatiotemporal variations and large errors in the observed NO₂, small sensitivities to the true profile (i.e. small averaging kernel), and large observation errors.

5.1.2 Comparison with satellite data

The data assimilation results are validated against independent data, as listed in Table 5 and shown in Fig. 5. The tropospheric NO₂ columns are compared with GOME-2, SCIAMACHY, and OMI data. Differences among the retrievals mainly reflect diurnal variations of chemical processes and emissions, because a very similar algorithm is used for the retrieval of these data. The viewing pixel size difference will not affect the comparison results too much, since these retrievals are gridded to the same resolution (2.5° × 2.5°), using weighting factors for the surface overlap between the satellite pixel and grid cell. The data assimilation largely improves the agreement with these data. The improvement is most pronounced in January. The data assimilation increases the spatial correlation by about 0.15–0.21, decreases the bias by about 85 % (except for SCIAMACHY), and decreases the RMSE by about 30 %. These improvements are mainly attributed to the increased NO₂ columns over East

China and Central Africa and the decreased columns over Europe (Fig. 8). The OSEs confirm that these improvements are mainly due to the OMI NO₂ data assimilation (Fig. 5). The assimilation of MLS O₃ and HNO₃ data also contributes significantly to the reduction of the negative NO₂ column bias compared to OMI and GOME-2 (in January only) data, by increasing the upper tropospheric NO₂ concentration. In contrast, the bias compared to independent GOME-2 (in July only) and SCIAMACHY data is increased by the data assimilation. The errors in the simulated diurnal NO₂ variations along with a bias between OMI and these retrievals may cause the bias to increase. The diurnal variations are especially important in the warmer seasons (e.g. in July in the Northern Hemisphere), when chemistry is sufficiently fast to make a difference between morning and early noon NO₂ columns (Boersma et al., 2009).

The global mean negative bias and large RMSE in the model simulation against the MOPITT CO are mostly (by 40–90 %) removed by the data assimilation, while a very high spatial correlation (0.92–0.97) is maintained. The reduced negative bias is primarily due to the enhanced concentrations over East Asia, North America, and northern Eurasian continent (Fig. 10). Because of the long lifetime of CO, the data assimilation system is able to capture the observed CO variability. This improvement is mostly achieved by the MOPITT CO data assimilation, while the assimilation of other data sets slightly (typically less than 5 %) affected the comparison through their influence on the OH fields.

The data assimilation also improves the spatial correlation with the independent TES CO data both at 700 and 300 hPa, reflecting enhanced concentrations over China, India, Central Africa, and North America, and reduced concentrations over South America (Fig. 10). However, the global mean bias and RMSE mostly increases due to the data assimilation, primarily reflecting too high concentrations at high latitudes. The bias increase is possibly due to a systematic bias between MOPITT and TES. Luo et al. (2007) showed that MOPITT CO version 3 data have higher values than TES CO version 2 data, with a mean difference of –4.8 % at 150 hPa and –0.2 % at 700 hPa. This result is consistent with our data assimilation result, but with different data versions. The assimilation of other species (i.e., NO₂, O₃, and HNO₃) data contributed slightly to improve the agreement.

The data assimilation greatly improves the agreement with the TES O₃ data in the upper troposphere (300 hPa), with a bias reduction of up to 65 % and an RMSE reduction of about 40–50 %. The improvements are mainly due to increased concentrations in the southern extratropics (Fig. 9), which are achieved by the TES O₃ and MLS O₃ data assimilation. The assimilation of OMI NO₂ data acts to reduce the bias compared to TES O₃ data at 700 hPa in July, implying the importance of optimizing O₃ precursors fields. However, the improvement is not significant in the simultaneous data assimilation fields in the lower troposphere. The TES O₃ sensitivity is reduced greatly in the lower troposphere especially due to

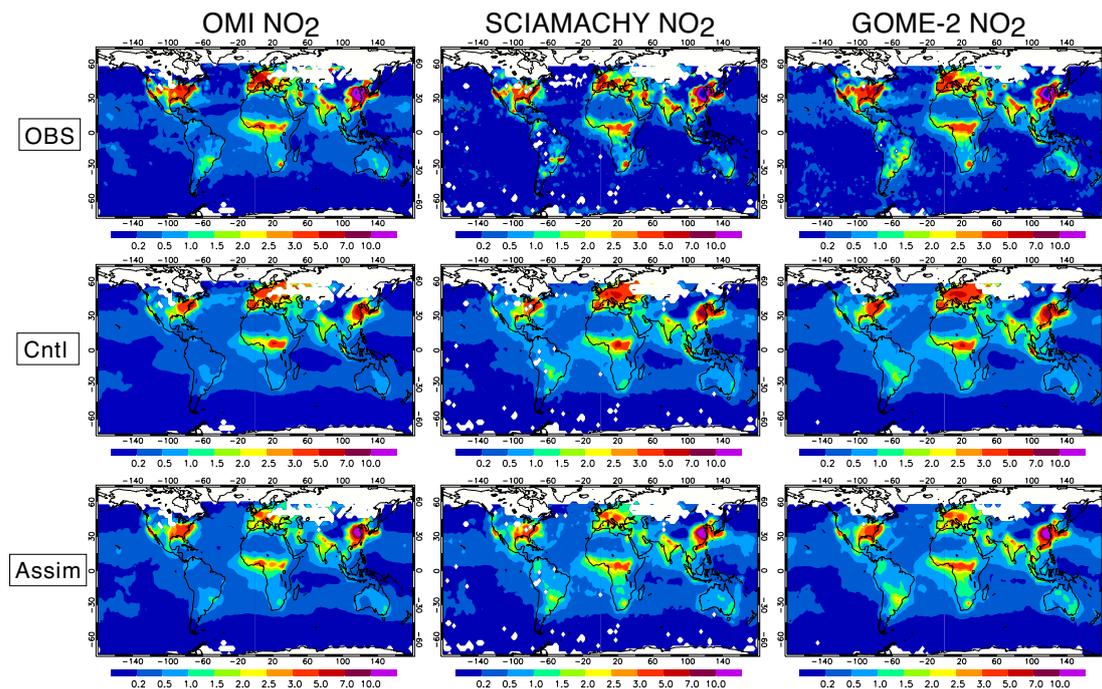


Fig. 8. Global distributions of the tropospheric NO_2 columns (in $10^{15} \text{ molec cm}^{-2}$), averaged over the period 16–30 January 2007. The results are shown for OMI (left columns), SCIAMACHY (middle columns), and GOME-2 (right columns). Upper rows show the tropospheric NO_2 columns obtained from the satellite retrievals (OBS); centre rows from the control run (Cntl); and lower rows from the data assimilation run (Assim). The averaging kernel of each retrieval is applied to the control run and data assimilation fields. The red (blue) colour indicates relatively high (low) values.

the presence of clouds and makes it difficult to improve the analysis. Validation or assimilation is virtually meaningless when the retrieval sensitivity is very low. Since we applied the averaging kernel and the a priori profile information in the comparison (Eq. 1), substantial adjustments in the assimilation or differences in the validation only occur when there is a meaningful signal (i.e. the retrieved profile minus the retrieval a-priori).

The MLS O_3 data assimilation is very effective in removing the positive O_3 model bias in the UTLS because of its wide and dense coverage and good quality, as similarly shown by Jackson (2007) and Feng et al. (2008). The global RMSE against MLS O_3 data is also reduced by TES O_3 data assimilation. However, the OSEs confirmed that the assimilated concentration becomes too high because of the TES O_3 assimilation compared to MLS O_3 data. The OSEs suggest that TES O_3 concentration is higher (lower) in the tropics (extratropics) than MLS O_3 concentration, with a mean difference of 20–40 ppb at altitudes between 200 and 80 hPa.

The bias and the RMSE compared with MLS HNO_3 data have also been largely removed by the data assimilation. The improvement is primarily due to MLS HNO_3 data assimilation, but MLS O_3 data assimilation also contributes to the improvement, as seen in reduced HNO_3 bias. This indicates that MLS O_3 data have meaningful information about the abundance of HNO_3 in the UTLS, through atmospheric transports

and the chemical link. In contrast, the decreased (increased) spatial correlation (RMSE) due to the MLS O_3 data assimilation may be related to errors in the background error covariance or poor data quality either in MLS HNO_3 or MLS O_3 data, especially in the upper troposphere.

The improved agreement with TOC data obtained from the independent MLS/OMI data is mainly attributed to the assimilation of TES O_3 data because of their strong sensitivity to tropospheric O_3 in the tropics. For instance, the high columns over the Atlantic and in the southern subtropics (from South Africa to Australia) are better captured by the data assimilation (Fig. 9). However, the data assimilation still has difficulty in reproducing the observed features. For instance, longitudinal variations with a persistent wave-1 pattern in the southern tropics are larger in both the model and assimilation compared to the OMI/MLS product. This may indicate a difficulty in correcting processes responsible for the enhanced ozone in the Atlantic (e.g. via rapid convective updraft). At the same time, there is large uncertainty in the retrieved TOC. Measuring tropospheric O_3 from space is challenging because of large amount of stratospheric O_3 in the total column, while the separation between the troposphere and stratosphere strongly depends on the tropopause definition (e.g., Bethan et al., 1996).

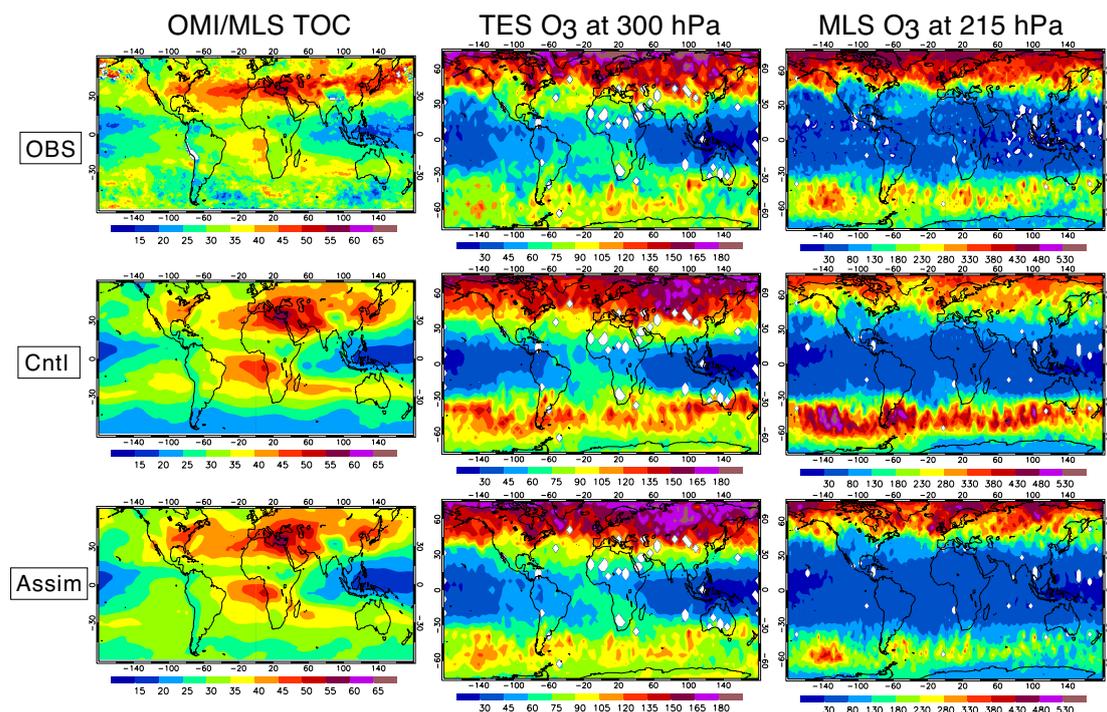


Fig. 9. Global distributions of the tropospheric O₃ columns (in DU) and O₃ mixing ratio (in ppb), averaged over 16–30 July 2007. The results are shown for OML/MLS O₃ columns (left columns), TES O₃ mixing ratio at 300 hPa (middle columns), and MLS O₃ mixing ratio at 215 hPa (right columns). Upper rows show the satellite retrievals (OBS); centre rows from the control run (Cntl); and lower rows from the data assimilation run (Assim). The red (blue) colour indicates relatively high (low) values.

5.1.3 Comparison with ozonesonde data

Figure 11 shows the comparison against the ozonesonde data. Without assimilation the global mean bias with the ozonesonde is large, up to 30 % in the free troposphere and 40 % in the lower stratosphere. The data assimilation removes most of the bias from the middle troposphere to the lower stratosphere, down to within 10 %. It also reduces the RMSE by about 20 % in the middle troposphere and by 50 % in the UTLS. Significant positive biases in simulated O₃ in the UTLS are also mostly removed by the data assimilation, whereas the simulated O₃ profiles suffer from errors in stratosphere-to-troposphere exchange (STE). The great improvements in the UTLS reflects the long chemical lifetime of O₃ and the fact that satellite retrievals capture the large scale variations of O₃ well. The effect of the data assimilation on the lower tropospheric O₃ below about 850 hPa is not obvious on a global scale, implying that further constraints are needed on the near surface O₃ and its precursors (e.g. VOCs). Parrington et al. (2009) demonstrated that the changes in the O₃ flux from the free troposphere into the planetary boundary layer (PBL) by the TES O₃ assimilation reduces the positive bias in the PBL indirectly over North America. Although this effect is not confirmed by our global analysis, it is of interest to survey the detailed spatial distributions resulting from the data assimilation.

The OSEs demonstrate that these improvements are mainly due to the assimilation of TES data in the free troposphere (between 750 to 200 hPa) and both TES and MLS O₃ data in the UTLS (between 200 and 90 hPa). TES data provide valuable constraints on the free tropospheric O₃. Although the MLS data do not extend down to altitudes below 260 hPa, the MLS assimilation influenced the ozone analysis even below this level through the vertical propagation of the observation signal mainly via the extratropical downward motion. It is emphasized that all the assimilated datasets contribute to reduce the global mean bias between 750 and 450 hPa and between 200 and 90 hPa in July. This indicates that the simultaneous assimilation of multiple chemical observations is effective to improve tropospheric O₃, by their influence on the precursor emissions and chemical processes that affect the O₃ concentrations. In contrast, the improvement by the non-O₃ data assimilation is not obvious in January. This may reflect the seasonal difference in the chemical links between O₃ and other species. Since most ozonesonde sites are located in the Northern Hemisphere, the greater improvement in July may be related to summertime active chemical processes in the Northern Hemisphere. Much less ozone is produced from the precursors in winter than in summer (e.g. Liu et al., 1987).

The assimilated O₃ fields show a persistent positive bias compared to the ozonesonde data, with a global mean bias

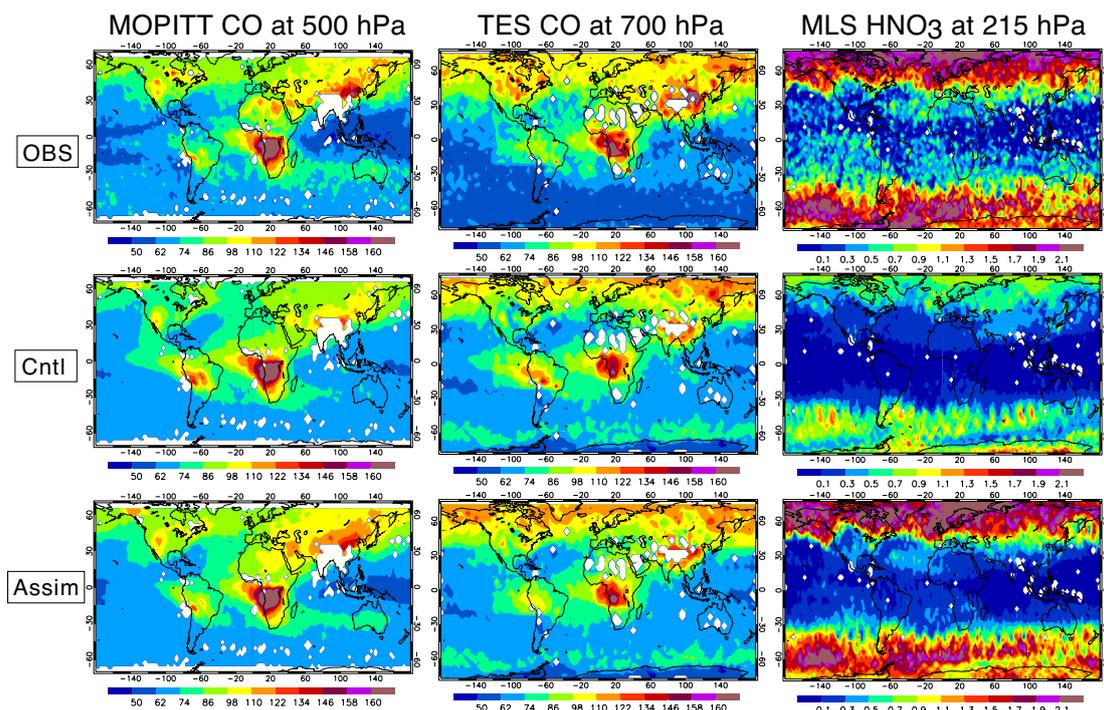


Fig. 10. Global distribution of the CO mixing ratio (in ppb) and HNO₃ mixing ratio (in ppb), averaged over 16–30 July 2007. The results are shown for MOPITT CO mixing ratio at 500 hPa (left columns), TES CO mixing ratio at 700 hPa (middle columns), and MLS HNO₃ mixing ratio at 215 hPa (right columns). Upper row shows the satellite retrievals (OBS); centre row from the control run (Cntl); and lower row from the data assimilation run (Assim). The red (blue) colour indicates relatively high (low) values.

of up to 15 %, below 300 (500) hPa in January (July). The OSEs demonstrate that the positive bias can be attributed to the assimilation of TES O₃ data. The positive bias in TES O₃ data compared to ozonesonde data is reported by Nassar et al. (2008) and Worden et al. (2009). A data assimilation experiment with a bias correction (a uniform 3.3 ppbv bias above 500 hPa and a 6.5 ppbv below 500 hPa, according to Worden et al., 2009) reduces the negative bias in the data assimilation (Fig. 12), demonstrating the importance of bias correction before data assimilation. However the effect of bias correction is not always positive, causing too low concentrations in the middle troposphere in both January and July. A more accurate estimation of the spatially-varying bias is thus required to improve the analysis.

5.1.4 Comparison with aircraft data

Comparisons with aircraft measurements from the INTEX-B campaign allow us to look into the effect of data assimilation on various chemical fields (Fig. 13). The observed NO_x concentrations show a decrease from the boundary layer to the free troposphere. Oxidation of NO_x to HNO₃ and other minor products dominates NO_x loss in the boundary layer, whereas conversions to HNO₃ and PAN dominate it in the free troposphere (e.g. Staudt et al., 2003). The increase in HNO₃ toward the surface is driven by chemical production

of HNO₃ in polluted areas, while HNO₃ is depleted in the troposphere because of deposition processes. Compared to the observed profiles, the simulated NO₂ is slightly lower in the boundary layer and too low in the free troposphere, while HNO₃ is too high by 500 pptv in the boundary layer. Observed PAN shows a maximum above the boundary layer and a minimum in the free troposphere, while the simulation overestimates (underestimates) it by 80 pptv (by 200 pptv) in the boundary layer (the upper troposphere). Observed O₃ shows a maximum near 900 hPa and decreases toward the lower free troposphere, while the simulation slightly underestimates it, except near the surface. Observed HO₂ and CH₂O decrease with altitude, reflecting the decrease in water vapor (Heikes, 1992) and the boundary layer source from oxidation of isoprene (Millet et al., 2006), respectively. The simulation captures the observed features of CH₂O well, but overestimates HO₂ by 10 pptv throughout the troposphere.

The data assimilation improves the agreement with the aircraft observations for NO₂, O₃, and PAN. Underestimations of these species concentrations are generally reduced by the data assimilation. Chemical production of O₃ is strongly related to the abundance of NO_x and OH. The NO_x emissions tend to increase OH via the NO and HO₂ reaction and the O₃ (and excited oxygen atoms) and H₂O reaction, while the CO emissions tend to decrease OH. Corresponding to the increased NO_x emissions and decreased CO emissions

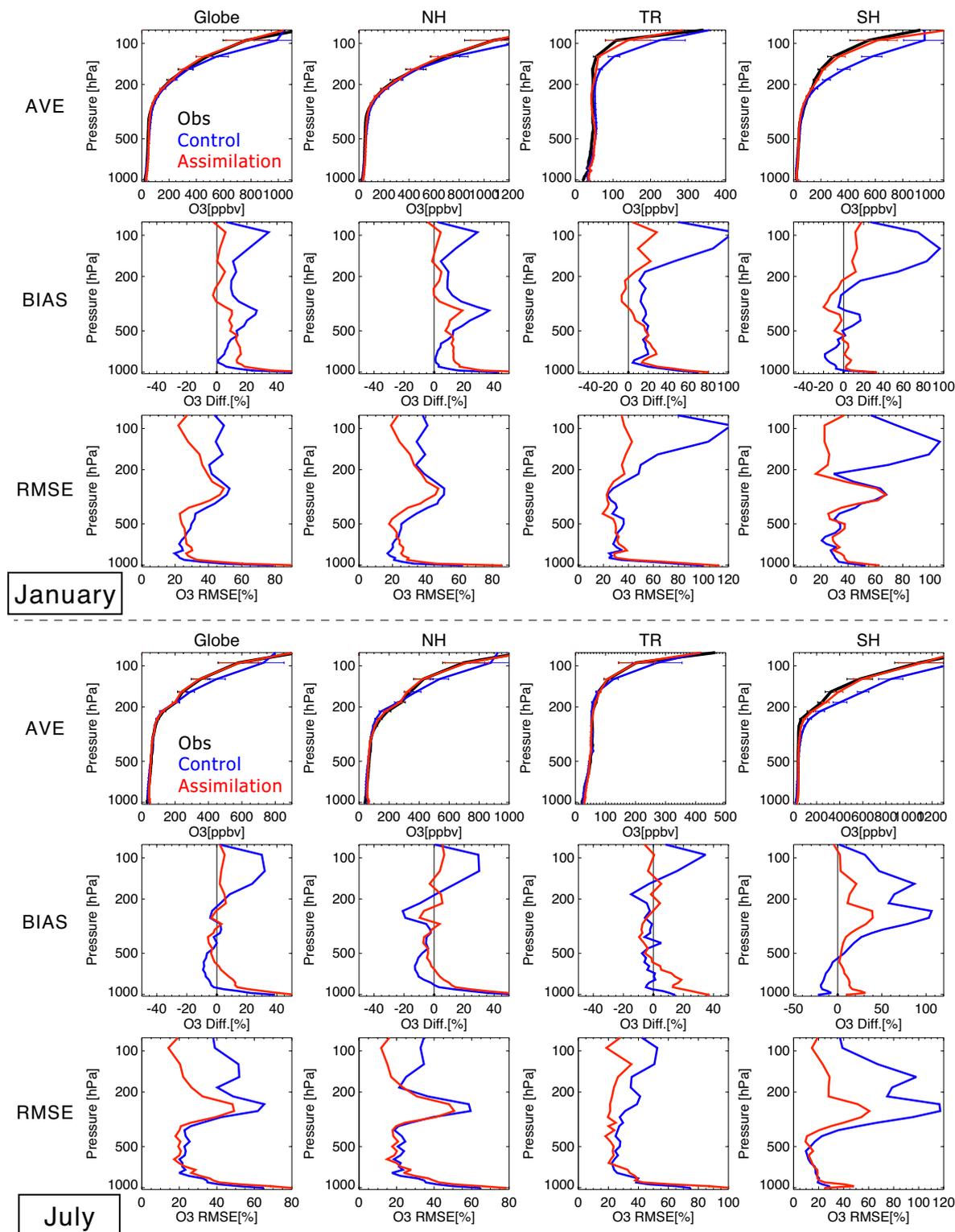


Fig. 11. Comparison of the vertical O₃ profiles between ozonesondes (black), the control run (blue), and the data assimilation (red) during 7–30 January 2007 (upper 9 panels) and 7–30 July 2007 (lower 9 panels). Upper row of each panel shows the mean profile; center and lower rows of each panel show the mean difference and the RMSE between the control run (the data assimilation run) and ozonesondes in blue (red).

Table 6. The 15-day means (from the 16–30 of each month) and the standard deviations (from the means) of the global and regional surface NO_x emissions ($e(\text{NO}_x)$, in TgNyr^{-1}), lightning NO_x emissions ($e(\text{LNO}_x)$, in TgNyr^{-1}), and surface CO emissions ($e(\text{CO})$, in TgCOyr^{-1}) obtained from the a priori emissions and the a posteriori emissions. GL is global (90°S – 90°N); NH is the Northern Hemisphere (20°N – 90°N); TR is the tropics (20°S – 20°N); and SH is the Southern Hemisphere (90°S – 20°S). The emissions optimized from the full assimilation run and the emission inversion run (in bracket) are presented.

		January		July	
		A priori	A posteriori	A priori	A posteriori
$e(\text{NO}_x)$	NH	24.5	24.6±0.4 (23.1±0.8)	31.5	32.3±1.3 (33.9±2.2)
	TR	15.4	14.4±0.6 (15.5±0.6)	12.9	16.7±0.6 (19.7±1.0)
	SH	3.0	3.9±0.2 (3.7±0.1)	2.4	3.0±0.1 (3.6±0.2)
	GL	42.8	42.9±0.6 (42.3±1.3)	46.7	52.0±1.4 (57.3±2.9)
$e(\text{CO})$	NH	544.8	660.7±10.5 (906.2±25.6)	720.8	630.2±30.5 (901.0±87.4)
	TR	495.2	358.8±49.0 (440.4±50.1)	443.6	307.6±8.7 (374.8±22.2)
	SH	56.3	40.8±4.0 (69.7±3.5)	53.0	95.5±9.2 (93.5±12.9)
	GL	1096.3	1060.3±58.5 (1416.3±37.7)	1217.4	1033.3±26.9 (1369.3±62.4)
$e(\text{LNO}_x)$	NH	0.6±0.2	0.8±0.3	2.9±0.8	3.4±0.9
	TR	2.8±0.5	3.3±0.7	2.5±0.5	3.6±0.7
	SH	1.0±0.3	1.1±0.3	0.3±0.0	0.3±0.1
	GL	4.4±0.7	5.2±0.9	5.7±0.9	7.3±1.1

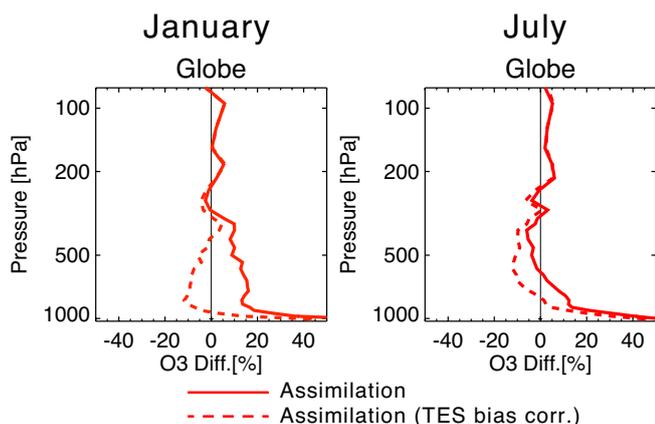


Fig. 12. The mean relative difference of the vertical O_3 profiles between ozonesondes and the data assimilation with (red dashed) and without (red solid) the bias correction for TES O_3 data during 7–30 January 2007 (left) and 7–30 July 2007 (right).

at low latitudes, the data assimilation increases OH and O_3 . Because of the low sensitivity of TES in the lower troposphere, the changes in near surface O_3 are largely attributed to the change in NO_x emissions, as will be further discussed in Sect. 5.3. The assimilated fields still underestimate the concentrations of NO and PAN in the free troposphere, and overestimate HO_2 throughout the troposphere. Martin et al. (2007) concluded that oxidation of lightning NO_x explains nearly 80 % of the HNO_3 concentration in the tropical upper troposphere. Increasing the lightning NO_x source also decreases HO_2 in the upper troposphere, while increasing OH (Hudman et al., 2007). The increased NO_x results in a reduction of the HO_2/OH ratio through the $\text{NO} + \text{HO}_2$ and

$\text{NO}_2 + \text{HO}_2$ reactions, and also results in an increased loss of OH via production of HNO_3 (DeCaria et al., 2005). Thus, it is likely that more NO_x sources in the free troposphere are required to reduce the negative bias of NO_2 , NO, PAN, and HNO_3 and the positive bias of HO_2 in the free troposphere. The overestimated H_2O may also contribute to the overestimation in the concentrations of OH and other HO_x species through its reaction with excited oxygen atoms. The data assimilation tends to increase the overestimation in HNO_3 concentration in the boundary layer, corresponding to the increased NO_2 concentration. Simultaneous adjustments for its removal processes (e.g. wet and dry depositions) might be important to further improve the analysis. Removal of HNO_3 by wet deposition processes occurs within a few days in the lower troposphere and results in the loss of HO_x species, which may also explain a part of the overestimation in HO_x species concentrations. Meanwhile, a large uncertainty in both observed and simulated OH concentrations in the free troposphere remains an important issue (e.g. Hudman et al., 2007). There are many other factors in the chemical transport processes affecting the overall model performance. They may obstruct further improvements by the data assimilation.

5.2 Estimated emission sources

The simultaneous optimization of multiple species leads to complex chemical interactions which together determine the estimated emissions. Especially, the imperfect representation of OH fields may cause large uncertainties in the NO_x and CO emissions inversion (Müller and Stavrou, 2005; Jones et al., 2009; Pison et al., 2009; Hooghiemstra et al., 2011). Müller and Stavrou (2005) demonstrated that the optimization of CO emissions constrained by both CO and

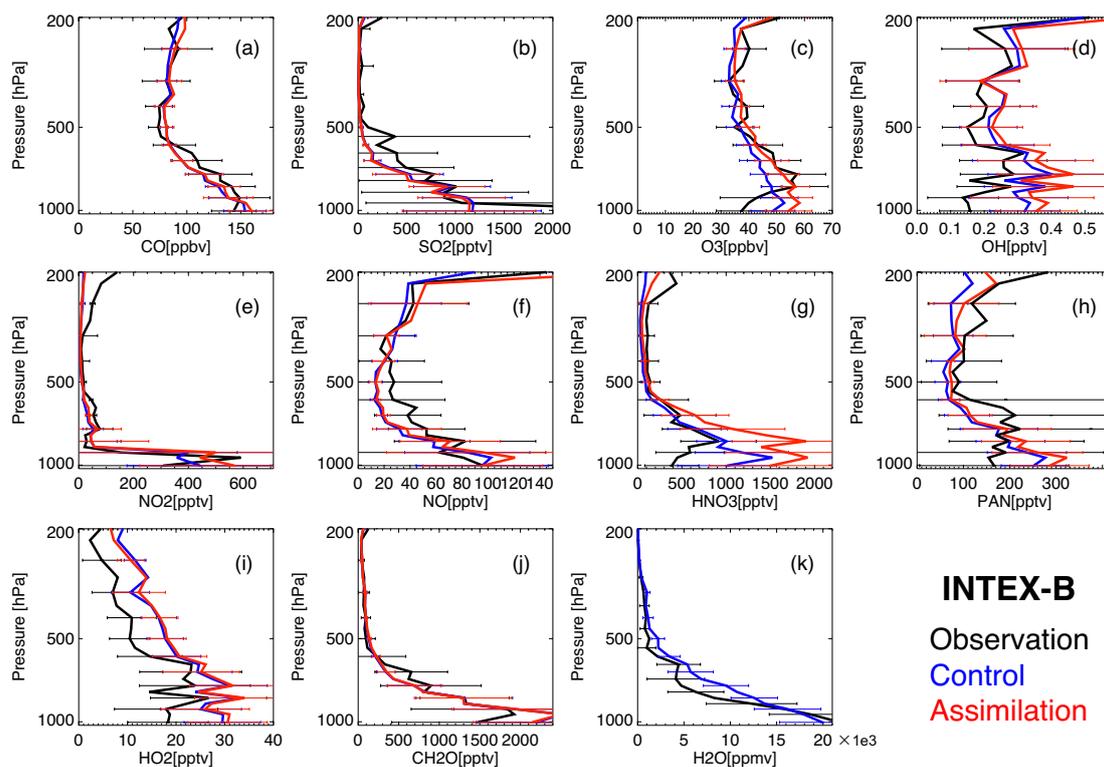


Fig. 13. Mean vertical profiles of (a) CO, (b) SO₂, (c) O₃, (d) OH, (e) NO₂, (f) NO, (g) HNO₃, (h) PAN, (i) HO₂, (j) CH₂O, and (k) H₂O, obtained from aircraft measurements (black), the control run (blue), and the data assimilation (red), during the INTEX-B campaign, March 2007. The error bars represent the standard deviation of all the data within one bin (with an interval of 30 hPa).

Table 7. The 15-day means (from the 16–30 of each month) and the standard deviations (from the means) of the regional surface NO_x emissions ($e(\text{NO}_x)$, in Tg N yr^{-1}) for Eastern China (110–123° E, 30–40° N), Europe (10° W–30° E, 35–60° N), the Eastern United States (95–71° W, 32–43° N), South America (70–50° W, 20° S–Equator), Northern Africa (20° W–40° E, Equator–20° N), Central Africa (10–40° E, 20° S–Equator), Southern Africa (26–31° E, 28–23° S), and Southeast Asia (96–105° E, 10–20° N) for January and July in 2007. The regional emissions obtained from the a priori emissions, the newer inventories (EDGAR version 4.2, GFED version 3.1, and GEIA), and those optimized from the full assimilation run and the emission inversion run (in bracket) are presented.

	January			July		
	A priori	A posteriori	newer inventories	A priori	A posteriori	newer inventories
Europe	5.2	3.3 ± 0.2 (3.7 ± 0.3)	4.5	6.7	5.6 ± 0.6 (5.9 ± 1.1)	4.6
E-USA	2.4	2.8 ± 0.2 (2.5 ± 0.3)	2.5	3.0	2.2 ± 0.3 (3.1 ± 0.3)	2.5
C-Africa	0.7	1.0 ± 0.0 (1.3 ± 0.0)	0.6	4.9	4.9 ± 0.4 (5.9 ± 0.5)	3.5
E-China	2.9	4.8 ± 0.3 (4.3 ± 0.2)	3.6	3.5	4.8 ± 0.7 (4.5 ± 0.7)	3.6
S-Africa	0.2	0.6 ± 0.1 (0.6 ± 0.0)	0.3	0.3	0.3 ± 0.1 (0.3 ± 0.1)	0.3
S-America	0.4	0.6 ± 0.1 (0.6 ± 0.0)	0.5	0.9	0.6 ± 0.1 (0.9 ± 0.1)	0.8
N-Africa	7.4	5.4 ± 0.1 (6.1 ± 0.5)	4.8	1.6	2.8 ± 0.2 (3.2 ± 1.2)	1.4
SE-Asia	0.5	0.9 ± 0.1 (0.7 ± 0.1)	0.9	0.3	0.4 ± 0.0 (0.5 ± 0.1)	0.3

NO₂ observations leads to a better agreement between modeled and observed values, through the chemical response of the CO-OH-NO_x-NMHC system to emission changes. Jones et al. (2009) also suggests that neglecting the influence of NO_x emissions on the CO chemistry could contribute to a significant bias in the CO source estimates.

In our system, as shown in Figs. 4 and 5 and discussed in Sect. 4.2, all the assimilated data significantly influence concentrations of OH, NO₂, and CO. The assimilation of OMI NO₂ data generally increases (decreases) the OH concentration in the tropics (extratropics) by 15 %, which affects the atmospheric CO lifetime and influence the CO emission inversion. Meanwhile, the higher CO emissions lead to a decrease

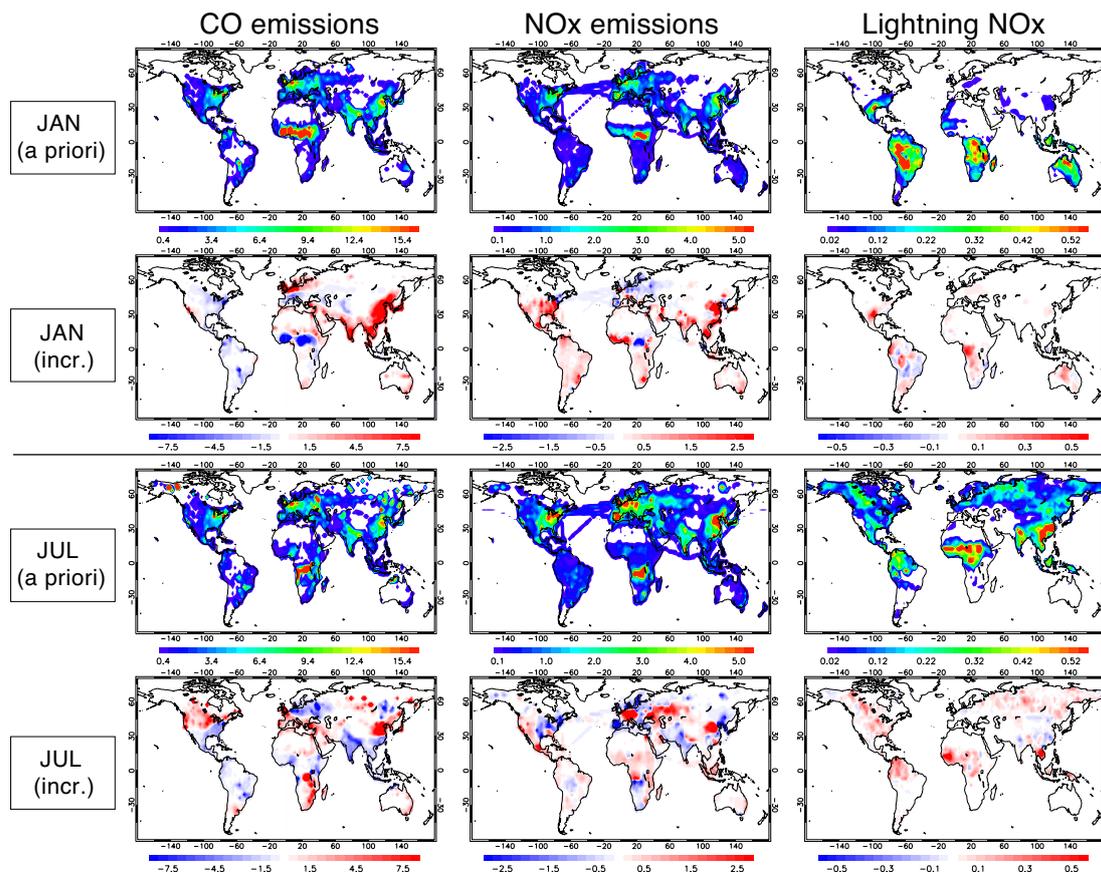


Fig. 14. Global distributions of the surface CO emissions (in $10^{-10} \text{ kg m}^{-2} \text{ s}^{-1}$) (left panels), the surface NO_x emissions (in $10^{-11} \text{ kg m}^{-2} \text{ s}^{-1}$) (centre panels), and the lightning NO_x emissions (in $10^{-11} \text{ kg m}^{-2} \text{ s}^{-1}$) (right panels), averaged over the 16–30 January (upper 6 panels) and July (lower 6 panels) in 2007. The a priori emissions (upper rows) and the analysis increment (lower rows), i.e. the difference between the a posteriori and the a priori, are shown for each panel. The red (blue) colour indicates an emission increase (decrease) for the analysis increment, respectively.

Table 8. Same as in Table 7, but for the regional surface CO emissions ($e(\text{CO})$, in Tg CO yr^{-1}).

	January			July		
	A priori	A posteriori	newer inventories	A priori	A posteriori	newer inventories
Europe	100.9	101.6 ± 4.0 (156.0 ± 5.3)	31.1	108.7	78.5 ± 4.4 (137.5 ± 21.1)	37.5
E-USA	55.4	23.4 ± 1.7 (47.5 ± 8.6)	28.5	60.3	22.8 ± 1.6 (39.4 ± 8.5)	28.6
C-Africa	21.9	16.5 ± 0.5 (17.1 ± 0.8)	15.4	206.5	201.3 ± 7.9 (252.4 ± 18.4)	212.8
E-China	66.8	175.3 ± 5.9 (198.0 ± 10.7)	62.4	66.8	115.7 ± 7.4 (131.0 ± 8.9)	62.5
S-Africa	2.6	1.7 ± 0.1 (3.7 ± 0.4)	2.0	3.6	11.3 ± 1 (11.3 ± 1.4)	5.7
S-America	13.0	6.2 ± 0.5 (5.8 ± 0.5)	4.1	16.6	5.5 ± 0.3 (5.8 ± 0.7)	39.0
N-Africa	306.1	174.4 ± 36.6 (177.9 ± 50.2)	292.7	59.1	28.6 ± 1.0 (39.2 ± 1.2)	33.8
SE-Asia	15.0	44.2 ± 5.3 (54.4 ± 0.6)	62.2	8.4	3.0 ± 0.3 (2.7 ± 0.3)	12.0

in OH abundances and slightly increases NO₂ concentration in the extratropics. The simultaneous data assimilation thus provides comprehensive constraints on the emission inversion. It is expected that the simultaneous data assimilation provides a better estimate of the emissions than the inversion run because the concentration assimilation may reduce some

of the model errors. However, this will not be the case for all model errors. For instance, errors in boundary layer venting or deposition may be compensated in our assimilation system by (incorrectly) changing the emissions. The a priori and a posteriori emissions estimated from data assimilation are shown in Fig. 14 and listed in Table 6, 7, and 8. Note

that the a priori surface emissions for the simulation years 2006–2007 were obtained by linearly temporal extrapolating the 1995 and 2000 inventories.

5.2.1 NO_x emissions

The data assimilation changes the global total NO_x emissions from 42.8 to 42.9 TgNyr⁻¹ in January and from 46.7 to 52.0 TgNyr⁻¹ in July. The a posteriori and the a priori emissions differ more significantly at the regional scale. The analysis increment is generally positive over Eastern China, North America (only in January), Australia, Northern India (only in January), Southeast Asia, and Southern Africa. An obvious increment is observed over Eastern China, with a factor of up to about 1.6 in January. Over the Eastern United States, the a posteriori emissions are higher than the a priori emissions in January, but are lower in July. The a posteriori emissions are lower than the a priori emissions over Europe, unlike over other industrial areas. Over Central Africa, the data assimilation increases the emissions in January. Over Northern Africa, the data assimilation decreases the emissions in January, but increases the emissions in July. Most of these features are also reported in Miyazaki et al. (2012). As a result of the data assimilation and the covariance inflation, the mean a posteriori error for the surface NO_x emissions typically ranges from 12 to 60 %, with smaller relative errors over polluted areas than over clean areas. The mean differences between the a priori and the a posteriori emissions are generally larger than both the a posteriori error and the variability (i.e., standard deviation) of the a posteriori emissions estimated during the analysis period.

The analysis increment structures obtained from the data assimilation strongly depend on the assumption made on the a priori emission. In CHASER, the 1995 and 2000 emission inventories are extrapolated to the simulation years 2006–2007. This procedure may give spurious results for certain regions, as described in Miyazaki et al. (2012). However, the bottom-up emissions obtained from the newer inventories (EDGAR version 4.2 (European Commission, 2011), GFED version 3.1, and GEIA) for the year 2007 show a similar difference with the a posteriori emissions. This indicates common problems in the emission inventories (e.g. too little emissions over Eastern China, the Eastern United States in January, Central Africa in January, Northern Africa in July, and Southern Africa in January, and too much emissions over Europe in January, the Eastern United States in July, South America in July). Note that the a posteriori emissions are closer to the newer inventories than the a priori emissions in some cases (e.g. over the Eastern United States, Eastern China, Europe, and Southeast Asia). In particular, the a priori emissions in Spain are unrealistically high, which are very different from both the a posteriori emissions and the newer inventories.

The simultaneous data assimilation system results in NO_x emissions somewhat different from the emission inver-

sion system in which only surface emissions are optimized (brackets in Tables 6 and 7). This indicates that the direct adjustment to the concentration fields by the data assimilation provides important effects upon the emission inversion, with a regional difference of up to 40 % over industrial areas and up to 30 % over biomass burning areas. For instance, the emissions over Central Africa in the simultaneous data assimilation are smaller than in the emission inversion, which is attributed to the increased NO₂ concentration in the middle and upper troposphere mainly due to the assimilation of TES O₃ data. The smaller emissions over the Eastern United States for July in the simultaneous data assimilation results from the larger NO₂ concentrations in the middle and upper troposphere, primarily by the adjustment made directly to the concentrations due to the assimilation of OMI NO₂ data.

The January and July mean global surface NO_x emissions of 47.4 TgNyr⁻¹ estimated from the data assimilation is slightly larger than the annual mean emissions estimated from previous studies (e.g. 42.1 TgNyr⁻¹, Müller and Stavrou, 2005, 40.3 TgNyr⁻¹, Jaeglé et al., 2005, 45.4 TgNyr⁻¹, Miyazaki et al., 2012). Differences in analysis years and the focus on only two months may primarily contribute to the difference in NO_x emission estimates. The NO_x emissions are generally larger over industrial areas in winter and over soil/desert areas in summer than in other seasons; this may also contribute to the larger NO_x emissions estimated from this study compared to the annual mean emissions. Meanwhile, the comparison against the January and July mean a priori emissions (44.7 TgNyr⁻¹) and the newer inventories (40.4 TgNyr⁻¹) implies general underestimations in the emission inventories. On the regional scale, the 11.0 TgNyr⁻¹ estimated over East Asia (80–150° E, 10–50° N) for July 2007 from OMI observations (Zhao and Wang, 2009) is comparable to our estimates of 10.2 TgNyr⁻¹. The 0.465 TgN estimated over the Eastern United States (102–64° W, 22–50° N) from the OMI observations for March 2006 (Boersma et al., 2008a) is also comparable to our estimate of 0.485 TgN.

5.2.2 CO emissions

Because of the long chemical lifetime of CO in the troposphere, the CO emission inversion requires an assimilation cycle with a long assimilation window (i.e. by using the 4D-Var assimilation technique (e.g. Hooghiemstra et al., 2011)) in order to obtain enough constraints from observations. The CO emissions estimated from this study, based on one-month calculation, may not have been sufficiently constrained by the observations. Further, the simultaneous data assimilation corrects the CO concentrations from the MOPITT data obtained at the 9 pressure levels and at the surface, whereas the emissions are optimized using the data only obtained at 700 hPa (see Sect. 2.1.3). Consequently, the simultaneous data assimilation system can start by adjusting concentrations, and then the emissions will adjust more slowly,

depending on the averaging kernel profile and the DOFs. The estimated CO emissions are thus more strongly constrained by the observations in the emission inversion run than in the full assimilation run. Therefore only the CO emissions estimated by the emission inversion system are presented, as depicted in Fig. 14 and listed in the brackets in Tables 6 and 8.

CHASER shows a large underestimation in the simulated CO fields in the northern extratropics, as commonly revealed by many CTMs (Shindell et al., 2006). The underestimated CO fields might be mostly attributed to an underestimation of the surface CO emissions along with an overestimation of OH. Correspondingly, the assimilation of MOPITT data largely increases the surface CO emissions in the northern extratropics both in January (+66 %) and July (+25 %). The large increase in the CO emissions are mainly attributed to the increase over industrial areas, especially over Eastern China in the both seasons with a factor of 2–3. The large positive increment is consistent with the results of Arellano et al. (2004), who showed that anthropogenic emissions over Asia are too low in EDGAR v3.2. The decreased emissions over North Africa and the increased emission over Australia and South Asia (especially in January) are also consistent with recent estimates (Jones et al., 2009; Fortems-Cheiney et al., 2011). The large increments obtained for Central and North Africa indicate a large uncertainty in biomass burning in the GFED2 inventory, as similarly suggested by Kopacz et al. (2010). The larger emissions in winter than in summer in the US, Europe, and East Asia are also commonly revealed from recent inversions, which could be due to a combination of emissions from residential heating and vehicle cold starts (e.g. Kopacz et al., 2010). The newer inventories show lower emission values than the a priori emissions over Europe, whereas the data assimilation further increases the emissions from the a priori emissions. In contrast, the a posteriori emissions are significantly larger than both the a priori emission and the newer inventories over Eastern China. Over Eastern United States, the data assimilation decreases the emissions; however, the newer inventories show even lower emissions. These results imply different error characteristics in the different bottom-emission inventories.

Our a posteriori January and July mean estimate for the surface CO emission is 1393 Tg yr^{-1} , which is about 20 % higher than the a priori emissions, mainly due to increased emissions by up to 60 % in the Northern Hemisphere in January. This is within 10 % of the results from previous estimates of $1342\text{--}1502 \text{ Tg CO yr}^{-1}$ (Arellano et al., 2004; Arellano and Hess, 2006), $1390 \text{ Tg CO yr}^{-1}$ (Hooghiemstra et al., 2011), $1391 \text{ Tg CO yr}^{-1}$ (Pison et al., 2009), $1393 \text{ Tg CO yr}^{-1}$ (Kopacz et al., 2010), $1440 \text{ Tg CO yr}^{-1}$ (Jones et al., 2009), $1504 \text{ Tg CO yr}^{-1}$ (Fortems-Cheiney et al., 2011). The a posteriori emissions are also much larger than the newer inventories (with the January and July mean global emission of $892 \text{ Tg CO yr}^{-1}$).

5.2.3 Lightning NO_x sources

The data assimilation provides strong constraint on the magnitude and the distribution of LNO_x. The global LNO_x amount is increased from 4.4 to 5.2 Tg N yr^{-1} in January and from 5.7 to 7.3 Tg N yr^{-1} in July. The estimated emissions are within the range of the annual global LNO_x source of $5 \pm 3 \text{ Tg N yr}^{-1}$ by Schumann and Huntrieser (2007). The large increase in July corresponds to the significant increase over the Eurasian continent, North America, Southeast Asia, the tropical South America, and Central Africa. The data assimilation also changed the vertical profile of the LNO_x sources both in the tropical and extratropics. The large changes in the three-dimensional distribution of LNO_x obtained from the data assimilation indicate that the Price and Rind (1992) lightning parameterization used in the simulation does not fully capture the observed distribution of lightning activity, as also suggested by Allen and Pickering (2002). In particular, the data assimilation generally increases LNO_x in the upper troposphere both in the tropics and the extratropics. This suggests that the C-shape vertical profile of lightning NO_x assumed in the parameterization may place too much mass near the surface and too little in the middle troposphere, as also suggested by Ott et al. (2010). Especially, TES O₃ and OMI NO₂ data provided particularly strong constraints on LNO_x sources. However, a lightning signal in satellite observations of NO₂ columns is often obscured by the high contributions from (anthropogenic) boundary layer pollution and biomass burning (Martin et al., 2002, 2007; Boersma et al., 2005). Also the increase of the NO/NO₂ ratio with height in the troposphere reduces the relative sensitivity to lightning produced NO₂. The lightning signal is also almost comparable to the measurement uncertainty for tropospheric NO₂. Thus, further careful considerations are required for LNO_x estimates, which will be discussed in a separate study.

5.3 Relative importance of the emission and concentration optimization on the tropospheric O₃ analysis

As a result of the simultaneous optimization of the emissions and the concentrations, the global tropospheric O₃ burden, which is calculated for the region below the tropopause height determined from the vertical temperature gradient (-2 K km^{-1}) in the model, is decreased by 2.5 % in January and is increased by 5.6 % in July in the full assimilation run (Table 9). The obvious increase in July, from 346.8 to 366.1 Tg O_3 , is almost equally attributed to the enhanced emissions of the precursors and the direct adjustment to the concentration fields. The simultaneous optimization provides important contributions to the tropospheric ozone budget analysis, by optimizing its precursor emissions and reducing model errors while taking the chemical feedback into account.

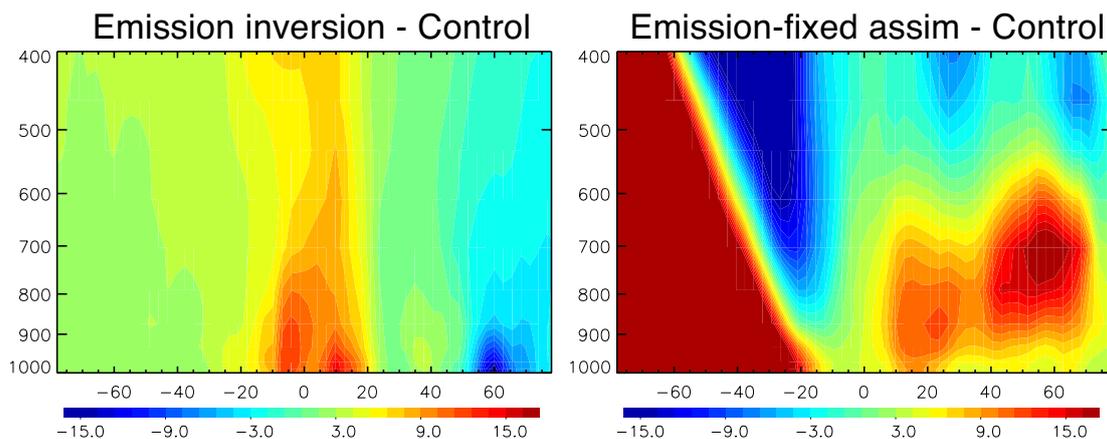


Fig. 15. The latitude-pressure distribution of the relative difference of zonal mean O_3 mixing ratio (in %) between the emission inversion run and the control run (left) and the emission-fixed assimilation and the control run (right) averaged over 16–30 July 2007. The red (blue) colour indicates relatively high (low) values in the inversion/assimilation run.

Figure 15 shows the relative importance of the emission inversion and the direct concentration assimilation on the vertical O_3 profiles. The emission inversion largely changes the O_3 profiles in the PBL, especially below 900 hPa. This demonstrates the importance of optimizing O_3 precursor fields in correcting the near surface O_3 . The obvious impact in the PBL, with a mean difference of up to 15 %, is found in the tropics and at northern mid-latitudes in July, associated with changes in biomass burning and anthropogenic emissions, respectively. The regional differences are more pronounced over the northern mid-latitudes polluted regions, central Africa, and south America in July, with a maximum difference of 30 % (figure not shown). Even in the free troposphere, the O_3 analysis is significantly affected by the emission changes through vertical transport of O_3 and its precursors. However, the direct concentration adjustment dominates the changes in the O_3 profiles in the free troposphere in the combined data assimilation. The simultaneous adjustment of the emissions and the concentrations is thus a powerful approach to optimize the whole tropospheric O_3 profiles.

The sum of these two individual effects mostly explains the difference between the full assimilation run and the control run (figure not shown). The O_3 changes in the tropical troposphere are an exception to this rule in that the changes estimated from the sum of these two individual effects are slightly (~ 15 %) larger than the changes estimated from the full assimilation run. This may indicate too large emission adjustments and resultant O_3 productions in the emission inversion run. The spatial pattern of the changes in the O_3 profiles obtained from emission inversion run and the fixed-emission assimilation run is very different (figure not shown), confirming the independent adjustments realized from the emission and concentration optimizations.

Table 9. The 1-day average (on the 19th of each month in 2007) global tropospheric O_3 burden (TgO_3) obtained from the control run, the emission inversion run, and the full data assimilation run.

	January	July
Control	317.1	346.8
Emission inversion	312.5	356.6
Full assimilation	309.4	366.1

5.4 Uncertainties

The EnKF data assimilation provides information about the uncertainty of the analysis. The ensemble spread, estimated as the standard deviation of the simulated concentrations across the ensemble, is a measure of the analysis uncertainty (e.g. Arellano et al., 2007). The uncertainty in the a posteriori fields represented by the analysis spread is reduced if the analysis converges to a true state. This spread is caused by errors in the model input data, chemical or physical parameters, parameterizations, the numerical scheme as well as errors in the measurements assimilated (Boynard et al., 2011).

Figure 16 shows the distributions of the analysis spread for O_3 , CO , and NO_2 . The analysis spread typically shows a reduction of the analysis errors due to effective (high quality, high sensitivity, good coverage) observations and an increase due to error growth as represented by the ensemble model forecast and the covariance inflation. Near the surface, the analysis spread of O_3 and CO is generally smaller in the tropics than in the extratropics, corresponding to the latitudinal dependence of TES O_3 and MOPITT CO retrieval sensitivities, respectively. The vertical profile shows that the analysis spread is effectively reduced in the middle troposphere, reflecting the maximum sensitivity of the retrievals at these altitudes. Within the free troposphere, the O_3 analysis

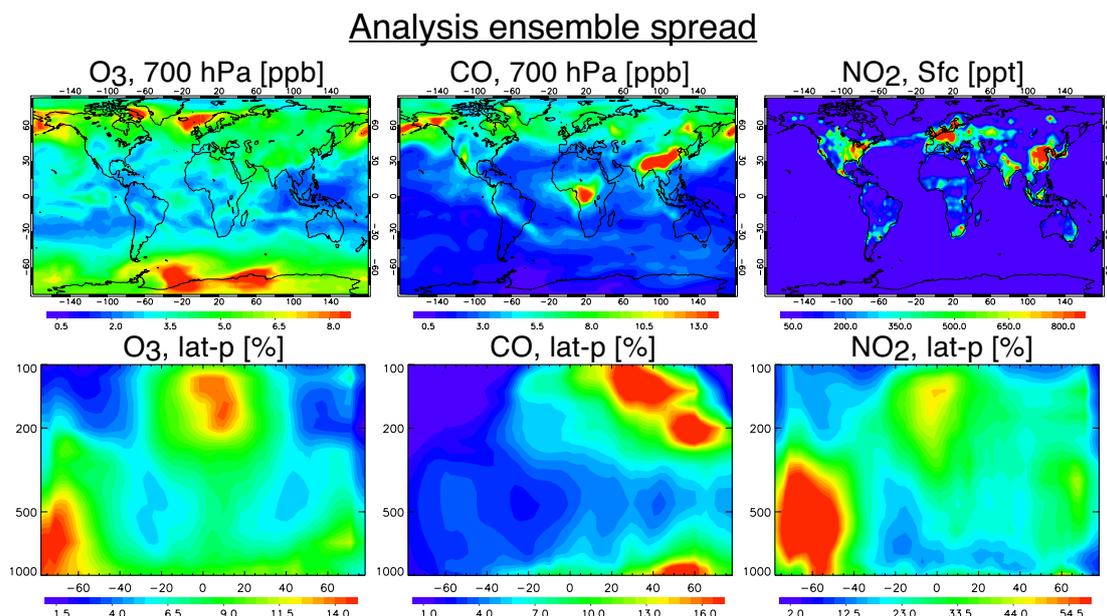


Fig. 16. Analysis ensemble spread of O₃ (left), CO (centre), and NO₂ (right) averaged over 16–30 July 2007. Upper panels show the global distribution at 700 hPa. Lower panels show the latitude-pressure distribution of the percentage ratio of the zonal mean analysis ensemble spread to the zonal mean analysis ensemble mean concentration. The red (blue) colour indicates relatively high (low) values.

spread is relatively high in the tropical upper troposphere. The OSEs showed that the assimilation of MLS HNO₃ data acts to increase the O₃ analysis spread in the tropical upper troposphere during the forecast, through its influence on the NO_y species fields during the analysis and because of its large observation errors. In the extratropical upper troposphere and around the subtropical jet streams, the downward propagation of the well constrained O₃ due to the assimilation of MLS O₃ data helps to reduce the analysis spread. The CO analysis spread is maximum at the northern mid-latitudes near the surface, related to large uncertainties in the analyzed CO emissions. The analysis spread of NO₂ is closely related to the emissions near the surface, while it also has strong latitude-vertical dependence. The analysis spread is generally maximum in the upper troposphere, which is a result of the low concentrations which are not well constrained by the OMI data. The large analysis spread in the southern extratropics is related to large relative observation errors of OMI NO₂ data related to the low concentrations. These large analysis spreads indicate requirements for further constraints from additional observations or higher quality data.

The assimilation system can also be used to diagnose model and/or observation errors. We use the difference between analysis and forecast, the so-called analysis increment, to represent short-term systematic errors in the model (Fig. 17). By assuming that the assimilated fields approximate the assimilated data after several assimilation cycles, the averaged analysis increment primarily relates to the persistent model bias. The increment thus represents the adjustment made in the analysis step to bring the model closer to

the observations, and the spatial distribution of the averaged increments shows where the model fields are frequently adjusted by the data assimilation. The positive increments obtained for O₃ and CO in the extratropical lower troposphere imply that CHASER tends to underestimate those concentrations in these regions compared to the assimilated data. Positive increments of O₃ are frequently observed over the northern Eurasian continent, around North America, and over the Southern Ocean. The OSEs confirms that the positive O₃ analysis increments in the lower and middle troposphere are due to the TES O₃ data, while the upper tropospheric negative increments are due to both the TES and MLS O₃ data. This implies that the model bias strongly varies with height, because of different contributions of transport and chemical processes. The data assimilation also tends to increase CO over East China and North America near the surface, as similarly shown by Elguindi et al. (2010), whereas it decreases CO in the tropics. The negative NO₂ increments in the tropical and high-latitudes troposphere are associated with the assimilation of very small or negative OMI NO₂ concentrations mainly over the oceans. The large positive NO₂ increments obtained for the extratropical UTLS reflects the fact that the assimilation of MLS O₃ and HNO₃ data tends to compensate for the model underestimation through the inter-species correlation. The knowledge of the model error structure is useful to identify sources of the model error. Geer et al. (2006) showed that the enhanced skill of the best performing analysis can usually be attributed to better modeling.

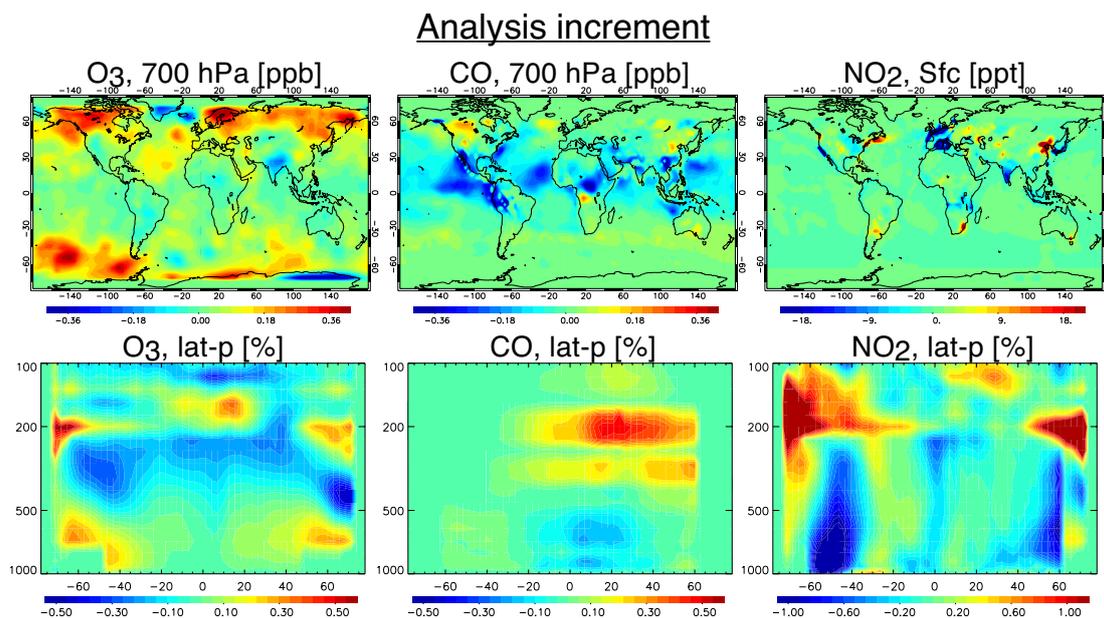


Fig. 17. Same as Fig. 16, but for the analysis increment. Lower panels show the latitude-pressure distribution of the percentage ratio of the zonal mean analysis increment to the zonal mean analysis ensemble mean concentration. The red (blue) colour indicates relatively positive (negative) values.

6 Conclusions

An advanced data assimilation system for tropospheric chemical compositions, the CHASER-DAS, is developed based on the CHASER model and the LETKF scheme. The data assimilation system is applied to integrate observation information obtained from multiple satellite measurements, namely, NO₂ data from OMI, O₃ data from TES, CO data from MOPITT, and O₃ and HNO₃ data from MLS. The data assimilation provides multiple constraints on tropospheric composition and allows us to simultaneously optimize the atmospheric distributions of various chemical compositions together with the emissions of O₃ precursors (NO_x and CO) while taking their chemical feedbacks in the CO-OH-NO_x-O₃ system into account. In the simultaneous data assimilation system, improved atmospheric concentrations of chemically-related species have the potential to improve the emission inversion, while the improved emissions estimates will benefit the atmospheric concentration analysis through a reduction in the model forecast error. A covariance localization technique is applied to neglect the covariance among non-related or weakly-related variables which may suffer significantly from errors in the ensemble sampling and the forecast model.

The improvement obtained by the assimilation demonstrates that multi-species data assimilation provides valuable information on various chemical fields. The OmF analysis confirmed significant error reductions for both bias and RMSE from the data assimilation. The standard deviation around the mean of the OmF is generally comparable to the

observation error, indicating that the data assimilation is successfully performed. Significant reductions of both bias (by 85 %) and RMSE (by 50 %) against independent data sets for various chemical fields show that multi-species data assimilation is a very effective way of combining observation information and compensating for systematic model errors. The improvements include enhanced tropospheric NO₂ columns over industrial areas (with a global mean bias reduction of 40–85 %), especially over China, reduced positive O₃ bias in the middle and upper troposphere (by 60 %), reduced negative CO bias in the Northern Hemisphere in the lower troposphere (by 40–90 %), especially over East Asia and North America, and a reduced negative HNO₃ bias in the extratropical UTLS (by 70–85 %). Comparisons against ozonesonde and aircraft data confirmed improvements in the vertical profiles of O₃ and its precursors in the free troposphere and the UTLS through the data assimilation. The data assimilation removes most of the bias from the middle troposphere to the lower stratosphere against ozonesonde data, from 30–40 % to within 10 %. The results confirm that the assimilated satellite data have highly valuable information about the tropospheric chemical processes, although further improvements are required for the lower tropospheric processes.

OSEs have been conducted to quantify the relative importance of each data set on constraining the emissions and concentrations. The assimilation of each individual dataset has a strong influence on both assimilated and non-assimilated species through the use of inter-species error correlations and through the chemical model. For instance, the assimilation of upper tropospheric O₃ and HNO₃ obtained from MLS was

useful to reduce the bias in the tropospheric NO_2 columns. Comparisons against independent ozonesonde data showed that both MLS and TES O_3 data largely improve the O_3 profiles in the free troposphere and the UTLS. Note that all the assimilated data contribute to the global mean O_3 bias reduction compared to ozonesonde data in the middle troposphere (between 750 and 450 hPa) in July, through their influences on various chemical states that affect O_3 variations. Especially this last result demonstrates the strength of the simultaneous assimilation of multiple datasets for different species. These inter-species influences can be tightly associated with the changes in OH. The simultaneous assimilation increased tropospheric OH concentrations in July by 5–15 % in the tropics and the Southern Hemisphere mainly due to the assimilation of OMI NO_2 and TES O_3 data, respectively. The large improvement in July may be related to summertime active chemical processes in the Northern Hemisphere.

In comparison to the a priori emissions based on bottom-up inventories (EDGAR3.2+GFED2.1+REAS1.1), the optimized emissions of both NO_x and CO are generally higher over most industrial areas, especially in the northern mid-latitudes, implying that the emission inventories underestimate sources. The NO_x emissions estimated from the simultaneous data assimilation are different from those from the emission inversion system in which only the emissions are optimized from observations. The results indicate a large uncertainty in the a posteriori NO_x emissions due to model errors when estimating from NO_2 data only, with an uncertainty of up to 40 % over industrial areas and up to 30 % over biomass burning areas, as measured by the impact of the concentration assimilation on the a posteriori emissions. The simultaneous assimilation of multiple chemical observations is very useful to represent the chemical processes in a realistic way by removing model errors, and it provides important effects upon the emission inversion. The CO emissions estimated in this study may not have enough constraint from observations, because the calculation period is too short and the observational information is insufficient. Nevertheless, comparison of our results to previous inverse modeling studies (e.g. Kopacz et al., 2010) is very encouraging. The uncertainties in the a priori emissions, based on an extrapolation of year 1995 and 2000 inventories, caused large increments especially over anthropogenic source areas. The data assimilation also increases the lightning NO_x sources over land, especially in boreal summer, indicating that the lightning parameterization used in the simulation has a large uncertainty.

As a result of the simultaneous optimization, the tropospheric O_3 burden is increased by 5.6 % in July, with almost equal contributions from the emission optimization and the direct adjustment to the concentration fields. The emission optimization dominated the changes in the O_3 profiles in the PBL in the tropics and at northern mid-latitudes, whereas the direct concentration adjustment was much more important in the free troposphere. This reveals the importance of the si-

multaneous adjustment of the emissions and concentrations for the tropospheric ozone budget and profile analyses.

7 Discussions: future challenges

The CHASER-DAS provides valuable information for the future development of both models and observations. The ensemble spread can be a measure of the analysis uncertainty. The observed large analysis spreads for O_3 and NO_2 in the tropical upper troposphere and near the surface indicate a requirement for further constraints from additional observations or high quality data to improve the analysis. The analysis increment obtained during the data assimilation cycle primarily relates to persistent model biases. The positive analysis increments obtained for O_3 and CO imply that CHASER tends to underestimate (overestimate) O_3 (CO) concentrations in lower/middle troposphere and tends to overestimate (underestimate) them in the upper troposphere. This information is useful to identify sources of the model error and improve the performance of both model and data assimilation. The large analysis spreads and increments near the surface also indicate a requirement for better emission data sets.

The simultaneous assimilation of multiple satellite datasets is an important development for improving chemical weather forecasting (e.g. Kaminski et al., 2008) and better understanding the processes controlling the atmospheric environment. However, further developments are still required. First, more observation data are required to constrain O_3 and its precursors, especially near the surface. Retrieval sensitivity to the lowermost troposphere is critical for the emission inversion and the near surface air quality analysis. For instance, adding the near infrared (NIR) channel to the MOPITT retrieval increases the near surface sensitivity (Deeter et al., 2010), which may help to improve the analysis, while the IASI retrievals may contain information on the spatial extent of plumes (Coheur et al., 2009). Also, the emissions of O_3 precursors other than NO_x and CO, such as VOCs, have a pronounced influence on tropospheric chemistry. Further constraints are required for these fields; in particular, satellite CH_2O data may provide a significant constraints on VOCs emissions. Apart from the lower tropospheric observations, high quality satellite observations in the UTLS are needed for O_3 and HNO_3 , potentially augmented with CO and PAN as well as short-lived gases such as NO_2 (ESA, 2012). Second, the model resolution is too coarse to describe accurately small scale processes. A chemical data assimilation requires observations with sufficient spatial and temporal resolution to capture the heterogeneous distribution of tropospheric composition. In order to better take into account the small scale information available in the dataset, it is important to increase the model resolution close to the data set's resolution, as suggested by Pajot et al. (2011) and demonstrated using regional data assimilation systems (e.g. Hanea et al., 2004). In addition, the combined use of satellite and

surface in-situ data may provide strong constraints on the near surface analysis at high resolution. Third, introduction of a reasonable bias correction scheme is important to improve the analysis, especially when multiple data sets are simultaneously assimilated (e.g. Dee, 2005).

Appendix A

System ability check based on synthetic observations

It is of great interest to test the ability of the data assimilation system to improve the O₃ analysis in the presence of an emission error. We conducted an idealized data assimilation experiment, the so-called twin experiment (e.g. Ghil et al., 1991), by perturbing both the initial condition and the NO_x emission. The purpose of this experiment is to demonstrate that the data assimilation is properly implemented and quantify how the emission optimization influences the O₃ analysis, as similarly performed by Constantinescu et al. (2007) and Messina et al. (2011). Under the assumption of the perfect model scenario (i.e. a forecast model provides a perfect representation of the atmosphere), the actual background error (\mathbf{P}^b without model errors \mathbf{Q}) and observation error (\mathbf{R}) statistics can be determined precisely, so that the perfect model experiment allows us to demonstrate the importance of the data assimilation without unexpected model and observation errors.

A time series of a reference solution (or true state) for O₃ and NO₂ fields was generated by the simulation (without any assimilation) of the CTM using unperturbed emissions (i.e. the a priori emissions used for the real data assimilation). The reference solution was used to obtain artificial observation data and initial conditions for ensemble simulations and to validate the analysis. The artificial observation data were obtained from the true state, with the addition of zero-mean Gaussian random noise as observation errors with standard deviations of 10 % of the reference concentration. It was assumed that observation stations were located at 6.25 % (12.5 %) of the model grid points for O₃ (NO₂) in the horizontal; the vertical partial column with 3 km resolution was assimilated every 6 h. The state vector includes O₃ concentration and NO_x emissions. The O₃ data was used to update the O₃ concentration, while the NO₂ data was used to update NO_x emissions. The number of the assumed observations was larger than that in the real world, and it will affect the data assimilation performance. However, this idealized setting helps to demonstrate the ability of the system with enough constraints from observations. The simulated fields on 1 November 2007, were used in the initial assimilation cycle, and the analysis for 7–8 November was evaluated. The background error covariance for the initial assimilation cycle was obtained from the lagged average forecast (Hoffman and Kalnay, 1983); the initial ensemble concentration fields

were obtained from the reference simulation during 28 October to 4 November 2007.

Because of the biased emissions (constructed based on the annual mean a priori emissions), the model simulation without data assimilation has large errors in the simulated O₃ fields in the lower troposphere. The mean O₃ RMSE normalized by the background concentration averaged over 10° S–50° N latitudinal bands at 950 hPa is 23.7 % for the model simulation, which is almost the same as the initial error of 25.0 %. The assimilation of O₃ data reduces it to 16.5 %. The assimilation of NO₂ data helped to improve the O₃ analysis by reducing the errors included in the O₃ simulation due to biased NO_x emissions; the normalized O₃ RMSE is 14.2 % with a regional mean NO_x emission bias (RMSE) reduction of 41 (30) %. The assimilation of both O₃ and NO₂ data provided the best performance analysis, with a normalized O₃ RMSE of 11.7 %, which is almost equivalent to the assumed observation error (i.e. 10 %). In contrast, in the free troposphere (e.g. at 500 hPa), assimilation of O₃ data provided a much more significant improvement of the O₃ analysis than that provided by the NO₂ data. These results confirm that the simultaneous optimization for O₃ concentration and its precursors emissions is a powerful framework for the tropospheric chemistry analysis.

Acknowledgements. We would like to thank the two anonymous reviewers and the editor for their valuable comments. This research was supported by the JSPS Postdoctoral Fellowships for Research Abroad and the Environment Research and Technology Development Fund (A-0903) of the Ministry of the Environment, Japan. Folkert Boersma acknowledges funding by the Netherlands Organisation for Scientific Research, NWO Vidi grant 864.09.001.

Edited by: M. Kopacz

References

- Allen, D. J. and Pickering, K. E.: Evaluation of lightning flash rate parameterizations for use in a global chemical transport model, *J. Geophys. Res.*, 107, 4711, doi:10.1029/2002JD002066, 2002.
- Arellano, A. and Hess, P.: Sensitivity of top-down estimates of CO sources to GCTM transport, *Geophys. Res. Lett.*, 33, L21807, doi:10.1029/2006GL027371, 2006.
- Arellano, A., Kasibhatla, P., Giglio, L., der Werf, G. V., and Randerson, J.: Top-down estimates of global CO using MOPITT measurements, *Geophys. Res. Lett.*, 31, L01104, doi:10.1029/2003GL018609, 2004.
- Arellano Jr., A. F., Raeder, K., Anderson, J. L., Hess, P. G., Emmons, L. K., Edwards, D. P., Pfister, G. G., Campos, T. L., and Sachse, G. W.: Evaluating model performance of an ensemble-based chemical data assimilation system during INTEX-B field mission, *Atmos. Chem. Phys.*, 7, 5695–5710, doi:10.5194/acp-7-5695-2007, 2007.
- Beer, R.: TES on the Aura mission: scientific objectives, measurements, and analysis overview, *IEEE T. Geosci. Remote*, 44, 1102–1105, 2006.

- Bethan, S., Vaughan, G. and Reid, S. J.: A comparison of ozone and thermal tropopause heights and the impact of tropopause definition on quantifying the ozone content of the troposphere, *Q. J. Roy. Meteor. Soc.*, 122, 929–944, doi:10.1002/qj.49712253207, 1996.
- Boersma, K. F., Eskes, H. J., and Brinksma, E. J.: Error analysis for tropospheric NO₂ retrieval from space, *J. Geophys. Res.*, 109, D04311, doi:10.1029/2003JD003962, 2004.
- Boersma, K. F., Eskes, H. J., Meijer, E. W., and Kelder, H. M.: Estimates of lightning NO_x production from GOME satellite observations, *Atmos. Chem. Phys.*, 5, 2311–2331, doi:10.5194/acp-5-2311-2005, 2005.
- Boersma, K. F., Eskes, H. J., Veefkind, J. P., Brinksma, E. J., van der A, R. J., Sneep, M., van den Oord, G. H. J., Levelt, P. F., Stammes, P., Gleason, J. F., and Bucsela, E. J.: Near-real time retrieval of tropospheric NO₂ from OMI, *Atmos. Chem. Phys.*, 7, 2103–2118, doi:10.5194/acp-7-2103-2007, 2007.
- Boersma, K. F., Jacob, D. J., Bucsela, E. J., Perring, A. E., Dirksen, R., van der A, R. J., Yantosca, R. M., Park, R. J., Wenig, M. O., Bertram, T. H., and Cohen, R. C.: Validation of OMI tropospheric NO₂ observations during INTEX-B and application to constrain NO_x emissions over the Eastern United States and Mexico, *Atmos. Environ.*, 42, 4480–4497, doi:10.1016/j.atmosenv.2008.02.004, 2008a.
- Boersma, K. F., Jacob, D. J., Eskes, H. J., Pinder, R. W., Wang, J., and van der A, R. J.: Intercomparison of SCIAMACHY and OMI tropospheric NO₂ columns: observing the diurnal evolution of chemistry and emissions from space, *J. Geophys. Res.*, 113, 1–14, doi:10.1029/2007JD008816, 2008b.
- Boersma, K. F., Jacob, D. J., Trainic, M., Rudich, Y., DeSmedt, I., Dirksen, R., and Eskes, H. J.: Validation of urban NO₂ concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities, *Atmos. Chem. Phys.*, 9, 3867–3879, doi:10.5194/acp-9-3867-2009, 2009.
- Boersma, K. F., Eskes, H. J., Dirksen, R. J., van der A, R. J., Veefkind, J. P., Stammes, P., Huijnen, V., Kleipool, Q. L., Sneep, M., Claas, J., Leitão, J., Richter, A., Zhou, Y., and Brunner, D.: An improved tropospheric NO₂ column retrieval algorithm for the Ozone Monitoring Instrument, *Atmos. Meas. Tech.*, 4, 1905–1928, doi:10.5194/amt-4-1905-2011, 2011.
- Boersma, K. F., Braak, R., and van der A, R. J.: Dutch OMI NO₂ (DOMINO) data product v2.0, HE5 data file user manual. Tropospheric Emission Monitoring Internet Service (TEMIS), Royal Netherlands Meteorological Institute (KNMI), available at: http://www.temis.nl/docs/OMI_NO2_HE5_2.0_2011.pdf, August 2011.
- Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H.: SCIAMACHY: mission objectives and measurement modes, *J. Atmos. Sci.*, 56, 127–150, 1999.
- Bowman, K. W., Worden, J., Steck, T., Worden, H. M., Clough, S., and Rodgers, C.: Capturing time and vertical variability of tropospheric ozone: a study using TES nadir retrievals, *J. Geophys. Res.*, 107, 4723, doi:10.1029/2002JD002150, 2002.
- Bowman, K. W., Rodgers, C. D., Sund-Kulawik, S., Worden, J., Sarkissian, E., Osterman, G., Steck, T., Luo, M., Eldering, A., Shephard, M. W., Worden, H., Clough, S. A., Brown, P. D., Rinsland, C. P., Lampel, M., Gunson, M., and Beer, R.: Tropospheric emission spectrometer: retrieval method and error analysis, *IEEE T. Geosci. Remote*, 44, 1297–1307, 2006.
- Bowman, K. W., Jones, D. B. A., Logan, J. A., Worden, H., Boersma, F., Chang, R., Kulawik, S., Osterman, G., Hamer, P., and Worden, J.: The zonal structure of tropical O₃ and CO as observed by the Tropospheric Emission Spectrometer in November 2004 – Part 2: Impact of surface emissions on O₃ and its precursors, *Atmos. Chem. Phys.*, 9, 3563–3582, doi:10.5194/acp-9-3563-2009, 2009.
- Boxe, C. S., Worden, J. R., Bowman, K. W., Kulawik, S. S., Neu, J. L., Ford, W. C., Osterman, G. B., Herman, R. L., Eldering, A., Tarasick, D. W., Thompson, A. M., Doughty, D. C., Hoffmann, M. R., and Oltmans, S. J.: Validation of northern latitude Tropospheric Emission Spectrometer stare ozone profiles with ARC-IONS sondes during ARCTAS: sensitivity, bias and error analysis, *Atmos. Chem. Phys.*, 10, 9901–9914, doi:10.5194/acp-10-9901-2010, 2010.
- Boynard, A., Beekmann, M., Foret, G., Ung, A., Szopa, S., Schmechtig, C., and Coman, A.: Assessment of regional ozone model uncertainty with a modelling ensemble using an explicit error representation, *Atmos. Environ.*, 45, 784–793, 2011.
- Bucsela, E. J., Perring, A. E., Cohen, R. C., Boersma, K. F., Celarier, E. A., Gleason, J. F., Wenig, M. O., Bertram, T. H., Wooldridge, P. J., Dirksen, R., and Veefkind, J. P.: Comparison of tropospheric NO₂ from in situ aircraft measurements with near-real-time and standard product data from OMI, *J. Geophys. Res.*, 113, D16S31, doi:10.1029/2007JD008838, 2008.
- Callies, J., Corpaccioli, E., Eisinger, M., Hahne, A., and Lefebvre, A.: GOME-2 – MetOp’s Second Generation Sensor for Operational Ozone Monitoring, No. 102, ESA Bulletin, 2000.
- Coheur, P.-F., Clarisse, L., Turquety, S., Hurtmans, D., and Clerbaux, C.: IASI measurements of reactive trace species in biomass burning plumes, *Atmos. Chem. Phys.*, 9, 5655–5667, doi:10.5194/acp-9-5655-2009, 2009.
- Coman, A., Foret, G., Beekmann, M., Eremenko, M., Dufour, G., Gaubert, B., Ung, A., Schmechtig, C., Flaud, J.-M., and Bergametti, G.: Assimilation of IASI partial tropospheric columns with an Ensemble Kalman Filter over Europe, *Atmos. Chem. Phys.*, 12, 2513–2532, doi:10.5194/acp-12-2513-2012, 2012.
- Constantinescu, E., Sandu, A., Chai, T., and Carmichael, G.: Ensemble-based chemical data assimilation. I: general approach, *Q. J. Roy. Meteor. Soc.*, 133, 1229–1243, doi:10.1002/qj.76, 2007.
- Daniel, J. S. and Solomon, S.: On the climate forcing of carbon monoxide, *J. Geophys. Res.*, 103, 13249–13260, 1998.
- DeCaria, A. J., Pickering, K. E., Stenchikov, G. L., and Ott, L. E.: Lightning-generated NO_x and its impact on tropospheric ozone production: a three-dimensional modeling study of a Stratosphere-Troposphere Experiment: Radiation, Aerosols and Ozone (STERAO-A) thunderstorm, *J. Geophys. Res.*, 110, D14303, doi:10.1029/2004JD005556, 2005.
- Dee, D. P.: Bias and data assimilation, *Q. J. Roy. Meteor. Soc.*, 132, 3323–3343, doi:10.1256/qj.05.137, 2005.
- Deeter, M. N.: MOPITT (Measurements of Pollution in the Troposphere) Version 5 Product User’s Guide, Tech. rep., National Center for Atmospheric Research, Boulder, CO, USA, 2011.
- Deeter, M. N., Emmons, L. K., Francis, G. L., Edwards, D. P., Gille, J. C., Warner, J. X., Khattatov, B., Ziskin, D., Lamarque, J.-F., Ho, S.-P., Yudin, V., Attié, J.-L., Packman, D., Chen, J., Mao, D.,

- and Drummond, J. R.: Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, *J. Geophys. Res.*, 108, 4399, doi:10.1029/2002JD003186, 2003.
- Deeter, M. N., Edwards, D. P., Gille, J. C., and Drummond, J. R.: Sensitivity of MOPITT observations to carbon monoxide in the lower troposphere, *Geophys. Res. Lett.*, 112, D24306, doi:10.1029/2007JD008929, 2007.
- Deeter, M. N., Edwards, D. P., Gille, J. C., Emmons, L. K., Francis, G., Ho, S.-P., Mao, D., Masters, D., Worden, H., Drummond, J. R., and Novelli, P. C.: The MOPITT version 4 CO product: algorithm enhancements, validation and long-term stability, *J. Geophys. Res.*, 115, D07036, doi:10.1029/2009JD013005, 2010.
- Edwards, D. P., Pétron, G., Novelli, P. C., Emmons, L. K., Gille, J. C., and Drummond, J. R.: Southern Hemisphere carbon monoxide interannual variability observed by Terra/Measurement of Pollution in the Troposphere (MOPITT), *J. Geophys. Res.*, 111, D16303, doi:10.1029/2006JD007079, 2006.
- Elbern, H. and Schmidt, H.: Ozone episode analysis by four-dimensional variational chemistry data assimilation, *J. Geophys. Res.*, 106, 3569–3590, 2001.
- Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, *Atmos. Chem. Phys.*, 7, 3749–3769, doi:10.5194/acp-7-3749-2007, 2007.
- Elguindi, N., Clark, H., Ordóñez, C., Thouret, V., Flemming, J., Stein, O., Huijnen, V., Moinat, P., Inness, A., Peuch, V.-H., Stohl, A., Turquety, S., Athier, G., Cammas, J.-P., and Schultz, M.: Current status of the ability of the GEMS/MACC models to reproduce the tropospheric CO vertical distribution as measured by MOZIC, *Geosci. Model Dev.*, 3, 501–518, doi:10.5194/gmd-3-501-2010, 2010.
- Errera, Q., Daerden, F., Chabrilat, S., Lambert, J. C., Lahoz, W. A., Viscardy, S., Bonjean, S., and Fonteyn, D.: 4D-Var assimilation of MIPAS chemical observations: ozone and nitrogen dioxide analyses, *Atmos. Chem. Phys.*, 8, 6169–6187, doi:10.5194/acp-8-6169-2008, 2008.
- ESA: Report for Mission Selection: PProcess Exploration through Measurements of Infrared and millimetre-wave Emitted Radiation (PREMIER), ESA SP-1324/3 (3 volume series), European Space Agency, Noordwijk, The Netherlands, 2012.
- Eskes, H. J. and Boersma, K. F.: Averaging kernels for DOAS total-column satellite retrievals, *Atmos. Chem. Phys.*, 3, 1285–1291, doi:10.5194/acp-3-1285-2003, 2003.
- Eskes, H. J., Velthoven, P. F. J. V., Valks, P. J. M., and Kelder, H. M.: Assimilation of GOME total-ozone satellite observations in a three-dimensional tracer-transport model, *Q. J. Roy. Meteor. Soc.*, 129, 1663–1681, doi:10.1256/qj.02.14, 2003.
- European Commission: Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL), Emission Database for Global Atmospheric Research (EDGAR), release version 4.2, <http://edgar.jrc.ec.europa.eu>, 2011.
- Feng, L., Brugge, R., Hólm, E. V., Harwood, R. S., O'Neill, A., Filipiak, M. J., Froidevaux, L., and Livesey, N.: Four-dimensional variational assimilation of ozone profiles from the Microwave Limb Sounder on the Aura satellite, *J. Geophys. Res.*, 113, D15S0, doi:10.1029/2007JD009121, 2008.
- Flemming, J., Inness, A., Flentje, H., Huijnen, V., Moinat, P., Schultz, M. G., and Stein, O.: Coupling global chemistry transport models to ECMWF's integrated forecast system, *Geosci. Model Dev.*, 2, 253–265, doi:10.5194/gmd-2-253-2009, 2009.
- Folberth, G., Hauglustaine, D. A., Ciais, P., and Lathière, J.: On the role of atmospheric chemistry in the global CO₂ budget, *Geophys. Res. Lett.*, 32, L08801, doi:10.1029/2004GL021812, 2005.
- Foret, G., Hamaoui, L., Schmechtig, C., Eremenko, M., Keim, C., Dufour, G., Boynard, A., Coman, A., Ung, A., and Beekmann, M.: Evaluating the potential of IASI ozone observations to constrain simulated surface ozone concentrations, *Atmos. Chem. Phys.*, 9, 8479–8491, doi:10.5194/acp-9-8479-2009, 2009.
- Fortems-Cheiney, A., Chevallier, F., Pison, I., Bousquet, P., Szopa, S., Deeter, M. N., and Clerbaux, C.: Ten years of CO emissions as seen from Measurements of Pollution in the Troposphere (MOPITT), *J. Geophys. Res.*, 116, D05304, doi:10.1029/2010JD014416, 2011.
- Froidevaux, L., Jiang, Y. B., Lambert, A., Livesey, N. J., Read, W. G., Waters, J. W., Browell, E. V., Hair, J. W., Avery, M. A., McGee, T. J., Twigg, L. W., Sumnicht, G. K., Jucks, K. W., Margitan, J. J., Sen, B., Stachnik, R. A., Toon, G. C., Bernath, P. F., Boone, C. D., Walker, K. A., Filipiak, M. J., Harwood, R. S., Fuller, R. A., Manney, G. L., Schwartz, M. J., Daffer, W. H., Drouin, B. J., Cofield, R. E., Cuddy, D. T., Jarnot, R. F., Knosp, B. W., Perun, V. S., Snyder, W. V., Stek, P. C., Thurstans, R. P., and Wagner, P. A.: Validation of Aura Microwave Limb Sounder stratospheric ozone measurements, *J. Geophys. Res.*, 113, D15S20, doi:10.1029/2007JD008771, 2008.
- Fu, T.-M., Jacob, D. J., Palmer, P. I., Chance, K., Wang, Y. X., Barletta, B., Blake, D. R., Stanton, J. C., and Pilling, M. J.: Space-based formaldehyde measurements as constraints on volatile organic compound emissions in East and South Asia, *J. Geophys. Res.*, 112, D06312, doi:10.1029/2006JD007853, 2007.
- Geer, A. J., Lahoz, W. A., Bekki, S., Bormann, N., Errera, Q., Eskes, H. J., Fonteyn, D., Jackson, D. R., Jukes, M. N., Massart, S., Peuch, V.-H., Rharmili, S., and Segers, A.: The ASSET intercomparison of ozone analyses: method and first results, *Atmos. Chem. Phys.*, 6, 5445–5474, doi:10.5194/acp-6-5445-2006, 2006.
- Ghil, M., and Malanotte-Rizzoli, P.: Data assimilation in meteorology and oceanography, *Adv. Geophys.* 33, 141–266, 1991.
- Gloudemans, A. M. S., de Laat, A. T. J., Schrijver, H., Aben, I., Meirink, J. F., and van der Werf, G. R.: SCIAMACHY CO over land and oceans: 2003–2007 interannual variability, *Atmos. Chem. Phys.*, 9, 3799–3813, doi:10.5194/acp-9-3799-2009, 2009.
- Graedel, T. E., Bates, T. S., Bouwman, A. F., Cunnold, D., Dignon, J., Fung, I., Jacob, D. J., Lamb, B. K., Logan, J. A., Marland, G., Middleton, P., Pacyna, J. M., Placet, M., and Veldt, C.: A compilation of inventories of emissions to the atmosphere, *Global Biogeochem. Cy.*, 7, 1–26, 1993.
- Grassi, B., Redaelli, G., and Visconti, G.: Assimilation of stratospheric ozone in the chemical transport model STRATAQ, *Ann. Geophys.*, 22, 2669–2678, doi:10.5194/angeo-22-2669-2004, 2004.
- Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., McKay, W., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J., and Zimmerman, P.: A global model of natural volatile organic compound emissions, *J. Geophys. Res.*, 100, 8873–8892, doi:10.1029/94JD02950, 1995.

- Hanea, R. G., Velders, G. J. M., Heemink, A.: Data assimilation of groundlevel ozone in Europe with a Kalman filter and chemistry transport, *J. Geophys. Res.*, 109, 2004.
- Hains, J., Boersma, F., Kroon, M., Dirksen, R., Cohen, R. C., Perring, A., Bucseles, E., Volten, H., Swart, D., Richter, A., Wittrock, F., Schoenhardt, A., Wagner, T., Ibrahim, O., Van Roozendaal, M., Pinardi, G., Gleason, J. F., Veeffkind, J., and Levelt, P.: Testing and improving OMI DOMINO tropospheric NO₂ using observations from the DANDELIONS and INTEX-B validation campaigns, *J. Geophys. Res.*, 115, 1–20, doi:10.1029/2009JD012399, 2010.
- Hanea, R. G., Velders, G. J. M., and Heemink, A.: Data assimilation of ground-level ozone in Europe with a Kalman filter and chemistry transport model, *J. Geophys. Res.*, 109, D10302, doi:10.1029/2003JD004283, 2004.
- Heald, C. L., Jacob, D. J., Jones, D. B. A., Palmer, P. I., Logan, J. A., and Streets, D. G.: Comparative inverse analysis of satellite (MOPITT) and aircraft (TRACE-P) observations to estimate Asian sources of carbon monoxide, *J. Geophys. Res.*, 109, D23306, doi:10.1029/2004JD005185, 2004.
- Heikes, B. G.: Formaldehyde and hydroperoxides at Mauna Loa observatory, *J. Geophys. Res.*, 97, 18001–18013, 1992.
- Ho, S. ., Edwards, D. P., Gille, J. C., Luo, M., Osterman, G. B., Kulawik, S. S., and Worden, H.: A global comparison of carbon monoxide profiles and column amounts from Tropospheric Emission Spectrometer (TES) and Measurements of Pollution in the Troposphere (MOPITT), *J. Geophys. Res.*, 114, D21307, doi:10.1029/2009JD012242, 2009.
- Hoffman, R. and Kalnay, E.: Lagged average forecasting, and alternative to Monte Carlo forecasting, *Tellus*, 35A, 100–118, 1983.
- Holloway, T., Levy, H., and Kasibhatla, P.: Global distribution of carbon monoxide, *J. Geophys. Res.*, 105, 12123–12147, doi:10.5194/acp-11-4705-2011, 2000.
- Hooghiemstra, P. B., Krol, M. C., Meirink, J. F., Bergamaschi, P., van der Werf, G. R., Novelli, P. C., Aben, I., and Röckmann, T.: Optimizing global CO emission estimates using a four-dimensional variational data assimilation system and surface network observations, *Atmos. Chem. Phys.*, 11, 4705–4723, doi:10.5194/acp-11-4705-2011, 2011.
- Houtekamer, P. L. and Mitchell, H. L.: Data Assimilation Using an Ensemble Kalman Filter Technique, *Mon. Weather Rev.*, 126, 796–811, 1998.
- Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, *Mon. Weather Rev.*, 129, 123–137, 2001.
- Hudman, R. C., Jacob, D. J., Turquety, S., Leibensperger, E. M., Murray, L. T., Wu, S., Gilliland, A. B., Avery, M., Bertram, T. H., Brune, W., Cohen, R. C., Dibb, J. E., Flocke, F. M., Fried, A., Holloway, J., Neuman, J. A., Orville, R., Perring, A., Ren, X., Sachse, G. W., Singh, H. B., Swanson, A., and Wooldridge, P. J.: Surface and lightning sources of nitrogen oxides over the United States: magnitudes, chemical evolution, and outflow, *J. Geophys. Res.*, 112, D12S05, doi:10.1029/2006JD007912, 2007.
- IPCC: Climate Change 2007, The Physical Science Basis, Cambridge University Press, 2007.
- Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter, *Physica D*, 230, 112–126, 2007.
- Jackson, D.: Assimilation of EOS MLS ozone observations in the Met Office data assimilation system, *Q. J. Roy. Meteor. Soc.*, 133, 1771–1788, doi:10.1002/qj.140, 2007.
- Jaeglé, L., Steinberger, L., Martin, R. V., and Chance, K.: Global partitioning of NO_x sources using satellite observations: relative roles of fossil fuel combustion, biomass burning and soil emissions, *Faraday Discuss.*, 130, 407–423, 2005.
- Jenkins, G. S. and Ryu, J.-H.: Linking horizontal and vertical transports of biomass fire emissions to the tropical Atlantic ozone paradox during the Northern Hemisphere winter season: climatology, *Atmos. Chem. Phys.*, 4, 449–469, doi:10.5194/acp-4-449-2004, 2004.
- Jones, D. B. A., Bowman, K. W., Palmer, P. I., Worden, J. R., Jacob, D. J., Hoffman, R. N., Bey, I., and Yantosca, R. M.: Potential of observations from the Tropospheric Emission Spectrometer to constrain continental sources of carbon monoxide, *J. Geophys. Res.*, 108, 4789, doi:10.1029/2003JD003702, 2003.
- Jones, D. B. A., Bowman, K. W., Logan, J. A., Heald, C. L., Liu, J., Luo, M., Worden, J., and Drummond, J.: The zonal structure of tropical O₃ and CO as observed by the Tropospheric Emission Spectrometer in November 2004 – Part 1: Inverse modeling of CO emissions, *Atmos. Chem. Phys.*, 9, 3547–3562, doi:10.5194/acp-9-3547-2009, 2009.
- Jourdain, L., Worden, H. M., Worden, J. R., Bowman, K., Li, Q. B., Eldering, A., Kulawik, S. S., Osterman, G., Boersma, F., Fisher, B., Rinsland, C. P., Beer, R., and Gunson, M.: Tropospheric vertical distribution of tropical Atlantic ozone observed by TES during the Northern African biomass burning season, *Geophys. Res. Lett.*, 34, L04810, doi:10.1029/2006GL028284, 2007.
- Kaminski, J. W., Neary, L., Struzewska, J., McConnell, J. C., Lupu, A., Jarosz, J., Toyota, K., Gong, S. L., Côté, J., Liu, X., Chance, K., and Richter, A.: GEM-AQ, an on-line global multiscale chemical weather modelling system: model description and evaluation of gas phase chemistry processes, *Atmos. Chem. Phys.*, 8, 3255–3281, doi:10.5194/acp-8-3255-2008, 2008.
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., et al.: NCEP-DOE AMIP-II Reanalysis (R-2), *B. Am. Meteorol. Soc.*, 83, 1631–1643, doi:10.1175/BAMS-83-11-1631, 2002.
- Kasibhatla, P., Logan, J., Palmer, P., and Novelli, P.: Top-down estimate of a large source of atmospheric carbon monoxide associated with fuel combustion in Asia, *Geophys. Res. Lett.*, 29, 1900, doi:10.1029/2002GL015581, 2002.
- Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 341 pp. 2003.
- Khattatov, B. V., Lamarque, J.-F., Lyjak, L. V., Menard, R., Levelt, P., Tie, X., Brasseur, G. P., and Gille, J. C.: Assimilation of satellite observations of long-lived chemical species in global chemistry transport models, *J. Geophys. Res.*, 105, 29135–29144, doi:10.1029/2000JD900466, 2000.
- Kopacz, M., Jacob, D. J., Henze, D. K., Heald, C. L., Streets, D. G., and Zhang, Q.: Comparison of adjoint and analytical Bayesian inversion methods for constraining Asian sources of carbon monoxide using satellite (MOPITT) measurements of CO columns, *J. Geophys. Res.*, 114, D04305, doi:10.1029/2007JD009264, 2009.
- Kopacz, M., Jacob, D. J., Fisher, J. A., Logan, J. A., Zhang, L., Megretskaia, I. A., Yantosca, R. M., Singh, K., Henze, D. K., Burrows, J. P., Buchwitz, M., Khlystova, I., McMillan, W. W.,

- Gille, J. C., Edwards, D. P., Eldering, A., Thouret, V., and Nedelec, P.: Global estimates of CO sources with high resolution by adjoint inversion of multiple satellite datasets (MOPITT, AIRS, SCIAMACHY, TES), *Atmos. Chem. Phys.*, 10, 855–876, doi:10.5194/acp-10-855-2010, 2010.
- Lahoz, W. A., Errera, Q., Swinbank, R., and Fonteyn, D.: Data assimilation of stratospheric constituents: a review, *Atmos. Chem. Phys.*, 7, 5745–5773, doi:10.5194/acp-7-5745-2007, 2007.
- Lamsal, L. N., Martin, R. V., van Donkelaar, A., Celarier, E. a., Bucsela, E. J., Boersma, K. F., Dirksen, C. L., and Wang, Y.: Indirect validation of tropospheric nitrogen dioxide retrieved from the OMI satellite instrument: insight into the seasonal variation of nitrogen oxides at northern midlatitudes, *J. Geophys. Res.*, 115, 1–15, doi:10.1029/2009JD013351, 2010.
- Lamsal, L. N., Martin, R. V., Padmanabhan, A., van Donkelaar, A., Zhang, Q., Sioris, C. E., Chance, K., Kurosu, T. P., and Newchurch, M. J.: Application of satellite observations for timely updates to global anthropogenic NO_x emission inventories, *Geophys. Res. Lett.*, 38, L05810, doi:10.1029/2010GL046476, 2011.
- Levelt, P. F., Hilsenrath, E., Leppelmeier, G. W., van den Oord, G. H. J., Bhartia, P. K., Tamminen, J., de Haan, J. F., and Veefkind, J. P.: Science objectives of the Ozone Monitoring Instrument, *Geosci. Remote Sens.*, 44, 1199–1208, doi:10.1109/TGRS.2006.872336, 2006.
- Lin, C., J. Zhu, and Wang, Z.: Model bias correction for dust storm forecast using ensemble Kalman filter, *J. Geophys. Res.*, 113, D14306, doi:10.1029/2007JD009498, 2008.
- Liu, S. C., M. Trainer, F. C. Fehsenfeld, D. D. Parrish, E. J. Williams, D. W. Fahey, G. Hübler, and P. C. Murphy: Ozone Production in the Rural Troposphere and the Implications for Regional and Global Ozone Distributions, *J. Geophys. Res.*, 92(D4), 4191–4207, doi:10.1029/JD092iD04p04191, 1987.
- Livesey, N. J., Read, W. G., Froidevaux, L., Lambert, A., Manney, G. L., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Cofield, R. E., Cuddy, D. T., Fuller, R. A., Jarnot, R. F., Jiang, J. H., Knosp, B. W., Stek, P. C., Wagner, P. A., and Wu, D. L.: Aura Microwave Limb Sounder (MLS). Version 3.3 Level 2 data quality and description document, Tech. Rep. JPL D-33509, Jet Propul. Lab., Pasadena, CA, 2011.
- Logan, J. A., Prather, M. J., Wofsy, S. C., and McElroy, M. B.: Tropospheric chemistry: a global perspective, *J. Geophys. Res.*, 86, 7210–7254, 1981.
- Lopez, J. P., Luo, M., Christensen, L. E., Loewenstein, M., Jost, H., Webster, C. R., and Osterman, G.: TES carbon monoxide validation during two AVE campaigns using the Argus and ALIAS instruments on NASA's WB-57F, *J. Geophys. Res.*, 113, D16S47, doi:10.1029/2007JD008811, 2008.
- Luo, M., Rinsland, C. P., Rodgers, C. D., Logan, J. A., Worden, H., Kulawik, S., Eldering, A., Goldman, A., Shephard, M. W., Gunson, M., and Lampel, M.: Comparison of carbon monoxide measurements by TES and MOPITT: influence of a priori data and instrument characteristics on nadir atmospheric species retrievals, *J. Geophys. Res.*, 112, D09303, doi:10.1029/2006JD007663, 2007.
- Mallet, V., and Sportisse, B.: A comprehensive study of ozone sensitivity with respect to emissions over Europe with a chemistry-transport model, *J. Geophys. Res.*, 110, D22302, doi:10.1029/2005JD006234, 2005.
- Martin, R., Jacob, D., Chance, K., Kurosu, T., Palmer, P., and Evans, M.: Global inventory of nitrogen oxide emissions constrained by space-based observations of NO₂ columns, *J. Geophys. Res.*, 108, 1–12, doi:10.1029/2003JD003453, 2003.
- Martin, R. V., Jacob, D. J., Logan, J. A., Bey, I., Yantosca, R. M., Staudt, A. C., Li, Q. B., Fiore, A. M., Duncan, B. N., Liu, H., Ginoux, P., and Thouret, V.: Interpretation of TOMS observations of tropical tropospheric ozone with a global model and in situ observations, *J. Geophys. Res.*, 107, 4351, doi:10.1029/2001JD001480, 2002.
- Martin, R. V., Sauvage, B., Folkins, I., Sioris, C. E., Boone, C., Bernath, P., and Ziemke, J.: Space-based constraints on the production of nitric oxide by lightning, *J. Geophys. Res.*, 112, 1–14, doi:10.1029/2006JD007831, 2007.
- Massart, S., Piacentini, A., and Pannekoucke, O.: Importance of using ensemble estimated background error covariances for the quality of atmospheric ozone analyses, *Q. J. Roy. Meteor. Soc.*, 138, 889–905, doi:10.1002/qj.971, 2012.
- Ménard, R. and Chang, L.-P.: Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part II: χ^2 -validated results and analysis of variance and correlation dynamics, *Mon. Weather Rev.*, 128, 2672–2686, 2000.
- Messina, P., D'Isidoro, M., Maurizi, A., and Fierli, F.: Impact of assimilated observations on improving tropospheric ozone simulations, *Atmos. Environ.*, 45, 6674–6681, doi:10.1016/j.atmosenv.2011.08.056, 2011.
- Millet, D. B., Jacob, D. J., Turquety, S., Hudman, R. C., Wu, S. L., Fried, A., Walega, J., Heikes, B. G., Blake, D. R., Singh, H. B., Anderson, B. E., and Clarke, A. D.: Formaldehyde distribution over North America: implications for satellite retrievals of formaldehyde columns and isoprene emission, *J. Geophys. Res.*, 111, D24S02, doi:10.1029/2005JD006853, 2006.
- Miyazaki, K., Eskes, H. J., and Sudo, K.: Global NO_x emission estimates derived from an assimilation of OMI tropospheric NO₂ columns, *Atmos. Chem. Phys.*, 12, 2263–2288, doi:10.5194/acp-12-2263-2012, 2012.
- Miyoshi, T. and Yamane, S.: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution, *Mon. Weather Rev.*, 135, 3841–3861, doi:10.1175/2007MWR1873.1, 2007.
- Müller, J.-F. and Stavrakou, T.: Inversion of CO and NO_x emissions using the adjoint of the IMAGES model, *Atmos. Chem. Phys.*, 5, 1157–1186, doi:10.5194/acp-5-1157-2005, 2005.
- Nassar, R., Logan, J. A., Worden, H. M., Megretskaia, I. A., Bowman, K. W., Osterman, G. B., Thompson, A. M., Tarasick, D. W., Austin, S., Claude, H., Dubey, M. K., Hocking, W. K., Johnson, B. J., Joseph, E., Merrill, J., Morris, G. A., Newchurch, M., Oltmans, S. J., Posny, F., Schmidlin, F. J., Vömel, H., Whiteman, D. N., and Witte, J. C.: Validation of Tropospheric Emission Spectrometer (TES) nadir ozone profiles using ozonesonde measurements, *J. Geophys. Res.*, 113, D15S17, doi:10.1029/2007JD008819, 2008.
- Ohara, T., Akimoto, H., Kurokawa, J., Horii, N., Yamaji, K., Yan, X., and Hayasaka, T.: An Asian emission inventory of anthropogenic emission sources for the period 1980–2020, *Atmos. Chem. Phys.*, 7, 4419–4444, doi:10.5194/acp-7-4419-2007, 2007.
- Olivier, J. G. J., Van Aardenne, J. A., Dentener, F., Ganzeveld, L., and Peters, J. A. H. W.: Recent trends in global greenhouse gas emissions: regional trends 1970–2000 and spatial distribution of

- key sources in 2000, *Environ. Sci.*, 2, 81–99, 2005.
- Osterman, G. B., Kulawik, S. S., Worden, H., Richards, N. A., Fisher, B. M., Eldering, A., Shephard, M. W., Froidevaux, L., Labow, G., Luo, M., Herman, R. L., Bowman, K. W., and Thompson, A. M.: Validation of Tropospheric Emission Spectrometer (TES) measurements of the total, stratospheric, and tropospheric column abundance of ozone, *J. Geophys. Res.*, 113, D15S16, doi:10.1029/2007JD008801, 2008.
- Osterman, G., Bowman, K., Eldering, A., Fisher, B., Herman, R., Jacob, D., Jourdain, L., Kulawik, S., Luo, M., Monarrez, R., Paradise, S., Payne, V., Poosti, S., Richards, N., Rider, D., Shepard, D., Shephard, M., Vilmotter, F., Worden, H., Worden, J., Yun, H., and Zhang, L.: TES Level 2 Data Users Guide, v4.0, Tech. Rep. JPL D-38042, Jet Propul. Lab., Pasadena, CA, USA, 2009.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A.: A local ensemble Kalman filter for atmospheric data assimilation, *Tellus A*, 56, 415–428, doi:10.1111/j.1600-0870.2004.00076.x, 2004.
- Ott, L. E., Pickering, K. E., Stenchikov, G. L., Allen, D. J., Decaria, A. J., Ridley, B., Lin, R.-F., Lang, S., and Tao, W.-K.: Production of lightning NO_x and its vertical distribution calculated from three-dimensional cloud-scale chemical transport model simulations, *J. Geophys. Res.-Atmos.*, 115, D04301, doi:10.1029/2009jd011880, 2010.
- Pajot, B., Massart, S., Cariolle, D., Piacentini, A., Pannekoek, O., Lahoz, W. A., Clerbaux, C., Coheur, P. F., and Hurtmans, D.: High resolution assimilation of IASI ozone data with a global CTM, *Atmos. Chem. Phys. Discuss.*, 11, 29357–29406, doi:10.5194/acpd-11-29357-2011, 2011.
- Parrington, M., Jones, D. B. A., Bowman, K. W., Horowitz, L. W., Thompson, A. M., Tarasick, D. W., and Witte, J. C.: Estimating the summertime tropospheric ozone distribution over North America through assimilation of observations from the Tropospheric Emission Spectrometer, *J. Geophys. Res.*, 113, D18307, doi:10.1029/2007JD009341, 2008.
- Parrington, M., Jones, D. B. A., Bowman, K. W., Thompson, A. M., Tarasick, D. W., Merrill, J., Oltmans, S. J., Leblanc, T., Witte, J. C., and Millet, D. B.: Impact of the assimilation of ozone from the Tropospheric Emission Spectrometer on surface ozone across North America, *Geophys. Res. Lett.*, 36, L04802, doi:10.1029/2008GL036935, 2009.
- Pickering, K. E., Wang, Y., Tao, W.-K., Price, C., and Müller, J.-F.: Vertical distributions of lightning NO_x for use in regional and global chemical transport models, *J. Geophys. Res.*, 103, 31203–31216, 1998.
- Pison, I., Bousquet, P., Chevallier, F., Szopa, S., and Hauglustaine, D.: Multi-species inversion of CH_4 , CO and H_2 emissions from surface measurements, *Atmos. Chem. Phys.*, 9, 5281–5297, doi:10.5194/acp-9-5281-2009, 2009.
- Price, C. and Rind, D.: A simple lightning parameterization for calculating global lightning distributions, *J. Geophys. Res.*, 97, 9919–9933, 1992.
- Randerson, J. T., van der Werf, G. R., Giglio, L., Collatz, G. J., and Kasibhatla, P. S.: Global Fire Emissions Database, Version 2 (GFEDv2.1), Data Set, available at: <http://daac.ornl.gov/>, last access: May 2012, 2007.
- Richter, A. and Burrows, J. P.: Retrieval of Tropospheric NO_2 from GOME Measurements, *Adv. Space Res.*, 29, 1673–1683, 2002.
- Richter, A., Burrows, J. P., Nusz, H., Granier, C., and Niemeier, U.: Increase in tropospheric nitrogen dioxide over China observed from space, *Nature*, 437, 129–132, doi:10.1038/nature04092, 2005.
- Rodgers, C. and Connor, B.: Intercomparison of remote sounding instruments, *J. Geophys. Res.*, 108, 4116, doi:10.1029/2002JD002299, 2003.
- Rodgers, C. D.: *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Science Publishing, London, 2000.
- Sandu, A. and Chai, T.: Chemical data assimilation – an overview, *Atmosphere*, 3, 426–463, 2011.
- Sawada, S. and Totsuka, T.: Natural and anthropogenic sources and fate of atmospheric ethene, *Atmos. Environ.*, 20, 821–832.
- Schoeberl, M. R., Ziemke, J. R., Bojkov, B., Livesey, N., Duncan, B., Strahan, S., Froidevaux, L., Kulawik, S., Bhartia, P. K., Chandrasekhar, S., Levelt, P. F., Witte, J. C., Thompson, A. M., Cuevas, E., Redondas, A., Tarasick, D. W., Davies, J., Bodeker, G., Hansen, G., Johnson, B. J., Oltmans, S. J., Voemel, H., Allaart, M., Kelder, H., Newchurch, M., Godin-Beekmann, S., Ancellet, G., Claude, H., Andersen, S. B., Kyrö, E., Parrondos, M., Yela, M., Zabolocki, G., Moore, D., Dier, H., von der Gathen, P., Viatte, P., Stübi, R., Calpini, B., Skrivankova, P., Dorokhov, V., De Backer, H., Schmidlin, F. J., Coetzee, G., Fujiwara, M., Thouret, V., Posny, F., Morris, G., Merrill, J., Leong, C. P., König-Langlo, G., and Joseph, E.: A trajectory-based estimate of the tropospheric ozone column using the residual method, *J. Geophys. Res.*, 112, D24S49, doi:10.1029/2007JD008773, 2007.
- Schumann, U. and Huntrieser, H.: The global lightning-induced nitrogen oxides source, *Atmos. Chem. Phys.*, 7, 3823–3907, doi:10.5194/acp-7-3823-2007, 2007.
- Segers, A. J., Eskes, H., van der A, R., van Oss, R., and van Velthoven, P.: Assimilation of GOME ozone profiles and a global chemistry-transport model using a Kalman filter with anisotropic covariance, *Q. J. Roy. Meteor. Soc.*, 131, 477–502, doi:10.1256/qj.04.92, 2005.
- Shindell, D. T., Faluvegi, G., Koch, D. M., Schmidt, G. A., Unger, N., and Bauer, S. E.: Improved attribution of climate forcing to emissions, *J. Geophys. Res.*, 326, 716–718, doi:10.1126/Science.1174760, 2009.
- Shindell, D. T., Faluvegi, G., Stevenson, D. S., Krol, M. C., Emmons, L. K., Lamarque, J.-F., Petron, G., Dentener, F. J., Ellingsen, K., Schultz, M. G., Wild, O., Amann, M., Atherton, C. S., Bergmann, D. J., Bey, I., Butler, T., Cofala, J., Collins, W. J., Derwent, R. G., Doherty, R. M., Drevet, J., Eskes, H. J., Fiore, A. M., Gauss, M., Hauglustaine, D. A., Horowitz, L. W., Isaksen, I. S. A., Lawrence, M. G., Montanaro, V., Mueller, J.-F., Pitari, G., Prather, M. J., Pyle, J. A., Rast, S., Rodriguez, J. M., Sander, M. G., Savage, N. H., Strahan, S. E., Sudo, K., Szopa, S., Unger, N., van Noije, T. P. C., and Zeng, G.: Multimodel simulations of carbon monoxide: comparison with observations and projected near-future changes, *J. Geophys. Res.*, 111, D19306, doi:10.1029/2006JD007100, 2006.
- Singh, H. B., Brune, W. H., Crawford, J. H., Flocke, F., and Jacob, D. J.: Chemistry and transport of pollution over the Gulf of Mexico and the Pacific: spring 2006 INTEX-B campaign overview and first results, *Atmos. Chem. Phys.*, 9, 2301–2318, doi:10.5194/acp-9-2301-2009, 2009.
- Singh, K., Jardak, M., Sandu, A., Bowman, K., Lee, M., and Jones, D.: Construction of non-diagonal background error co-

- variance matrices for global chemical data assimilation, *Geosci. Model Dev.*, 4, 299–316, doi:10.5194/gmd-4-299-2011, 2011.
- Smit, H. G. J. and Kley, D.: Ozone Sonde Intercomparison Experiment (JOSIE), WMO Global Atmosphere Watch report series, No. 130 (Technical Document No. 926), World Meteorological Organization, Geneva, 1998.
- Stajner, I., Wargan, K., Pawson, S., Hayashi, H., Chang, L. P., Hudman, R. C., Froidevaux, L., Livesey, N., Levelt, P. F., Thompson, A. M., Tarasick, D. W., Stübi, R., Andersen, S. B., Yela, M., König-Langlo, G., Schmidlin, F. J., and Witte, J. C.: Assimilated ozone from EOS-Aura: evaluation of the tropopause region and tropospheric columns, *J. Geophys. Res.*, 113, D16S32, doi:10.1029/2007JD008863, 2008.
- Staudt, A. C., Jacob, D. J., Ravetta, F., Logan, J. A., Bachiochi, D., Krishnamurti, T. N., Sandholm, S., Ridley, B., Singh, H. B., and Talbot, B.: Sources and chemistry of nitrogen oxides over the tropical Pacific, *J. Geophys. Res.*, 108, 8239, doi:10.1029/2002JD002139, 2003.
- Stavrakou, T. and Müller, J.-F.: Grid-based versus big region approach for inverting CO emissions using Measurement of Pollution in the Troposphere (MOPITT) data, *J. Geophys. Res.*, 111, D15304, doi:10.1029/2005JD006896, 2006.
- Sudo, K., Takahashi, M., and Akimoto, H.: CHASER: a global chemical model of the troposphere 2. Model results and evaluation, *J. Geophys. Res.*, 107, 4586, doi:10.1029/2001JD001114, 2002.
- Thompson, A. M.: The oxidizing capacity of the Earth's atmosphere: probable past and future changes, *Science*, 256, 1157–1168, 1992.
- Thornton, J. A., Wooldridge, P. J., Cohen, R. C., Williams, E. J., Hereid, D., Fehsenfeld, F. C., Stutz, J., and Alicke, B.: Comparisons of in situ and long path measurements of NO₂ in urban plumes, *J. Geophys. Res.*, 108, 4496, doi:10.1029/2003JD003559, 2003.
- Turquety, S., Hadji-Lazaro, J., Clerbaux, C., Hauglustaine, D. A., Clough, S. A., Cassé, V., Schlüssel, P., and Mégie, G.: Operational trace gas retrieval algorithm for the Infrared Atmospheric Sounding Interferometer, *J. Geophys. Res.*, 109, D21301, doi:10.1029/2004JD004821, 2004.
- van der A, R. J., Eskes, H. J., Boersma, K. F., van Noije, T. P. C., van Roozendaal, M., De Smedt, I., Peters, D. H. M. U., Kuenen, J. J. P., and Meijer, E. W.: Trends, seasonal variability and dominant NO_x source derived from a ten year record of NO₂ measured from space, *J. Geophys. Res.*, 113, 1–12, doi:10.1029/2007JD009021, 2008.
- van Loon, M., Bultjes, P., and Segers, A.: Data assimilation of ozone in the atmospheric transport chemistry model LOTOS, *Environ. Modell. Softw.*, 15, 603–609, 2000.
- Waters, J. W., Froidevaux, L., Harwood, R., Jarnot, R., Pickett, H., Read, W., Siegel, P., Cofield, R., Filipiak, M., Flower, D., Holden, J., Lau, G., Livesey, N., Manney, G., Pumphrey, H., Santee, M., Wu, D., Cuddy, D., Lay, R., Loo, M., Perun, V., Schwartz, M., Stek, P., Thurstans, R., Boyles, M., Chandra, S., Chavez, M., Chen, G. S., Chudasama, B., Dodge, R., Fuller, R., Girard, M., Jiang, J., Jiang, Y., Knosp, B., LaBelle, R., Lam, J., Lee, K., Miller, D., Oswald, J., Patel, N., Pukala, D., Quintero, O., Scaff, D., Snyder, W., Tope, M., Wagner, P., and Walch, M.: The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite, *IEEE Trans. Geosci. Remote Sens.*, 44, 1106–1121, 2006.
- Worden, J., Kulawik, S. S., Shephard, M. W., Clough, S. A., Worden, H., Bowman, K., and Goldman, A.: Predicted errors of tropospheric emission spectrometer nadir retrievals from spectral window selection, *J. Geophys. Res.*, 109, D09308, doi:10.1029/2004JD004522, 2004.
- Worden, J., Liu, X., Bowman, K., Chance, K., Beer, R., Eldering, A., Gunson, M., and Worden, H.: Improved tropospheric ozone profile retrievals using OMI and TES radiances, *J. Geophys. Res.*, 34, L01809, doi:10.1029/2006GL027806, 2007.
- Worden, J., Jones, D. B. A., Liu, J., Parrington, M., Bowman, K., Stajner, I., Beer, R., Jiang, J., Thouret, V., Kulawik, S., Li, J. L. F., Verma, S., and Worden, H.: Observed vertical distribution of tropospheric ozone during the Asian summertime monsoon, *J. Geophys. Res.*, 114, D13304, doi:10.1029/2008JD010560, 2009.
- Zhao, C. and Wang, Y.: Assimilated inversion of NO_x emissions over east Asia using OMI NO₂ column measurements, *Geophys. Res. Lett.*, 36, 1–5, doi:10.1029/2008GL037123, 2009.
- Zhao, Y., Nielsen, C. P., Lei, Y., McElroy, M. B., and Hao, J.: Quantifying the uncertainties of a bottom-up emission inventory of anthropogenic atmospheric pollutants in China, *Atmos. Chem. Phys.*, 11, 2295–2308, doi:10.5194/acp-11-2295-2011, 2011.
- Ziemke, J. R., Chandra, S., Duncan, B. N., Froidevaux, L., Bharatia, P. K., Levelt, P. F., and Waters, J. W.: Tropospheric ozone determined from Aura OMI and MLS: evaluation of measurements and comparison with the Global Modeling Initiatives Chemical Transport Model, *J. Geophys. Res.*, 111, D19303, doi:10.1029/2006JD007089, 2006.