We have responded to each comment below. Our replies are in blue, and the revised manuscript text is written in bold.

Response to review 1

Qu et al. have studied the impact of top-down $NO_x$ emission estimates derived from two OMI $NO_2$ satellite data sets (NASA SP v3 and DOMINO v2) on $NO_2$ and O3 simulations with the GEOS-Chem model. Previous work already showed (e.g. Verstraeten et al. [2015], studies by Miyazaki et al.) that O3 in the troposphere is generally better understood when $NO_x$ emissions are derived from satellite $NO_2$ data than when taken from emission inventories.

Here, Qu et al. find substantial differences in the agreement of $NO_2$ and O3 simulations against independent measurements depending on whether data set NASA or data set DOMINO is used. This was to be expected given that it is well-known that the NASA and DOMINO datasets have considerable differences. A useful aspect of the study is that the authors now quantify the consequences of these differences, which is relevant because satellite data is increasingly used to improve model understanding of atmospheric composition.

What is disappointing however is that we do not learn much new. Simulations with the NASA emissions compare better to some metrics, and worse to others, but the authors do not explain why. This makes the manuscript a technical document, where it is left to the reader to figure out what emissions could work best for his/her particular purpose, without actual guidance on why that would be. The authors should do more to investigate why using one dataset leads to better agreement e.g. for surface O3 at remote sites, and the other for polluted sites. Aspects of spatial resolution, temporal representativeness, and vertical sensitivity should be taken into account when providing this guidance to potential users.

We appreciate the comments from the reviewer. We have modified the title, abstract, and the details in the manuscript accordingly to address these concerns. The title is now changed to "**Impacts of** global $NO_x$ inversions **on $NO_2$ and ozone simulations**."

"**Abstract.** Tropospheric **$NO_2$ and** ozone simulations have large uncertainties, but their biases, seasonality and trends can be improved with **$NO_2$ assimilations**. We perform global top-down estimates of monthly $NO_x$ emissions using two OMI $NO_2$ retrievals (NASAv3 and DOMINOv2) from 2005 to 2016 through a hybrid 4D-Var / mass balance inversion. **Discrepancy in $NO_2$ retrieval products is a major source of uncertainties in the top-down $NO_x$ emission estimates.** The 12-year averages of regional $NO_x$ budgets from the NASA posterior emissions are 37% to 53% smaller than the DOMINO posterior**. Consequently, the DOMINO posterior surface $NO_2$ simulations greatly reduced the negative biases in China (by 15%) and the US (by 22%)** compared to surface $NO_2$ measurements. **Posterior $NO_x$ emissions show consistent trend over China, US, India, and Mexico constrained by the two retrievals. Emission trends are less robust over South America, Australia, Western Europe and Africa, where the two retrievals show less consistency.** $NO_2$ trends have more consistent decreases (by 26%) with the measurements (by 32%) in the US from 2006 to 2016 when using the NASA posterior. **The performance of posterior ozone simulations has spatial heterogeneities from region to**

**region. On a global scale, ozone simulations using NASA-based emissions alleviates the double peak in the prior simulation of global ozone seasonality. The higher abundances of $NO_2$ from the DOMINO posterior increase the global background ozone concentrations and therefore reduce the negative biases more than the NASA posterior in the GEOS-Chem v12 simulations at remote sites. Compared to surface ozone measurements, posterior simulations have more consistent magnitude and interannual variations than the prior estimates, but the performance from the NASA-based and DOMINO-based emissions varies across ozone metrics. The current hard-constraints on $NO_x$ diurnal variations and limited availability of remote sensing data hinder improvement of ozone diurnal variations from the assimilation, and therefore have mixed performance on improving different ozone metrics. Additional improvements in posterior $NO_2$ and ozone simulations require more precise and consistent $NO_2$ retrieval products, more accurate diurnal variations of $NO_x$ and VOC emissions, and improved simulations of ozone chemistry and depositions."**

From a data user perspective, this work quantifies how differences in $NO_2$ retrieval products propagate to the downstream estimates in top-down $NO_x$ emissions and ozone simulations. The discrepancy found in this study is larger than uncertainties caused by data assimilation methods (4D-Var versus Kalman Filter) and chemical transport models [Koukouli et al., 2020], and is therefore a unique contribution of this work. Detailed investigation of the origin of differences in the NASA and KNMI $NO_2$ retrieval products goes beyond scope of this study. We do note however "**The GEOS-Chem $NO_2$ SCDs converted using scattering weight from the NASA product are larger than the SCDs calculated using the DOMINO scattering weight and the same GEOS-Chem VCDs (See Fig. S2). These can be explained by the use of different surface albedo and cloud product in the two retrievals.**" (Added in Section 3)

Another criticism is that the chain of technicalities is very long and that the experiments are set-up in a sub-optimal manner (for example comparing 2.5 degree simulations of surface $NO_2$ to surface stations that are representative for much smaller domains).

Comparing $NO_2$ simulations at 2.5° with in-situ measurements is sub-optimal, but this is the highest resolution we can perform global 4D-Var assimilation using this model.

A major concern I have is with the lack of detail and clarity on how the adjoint incorporates the information from the satellite retrievals. From the manuscript I first suspected that monthly mean column $NO_2$ data was simply used to estimate the emissions, suggesting that the highly variable and non-linear vertical sensitivities of the retrievals have not been used to interface the model with the satellite data. There are various studies pointing out how crucial it is to account for the vertical sensitivity of the $NO_2$ retrievals, e.g. Miyazaki et al. [2017], Boersma et al. [2016] to name a few. Then I read the supplementary material and there the impression was given that at least the a priori profile shapes are made consistent between the NASA and DOMINO retrievals, but it remains unclear to what extent this has harmonized the data, and to what extent vertical sensitivities between the two datasets are still fundamentally different.

To clarify, we added the following sentences to Section 2.2:

**“We converted GEOS-Chem $NO_2$ VCD to SCD using scattering weight (NASA product) and averaging kernel (DOMINO and QA4ECV product) from the OMI retrievals and then compare GEOS-Chem SCD with SCD retrieved from OMI. A cost function is defined as the observation error weighted differences between simulated and retrieved $NO_2$ SCD, plus the prior error weighted departure of the emission scaling factors from the prior estimates. We minimize the cost function using the quasi-Newton L-BFGS-B gradient-based optimization technique [Byrd et al., 1995; Zhu et al., 1994], in which the gradient of the cost function with respect to the control parameter is calculated using the adjoint method. Details of the** assimilation of $NO_2$ slant column densities (SCDs), **how vertical sensitivities of satellite retrievals are accounted for**, and the hybrid 4D-Var / mass balance inversion of $NO_x$ emissions **are described in** Qu et al. [2017].”

More detailed technicalities have been described in our previous publications cited in the manuscript and are therefore not the focus of this manuscript. The focus here is to apply this method for global $NO_x$ inversion, evaluate the impact of different retrieval products on top-down emission estimates, and how the changes in $NO_x$ emissions affect ozone simulations. Therefore, we did not repeat all the technical details that can be found in the cited publications. Please see our detailed responses below for all the concerns raised by the reviewer.

Specific comments

P2, L42-43: the formation depends not only on the local $NO_x$ and VOC concentrations, but also on the radiative regime in which these occur.

Changed to “Ozone formation and trends depend nonlinearly on the local relative abundances of $NO_x$ and VOCs **and the radiative regime in which these occur**.”

P2, L65: different → differ

Modified as suggested.

P3, L72: import →importer

Modified as suggested.

P3, L78-81: Zhang et al. [2008] and Verstraeten et al. [2015] already showed that through optimizing $NO_x$ emissions in China, the simulated O3 over the Pacific and over the western US indeed improved.

We changed this sentence to:

“Optimization of $NO_x$ emissions in the upwind regions can improve remote ozone simulations in downwind regions after transport of intercontinental pollution plumes from the free troposphere to the surface **[Zhang et al., 2008; Verstraeten et al., 2015]**.”

Section 2.1 It is unclear in this manuscript how the adjoint accounts for (a) vertical sensitivity of the satellite retrievals, and (b) the diurnal cycle of $NO_x$ emissions. These aspects are important enough to describe in the manuscript, for (a) useful information is provided in the supplement but it is not clear whether the replacing of the a priori profiles by GEOS-Chem prior profiles was also applied in the research to estimate the emissions. The authors should clarify this in section 2, and also briefly quantify to what extent the differences in prior simulations have been minimized by this approach.

Many of these aspects have been described in details in a previous publication cited in Section 2.2 (Qu et al. 2017), so we do not repeat the same information in this manuscript. To clarify, we added a brief summary of our inversion in Section 2.2, see our response above.

For the reviewer's information, The comparison of SCDs ($VCD_{GC}AMF_{GC} - SCD_{OMI}$) is theoretically equivalent to comparisons of VCDs ($VCD_{GC} - \frac{SCD_{OMI}}{AMF_{GC}}$). These have been described in Qu et al. [2017], pasted below:

"In all of our simulations, we calculate the air mass factor (AMF) for GEOS-Chem simulated $NO_2$ columns ($AMF_{GC}$) following Equations 1 to Equation 4 in Bucsela et al. [2013]. Here, $AMF_{GC}$ is expressed as the ratio of the sum of slant sub-columns in the troposphere (*S*) to the sum of vertical sub-columns in the troposphere (*V*):

$$AMF_{GC}(i,j) = \frac{S}{V}$$

where

$$S = \sum_{l \text{ in the troposphere}} MR(i,j,l)(P(i,j,l) - P(i,j,l+1))SCW_{OMI}(i,j,l)$$

$$V = \sum_{l \text{ in the troposphere}} MR(i,j,l)(P(i,j,l) - P(i,j,l+1))$$

Here, *MR* is the mixing ratio of $NO_2$, *P* is the pressure at the center of the GEOS-Chem grid, $SCW_{OMI}$ is the scattering weight linearly interpolated from the OMI product to GEOS-Chem grid using the scattering weight pressure from the Level 2 product and pressure at the center of each model grid cell, with application of temperature correction following Equation 4 of Buscela et al. [2013]. $AMF_{GC}$ is then used for conversion of GEOS-Chem $NO_2$ vertical column densities to SCDs, which are directly comparable to SCDs retrieved from OMI,

$$SCD_{GC}(i,j) = AMF_{GC}(i,j) \sum_{l \text{ in the troposphere}} (c(i,j,l) \times h(i,j,l))$$

where $c$ is simulated $NO_2$ concentration [molecules $cm^{-3}$] and $h$ is the height of the box."

We added the following sentence to the first paragraph of Section 3:

**"The cost function has reduced by 6% - 29% in the monthly inversion."**

For (b), some info is given but only late in the game (P7: The diurnal variations of $NO_x$ emission are constrained to be those of the prior emissions), and we do not learn what the diurnal cycle is in the first place. Please revise section 2 thoroughly with this in mind.

We added the following sentence to Section 2.1:

**"The diurnal variation of $NO_x$ emissions is derived from EDGAR hourly variations ( http://wiki.seas.harvard.edu/geos-chem/index.php/Scale_factors_for_anthropogenic_emissions#Diurnal_Variation) and is not optimized in the inversion."**

Then I have other questions:

- how does the adjoint approach account for other relevant aspects of data assimilation?

Details of our 4D-Var inversion are in Qu et al. [2017]. In brief, we define a cost function as described in Section 3 of Qu et al. [2017]. Then, "We minimize the cost function using the quasi-Newton L-BFGS-B gradient-based optimization technique [Byrd et al., 1995; Zhuetal., 1994], in which the gradient of the cost function J($\boldsymbol{\sigma}$) with respect to the control parameter $\boldsymbol{\sigma}$ is calculated using the adjoint method. The adjoint model is driven by a forcing term, which is the error weighted difference between predicted and simulated $NO_2$ slant columns. Inversions are considered to have converged when the cost function decreases by less than 1% in three consecutive iterations."

- how is the OMI data averaged spatially to the grid of GEOS-Chem, and how are superobservation errors incorporated?

We did not average OMI data or use super-observations. Instead, we assimilate each OMI retrieval separately and compare it with GEOS-Chem simulations at the corresponding hour, with corresponding averaging kernel applied. Please see Section 3 in Qu et al. [2017] for more details, which state:

"Slant column densities from OMI at each observation time and site are used to constrain monthly anthropogenic $NO_x$ emissions. The observation error covariance matrix, $\mathbf{S_{obs}}$, is assumed to be diagonal. Absolute uncertainties of these diagonal values are read from NASA $OMNO_2$ L2 products for each individual OMI observation. On average, the tropospheric slant column uncertainty of OMI is estimated to be $\sim 0.7 \times 10^{15}$ molecules $cm^{-2}$ [Boersma et al., 2008; Castellanos and Boersma, 2012]. To reduce the influence of observations below the OMI detection limit, which mainly occur in remote locations, we conservatively assume an absolute uncertainty of $1.0 \times 10^{15}$molecules $cm^{-2}$, and we add this value to $\mathbf{S}_{obs}$."

- did the authors only use the mostly cloud-free OMI retrievals?

Yes, only retrievals with cloud fraction less than 0.2 are used. This has been stated in section 2.2 of this manuscript:

"We screen all OMI $NO_2$ retrievals using data quality flags and by the criteria of positive tropospheric column, cloud fraction < 0.2, solar zenith angle < 75°, and viewing zenith angle < 65°."

Section 2.2: OMI is suffering from the so-called row anomaly, which was absent until mid-2007, and then became gradually more important. How did the authors ensure that the growing impact of the row anomaly did not unduly affect their trends in $NO_x$ emissions?

The OMI data affected by row anomaly are filtered out using the quality flag. We added the following sentences to section 2.2:

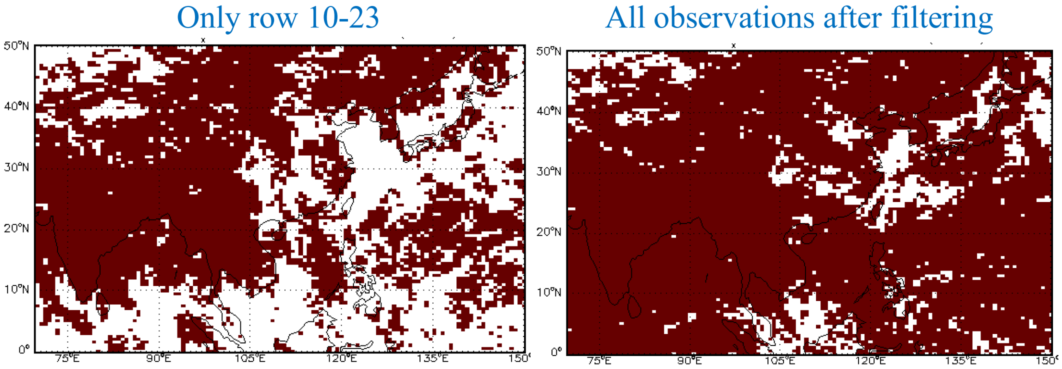
**"We excluded all retrievals that are affected by row anomaly."**

We have tested the differences between annual mean OMI $NO_2$ column densities without data filling after excluding pixels affected by row anomaly and when filling missing data by linearly interpolating column densities from adjacent years in Qu et al. [2017]; we found the filling to impact annual mean SCDs by less than 10% for all regions shown in Figure 8 of Qu et al. [2017]. Differences in these two SCDs for all studied years are less than 1% in mainland China.

Another approach to mitigate inconsistent sampling of the data is to follow Duncan et al. [2013] and consider the trend in $NO_2$ columns from only rows 10 to 23 of the NASA standard product, which are unaffected by the row anomaly throughout the period. These are shown in the grey lines in Figure 8 of Qu et al. [2017]. Please also note that even though we are using the same rows each year, this doesn't necessarily mean that the number of observations is the same after screening according to our other filtering criteria, nor does it mean the same geographical locations are observed throughout the period. The correlation of this dataset with OMI data from the standard NASA product in all rows is >0.75 in most regions.

Though we recognize the benefits of using a consistent number of observations to analyze the trend of $NO_2$ columns alone, this is not necessarily the case for a Bayesian inversion of $NO_x$ emissions. The inversion is forced by the residual model error summed over all available observations; fewer observations in some years or locations will thus naturally result in greater dependence on the prior emissions. If we exclude observations to maintain consistency in the rows used, emissions in many grid cells do not get updated due to lack of observations (see Fig. R1). This would lead to spatial trends in posterior emissions that could have been avoided if using all available observations (after data screening).

We think the two approaches to invert $NO_x$ emissions, maintaining consistency in rows used or not, both have their pros and cons. Since the goal of this work is to derive top-down emissions, which would benefit from broader observation coverage (in the example of January 2006 below, we would not be able to get posterior emissions for regions covered in white if eliminating those rows affected by row anomaly throughout) and the trend of $NO_2$ columns between these two does not differ much, we chose to use all observations available after data selection.


Only row 10-23

All observations after filtering

Figure R1. Data coverage in January, 2006, using only rows 10 to 23 (left) and all rows (right), where, red color stands for grid cells that have at least one observation during the month.

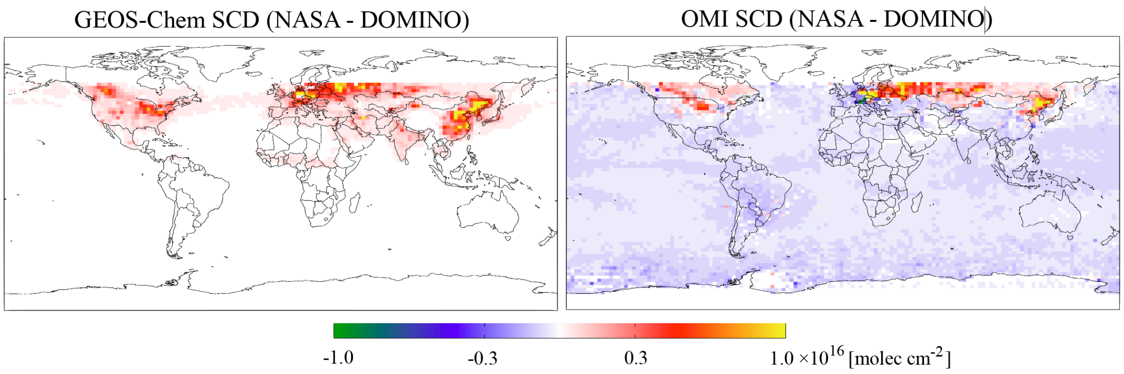
Section 2.3: it remains unclear what type of surface station was used for the GEOSChem surface evaluation. Using urban background and regional stations seems appropriate to evaluate the large GC grid cells, but urban street stations should be excluded.

We checked the monitoring site lists and a document defining the site category (http://www.bjmemc.com.cn/xgzs_getOneInfo.action?infoID=1661). None of the sites included in this study was listed as roadway sites. We added the following sentence to Section 2.3:

**"The city monitoring sites included in the analysis represent either urban background or the averaged pollutant concentrations over the city."**

P5, L152-154: what explains the OMI-driven differences between the posterior $NO_x$ emissions, differences in tropospheric slant columns or in the AMFs? Presumably the latter, but since the a priori profile differences have been "minimized", the differences must be in the assumptions on surface albedo and clouds. It would be best if the authors could shed more light on how the scattering weights or averaging kernels are different between the OMI $NO_2$ retrievals. Please clarify.

We added a new Figure S2 to the supporting information:



“

**Figure S2. Differences in tropospheric NO$_2$ SCDs between the NASA and the DOMINO products in January 2010. The differences in GEOS-Chem SCDs (left figure) are calculated by converting the same GEOS-Chem VCD using scattering weight and averaging kernel from the two products. In the right figure, AMFs provided by the two products are applied to their corresponding VCDs to calculate the differences in SCDs. ”**

We also added the following sentences to the cited paragraph:

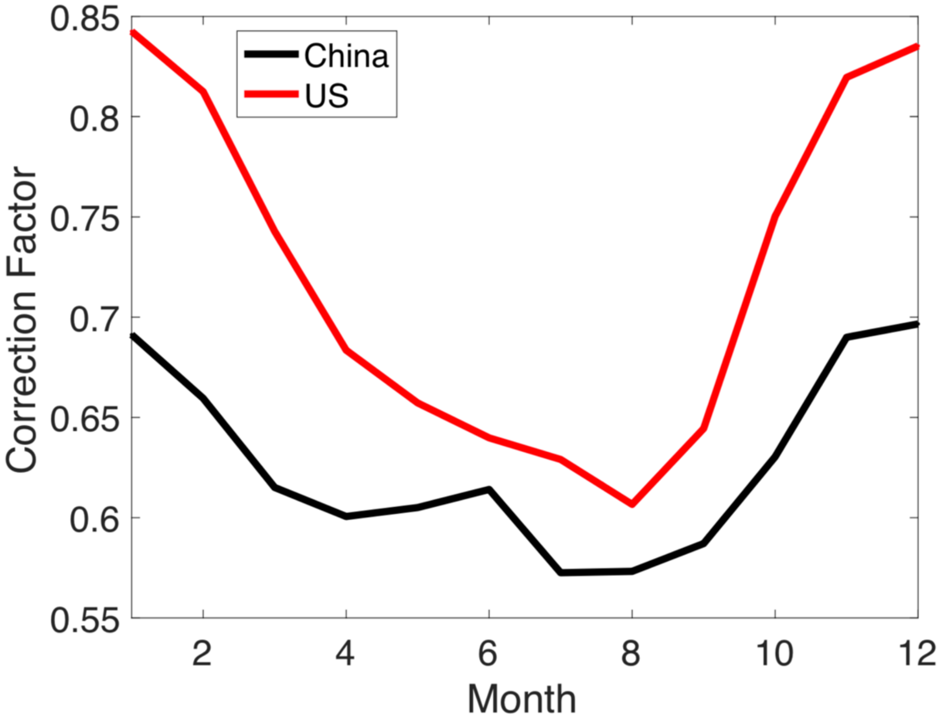
**“The GEOS-Chem NO$_2$ SCDs converted using scattering weight from the NASA product are larger than the SCDs calculated using the DOMINO scattering weight and the same GEOS-Chem VCDs (See Fig. S2). These can be explained by the use of different surface albedo and cloud product in the two retrievals. The retrieved NO$_2$ SCDs from the NASA product are mostly smaller than the DOMINO retrieval except for some regions between 40°N – 60°N in January 2010. The smaller magnitude in OMI SCD and the larger magnitude in GEOS-Chem SCD using the NASA scattering weight lead to smaller magnitude of posterior NO$_x$ emissions than inversions from the DOMINO product.”**

P6, L173-174: the statement that “NO$_2$ column simulations at 2° × 2.5° in this study are likely to be underestimated and lead to high biases of posterior NO$_x$ emissions to match satellite NO$_2$column concentrations” needs more evidence. The hypothesis that instant dilution leads to too much OH (by Valin et al. [2011]) may be valid for isolated NO$_x$ sources in otherwise pristine areas, but instant dilution of NO$_x$ emissions situated in high-background NO$_2$ regions such as the eastern US or western Europe is probably of less concern.

We removed the cited sentences.

P6, L193: what is the magnitude of the correction factors over China and the US? How do they vary by season?

We added the following figure in the SI:



**“Figure S3. Seasonal variation of the $NO_2$ correction factors in China (black) and the US (red) calculated following Lamsal et al. [2008]. “**

We added the following sentences to the cited paragraph:

**“The correction factors are generally higher in the US than in China, but have similar seasonality (see Fig. S3).”**

P7, L195-199: this part is rather inconclusive. The GEOS-Chem simulations have been corrected for resolution (an increase) and surface measurements have been corrected down for molybdenum interference, and still GEOS-Chem with posterior emissions is biased low by 20%-50%. What explains the persistent low bias?

We added the following sentences to this paragraph:

**“These remaining negative biases reflect the unrepresentativeness of 0.1° pseudo measurements for real point measurements for resolution bias correction, comparison of $NO_2$ concentrations averaged over 2°×2.5° simulation to limited measurements, the underestimates of $NO_2$ retrievals using coarse resolution a priori, and the inability of data assimilation to increase emissions at grid cell where $NO_2$ retrievals are below the detection limit of OMI.”**

Also, we do not expect the posterior simulations to be completely unbiased given the potential biases from model and satellite retrieval.

P7, L224-225: OMI measurements frequently miss the high values of $NO_2$ column densities that occur before or after its overpassing time. OMI was never designed to measure $NO_2$ before or after its overpass time, so to say that OMI misses these high values is misleading. Please rephrase.

We changed the sentence to:

**"The daily $NO_2$ column densities from OMI are also underestimated compared to the diurnally varying ground-based retrievals [Herman et al., 2019]."**

P7, L226: twice-per-day constraints on $NO_x$ emissions have been achieved in earlier studies based on SCIAMACHY + OMI (Boersma et al. [2008], GOME-2 + OMI [Lin et al., 2011], including via sophisticated assimilation schemes [Miyazaki et al., 2017].

We changed the sentence to "**Assimilating $NO_2$ observations from instruments overpassing at different times of the day [e.g., Boersma et al., 2008; Lin et al., 2010; Miyazaki et al., 2017] and using** hourly constraints from the geostationary satellite data (e.g., Geo-stationary Environmental Monitoring Spectrometer (GEMS), Tropospheric Emissions: Monitoring of Pollution (TEMPO) [Zoogman et al., 2017] and Sentinel-4) have the potential to improve simulations of ozone diurnal variations and different ozone metrics, **although the ratio of $NO_2$ column densities from satellites that overpass in the morning and afternoon are generally lower than the same ratio from surface measurements [Penn and Holloway, 2020]."**

P8, L237: the June peak in $NO_2$ over China can be easily traced back to crop residu burning in that month – e.g. Stavrakou et al. [2016].

We added the following sentence:

**"The June peak in China has been explained by the crop residual burning [Stavrakou et al., 2016]."**

P8, L238-240: can you explain more why the DOMINO product would be more sensitive to soil $NO_x$ emissions? It's not because of the different a priori profiles assumed in the NASA and DOMINO retrievals?

As the reasons are not entirely clear, we changed the cited sentence to:

"The peak of the DOMINO posterior $NO_x$ emissions in the United States and Mexico shifted earlier in the year to June and July compared to the prior and NASA posterior emissions, similar to the results from Miyazaki et al. [2017]. **The peak in DOMINO posterior emissions corresponds to the time of high soil $NO_x$ emissions, which are reported to be underestimated in high-temperature agricultural systems in the bottom-up inventory [Oikawa et al., 2015; Miyazaki et al., 2017]."**

P8, L243-244: please see my previous comment. The authors seem to know something very interesting here, but they don't show it. Is there any evidence that one retrieval would be more sensitive to $NO_x$ sources than the other? That would be extremely relevant to know more about. Since the satellite measurements are identical for the NASA and OMI retrievals, it must have to do with AMF differences , see e.g. Lorente et al. [2017]. But what drives the apparent difference in sensitivity – albedo, cloud fraction, cloud pressure?

The two retrievals have the same spectrum but the retrieved tropospheric SCDs are not exactly the same (for instance, the two products use different stratosphere-troposphere separation), see our previous response and the new figure S2. All of the factors the reviewer mentioned here are different between the two products. It is hard for us to pinpoint which of the albedo, cloud fraction, or cloud pressure drives the sensitivity without running the radiative transfer model and performing the retrieval ourselves, which is beyond the scope of this study.

We changed the cited sentence to:

“**These retrieval products have similar number of observations and spatial distributions of observation densities after the filtering.** The different seasonal variations in the posterior $NO_x$ emissions may reflect the **AMF structural uncertainties when the retrieved $NO_2$ column densities use different ancillary data [Lorente et al., 2017]**. **For instance, the GEOS-Chem $NO_2$ SCDs converted using the scattering weight from the NASA product have larger seasonal variations than the SCDs converted using the DOMINO averaging kernel in the US, reflecting the different seasonal variations of vertical sensitivities from the two retrievals.**”

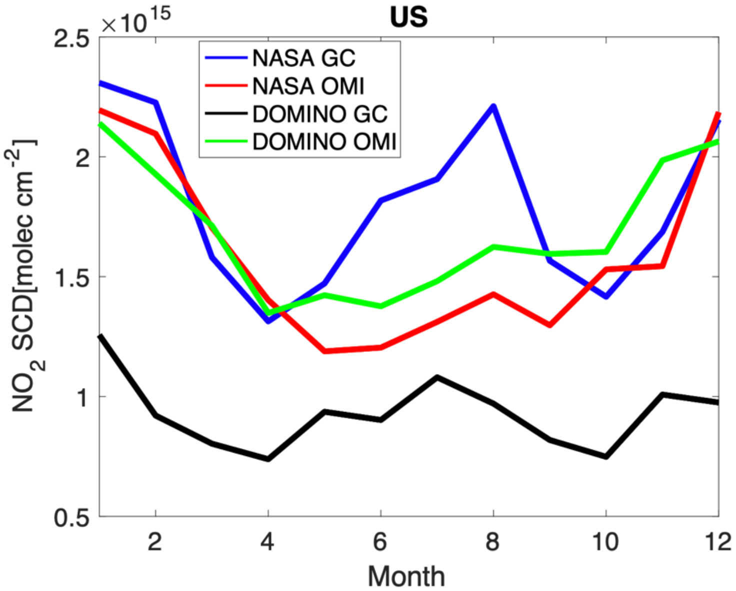


Figure R2. Seasonal variations of OMI $NO_2$ SCDs from NASA (red) and DOMINO (green) retrievals, and the GEOS-Chem simulated $NO_2$ SCDs using scattering weight from the NASA (blue) and the DOMINO (black) products.

P8, L246-256: Figure 5 – the daytime O3 simulations in China all seem strongly low biased relative to the observations. The other ozone metrics in China and all in the US match much better. Why is this?

Thanks for pointing this out. There was a bug in processing daytime ozone in China, which is fixed now. Please see the revised Figure 5 below.

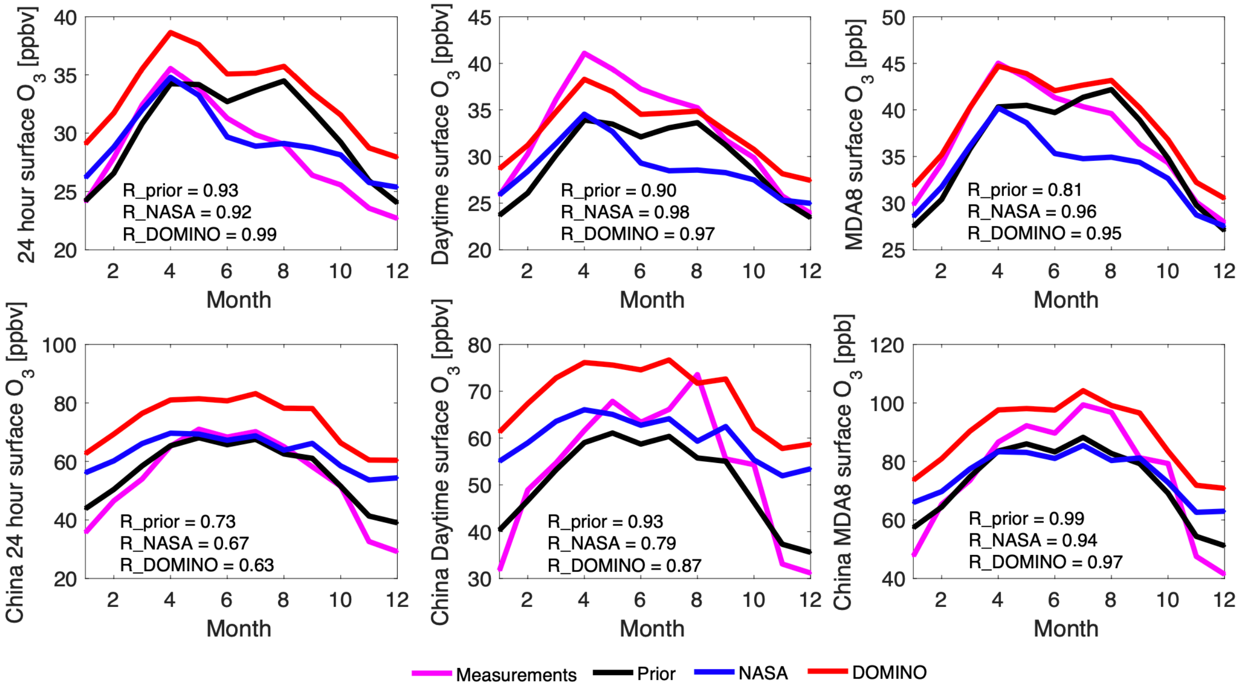


Figure 5. Seasonality of surface ozone concentration at 2 meters in 2010 compared with TOAR (top) and in 2015 compared with CNEMC (bottom). Surface measurements are shown in magenta lines. Simulations are performed using GCv12 with $NO_x$ emissions from CEDS (black line), NASA posterior (blue line) and DOMINO posterior (red line).

P9, L271-272: “also not reflected”?

Changed to “not reflected”

P9, L276: no reduction of $NO_x$ emissions in Europe? This is strange – $NO_2$ tropopsheric columns are decreasing over Europe, and Miyazaki et al. [2017] showed reductions in for $NO_x$ emissions. Overall, Figure 6 looks very odd to me. DOMINO $NO_2$ columns are 40% higher than NASA, but the $NO_x$ emissions inferred from DOMINO are more than 40% higher than the emissions inferred with NASA (L278-281). Also, Miyazaki et al. [2017] (Figure 9) still find reductions in $NO_x$ emissions over Europe between 2005 and 2014 based on the same DOMINO data, so how can you find increases? Please clarify.

We do not expect the relative differences in the direct comparison of $NO_2$ column densities from the two OMI products to have similar magnitude with the differences in their posterior emissions. As shown in the newly added Figure S2, the adjustment in $NO_x$ emissions are determined not only by the differences in $NO_2$ SCDs from OMI retrievals but also by the GEOS-Chem SCDs after applying scattering weight / averaging kernel (equivalent to converting OMI SCD to VCD using a new AMF based on GEOS-Chem profile and compare with GEOS-Chem simulated VCD). The smaller magnitude in OMI SCD and the larger magnitude in GEOS-Chem SCD using the NASA scattering weight leads to even smaller magnitude of posterior $NO_x$ emissions than the posterior constrained by the DOMINO product.

As for the posterior emissions in Europe, the result from Miyazaki et al. (screenshot shown in the left panel of Figure R3) shows large fluctuations around 0 throughout 2005 and 2014, and it is hard to say there is a decreasing trend from their Figure 9. The relative change from 2005 to 2014 in this study, shown in the right panel of Figure R2, is also negative (-1.3%), consistent with results in Miyazaki et al. [2017]. The slight upward fluctuation of posterior $NO_x$ emissions in this study happened after 2014, which is not included in the time range of Miyazaki et al. [2017].

We changed the cited sentence to:

**"In Western Europe and Africa, posterior $NO_x$ emissions fluctuate and do not have a significant consistent trend from the two inversions."**

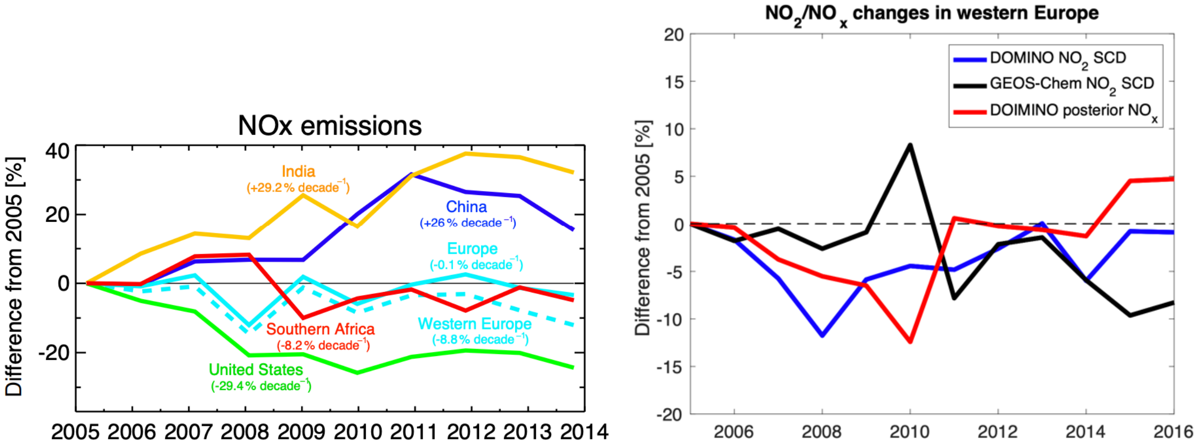


Figure R3. Relative changes in $NO_x$ from Miyazaki et al. [2017] (left) and this study (right).

P10, L295-297: I'm missing an explanation or hypothesis why $NO_x$ emissions from one dataset would do better than the other for different ozone metrics.

We added the following sentences to the cited paragraph:

**"The different performance of $NO_x$ emission datasets for different ozone metrics is a consequence of the hard constraint on $NO_2$ diurnal variations within the assimilation (and the lack of sufficient observations to constrain this). This can lead to better agreement of mean ozone concentration with measurements over particular hours but worse mean concentrations averaged over other hours."**

P10, L304 and L315: please clarify how the impact of meteorology and non-$NO_x$ sources on O3 changes was evaluated.

We changed the original sentence on L304 to "The trends of simulated annual MDA8 ozone concentrations are correlated with impacts from meteorology and non-$NO_x$ sources **based on simulations (shown as green lines) that use the same anthropogenic $NO_x$ emissions for all years and simulations that use interannually varied anthropogenic $NO_x$ emissions,** leading to …"

We added the following sentences to the original sentence on L315:

"…as well as meteorology and non-$NO_x$ sources. **The second trend is calculated through simulations that use constant $NO_x$ emissions throughout the studied years. It has similar trend from GCv12 and GCadj as shown in the green lines in Fig. 9. The trend caused by $NO_x$ emissions is obtained by subtracting the second trend from the ozone trend simulated using $NO_x$ emissions at each corresponding year.** The ozone trends…"

L306-307: "The trends of simulated MDA8 ozone are similar when using the NASA and the DOMINO posterior $NO_x$ emissions as inputs" – yes, but please also explain why the magnitude of the NASA-derived MDA8 O3 levels are biased high then.

The blue and red colors in this figure now represent ozone simulations from different models. The differences from $NO_2$ retrievals are now represented in the error bars. The NASA-derived MDA8 ozone are actually lower than the DOMINO-derived one. We added the following sentences to this paragraph:

**"The DOMINO-derived MDA8 ozone concentrations are higher than the NASA-derived ones in all studied regions, represented by the upper and lower limit of the error bars respectively. GCv12 simulated ozone concentrations are smaller than simulations from GC-adj, especially over relatively less polluted regions, consistent with the inclusion of halogen chemistry in GCv12, which depleted ozone."**

P11, L332-333: the prior simulated O3 profiles in Figure 10 agree much better with the O3 sondes between 800-400 hPa than the assimilated profiles. I don't understand why that is, since the effect of the updated $NO_x$ emissions should be mostly felt in the lower 2 kms of the atmosphere. Or is this the impact of changes in background O3 in response to changing Asian emissions?

The reviewer must have been mistaken when considering this figure, as it is not true that all nor even most prior simulations (black dotted and black solid lines) agree better with ozone sondes (magenta solid) in Figure 10. In the 800-400 hPa range, the figure shows the GC-adj simulation using the DOMINO posterior $NO_x$ emissions (dashed red) is almost always the closest to the sonde data. More detailed statistics of ozone profiles between 700-900 hPa, where ozone is mainly impacted by Asian emissions (Figure S8), show that the posterior O3 from GC-adj have smaller NMB and NMSE than the prior at 4 of 6 sites.

We added the following sentence to the title of Figure 10:

**“The six sites are over remote regions and are used to evaluate the intercontinental transport of ozone.”**

P13, L394-395: one important difference between this research and the work done by Miyazaki in a number of papers, is that the latter assimilates also other species relevant for $NO_x$ inversions and O3 simulations (e.g. CO, HNO3, SO2). It would be interesting to also discuss to what extent these additional constraints can help explain the “remaining differences between simulated and measured ozone”.

We added the following sentence:

**“Assimilation of multiple species (e.g, ozone, CO, $HNO_3$ and $SO_2$) together with $NO_2$ may improve posterior ozone simulations, but the performance of posterior simulations may depend on the chemical transport model, as shown in Miyazaki et al. [2020], where the GEOS-Chem adjoint model v35 shows mixed performance in correcting the bias between ozonesonde and posterior simulations between 850-500 hPa at different latitude band.”**

P13, L398-400: the statement “Both OMI $NO_2$ retrievals employed in this study use $NO_2$ vertical shape factors from coarse resolution simulations, and therefore are biased low compared to in-situ measurements [Goldberg et al., 2017].” Brought up the question (again) whether both OMI $NO_2$ retrievals are at least consistent now in their use of the same coarse-resolution vertical shape factors (i.e. those from GEOS-Chem).

Yes, we converted GEOS-Chem VCD to SCD using scattering weight from these two products for comparison, but mathematically they are equivalent to replacing the shape factor with the same GEOS-Chem one. Please see more detail in our response to previous comment.

P13, L401: “retrievals also have not explicitly accounted for the aerosol optical effects, which are demonstrated to degrade the accuracy of $NO_2$ column concentrations”. This is an overstatement. Only when AOD is very high (>0.5-1.0) there are indications that implicit corrections break down. Even in Liu et al. [2019] accounting for AOD did not solve the low bias in tropospheric $NO_2$ which was not apparent in the DOMINO scheme without an explicit aerosol correction.

We change this sentence to “which are demonstrated to degrade the accuracy of $NO_2$ column concentrations **when AOD is very high**”.

References

Boersma, K. F., Vinken, G. C. M., and Eskes, H. J.: Representativeness errors in comparing chemistry transport and chemistry climate models with satellite UV–Vis tropospheric column retrievals, Geosci. Model Dev., 9, 875-898, doi:10.5194/gmd-9-875- 2016, 2016.

Miyazaki, K., Eskes, H., Sudo, K., Boersma, K. F., Bowman, K., and Kanaya, Y.: Decadal changes in global surface $NO_x$ emissions from multi-constituent satellite data assimilation, Atmos. Chem. Phys., 17, 807–837, https://doi.org/10.5194/acp-17-807- 2017, 2017.

Stavrakou, T., Müller, J.-F., Bauwens, M., De Smedt, I., Lerot, C., Van Roozendael, M., Coheur, P.-F., Clerbaux, C., Boersma, K. F., van der A., R. J., and Song, Y.: Substantial underestimation of post-harvest burning emissions in the North China Plain revealed by multi-species space observations, Sci. Rep., 6(32307), doi:10.1038/srep32307, 2016.

Response to review 2

This manuscript has presented top-down estimates of global $NO_x$ emissions using two OMI satellite $NO_2$ products over 2005-2016 and using the GEOS-Chem adjoint inversion method. Considerable differences are found between the two top-down emission estimates. Implementing the top-down $NO_x$ emissions to the GEOS-Chem atmospheric chemistry model shows some improvements on the model simulation of tropospheric ozone. The study also points out that model improvements largely depend on the top-down emissions, the ozone metrics used, and model versions. The manuscript is in general well organized and meets the scope of ACP. One main concern is that the manuscript has been presented as a model evaluation paper that comparing several model simulations with different $NO_x$ emissions with surface and sonde ozone measurements. It lacks some analyses in depth to understand the driving factors of the differences. The key findings of this study are also not clear. Do we have a better understanding of the $NO_x$ emission trends as constrained by the satellite measurements, or how $NO_x$ emission changes affect tropospheric ozone? I think the concern and the following specific comments should be addressed before considering publish.

We appreciate the comments from the reviewer. We added the following sentences in the abstract to address the concern on the $NO_x$ emission trends:

**“Posterior $NO_x$ emissions show consistent trend over China, US, India, and Mexico constrained by the two retrievals. Emission trends are less robust over South America, Australia, Western Europe and Africa, where the two retrievals show less consistency.”**

Limited by the availability of surface measurements, we cannot claim that $NO_x$ emission trends are improved everywhere. However, we demonstrate in this study that there are several regions where top-down $NO_x$ emission trends are consistent across different retrievals and we are more confident about these.

The impact of $NO_x$ emission on ozone simulations have spatial heterogeneity due to the nonlinear response of ozone to $NO_x$ and our different understanding of local sources, physics, and chemistry. So, there is no generalized conclusion at global scale. We added the following sentences to the abstract to summarize our findings from this work:

**“The performance of posterior ozone simulations is spatially heterogeneous from region to region. On a global scale, ozone simulations using NASA-based emissions remove the double peak in the prior simulation of global ozone. The higher abundances of $NO_2$ from the DOMINO posterior increase the global background ozone concentrations and therefore reduce the negative biases more than the NASA posterior in the GEOS-Chem v12 simulations at remote sites. Compared to surface ozone measurements, posterior simulations have more consistent magnitude and interannual variations than the prior, but the performance from the NASA-based and DOMINO-based emissions varies across ozone metrics. The current hard-constraints on $NO_x$ diurnal variations and limited availability of remote sensing data hinder improvement of ozone diurnal variations from the assimilation,**

**and therefore have mixed performance on improving different ozone metrics. Additional improvements in posterior $NO_2$ and ozone simulations require more precise and consistent $NO_2$ retrieval products, more accurate diurnal variations of $NO_x$ and VOC emissions, and improved simulations of ozone chemistry and depositions."**

Specific comments:

1) Page 1, Line 24-25 in the Abstract: The statement "using $NO_x$ emission datasets that have the best performance . . ." is not clear. As ozone simulation is affected by many other factors, the $NO_x$ emissions that have the best performance on ozone simulation may not be the correct one. Some results in this study also showed that satellite constrained $NO_x$ emissions did not necessarily improve ozone simulation (e.g., China daytime surface ozone in Figure 5)

We have revised the abstract, see response above.

2) Page 3, Section 2.1: What was the spin-up time for the model simulations? Were you using the same initial conditions? Please clarify.

The initial conditions are different for each $NO_x$ emission datasets. We added the following sentence to the last paragraph of Section 2.1:

**"For each $NO_x$ emission dataset, the model is spun-up for 6 months, starting from July 2005. Therefore, we derive $NO_x$ emissions from 2005, but only evaluate simulations with measurements from 2006."**

3) Page 6, Line 179: Should here "the average of GEOS-Chem simulated $NO_2$ column density" be OMI observed $NO_2$ column density over 2x2.5 grid cell? Here you are generating pseudo measurements in the statement. The ratio should be calculated by OMI observations to avoid the OMI vs. model biases.

Thanks for pointing this out. We are calculating in the way the reviewer suggested, but did not describe it correctly. We changed the sentence to:

"…by the ratio of OMI $NO_2$ column density gridded at 0.1° × 0.1° to the **OMI $NO_2$ column density gridded at** 2° × 2.5° grid cell"

4) Page 8, Line 240-245: The large differences in seasonal variations of DOMINO and NASA posterior $NO_x$ emissions seem interesting. Here you explained that the DOMINO posterior may better constrain soil emissions. Do you have any evidence or support for that?

We changed the cited sentence to:

"The peak of the DOMINO posterior $NO_x$ emissions in the United States and Mexico shifted earlier in the year to June and July compared to the prior and NASA posterior emissions, similar to the results from Miyazaki et al. [2017]. **The peak in DOMINO posterior emissions corresponds to the time of high soil $NO_x$ emissions, which are reported to be underestimated in high-temperature agricultural systems in the bottom-up inventory [Oikawa et al., 2015; Miyazaki et al., 2017].**

5) Page 8, Line 250-256: Here you showed that prior simulated surface ozone concentrations had double maxima in April and August, and the posterior results partly corrected the biases. What cause the double maxima in the prior simulation? And how $NO_x$ emission changes correct the August maximum? Please clarify.

We added the following sentences to the cited paragraph:

**"The August ozone peak in the prior simulation is mainly due to the high ozone concentrations in North China, Southwest China, and North India. The NASA and DOMINO posterior simulations have both reduced surface ozone concentrations in North China Plain and Northeast China in August due to the larger posterior $NO_x$ emissions than the prior in these high-$NO_x$ regions. Both posterior ozone simulations are also smaller than the prior in Tibet and North India due to the reductions of posterior $NO_x$ emissions in low-$NO_x$ region. The August ozone peak in the DOMINO posterior comes from the higher ozone concentrations in Angola and Democratic Republic of the Congo compared to the NASA posterior and prior simulations in the same month and DOMINO posterior simulations in the previous months. This can be explained by the larger upward adjustment of DOMINO posterior $NO_x$ emissions in South Africa in August. These results show the large spatial heterogeneities on the responses of ozone seasonality to the changes in $NO_2$ abundances on a global scale."**

6) Page 9, Line 269-271: As indicated in Figure 6, interannual changes in the two posterior $NO_x$ emissions in Australia over 2005-2016 are not that consistent. The DOMINO results show large reduction over 2006-2010 and then increase afterwards. Do you have any explanation why the two satellite products show different interannual variation and trends over some regions?

The different trends in posterior $NO_x$ emissions are propagated from the different trend in $NO_2$ column densities retrieved from these two products, as shown in Figure R4. This could possibly be caused by the differences in scattering weight / averaging kernel, but it is hard for us to pinpoint what is the exact cause. We made the following modification to the cited sentence:

"In **Mexico**, the two posterior $NO_x$ emissions consistently increased by 6% (NASA) and 13% (DOMINO) from 2005 to 2016. **The DOMINO posterior shows more obvious increase** in Mexico from 2010 to 2016. ... **In Australia, the NASA posterior increases by 10% from 2005 to 2016. In comparison, the DOMINO posterior decreases from 2005 to 2010 and increases afterwards, consistent with the posterior trend from Miyazaki et al. [2017]. The different trends in posterior $NO_x$ emissions are propagated from the trends in the two OMI $NO_2$**

**retrieval products. The discrepancies are likely due to the different surface albedo and cloud products used in the two retrievals, which affect averaging kernel sensitivities**."

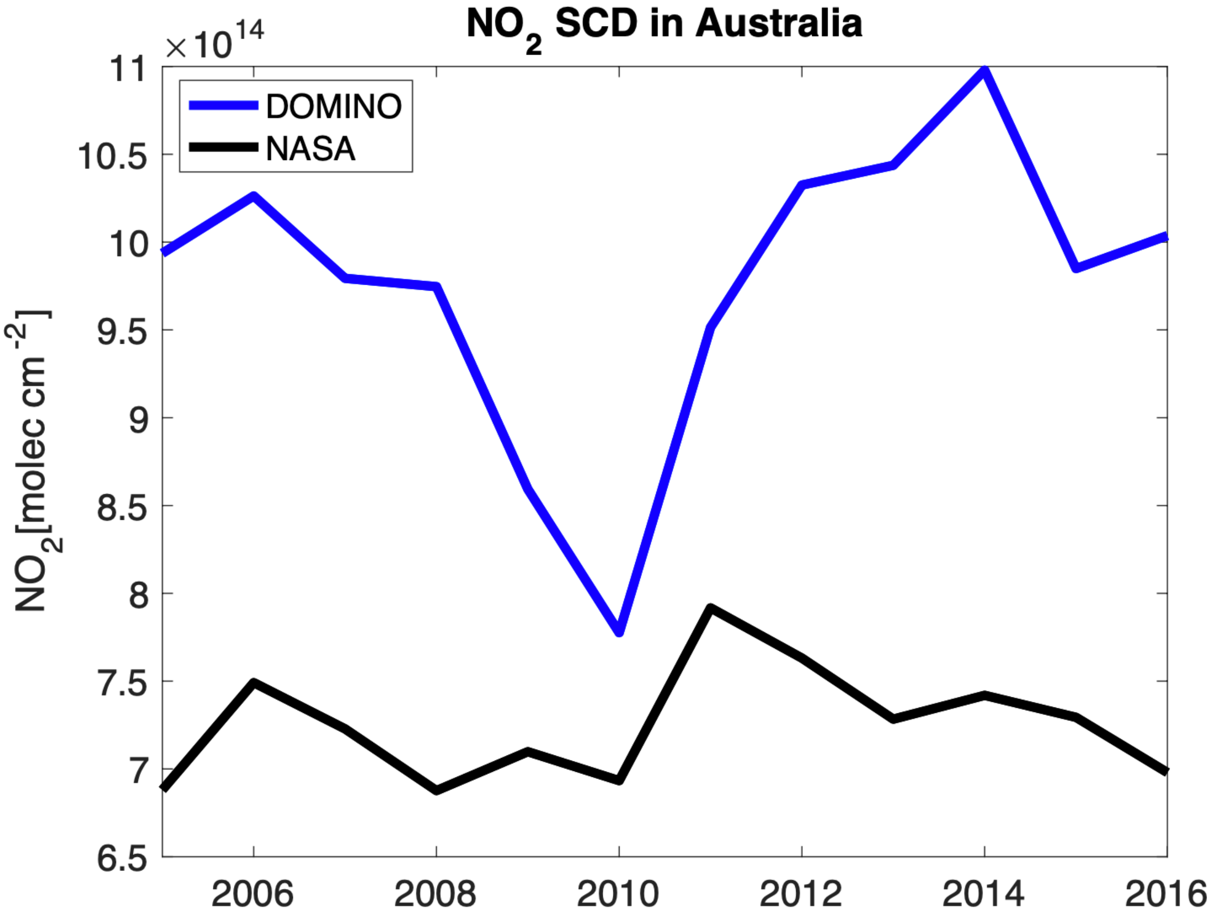


Figure R4. $NO_2$ column densities in Australia from OMI.

7) Page 10, Line 319: "Ozone measurements in 2014 decreased compared to the 2006 level in China, the US, South America and Mexico". I do not see from Figure 9 that in China ozone concentration in 2014 was lower than 2006.

Thanks for pointing this out. That statement comes from an earlier analysis that used all available TOAR sites at each year, not just sites that have continuous measurements throughout the years. We removed China and South America from that sentence.

8) Page 10, Line 314-316: How did you separate the ozone trends caused by $NO_x$ emissions vs. meteorology? A description in the main text is needed. Also, you may calculate the meteorology (non-$NO_x$) effects using either GC-adj or GCv12 results? Which one did you use in Figure 9, and how they differed?

We added the following sentences to this paragraph:

**“The second trend is calculated through simulations that use constant $NO_x$ emissions throughout the studied years. It has a similar trend from GCv12 and GC-adj as shown in the green lines in Fig. 9. The trend caused by $NO_x$ emissions is obtained by subtracting the second trend from the ozone trend simulated using $NO_x$ emissions at each corresponding year.”**

We also added dotted green lines in Fig. 9 to separately show simulated trend from non-$NO_x$ sources from GCv12 and GCadj.

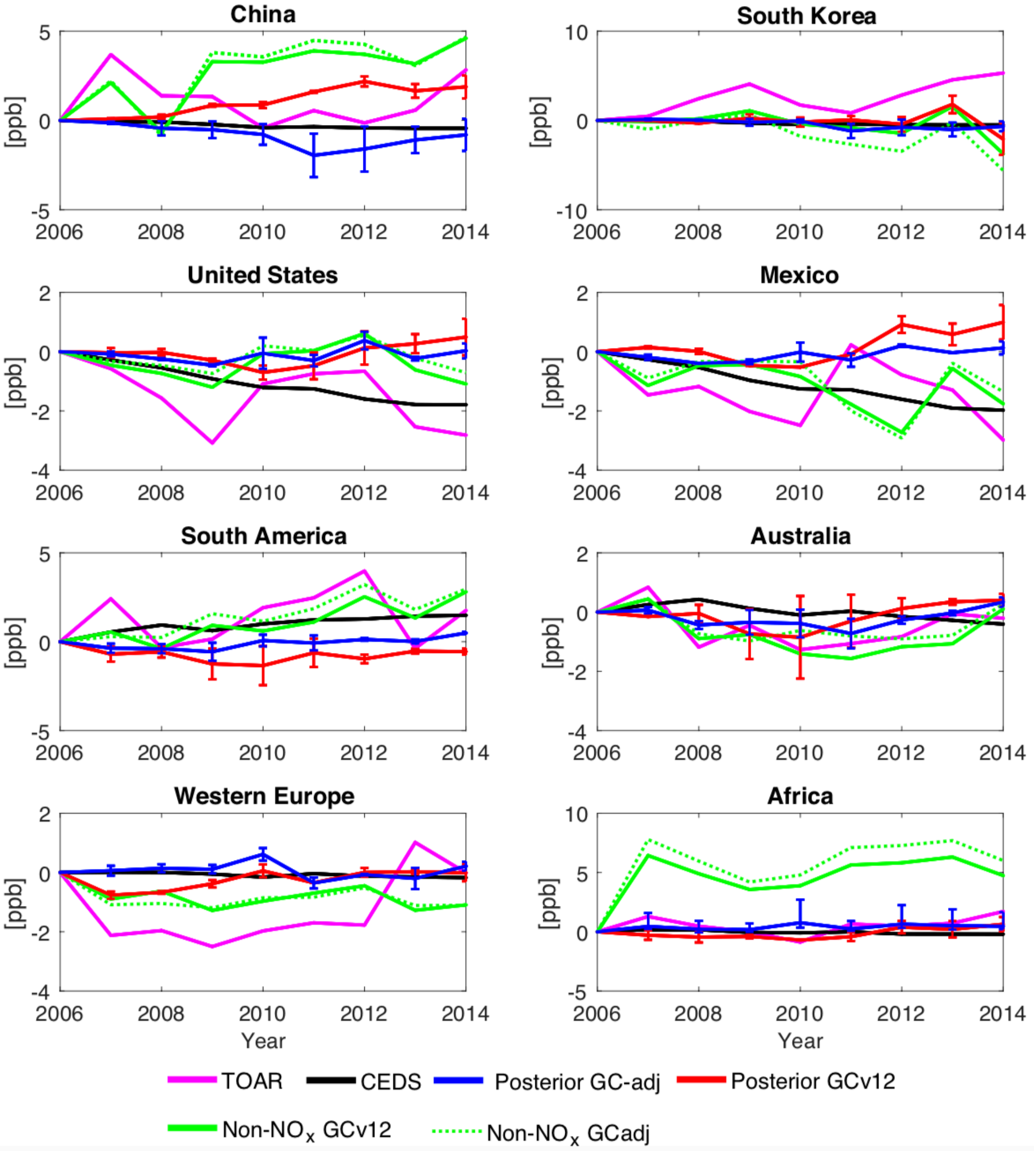


**Figure 9. Changes of regional mean annual MDA8 ozone concentrations compared to 2006 from TOAR measurements (magenta line), due to changes in bottom-up $NO_x$ emissions (black), due to changes in top-down $NO_x$ emissions (blue lines for simulations from GC-adj and red lines for simulations from GCv12), and due to changes in meteorology and non-**

**$NO_x$ emissions (green lines). Only sites that have continuous measurements throughout the 9 years are included. The vertical bars represent the spread of changes from simulations using the NASA and the DOMINO posterior $NO_x$ emissions. The impact of meteorology and natural sources are removed from black, blue and red lines by subtracting simulations using 2010 bottom-up anthropogenic emissions for all years from simulations that use bottom-up $NO_x$ emissions corresponding to each year.**

9) Page 11, Line 338: It is surprising that the model versions (GCadj and GCv12) simulate very different ozone vertical profiles. GCv12, which is a more updated version, has much large biases in the upper troposphere, in particular with the updated $NO_x$ emissions. Can you explain why in GCv12 changes in surface $NO_x$ emissions would lead to large ozone changes in the upper troposphere?

GCv12 includes halogen chemistry, which is not included in GCadj. This chemistry depletes ozone. Its impact is especially larger at locations away from $NO_x$ sources, e.g., upper troposphere, leading to much lower ozone concentrations in the GCv12 simulations.

We modified the following sentences in the cited paragraph:

"The different biases in ozone simulations **close to surface** can be explained by the usage of different emission inventories (e.g., different biogenic emissions) **and** different boundary layer mixing scheme (non-local mixing [Lin and McElroy, 2010] in GCv12 and full mixing in GCadj). **The different** chemical **mechanisms in the two model versions affect the different model biases especially in the upper troposphere**. For instance, …"

We also added the following sentences to the cited paragraph:

**"The prior simulations in GCv12 applies $NO_x$ emissions at different altitude, whereas the posterior GCv12 and all GC-adj simulations apply all $NO_x$ emissions to the surface. This leads to different transport and formation of ozone at different model layers and therefore causes larger differences in ozone simulations in the upper troposphere."**

# Impacts of global $NO_x$ inversions on $NO_2$ and ozone simulations

**Deleted:** roving

**Deleted:** $NO_2$ and ozone simulations through

**Deleted:** emission

Zhen Qu[1,2], Daven K. Henze[1], Owen R. Cooper[3,4], Jessica L. Neu[5]

[1]Department of Mechanical Engineering, University of Colorado Boulder, Boulder, CO, 80309, USA
[2]School of Engineering and Applied Science, Harvard University, Cambridge, MA, 02138, USA
[3]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, 80309, USA
[4]NOAA Chemical Sciences Laboratory, Boulder, CO, 80305, USA
[5]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 91109, USA

*Correspondence to*: Zhen Qu (zhen.qu@colorado.edu)

**Abstract.** Tropospheric NO2 and ozone simulations have large uncertainties, but their biases, seasonality and trends can be improved with NO2 assimilations. We perform global top-down estimates of monthly $NO_x$ emissions using two OMI $NO_2$ retrievals (NASAv3 and DOMINOv2) from 2005 to 2016 through a hybrid 4D-Var / mass balance inversion. Discrepancy in $NO_2$ retrieval products is a major source of uncertainties in the top-down $NO_x$ emission estimates. The 12-year averages of regional $NO_x$ budgets from the NASA posterior emissions are 37% to 53% smaller than the DOMINO posterior. Consequently, the DOMINO posterior surface $NO_2$ simulations greatly reduced the negative biases in China (by 15%) and the US (by 22%) compared to surface $NO_2$ measurements. Posterior $NO_x$ emissions show consistent trend over China, US, India, and Mexico constrained by the two retrievals. Emission trends are less robust over South America, Australia, Western Europe and Africa, where the two retrievals show less consistency. $NO_2$ trends have more consistent decreases (by 26%) with the measurements (by 32%) in the US from 2006 to 2016 when using the NASA posterior. The performance of posterior ozone simulations has spatial heterogeneities from region to region. On a global scale, ozone simulations using NASA-based emissions alleviates the double peak in the prior simulation of global ozone seasonality. The higher abundances of $NO_2$ from the DOMINO posterior increase the global background ozone concentrations and therefore reduce the negative biases more than the NASA posterior in the GEOS-Chem v12 simulations at remote sites. Compared to surface ozone measurements, posterior simulations have more consistent magnitude and interannual variations than the prior estimates, but the performance from the NASA-based and DOMINO-based emissions varies across ozone metrics. The current hard-constraints on $NO_x$ diurnal variations and limited availability of remote sensing data hinder improvement of ozone diurnal variations from the assimilation, and therefore have mixed performance on improving different ozone metrics. Additional improvements in posterior $NO_2$ and ozone simulations require more precise and consistent $NO_2$ retrieval products, more accurate diurnal variations of $NO_x$ and VOC emissions, and improved simulations of ozone chemistry and depositions.

**Deleted:** more accurate estimates of precursor gas emissions

**Deleted:** C

**Deleted:** ,

**Deleted:** GEOS-Chem adjoint $NO_2$ simulations using the DOMINO posterior $NO_x$ emissions have smaller biases in China (by 15%) and the US (by 22%), but

## 1 Introduction

Tropospheric ozone is a harmful secondary air pollutant affecting human health, sensitive vegetation, and ecosystems [NRC, 1991; Monks et al., 2015]. Long-term ozone ($O_3$) exposure is estimated to cause 1.04 – 1.23 million respiratory deaths in adults [Malley et al., 2017]. Short-term exposure to high ambient ozone is associated with respiratory and cardiovascular mortality [Turner et al., 2016; Fleming et al., 2018]. Accurate simulations of ozone in highly polluted regions are important for better pollution forecasts and more effective emission regulations. Tropospheric ozone is formed through photochemical reactions between nitrogen oxide ($NO_x$ = NO + $NO_2$), carbon monoxide (CO), methane ($CH_4$), and volatile organic compounds (VOCs) in the presence of sunlight [Crutzen, 1973; Derwent et al., 1996]. These precursor gases are mainly emitted from fossil fuel combustion, biomass burning, oil and gas production, industry, agriculture, and biogenic activities. Tropospheric ozone can also be transported from the stratosphere through stratosphere-troposphere exchange [Stohl et al., 2003; Hsu and Prather, 2009; Stevenson et al., 2006], but this magnitude is smaller than the amount from chemical production by a factor of 5 – 7 [Young et al., 2013]. Ozone is removed from the troposphere through deposition [Fowler et al., 2009], photo-dissociation, and reactions with $HO_2$, $NO_2$, unsaturated VOCs, halogens, and aerosols [Crutzen, 1973].

From 1850 to 2000, global mean tropospheric ozone burden has increased by 29% [Young et al., 2013]. Human activities are major sources of ozone precursor gases, contributing to 9% (24.98 Tg) increase of the global tropospheric ozone burden from 1980 to 2010 [Zhang et al., 2016]. Ozone formation and trends depend nonlinearly on the local relative abundances of $NO_x$ and VOCs and the radiative regime in which these occur. Previous studies have shown that changes in surface ozone are dominated by regional emission trends of precursor gases [Zhang et al., 2016]. At the global scale, 77% of $NO_x$ emissions are from anthropogenic sources, according to the HTAP 2010 inventory [Janssens-Maenhout, 2015]. Anthropogenic $NO_x$ emissions have been decreasing in North America and Europe due to transportation and energy transformations [Simon et al., 2015], but have been increasing in China up until 2011 according to bottom-up emission inventories [Liu et al., 2016; Hoesly et al., 2018]. Top-down $NO_x$ emission estimates using satellite observations from the Ozone Monitoring Instrument (OMI) showed a similar turning point in China [Miyazaki et al., 2017; Qu et al., 2017], but a slowdown in reductions in the US compared to bottom-up estimates [Miyazaki et al., 2017; Jiang et al., 2018]. However, in India and the Middle East, where ozone production is more efficient than higher latitude regions [Zhang et al., 2016], $NO_2$ column densities from OMI are continuing to increase [Krotkov et al., 2016].

Top-down methods have the advantage of being able to update emissions in a more timely fashion than the bottom-up approaches; still, top-down approaches can contain large differences and uncertainties. For instance, the magnitude of tropospheric $NO_2$ column densities from two global retrievals from the National Aeronautics and Space Administration (NASA) and the Royal Netherlands Meteorological Institute (KNMI) differ by 50%, and have different trends at the regional scale [Zheng et al., 2014; Canty et al., 2015; Qu et al., 2017]. These differences in column densities can propagate to

**Deleted:** The two posterior $NO_x$ emissions datasets have different strengths with respect to simulation of ozone concentrations. Simulations using NASA-based emissions provide better agreement for polluted conditions. Ozone from these shows the highest correlation ($R^2$ = 0.88) with annual MDA8 ozone trends and alleviates the double peak in the prior simulation of global ozone seasonality. The DOMINO-based emissions, on the other hand, are better for simulating ozone at remote sites, making them well-suited to generation of boundary conditions for regional models, and in capturing the interannual variability of daytime-average ($R^2$ = 0.72 - 0.81) and 24-hour average ($R^2$ = 0.88 – 0.96) surface ozone. We recommend using $NO_x$ emission datasets that have the best performance in the corresponding spatial domain and temporal focus to improve ozone simulations.¶

**Formatted:** No underline

**Deleted:** s

differences in top-down $NO_x$ emission estimates [e.g., Miyazaki et al., 2017; Qu et al., 2017]. In this study, we assess the importance of these discrepancies in $NO_x$ emissions for the simulation of ozone. We derive global top-down $NO_x$ emissions from 2005 to 2016 using two widely used products (OMNO2 v3 and Dutch OMI $NO_2$ (DOMINO) v2) based on the same inversion process for consistent evaluations (Sect. 3). We also evaluate a new OMI $NO_2$ retrieval product, the Quality Assurance for the Essential Climate Variables (QA4ECV) [Boersma et al., 2018], and apply it to derive monthly $NO_x$ emissions in 2010. We do not repeat our entire set of ozone evaluations with this product given that its magnitude and seasonality does not significantly differ from the other two products. We further explore the impact of adjusting $NO_x$ emissions on ozone simulations, by evaluating the ozone simulations produced from bottom-up and top-down $NO_x$ emissions against global surface measurements from the Tropospheric Ozone Assessment Report (TOAR) database and the China National Environmental Monitoring Center (CNEMC) network.

**Deleted:** ent

In addition to local sources, the lifetime of ozone (~22 days on global average) is sufficiently long enough for intercontinental transport [UNECE, 2010]. Consequently, every country is an exporter as well as an importer of ozone pollution. Transport from East Asia can be an important contributor to ozone exceedances in the western US [Goldstein et al., 2004; Zhang et al., 2009; Zhang et al., 2014; Fiore et al., 2014; Verstraeten et al., 2015; Lin et al., 2017; Jaffe et al., 2018]. The influence of intercontinental ozone transport is strongest in spring and summer, when background ozone concentrations reach 50 ppbv at the west coast of the US [Jaffe et al., 2018]. The impact of background ozone is increasingly important and challenging due to the decreased local sources of precursor gases in the US [Hoesly et al., 2018] and the recent stricter ozone standard in the US lowering the annual 4th highest maximum daily 8-hour average ozone concentration from 75 ppbv to 70 ppbv in 2015 [Cooper et al., 2015]. Optimization of $NO_x$ emissions in the upwind regions can improve remote ozone simulations in downwind regions after transport of intercontinental pollution plumes from the free troposphere to the surface [Zhang et al., 2008; Verstraeten et al., 2015]. Therefore, we also evaluate the model simulations of remote ozone at the west coast of the United States using bottom-up and top-down $NO_x$ emissions in Sect. 4.

**Deleted:** potentially

## 2 Methods

### 2.1 GEOS-Chem and its adjoint model

The GEOS-Chem adjoint model [Henze et al., 2007] v35k is used to derive global $NO_x$ emission estimates at 2° × 2.5° resolution. It was developed for inverse modelling of aerosol and gas emissions using the 4D-Var method by Henze et al. [2007, 2009] based on version 8 of GEOS-Chem, with bug fixes and updates up to version 10. Simulations in this study are driven by Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) meteorological fields from NASA Global Modeling and Assimilation Office (GMAO). Anthropogenic emissions of $NO_x$, $SO_2$, $NH_3$, CO, NMVOCs and primary aerosol from the HTAP 2010 inventory version 2 [Janssens-Maenhout et al., 2015] are used to drive all prior simulations from 2005 to 2017. The diurnal variation of $NO_x$ emissions is derived from EDGAR hourly variations (

**Formatted:** Left

**Deleted:** modeling

http://wiki.seas.harvard.edu/geos-chem/index.php/Scale_factors_for_anthropogenic_emissions#Diurnal_Variation) and is not optimized in the inversion. The use of non-anthropogenic emissions and other setups follow Qu et al. [2017, 2019]. In the following analyses, we refer to this model as "GC-adj."

GC-adj does not include several halogen chemistry mechanisms that affect ozone depletions primarily over the oceans [Sherwen et al., 2016a; Wang et al., 2019] and at high altitude regions [Sherwen et al., 2016a]. Given their impact on the global background ozone concentrations, we also use GEOS-Chem v12.1.1 to evaluate ozone simulations at 2° × 2.5° resolution driven by the MERRA-2 meteorological fields. The chemistry updates include the stratospheric chemistry from the Universal tropospheric-stratospheric Chemistry eXtension (UCX) [Eastham et al., 2014], halogen chemistry [Bell et al., 2002; Parrella et al., 2012; Sherwen et al., 2016a, 2016b; Schmidt et al., 2016; Sherwen et al., 2017], and updated isoprene and monoterpene chemistry [Chan Miller et al., 2017; Fisher et al., 2016; Marais et al., 2016; Travis et al., 2016]. The Harvard-NASA Emission Component (HEMCO) is employed to process emissions in this version of GEOS-Chem [Keller et al., 2014]. We use 72 levels of vertical grid and global anthropogenic emissions from the Community Emissions Data System (CEDS) [Hoesly et al., 2018]. Top-down $NO_x$ emissions derived using GC-adj are also input to this model to evaluate the impact of $NO_2$ data assimilation on ozone simulations under different chemical mechanisms. We refer to this model as "GCv12" in this manuscript.

For each $NO_x$ emission dataset, the model is spun-up for 6 months, starting from July 2005. Therefore, we derive $NO_x$ emissions from 2005, but only evaluate simulations with measurements from 2006. To avoid high biases when comparing simulated ozone averaged over the first vertical model layer (~100 m in box height) with surface measurements, 2-meter ozone mixing ratios are calculated by scaling simulated ozone mixing ratios in the first layer using adjusted dry deposition velocities at 2 meters following Zhang et al. [2012] and Lapina et al. [2015].

### 2.2 Satellite observations and global top-down $NO_x$ emissions

We estimate global top-down $NO_x$ emissions at the surface from 2005 to 2016 at 2° × 2.5° resolution using tropospheric $NO_2$ column densities from OMI. OMI is an Ultraviolet/Visible nadir solar backscatter spectrometer aboard the NASA Aura satellite. It has a local overpass time of about 13:45 and a nadir resolution of 13 km × 24 km. OMI was launched in July 2004 and has provided operational data products since October 2004. Two Level 2 $NO_2$ retrieval products are used to derive long-term top-down $NO_x$ emissions in this study: the NASA standard product OMNO2 version 3 [Krotkov et al., 2017] and the DOMINO version 2 from KNMI [Boersma et al., 2011]. A new OMI $NO_2$ retrieval, the Quality Assurance for the Essential Climate Variables (QA4ECV) [Boersma et al., 2018], has recently become available. This product is jointly developed by KNMI, the Belgian Institute for Space Aeronomy (BIRA-IASB), University of Bremen, Max-Plank Institute for Chemistry, and Wageningen University. We evaluate the magnitude of $NO_2$ column densities and the seasonality of posterior $NO_x$ emissions in 2010 from this product. We screen all OMI $NO_2$ retrievals using data quality flags and by the criteria of positive tropospheric column, cloud fraction < 0.2, solar zenith angle < 75°, and viewing zenith angle < 65°. We excluded all retrievals that are affected by row anomaly.

We converted GEOS-Chem $NO_2$ VCD to SCD using scattering weight (NASA product) and averaging kernel (DOMINO and QA4ECV product) from the OMI retrievals and then compare GEOS-Chem SCD with SCD retrieved from OMI. A cost function is defined as the observation error weighted differences between simulated and retrieved $NO_2$ SCD, plus the prior error weighted departure of the emission scaling factors from the prior estimates. We minimize the cost function using the quasi-Newton L-BFGS-B gradient-based optimization technique [Byrd et al., 1995; Zhu et al., 1994], in which the gradient of the cost function with respect to the control parameter is calculated using the adjoint method. Details of the assimilation of $NO_2$ slant column densities (SCDs), how vertical sensitivities of satellite retrievals are accounted for, and the hybrid 4D-Var / mass balance inversion of $NO_x$ emissions are described in Qu et al. [2017]. We use top-down $NO_x$ emissions estimated from the NASA standard product and the DOMINO product in the evaluations of ozone simulations.

**Deleted:** T

**Deleted:**

**Deleted:** follows

### 2.3 Surface measurements

We evaluate surface $NO_2$ simulations with measurements from the Environmental Protection Agency (EPA) Air Quality System (AQS) in the US and the China National Environmental Monitoring Center (CNEMC) network in China. The city monitoring sites included in the analysis represent either urban background or the averaged pollutant concentrations over the city. Simulated ozone mixing ratios from 2006 to 2016 are compared to surface observations from the TOAR Surface Ozone Database [Schultz et al., 2017] at the global scale and the CNEMC network in China. TOAR has produced a relational database of global surface ozone observations at all available sites; see Gaudel et al. [2018] for illustrations of the global coverage of the TOAR data. Precompiled TOAR data (https://doi.pangaea.de/10.1594/PANGAEA.876108, available from 1995 to 2014) at each individual site are used in this study. Given the sparse TOAR data coverage of only 32 sites over China, hourly surface ozone measurements from the CNEMC (http://106.37.208.233:20035/) are used to evaluate simulations in China from 2014 to 2016. The CNEMC national network was designed for urban and suburban air pollution monitoring. The archive contains hourly observations of ozone, carbon monoxide, nitrogen dioxide, sulfur dioxide and fine particulate matter across mainland China since 2013.

### 2.4 Ozonesonde measurements

Ozone profile measurements from the Intercontinental Chemical Transport Experiment Ozonesonde Network Study (IONS-2010) [Cooper et al., 2011] are used to evaluate the continental inflow of ozone along the west coast of the United States, where air masses are not influenced by recent US emissions. IONS-2010 was a component of the California Research at the Nexus of Air Quality and Climate Change (CalNex) 2010 experiment [Ryerson et al., 2013] and was a continuation of previous IONS experiments to measure tropospheric ozone variability across North America [Thompson et al., 2007, 2008; Cooper et al., 2007]. Balloon-borne electrochemical cell sensors were used to measure ozone profiles with an accuracy of +/-

10% in the troposphere [Johnson et al., 2002; Smit et al., 2007]. All six sites in California from IONS-2010 (Trinidad Head, Point Reyes, Point Sur, San Nicolas, Joshua Tree, and Shasta) are included in this study. These measurements are made in the mid-afternoon (95% occurring between 14:00 and 16:59 local time) over a six-week period from May 10 to June 19, 2010. There are 34-37 profiles for all sites except for San Nicolas Island, where only 26 profiles are available due to multiple instrument failures. Measurements made between 700 – 800 hPa are used to evaluate remote ozone simulations.

## 3 Magnitude, seasonality and trend of $NO_x$ emissions, surface $NO_2$ and surface ozone

Differences between the prior and posterior $NO_x$ emission estimates are mainly driven by the differences between simulated and retrieved tropospheric $NO_2$ vertical column densities (VCDs), which are compared in Sect. S1 in the supporting information. The GEOS-Chem $NO_2$ SCDs converted using scattering weight from the NASA product are larger than the SCDs calculated using the DOMINO scattering weight and the same GEOS-Chem VCDs (See Fig. S2). These can be explained by the use of different surface albedo and cloud product in the two retrievals. The retrieved $NO_2$ SCDs from the NASA product are mostly smaller than the DOMINO retrieval except for some regions between 40°N – 60°N in January 2010. The smaller magnitude in OMI SCD and the larger magnitude in GEOS-Chem SCD using the NASA scattering weight lead to smaller magnitude of posterior $NO_x$ emissions than inversions from the DOMINO product. The cost function has reduced by 6% - 29% in the monthly inversion.

### 3.1 Annual average

As shown in Table 1, the global budgets of $NO_x$ emissions from the NASA posterior in 2010 is 0.7% smaller than the prior; DOMINO posterior is 18% larger than the prior; QA4ECV posterior is 11% larger than the prior. The positive increment in the DOMINO posterior emissions is consistent with the +26% increments of 10-year mean posterior $NO_x$ emissions in Miyazaki et al. [2017]. The annual global $NO_x$ emissions from Miyazaki et al. [2017] are between 46.7 Tg N $yr^{-1}$ and 50.9 Tg N $yr^{-1}$ from 2005 to 2014, which are within 31% from the DOMINO posterior emissions in 2010 in this study.

**Deleted:** Differences between the prior and posterior $NO_x$ emission estimates are mainly driven by the differences between simulated and retrieved tropospheric $NO_2$ vertical column densities (VCDs), which are compared in Sect. S1 in the supporting information.

As shown in Fig. 1, the NASA posterior $NO_x$ emissions are less than the prior $NO_x$ emissions in the northeast US, northeast China, and southeast China. The DOMINO posterior $NO_x$ emissions are larger than the prior in most regions except for North Mexico and most parts of the tropics. The QA4ECV posterior $NO_x$ emissions have more consistent negative increments in Eastern China with the NASA posterior emissions and more consistent positive increments in the United States, India, Europe, and Australia with the DOMINO posterior emissions. At the regional scale, NASA posterior increments are -3% in China, -1% in the US, +0.3% in India, and -1% in Western Europe. The increments from the DOMINO posterior emissions are +21% in China, +31% in the US, +28% in India, and +38% in Western Europe. The different changing directions in the above two posterior $NO_x$ emissions are consistent with the reportedly higher magnitude of $NO_2$ column densities in the DOMINO product than the NASA product in densely populated and industrial regions

[Zheng et al., 2014; Canty et al., 2015; Qu et al., 2017]. The increments from the QA4ECV posterior emissions are +5% in China, +19% in the US, +18% in India, and +14% in Western Europe.

To evaluate the magnitude of the posterior $NO_x$ emissions, we compare simulations of surface $NO_2$ concentrations using the NASA and DOMINO based $NO_x$ emissions with surface measurements in the US and China. Surface $NO_2$ simulations at coarse resolution are usually biased low compared to measurements at urban sites, due to the short lifetime of $NO_x$. We therefore start with analysing this resolution error by generating high-resolution pseudo surface measurements at 0.1° × 0.1° and compare them with low-resolution model simulations at 2° × 2.5°. We generate high-resolution surface $NO_2$ concentrations by scaling simulated surface $NO_2$ concentrations at 2° × 2.5° grid cells by the ratio of OMI $NO_2$ column density gridded at 0.1° × 0.1° to the OMI $NO_2$ column density gridded at 2° × 2.5° grid cell. We identify 0.1° × 0.1° grid cells that include surface monitoring sites and treat downscaled surface $NO_2$ concentrations at these grid cells as the pseudo surface measurements. Comparisons of pseudo surface measurements and $NO_2$ simulations at 2° × 2.5° purely reflect differences caused by comparing $NO_2$ concentrations at 2° × 2.5° with higher resolution surface measurements at urban regions. The mean of the pseudo $NO_2$ measurements is 32% higher than the low-resolution simulations in the US, and it is 18% higher than the low-resolution simulations in China. The real surface measurements, which represent a single point within the 0.1° × 0.1° grid cell, are expected to have even larger biases than the values calculated here, where we assume the measurements are at 0.1° × 0.1° grid cells. The smaller bias in China in comparison to the US is related to the higher background $NO_2$ concentrations in China.

Figure 2 shows the comparisons of annual mean surface $NO_2$ concentrations in 2015 from measurements and simulations using different $NO_x$ emission inputs. The selection of this year is due to the limited availability of nation-wide surface $NO_2$ measurements in China. Surface $NO_2$ concentrations in both China and the US are measured by chemiluminescence analyzers, each equipped with a molybdenum converter, which converts additional $NO_y$ compounds to NO and leads to a positive bias in $NO_2$ measurements [Dunlea et al., 2007; Steinbacher et al., 2007]. We therefore calculate a correction factor following Lamsal et al. [2008] for each GEOS-Chem simulation and divide the simulated $NO_2$ concentrations by this correction factor to convert simulated $NO_2$ to the measured species. The correction factors are generally higher in the US than in China, but have similar seasonality (see Fig. S3). Subtracting the resolution bias from the statistics shown on Fig. 2, the equivalent normalized mean bias (NMB) of surface $NO_2$ concentrations using the NASA posterior is -54% in China and -41% in the US. The equivalent NMB using the DOMINO posterior is -38% in China and -19% in the US. These remaining negative biases reflect the unrepresentativeness of 0.1° pseudo measurements for real point measurements for resolution bias correction, comparison of $NO_2$ concentrations averaged over 2°×2.5° simulation to limited measurements, the underestimates of $NO_2$ retrievals using coarse resolution a priori, and the inability of data assimilation to increase emissions at grid cell where $NO_2$ retrievals are below the detection limit of OMI. Although we have not performed a $NO_x$ emission inversion using

**Deleted:** Valin et al. [2011] showed negative biases in the coarse resolution simulations of $NO_2$ column densities due to inaccurate $NO_2$ - OH feedback. Therefore, $NO_2$ column simulations at 2° × 2.5° in this study are likely to be underestimated and lead to high biases of posterior $NO_x$ emissions to match satellite $NO_2$ column concentrations. In addition to this bias in column concentration,

**Deleted:** s

**Deleted:** analyzing

**Deleted:** average of

**Deleted:** GEOS-Chem simulated $NO_2$ column density over the QA4ECV product for 2015, we expect its bias to lie between the results from the NASA and DOMINO products, based on the magnitude of $NO_x$ emissions in 2010.

We evaluate the simulated ozone concentrations with global surface measurements from the TOAR database using three ozone metrics: maximum daily 8-hour average (MDA8) ozone, daytime average ozone (8:00 – 20:00 local time), and 24-hour average ozone. In addition to the GC-adj simulation, with which we derived top-down $NO_x$ emissions, we also input the same top-down emissions to GCv12 and evaluate ozone simulations from this more recent version of the GEOS-Chem that includes updated halogen and isoprene chemistry.

All GC-adj simulations of 2-meter ozone concentrations have a high bias compared to the TOAR measurements in 2010. NMB and Normalized Mean Square Error (NMSE) are largest for 24-hour ozone concentrations. Simulations using posterior $NO_x$ emissions have slightly better agreement with the measurements from TOAR in 2010 (Fig. 3). In particular, simulations using the DOMINO posterior $NO_x$ emissions have the smallest NMB in all ozone metrics and the smallest NMSE in all metrics except for the North Hemisphere (NH) summertime MDA8 ozone. Simulations using the NASA posterior $NO_x$ emissions have the best spatial correlation when compared with measurements for all metrics except for the NH summer daytime ozone and annual MDA8 ozone, for which DOMINO posterior simulations have the largest correlation coefficient (Fig. S4).

**Deleted:** 2

In comparison, GCv12 simulations have a low bias in daytime ozone, but high bias in 24-hour average ozone, reflecting the potential underestimate of ozone loss at night. The impact of $NO_2$ assimilation on improving estimates of surface ozone simulations in GCv12 depends upon the ozone metric, as shown in the bottom left panel of Fig. 3. Simulations using the DOMINO posterior emissions have the smallest NMB for annual mean daytime ozone; simulations using bottom-up $NO_x$ emissions have the smallest NMB for annual mean MDA8 ozone; simulations using the NASA posterior emissions have the smallest NMB for annual mean 24-hour averaged ozone. These results suggest that the simulated diurnal variations of surface ozone concentrations may not be correct. The current constraints on $NO_x$ emissions use observations from OMI, which overpasses the same location approximately once per day. The diurnal variations of $NO_x$ emission are constrained to be those of the prior emissions. The daily $NO_2$ column densities from OMI are also underestimated compared to the diurnally varying ground-based retrievals [Herman et al., 2019]. Assimilating $NO_2$ observations from instruments overpassing at different time of the day [e.g., Boersma et al., 2008; Lin et al., 2010; Miyazaki et al., 2017] and using hourly constraints from the geostationary satellite data (e.g., Geo-stationary Environmental Monitoring Spectrometer (GEMS), Tropospheric Emissions: Monitoring of Pollution (TEMPO) [Zoogman et al., 2017] and Sentinel-4) have the potential to improve simulations of ozone diurnal variations and different ozone metrics, although the ratio of $NO_2$ column densities from satellites that overpass in the morning and afternoon are generally lower than the same ratio from surface measurements [Penn and Holloway, 2020]. Simulated MDA8 ozone values are mostly biased low in NH summer but biased

**Deleted:** OMI measurements frequently miss the high values of $NO_2$ column densities that occur before or after its overpassing time (as shown from comparisons with PANDORA measurements [Herman et al., 2019]). Future

**Deleted:** upcoming

**Deleted:** $NO_2$,

**Deleted:** ,

high in annual mean concentrations, reflecting different seasonal variations in simulated and measured ozone concentrations, which will be further discussed in Sect. 3.2. Evaluations with the CNEMC ozone measurements in China are in Sect. S2.

### 3.2 Seasonal variation

The seasonal variations of monthly $NO_x$ emissions are consistent between the prior and the NASA posterior emissions (Fig. 4). The DOMINO posterior emissions show different seasonal variations in several regions. In China, the prior and the NASA posterior $NO_x$ emissions show summer peaks, which are mainly caused by the increase of natural sources when temperatures are high and lightning occurs more often [Qu et al., 2017]. The DOMINO posterior emissions have the largest values in January and June in China, consistent with the posterior seasonality from Miyazaki et al. [2017] constrained by the same OMI $NO_2$ product. The June peak in China has been explained by the crop residual burning [Stavrakou et al., 2016]. The peak of the DOMINO posterior $NO_x$ emissions in the United States and Mexico shifted earlier in the year to June and July compared to the prior and NASA posterior emissions, similar to the results from Miyazaki et al. [2017]. The peak in DOMINO posterior emissions corresponds to the time of high soil $NO_x$ emissions, which are reported to be underestimated in high-temperature agricultural systems in the bottom-up inventory [Oikawa et al., 2015; Miyazaki et al., 2017]. The differences between the DOMINO posterior and the other two sets of emissions are especially large during the springtime in India, when biomass burning activity increases [Miyazaki et al., 2017; Venkataraman et al., 2006]. These retrieval products have similar number of observations and spatial distributions of observation densities after the filtering. The different seasonal variations in the posterior $NO_x$ emissions may reflect the AMF structural uncertainties when the retrieved $NO_2$ column densities use different ancillary data [Lorente et al., 2017]. For instance, the GEOS-Chem $NO_2$ SCDs converted using the scattering weight from the NASA product have larger seasonal variations than the SCDs converted using the DOMINO averaging kernel in the US, reflecting the different seasonal variations of vertical sensitivities from the two retrievals. The seasonal variations of simulated surface $NO_2$ concentrations are similar with measurements in China and the US (see Fig. S6).

**Deleted:** , and could possibly be explained by better constraints on soil $NO_x$ emissions in rural regions in the DOMINO posterior.

**Deleted:** se

**Deleted:** different

**Deleted:** sensitivity of OMI $NO_2$ retrievals to $NO_x$ sources and regions

**Deleted:** 4

Seasonal variations of 2-meter ozone concentrations simulated by the GC-adj are also similar despite different $NO_x$ emission inputs: the differences in correlation coefficients of the simulated and the measured monthly ozone concentrations are less than 9%. The simulations of 2-meter ozone concentrations from GCv12 show better seasonality when using the posterior $NO_x$ emissions than using the prior, as shown in Fig. 5. Simulations using the CEDS inventory show double maxima in April and August, whereas surface measurements only show a single maximum in April. Assimilation of NASA $NO_2$ concentrations alleviates this difference and leads to the largest correlation with measured MDA8 and 24-hour average ozone; simulations using the DOMINO posterior emissions have the largest correlation coefficient for daytime ozone. That being said, the correlation coefficients are not notably different. The August ozone peak in the prior simulation is mainly due to the high ozone concentrations in North China, Southwest China, and North India. The NASA and DOMINO posterior simulations have both reduced surface ozone concentrations in North China Plain and Northeast China in August due to the larger posterior NO$_x$ emissions than the prior in these high-NO$_x$ regions. Both posterior ozone simulations are also smaller than the prior in Tibet and North India due to the reductions of posterior NO$_x$ emissions in low-NO$_x$ region. The August ozone peak in the DOMINO posterior comes from the higher ozone concentrations in Angola and Democratic Republic of the Congo compared to the NASA posterior and prior simulations in the same month and DOMINO posterior simulations in the previous months. This can be explained by the larger upward adjustment of DOMINO posterior NO$_x$ emissions in South Africa in August. These results show the large spatial heterogeneities on the responses of ozone seasonality to the changes in NO$_2$ abundances on a global scale. Compared with CNEMC measurements in China, simulations using the prior emissions have the most consistent seasonal variations and smallest NMSE. All simulations have smaller seasonal variations than the measurements in daytime ozone.

### 3.3 Inter-annual variations

The three different versions of NO$_x$ emissions have different regional trends from 2005 to 2016 as shown in Fig. 6. In China, the NASA posterior NO$_x$ emissions increased by 32% and the DOMINO posterior NO$_x$ emissions increased by 32% from 2005 to 2011. From 2011 to 2016, they decreased by 20% (NASA) and 11% (DOMINO). This turning point reflects the regulation of NO$_x$ emissions in China since the “11$^{th}$ 5-year plan” in 2011. In India, both posterior NO$_x$ emissions showed continuous increases (by 24% from the NASA posterior and 34% from the DOMINO posterior) from 2005 to 2016. The sources of NO$_x$ emissions in India are mainly from thermal power and transportation and are expected to continue increasing in the near future under current regulations [Venkataraman et al., 2018]. In the US, NO$_x$ emissions decreased by 24% (NASA) and 19% (DOMINO) from 2005 to 2010 and then flattened from 2010 to 2016. This slowdown in the total top-down NO$_x$ emissions was attributed to the growing contribution from industrial, areal, and off-road mobile sources as well as the slower than expected decreases in on-road diesel NO$_x$ emissions by Jiang et al. [2018]. Silvern et al. [2019], however, argued that the slowdown was driven by the weaker decreases in background sources of NO$_x$, which has increasing contribution with the decrease of anthropogenic NO$_x$ sources. In Mexico, the two posterior NO$_x$ emissions consistently increased by 6% (NASA) and 13% (DOMINO) from 2005 to 2016. The DOMINO posterior shows more obvious increase in Mexico from 2010 to 2016. This increase in Mexico is not reflected in the bottom-up estimates from the EPA National Emissions Inventory. In Australia, the NASA posterior increases by 10% from 2005 to 2016. In comparison, the DOMINO posterior decreases from 2005 to 2010 and increases afterwards, consistent with the posterior trend from Miyazaki et al. [2017]. The different trends in posterior NO$_x$ emissions are propagated from the trends in the two OMI NO$_2$ retrieval products. The discrepancies are likely due to the different surface albedo and cloud products used in the two retrievals, which affect averaging kernel sensitivities. The trends of NO$_x$ emissions in South America are different in the two posterior estimates after 2012, when the NASA posterior emissions started to decrease by 27% and the DOMINO posterior emissions started to increase by 11% up until 2016. In Western Europe and Africa, posterior NO$_x$ emissions fluctuate and do not have a significant consistent trend from the two inversions.

**Deleted:** and Australia

**Deleted:** , and 11% (NASA) and 4% (DOMINO)

**Deleted:** , respectively

**Deleted:** The increase

**Deleted:** is more obvious

**Deleted:** from the DOMINO posterior

**Deleted:** also

**Deleted:** Throughout the period, posterior NO$_x$ emissions have different trends in Western Europe and Africa where the NASA posterior showed decreases by 11% and 0.3%, and the DOMINO posterior showed increases by 5% and 11%, respectively.

The magnitudes of DOMINO posterior $NO_x$ emissions are consistently larger than the NASA ones throughout the period. The 12-year averages of annual $NO_x$ budgets from NASA posteriors are 37% (China), 53% (India), 43% (US), 50% (Mexico), 45% (Australia), 58% (South America), 47% (Western Europe), and 46% (Africa) smaller than the DOMINO posterior.

We evaluate the trend of simulated surface $NO_2$ concentrations in the US with AQS measurements due to its availability throughout the study period (Fig. 7). From 2006 to 2016, the surface $NO_2$ concentrations show consistent decreases in the AQS measurements (by 32%) and GC-adj simulations (by 26% using the NASA posterior, by 10% using the DOMINO posterior, and by 7% using the prior emissions). Since we use the same anthropogenic emissions throughout 2006-2016 in the prior simulations, the variations in the black line reflect changes from natural sources and the impact of meteorological factors (e.g., temperature, humidity, wind, etc.). Surface $NO_2$ simulations using the NASA posterior $NO_x$ emissions also have the largest correlation coefficient when compared to the measurements ($R^2 = 0.93$ for the NASA posterior, $R^2 = 0.81$ for the DOMINO posterior, and $R^2 = 0.74$ for the prior). The more consistent trends and correlations in surface $NO_2$ simulations using the NASA posterior emissions are consistent with the larger decrease of NASA posterior $NO_x$ emissions in the US (by 20%, or for comparison a decrease of 1% in the DOMINO posterior) from 2006 to 2016, as shown in Fig. 6.

The interannual variability of global simulations of 2-meter ozone sampled at the TOAR locations is similar between GC-adj and GCv12. During the NH summer, simulations using the DOMINO posterior $NO_x$ emissions have the most consistent trend in daytime and 24-hour average ozone in both models (see Table S1); GC-adj simulations using the NASA posterior emissions have the best consistency with the measured trend of MDA8 ozone. The different performance of $NO_x$ emission datasets for different ozone metrics is a consequence of the hard constraint on $NO_2$ diurnal variations within the assimilation (and the lack of sufficient observations to constrain this). This can lead to better agreement of mean ozone concentration with measurements over particular hours but worse mean concentrations averaged over other hours. Detailed analyses of global ozone trends are in Sect. S3. At the regional scale, shown in Fig. 8, surface ozone measurements from TOAR mostly fall within the ranges of assimilation results. The interannual variations of simulated ozone over the whole region (black dotted lines) are generally smaller than the ones at grid cells that include surface measurements (black solid lines). The number of years that ozone measurements are available in each grid cell is shown in Fig. S8. The overlap of solid black and green lines in Fig. 8 suggests that interannual variations of anthropogenic $NO_x$ emissions from CEDS do not have a large impact on surface ozone simulations. The trends of simulated annual MDA8 ozone concentrations are correlated with impacts from meteorology and non-$NO_x$ sources based on simulations (shown as green lines) that use the same anthropogenic $NO_x$ emissions for all years and simulations that use interannually varied anthropogenic $NO_x$ emissions, leading to ozone changes of up to 4 ppbv (China), 5 ppbv (South Korea), 1ppbv (US), 2 ppbv (Mexico), 1 ppbv (South America), 1 ppbv (Australia), 1 ppbv (Western Europe), and 6 ppbv (Africa) from one year to the next. The trends of simulated MDA8 ozone are similar when using the NASA and the DOMINO posterior $NO_x$ emissions as inputs. The

**Deleted:**

**Deleted:** 6

**Deleted:** ,

DOMINO-derived MDA8 ozone concentrations are higher than the NASA-derived ones in all studied regions, represented by the upper and lower limit of the error bars respectively. GCv12 simulated ozone concentrations are smaller than simulations from GC-adj, especially over relatively less polluted regions, consistent with the inclusion of halogen chemistry in GCv12, which depleted ozone. The simulated MDA8 ozone trends in grid cells that include measurements in the US and Australia are more consistent with the TOAR measurements than the other regions, with coefficients of determination ($R^2$) larger than 0.45. The larger differences in ozone between the prior and posterior emissions as well as variability between the two top-down $NO_x$ emissions in GCv12 suggest a larger responsiveness of the ozone chemistry to changes in $NO_x$. We do not expect simulated ozone trends to be completely consistent with the measurements in the TOAR database due to errors in the model's transport, chemical mechanism, and VOC emissions.

We further separate the ozone trends in grid cells that include measurements into changes caused by $NO_x$ emissions as well as meteorology and non-$NO_x$ sources. The second trend is calculated through simulations that use constant $NO_x$ emissions throughout the studied years. It has similar trend from GCv12 and GC-adj as shown in the green lines in Fig. 9. The trend caused by $NO_x$ emissions is obtained by subtracting the second trend from the ozone trend simulated using $NO_x$ emissions at each corresponding year. The ozone trends due to changes in meteorology and non-$NO_x$ sources (green lines) are moderately correlated (R > 0.5) with measurements from TOAR in Australia, the US, South America, and India. The ozone trends due to changes in posterior $NO_x$ emissions (red and blue lines) only have positive correlations with TOAR measurements in both GC-adj and GCv12 simulations in Africa and Australia. Ozone measurements in 2014 decreased compared to the 2006 level in the US and Mexico. GC-adj simulations do not have big trends in these regions, whereas GC-v12 simulations show increases in China, the US, and Mexico. Meteorological and non-$NO_x$ sources lead to larger inter-annual variations in ozone simulations than those driven by $NO_x$ emissions in South America, Australia, and Africa, where anthropogenic activities are much less than the other regions. These underscore the challenges of attributing observed variations in ozone to changes in $NO_x$ emissions at regional scales.

**Deleted:** in Fig. 9.

**Deleted:** T

**Deleted:** China,

**Deleted:** , South America

## 4 Western US remote ozone

Assimilations of ozone precursor gases have the potential to improve remote ozone simulations, which can be used to provide boundary conditions for regional air quality models and to quantify and attribute sources of background ozone. We therefore focus specifically on remote regions in the US in this section to evaluate the vertical profile and surface concentrations of ozone simulations.

### 4.1 Evaluations with ozonesonde profiles

Field campaigns and routine observations of ozone concentrations along the west coast of the US have provided opportunities to understand regional and intercontinental influences on surface air quality [Cooper et al., 2015]. Evaluations with the IONS-2010 measurements in Fig. 10 show that the GCv12 simulations of ozone vertical profiles have negative biases (NMB between -8% and -32%) above all six sites. The standard deviations of ozonesonde and simulated profiles overlap with each other (see Fig. S9). The GC-adj simulations have positive biases at San Nicolas and Trinidad Head and have smaller negative biases (NMB between -3% and -11%) at the remaining sites than the GCv12 simulations. The magnitudes of the NMSE and NMB of the GCv12 simulations at 700 – 900 hPa are also larger than those of the GC-adj simulations (see Fig. S10). The prior simulations in GCv12 applies $NO_x$ emissions at different altitude, whereas the posterior GCv12 and all GC-adj simulations apply all $NO_x$ emissions to the surface. This leads to different transport and formation of ozone at different model layers and therefore causes larger differences in ozone simulations in the upper troposphere. The air masses at this altitude in the eastern Pacific are demonstrated to impact inland near surface ozone concentrations [Cooper et al., 2011; Lin et al., 2012; Yates et al., 2015]. The different biases in ozone simulations close to surface can be explained by the usage of different emission inventories (e.g., different biogenic emissions) and different boundary layer mixing scheme (non-local mixing [Lin and McElroy, 2010] in GCv12 and full mixing in GCadj). The different chemical mechanisms in the two model versions affect the different model biases especially in the upper troposphere. For instance, inclusion of halogen chemistry and additional chlorine chemistry in GCv12 leads to 19% and 7% decreases of global tropospheric ozone burden [Sherwen et al., 2016a; Wang et al., 2019]. GCv12 simulations using the CEDS emissions have smaller NMSE and NMB than the simulations using the posterior $NO_x$ emissions in all 6 sites in 2010. In comparison, the GC-adj simulations using the DOMINO posterior $NO_x$ emissions have the smallest NMSE and NMB at all sites except for San Nicolas and Trinidad Head, where the prior simulations have the smallest error and bias. Further evaluations with ozonesondes at Trinidad Head in 2016 are shown in Sect. S4.

**Deleted:** 7

**Deleted:** 8

**Deleted:** ,

**Deleted:** ,

**Deleted:** and differences

**Deleted:** in

**Deleted:** .

### 4.2 Evaluations with TOAR surface ozone measurements at remote sites

To further evaluate the model performance under different geographical scenarios, we compare surface ozone simulations from GC-adj and GCv12 with observations from simple to complex environments. These include 1) Mauna Loa Observatory and Mt Bachelor Observatory at night, which represent the lower free troposphere; 2) Mt. Bachelor Observatory, Lassen Volcanic National Park, Great Basin National Park, and Sequoia / Kings Canyon National Park at daytime, representing high elevation rural sites during well-mixed daytime conditions. The coefficients of determination (see Table S2) between the simulations and the measurements are larger than 0.6 for all daytime ozone comparisons except for Mt. Bachelor Observatory. The correlation coefficients are smaller than 0.5 for all nighttime comparisons, reflecting the need to further improve simulations of nighttime chemistry and atmospheric processes.

In Fig. 11, the surface ozone concentrations from both GC-adj and GCv12 simulations have low biases compared to the surface measurements at remote sites. These low biases in the GCv12 simulations are consistent with its performances when evaluated with ozonesondes from IONS-2010 and with daytime surface ozone at the global scale. However, the low biases in the GC-adj simulations are different from its high biases when compared with the global surface ozone concentrations and the ozone profiles at San Nicolas and Trinidad Head. This demonstrates the different biases in ozone simulations at rural and urban sites. Simulations using the DOMINO posterior emissions have the smallest NMSE and NMB at all remote sites except for the GCv12 simulations at Mauna Loa at night and Great Basin during the daytime.

## 5 Discussion and conclusions

We performed global inversions of $NO_x$ emissions from 2005 to 2016 using two widely used OMI $NO_2$ retrievals from NASA (OMNO2 v3) and KNMI (DOMINO v2). The DOMINO posterior $NO_x$ emissions have larger magnitude than the prior and the NASA posterior. Consequently, GC-adj simulations using the DOMINO posterior $NO_x$ emissions have the smallest negative bias in surface $NO_2$ and the smallest positive bias in 2-meter ozone. The impact of $NO_2$ assimilations on improving estimates of the GCv12 surface ozone simulations depends upon the ozone metrics, suggesting inaccurate diurnal variations in the surface ozone simulations. GEOS-Chem simulations using the DOMINO posterior emissions have the largest coefficients of determination for summertime daytime ($R^2$=0.81) and summertime 24-hour ($R^2$=0.96) ozone. Simulations using the NASA posterior emissions have the smallest bias and error for all ozone metrics and the largest correlation for summertime MDA8 ozone ($R^2$ = 0.88). Ozone simulations with GEOS-Chem v12.1.1 using the DOMINO posterior $NO_x$ emissions lead to the most consistent seasonality in 24-hour average ozone ($R^2$ = 0.99) with TOAR measurements, while the NASA posterior emissions lead to the best agreement in seasonal variations of MDA8 (R = 0.96) and daytime ozone (R = 0.98). The interannual variations of posterior $NO_x$ emissions from the two products are similar in China, India, the US, Mexico and Australia, but different in South America, West Europe and Africa. Surface $NO_2$ simulations using the NASA posterior have the best agreement with measurements in the US. Daytime and 24-hour average ozone simulations using the DOMINO posterior also have the best trend (R = 0.72 and 0.88) in the Northern Hemisphere summer. The GC-adj simulations using the NASA posterior $NO_x$ emissions have the best trend in MDA8 ozone in NH summer.

Posterior $NO_x$ emissions lead to improved simulations of ozone at several remote sites in the western US. The GC-adj simulations using the DOMINO posterior emissions have the smallest NMSE and NMB compared to ozonesonde measurements during IONS-2010, except for the San Nicolas and Trinidad Head sites. At the remote surface sites evaluated in this study, surface ozone simulations using the DOMINO posterior emissions have the best performance except for GCv12 simulations at Mauna Loa at night and Great Basin during the daytime. The reduced negative biases in daytime surface ozone simulations using the DOMINO posterior emissions at these remote sites and at most IONS-2010 sites are consistent with the increases of daytime remote ozone in the western US through $NO_2$ and ozone data assimilation in Huang et al. [2015]. Simulations using the DOMINO posterior emissions are demonstrated to provide more precise magnitudes at these remote sites and can potentially be used as boundary conditions for regional air quality models for further air pollution and health studies.

The remaining differences between simulated and measured ozone can be explained by the roles of VOCs, errors in satellite retrievals, and uncertainties in the chemical and physical processes in the model simulations. In addition to $NO_x$, emissions of other ozone precursors also impact the accuracy of ozone simulations. For instance, inversion of isoprene emissions over the southeast US decreases surface ozone simulations by 1-3 ppbv [Kaiser et al., 2018]. Inversion of non-methane VOC emissions changes surface afternoon ozone simulations by up to 10 ppbv in China [Cao et al., 2018]. Assimilation of multiple species (e.g, ozone, CO, $HNO_3$ and $SO_2$) together with $NO_2$ may improve posterior ozone simulations, but the performance of posterior simulations may depend on the chemical transport model, as shown in Miyazaki et al. [2020], where the GEOS-Chem adjoint model v35 shows mixed performance in correcting the bias between ozonesonde and posterior simulations between 850-500 hPa at different latitude band. Both OMI $NO_2$ retrievals employed in this study use $NO_2$ vertical shape factors from coarse resolution simulations, and therefore are biased low compared to in-situ measurements [Goldberg et al., 2017]. These retrievals also have not explicitly accounted for the aerosol optical effects, which are demonstrated to degrade the accuracy of $NO_2$ column concentrations when AOD is very high [Chimot et al., 2016; Liu et al., 2019; Cooper et al., 2019]. The differences in the magnitude of ozone concentrations from GC-adj and GCv12 reflect the impact of other species emissions and chemical mechanisms on the bias of ozone simulations. Previous studies also show that global simulations at coarse resolution are not able to capture the observed persistence of chemical plumes in the free troposphere on intercontinental scales, therefore leading to underestimates of remote ozone concentrations [Hudman et al., 2004; Zhuang et al., 2018].

Although biases, errors, seasonalities and inter-annual variations of ozone simulations have been improved in several cases through constraints on $NO_x$ emissions, there are still large discrepancies in the vertical profile and diurnal variations between ozone simulations and measurements. For instance, the different performances of each set of $NO_x$ emissions on the simulations of different ozone metrics reflect errors in the ozone diurnal simulations. The differences in ozone vertical profiles suggest errors in vertical transport in the model. These discrepancies could not be improved by adjusting only surface $NO_x$ emissions using observations at one time of the day, as performed in this study. Future geostationary satellite observations will provide opportunities to update $NO_x$ emissions at every hour. Separately constraining $NO_x$ emissions from surface (e.g., anthropogenic sources) and upper atmosphere (e.g., lightning sources [Pickering et al., 2016]) and implementing these posterior $NO_x$ emissions at their corresponding vertical levels can potentially improve the vertical profile of ozone simulations.

## Data Availability

The OMI $NO_2$ NASA standard product is downloaded from GES DISC (https://atrain.gesdisc.eosdis.nasa.gov/data/OMI/OMNO2_CPR.003/). The DOMINO and QA4ECV $NO_2$ retrievals are from KNMI (http://www.temis.nl/airpollution/no2col/no2regioomi_v2.php, http://www.temis.nl/airpollution/no2col/no2regioomi_qa.php). Ozonesonde profiles from Shasta, Big Sur, Point Reyes, Joshua Tree and San Nicolas Island are available from the NOAA Global Monitoring Laboratory (ftp://aftp.cmdl.noaa.gov/data/ozwv/Ozonesonde/2_Field%20Projects/CALNEX/ Ozonesondes from Trinidad Head are also available from the NOAA Global) Monitoring Laboratory( ftp://aftp.cmdl.noaa.gov/data/ozwv/Ozonesonde/Trinidad%20Head,%20California/100%20Meter%20Average%20Files/). Precompiled TOAR ozone data were downloaded from: https://doi.pangaea.de/10.1594/PANGAEA.876108.

## Author contribution

Z.Q., D.K.H., O.R.C, and J.L.N. designed the research; Z. Q. performed the research and prepared the paper with help from all authors.

## Acknowledgements

Z. Qu and D. K. Henze acknowledge funding support from National Aeronautics and Space Administration (NASA) HAQAST NNX16AQ26G and NASA ACMAP NNX17AF63G. Part of the computing resources supporting this work was provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center. Z. Qu would also like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation.

## References

Bell, N., L. Hsu, D. J. Jacob, M. G. Schultz, D. R. Blake, J. H. Butler, D. B. King, J. M. Lobert, and E. Maier-Reimer: Methyl iodide: Atmospheric budget and use as a tracer of marine convection in global models, J. Geophys. Res. Atmos., 107(D17), ACH 8–1–ACH 8–12, doi:10.1029/2001JD001151, 2002.

Boersma, K. F., D. J. Jacob, H. J. Eskes, R. W. Pinder, J. Wang, R. J. van der A: Intercomparison of SCIAMACHY and OMI tropospheric NO2 columns: Observing the diurnal evolution of chemistr and emissions from space, J. Geophys. Res. Atmos., 113(D16), doi:10.1029/2007JD008816, 2007.

Boersma, K. F. et al.: An improved tropospheric NO2 column retrieval algorithm for the Ozone Monitoring Instrument, Atmos. Meas. Tech., 4(9), 1905–1928, doi:10.5194/amt-4-1905-2011, 2011.

Boersma, K. F. et al.: An improved tropospheric NO2 column retrieval algorithm for the Ozone Monitoring Instrument, Atmos. Meas. Tech., 4(9), 1905–1928, doi:10.5194/amt-4-1905-2011, 2011.

Boersma, K. F. et al.: Improving algorithms and uncertainty estimates for satellite NO2 retrievals: results from the quality assurance for the essential climate variables (QA4ECV) project, Atmos. Meas. Tech., 11(12), 6651–6678, 2018.

Canty, T. P., L. Hembeck, T. P. Vinciguerra, D. C. Anderson, D. L. Goldberg, S. F. Carpenter, D. J. Allen, C. P. Loughner, R. J. Salawitch, and R. R. Dickerson: Ozone and NOx chemistry in the eastern US: evaluation of CMAQ/CB05 with satellite (OMI) data, Atmos. Chem. Phys., 15(19), 10965–10982, doi:10.5194/acp-15-10965-2015, 2015.

Cao, H. et al.: Adjoint inversion of Chinese non-methane volatile organic compound emissions using space-based observations of formaldehyde and glyoxal, Atmos. Chem. Phys., 18(20), 15017–15046, 2018.

Chan Miller, C. et al.: Glyoxal yield from isoprene oxidation and relation to formaldehyde: chemical mechanism, constraints from SENEX aircraft observations, and interpretation of OMI satellite data, Atmos. Chem. Phys., 17(14), 8725–8738, doi:10.5194/acp-17-8725-2017, 2017.

Chimot, J., T. Vlemmix, J. P. Veefkind, J. F. de Haan, and P. F. Levelt: Impact of aerosols on the OMI tropospheric NO2 retrievals over industrialized regions: how accurate is the aerosol correction of cloud-free scenes via a simple cloud model? Atmos. Meas. Tech., 9(2), 359–382, doi:10.5194/amt-9-359-2016, 2016.

Cooper, M. J., R. V. Martin, M. S. Hammer, and C. A. McLinden: An Observation-Based Correction for Aerosol Effects on Nitrogen Dioxide Column Retrievals Using the Absorbing Aerosol Index, Geophys. Res. Lett., 46(14), 8442–8452, doi:10.1029/2019GL083673, 2019.

Cooper, O. R. et al.: Evidence for a recurring eastern North America upper tropospheric ozone maximum during summer, J. Geophys. Res. Atmos., 112(D23), doi:10.1029/2007JD008710, 2007.

Cooper, O. R. et al.: Measurement of western U.S. baseline ozone from the surface to the tropopause and assessment of downwind impact regions, J. Geophys. Res. Atmos., 116(D21), doi:10.1029/2011JD016095, 2011.

Cooper, O. R., A. O. Langford, D. D. Parrish, and D. W. Fahey: Challenges of a lowered U.S. ozone standard, Science, 348(6239), 1096–1097, doi:10.1126/science.aaa5748, 2015.

Crutzen, P.: A discussion of the chemistry of some minor constituents in the stratosphere and troposphere, Pure App. Geophys., 106(1), 1385–1399, 1973.

Derwent, R. G., M. E. Jenkin, and S. M. Saunders: Photochemical ozone creation potentials for a large number of reactive hydrocarbons under European conditions, Atmos. Environ., 30(2), 181–199, 1996.

Duncan, B. N., L. N. Lamsal, A. M. Thompson, Y. Yoshida, Z. Lu, D. G. Streets, M. M. Hurwitz, and K. E. Pickering: A space-based, high-resolution view of notable changes in urban NOx pollution around the world (2005–2014), J. Geophys. Res. Atmos., 121(2), 976–996, doi:10.1002/2015JD024121, 2016.

Dunlea, E. J. et al.: Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment, Atmos. Chem. Phys., 7(10), 2691–2704, 2007.

Eastham, S. D., D. K. Weisenstein, and S. R. H. Barrett: Development and evaluation of the unified tropospheric–stratospheric chemistry extension (UCX) for the global chemistry-transport model GEOS-Chem, Atmos. Environ., 89, 52–63, 2014.

Fiore, A. M., J. T. Oberman, M. Y. Lin, L. Zhang, O. E. Clifton, D. J. Jacob, V. Naik, L. W. Horowitz, J. P. Pinto, and G. P. Milly: Estimating North American background ozone in U.S. surface air with two independent global models: Variability, uncertainties, and recommendations, Atmos. Environ., 96, 284–300, 2014.

Fiore, A. M., V. Naik, and E. M. Leibensperger: Air Quality and Climate Connections, J. Air Waste Manage., 65(6), 645–685, doi:10.1080/10962247.2015.1040526, 2015.

Fisher, J. A. et al.: Organic nitrate chemistry and its implications for nitrogen budgets in an isoprene- and monoterpene-rich atmosphere: constraints from aircraft (SEAC4RS) and ground-based (SOAS) observations in the Southeast US, Atmos. Chem. Phys., 16(9), 5969–5991, doi:10.5194/acp-16-5969-2016, 2016.

Fleming, Z. L. et al.: Tropospheric Ozone Assessment Report: Present-day ozone distribution and trends relevant to human health, Elem Sci Anth, 6(1), 12, doi:doi.org/10.1525/elementa.273, 2018.

Fowler, D. et al.: Atmospheric Composition Change: ecosystems-Atmosfere Interactions, Atmos. Environ., 43, 5193–5267, doi:10.1016/j.atmosenv.2009.07.068, 2009.

Gaudel, A. et al.: Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation, Elem. Sci. Anth., 6(1), 39, doi:10.1525/elementa.291, 2018.

Goldberg, D. L., L. N. Lamsal, C. P. Loughner, W. H. Swartz, Z. Lu, and D. G. Streets: A high-resolution and observationally constrained OMI NO2 satellite retrieval, Atmos. Chem. Phys., 17(18), 11403–11421, doi:10.5194/acp-17-11403-2017, 2017.

Goldstein, A. H., D. B. Millet, M. McKay, L. Jaegle, L. Horowitz, O. Cooper, R. Hudman, D. J. Jacob, S. Oltmans, and A. Clarke: Impact of Asian emissions on observations at Trinidad Head, California, during ITCT 2K2, J. Geophys. Res. Atmos., 109(D23), doi:10.1029/2003JD004406, 2004.

Henze, D. K., A. Hakami, and J. H. Seinfeld: Development of the adjoint of GEOS-Chem, Atmos. Chem. Phys., 7(9), 2413–2433, doi:10.5194/acp-7-2413-2007, 2007.

Henze, D. K., J. H. Seinfeld, and D. T. Shindell: Inverse modeling and mapping US air quality influences of inorganic PM2.5 precursor emissions using the adjoint of GEOS-Chem, Atmos. Chem. Phys., 9(16), 5877–5903, 2009.

Herman, J., N. Abuhassan, J. Kim, M. Dubey, M. Raponi, and M. Tzortziou: Underestimation of Column NO2 Amounts from the OMI Satellite Compared to Diurnally Varying Ground-Based Retrievals from Multiple Pandora Spectrometer Instruments, AMTD, 2019, 1–35, doi:doi.org/10.5194/amt-2019-123, 2019.

Hoesly, R. M. et al.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geosci. Model Dev., 11(1), 369–408, doi:10.5194/gmd-11-369-2018, 2018.

Hsu, J., and M. J. Prather: Stratospheric variability and tropospheric ozone, J. Geophys. Res. Atmos., 114(D6), doi:10.1029/2008JD010942, 2009.

Huang, M., K. W. Bowman, G. R. Carmichael, M. Lee, T. Chai, S. N. Spak, D. K. Henze, A. S. Darmenov, and A. M. da Silva: Improved western U.S. background ozone estimates via constraining nonlocal and local source contributions using Aura TES and OMI observations, J. Geophys. Res. Atmos., 120(8), 3572–3592, doi:10.1002/2014JD022993, 2015.

Hudman, R. C. et al.: Ozone production in transpacific Asian pollution plumes and implications for ozone air quality in California, J. Geophys. Res. Atmos., 109(D23), doi:10.1029/2004JD004974, 2004.

Jaffe, D. A., O. R. Cooper, A. M. Fiore, B. H. Henderson, G. S. Tonneson, A. G. Russell, D. K. Henze, A. O. Langford, M.

Lin, and T. Moore: Scientific assessment of background ozone over the U.S.: Implications for air quality management, Elem Sci Anth, 6(1), 56, doi:http://doi.org/10.1525/elementa.309, 2018.

Janssens-Maenhout, G. et al.: HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, Atmos. Chem. Phys., 15(19), 11411–11432, doi:10.5194/acp-15-11411-2015, 2015.

Jiang, Z. et al.: Unexpected slowdown of US pollutant emission reduction in the past decade, Proc. Natl. Acad. Sci., 115(20), 5099–5104, doi:10.1073/pnas.1801191115, 2018.

Johnson, B. J., S. J. Oltmans, H. Vömel, H. G. J. Smit, T. Deshler, and C. Kröger: Electrochemical concentration cell (ECC) ozonesonde pump efficiency measurements and tests on the sensitivity to ozone of buffered and unbuffered ECC sensor cathode solutions, J. Geophys. Res. Atmos., 107(D19), ACH 8–1–ACH 8–18, doi:10.1029/2001JD000557, 2002.

Kaiser, J. et al.: High-resolution inversion of OMI formaldehyde columns to quantify isoprene emission on ecosystem-relevant scales: application to the southeast US, Atmos. Chem. Phys., 18(8), 5483–5497, 2018.

Keller, C. A., M. S. Long, R. M. Yantosca, A. M. Da Silva, S. Pawson, and D. J. Jacob: HEMCO v1.0: a versatile, ESMF-compliant component for calculating emissions in atmospheric models, Geosci. Model Dev., 7(4), 1409–1417, doi:10.5194/gmd-7-1409-2014, 2014.

Krotkov, N. A. et al.: Aura OMI observations of regional $SO_2$ and $NO_2$ pollution changes from 2005 to 2015, Atmos. Chem. Phys., 16, 4605–4629, doi:10.5194/acp-16-4605-2016, 2016.

Krotkov, N. A., L. N. Lamsal, E. A. Celarier, W. H. Swartz, S. V. Marchenko, E. J. Bucsela, K. L. Chan, M. Wenig, and M. Zara: The version 3 OMI NO2 standard product, Atmos. Meas. Tech., 10(9), 3133–3149, doi:10.5194/amt-10-3133-2017, 2017.

Lamsal, L. N., R. V. Martin, A. van Donkelaar, M. Steinbacher, E. A. Celarier, E. Bucsela, E. J. Dunlea, and J. P. Pinto: Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument, J. Geophys. Res. Atmos., doi:10.1029/2007JD009235, 2008.

Lin, J. and M. B. McElroy: Impacts of boundary layer mixing on pollutant vertical profiles in the lower troposphere: Implications to satellite remote sensing, Atmos. Environ., 44, 1726-1739, doi:10.1016/j.atmosenv.2010.02.009, 2010.

Lin, J., M. B. McElroy, and K. F. Boersma: Constraint of anthropogenic $NO_x$ emissions in China from different sectors: a new methodology using multiple satellite retrievals, Atmos. Chem. Phys., 10, 63-78, 2010, doi:10.5149/acp-10-63-

2010, 2010.

Lin, M. et al.: Transport of Asian ozone pollution into surface air over the western United States in spring, J. Geophys. Res. Atmos., 117, D00V07, doi:10.1029/2011JD016961, 2012.

Lin, M., L. W. Horowitz, R. Payton, A. M. Fiore, and G. Tonnesen: US surface ozone trends and extremes from 1980 to 2014: quantifying the roles of rising Asian emissions, domestic controls, wildfires, and climate, Atmos. Chem. Phys., 17(4), 2943–2970, doi:10.5194/acp-17-2943-2017, 2017.

Liu, F., Q. Zhang, R. J. van der A, B. Zheng, D. Tong, L. Yan, Y. Zheng, and K. He: Recent reduction in NO x emissions over China: synthesis of satellite observations and emission inventories, Environ. Res. Lett., 11(11), 114002, doi:10.1088/1748-9326/11/11/114002, 2016.

Liu, M. et al.: Improved aerosol correction for OMI tropospheric NO2 retrieval over East Asia: constraint from CALIOP aerosol vertical profile, Atmos. Meas. Tech., 12(1), 1–21, 2019.

Malley, C. S., D. K. Henze, J. C. I. Kuylenstierna, H. W. Vallack, Y. Davila, S. C. Anenberg, M. C. Turner, and M. R. Ashmore: Updated global estimates of respiratory mortality in adults ≥ 30 years of age attributable to long-term ozone exposure, Environ. Health Perspect., 125(8), 087021, doi:10.1289/EHP1390, 2017.

Marais, E. A. et al.: Aqueous-phase mechanism for secondary organic aerosol formation from isoprene: application to the southeast United States and co-benefit of SO2 emission controls, Atmos. Chem. Phys., 16(3), 1603–1618, 2016.

Miyazaki, K., H. Eskes, K. Sudo, K. F. Boersma, K. Bowman, and Y. Kanaya: Decadal changes in global surface $NO_X$ emissions from multi-constituent satellite data assimilation, Atmos. Chem. Phys., 17(2), 807–837, doi:10.5194/acp-17-807-2017, 2017.

Monks, P. S. et al.: Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer, Atmos. Chem. Phys., 15(15), 8889–8973, doi:10.5194/acp-15-8889-2015, 2015.

National Research Council (NRC): Rethinking the ozone problem in urban and regional air pollution, National Academy Press, Washington D.C, 1991.

Parrella, J. P. et al.: Tropospheric bromine chemistry: implications for present and pre-industrial ozone and mercury, Atmos. Chem. Phys., 12(15), 6723–6740, doi:10.5194/acp-12-6723-2012, 2012.

Penn, E. and T. Holloway: Evaluating current satellite capability to observe diurnal change in nitrogen oxides in preparation for geostationary satellite missions, Environ. Res. Lett., 15, doi:10.1088/1748-9326/ab6b36, 2020.

Pickering, K. E., E. Bucsela, D. Allen, A. Ring, R. Holzworth, and N. Krotkov: Estimates of lightning NOx production based on OMI NO2 observations over the Gulf of Mexico, J. Geophys. Res. Atmos., 121(14), 8668–8691, doi:10.1002/2015JD024179, 2016.

Qu, Z., D. K. Henze, S. L. Capps, Y. Wang, X. Xu, J. Wang, and M. Keller: Monthly top-down NOx emissions for China (2005–2012): A hybrid inversion method and trend analysis, J. Geophys. Res. Atmos., 122(8), 4600–4625, 2017.

Qu, Z., D. K. Henze, N. Theys, J. Wang, and W. Wang: Hybrid Mass Balance/4D-Var Joint Inversion of NOx and SO2 Emissions in East Asia, J. Geophys. Res. Atmos., 124(14), 8203–8224, doi:10.1029/2018JD030240, 2019.

Ryerson, T. B. et al.: The 2010 California Research at the Nexus of Air Quality and Climate Change (CalNex) field study, J. Geophys. Res. Atmos., 118, 5830-5866, doi:10.1002/jgrd.50331, 2013.

Schmidt, J. A. et al.: Modeling the observed tropospheric BrO background: Importance of multiphase chemistry and implications for ozone, OH, and mercury, J. Geophys. Res. Atmos., 121(19), 11–819–11–835, doi:10.1002/2015JD024229, 2016.

Schultz, M. G. et al.: Tropospheric Ozone Assessment Report: Database and metrics data of global surface ozone observations, Elem. Sci. Anth., 5, 58, doi:10.1525/elementa.244, 2017.

Sherwen, T. et al.: Global impacts of tropospheric halogens (Cl, Br, I) on oxidants and composition in GEOS-Chem, Atmos. Chem. Phys., 16(18), 12239–12271, doi:10.5194/acp-16-12239-2016, 2016a.

Sherwen, T. et al.: Iodine's impact on tropospheric oxidants: a global model study in GEOS-Chem, Atmos. Chem. Phys., 16(2), 1161–1186, doi:10.5194/acp-16-1161-2016, 2016b.

Sherwen, T., M. J. Evans, L. J. Carpenter, J. A. Schmidt, and L. J. Mickley: Halogen chemistry reduces tropospheric O3 radiative forcing, Atmos. Chem. Phys., 17(2), 1557–1569, doi:10.5194/acp-17-1557-2017, 2017.

Silvern, R. F. et al.: Using satellite observations of tropospheric NO2 columns to infer long-term trends in US NOx emissions: the importance of accounting for the free tropospheric NO2 background, Atmos. Chem. Phys., 19(13), 8863–8878, 2019.

Simon, H., A. Reff, B. Wells, J. Xing, and N. Frank: Ozone Trends Across the United States over a Period of Decreasing NOx and VOC Emissions, Environ. Sci. Technol., 49(1), 186–195, doi:10.1021/es504514z, 2015.

Smit, H. G. J. et al.: Assessment of the performance of ECC-ozonesondes under quasi-flight conditions in the environmental simulation chamber: Insights from the Juelich Ozone Sonde Intercomparison Experiment (JOSIE), J. Geophys. Res. Atmos., 112(D19), doi:10.1029/2006JD007308, 2007.

Stavrakou, T., Müller, J.-F., Bauwens, M., De Smedt, I., Lerot, C., Van Roozendael, M., Coheur, P.-F., Clerbaux, C., Boersma, K. F., van der A., R. J., and Song, Y.: Substantial underestimation of post-harvest burning emissions in the North China Plain revealed by multi-species space observations, Sci. Rep., 6(32307), doi:10.1038/srep32307, 2016.

Steinbacher, M., C. Zellweger, B. Schwarzenbach, S. Bugmann, B. Buchmann, C. Ordóñez, A. S. H. Prevot, and C. Hueglin: Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques, J. Geophys. Res. Atmos., 112(D11), doi:10.1029/2006JD007971, 2007.

Stevenson, D. S. et al.: Multimodel ensemble simulations of present-day and near-future tropospheric ozone, J. Geophys. Res. Atmos., 111(D8), doi:10.1029/2005JD006338, 2006.

Stohl, A. et al.: Stratosphere-troposphere exchange: A review, and what we have learned from STACCATO, J. Geophys. Res. Atmos., 108(D12), doi:10.1029/2002JD002490, 2003.

Thompson, A. M. et al.: Intercontinental Chemical Transport Experiment Ozonesonde Network Study (IONS) 2004: 1. Summertime upper troposphere/lower stratosphere ozone over northeastern North America, J. Geophys. Res. Atmos., 112(D12), doi:10.1029/2006JD007441, 2007.

Thompson, A. M., J. E. Yorks, S. K. Miller, J. C. Witte, K. M. Dougherty, G. A. Morris, D. Baumgardner, L. Ladino, and B. Rappenglück: Tropospheric ozone sources and wave activity over Mexico City and Houston during MILAGRO/Intercontinental Transport Experiment (INTEX-B) Ozonesonde Network Study, 2006 (IONS-06), Atmos. Chem. Phys., 8(17), 5113–5125, 2008.

Travis, K. R. et al.: Why do models overestimate surface ozone in the Southeast United States? Atmos. Chem. Phys., 16(21), 13561–13577, 2016.

Turner, M. C. et al.: Long-Term Ozone Exposure and Mortality in a Large Prospective Study, Am. J. Respir. Crit. Care Med., 193(10), 1134–1142, doi:10.1164/rccm.201508-1633OC, 2016.

United Nations Economic Commission For Europe (UNECE): Hemispheric Transport of Air Pollution 2010, 2010.

Valin, L. C., A. R. Russell, R. C. Hudman, and R. C. Cohen: Effects of model resolution on the interpretation of satellite $NO_2$ observations, Atmos. Chem. Phys., 11(22), 11647–11655, 2011.

Venkataraman, C. et al.: Source influence on emission pathways and ambient PM2.5 pollution over India (2015–2050), Atmos. Chem. Phys., 18(11), 8017–8039, doi:10.5194/acp-18-8017-2018, 2018.

Venkataraman, C., G. Habib, D. Kadamba, M. Shrivastava, J. F. Leon, B. Crouzille, O. Boucher, and D. G. Streets: Emissions from open biomass burning in India: Integrating the inventory approach with high-resolution Moderate Resolution Imaging Spectroradiometer (MODIS) active-fire and land cover data, Global Biogeochem. Cy., 20(2), doi:10.1029/2005GB002547, 2006.

Verstraeten, W. W., J. L. Neu, J. E. Williams, K. W. Bowman, J. R. Worden, and K. F. Boersma: Rapid increases in tropospheric ozone production and export from China, Nat. Geosci., 8(9), 690–695, doi:10.1038/ngeo2493, 2015.

Wang, X. et al.: The role of chlorine in global tropospheric chemistry, Atmos. Chem. Phys., 19(6), 3981–4003, doi:10.5194/acp-19-3981-2019, 2019.

Yates, E. L., L. T. Iraci, D. Austerberry, R. B. Pierce, M. C. Roby, J. M. Tadić, M. Loewenstein, and W. Gore: Characterizing the impacts of vertical transport and photochemical ozone production on an exceedance area, Atmos. Environ., 109, 342–350, doi:10.1016/j.atmosenv.2014.09.002, 2015.

Young, P. J. et al.: Pre-industrial to end 21st century projections of tropospheric ozone from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., 13(4), 2063–2090, doi:10.5194/acp-13-2063-2013, 2013.

Zhang, L. et al.: Transpacific transport of ozone pollution and the effect of recent Asian emission increases on air quality in North America: an integrated analysis using satellite, aircraft, ozonesonde, and surface observations, Atmos. Chem. Phys., 8, 6117-6136, doi: 10.5194/acp-8-6117-2008.

Zhang, L., D. J. Jacob, M. Kopacz, D. K. Henze, K. Singh, and D. A. Jaffe: Intercontinental source attribution of ozone pollution at western U.S. sites using an adjoint method, Geophys. Res. Lett., 36(11), doi:10.1029/2009GL037950, 2009.

Zhang, L., D. J. Jacob, X. Yue, N. V. Downey, D. A. Wood, and D. Blewitt: Sources contributing to background surface ozone in the US Intermountain West, Atmos. Chem. Phys., 14(11), 5295–5309, doi:doi.org/10.5194/acp-14-5295-2014, 2014.

Zhang, Y., O. R. Cooper, A. Gaudel, A. M. Thompson, P. Nédélec, S.-Y. Ogino, and J. J. West: Tropospheric ozone change from 1980 to 2010 dominated by equatorward redistribution of emissions, Nat. Geosci., 9, 875–879, doi:10.1038/NGEO2827, 2016.

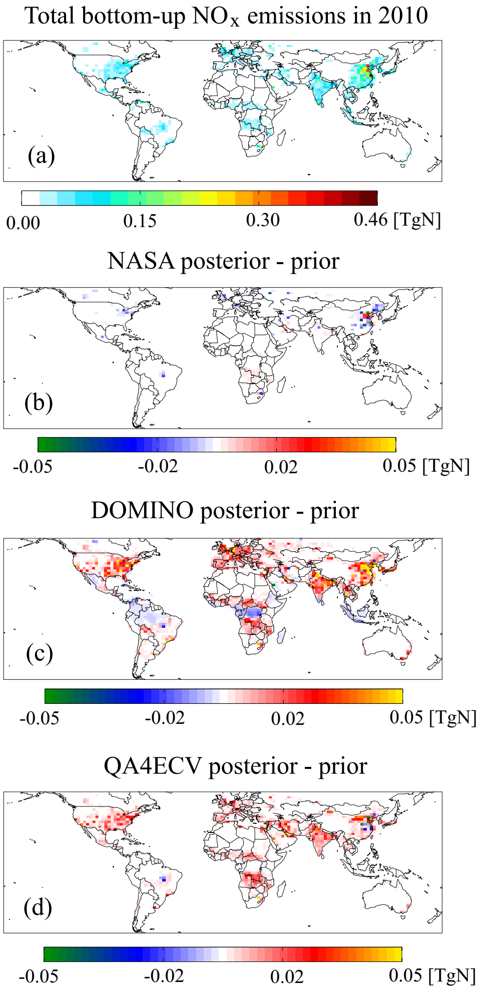
Zheng, F., T. Yu, T. Cheng, X. Gu, and H. Guo: Intercomparison of tropospheric nitrogen dioxide retrieved from Ozone Monitoring Instrument over China, Atmos. Pollut. Res., 5(4), 686–695, 2014.

Zhuang, J., D. J. Jacob, and S. D. Eastham: The importance of vertical resolution in the free troposphere for modeling intercontinental plumes, Atmos. Chem. Phys., 18(8), 6039–6055, doi:10.5194/acp-18-6039-2018, 2018.

Zoogman, P. et al.: Tropospheric emissions: Monitoring of pollution (TEMPO), J. Quant. Spectrosc. Ra. , 186, 17-39, doi:10.1016/j.jqrst.2016.05.008, 2017.

**Table 1. Total $NO_x$ emission (anthropogenic + natural) budgets in 2010 [Tg N $yr^{-1}$]**

| | Bottom-up | NASA posterior | DOMINO posterior | QA4ECV posterior |
|---|---|---|---|---|
| Global | 52.20 | 51.86 | 61.36 | 57.97 |
| China | 9.85 | 9.57 | 11.94 | 10.30 |
| US | 5.69 | 5.63 | 7.45 | 6.78 |
| India | 4.03 | 4.04 | 5.16 | 4.74 |
| Western Europe | 3.13 | 3.09 | 4.33 | 3.57 |


Total bottom-up $NO_x$ emissions in 2010

NASA posterior - prior

DOMINO posterior - prior

QA4ECV posterior - prior

**Figure 1. (a) Global total $NO_x$ emissions from the bottom-up inventory and the differences between 4D-Var posterior and bottom-up estimates constrained by (b) NASA standard product v3, (c) DOMINO product v2, and (d) QA4ECV product in 2010.**

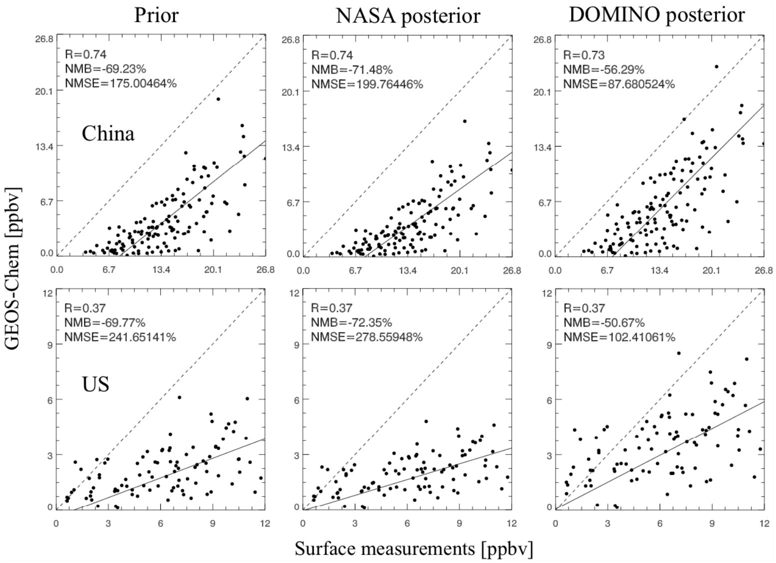


**Figure 2. Evaluation of annual mean surface $NO_2$ mixing ratios with measurements in China (top) and the US (bottom) in 2015.**

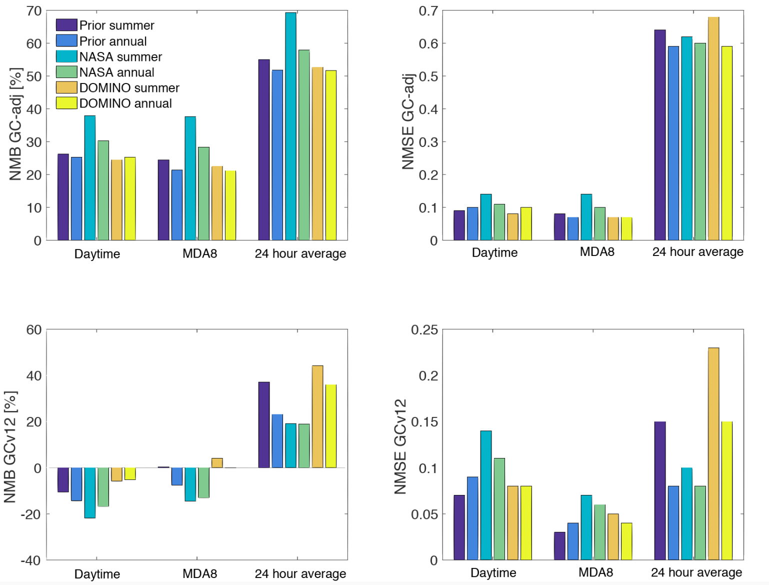
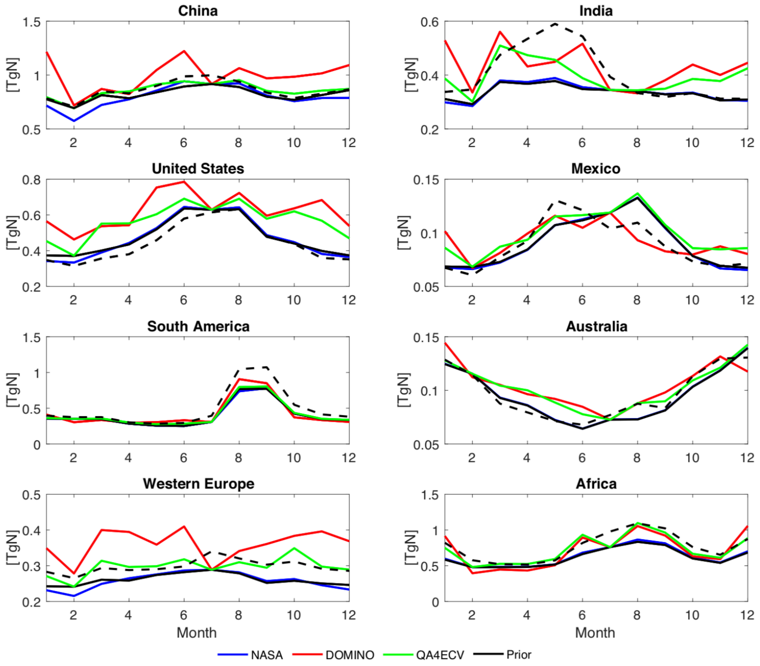
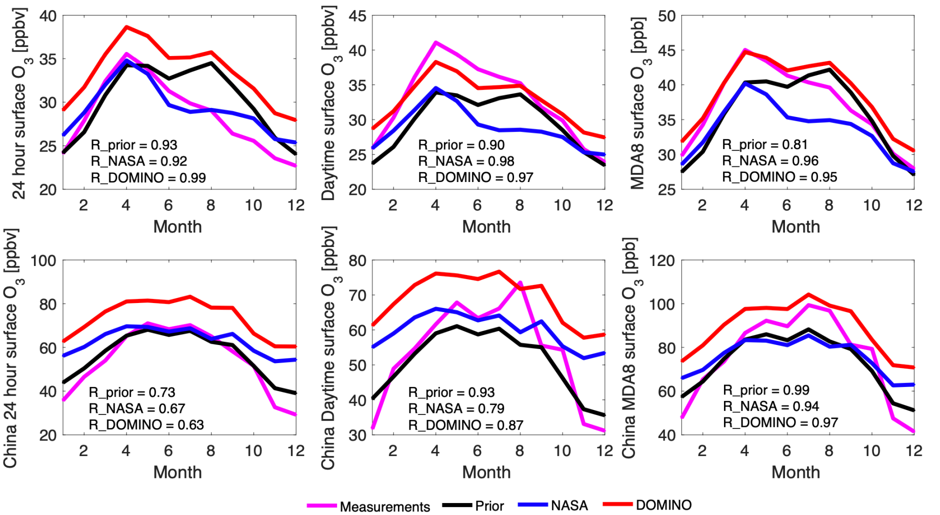


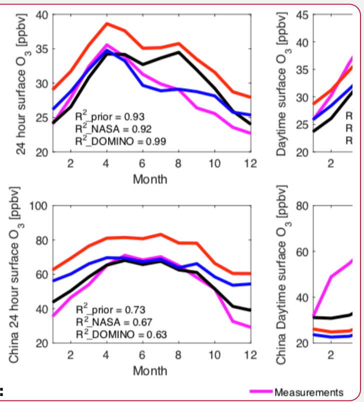
**Figure 3. NMB and NMSE of annual mean and NH summertime surface ozone concentrations when comparing all measurements from TOAR in 2010 with GC-adj (top) and GCv12 (bottom) simulations. The simulations are input with three sets of NO$_x$ emissions: CEDS bottom-up inventory (HTAP for GC-adj and CEDS for GCv12), posterior emissions constrained by the NASA product, and posterior emissions constrained by the DOMINO product.**



845 **Figure 4. Seasonal variations of total 4D-Var posterior $NO_x$ emissions in 2010. The black lines are prior emissions from bottom-up inventories (solid lines are from GC-adj, dashed lines are from GCv12). The blue lines are the emissions constrained by OMI $NO_2$ NASA product. The red lines are emissions constrained by OMI $NO_2$ DOMINO product. The green lines are emissions constrained by OMI $NO_2$ QA4ECV product.**



850 **Figure 5. Seasonality of surface ozone concentration at 2 meters in 2010 compared with TOAR (top) and in 2015 compared with CNEMC (bottom). Surface measurements are shown in magenta lines. Simulations are performed using GCv12 with $NO_x$ emissions from CEDS (black line), NASA posterior (blue line) and DOMINO posterior (red line).**



**Deleted:**

**Formatted:** English (US)

855

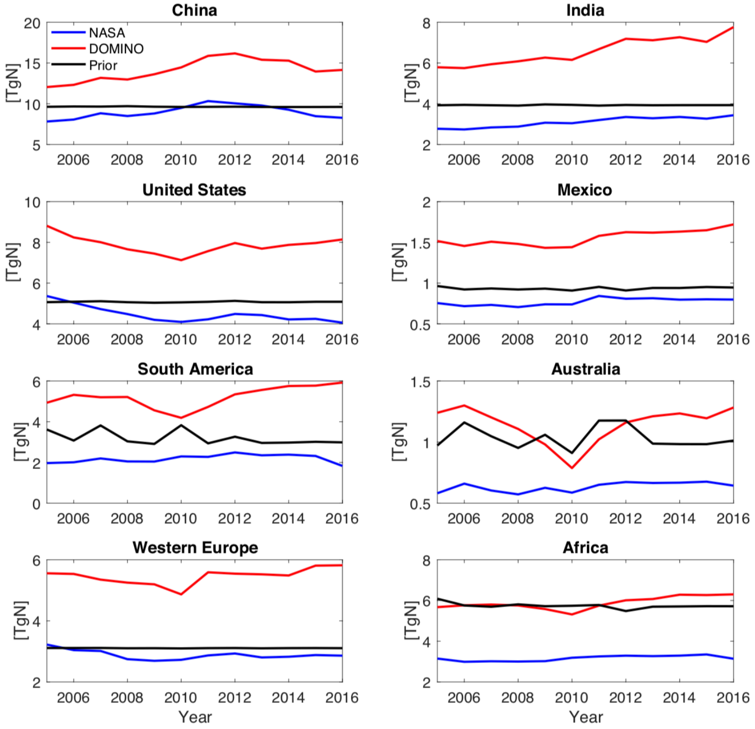


**Figure 6. Annual total posterior $NO_x$ emissions from 2005 to 2016. The black lines show prior total $NO_x$ emissions from bottom-up inventories, which use HTAP anthropogenic emissions in 2010 for all years. The blue lines represent the emissions constrained by the OMI $NO_2$ NASA product. The red lines represent emissions constrained by the OMI $NO_2$ DOMINO product.**

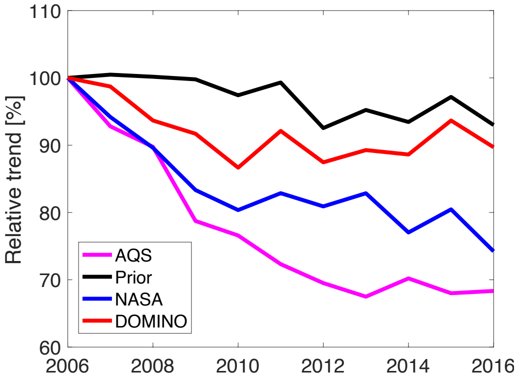


**Figure 7. The trend of annual mean surface $NO_2$ concentrations over the US from 2006 to 2016, expressed as a percent of the 2006 values. Surface measurements are from EPA AQS sites (magenta line). GEOS-Chem simulations are performed using prior emissions (black line) with constant anthropogenic emissions throughout the years, posterior $NO_x$ emissions constrained by NASA product (blue line), and posterior $NO_x$ emissions constrained by DOMINO product (red line).**

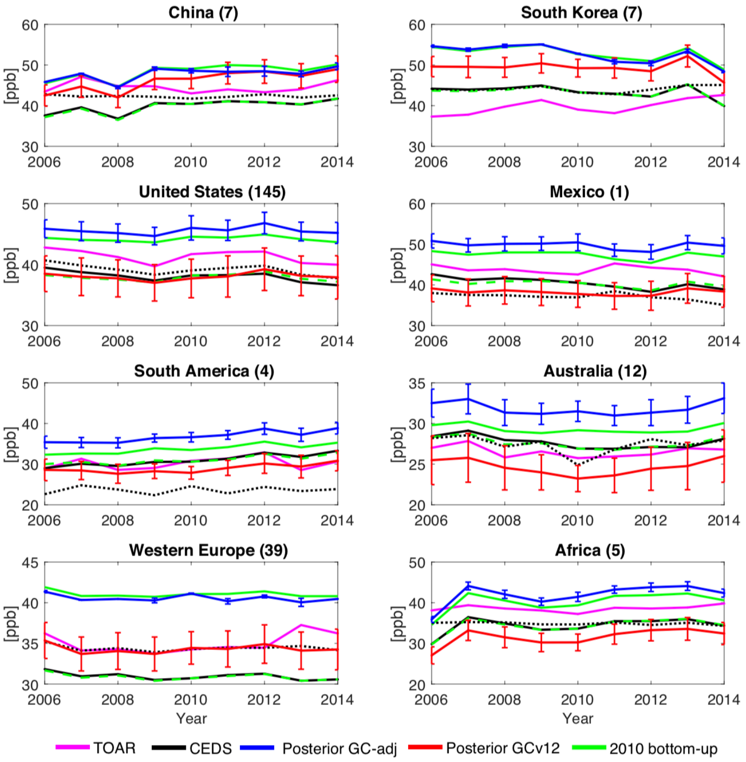


**Figure 8. The trends of regional mean annual MDA8 ozone concentrations from 2006 to 2014. Surface measurements are from the TOAR database (magenta line). Only sites that have continuous measurements throughout the 9 years are included. The numbers in the parenthesis are the number of 2° × 2.5° grid cells that include monitoring sites in each region. The black dotted lines show national mean of surface ozone from GCv12 simulations using the CEDS inventory. The other lines are simulations from GC-adj and GCv12 averaged over the 2° × 2.5° grid cells that include monitoring sites. Black lines show ozone simulations using the bottom-up $NO_x$ emissions from CEDS in each corresponding year. Green lines show ozone simulations using 2010 bottom-up $NO_x$ emissions for all years (HTAP 2010 for GC-adj shown in solid lines, CEDS 2010 for GCv12 shown in dashed lines). The vertical bars represent the spread of simulated surface ozone concentrations using the NASA and the DOMINO posterior $NO_x$ emissions.**

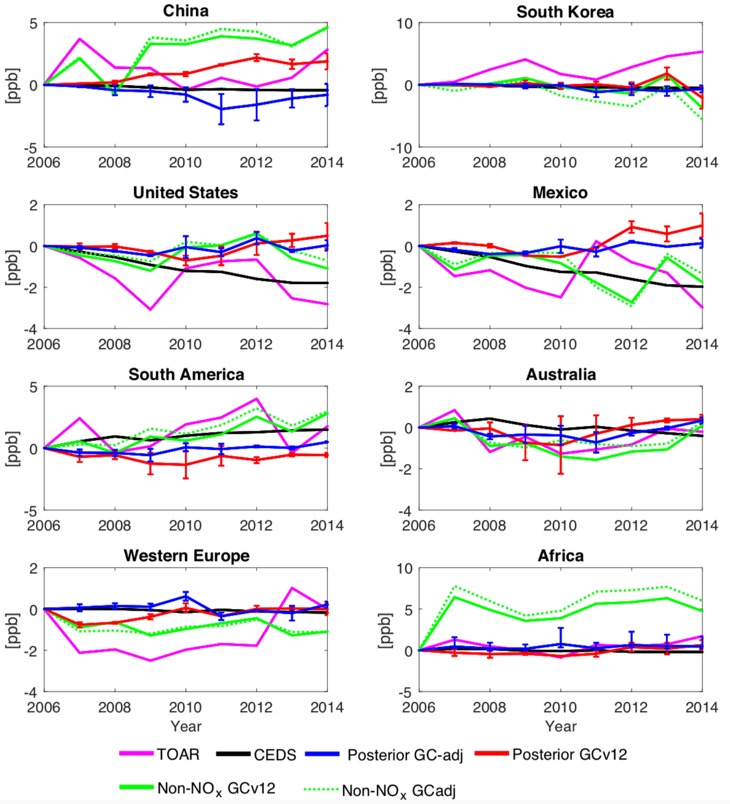
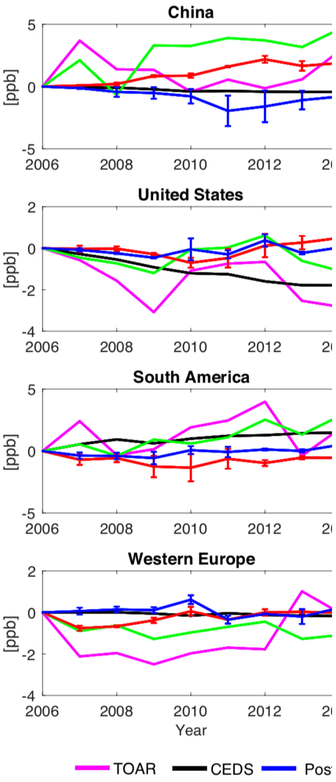


**Figure 9.** Changes of regional mean annual MDA8 ozone concentrations compared to 2006 from TOAR measurements (magenta line), due to changes in bottom-up NO$_x$ emissions (black), due to changes in top-down NO$_x$ emissions (blue lines for simulations from GC-adj and red lines for simulations from GCv12), and due to changes in meteorology and non-NO$_x$ emissions (green lines). Only sites that have continuous measurements throughout the 9 years are included. The vertical bars represent the spread of changes from simulations using the NASA and the DOMINO posterior NO$_x$ emissions. The impact of meteorology and natural sources are removed from black, blue and red lines by subtracting simulations using 2010 bottom-up anthropogenic emissions for all years from simulations that use bottom-up NO$_x$ emissions corresponding to each year.

**Deleted:**



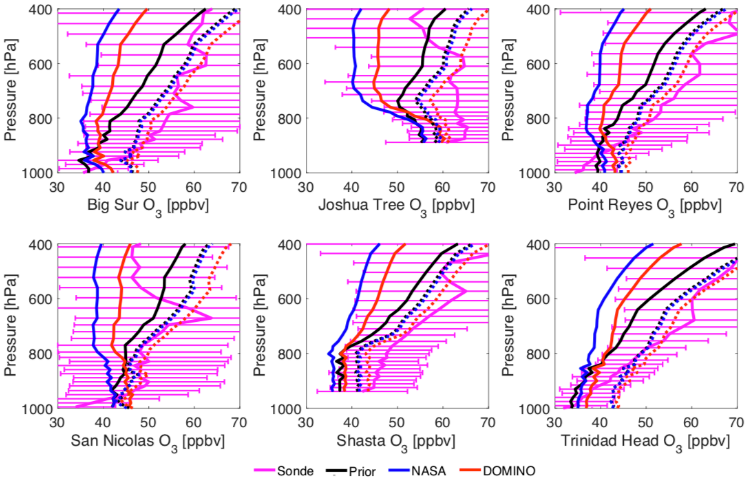


**Figure 10. Ozone vertical profiles averaged over May and June of 2010 from 6 ozonesonde measurement sites from the IONS-2010 field experiment in California. The six sites are over remote regions and are used to evaluate the intercontinental transport of ozone. Solid black (prior), blue (NASA posterior) and red (DOMINO posterior) lines are from the GCv12 simulations (prior anthropogenic emission from CEDS), whereas dashed lines are from the GC-adj simulations (prior anthropogenic emission from HTAP). The horizontal bars show the standard deviations of the measurements at each vertical layer.**

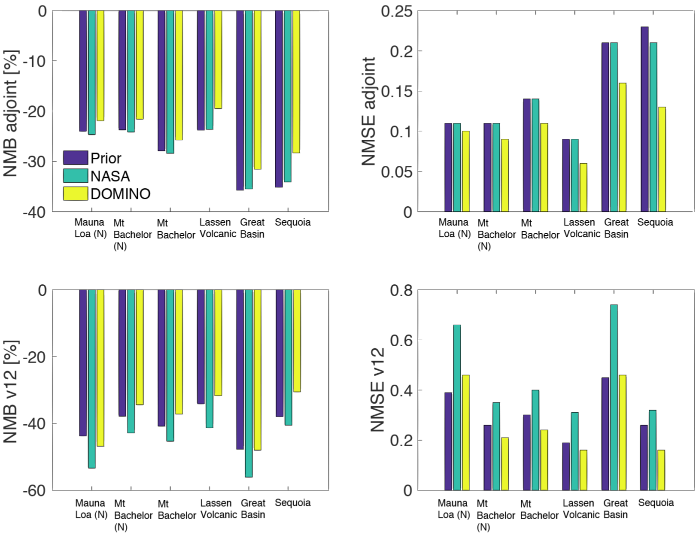


**Figure 11. NMSE and NMB of GC-adj (top) and GCv12 (bottom) ozone simulations in 2010 -2014 evaluated with surface measurements at remote sites. Three sets of $NO_x$ emissions, i.e., bottom-up inventory (HTAP for GC-adj, CEDS for GCv12), DOMINO posterior, and NASA posterior, are input in each model.**