A comparison of the impact of TROPOMI and OMI tropospheric NO$_2$ on global chemical data assimilation

Takashi Sekiya$^1$, Kazuyuki Miyazaki$^2$, Henk Eskes$^3$, Kengo Sudo$^{4,1}$, Masayuki Takigawa$^1$, and Yugo Kanaya$^1$

1Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan
2Jet Propulsion Laboratory/California Institute for Technology, Pasadena, CA, USA
3Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands
4Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan

Correspondence: Takashi Sekiya (tsekiya@jamstec.go.jp)

Received: 24 November 2021 – Discussion started: 2 December 2021
Revised: 1 February 2022 – Accepted: 2 February 2022 – Published: 23 March 2022

Abstract. This study gives a systematic comparison of the Tropospheric Monitoring Instrument (TROPOMI) version 1.2 and Ozone Monitoring Instrument (OMI) QA4ECV tropospheric NO$_2$ column through global chemical data assimilation (DA) integration for the period April–May 2018. DA performance is controlled by measurement sensitivities, retrieval errors, and coverage. The smaller mean relative observation errors by 16% in TROPOMI than OMI over 60$^\circ$N–60$^\circ$S during April–May 2018 led to larger reductions in the global root-mean-square error (RMSE) against the assimilated NO$_2$ measurements in TROPOMI DA (by 54%) than in OMI DA (by 38%). Agreements against the independent surface, aircraft-campaign, and ozonesonde observation data were also improved by TROPOMI DA compared to the control model simulation (by 12%–84% for NO$_2$ and by 7%–40% for ozone), which were more obvious than those by OMI DA for many cases (by 2%–70% for NO$_2$ and by 1%–22% for ozone) due to better capturing spatial and temporal variability by TROPOMI DA. The estimated global total NO$_x$ emissions were 15% lower in TROPOMI DA, with 2%–23% smaller regional total emissions, in line with the observed negative bias of the TROPOMI version 1.2 product compared to the OMI QA4ECV product. TROPOMI DA can provide city-scale emission estimates, which were within 10% differences with other high-resolution analyses for several limited areas, while providing a globally consistent analysis. These results demonstrate that TROPOMI DA improves global analyses of NO$_2$ and ozone, which would also benefit studies on detailed spatial and temporal variations in ozone and nitrate aerosols and the evaluation of bottom-up NO$_x$ emission inventories.

1 Introduction

Satellite measurements from the Global Ozone Monitoring Experiment (GOME) (Burrows et al., 1999), the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) (Bovensmann et al., 1999), the Ozone Monitoring Instrument (OMI) (Levelt et al., 2006), and GOME-2 (Callies et al., 2000) have provided long-term global pictures of tropospheric NO$_2$ columns since 1996. Tropospheric NO$_2$ is important for air quality, atmospheric chemistry, and climate change as the main precursor of tropospheric ozone and nitrate aerosols (IPCC, 2022). Although these measurements have provided unprecedented information on global and regional NO$_2$ variations associated with changes in human and natural activity, their spatial coverage and accuracy limited their ability for a range of applications. Since October 2017, the Tropospheric Monitoring Instrument (TROPOMI) aboard the Sentinel-5 Precursor (Veefkind et al., 2012) has been measuring tropospheric NO$_2$ columns at higher spatial resolutions of $7 \times 3.5$ km$^2$ ($5.5 \times 3.5$ km$^2$ after 6 August 2019) and improved signal-to-noise ($S/N$) ratio, compared to previous satellite measurements such as OMI (Eskes et al., 2019; van Geffen et al., 2020).

Satellite NO$_2$ observations have proven useful for constraining NO$_x$ emissions, for instance, through fitting downwind line densities (e.g., Lin et al., 2012; de Foy et al.,
Using TROPOMI NO$_2$, surface NO$_2$ emissions have been estimated at high spatial and temporal resolutions, but studies are mostly limited to specific areas at point source to urban scales (Beirle et al., 2019; Goldberg et al., 2019; Lorente et al., 2019; van der A et al., 2020; Huber et al., 2020; Lange et al., 2021). In recent studies, TROPOMI NO$_2$ retrievals have also been used to provide a detailed understanding of regional and global emission reductions during the COVID-19 lockdowns (Ding et al., 2020; Miyazaki et al., 2020b, 2021; Kim et al., 2021; Zhang et al., 2021). These studies demonstrate the great potential of TROPOMI NO$_2$ for improving the spatial distribution and temporal variability of emissions. Nevertheless, its relative advantage over previous satellite measurements, such as OMI NO$_2$, in emission estimations for different regions of the world has not been clearly addressed in a quantitative and consistent manner.

Impacts of individual measurements can be evaluated using state-of-the-art data assimilation (DA) techniques, which have widely been used in numerical weather forecast applications (e.g., Gelaro and Zhu, 2009). Chemical DA systems have been used to address measurement impacts on atmospheric composition analysis, including the evaluation of air pollutant emissions (Fortems-Cheney et al., 2009; Barré et al., 2014, 2015; Emili et al., 2014; Miyazaki et al., 2012b, 2017, 2019; Zhang et al., 2019). A multi constituent chemical DA system developed by our group assimilates multiple satellite measurements simultaneously to improve emissions and concentrations of various species (e.g., Miyazaki et al., 2017, 2020a; Sekiya et al., 2021), which allows us to evaluate the relative value of TROPOMI and OMI retrievals in a consistent framework.

In this study, we compared concentration and emission analyses derived from the assimilation of TROPOMI and OMI tropospheric NO$_2$ retrievals, which simultaneously optimize tropospheric NO$_2$, ozone concentrations, and NO$_x$ emissions at 0.56° resolution for the globe. Although this resolution is still insufficient to resolve point source to urban scales, it has the advantage of providing globally consistent analyses on a megacity scale (Sekiya et al., 2021). The DA analyses were validated against assimilated and independent observations. The systematic comparison of TROPOMI DA and OMI DA reveals relative advantages of DA using TROPOMI over OMI, which benefit studies in particular on the evaluation of bottom-up emission inventories and formation processes of ozone and nitrate aerosols. The remainder of this paper is structured as follows. Section 2 describes the observation data used for the assimilation and validation and the DA system. Section 3 validates tropospheric NO$_2$ concentration analyses against assimilated and independent observations. Sections 4 and 5 present surface NO$_x$ emission analyses and their impacts on the ozone analyses, respectively. Section 6 provides a summary of the study.

## Data and methods

### 2.1 TROPOMI and OMI satellite observations of tropospheric NO$_2$ for assimilation

The TROPOMI and OMI instruments are ultraviolet/visible nadir-scanning solar-backscatter spectrometers (Levelt et al., 2006; Veffkind et al., 2012). The local Equator crossing time is approximately 13:40 LT (local time) for both instruments. The TROPOMI and OMI ground pixel sizes are $3.5 \times 7$ km$^2$ (3.5 $\times$ 5.5 km$^2$ after 6 August 2019) and 13 $\times$ 24 km$^2$, respectively. TROPOMI and OMI provide nearly global daily coverage. We used the TROPOMI NO$_2$ unofficial reprocessing product (version 1.2 beta), which is similar to the official version 1.2.2 reprocessing product (ESA, 2018; van Geffen et al., 2020), and the OMI QA4ECV version 1.1 product (Boersma et al., 2017, 2018) for the period 1 April–31 May 2018. These products were retrieved based on the differential optical absorption spectroscopy (DOAS) approach using the same wavelength window of 405–465 nm, with slight differences in the detailed settings, such as the formulation of modeled reflectance, the fitting methods, and the intensity offset correction (van Geffen et al., 2020). The TROPOMI slant column density (SCD) error for a single pixel is 30 % (20 % after 6 August 2019) lower than that of the OMI retrievals (van Geffen et al., 2020). A priori NO$_2$ profiles for TROPOMI and OMI were obtained from the TM5-MP data assimilation system at 1° $\times$ 1° resolution. Temporal changes in row anomalies (after 2007), stripes, and instrument radiometric degradation increase the uncertainty of the OMI NO$_2$ SCD by 1 % yr$^{-1}$ –2 % yr$^{-1}$ and decrease the coverage area fraction (Schenkeveld et al., 2017; Zarra et al., 2018). Therefore, the relative advantages of TROPOMI over OMI in 2018 need to be evaluated with caution.

The TROPOMI retrievals with quality assurance (QA) values of $> 0.75$ were used, which corresponds to good-quality retrievals over (nearly) cloud-free scenes. These screening criteria are similar to the criteria applied for OMI: cloud radiance fraction (CRF) of $< 0.5$, solar zenith angle (SZA) of $< 81^\circ$, surface albedo of $< 0.3$, quality flag of $= 0$, and ratio of tropospheric air mass factor (AMF) to geometric AMF of $> 0.05$. For OMI, retrievals affected by row anomalies were excluded using a quality flag. Cloud-covered scene retrievals with CRF of $> 0.5$ were separately used for optimizing lightning NO$_x$ sources, following the method proposed by Miyazaki et al. (2014).

Negative biases (by up to 50 %) against surface remote sensing observations in the TROPOMI versions 1.2 and 1.3 products were reported by Verhoelst et al. (2021). However, a large fraction of the negative biases might arise from the vertical profile shape of NO$_2$ assumed for retrievals, as reported by Dimitropoulou et al. (2020) for Uccle, Belgium. Compared to the OMI QA4ECV product, the tropospheric NO$_2$ column in the TROPOMI version 1.2 and 1.3 products is systematically lower especially for winter, as reported by


Lambert et al. (2021), which is largely attributed to a negative cloud height bias in the Fast Retrieval Scheme for Clouds from Oxygen A band (FRESCO) implementation (van Geffen et al., 2021).

2.2 Independent observations for validation

Vertical profiles and surface concentrations of NO2 and ozone derived from TROPOMI DA and OMI DA were validated against independent observations. The DA analysis fields at the closest time to measurement were sampled using two-hourly analysis output and then linearly interpolated to the observation locations from the surrounding grids in the horizontal and vertical directions. Vertical profiles were compared by averaging data within each vertical pressure bin, namely seven bins from 850 (800–900) to 250 (200–300) hPa.

2.2.1 NASA ATom aircraft-campaign observations

Vertical profiles of NO2 were obtained from the NASA Atmospheric Tomography mission 4 (ATom-4) aircraft campaign (Wofsy et al., 2018). The ATom-4 campaign was conducted using a NASA DC-8 aircraft from 24 April to 21 May 2018. The DC-8 flight tracks covered regions between 85° S and 83° N over the Pacific, Atlantic, and United States. The NO2 concentrations were measured via the NOAA NOyO3 four-channel chemiluminescence instrument per 1 s with precision of 5–10 pptv (https://espoarchive.nasa.gov/instrument/NOyO3, last access: 1 February 2022). The merged dataset of flight data with 10 s means was used for the validation. To evaluate the DA performance for vertical NO2 profiles over polluted areas, we used data averaged over coastal regions of the western United States (32–37° N, 117.25–122.5° W) from three flights on 24 and 27 April and on 21 May.

2.2.2 Surface in situ observations

We used surface NO2 and ozone concentrations from 3255 sites over Europe obtained from the European air quality database (AirBase) of the European Environmental Agency (EEA), 404 sites over the United States obtained from the Air Quality System (AQS) of the United States Environmental Protection Agency (US EPA), and 1246 sites over Japan obtained from Japanese continuous measurement data of general air pollution at ground level compiled by the National Institute for Environmental Studies (NIES). We excluded sites in high-traffic and industrial locations, because the 0.56° resolution model grids cannot resolve NO2 enhancement at roadside and individual point sources. For AirBase and AQS, sites with station types of “industrial” and “traffic” and with land use of “industrial” and “mobile” were excluded, respectively. For Japan, we excluded measurement sites for automobile exhaust gases. More than 97% of the observed NO2 concentrations used in this study were measured by commercial chemiluminescence analyzers, with typical measurement errors of 1%–5% (Gluck et al., 2003). These analyzers overestimate the ambient NO2 concentrations, because the measurements contain interference from reactive nitrogen compounds other than NO2 (e.g., Dickerson et al., 2019). Thus, correction factors proposed by Lamsal et al. (2008) using concentration analyses of HNO3, PAN, and ∑ANs were applied to the observations derived from the commercial chemiluminescence analyzers in the manner described by Sekiya et al. (2021):

\[
CF = \frac{\text{NO2}}{\text{NO2} + \sum \text{ANs} + 0.95 \times \text{PAN} + 0.35 \times \text{HNO3}}.
\]

where \(\sum \text{AN}\) is the sum of all alkyl nitrate concentrations, and PAN is the peroxyacetyl nitrate concentrations.

2.2.3 Ozonesonde observations

The observed vertical profiles of ozone were obtained from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC, https://woudc.org, last access: 1 February 2022), Southern Hemisphere Additional Ozonesondes (SHADOZ; Sterling et al., 2018; Thompson et al., 2017; Witte et al., 2017, 2018) and the National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL) Global Monitoring Division (GMD, https://gml.noaa.gov., last access: 28 February 2022). We used 127 profiles at 19 stations for the northern extratropics (20–90° N), 45 profiles at 9 stations for the tropics (20° S–20° N), and 36 profiles at 7 stations for the southern extratropics (20–90° S).

2.3 Data assimilation system

2.3.1 CHASER chemical transport model

We used the global chemical transport model, CHASER V4.0, at a resolution of 0.56° with 32 vertical layers (Sudo et al., 2002; Sekiya et al., 2018) as the forecast model, which simulates tracer transport, emission, dry and wet deposition, and chemical processes (92 species and 262 reactions) including the ozone–HOx–NOx–CO–VOCs system. The meteorological fields simulated by the dynamical and physical modules of CHASER (i.e., MIROC-AGCM atmospheric general circulation model; K-1 model developers, 2004) were nudged to the 6-hourly ERA-Interim reanalysis data (Dee et al., 2011) with a relaxation time of 5 d for temperature and 0.7 d for horizontal winds and used in the chemical module of CHASER at every time step (1–4 min). We demonstrated that increasing model resolution from the conventional resolution (2.8°) to 0.56° resolution substantially improves the model performance over polluted regions (Sekiya et al., 2018).

The a priori surface NOx emissions were obtained from the HTAPv2.2 anthropogenic emission inventory (at 0.1° resolution) for 2010 (Janssens-Maenhout et al., 2015), the Global Fire Emissions Database (GFED) version 4.1s...
monthly-based biomass burning emission inventory (at 0.25° resolution) for 2018 (Randerson et al., 2018), and the Global Emissions Initiative (GEIA) soil NOx emission inventory (at 0.5° resolution) (Yienger and Levy, 1995). These emissions are released at the lowest model layers. The a priori lightning NOx sources were calculated in the model at each model time step using the parameterization proposed by Price and Rind (1992), with the assumption for vertical distribution of lightning NOx source based on the C-shaped profile given by Pickering et al. (1998).

2.3.2 Ensemble Kalman filter data assimilation

We developed a state-of-the-art chemical DA system (e.g., Miyazaki et al., 2019, 2020a) using the local ensemble transform Kalman filter (LETKF) technique (Hunt et al., 2007), which optimizes ozone and related chemical species’ concentrations and ozone precursors’ emissions simultaneously. The LETKF uses an ensemble model forecast to estimate background error covariance assuming that the background ensemble perturbations sample the forecast model errors. The background ensemble model fields were converted into observation space by applying the observation operator which includes a spatial interpolation operator and an averaging kernel. The inclusion of averaging kernels in the observation operator describes the vertically dependent sensitivities and removes the influence of a priori profile shape (Eskes and Boersma, 2003). The analysis ensemble mean \( \tilde{x}^a \) was obtained by combining the background ensemble mean \( \tilde{x}^b \) and assimilated observations \( y^o \) with relative weights, which were determined using background and observation error covariance matrices \( X^b \) and \( R \), respectively:

\[
\tilde{x}^a = \tilde{x}^b + X^b P^a \left( Y^b \right)^T R^{-1} \left( y^o - \tilde{y}^b \right),
\]

where \( P^a \) is the local analysis error covariance in the ensemble space, while \( \tilde{y}^b \) and \( Y^b \) are the background ensemble mean and error covariance in the observation space, respectively. The local analysis error covariance was estimated as

\[
P^a = \left( \frac{k - 1}{1 + \Delta} I + \left( Y^b \right)^T R^{-1} Y^b \right)^{-1},
\]

where \( \Delta \) is a covariance inflation factor (≈ 7 % per DA cycle) and \( k \) is the ensemble size (32 or 64 in this study; see Table 1 for details).

Surface and lightning NOx emissions were estimated based on a state argumentation method (e.g., Evensen, 2009) using the relationship between NO2 concentrations and NOx emissions in the background error covariance matrix generated based on ensemble model simulations. For lightning NOx, multiplication factors for the lightning NO production rate were adjusted differently at different model layers using the method proposed by Miyazaki et al. (2014) and the background error covariance matrix. The initial a priori error was set as 40 % and 60 % for the surface and lightning NOx sources, respectively. In the analysis step, the standard deviation of emission ensembles was artificially inflated to a predefined minimum value obtained through sensitivity calculations (i.e., 56 % of a priori emissions) to prevent covariance underestimation.

Our previous study (Sekiya et al., 2021) demonstrated that DA improvements were larger by factors of 1.5–3 at 0.56° resolution than at 2.8° resolution over polluted regions in comparison to the assimilated NO2 observations. This high resolution leads to reduced spatial representativeness errors (due to an increased average coverage fraction per grid cell at 0.56° resolution by a factor of 2, compared to 2.8° resolution). The 0.56° resolution ensemble model simulation also generates a background error covariance matrix which describes small (0.56°) scale features. Because of distinct non-linear transport and chemical processes, assimilation considering the background error covariance would also be essential for making the best use of observational information. The multi-constituent DA systems have been used to assimilate ozone, NO2, CO, SO2, and HNO3 (Miyazaki et al., 2020a). Nevertheless, in this study, only TROPOMI and OMI NO2 were assimilated to emphasize the impact of assimilation of tropospheric NO2 retrievals.

2.3.3 Super-observation approach

The super-observation approach (Eskes et al., 2003; Miyazaki et al., 2012a) was used for generating satellite observation data representative to the model grid size (i.e., 0.56°). The super-observation approach can minimize spatial representativeness errors for spatially varying concentrations of short-lived tracers on sub-grid scales, such as NO2, compared to the thinning approach which randomly selects an observation per grid cell (Boersma et al., 2016). The resolution of super-observation was set to be identical to the forecast model resolution. The super-observation concentration was generated by averaging all data within a super-observation grid cell, while applying a weighting function based on the coverage area of overlap with the super-observation grid cell. The super-observation error was calculated as a combination of measurement and spatial representativeness errors \( \sqrt{\sigma_m^2 + \sigma_r^2} \). In our approach, the super-observation measurement error \( \sigma_m \) was estimated as

\[
\sigma_m = \sqrt{(1 - c) \sum_{i=1}^{n} w_i^2 \sigma_{m,i}^2 + c \sum_{i=1}^{n} (\sigma_{m,i})^2},
\]

where \( \sigma_{m,i} \) is the mean measurement error at individual pixels related to total slant column density (SCD), stratosphere–troposphere separation (STS), and tropospheric AMF; \( c \) is the error correlation coefficient among the individual retrieval data for these error components; and \( n \) is the number of measurements with non-zero overlap with the chosen grid cell. This approach explicitly accounts for spatial correlations, \( c \), between observation errors, which depend on the
mean relative super-observation errors (i.e., errors divided averaged over 60° S–60° N) during April–May 2018, with lower concentrations in TROPOMI by 15% well correlated (\( r \)) TROPOMI and OMI super-observation concentrations were compared in Fig. 1 and Table 2. The super-observation concentrations and errors can affect DA results, which are compared in Fig. 1 and Table 2. The TROPOMI and OMI super-observation concentrations were well correlated (\( r = 0.82 \) over 60° S–60° N) during April–May 2018, with lower concentrations in TROPOMI by 15% averaged over 60° S–60° N without applying averaging kernels of each other. The mean super-observation errors and mean relative super-observation errors (i.e., errors divided

### Validation results

#### 3.1 Data characteristics

Super-observation concentrations and errors can affect DA results, which are compared in Fig. 1 and Table 2. The TROPOMI and OMI super-observation concentrations were well correlated (\( r = 0.82 \) over 60° S–60° N) during April–May 2018, with lower concentrations in TROPOMI by 15% averaged over 60° S–60° N without applying averaging kernels of each other. The mean super-observation errors and mean relative super-observation errors (i.e., errors divided

---

### Table 1. List of data assimilation and control model simulation runs performed in this study.

<table>
<thead>
<tr>
<th>No.</th>
<th>Experiment</th>
<th>Period</th>
<th>Assimilated observation</th>
<th>Ensemble size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TROPOMI DA (2018, ( N = 64 ))</td>
<td>April–May 2018</td>
<td>TROPOMI v1.2 beta</td>
<td>64</td>
</tr>
<tr>
<td>2</td>
<td>OMI DA (2018, ( N = 64 ))</td>
<td>April–May 2018</td>
<td>OMI QA4ECV v1.1</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>OMI DA (2018, ( N = 32 ))</td>
<td>April–May 2018</td>
<td>OMI QA4ECV v1.1</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>OMI DA (2005)</td>
<td>April–May 2005</td>
<td>OMI QA4ECV v1.1</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>Control model simulation</td>
<td>April–May 2018</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>April–May 2005</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

The spatial representativeness error \( \sigma _{r} \) was calculated as a function of coverage area fraction in the same way as Boersma et al. (2016).

### 2.4 Experimental setup

As summarized in Table 1, four DA runs from April–May (61 d) were performed. Firstly, we conducted two DA calculations for TROPOMI and OMI separately at an ensemble size of 64. This comparison was used to investigate how TROPOMI DA improves agreements with assimilated and independent observations, compared to OMI DA (Sects. 3.2–3.4, 4, and 5). Secondly, we compared OMI DA calculations for two different years (2005 and 2018) at an ensemble size of 32. This demonstrates the impacts of OMI instrumental degradation and row anomalies, which significantly reduce daily coverage (see Sect. 2.1), on the DA performance (Sect. 3.5). In addition, a control model simulation without any DA was conducted to measure the DA impacts in each case. We chose the calculation period of April–May 2018 because of the ATom-4 aircraft-campaign data availability (see Sect. 2.2.1). Furthermore, systematic biases between the TROPOMI and OMI retrievals are known to be smaller in the summer season than those in the winter season (Lambert et al., 2020). We analyzed the DA results for the period 15 April–31 May after a 2-week-long spin-up.

### 3 Validation results

#### 3.1 Data characteristics

Super-observation concentrations and errors can affect DA results, which are compared in Fig. 1 and Table 2. The TROPOMI and OMI super-observation concentrations were well correlated (\( r = 0.82 \) over 60° S–60° N) during April–May 2018, with lower concentrations in TROPOMI by 15% averaged over 60° S–60° N without applying averaging kernels of each other. The mean super-observation errors and mean relative super-observation errors (i.e., errors divided

---

The spatial coverage per super-observation grid cell of TROPOMI (72%) was larger than that of OMI (69%) mainly because of OMI row anomalies, which led to smaller spatial representativeness errors of TROPOMI (7%) than those of OMI (10%). The mean relative super-observations errors of OMI were 8% smaller in 2005 than in 2018 (figure not shown), which is attributed to the temporal changes in OMI row anomalies, stripes, and instrument radiometric degradation (see Sect. 2.1). The averaging kernel values in the lower troposphere (below 850 hPa) were higher by 44% in TROPOMI averaged over 60° S–60° N than those in OMI, because mean CRF over 60° S–60° N is 15% smaller in TROPOMI due to better resolving small-scale cloud-free scenes.

---


Figure 1. Global distribution of tropospheric NO$_2$ column (a, d, g), super-observation errors (b, e, h), and relative super-observation errors (c, f, i) obtained from the Tropospheric Monitoring Instrument (TROPOMI; a–c) and the Ozone Monitoring Instrument (OMI; d–f) from April–May 2018, as well as the differences between TROPOMI and OMI (g–i). The values are mapped onto 0.56° resolution grids. The units of the tropospheric NO$_2$ column, super-observation errors, and relative super-observation errors are $\times 10^{15}$ molec. cm$^{-2}$, $\times 10^{15}$ molec. cm$^{-2}$, and percentage (%), respectively.

Table 2. Mean tropospheric NO$_2$ column ($\times 10^{-15}$ molec. cm$^{-2}$), super-observation error ($\times 10^{-15}$ molec. cm$^{-2}$), relative super-observation error (%) over 60°S–60°N in the Tropospheric Monitoring Instrument (TROPOMI) from April to May 2018, and the Ozone Monitoring Instrument (OMI) from April to May in 2005 and 2018. Values in brackets are calculated from TROPOMI and OMI data with co-location criteria of < 60 km in space and < 2 h in time.

<table>
<thead>
<tr>
<th>Satellite observations</th>
<th>Tropospheric column</th>
<th>Super-observation error</th>
<th>Relative super-observation error</th>
</tr>
</thead>
<tbody>
<tr>
<td>TROPOMI</td>
<td>0.52 (0.54)</td>
<td>0.28 (0.27)</td>
<td>53 (51)</td>
</tr>
<tr>
<td>OMI (2018)</td>
<td>0.63 (0.64)</td>
<td>0.40 (0.38)</td>
<td>63 (60)</td>
</tr>
<tr>
<td>OMI (2005)</td>
<td>0.57</td>
<td>0.33</td>
<td>58</td>
</tr>
</tbody>
</table>

3.2 Self-consistency

The performance of TROPOMI DA and OMI DA was confirmed by the $\chi^2$ test (Ménard and Chang, 2000; Zupanski and Zupanski, 2006). The $\chi^2$ value is used to diagnose balance between actual errors and estimated errors. When the $\chi^2$ value is larger (smaller) than the ideal value of 1, it suggests underestimated (overestimated) background error covariance or observation errors. The $\chi^2$ value is diagnosed from the ratio of the observation minus forecast (OmF; i.e., $y^o - H(x^b)$) to estimated error covariance in the observational space $(H P^b H^T + R)$ as

$$Y = \frac{1}{\sqrt{N}} \left( H P^b H^T + R \right)^{-1/2} \left( y^o - H(x^b) \right)$$

(5)
\[ \chi^2 = \text{trace } YY^T. \]  

The mean values of estimated \( \chi^2 \) with standard deviation range over polluted regions (> \( 1 \times 10^{15} \) molec. cm\(^{-2}\)) after inflation factor tuning was 0.99 \( \pm \) 0.25 for TROPOMI DA, whereas the mean \( \chi^2 \) of 1.17 \( \pm \) 0.19 for OMI DA is larger than the ideal value.

We also evaluated the self-consistency with the assimilated observations based on reductions in root-mean-square error (RMSE) by DA (\( \Delta \text{RMSE} \)) using daily maps sampled at observation locations as

\[
\Delta \text{RMSE} = - \frac{\left( \sum_{i=1}^{N} (A_{\text{cassim}} - V) - \sum_{i=1}^{N} (A_{\text{ctl}} - V) \right)^2}{N}
\]

\[
+ \frac{\sum_{i=1}^{N} (A_{\text{cassim}} - V)^2}{N} \right) / \left( \sqrt{\frac{\sum_{i=1}^{N} (A_{\text{ctl}} - V)^2}{N}} \right). \tag{7}
\]

where \( V \) and \( A \) are the observed tropospheric NO\(_2\) column and corresponding averaging kernels, respectively, used for DA; \( c_{\text{cassim}} \) and \( c_{\text{ctl}} \) are NO\(_2\) concentration fields obtained from the DA runs and control model simulations, respectively; and \( N \) is the number of super-observation data. The level of significance of \( \Delta \text{RMSE} \) was determined using the Mann–Whitney \( U \) test (Mann and Whitney, 1947).

As shown in Fig. 2 and Table 3, the RMSE for TROPOMI DA over 60\(^\circ\)S–60\(^\circ\)N was reduced by 54 \% compared to that for the control model simulation, with larger RMSE reductions over polluted regions (by 60 \%) than over remote regions (by 37 \%). The RMSE reductions were substantial over most regions in the tropics and northern midlatitudes, whereas improvements are not clear over the northern high latitudes, Tarim Basin, Arabian Sea, northern Australia, South America, and parts of the southern midlatitudes. Mean RMSE reductions were larger for TROPOMI DA than OMI DA (by 38 \%). The differences in RMSE reductions between TROPOMI DA and OMI DA over the tropics and northern midlatitudes were statistically significant at the 95 \% confidence level. These differences can be explained by the reduced relative super-observation errors in TROPOMI. Over the oceans in the tropics and midlatitudes, higher vertical sensitivity (i.e., averaging kernels) in TROPOMI than OMI in the lower troposphere and above clouds contributed to the improved performance, through ship and lightning NO\(_x\) emission corrections and direct NO\(_2\) concentration modifications. In contrast, the differences in RMSE reduction between TROPOMI DA and OMI DA were statistically insignificant over most regions with larger relative super-observation errors in TROPOMI.

The two-dimensional histogram of grid-level relative super-observation errors and RMSE reductions (Fig. 3a and b) shows clear decreases in RMSE reductions with increasing relative super-observation errors for both TROPOMI and OMI DA. Steep RMSE decreases occurred around relative super-observation errors of 20–50 \%, which reflected areas over and downwind of polluted regions. Over polluted regions, observational information is more effectively assimilated into the model, because of the large uncertainty (i.e., background error covariance) of estimated NO\(_x\) emissions over these regions. As shown in Fig. 3c, mean relative super-observation errors at individual grids were smaller than those in TROPOMI in OMI by 16 \%. Corresponding to these smaller super-observation errors, the mean RMSE reductions by TROPOMI DA at individual grids (by 54 \%) were larger than those by OMI DA (by 38 \%), with large differences in frequency of RMSE reductions between TROPOMI DA and OMI DA for RMSE reductions of > 10 \% (Fig. 3d). These results confirm that improved RMSE reductions by TROPOMI DA compared to OMI DA can be attributed to the reduced relative super-observation errors in TROPOMI. Meanwhile, the obtained result suggests that the DA efficiency by TROPOMI was evaluated based on RMSE against assimilated observation itself. It is determined by the amount and quality of TROPOMI data, regardless of the TROPOMI low bias.

### 3.3 Validation against independent observations

#### 3.3.1 ATom aircraft-campaign data

Figure 4 and Table 4 validated the vertical profiles of NO\(_2\) with the ATom-4 aircraft-campaign observations on 24 and 27 April and 21 May when the DC-8 aircraft flew over coastal areas of the western United States (32–37\(^\circ\)N, 117.25–122.5\(^\circ\)W). For the comparison, model simulation and data assimilation results were sampled at observation locations, and then the observation data, the control model simulation, and the data assimilation were averaged on each day over the coastal areas of the western United States. The control model simulation overestimated the observed concentrations in the lower troposphere (700–900 hPa) by factors of 1.4–4 in all cases, while underestimating the NO\(_x\) concentrations in the middle and upper troposphere (300–700 hPa) by 48 \%–70 \%, except on 27 April. The positive model biases were particularly large at 750 hPa on 24 April and at 850 hPa on 27 April and 21 May. The use of 2010 anthropogenic NO\(_x\) emissions could explain the positive model biases. In addition, on 24 April, the simulated planetary boundary layer (PBL) height was 30 \% higher than that in the ERA-Interim reanalysis, which could in turn increase NO\(_2\) bias at 750 hPa.

On 24 April, TROPOMI DA increased negative bias at 850 hPa, while it reduced positive bias at 750 hPa, which could also be attributable to model biases in PBL height. The mean bias in the lower troposphere (below 700 hPa) was largely reduced by TROPOMI DA (by 84 \%) on 24 April. The improvements were small (by 17 \%) on 27 April when the DC-8 aircraft NO\(_2\) measurements were conducted in the early morning before the TROPOMI overpass time, whereas...
TROPOMI DA reduced positive model biases by 78 % in the lower troposphere on 21 May. In the middle and upper troposphere, TROPOMI DA reduced the model biases by 12 %–53%. These bias reductions were larger by 52 %–70 % for the lower troposphere and by up to 31 % for the middle and upper troposphere in TROPOMI DA than OMI DA, except for the lower troposphere on 27 April. Because of the large variability in the observed concentrations, these differences in bias were statistically insignificant based on a two-sample t test, except for the upper troposphere on 21 May.

### 3.3.2 Surface in situ data

Surface in situ observation data at 14:00 LT were used for validation to evaluate assimilation impacts just after their overpass times. Validation was conducted after filtering out model grids where water bodies cover > 50 % of a grid box area using annual Moderate Resolution Imaging Spectroradiometer (MODIS) land cover data (Friedl and Sulla-Menashe, 2015) for 2018, considering large representativeness errors. Over Europe, the regional mean model bias and RMSE of NO₂ were ~18 % and 145 %, respectively (absolute errors are shown in Table 5). The model biases vary with regions, with positive biases of 12 %–115 % over the United Kingdom (UK), Belgium, and the Netherlands and negative biases of 42 %–78 % over Italy, Serbia, and Romania (Fig. 5). Over the United States, regional mean model bias and RMSE were 37 % and 268 %, respectively, with larger positive biases over urban areas such as New York, Los Angeles, and Chicago. The regional mean bias and RMSE over Japan were ~23 % and 124 %, respectively.

TROPOMI DA reduced the regional RMSE over Europe by 29 %, with larger RMSE reductions by 45 % and 47 % over the UK and the Netherlands, respectively, reflecting improvements in spatial and temporal variability by TROPOMI DA (Fig. 6). Because of the small RMSEs in the control model simulation, RMSE reductions by TROPOMI DA were not obvious over Italy, Spain, and Portugal. Over the United States, TROPOMI DA reduced the regional mean bias and RMSE by 46 % and 50 %, respectively. In contrast to the
large RMSE reductions over the eastern United States and western coastal areas, RMSEs increased over Colorado and Wyoming again due to the small RMSEs in the control model simulations. Over Japan, TROPOMI DA reduced RMSE by 23 % due to better capturing spatial and temporal variations but increased negative model bias by 68 %. Error reductions were smaller in OMI DA overall. The RMSE over Europe was increased by OMI DA by 5 % mainly due to the increased errors over the Netherlands. Over the United States and Japan, the RMSE reductions for megacities such as New York, Los Angeles, and Tokyo were 25 %–70 % larger in TROPOMI DA than in OMI DA. The regional RMSE reduc-


Figure 3. Two-dimensional (2-D) histogram (%) as a function of relative super-observation errors (%) and root-mean-square error (RMSE) reductions (%) for tropospheric NO$_2$ column made to data assimilation (DA) (a, b) obtained from Tropospheric Monitoring Instrument (TROPOMI) DA (a, c) and Ozone monitoring Instrument (OMI) DA (b, d). The values in the top panels are regression coefficients, intercepts, and correlation coefficients between relative super-observation errors and RMSE reductions against assimilated observations by DA on the grid scale. The values in brackets were calculated using data with relative super-observation errors of < 50 %. One-dimensional (1-D) histograms as a function of relative super-observation errors (c) and RMSE reductions (d) are exhibited in the bottom row. The black and red lines are taken from TROPOMI DA and OMI DA, respectively.

Figure 4. Vertical profiles of NO$_2$ (pptv) on 24 April (a), 27 April (b), and 21 May (c) over coastal areas of the western United States (32–37° N, 117.25–122.5° W). The results were obtained from the ATom-4 aircraft-campaign observations (black), Tropospheric Monitoring Instrument (TROPOMI) data assimilation (DA) (red), Ozone Monitoring Instrument (OMI) DA (blue), and the control model simulation (yellow). The error ranges are the standard deviations of individual values in each pressure bin.
Table 4. Mean bias (MB) against the ATom-4 aircraft-campaign observations in the lower (700–900 hPa) and middle–upper troposphere (300–700 hPa) over coastal areas of the western United States (32–37° N, 117.25–122.5° W) on 24 and 27 April and 21 May 2018. MB is expressed in parts per trillion by volume (pptv). Boldface denotes the best agreement with in situ observations. The ranges of MB are the standard deviations of individual values in each pressure bin. “Model” indicates the control model simulation.

<table>
<thead>
<tr>
<th></th>
<th>24 April</th>
<th>27 April</th>
<th>21 May</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>700–900 hPa</td>
<td>300–700 hPa</td>
<td>700–900 hPa</td>
</tr>
<tr>
<td>TROPOMI DA</td>
<td>−33.5 ± 516.0</td>
<td>−21.2 ± 144.1</td>
<td>166.0 ± 295.1</td>
</tr>
<tr>
<td>OMI DA</td>
<td>103.5 ± 429.9</td>
<td>−22.7 ± 144.8</td>
<td>159.2 ± 278.8</td>
</tr>
<tr>
<td>Model</td>
<td>216.0 ± 487.4</td>
<td>−26.4 ± 146.0</td>
<td>200.4 ± 419.1</td>
</tr>
</tbody>
</table>

Figure 5. Mean model biases against surface in situ observations for surface NO₂ concentrations (%) at 14:00 LT (local time) derived from the control model simulation (left column), Tropospheric Monitoring Instrument (TROPOMI) data assimilation (DA) (middle column), and Ozone Monitoring Instrument (OMI) DA (right column) over Europe (top), the United States (middle), and Japan (bottom). The values are mapped onto 0.56° resolution grids.

3.4 Regional performance over Los Angeles

The magnitude of improvements by DA can be affected by meteorological conditions (e.g., Miyazaki et al., 2019). We evaluated impacts of meteorological conditions on the self-consistency over Los Angeles where both independent surface and aircraft-campaign observations were available (Sect. 3.3.1 and 3.3.2). During 15 April–31 May, southwesterly winds were predominant over Los Angeles, while wind speed varied (see Fig. 8). As shown in Fig. 7, the RMSEs...
Table 5. Mean bias (MB) and root-mean-square error (RMSE) for surface NO\textsubscript{2} at 14:00 LT (local time) and daily maximum 8 h average (MDA8) ozone against in situ observations during 15 April–31 May 2018. The units of MB and RMSE are parts per billion by volume (ppbv). Boldface denotes the best agreement with in situ observations. The ranges of MB and RMSE are the standard deviation of the time series. “Model” indicates the control model simulation.

<table>
<thead>
<tr>
<th>Region</th>
<th>Run</th>
<th>NO\textsubscript{2}</th>
<th>MDA8 ozone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MB</td>
<td>RMSE</td>
</tr>
<tr>
<td>Europe</td>
<td>TROPOMI DA</td>
<td>$-0.43 \pm 0.44$</td>
<td>$1.71 \pm 0.54$</td>
</tr>
<tr>
<td></td>
<td>OMI DA</td>
<td>$-0.12 \pm 1.02$</td>
<td>$2.53 \pm 0.46$</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>$-0.30 \pm 0.80$</td>
<td>$2.41 \pm 0.46$</td>
</tr>
<tr>
<td>United States</td>
<td>TROPOMI DA</td>
<td>$-0.29 \pm 0.51$</td>
<td>$1.93 \pm 0.56$</td>
</tr>
<tr>
<td></td>
<td>OMI DA</td>
<td>$-0.08 \pm 0.55$</td>
<td>$2.02 \pm 0.57$</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>$0.53 \pm 1.55$</td>
<td>$3.87 \pm 0.47$</td>
</tr>
<tr>
<td>Japan</td>
<td>TROPOMI DA</td>
<td>$-0.80 \pm 0.93$</td>
<td>$1.91 \pm 0.61$</td>
</tr>
<tr>
<td></td>
<td>OMI DA</td>
<td>$-0.49 \pm 0.95$</td>
<td>$1.98 \pm 0.61$</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>$-0.47 \pm 1.35$</td>
<td>$2.47 \pm 0.50$</td>
</tr>
</tbody>
</table>

Figure 6. Root-mean-square errors (RMSEs) against surface in situ observations for surface NO\textsubscript{2} concentration fields (ppbv) at 14:00 LT (local time) in the control model simulation (left column) and their reductions (%) by Tropospheric Monitoring Instrument (TROPOMI) data assimilation (DA) and Ozone Monitoring Instrument (OMI) DA (middle and right columns, respectively) over Europe (top), the United States (middle), and Japan (bottom). The values are mapped onto 0.56° resolution grids. For the middle and right columns, grids with open circles indicate RMSE reductions that are statistically significant at the 95% confidence level using the Mann–Whitney U test.
Figure 7. Root-mean-square error (RMSE) reduction for tropospheric NO$_2$ concentration fields against assimilated observations by data assimilation (DA) over Los Angeles under windy (wind speed > 2.5 m s$^{-1}$, top row) and calm (bottom row) conditions. The left, middle, and right columns show Tropospheric Monitoring Instrument (TROPOMI) DA, Ozone Monitoring Instrument (OMI) DA, and the differences between them, respectively. The unit is percentage (%). Arrows in the left and middle columns show surface winds derived from ERA-Interim reanalysis data. For the right column, grids with open circles indicate the differences in RMSE reductions between TROPOMI DA and OMI DA that are statistically significant at the 95 % confidence level using the Mann–Whitney $U$ test. The black circle indicates the location of Los Angeles city center.

over the city of Los Angeles (black rectangles of Fig. 7) were reduced by TROPOMI DA in windy conditions (wind speed > 2.5 m s$^{-1}$) by 46 % and calm conditions by 37 %. Over the city of Los Angeles, the RMSE reductions by TROPOMI DA were larger under the windy conditions (by a factor of 1.3) compared with OMI DA, with statistical significance at the 99 % confidence level; the RMSE reductions were comparable under the calm conditions (within 5 %). The TROPOMI measurements with high vertical sensitivity (i.e., averaging kernels) in the lower troposphere captured the dilution of NO$_2$ over the city of Los Angeles during the windy conditions better than OMI, which resulted in better improvements by TROPOMI DA than OMI DA under the windy conditions. The vertical sensitivity of TROPOMI in the lower troposphere over the city of Los Angeles was 36 % higher than that of OMI, reflecting a smaller CRF in TROPOMI than in OMI (by 17 %), because TROPOMI has higher spatial resolution and better resolves small-scale cloud-free scenes over Los Angeles. In contrast, the super-observation errors and quantities in TROPOMI and OMI during the windy conditions were comparable to those during the calm conditions (figure not shown). Meanwhile, over the surrounding areas of the city of Los Angeles, RMSE reductions by TROPOMI DA and OMI DA were comparable under both the windy and calm conditions.

As summarized in Table 6, positive model bias in NO$_2$ concentrations against in situ observations for 15 April–31 May 2018 was 33 %. Temporal correlation coefficient between observed and simulated concentrations was 0.49. TROPOMI DA introduced negative bias while improving the temporal correlation to 0.63. RMSEs were reduced by 37 % and 26 % during the calm and windy conditions, respectively. The negative bias was larger in TROPOMI DA than in OMI DA, whereas temporal correlation coefficients in TROPOMI DA ($r = 0.63$) were larger than those in OMI DA ($r = 0.25$). The RMSE reductions by TROPOMI DA were 8 % larger than those by OMI DA under the windy conditions, whereas the RMSE reductions by TROPOMI DA and OMI DA were comparable under the calm conditions. These results suggest similar improvements by TROPOMI DA compared to OMI DA under the windy and calm conditions, while meteorological conditions slightly affect the magnitude of improvements in NO$_2$ concentrations by TROPOMI DA compared to OMI DA.
Table 6. Mean bias (MB), temporal correlation coefficient (T-Corr.), and root-mean-square error (RMSE) for surface NO$_2$ at 14:00 LT (local time) against in situ observations from the Air Quality System (AQS) over Los Angeles from 15 April–31 May 2018 under all, calm (wind speed $\leq$ 2.5 m s$^{-1}$), and windy conditions. Both MB and RMSE are expressed in parts per billion by volume (ppbv). Boldface denotes the best agreement with in situ observations. The range of the MB is the standard deviation of the time series. “Model” indicates the control model simulation.

<table>
<thead>
<tr>
<th></th>
<th>All conditions</th>
<th>Calm condition</th>
<th>Windy condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>TROPOMI DA</td>
<td>$-1.69 \pm 1.60$</td>
<td>0.63</td>
<td>2.33</td>
</tr>
<tr>
<td>OMI DA</td>
<td>$-1.11 \pm 2.28$</td>
<td>0.25</td>
<td>2.54</td>
</tr>
<tr>
<td>Model</td>
<td>1.88 $\pm 2.80$</td>
<td>0.49</td>
<td>3.37</td>
</tr>
</tbody>
</table>

Figure 8. Surface NO$_2$ concentrations (ppbv) at 14:00 LT over Los Angeles. The results were obtained from in situ observations (black), Tropospheric Monitoring Instrument (TROPOMI) data assimilation (DA) (red), Ozone Monitoring Instrument (OMI) DA (blue), and the control model simulation (yellow). The periods filled in pink are windy conditions (wind speed $> 2.5$ m s$^{-1}$).

3.5 Impact of OMI instrumental degradation

Temporal changes in OMI row anomalies, stripes, and instrument radiometric degradation from 2005 to 2018 could affect OMI DA results. Thus, we compared OMI DA results between 2005 and 2018. As summarized in Table 3, the RMSE reduction over polluted regions in 2018 (by 48 %) was larger than that in 2005 (by 41 %) with statistical significance at the 99 % confidence level. The multi-year difference in DA performance is likely driven by inter-annual variations in meteorological conditions rather than by OMI degradation. Over Europe, the United States, and China, the number of cloud-free scenes in 2018 was increased by 11 %–19 % compared to those in 2005. In contrast, RMSE reductions over remote regions were similar (23 % in 2018 and 25 % in 2005). Such interannual changes in cloud cover can affect the overall OMI DA performance, which needs to be considered in the TROPOMI and OMI comparison results for 2018. Nevertheless, the improvements against assimilated observations by TROPOMI DA (by 54 %) were larger than those by OMI DA for both years 2005 (by 34 %) and 2018 (by 38 %).

TROPOMI DA clearly shows better performance compared to OMI DA for the periods before instrumental degradation, even when considering inter-annual variations in meteorological conditions.

4 NO$_x$ emission estimates

The top-down estimates provided by TROPOMI DA significantly differed from the a priori emissions (Fig. 9 and Table 7). TROPOMI DA tends to decrease emissions over the eastern United States, China, northern India, and Central Africa. Large positive increments (by 42 % on average) were found over regions where soil emissions are dominant (> 50 % in a priori emissions), such as on remote areas of Spain, Turkey, the Midwestern United States, Kazakhstan, and the Sahel regions. This suggests underestimated soil emissions in a prior inventories, as commonly reported by previous studies (Vinken et al., 2014; Oikawa et al., 2015; Visser et al., 2019). The country and regional total emissions were decreased by 14 % in the United States, 38 % in China, 17 % in India, and 22 % in Central Africa, and they were increased by 12 % in Europe, 39 % in the Middle East, and 44 % in Southeast Asia.

The global total NO$_x$ emissions were 15 % smaller in TROPOMI DA than in OMI DA, with 3 %–18 % smaller regional total emissions for polluted regions and 22 %–23 % smaller regional total emissions for biomass burning regions, which led to smaller surface NO$_2$ concentrations (Fig. 9). These differences reflect the low bias of TROPOMI retrievals compared to OMI retrievals. The low bias of the TROPOMI retrievals compared to the OMI retrievals also affects OH concentrations. Assimilation of lower NO$_2$ retrievals, through NO$_x$ emission and NO$_2$ concentration optimization, led to weaker chemical production of HO$_2$ and conversion from HO$_2$ to OH. This effect resulted in 2 %–21 % smaller regional-mean OH concentrations in the lower troposphere in TROPOMI DA, except for South Africa. In contrast, differences in the regional total emissions over India and the Middle East between TROPOMI DA and OMI DA were small (4 %–5 %), reflecting small differences in regional-mean concentrations between the TROPOMI and
OMI retrievals (4%–6% lower in TROPOMI). Compared to the EDGARv5 (Crippa et al., 2019) and REASv3.2 (Kurokawa and Ohara, 2020) bottom-up emission inventories for 2015, the regional total emissions from TROPOMI DA and OMI DA over major polluted regions, except for Europe, were smaller by 17%–35% and 9%–21%, respectively. These results suggest that the emission estimates from OMI DA are closer to the EDGARv5 and REASv3.2 bottom-up emission inventories than those from TROPOMI DA (using the TROPOMI v1.2 beta product).

The NO$_x$ emissions derived from TROPOMI DA were compared with previous estimates over large urban areas based on statistical fits of NO$_2$ line density data with the exponentially modified Gaussian (EMG) function using TROPOMI NO$_2$ (Beirle et al., 2019; Lorente et al., 2019; Goldberg et al., 2019; Lange et al., 2021). We focused on large urban areas where at least two estimates were avail-
able. For this comparison, a posteriori emissions from our TROPOMI DA estimates were integrated inside a square of 100 × 100 km² centered on the selected urban area, while the uncertainty information was obtained from the analysis ensemble spreads. As summarized in Table 8, our estimates were in good agreement with the previous estimates within 10 % differences for Riyadh, Chicago, and New York compared to the estimates from Beirle et al. (2019) and Lange et al. (2021). Nevertheless, lower emissions in our estimates by 18 %-66 % over Chicago, New York, and Toronto than the estimates from Goldberg et al. (2019) could be explained by the difference in the TROPOMI NO₂ AMF calculation, which was replaced by Environment and Climate Change Canada (ECCC) with their high-resolution regional chemical transport model (CTM) and the MODIS surface reflectance (Griffin et al., 2019). For Paris, our estimate in late spring was lower by 35 % and 41 % than those by Lorente et al. (2019) and Lange et al. (2021), respectively, but analyzed for different time periods in winter and annually, respectively. Increases in NOₓ emissions during cold seasons are because of residential heating (Lorente et al., 2019). Overall, these results imply that top-down NOₓ emission estimates using TROPOMI version 1.2–1.3 products could be affected by the TROPOMI low biases compared to OMI, while top-down estimates using TROPOMI have the potential for constraints on detailed spatial and temporal variations based on validation results (see Sect. 3.3).

Cloud-covered scenes of satellite NO₂ retrievals were used to optimize lightning NOₓ sources following the method of Miyazaki et al. (2014), which provide important constraints on tropospheric chemistry including ozone (e.g., Boersma et al., 2005; Miyazaki et al., 2014; Allen et al., 2021). Because of its high spatial resolution, TROPOMI NO₂ retrievals offer the potential for better resolving small-scale cloud-covered scenes (Marais et al., 2021) and constraining lightning NOₓ sources. As a result, the difference between TROPOMI DA and OMI DA can be attributed to 75 %–92 % higher vertical sensitivities above the cloud height for cloud-covered scenes (CRF > 0.5). As shown in Table 7, the global total production of lightning NO estimated by TROPOMI DA was 13 % larger than that estimated by OMI DA, with larger regional total production by 14 %-52 % over North and South America, Southeast Asia, the Atlantic, and the Indian Ocean. The impacts of TROPOMI on lightning NOₓ source estimation will be investigated in more detail in a separate study.

5 Impacts on ozone analysis

The NO₂ DA plays an important role in improving the representation of tropospheric chemistry, including ozone. We evaluated the relative values of TROPOMI and OMI NO₂ DA on surface and tropospheric ozone analysis.

5.1 Validation using surface in situ data

Daily maximum 8 h average (MDA8) ozone concentrations were validated using surface in situ observation data in the same manner as NO₂. As summarized in Table 5, the regional mean bias and RMSE of the control model simulation against surface in situ ozone observations over Europe were 22 % and 29 %, respectively, with large RMSEs over southern Europe (Fig. 10). Over the United States, the mean model bias and RMSE were 10 % and 22 %, respectively, reflecting large model biases over the eastern United States. The mean bias and RMSE over Japan were 7 % and 18 %, respectively. Positive model biases in surface ozone over polluted regions are commonly reported in other global CTMs (Schnell et al., 2015; Turnock et al., 2020).

TROPOMI DA increased the mean bias and RMSE of surface ozone over Europe by 14 % and 8 %, respectively, with large error increases over southern Europe. Increased bias and RMSE were also found over the western United States. Many factors can lead to increased errors in ozone, including model errors in ozone precursors’ emissions other than NOₓ, chemical processes, and meteorological processes such as PBL vertical mixing. For example, ozone responses to NOₓ emissions strongly depend on the choice of CTMs (Miyazaki et al., 2020c), which affects the impacts of NOₓ emission corrections on ozone analyses. Over the eastern United States, RMSEs were typically reduced by 5–10 ppbv by TROPOMI DA, resulting in improved bias by 14 % and RMSE by 16 % at the country scale. Over Japan, the mean bias and RMSE were reduced by TROPOMI DA by 54 % and 7 %, respectively.

For most regions, better agreement with surface ozone data was obtained from TROPOMI DA than from OMI DA. OMI DA resulted in larger increases in the positive bias and RMSE over Europe by 32 % and 22 %, respectively. Over the United States, the mean bias and RMSE were slightly increased by OMI DA by 8 % and 1 %, respectively. The bias reduction over Japan by TROPOMI DA was larger than that by OMI DA (by 18 %), while RMSE over Japan was increased by OMI DA (by 4 %). The better agreement in surface ozone by TROPOMI DA coincides with that in surface NO₂ (see Sect. 3.3.2). This confirms that the better representation of NO₂ through assimilation of advanced NO₂ satellites is essential to improve surface ozone analysis for many regions of the world. Meanwhile, any biases in satellite NO₂ retrievals affect surface ozone analysis. Surface ozone analysis biases are expected to be increased for a NOₓ-limited ozone chemical regime when using updated retrievals with reduced TROPOMI NO₂ negative bias.

5.2 Validation using ozonesonde data

Here we focus on the NO₂ DA impacts on free tropospheric ozone. Mean negative model biases of ozone at 500 and 800 hPa against ozonesonde observations were 9.5 % and
Table 8. NOx emission estimates in large urban areas obtained from this study and previous studies. The unit is megagrams (Mg) per hour. Emissions are averaged for the period 15 April–31 May 2018 for this study, March 2018–November 2020 for Lange et al. (2021), December 2017–October 2018 for Beirle et al. (2019), May–September 2018 for Goldberg et al. (2019), and February–June 2018 for Lorente et al. (2019). The ranges of the emissions are emission analysis spreads in this study, while the ranges are the errors estimated by individual previous studies.

<table>
<thead>
<tr>
<th>City</th>
<th>This study</th>
<th>Lange et al. (2021)</th>
<th>Beirle et al. (2019)</th>
<th>Goldberg et al. (2019)</th>
<th>Lorente et al. (2019)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riyadh (24.6° N, 46.7° E)</td>
<td>21.5 ± 0.9</td>
<td>21.8 ± 0.8</td>
<td>23.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicago (41.8° N, 87.8° W)</td>
<td>11.0 ± 2.4</td>
<td>12.1 ± 1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York (40.7° N, 74.0° W)</td>
<td>14.6 ± 2.5</td>
<td>14.7 ± 1.5</td>
<td>17.9 ± 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toronto (43.7° N, 79.4° W)</td>
<td>4.9 ± 2.9</td>
<td>7.6 ± 0.5</td>
<td>14.3 ± 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paris (48.9° N, 2.3° E)</td>
<td>5.2 ± 2.7</td>
<td>8.0 ± 0.5</td>
<td></td>
<td></td>
<td>8.8</td>
</tr>
</tbody>
</table>

Figure 10. Root-mean-square errors (RMSEs) against surface in situ observations (ppbv) for daily maximum 8 h average (MDA8) ozone concentrations in the control model simulation (left), as well as their reductions by Tropospheric Monitoring Instrument (TROPOMI) data assimilation (DA) and Ozone Monitoring Instrument (OMI) DA (middle and right, respectively) over Europe (top), the United States (middle), and Japan (bottom). The values are mapped onto 0.56° resolution grids. For the middle and right columns, grids with open circles indicate the RMSE reductions that are statistically significant at the 95 % confidence level using the Mann–Whitney U test.
We compared DA analyses of NO\(_2\), ozone concentrations, and NO\(_x\) emissions derived from the assimilation of the TROPOMI and OMI tropospheric NO\(_2\) column retrievals. To generate observation data representative to the model grid size, we employed a super-observation approach that explicitly accounts for spatial correlations between observation errors. Because of 16% smaller relative super-observation errors in TROPOMI than in OMI, the DA self-consistency, as measured by RMSE reductions against the assimilated observations, was improved in TROPOMI DA by 54%, which was larger than OMI DA (by 38%). Agreements against the independent ATom-4 aircraft-campaign and surface in situ NO\(_2\) data were also improved by 12%–84% and 23%–50%, respectively, which was larger than those for OMI DA (by up to 70% and 47%, respectively) for many cases. The improved NO\(_2\) led to improved agreement with surface in situ MDA8 ozone over the United States and Japan in TROPOMI DA (by 7%–40%) than in OMI DA (by 1%–22%). Agreements with ozonesonde data at 500 and 800 hPa were also improved by TROPOMI DA by 14%–40% for most regions, except for the tropics, which was larger than those by OMI DA (by 12%–22%).

Global total NO\(_x\) emission for 15 April–31 May 2018 was increased from 43.5 Tg N yr\(^{-1}\) in a priori emissions to 46.2 Tg N yr\(^{-1}\) by TROPOMI DA, which was 15% smaller than those derived from OMI DA (54.2 Tg N), with 3%–23% smaller regional total emissions for major polluted and biomass burning areas. The city-scale emissions derived from TROPOMI DA were generally consistent with previous estimates using limited-area high-resolution analyses (within 10% differences for Riyadh, New York, and Chicago). The global emission estimates constrained by the more accurate and dense TROPOMI measurements provide complementary information about emission variability, especially where accurate and detailed information on activity data and emission factors is missing when developing bottom-up inventories (Elguindi et al., 2020). This would also benefit model simulations of tropospheric ozone (e.g., Miyazaki et al., 2019; Visser et al., 2019; Bae et al., 2020; Qu et al., 2020) and estimations of nitrate aerosols and their deposition flux (Nowlan et al., 2014; Geddes and Martin, 2017). These improvements are important for productivities and diversities of terrestrial and marine ecosystems.

The DA performance comparisons provide a systematic evaluation of TROPOMI and OMI retrievals, independent from their averaging kernels and a priori profiles. The improved agreements with independent observations in TROPOMI DA demonstrate the importance of improved spatial coverage and reduced retrieval uncertainty for many science applications. Along with the improved spatial patterns, the impact of systematic biases in the retrievals needs to be carefully evaluated. Validation against surface NO\(_2\) measurements confirmed lower bias in TROPOMI retrievals compared to OMI retrievals by 15% for the United States, Europe, and Japan, consistent with previous validation results (Lambert et al., 2021). The smaller estimated NO\(_2\) emissions also confirm the low biases in TROPOMI NO\(_2\) relative to OMI NO\(_2\) globally, which in turn affected ozone analysis.

The systematic differences of TROPOMI version 1.2 compared to ground-based remote sensing and OMI are larger in winter than in other seasons over the polluted regions (Verhoelst et al., 2021; Lambert et al., 2021), consistent with Appendix A. The influence of negative biases related to the a priori profile shape area removed by using averaging kernels. However, because of the larger TROPOMI (version 1.2) negative bias in winter than in April–May, the relative DA performance between TROPOMI and OMI will depend on the season, especially over heavily polluted areas. Because of the availability of aircraft-campaign observational data for validation and the active photochemical production during the target period, this study focused on April–May 2018 only, and the impact of seasonally varying relative biases between OMI and TROPOMI has not been investigated. The low bias in TROPOMI version 1.2.2 compared to OMI is largely attributed to a negative cloud height bias in the FRESCO cloud retrieval algorithm (van Geffen et al., 2021). The updated TROPOMI NO\(_2\) products were introduced in Decem-

Table 9. Mean bias (MB) and root-mean-square error (RMSE) for ozone concentrations at 500 and 800 hPa against ozonesonde observations over three latitude bands from 15 April–31 May 2018. The units of MB and RMSE are in parts per billion by volume (ppbv). Boldface denotes the best agreement with in situ observations. “Model” indicates the control model simulation.

<table>
<thead>
<tr>
<th>Latitude bands</th>
<th>Run</th>
<th>500 hPa</th>
<th>800 hPa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MB</td>
<td>RMSE</td>
<td>MB</td>
</tr>
<tr>
<td>20–90° N</td>
<td>TROPOMI DA</td>
<td>-0.10</td>
<td>12.22</td>
</tr>
<tr>
<td></td>
<td>OMI DA</td>
<td>0.56</td>
<td>12.47</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>-6.59</td>
<td>16.03</td>
</tr>
<tr>
<td>20° S–20° N</td>
<td>TROPOMI DA</td>
<td>7.00</td>
<td>10.70</td>
</tr>
<tr>
<td></td>
<td>OMI DA</td>
<td>10.22</td>
<td>13.78</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>-5.93</td>
<td>10.05</td>
</tr>
<tr>
<td>20–90° S</td>
<td>TROPOMI DA</td>
<td>1.51</td>
<td>3.88</td>
</tr>
<tr>
<td></td>
<td>OMI DA</td>
<td>2.46</td>
<td>5.21</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>-5.11</td>
<td>6.48</td>
</tr>
</tbody>
</table>

Appendix A: Seasonal cycles in the TROPOMI tropospheric NO$_2$ and super-observation errors

As shown in Fig. A1, the negative biases in TROPOMI tropospheric NO$_2$ column compared to OMI are larger in December 2018–February 2019 (by 25 %, 19 %, and 26 % over Europe, the United States, and China, respectively) than in April–May 2018 (by 10 %, 17 %, and 16 % over Europe, the United States, and China, respectively). In contrast, the differences in super-observation errors between TROPOMI and OMI are relatively constant over time. The differences in the relative super-observation errors (i.e., errors divided by concentrations) obtained from TROPOMI and OMI are smaller in winter than in other seasons over Europe and China because of the larger bias of the TROPOMI tropospheric NO$_2$ column compared to OMI in winter than in other seasons. The strong negative biases in TROPOMI retrievals in winter would increase the negative bias in NO$_2$ concentration analysis and reduce the estimated NO$_x$ emissions. Meanwhile, the differences in relative super-observation errors of TROPOMI retrievals between winter and other seasons suggest that TROPOMI DA might provide less constraints on spatial and temporal variations in NO$_2$ in winter than in other seasons and would still provide better constraints than OMI DA.

Consequently, the evaluation of individual satellite measurement through DA integration provides unique and detailed information on possible retrieval errors and its characteristics, including their spatiotemporal structures, which in turn provides a platform to evaluate different retrievals and supports satellite retrieval developments. Meanwhile, application of bias correction in DA analysis is essential for the combined use of observational information from multiple sensors, including those from other polar orbit satellites such as OMPS and advanced geostationary satellites such as GEMS, TEMPO, and Sentinel-4.
Appendix B: Characteristics of the TROPOMI version 1.2 and 2.3 products

In the latest version of the TROPOMI NO2 product, the low bias compared to OMI QA4ECV is largely improved from the previous versions (ESA, 2021; van Geffen et al., 2021). To discuss the potential impacts of the algorithm updates on the DA performance, Fig. B1 compares global distributions of tropospheric NO2 column, super-observation errors, and relative super-observation errors (i.e., errors divided by concentrations) obtained from the TROPOMI version 1.2 beta product, which was used in this study, and the S5P-PAL reprocessing product (processed with same processor as version 2.3.1), which was released more recently, for May 2018. The S5P-PAL reprocessing product data were obtained from the S5P-PAL data portal (https://data-portal.s5p-pal.com, last access: 1 February 2022). The algorithm updates from versions 1.2 to 2.3 led to increases in tropospheric NO2 column amounts typically by 6 % over polluted areas due to the algorithm updates from versions 1.2 to 2.3. These increases are mainly attributable to the improved FRESCO cloud retrievals (van Geffen et al., 2021). In contrast, the relative super-observation errors over most regions except for the southern midlatitudes are comparable between the products, with less than 0.2 % differences in the mean relative super-observation error over 60° N–60° S. These differences are much smaller than the differences between the TROPOMI version 1.2 beta and OMI QA4ECV products (by 19 % in May 2018).

The improved TROPOMI retrievals would reduce the negative bias in NO2 concentration analysis and increase the estimated NOx emissions for areas with weak chemical non-linearity. The increase in NOx emissions would reduce the negative biases in ozone analysis under NOx-limited ozone chemical regime. Meanwhile, the relative super-observation errors of TROPOMI retrievals were almost identical between versions 1.2 beta and 2.3.1. This suggests that the DA efficiency, for example, to constrain detailed temporal and spatial variations, might not be largely affected by the algorithm updates.

**Figure B1.** Global distribution of tropospheric NO$_2$ column (a, d, g), super-observation errors (b, e, h), and relative super-observation errors (c, f, i) obtained from the Tropospheric Monitoring Instrument (TROPOMI) version 1.2 beta product (a, b, c) and TROPOMI S5P-PAL (same as version 2.3.1) product (d, e, f), as well as the differences between TROPOMI version 1.2 beta and SSP-PAL products (g, h, i) during May 2018. The units of the tropospheric NO$_2$ column, super-observation errors, and relative super-observation errors are $\times 10^{15}$ molec. cm$^{-2}$, $\times 10^{15}$ molec. cm$^{-2}$, and percentage (%), respectively.

**Code availability.** The source code is not publicly available because of license restriction. The source code is available from Kengo Sudo (kengo@nagoya-u.jp) upon request. The source code for the data assimilation system is available from Kazuyuki Miyazaki (kazuyuki.miyazaki@jpl.nasa.gov) upon request.

**Data availability.** The data assimilation results are available online (https://figshare.com/projects/Sekiya_et_al_2021/126770, last access: 1 February 2022; Sekiya et al., 2021, https://doi.org/10.1029/2020MS002180). The TROPOMI and OMI satellite retrievals are publicly available at the TEMIS website (https://www.temis.nl/airpollution/no2col/no2regio_tropomi.php, last access: 1 February 2022; TEMIS, 2022) and the SSP-PAL data portal (https://data-portal.s5p-pal.com/browser/, last access: 1 February 2022, ESA, 2022). The AirBase data, AQS data, and Japanese continuous measurement data were provided by the EEA (https://discomap.eea.europa.eu/mapfme/AirQualityExport.htm, last access: 1 February 2022; EEA, 2022), the US EPA (https://aqs.epa.gov/aqsweb/airdata/download_files.html#Raw, last access: 1 February 2022; U.S. Environmental Protection Agency, 2022), and NIES (https://www.nies.go.jp/green/, last access: 1 February 2022; National Institute for Environmental Studies, 2022), respectively. The ATom-4 observation data were taken from the NASA Ames Earth Science Project Office (https://espoarchive.nasa.gov/archive/browse/atom/id14/DC8, last access: 1 February 2022; NASA Ames Earth Science Project Office, 2022). Ozonesonde observation data were obtained from https://woudc.org/data/explore.php?lang=en (last access: 1 February 2022; World Ozone and Ultraviolet Radiation Data Centre, 2022), SHADOZ (https://tropo.gsfc.nasa.gov/shadoz/Archive.html, last access: 1 February 2022; National Aeronautics and Space Administration, 2022), and NOAA ESRL GMD (https://gml.noaa.gov/afips/data/ozwv/Ozonesonde/, last access: 28 February 2022; National Oceanic and Atmospheric Administration, 2022). MCD12C1 MODIS/Terra+Aqua Land Cover Type data (Yearly L3 Global 0.05Deg CMG) were obtained from NASA Earthdata website (https://earthdata.nasa.gov/, last access: 1 February 2022; Friedl and Sulla-Menashe, 2015, https://doi.org/10.5067/MODIS/MCD12C1.006).
Author contributions. TS, KM, and HE designed this study. TS conducted data assimilation calculations and analyzed the results. HE provided the TROPOMI and OMI NO₂ data. All the coauthors commented upon and helped improve the manuscript.

Competing interests. The contact author has declared that neither they nor their co-authors have any competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. A part of the research was conducted at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. This work contains the modified Copernicus Sentinel data (2018), processed by ESA, for TROPOMI. Earth Simulator was used for simulation and data assimilation calculations with the support of JAMSTEC.

Financial support. This research has been supported by MEXT (grant no. JPMXP1020200305) as the Program for Promoting Researches on the Supercomputer Fugaku (Large Ensemble Atmospheric and Environmental Prediction for Disaster Prevention and Mitigation) and by JSPS KAKENHI (grant nos. 18H01285 and 18KK0102).

Review statement. This paper was edited by Helen Worden and reviewed by two anonymous referees.

References


