We would like to thank both reviewers for the helpful suggestions and questions and have made changes to the manuscript to reflect the suggestions made. Individual comments from the reviewer are bolded below with our response in italics. The comments and responses are followed by the revised manuscript with the tracked changes highlighted.

Reviewer 1:

The paper studies the impact of typical (relatively coarse) spatial resolutions of past, present and planned satellite missions on tropospheric NO₂ retrievals over areas characterized by a strong spatiotemporal variability in the NO₂ field. High resolution airborne GEOTASO NO₂ VCDs as well as TEMPO, TROPOMI and OMI VCDs, simulated from the GEOTASO observations, are compared with coincident observations at 10 PANDORA sites at the western shore of Lake Michigan and over the Los Angeles Basin. By combination of the different data sets, the work provides an interesting insight into the spatiotemporal tropospheric NO₂ variability over polluted areas. The impact of mismatched spatial representation has been quantified. The results and discussions are valuable in the assessment of validation strategies for the future generation of air pollution satellites. The scientific content of the paper fits well within the scope of AMT, and the manuscript is well-written and generally well-structured. However, some revisions (detailed below) should need to be conducted in the paper before publication.

General comments

The first two sections are missing some essential information that are valuable to the reader to better interpret the results and the geophysical parameters. I suggest the following:

-Please add a "Campaign" section. Some information about the campaigns is scattered in the manuscript, but a clear campaign section shortly discussing the geophysical sites, number of flights, time and duration of flights, SZA change during flights, environmental conditions, e.g. cloud fraction, etc. would improve interpretation. Information on the exact dates are for example provided for the first time on p. 5, L.9 in the context of a discussion on BRDF derived albedo.

We added a campaign section to isolate information about these campaigns and the details about the measurements taken in each region. We also added a table that summarizes details from each flight in terms of SZA, areas mapped, pollution scales, and cloud conditions.

-Section 2.1 is unbalanced when compared to 2.2 and 2.3 and not structured well. Maybe make a separation between instrument and retrieval description. I also suggest to move parts to campaign section (e.g. P.4,L.13 to L.18).

We moved the relevant information to the campaign section and reorganized Section 2 through the use of additional headers for separating subsections. We agree that the sections between OMI, Pandora, and GeoTASO are not equally weighted, but did not see a way to shorten the GeoTASO section as the retrieval is not yet standard whereas the Pandora and OMI products are more standardized and better characterized through the literature.

However, we moved the GeoTASO section after the Pandora and OMI and campaign sections, which we think helps address this comment by providing a smoother transition between the retrieval descriptions and the results.

-Discussion of rasters on p. 5, L.33 should better be moved to another section (new sub-section under section 3 "Results") as it shows actual results, PANDORA locations, etc. Moreover some details are lacking again, e.g. "morning flight", is this at 6 AM or 10 AM or…? This is important for interpretation of for example traffic plumes.
This figure and text aid discussion about the retrieval sensitivity and uncertainty, which we feel is needed in the data section. This makes the data section longer, which isn't ideal, but through the reorganization of section 2 (see previous comment) we think the intent of the comment has been addressed by improving the flow. We also added times to the captions of each of the rasters. Both flights are representative of mid-morning in their respective local times.

-Introduction: please provide a more detailed overview of currently existing UV-VIS mapping instruments for completeness of the literature overview.
We added a brief literature review of recent work using other mapping spectrometers in the introduction.

P.2, L.35: Scanning at different azimuth angles can improve the representativeness of the ground-based data when compared to satellite retrievals. Even though the focus is on direct-sun observations in this work, the potential of multi-azimuth scanning to cope with the representativeness problem should be discussed in the introduction and/or conclusion. Especially as you mention "Best practices for satellite validation strategies" in the abstract.

We agree that scanning at different azimuth angles adds a unique dimension to a future validation studies and should be explored in future work. We have added some discussion in section 3.3 in the context of a literature review of MAX-DOAS comparisons to OMI. Because the direct-sun NO_2 retrieval from Pandora is what is currently available as a standard product, we emphasize in the revised manuscript that these conclusions are drawn from comparisons to direct-sun retrievals.

P.4, L.3: Please explain why this small window for NO2 retrievals is selected and not a window which is better comparable with OMI, TROPOMI NO2 retrievals. Moreover, in Nowlan et al. (2016) the fitting window 420-465 nm was used. Please properly refer to where the DOAS retrieval settings can be found or provide them in the manuscript.
We apologize as there was a typographical error in the manuscript. The DOAS window spans from 425-460 nm, not 435-460 nm. We fixed this in this revision. This window is similar to Nowlan et al. (2016) and the same as Lamsal et al. (2017). We also added more details on the DOAS retrieval within section 2.4.1 in the revised manuscript.

It would be interesting for Section 3.2 to differentiate between PANDORA stations with a heterogeneous and rather homogeneous NO2 distribution around the station (semi-background stations). This could be done based on the spatial distribution around the station observed by GEOTASO. The impact of the mismatched spatial representation, as reported in section 3.2, is expected to decrease in case of a more homogeneous distribution. An effort in this direction is done by differentiating based on the magnitude of PANDORA TropVCs and assuming that high NO2 values can be associated with localized features and thus strong heterogeneity (P.17, L.11). This is true, but low PANDORA TropVCs do not necessarily mean that the NO2 field is semi-background and that there are no fine-scale plumes around the site. It could be depending on the viewing geometry of the direct-sun measurement which is missing a plume or plumes present around the station.

You are correct that low TropVCs do not necessarily reflect a homogeneous environment. The difference between individual coincident GeoTASO and Pandora measurements may be larger than the uncertainty of both measurements due to spatial representation mismatches. In fact, none of the Pandora sites are 100% homogeneous in all cases based on looking at the subpixel variability around the sites from the airborne spectrometer during these campaigns. Thank you for this comment as it sparked additional discussion of this section as well as the conclusions. Please see the revised manuscript for these changes.

Conclusion: Related to the mismatched spatial representation reported in 3.2, please provide as well suggestions to solve this for future operational validation of satellite data such as TROPOMI and TEMPO, based on ground-based stations. As airborne data is collected on campaign basis, it will not always be available. Do you consider more viewing angles (thus not only direct-sun observations) to add as an additional constraint? Maybe using AQ model data around the stations, providing knowledge on the emission sources and direction of the plumes? This could help to assess if the viewing direction is hitting a (localized) plume or not.

We did not consider more viewing angles as we do not have measurements from Pandora to consider this from these campaigns and therefore cannot make conclusions about how this information would add to the results. However, the focus here is on the now-standard Pandora direct-sun measurements that can be accessed in near-real time for any operating instrument. Overall, these results show that at a TROPOMI spatial scale or finer, direct-sun observations are highly correlated with satellite-like products even in polluted environments (that can still be subject to variability).

To specifically address the comment about adding recommendations for future validation work with direct-sun columns from Pandora, we have added additional suggestions to the final paragraph of the Conclusions section. Please see the revised manuscript for these changes.

Minor comments
P.2, L.44: Replace "the" by "a" unique perspective. Mobile-DOAS measurements for example can provide as well a unique insight in the spatial variability around a ground-based station.

Change made as suggested.

P.4, L.9: The overlapping GEOTASO retrievals also allow an interesting way to compare coinciding VCDs and assess the GEOTASO product quality (even if we know that the NO2 field is changing). Has this been done?
There are a few pixels that overlap between adjacent flight lines, and we have done some preliminary analysis but have not yet compared the overlapping pixels quantitatively. It does not appear to be a large effect in the data reported here. It could be an interesting study in the future as an interesting way to evaluate the impact of BRDF on the retrieval.

P.4, L.25: I have a bit hard time to interpret this. You are discussing larger retrieval uncertainty due to less signal (low albedo, large SZA)? The multi-linear regression is applied on which data? Details are lacking on time of flights (or SZA) to properly interpret this. Moreover, I assume flights took place with SZA smaller than 60°, so the uncertainty related to SZA should be smaller as reported?
We reworded this paragraph to clarify to which data the statistics are applied. Additional data about the flight times/SZAs have also been added in the campaign section for more interpretation.

P.5, L.26: "simplified from.." not sure if it is needed to cite this reference as it is a commonly used equation and appears in earlier publications.
Removed (simplified from……)

P.5, L.30: Please provide some more info on the reference used. Do you average spectra over a certain period to reduce noise? Do you use a different reference per spatial pixel in order to reduce striping effects? Is the instrument stable enough to use a single reference for the whole campaign period?

We added information (dates/areas/sza/time) on the references and noted that these were clean/homogeneous regions. You are correct that we use a difference reference for each spatial pixel to avoid striping. The instrument was stable throughout the 2017 flights and no changes were made to the instrument during the campaign, which allowed us to use just one reference for each region and therefore reduce bias associated with having a suite of different references.

P.15, L.35: To improve readability, please repeat again explicitly which exact data sets (which days) you are comparing here.
We reworded the opening sentences to that paragraph to clarify which datasets are used and how they are compared to the upscaled version.

P.15, L.30: Could you see any consistency with traffic peak times (or diurnal photochemistry) when looking at TropVC values in the 4 grids acquired on the same day?
There is a diurnal pattern that occurs with the 4 rasters on the same day in this case. We saw the peak in NO_2 early-mid morning and it decreased through the day. This seems to match

traffic patterns in the area, but in-depth analysis is pending updated emission inventories being developed for the Chicago region. We have attempted to clarify the wording of this sentence.

P.15, L.47: True for ground-based vs satellite retrievals. Maybe highlight here again the advantage of airborne measurements, able to fully cover satellite pixels at high resolution.
We added text to highlight the advantage of the airborne measurements in these types of experiments.

Figure 9 and 11: Please provide the fit parameters and correlation as well in the plot or legend.
We added the statistics to Figure 11, however felt they could not be legibly added to Figure 9 so instead added to the Figure 9 capture a referral to Table 3, which presents the statistics.

Figure 11: Pandora min and max data during the overpass +- 5 minutes can sometimes show large variations and maybe too large to be fully attributed to temporal variations. Can you shortly explain this? Outliers? I assume you lack good statistics to use 2 x st.dev. or 10-90 percentile for the whiskers.
Your assumption is correct. In a 10-minute window, there is a maximum of 7 measurements from Pandora, which is not sufficient for reporting statistics other than max/min. We have explored these max/min values at times when they have a very large range in this short time window, and the variation at these sites (most often the Schiller Park location) appears to be real. We have not found individual outliers suggestive of some sort of instrument or retrieval error, but rather large swings in the magnitude over multiple measurements. A couple examples of this are shown in Figures 4 and 6.

In general for Section 3.3: Please compare your results as well with other studies that have done efforts to compare OMI with ground-based measurements, e.g. with MAX-DOAS and assess if your findings are consistent.
We extended the discussion in Section 3.3 to include previous studies comparing measurements with OMI. Please refer to the revised text within section 3.3.

Section 3.3: Did you make use of the OMI averaging kernels to smooth the PANDORA VCDs in order to take into account differences in sensitivity?
We did not. As the Pandora VCDs are not dependent on their own a priori assumptions (equally sensitive vertically and AMF is geometric), this is not necessary as it would be when comparing to measurements requiring their own a priori and sensitivity profiles (i.e., MAX-DOAS or model comparisons).

Technical corrections
P.5, L.25: Remove "the"
P.7, L.36: Formulation is confusing. Maybe mention that these are the DSCD precision and accuracy or provide a typical value for the AMF, e.g. "assuming an AMF…."
P.18, L.29: Please change "city-to-regional spatial scale" to "regional spatial scale"

P.18, L.31: Please remove "very". A priori profiles and surface reflectances can be retrieved at much higher resolutions.

All technical corrections were made as suggested.

Reviewer 2:

Review of Judd et al. -- Evaluating the impact of spatial resolution on tropospheric NO2 column comparisons within urban areas using high-resolution airborne data

The authors investigate the impact of spatial variability on correlative studies for the validation of satellite trace gas products with ground-based instruments. High resolution airborne imaging DOAS measurements from GeoTASO, ground-based Pandora, and two OMI satellite products are used. The paper is generally well written and of significance for the validation of satellite trace gas retrievals. I therefore recommend publication in AMT after some minor revisions.

General comments

The introduction should contain an overview of existing airborne imaging DOAS systems

We have added a short literature review of other mapping spectrometers in the introduction.

Information about the campaign is scattered in the manuscript. A solution could be a campaign sections, with a description of the target sites (urban/rural, # of inhabitants, industrial emitters...), as well as a description of the measurement conditions (Date, time of day, SZA, AOD, meteorology...) maybe as a table... Here you should also give an overview of the flights presented in this study to help the reader

We have added a campaign section at the beginning of section 2. This includes an overview of the purpose of each campaign, number of flights, total hours, areas measured as well as some discussion on typical meteorology and SZA and a table that summarizes some details about each flight.

I could not find any information about the DOAS fit settings used (except the fit window). Please provide that information (cross-sections, I0, Ring, ...), e.g. in a table.

We have added information about the DOAS fit under the Airborne NO$_2$ retrieval section (now Section 2.4.1).

You often refer to differences in spatial resolution of the a priori inputs. It would be nice if you could provide the spatial (and temporal?) resolution of the SP and BEHR products.

The spatial resolution of the a priori input to the AMF calculation is finer for BEHR than SP. Discussion about the specific differences on surface reflectivity, terrain pressure, and NO2 profile are added to the OMI section.

You often state that the Pandora measurements are representative up to a certain pollution scale. In my opinion this statement is not correct. The representativeness depends on the spatio-temporal variability of NO2 at the Pandora location. You use an NO2 threshold to filter out data with large variability, but the magnitude of NO2 itself is not an indicator for the representativeness. I think you

should amend the manuscript to reflect the differences between the physical reasons (variability) and the methodology (filtering by threshold).
We agree with this statement and in hindsight believe the original statement was worded poorly (and backwards, as it is the spatial scale of the satellite retrieval that may not be representative of the features Pandora is observing). We amended the text to reflect these points and attempted to more clearly express the purpose of applying the thresholds. Please see the appropriate areas of the abstract, Sect 3.2, and conclusions in the revised submission to see how these points were edited to emphasize these conclusions.

Detailed comments & technical corrections: Page	Line	Comment
2	26ff	**You mention: "development of […] instruments" but you then only write about GeoTASO. I think you also had GCAS in mind. I suggest to explicitly mention it. Here you could also refer to other instrument previously used.** *Added GCAS as well since this is also a NASA supported instrument. There is also now a short literature review on other airborne spectrometer research efforts as well.*
2	35	**"… such as NO2." NO2 is not a product, but a chemical species. Suggestion: NO2 tropospheric vertical column densities.** *Added tropospheric vertical column densities as suggested.*
3	21	**What is the field of view in degrees?** *It is 45 degrees, which is stated within the section.*
4	3	**Is the spectral resolution constant over the spatial dimension? If not, how does it vary?** *The spectral resolution does not vary over the spatial dimension (within 0.01nm).*
4	7	**How many spectra are co-added for the 250m (or what is the speed of the aircraft)** *The across track dimension is separated into 33 across track positions with ~30 images per bin. For the along track dimension, the code considers the median aircraft speed at altitude (typically about 100 m/s) and then calculates how many along track spectra are needed to get closest to 250m. Typically, it comes results in around 300 spectra coadded to get 250 x 250 m.*

_____

Evaluating the impact of spatial resolution on tropospheric NO₂ column comparisons within urban areas using high-resolution airborne data

Laura M. Judd[1,2], Jassim A. Al-Saadi[1], Scott J. Janz[3], Matthew G. Kowalewski[3,4], R. Bradley Pierce[5], James J. Szykman[6], Lukas C. Valin[6], Robert Swap[3], Alexander Cede[7], Moritz Mueller[7,8], Martin Tiefengraber[7,8], Nader Abuhassan[3,9], David Williams[6]

[1]NASA Langley Research Center, Hampton, VA, 23681, United States
[2]NASA Postdoctoral Program, Hampton, VA, 23681, United States
[3]NASA Goddard Space Flight Center, Greenbelt, MD, 20771, United States
[4]Universities Space Research Association, Columbia, MD, 21046, United States
[5]University of Wisconsin-Madison Space Science and Engineering Center, Madison, WI, 53706, United States
[6]United States Environmental Protection Agency Office of Research and Development, Triangle Research Park, NC, 27709, United States
[7]LuftBlick, Kreith, Austria
[8]Department of Atmospheric and Cryospheric Science, University of Innsbruck, Innsbruck, Austria
[9]Joint Center for Earth Systems Technology, University of Maryland-Baltimore County, Baltimore, MD, 21228, United States

Correspondence to: Laura M. Judd (laura.m.judd@nasa.gov)

Abstract. NASA deployed the GeoTASO airborne UV-Visible spectrometer in May-June 2017 to produce high resolution (approximately 250×250 m) gapless NO₂ datasets over the western shore of Lake Michigan and over the Los Angeles Basin. The results collected show that the airborne tropospheric vertical column retrievals compare well with ground-based Pandora spectrometer column NO₂ observations (r^2=0.91 and slope of 1.03). Apparent disagreements between the two measurements can be sensitive to the coincidence criteria and are often associated with large local variability, including rapid temporal changes and spatial heterogeneity that may be observed differently by the sunward viewing Pandora observations. The gapless mapping strategy executed during the 2017 GeoTASO flights provides data suitable for averaging to coarser areal resolutions to simulate satellite retrievals. As simulated satellite pixel area increases to values typical of TEMPO, TROPOMI, and OMI, the agreement with Pandora measurements degraded, particularly for the most polluted columns as localized large pollution enhancements observed by Pandora and GeoTASO are spatially averaged with nearby less-polluted locations within the larger area representative of the satellite spatial resolution (aircraft-to-Pandora slope: TEMPO scale=0.88; TROPOMI scale=0.77; OMI scale=0.57). In these two regions, Pandora and TEMPO or TROPOMI have the potential to compare well at least up to pollution scales of 30×10^{15} molecules cm⁻². Two publicly available OMI tropospheric NO₂ retrievals are both found to be biased low with respect to these Pandora observations. However, the agreement improves when higher resolution a priori inputs are used for the tropospheric air mass factor calculation (NASA V3 Standard Product slope = 0.18 and Berkeley High Resolution Product slope=0.30). Overall, this work explores best practices for satellite validation strategies with Pandora direct-sun observations by showing the sensitivity to product spatial resolution and demonstrating how the high spatial resolution NO₂ data retrieved from airborne spectrometers, such as GeoTASO, can be used with high temporal resolution ground-based column observations to evaluate the influence of spatial heterogeneity on validation results.

1 Introduction

Nitrogen oxides (NOₓ: NO + NO₂) are primarily emitted via fossil fuel combustion, soil microbial processes, biomass burning and lightning. NO₂ is an important precursor of ozone and particulate matter, making it one of the six criteria air

Deleted: an
Deleted: /
Deleted: , GeoTASO,
Deleted:),
Deleted: Results
Deleted: also
Deleted: is
Deleted: polluted plumes
Deleted: over
Deleted: areas
Deleted: This behavior suggests that satellite products are representative of individual Pandora observations up to a certain pollution scale that depends on satellite spatial resolution.
Deleted: (NASA V3 Standard Product slope = 0.18 and Berkeley High Resolution Product slope=0.30).
Deleted: .
Deleted: demonstrates
Deleted: surface

pollutants monitored by the United States Environmental Protection Agency (EPA: https://www.epa.gov/criteria-air-pollutants). Unlike less reactive trace gases having atmospheric lifetimes of days or longer, the atmospheric lifetime of NO_x is reported to be on the order of hours in the daytime polluted boundary layer (Liang et al. 1998; Beirle et al., 2011; Liu et al., 2016). This short lifetime along with large variations in emission rates from sources causes the spatial distribution of

5 NO_2 in polluted regions to be highly heterogeneous, making it difficult to characterize over urban areas without high spatiotemporal observations.

Since the 1990s, NO_2 column densities have been monitored globally from sun-synchronous satellite platforms utilizing the differential optical absorption spectroscopy (DOAS) methodology applied to earth-shine spectra in the visible-blue wavelengths. The nadir spatial resolution of these sensors has generally improved over time—beginning with the

10 Global Ozone Monitoring Experiment (GOME) at 40 x 320 km in 1995 (Burrows et al., 1999), dramatically refining a decade later with the Ozone Monitoring Instrument (OMI) at 13 x 24 km in 2004 (Levelt et al., 2006), and improving most recently to a sub-city spatial scale from the TROPOspheric Monitoring Instrument (TROPOMI) instrument aboard Sentinel-5P launched October 2017 with a nadir spatial resolution of 3.5 x 7 km (van Geffen et al., 2019).

OMI has been a prominent resource for understanding the global distribution of tropospheric NO_2 since its launch.

15 However, short-falls have been documented regarding its inability to capture the spatial variability within polluted regions (e.g., Valin et al., 2011a, Valin et al., 2011b; Broccardo et al., 2018). This inability is further hindered by the use of coarse a priori assumptions in the air mass factor (AMF) calculation for slant-to-vertical column conversion (Heckel et al., 2011; Russell et al., 2011; Goldberg et al., 2017). Though TROPOMI observes NO_2 at a spatial resolution an order of magnitude finer than OMI, its sun-synchronous orbit still limits lower- to mid-latitude observations to the early afternoon hours.

20 However, within the next decade three geostationary air quality monitoring missions will be launched that will be able to monitor daytime air quality hourly at spatial scales of less than 10 km. These missions include observations of North America with Tropospheric Emissions: Monitoring Pollution (TEMPO) (Zoogman et al., 2017), Asia with Geostationary Environment Monitoring Spectrometer (GEMS) (Kim et al., 2017), and Europe with Sentinel-4/Ultraviolet/Visible/Near-Infrared Instrument (UVN) (Ingmann et al., 2012). These geostationary measurements over industrialized regions of the

25 Northern Hemisphere along with ongoing global daily sun-synchronous measurements will be important contributors to the goal of creating an atmospheric composition global observing network (IGACO, 2004; CEOS, 2011).

To prepare for these planned geostationary air quality missions, NASA supported the development of airborne UltraViolet-VISible (UV-VIS) mapping instruments, (Geostationary Trace Gas and Aerosol Sensor Optimization: GeoTASO and GEO-CAPE Airborne Simulator: GCAS), to help determine the satellite instrument requirements for measurements

30 relevant to air quality and to facilitate retrieval algorithm development at fine spatial resolutions at all times of day (Leitch et al., 2014; Kowalewski and Janz, 2014; Nowlan et al., 2016; Lamsal et al., 2017; Nowlan et al., 2018). These instruments have the capability of retrieving NO_2 at sub-kilometer spatial resolutions which can be useful in assessing trace gas heterogeneity at spatial scales finer than those of current and planned space-based retrievals. Similar instruments have also been developed by other countries. In the last decade, airborne spectrometers have mapped high resolution NO_2 in parts of

35 Europe (Popp et al., 2012; Schönhardt et al 2015; Lawrence et al., 2015; Meier et al., 2017; Tack et al., 2017; Tack et al 2019), the United States (Nowlan et al., 2016; Lamsal et al., 2017; Nowlan et al., 2018; Judd et al., 2018), Asia (Judd et al., 2018), and Africa (Broccardo et al., 2018). Results from these efforts have illustrated the capability of airborne spectrometers to observe the impact of emission sources and meteorology on the spatial distribution of NO_2 (Popp et al., 2012, Schönhardt et al., 2015; Judd et al., 2018, and Tack et al., 2019) and have shown utility for evaluating NOx emissions

40 (Schönhardt et al 2015; Souri et al 2018) as these high resolution measurements can resolve detailed NO_2 spatial patterns that satellites, at their current spatial resolutions, cannot (Broccardo et al., 2018; Lamsal et al., 2017; Judd et al., 2018).

Measurements from these airborne UV-VIS mapping instruments can provide a transfer standard between the space-based sensor footprint and ground-based column measurements used for validation of the satellite trace gas products, such as NO_2 tropospheric vertical columns. The direct-sun DOAS technique used to retrieve NO_2 from ground-based

45 Pandora spectrometer measurements has been shown to be highly precise and accurate due to little uncertainty in the path light travels through the atmosphere (i.e., the air mass factor) at solar zenith angles less than 80 degrees (Herman et al., 2009). Thus, Pandora instruments that operate in direct-sun mode are a strong candidate for providing a validation standard for trace gas retrievals from geostationary sensors like TEMPO as well as low Earth orbiting sensors like TROPOMI. NASA and EPA have been working toward creating long-term measurement sites across the United States to prepare for and

50 aid validation of the satellite products and their usage in air quality management activities (EPA, 2019a). Developments of

2

similar validation site activities are also underway in East Asia and Europe. However, in heterogeneously polluted areas such as cities, the relatively local measurements at such sites may not necessarily be representative of the spatial scales observed by space-based platforms. High resolution airborne mapping observations provide a unique perspective to assess the impact of spatial scale mismatches between satellite observations and local validation site measurements.

5   This work compares high spatial resolution NO_2 vertical columns retrieved from GeoTASO to those measured from small networks of Pandora spectrometers operating in direct-sun mode installed in two regions: the western shore of Lake Michigan and the Los Angeles (LA) Basin. The high spatial resolution GeoTASO data are then upscaled to spatial resolutions typical of past, present, and planned space-based sensors to demonstrate sensitivity of these comparisons to satellite pixel size in an idealized framework. Finally, these results are compared to two publicly available OMI NO_2
10 retrievals to provide real-world context. Overall, this work shows the sensitivity of satellite product validation strategies to spatial resolution and begins to demonstrate how the high spatial resolution NO_2 data retrieved from airborne mapping observations can be used with planned high temporal resolution ground-based column observations to evaluate the influence of spatiotemporal heterogeneity on satellite-based trace gas product validation.

2 Data

15 ### 2.1 Campaign Overview

     In 2017, one of NASA's airborne UV-VIS mapping instruments, GeoTASO, was flown aboard the NASA Langley Research Center UC-12b aircraft as part of the Lake Michigan Ozone Study (LMOS; https://www-air.larc.nasa.gov/missions/lmos/index.html) and during the Student Airborne Research Program (SARP; https://airbornescience.nasa.gov/content/Student_Airborne_Research_Program). Table 1 summarizes all GeoTASO flights
20 that occurred from May 22nd-June 27th, 2017. GeoTASO during LMOS was used to characterize the influence of NO_2 emissions and transport on ozone exceedances along the western shore of Lake Michigan. During LMOS, GeoTASO flew on 21 research fights totaling over 90 hours. The left map in Figure 1 shows the areas mapped by GeoTASO during the month-long LMOS campaign in grey. The entire LMOS domain cannot be mapped during a single flight, therefore individual flight plans focused on sub-portions of the domain where air quality and meteorological forecasts suggested that science goals
25 could be met for each flight. Eight of the research flights focused on the city of Chicago, whereas the rest were north of the city along the western shore of Lake Michigan. Each flight lasted approximately four hours and up to two flights were flown per day (one morning and one afternoon). Flights were optimized for times where clear skies were expected and any cloudy scenes that resulted are filtered out for this analysis. The median solar zenith angle for all flights was 36° with a minimum/maximum of 19°/70°.
30   After LMOS, five research flights were conducted on June 26th and 27th, 2017, totaling 15 hours, in the Los Angeles (LA) Basin during SARP. In 2017, the LA Basin had the highest 8 h and 1 h ozone concentrations in the United States, as well as the highest annual mean of NO_2 in the United States (EPA, 2019b). In the LA Basin, a single flight plan was developed to maximize coincidences with Pandora spectrometers installed in the region and extended south to the industrial area near Long Beach, CA (Fig. 1 right). This plan was completed once per flight with up to three flights per day
35 (each flight lasting approximately three hours). Both flight days in the LA Basin were cloud-free. LA Basin flights spanned a larger range of solar zenith angles than LMOS with a median of 49° and a minimum/maximum of 10°/70°.
     One Cimel sunphotometer was operating in both study domains during flight days (https://aeronet.gsfc.nasa.gov/cgi-bin/bamgomas_interactive). At the Zion site during LMOS, AOD was generally below 0.1 on flight days with a peak reported of ~0.2 after cloud filtering. In the LA Basin, there were no data reported for the June
40 26th, 2017 flight day, but AODs were less than 0.1 on June 27th, 2017 and both days were similar meteorologically.
     The LMOS region and the LA Basin each had five Pandora instruments operating in direct-sun mode within the research areas during GeoTASO flight days (hexagons in Fig. 1). The number of valid coincidences for each flight are labeled in Table 1. All flight plans had planned overflights of multiple Pandora sites, but cloud cover or instrument issues may have resulted in a fewer (or zero) valid coincidences. Pandora site names and coordinates and observed pollution
45 ranges during coincidences are listed in Table 2. The locations of these instruments span a wide variety of pollution environments. Urban locations include all those located in the LA Basin and one near the Chicago O'Hare Airport (Schiller

3

Park). Pandora spectrometers located along the western shore of Lake Michigan north of Chicago during LMOS were the cleanest (Zion, Milwaukee, Grafton, Sheboygan).

Table 1: GeoTASO flight summary

Flight	Date	Time	Location	SZA Range (°)	Pollution Scale (95th percentile x10^15 molecules cm^-2)	Percent Cloudy Pixels	# Valid Pandora Coincidences
1	2017-05-22	12.4-16.3 UTC	Chicago, IL	29-70	12.6	1%	3
2	2017-05-22	17.7-21.2 UTC	Chicago, IL	21-47	3.6	29%	1
3	2017-05-27	16.2-19.9 UTC	Zion-to-Sheboygan	20-33	2.4	3%	7
4	2017-06-01	13.3-17.1 UTC	Chicago, IL	22-59	17.1	0%	2
5	2017-06-01	19.1-22.8 UTC	Zion-to-Sheboygan	26-64	6.8	0%	7
6	2017-06-02	19.1-23.0 UTC	Zion-to-Sheboygan	26-66	9.5	0%	4
7	2017-06-04	19.3-22.6 UTC	Chicago, IL	26-64	10.1	7%	3
8	2017-06-07	12.9-16.9 UTC	Zion Area	23-63	3.4	0%	4
9	2017-06-07	18.7-22.7 UTC	Zion Area	23-62	6.0	0%	3
10	2017-06-08	12.8-16.6 UTC	Zion Area	26-64	5.0	0%	3
11	2017-06-12	18.9-22.5 UTC	Zion Area	24-59	6.0	7%	1
12	2017-06-13	16.4-19.7 UTC	Chicago, IL	18-29	5.5	9%	1
13	2017-06-14	13.3-17.0 UTC	Chicago, IL	21-59	5.9	8%	3
14	2017-06-15	18.8-22.7 UTC	Sheboygan Area	22-47	2.1	3%	2
15	2017-06-15	18.7-22.7 UTC	Sheboygan Area	22-61	5.2	2%	9
16	2017-06-17	16.3-20.3 UTC	Sheboygan Area	20-36	0.1	24%	1
17	2017-06-18	13.3-17.3 UTC	Chicago, IL	21-60	3.2	14%	2
18	2017-06-19	13.4-17.1 UTC	Chicago, IL	22-58	8.3	25%	1
19	2017-06-21	13.2-17.1 UTC	Zion-to-Sheboygan	22-60	7.1	2%	1
20	2017-06-21	19.0-22.0 UTC	Zion-to-Sheboygan	24-53	5.7	16%	0
21	2017-06-26	15.1-17.4 UTC	LA Basin	35-62	25.7	0%	5
22	2017-06-26	22.7-1.5 UTC	LA Basin	38-70	17.3	0%	4
23	2017-06-27	14.7-17.2 UTC	LA Basin	37-67	29.8	0%	5
24	2017-06-27	18.7-21.1 UTC	LA Basin	10-20	17.0	0%	3
25	2017-06-27	22.7-1.2 UTC	LA Basin	38-70	22.0	0%	5

5

Figure 1: Maps of the LMOS and LA Basin regions showing the areas mapped by GeoTASO during summer 2017. The labeled hexagons indicate locations of Pandora spectrometers operating during the flight days.

5   **Table 2: Pandora site names, location, and TropVC Range during GeoTASO coincidences in the LMOS and LA Basin domains. These locations are also indicated by hexagons on the maps in Figure 1.**

Pandora Site	Latitude, Longitude (°)	TropVC Range during GeoTASO Coincidences x10^15 molecules cm^-2
LA Main Street	34.066, -118.227	7.4-28.5
Ontario	34.068, -117.526	7.5-30.1
UCLA	34.074, -118.441	5.0-24.0
Pico Rivera	34.010, -118.069	12.1-18.3
CalTech	34.136, -118.127	3.4-14.2
Sheboygan	43.746, -87.709	-0.9-4.3
Grafton	43.343, -87.920	-0.2-4.0
Zion	42.468, -87.810	1.5-8.0
Schiller Park	41.965, -87.876	3.5-45.1
Milwaukee	43.061, -87.914	4.1-5.9

2.2 Pandora

Pandora is a UV-VIS ground-based spectrometer used to retrieve trace gas column amounts (Herman et al., 2009; Herman et al., 2015). Although the instrument has the capability to make both direct-sun and all-sky radiance measurements from which trace gas amounts can be retrieved, only direct-sun retrievals were collected during these campaigns. Direct-sun NO_2 columns are reported to have high precision and accuracy and they are availability in near-real time as a standard product. The direct-solar beam is measured by Pandora via an optical head sensor attached to a solar tracker. The solar beam is carried to the UV-VIS spectrometer by a fiber optic cable attached to the head sensor. NO_2 is retrieved by applying a spectral fitting algorithm (Cede, 2017) using a near-noon reference spectrum, from which the NO_2 slant column amount is derived by a statistical calibration approach (Herman et al., 2009). A geometric AMF is used for the slant-to-vertical column conversion as the path length through the atmosphere is dominated by the direct solar beam rather than scattered light through the atmosphere (Cede, 2017). The precision and accuracy of the direct sun NO_2 vertical columns are reported as 2.67×10^{14} and 2.67×10^{15}/AMF molecules cm^{-2}, respectively, by Herman et al. (2009). In this work the Pandora retrievals are screened to exclude observations with vertical column error greater than 0.05 DU and normalized RMS greater than 0.005, to limit the retrieval uncertainty to approximately 10%.

2.3 Ozone Monitoring Instrument (OMI)

OMI is a space-based UV-VIS instrument launched in 2004 aboard the EOS-Aura satellite in a sun synchronous orbit with an equator crossing time in the early afternoon (Levelt et al., 2006). This work uses the NASA Standard Product (SP) Version 3 (Krotkov et al., 2017) and the Berkeley High Resolution (BEHR) Product (Russell et al., 2011; Laughner et al., 2018b; Laughner et al., 2019) to assess how their tropospheric column retrievals compare with measurements from Pandora spectrometers deployed during summer 2017 in the LMOS and the LA Basin domains.

The two vertical column products are based on the same slant columns (produced by the SP retrieval) but they differ in their a priori inputs for the tropospheric air mass factor calculation. As the name suggests, BEHR a priori have a higher spatial resolution than the NASA SP (Russell et al., 2011; Bucsela et al., 2013; Krotkov et al., 2017; Laughner et al., 2018a; Laughner et al., 2019). Both products use monthly NO_2 a priori profile assumptions, but NASA SP V3 profiles are at a spatial resolution of 1° latitude x 1.25° longitude from the Global Modeling Initiative chemical transport model whereas BEHR profiles are from a WRF-Chem model simulation at 12 km resolution. Assumptions about surface reflectivity in the NASA SP are from an OMI based climatology at 0.5° x 0.5° (Kleipool et al., 2008) and the BEHR product uses a bidirectional reflectance factor from the Moderate Resolution Imaging Spectroradiometer (MODIS) over land at 30 arcsec resolution and an ocean reflectivity model over water. Terrain pressure in BEHR is assumed from WRF-Chem pressure profiles adjusted with the Global Land One-kilometer Base Elevation database (Hastings and Dunbar, 1999) and the hypsometric equation whereas the NASA SP uses a 3 km digital elevation model (Boersma et al., 2011).

OMI data used in this analysis are filtered to exclude cloud fractions (derived from MODIS) greater than 20% (following criteria defined in Laughner et al. (2019)) and data from OMI's row anomaly (http://projects.knmi.nl/omi/research/product/rowanomaly-background.php). From 2004 until October 2017, OMI had the finest spatial resolution of any NO_2 retrieving sensor in low Earth orbit. Before the appearance of the row anomaly in 2007, the finest pixel size of OMI was approximately 13 x 24 km at nadir. In January 2009, the row anomaly extended to affect nadir pixels and during summer 2017 the pixel areas coincident with the campaign Pandora observations ranged from 365 km^2 near-nadir to approximately 4600 km^2 at the edge of the swath.

2.4 GeoTASO

Geostationary Trace gas and Aerosol Sensor Optimization (GeoTASO) is a hyperspectral mapping instrument built by Ball Aerospace (Leitch et al, 2014) to acquire data for optimizing and testing new high-resolution retrievals of trace gases (NO_2, O_3, and HCHO) and aerosols in preparation for geostationary air quality observations. The instrument is composed of a reflective telecentric telescope that focuses scattered light within its field of view through a photo-elastic modulator for depolarization and lastly into an Offner spectrometer. Within the spectrometer the second order and first order diffracted beams are used to form the ultraviolet (UV) and visible (VIS) spectral ranges, respectively. Each spectral range (UV, 300

Deleted: air quality

Deleted: ,

nm to 380 nm and VIS: 410 nm to 690 nm) is focused onto a separate 2-dimensional charge coupled device (CCD) (1056 pixels in the wavelength dimension x 1033 pixels in the spatial dimension) operated at an integration time of 250 ms for all flights. NO_2 is retrieved using a spectral window within the VIS channel, which has a spectral resolution (FWHM) of 0.88 nm and a spectral sampling of 3.1 pixels (Nowlan et al., 2016). GeoTASO observes as a push-broom sensor with an across-track nadir field of view of 45°, providing a swath-width of approximately 7 km from a nominal altitude of 8.5 km. Spectra (~300 images) are coadded to an approximate ground pixel size of 250 x 250 m.

During GeoTASO flights, gapless maps (otherwise referred to as rasters) were created over an area of interest by executing flight plans composed of parallel flight legs spaced 6 km apart. The overlap in swath between these flight legs provided a tolerance for maintaining gapless coverage during instances of strong cross-winds or when a lower flight altitude was selected in order to operate below high-level thin cirrus clouds. This raster strategy provides data that can easily be spatially co-added to evaluate the influence of spatial resolution on tropospheric NO_2 column measurements, as is shown herein. Additional details and estimated uncertainties of the GeoTASO NO_2 retrieval are elaborated in the following sections.

2.4.1 Airborne NO₂ Slant Column Retrieval and Uncertainty

NO_2 differential slant columns (DSCs) are retrieved using QDOAS, an open-source DOAS computing software developed by the Royal Belgian Institute for Space Aeronomy (http://uv-vis.aeronomie.be/software/QDOAS). Fitted trace gas and interference cross sections include NO_2 (Vandaele et al., 1998), O_4 (Thalman and Volkamer, 2013), H_2O (Rothman et al., 2009), CHOCHO (Volkamer et al., 2005), Ring spectrum (Chance and Kurucz, 2010), and a fifth-order polynomial in the spectral window of 425-460 nm. The resultant DSCs can be physically described as the additional total NO_2 absorption along the path sunlight travels through the atmosphere to the GeoTASO instrument relative to the reference atmosphere.

Reference spectra are collected in-flight using nadir-viewing observations for each across track position in the swath over a homogeneous area with minimal NO_2 absorption. One reference set was collected for each region. The LMOS reference spectrum was observed during nadir observations on May 27th, 2017 at 19:00 UTC (SZA=35°) over Lake Michigan with an estimated below aircraft column of $2.3x10^{15}$ molecules cm^{-2} (estimated from the NAM-CMAQ model described below). For the LA Basin flights, the reference was collected over a clean homogeneous area north of the LA Basin on June 27th, 2017 also at 19:00 UTC (SZA=16°) with an estimated below aircraft column of approximately $1x10^{15}$ molecules cm^{-2}.

The average spectral fitting uncertainty for the NO_2 DSCs over all flights is $1.2x10^{15}$ molecules cm^{-2} with a standard deviation of $0.25x10^{15}$ molecules cm^{-2}. Higher solar zenith angles (SZA) and lower surface albedos both result in slightly higher uncertainties due to their impacts on instrument signal-to-noise ratio. A multi-linear regression applied to the dependent variable of NO_2 DSC uncertainty vs. the independent variables of SZA and surface albedo, for all cloud-free data, indicates that 33% of the variability in DSC uncertainty is associated with changes in SZA (uncertainty increases with increasing SZA by $0.011x10^{15}$ molecules cm^{-2} per degree) and almost 15% is associated with variations in surface albedo (uncertainty increases with decreasing albedo by $0.033x10^{15}$ molecules cm^{-2} per 0.01 decrease in albedo). Additional uncertainty due to using one NO_2 absorption cross section at a single temperature (294 K) leads to a potential bias of -0.6±1.7%, estimated using the NO_2 profile weighted effective temperature profile in Eq. 4 in Bucsela et al. (2013).

2.4.2 Slant-to-Vertical Column Conversion

Further processing of DSCs into vertical columns (VCs) requires computations of the AMF: the ratio of the pathlength of light through the atmosphere due to scattering and geometry to the vertical pathlength. AMFs depend on a priori assumptions such as the vertical distribution of NO_2, surface reflectivity, clouds, aerosols, sun angle, and viewing geometry (Palmer et al., 2001; Lamsal et al., 2017). Mathematically, AMFs are the integrated product of (1) scattering weights (the vertical distribution of the instrument sensitivity calculated by a radiative transfer model) and (2) the NO_2 profile shape factor (the relative vertical distribution of NO_2) (Palmer et al., 2001). Scattering weight calculations are most sensitive to inputs that influence light transmission through the atmosphere, most notably from solar and viewing geometry, surface reflectivity, and aerosols (Lamsal et al. 2017; Meier et al., 2017). This work uses Harvard Smithsonian Astrophysical Observatory's AMF Tool which packages the VLIDORT radiative transfer model to calculate scattering weights (Spurr,

7

2006; Nowlan et al., 2016; Nowlan et al., 2018). Inputs include atmospheric profiles of temperature, pressure, ozone, NO_2, aerosol, aircraft altitude, viewing geometry, solar geometry, and surface reflectance.

Tropospheric NO_2 profiles are taken from a 12 km hourly analysis of a parallel developmental simulation of the North American Model-Community Multiscale Air Quality (NAM-CMAQ) model from the National Air Quality

5 Forecasting Capability (NAQFC; Stajner et al., 2011). Stratospheric profiles are estimated using the PRATMO chemical box model of stratospheric NO_2 profile climatology (Prather, 1992; McLinden et al. 2000) which estimates NO_2 between approximately 10 and 60 km as a function of month, latitude, and solar zenith angle. This stratospheric model has been used in previous GeoTASO retrievals and has an estimated uncertainty of ~30% in the stratospheric column (Bourassa et al., 2011; Nowlan et al., 2016). Temperature and pressure profiles from the NAM-CMAQ in the troposphere and the Realtime

10 Air Quality Modeling System (RAQMS; Pierce et al., 2009) up to 60 km are extracted and merged, and ozone profiles are extracted from the NAM-CMAQ analysis for the troposphere and OMI gridded monthly climatology in the stratosphere (Liu et al. 2010) for May and June 2017.

Noguchi et al. (2014) evaluated the influence of surface anisotropy on NO_2 AMFs for geostationary scale measurements using a 1 km MODIS BRDF product and found that not accounting for BRDF in the AMF calculation,

15 especially in areas with high NO_2 near the surface, can lead to large errors. For this reason, this work uses the BRDF isometric, volumetric, and geometric kernels retrieved in Band 3 by the MODIS MCD43A1 daily L3 500m v006 product (Lucht et al., 2000; Schaaf and Wang, 2015). To fill gaps and decrease noise, the daily product is averaged over a month in the regions GeoTASO measured (May 22nd-June 21nd for LMOS and June 15th-July 14th for the LA Basin). Remaining gaps are small and are filled using linear interpolation of nearby pixels. Variations in BRDF derived albedo compare well

20 with the brightness of surface features. Li et al. (2018) found the v006 MCD43A1 product to compare very well with Multi-angle Imaging SpectroRadiometer (MISR) land surface reflectance with little bias and a high correlation (r=0.95 in a case study in northeast Asia).

Clouds and aerosols are not considered in the AMF calculation in this work, as cloudy scenes are filtered by removing pixels having detector count rates of greater than 2.5×10^4 counts s^{-1} within the DOAS NO_2 spectral window and

25 aerosol loading was relatively low during flight days. Given the relatively low aerosol loadings, errors in AMF due to neglecting aerosols are expected to be small for these flights (< 5% using sensitivities shown by Lamsal et al. (2017)).

Differential slant columns are converted into below aircraft vertical columns using the following equation:

$$VC_{below} \ or \ TropVCs = \frac{DSC - VC_{above}AMF_{above} + reference \ slant \ column}{AMF_{below}}, \qquad (1)$$

Where DSC is the differential slant column, VC_{above} and VC_{below} are respectively the vertical column above and below the

30 aircraft, and AMF_{above} and AMF_{below} are respectively the calculated AMF above and below the aircraft. As NO_2 column variations in the troposphere are largely confined to the boundary layer, this work assumes that residual tropospheric variations in the above-aircraft column are negligible and the below aircraft vertical columns are referred to as tropospheric vertical columns (TropVCs).

2.4.3 Vertical Column Retrieval Sensitivity and Uncertainty

35 Figure 2 shows examples of NO_2 TropVCs (top) and tropospheric AMFs (middle) for a morning flight (~08:30-12:00 LDT) over Chicago, IL on June 1st, 2017 (left) and a morning flight (~8:40-10:05 LDT) over the LA Basin from June 26th, 2017 (right). The bottom row of Fig. 2 shows base maps from Google Earth Pro to aid identification of surface features in the maps above. In the areas mapped in Fig. 2, NO_2 TropVC spans roughly two orders of magnitude and spatial patterns are consistent with emission sources, such as busy roadways, airports, and industrial regions. Both mornings in each region had

40 relatively stagnant conditions.

8

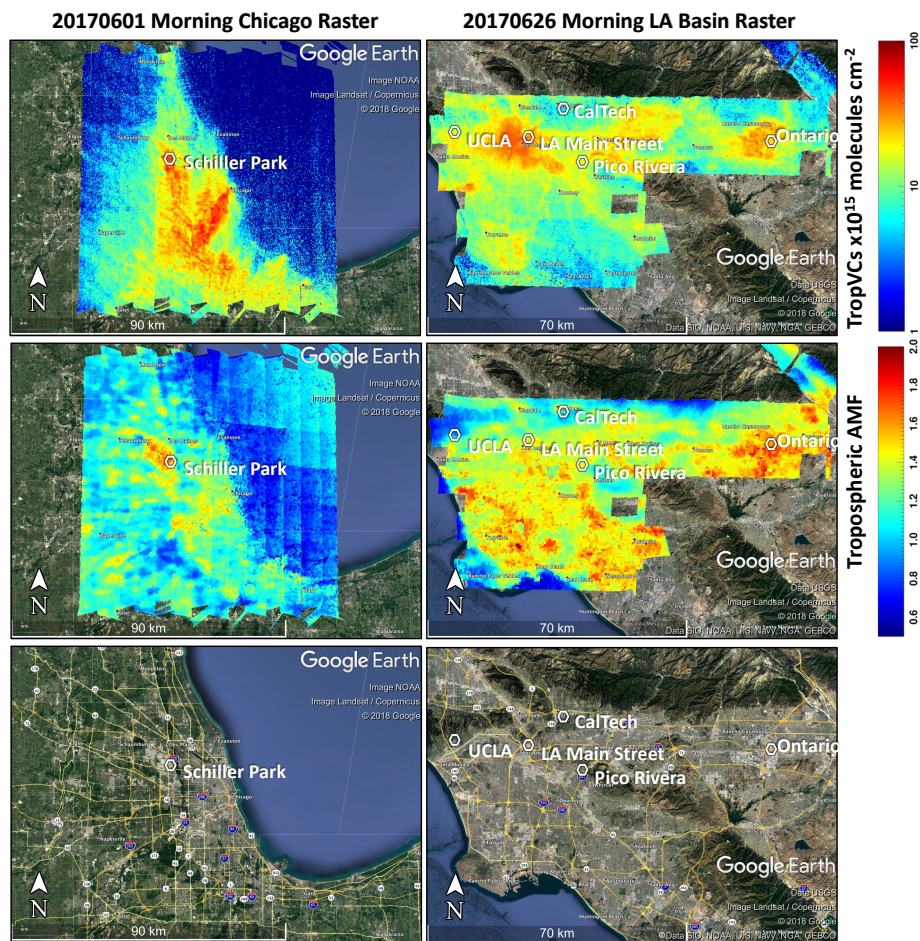**20170601 Morning Chicago Raster** **20170626 Morning LA Basin Raster**

Figure 2: Tropospheric NO₂ vertical column on a log10 color scale (top), tropospheric AMF on a linear color scale (middle), and base map (bottom) for a Chicago raster on the morning (08:30-12:00 LDT) of June 1st, 2017 (left) and an LA Basin raster on the morning (08:40-10:05 LDT) of June 26th, 2017 (right). The base maps show roadways, urbanized areas, and surface characteristics. The labeled hexagons indicate locations of Pandora spectrometers.

9

Over land, large spatial variations of tropospheric AMF are evident, and associated primarily with the surface reflectance characteristics. A multi-variable linear regression of these data indicates that surface albedo explains 64% of the variability in tropospheric AMF, illustrating the importance of accurately treating surface reflectance in these high spatial resolution retrievals. Over an area with homogeneous surface reflectance (like Lake Michigan east of Chicago), the subtle influence of a varying NO_2 shape factor is visible in the AMF, but has little impact on the magnitude of TropVC because there is little NO_2 in this location. The a priori NO_2 profile accounts for 16% of the variability in tropospheric AMF. Viewing geometry has smaller effects, with SZA accounting for 1.8% of the variability and the relative azimuth angle and viewing zenith angle causing even smaller amounts.

To better gauge the sensitivity of AMF and TropVC to the NO_2 vertical profile in the troposphere, the AMFs calculated for the most polluted LMOS flight (June 1st, 2017; Fig. 2) were compared to AMFs calculated using a single median NAM-CMAQ NO_2 vertical profile (not shown). This median profile is representative of a moderately polluted urban area with a tropospheric NO_2 column of $3.6x10^{15}$ molecules cm^{-2}, most of the NO_2 in the lowest km of the atmosphere, and an NO_2 maximum at the surface of approximately 4 ppbv. The resulting differences in AMF can be large, ranging from -30% to 10%, however the largest AMF differences occur over relatively clean areas in which the median polluted profile shape is less representative. The differences in TropVCs are about $\pm1x10^{15}$ molecules cm^{-2} at the 5th/95th percentiles with 50% of the points having differences between $-0.2x10^{15}$ and $0.5x10^{15}$ molecules cm^{-2}. These values are on the same order of magnitude as the uncertainty in DSCs.

To estimate total uncertainty in the TropVC, error propagation was applied to Eq. (1). Assuming a 50% uncertainty in the NAM-CMAQ profile, 30% in the PRATMO profiles, and a conservative 30% uncertainty in the below aircraft AMF (largest percent variation in the sensitivity study discussed in the preceding paragraph), the uncertainty in the reference slant column is estimated to be $2.3x10^{15}$ molecules cm^{-2} for LMOS and $1.9x10^{15}$ molecules cm^{-2} for SARP. The average uncertainty in the above aircraft vertical column is $1.3x10^{15}$ molecules cm^{-2} with a standard deviation of $0.08x10^{15}$ molecules cm^{-2}. The median total uncertainty for TropVCs for both regions combined is $2.3x10^{15}$ molecules cm^{-2}. Percent uncertainty (5th-95th percentile) ranges from 46-143% for relatively unpolluted columns (< $5x10^{15}$ molecules cm^{-2}), from 32-60% for moderately polluted columns (5-15x10^{15} molecules cm^{-2}), and ~30% for polluted columns (> $15x10^{15}$ molecules cm^{-2}). These values are similar to those calculated by Nowlan et al. (2016).

3 Results

3.1 Comparison of airborne GeoTASO and ground-based Pandora retrievals

Pandora spectrometer measurements have been previously used to validate retrieval products from GeoTASO and similar airborne mapping spectrometers during intensive campaigns (Nowlan et al., 2016; Lamsal et al., 2017; Nowlan et al., 2018). Figure 3 shows the comparison between the 80 coincident Pandora and GeoTASO TropVC observations during LMOS and the LA Basin flights colored by ground site. The first five ground sites in the legend correspond to Pandora spectrometers installed in the LA Basin and the last five are located in the LMOS domain. Coincidences are identified as the median of all cloud-free GeoTASO TropVCs having pixel centers within a 750 m radius of a Pandora for each overpass (minimum requirement of 16 valid cloud-free GeoTASO 250 x 250 m pixels for a spatial coverage of at least 1 km²) and the Pandora column observed closest in time to the GeoTASO overpass (within ± 5 minutes). A Pandora TropVC is derived by subtracting the stratospheric vertical column calculated with the PRATMO climatology from the Pandora total column, following the same approach used with GeoTASO TropVCs. Coincidences without at least two valid Pandora data points within ± 5 minutes of the GeoTASO overpass are excluded because acquisition of only a single Pandora data point within ten minutes indicates likely periodic clouds or poor solar tracking by the instrument. TropVCs from GeoTASO compare well with Pandora, with a high correlation (r^2=0.91), a slope of 1.03, and an offset of $0.52x10^{15}$ molecules cm^{-2}.

The whiskers in Fig. 3 indicate the 10th and 90th percentiles of GeoTASO TropVCs within the 750 m radius for GeoTASO (vertical whiskers) and the maximum and minimum Pandora TropVC within ± 5 minutes from the GeoTASO overpass (horizontal whiskers, representing 6-7 valid Pandora measurements for most of these coincidences), providing a glimpse of the spatiotemporal variability at the time of the coincidence. Generally, spatial variability (vertical whiskers) increases as the magnitude of NO_2 increases. An exception to this pattern occurs at the Zion site. Zion has the lowest surface

10

albedo of any site, ranging from 1.3-2.6% depending on SZA. The lower albedo over the Zion region together with the relatively low NO$_2$ amounts lead to a lower signal-to-noise ratio in the GeoTASO observations, resulting in a higher uncertainty in the DOAS spectral fit (as discussed in Sect. 2.1). Therefore, the large vertical whiskers at this site are indicative of increased uncertainty in the retrieval in addition to spatial heterogeneity.
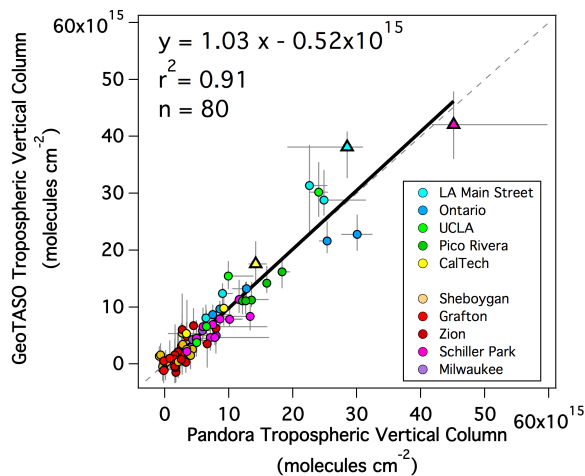

5

Figure 3: Scatter plot of GeoTASO NO$_2$ TropVCs vs. Pandora NO$_2$ TropVCs colored by Pandora site. Vertical whiskers show the 90th and 10th percentiles of GeoTASO TropVCs within the 750 m radius of the Pandora site. Horizontal whiskers show the maximum and minimum Pandora TropVCs within the ± 5-minute coincidence window. Triangles indicate coincidences discussed within Sect. 3.1. The grey dashed line indicates the 1:1 line.

10

The largest temporal variability (horizontal whiskers) occurs at the Schiller Park site in Chicago, IL, which is located along a major highway near the end of a Chicago O'Hare airport runway. High temporal variability is common at the site and is likely due to local emissions associated with air traffic and nearby roadways. The most polluted point of all coincidences occurs at Schiller Park on June 1st, 2017 (pink triangle in Fig. 3) and this case is further explored in Fig. 4.

15 Figure 4(a) shows a map of the GeoTASO NO$_2$ TropVCs and a hexagon depicting the Pandora location which is colored by the NO$_2$ column observed by Pandora nearest in time to the GeoTASO overpass. The 750 m radius used to define coincident GeoTASO data and an arrow depicting the direction in which Pandora is observing are overlaid. GeoTASO observed a median of 42×10^{15} molecules cm^{-2}, whereas the Pandora observed 45×10^{15} molecules cm^{-2}. Inside the 750 m radius, NO$_2$ columns range between 30 and 50×10^{15} molecules cm^{-2}, and the highest columns observed occur along the viewing direction

20 of Pandora. Figure 4(b) shows a time series of the Pandora TropVC surrounding this overpass and the GeoTASO coincidence. Within an hour prior to the overpass, large temporal variations in NO$_2$ column are observed by the Pandora, suggestive of small-scale plumes influencing the site. In the 10-minute Pandora window (grey section), Pandora peaks at over 60×10^{15} molecules cm^{-2}. If any temporal averaging is applied to Pandora data in this case, the coincidence would yield a larger apparent difference between GeoTASO and Pandora. The tremendous amount of variability at this highly heterogeneous location suggests caution for use in applications such as satellite validation.

25 Although the 80 GeoTASO/Pandora coincidences are highly correlated with a slope near 1:1, there are a few outliers that deviate from the linear regression line by amounts larger than the range of observed variability by either instrument. Some of these outliers can be attributed to mismatches in spatial representativeness, as the viewing geometry of the Pandora spectrometer is always oriented in the direction of the sun while the GeoTASO median is constructed from

11

pixels within the 750 m radius centered on the Pandora location extending into all directions. Figure 5 highlights one of these cases (yellow triangle in Fig. 3). Over the CalTech site in the LA Basin on June 27th, 2017, Pandora observes a column of approximately 14×10^{15} molecules cm^{-2}, whereas the GeoTASO spatial median is 18×10^{15} molecules cm^{-2}. The GeoTASO TropVC map (Fig. 5 (a)) shows that Pandora is observing in a direction where NO$_2$ values are at a local minimum, such that the GeoTASO median within the 750 m radius is about 30% higher than Pandora, though GeoTASO TropVCs along the viewing direction of Pandora at that time are more similar to the Pandora observations. Pandora TropVCs are varying slowly in time surrounding the GeoTASO overpass (Fig. 5(b)). Together these characteristics suggest that the differences between these two observations are more likely associated with the spatial criteria assumed for this comparison rather than the temporal variability near the coincidence time. Some of the other coincidences showing notable differences, including the two most polluted Ontario points (both lying below the 1:1 line with concentrations of $20\text{-}25\times10^{15}$ molecules cm^{-2}) also appear to be associated with the assumption in spatial criteria (not shown). Future comparisons of Pandora direct-sun retrievals to high-spatial resolution airborne data could consider the Pandora's viewing geometry in the coincidence criteria, but this is not a viable option for validating satellite products that are not as spatially refined
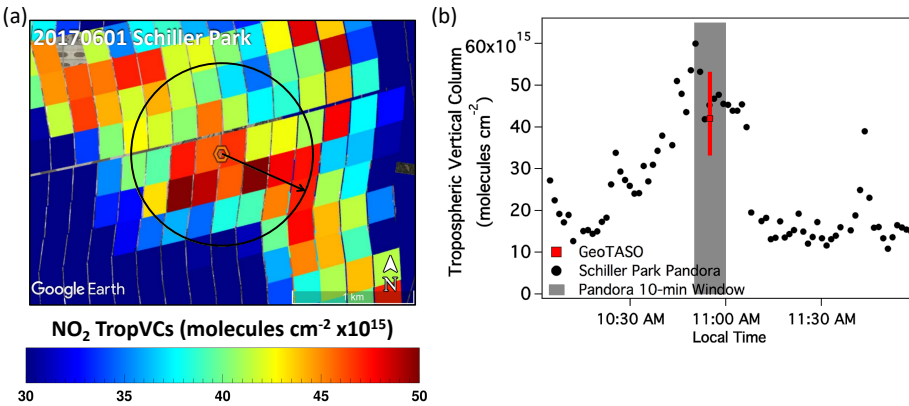

NO$_2$ TropVCs (molecules cm^{-2} x10^{15})

Figure 4: (a) Map of GeoTASO TropVCs on a linear color scale for the Schiller Park overflight on June 1st, 2017 at 15:55 UTC (10:55 LDT) (pink triangle in Figure 3) with the 750 m radius considered in the spatial binning of GeoTASO overlaid and an arrow depicting the Pandora viewing direction (solar azimuth angle) during the overpass time. The Pandora hexagon is colored by the NO$_2$ TropVC measured by Pandora during the overpass. (b) Time series showing Pandora data (black points) within approximately ± 1 hour of the GeoTASO overpass. The Pandora temporal window for the coincidence is shaded in grey and the GeoTASO TropVC and 10th-90th percentiles from the overpass are shown in red.

Deleted: On June 27th, 2017, over

Deleted: ,

Deleted: during the overpass

Deleted: while the

Deleted: at the time
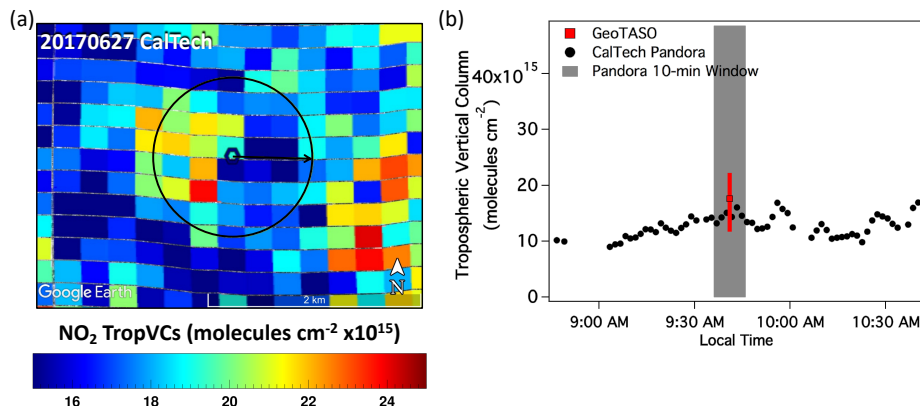
Deleted: Pandora

Deleted: which

Deleted: .

Figure 5: (a) Map of GeoTASO TropVCs on a linear color scale for the CalTech overflight on June 27th, 2017 at 16:41 UTC (09:41 LDT) (yellow triangle in Figure 3) with the 750 m radius considered in the spatial binning of GeoTASO overlaid and an arrow depicting the Pandora viewing direction (solar azimuth angle) during the overpass time. The Pandora hexagon is colored by the NO$_2$ TropVC measured by Pandora during the overpass. (b) Time series showing Pandora data (black points) within approximately ± 1 hour of the GeoTASO overpass. The Pandora temporal window for the coincidence is shaded in grey and the GeoTASO TropVC and 10th-90th percentiles from the overpass are shown in red.
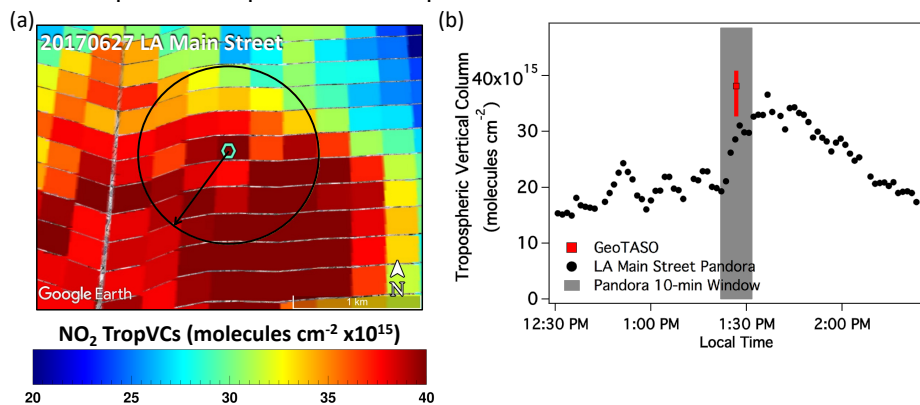

Figure 6: (a) Map of GeoTASO TropVCs on a linear color scale for the LA Main Street overflight on June 27th, 2017 at 20:26 UTC (13:26 LDT) (cyan triangle in Figure 3) with the 750 m radius considered in the spatial binning of GeoTASO overlaid and an arrow depicting the Pandora viewing direction (solar azimuth angle). The Pandora hexagon is colored by the NO2 TropVC measured by Pandora during the overpass. (b) Time series showing Pandora data (black points) within approximately ± 1 hour of the GeoTASO overpass. The Pandora temporal window for the coincidence is shaded in grey and the GeoTASO TropVC and 10th-90th percentiles from the overpass are shown in red.

13

The largest outlier from the regression line of Pandora-to-GeoTASO coincidences occurs at LA Main Street on the afternoon of June 27th, 2017 (blue triangle in Fig. 3), where GeoTASO observes a TropVC of 38×10^{15} molecules cm^{-2} and the Pandora TropVC is only 28×10^{15} molecules cm^{-2}. This coincidence also shows large gradients both spatially and temporally. What is unique about this case is the environmental/meteorological conditions present during this time period that make data comparisons and GeoTASO retrieval challenging. Figure 6 has the same formatting as Fig. 4 and Fig. 5.
5 Focusing on the ten-minute window used in the temporal matching of Pandora data, Pandora NO$_2$ TropVC increases from 20×10^{15} molecules cm^{-2} to over 30×10^{15} molecules cm^{-2}. This feature is caused by the arrival of a sea breeze front passing through from the southwest at this time that acts as a convergence zone accumulating NO$_2$ along this front as it passes through the LA Basin (Judd et al., 2018). In this case, because the Pandora data show a rapid increase, any temporal
10 averaging window placed on the Pandora data for data comparison would not be representative of the conditions during the overpass (similar to the coincidence shown in Fig. 4). Pandora reaches a peak within this 10-minute window of 31×10^{15} molecules cm^{-2} approximately one minute after the GeoTASO overpasses, which is closer to the magnitude observed by GeoTASO but still about 7×10^{15} molecules cm^{-2} lower. Within 5-10 minutes after the overpass, Pandora values rise again to a range of 33-37×10^{15}, almost as large as GeoTASO median value.
15 Figure 6(a) indicates that directionality does not reconcile the observations as GeoTASO NO$_2$ TropVCs in the direction Pandora is viewing are consistently larger than Pandora observations, suggesting a potential high bias in the GeoTASO TropVC in this case. High bias could be caused by an underestimation of the AMF; the tropospheric AMF would have to increase from 1.3 to ~1.75 for this area to yield a TropVC similar to Pandora. Given the unique circumstances of the sea breeze front, and the relatively coarse resolution of the NAM-CMAQ model used to obtain the NO$_2$ profile (12 km), we
20 speculate that perhaps the shape factor used in the AMF calculation may not be representative of the actual NO$_2$ vertical profile within the sea breeze circulation during this time. The rapid fluctuations in Pandora observations through this period also suggest that the airmass is not well mixed, i.e., eddies likely exist in the vicinity of this front which could affect vertical structure at very localized scales. Another contribution could be inaccuracies in the AMF associated with surface reflectance at this site. All LA Main Street coincidences lie above the 1:1 line in Fig. 3 (GeoTASO TropVC is consistently larger than
25 Pandora), which could be caused by a low bias in the BRDF weighting kernels. LA Main Street is in an area that has fewer valid retrievals from the MCD43A1 product during the monthly average used for BRDF input to VLIDORT. Further investigation of systematic biases by site due to BRDF uncertainty would require more consistent sampling at the site, which will be possible with satellite products and long-term Pandora measurements used for product validation in the future.

Table 3: Statistics between Pandora and the GeoTASO data at the nominal pixel size shown in Figure 3 as well as the simulated pixel
30 **sizes for TEMPO, TROPOMI, and OMI products shown in Figure 9.**

| Pixel Size	All Points				Pandora < 40×10^{15} molecules cm^{-2}				Pandora < 20×10^{15} molecules cm^{-2}			
	slope	Intercept x10^{15}	r^2	N	slope	Intercept x10^{15}	r^2	N	slope	Intercept x10^{15}	r^2	N
Nominal:	1.03	0.52	0.91	80	1.08	-0.75	0.89	79	0.96	-0.30	0.84	73
TEMPO: 3 x 3 km	0.88	-0.07	0.86	79	1.03	-0.72	0.89	78	0.93	-0.35	0.85	72
TROPOMI: 5 x 5 km	0.77	0.11	0.88	76	0.88	-0.32	0.89	75	0.87	-0.36	0.81	69
OMI: 18 x 18 km	0.57	0.9	0.61	66	0.73	0.01	0.83	65	0.89	-0.7	0.78	60

Deleted: which

Deleted: 2
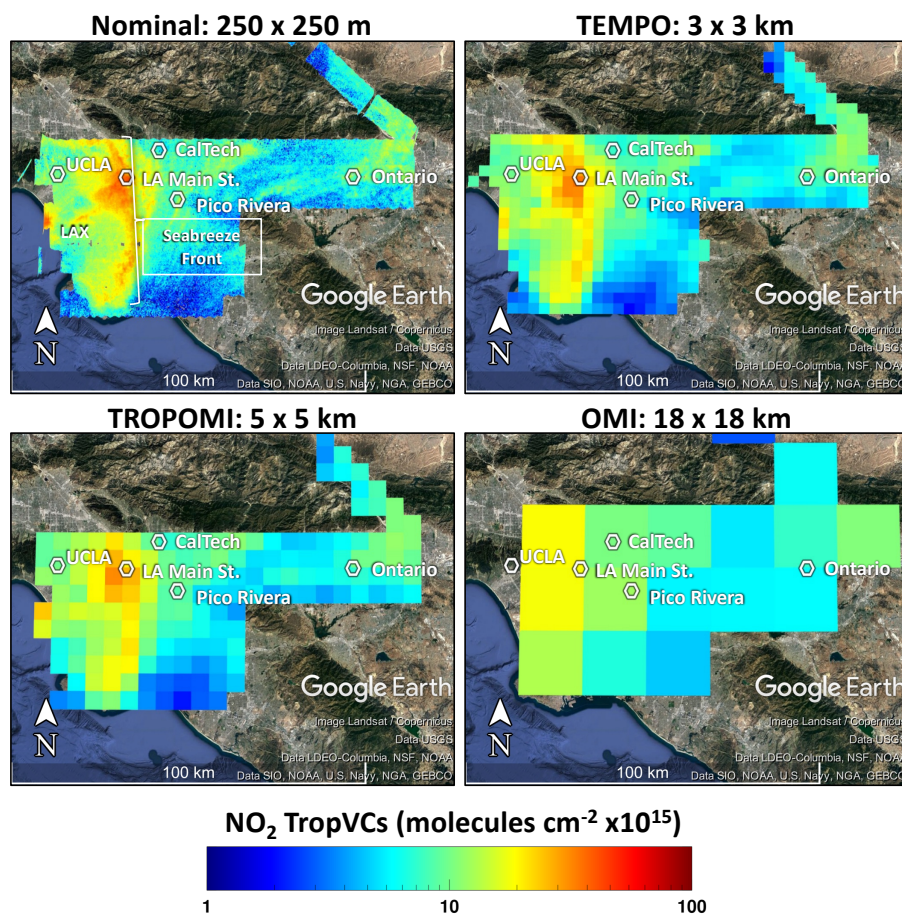
Figure 7: Maps of GeoTASO NO₂ TropVC on a log10 color scale averaged to the labeled pixel sizes to demonstrate the spatial detail that would be observed from the midday raster in Los Angeles on June 27ᵗʰ, 2017. All pixels shown have at least 30% of their area mapped by GeoTASO during this timeframe.
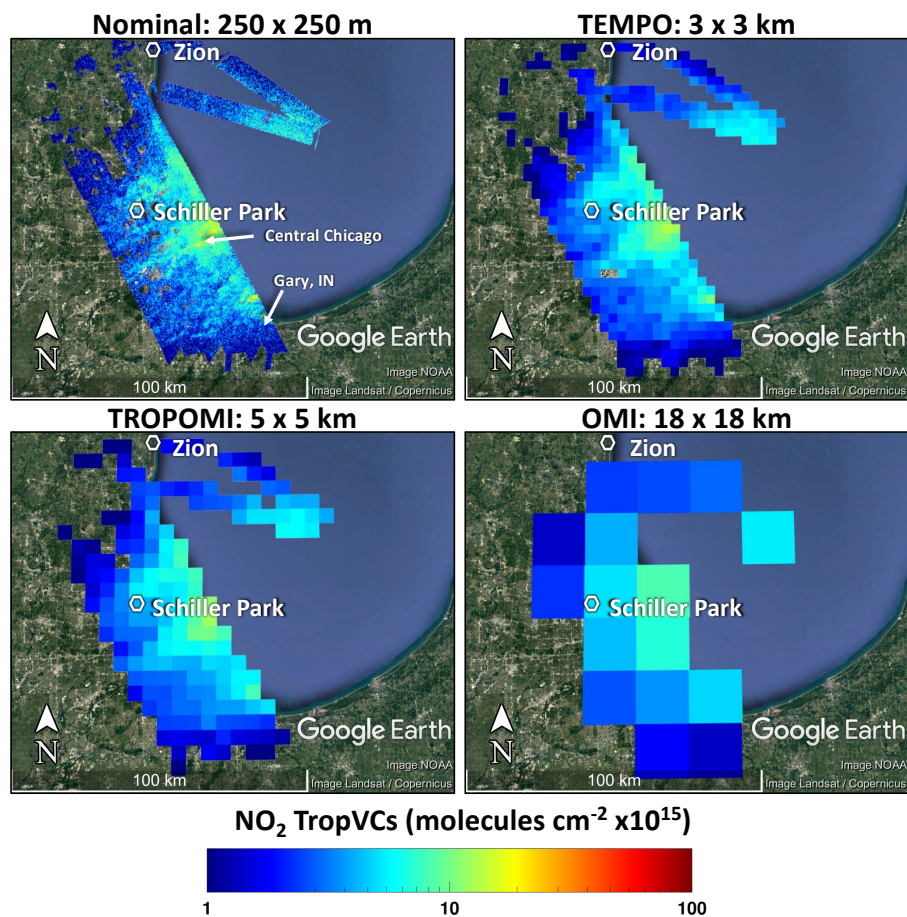
Figure 8: Maps of GeoTASO NO₂ TropVC on a log10 color scale averaged to the labeled pixel sizes to demonstrate the spatial detail that would be observed from the midday in the Chicago area on May 22ⁿᵈ, 2017. All pixels shown have at least 30% of their area mapped by GeoTASO during this timeframe.

5

3.2 Scaling GeoTASO to satellite product footprints

Past, present, and planned satellite instruments have spatial resolutions that are coarser than the GeoTASO airborne measurements. The gapless mapping strategy executed with GeoTASO during summer 2017 provides fine-resolution data suitable for averaging to coarser areal resolutions typical of these current and next-generation satellite retrievals. As

5 demonstrated by the good agreement between Pandora and GeoTASO shown in Sect 3.1, the GeoTASO observations during these rasters capture the spatiotemporal variability existing in these areas. Comparisons of upscaled GeoTASO retrievals with Pandora measurements provide an early assessment of the capabilities of the next-generation sensors for resolving urban scale NO_2 and also give insight toward validation strategies in urban regions, where sub-pixel variability within the satellite product footprint may be significant.

10   To simulate the satellite products, each GeoTASO raster is averaged to fixed grids with pixel areas representative of near-nadir observations from TEMPO, TROPOMI, and OMI. These grids are simplified by making each pixel square in shape to avoid introducing an orientation bias. Simulated pixels are 3 x 3 km (9 km^2) for TEMPO (literature reported 9.24 km^2 per pixel, Zoogman et al., 2017), 5 x 5 km (25 km^2) for TROPOMI (literature reported 24.5 km^2, van Geffen et al., 2019), and 18 x 18 km (324 km^2) for OMI (FoV75Area for nadir is 338.4 km^2).

15   GeoTASO data are remapped to these fixed grids by computing a weighted average of the GeoTASO TropVCs based on the areal overlap within each simulated satellite grid pixel for each raster. Figure 7 shows an example of this process for the midday Los Angeles raster on June 27th, 2017, at the nominal GeoTASO resolution and each simulated satellite product resolution. At this time, a marine/land airmass convergence zone was created by the development of a sea breeze that resulted in an accumulation of NO_2 seen as a line of enhanced NO_2 in excess of 30×10^{15} molecules cm^{-2}

20 extending north and south of LA Main Street (discussion of Fig. 6, Sec. 3.1; Judd et al., 2018). A plume is also seen extending inland from the Los Angeles International Airport (LAX). As expected, fine scale features evident in the native GeoTASO observations are increasingly blurred as the spatial resolution is degraded. The airport plume is spatially distinct up to the TEMPO spatial scale and the sea breeze is resolved through the TROPOMI spatial scale, but neither are visible at the spatial scale of OMI.

25   Another example is shown in Fig. 8 for a Chicago raster near midday on May 22nd, 2017. In this example, the southwesterly wind throughout the day quickly advected local emissions out over Lake Michigan, allowing visualization of fresh emission source regions and broadening and accumulation of the resulting pollution plume as it passes over the Chicago metropolitan area. The native-resolution GeoTASO retrievals show three areas of enhanced NO_2: near Schiller Park-O'Hare Airport, central Chicago, and the industrial shore near South Chicago and Gary, IN. This raster was repeated 3

30 other times on this day and all four rasters show a similar spatial pattern but magnitude of NO_2 TropVCs decreases throughout the day, likely indicating that morning emissions are larger than afternoon emissions on this day (not shown). These spatial features are discernible up through the areal resolution of TROPOMI but the OMI spatial scale is larger than the physical separation of these three enhanced regions leading to loss of this spatial information. The maximum TropVC observed at an OMI spatial resolution is even displaced, such that it does not accurately represent the location of the highest

35 NO_2 TropVCs.
         This upscaling technique was applied to all GeoTASO flights listed in Table 1. To assess the impact of spatial resolution on comparisons to Pandora, each Pandora coincidence from Sect. 3.1 is matched to its equivalent upscaled pixel. Fig. 9 shows the scatter plots of these coincidences and the associated linear regression statistics are shown in Table 3. To avoid excess extrapolation, only upscaled grid cells that are at least 30% sampled by GeoTASO are considered in this

40 analysis, which is why the number of coincidences decrease as pixel size increases (Table 3).
         Increasing pixel size progressively worsens 1:1 Pandora agreement with the slope decreasing from 1.03 at native resolution to 0.88, 0.77, and 0.57, respectively at TEMPO, TROPOMI, and OMI scales. The decrease in slope is partially driven by the most polluted coincidence at Schiller Park. Figure 10 shows this coincidence in context (zoomed out map from Fig. 4) with the upscaled equivalent pixels from TEMPO, TROPOMI, and OMI overlaid. The two nearest TEMPO pixels

45 are shown because the Pandora location is near a pixel edge, just inside the left TEMPO pixel. The GeoTASO observations show that the Pandora is located within a narrow region of strongly enhanced NO_2. The size of the feature observed by GeoTASO and Pandora in this case is smaller than the spatial scale of TEMPO pixels, and because the TEMPO pixels also include regions of much lower NO_2 values, the simulated-satellite values decrease from 42×10^{15} molecules cm^{-2} at native GeoTASO resolution to less than 30×10^{15} molecules cm^{-2} at simulated TEMPO resolution. This case illustrates the challenge

17

Deleted: to

Deleted: which

Deleted: ,

Deleted: differing magnitudes

Deleted: not at

Deleted: .

Deleted: quantify

Deleted: relationship between the simulated satellite pixels and

Deleted: data, the

Deleted: -based analysis of

Deleted: repeated for each

Deleted: resolution

Deleted: 2

Deleted: decreases

Deleted: 2

Deleted: As the

Deleted: increases, the inability to resolve smaller scale polluted plumes results in larger deviation from the

Deleted: at a TEMPO scale to

Deleted: , respectively

Deleted: plume

Deleted: a TEMPO pixel with a

Deleted: value that decreases quickly

Deleted: (Pandora is located inside the left TEMPO pixel in the map though the two nearest TEMPO pixels are shown in Fig. 10 to contemplate placement of Pandora in the TEMPO grid as its placement is near the pixel edge and does not appear to be a factor). This

of determining the difference between retrieval bias and spatial misrepresentation but also the advantage of having high-spatial resolution airborne spectrometer observations as a transfer standard between satellite measurements and localized column measurements from Pandora. In this analysis, the native resolution retrieval does not appear to be biased (Sect. 3.1). Therefore, the changes in the relationship between upscaled GeoTASO observations and Pandora are solely due to spatial representation. In the case of this coincidence with its high degree of spatial and temporal variability, none of the simulated satellite products are capable of resolving this dominate sub-grid scale feature observed by GeoTASO and Pandora, and without careful analysis the satellite products could appear to have a low bias if extrapolating validation datasets are extended to include column amounts that are likely representative of extreme sub-grid features too large for the satellite retrieval to resolve.

Figure 9 illustrates that a linear relationship appears to degrade above some threshold value of NO_2 in these cases. Above the threshold, the more polluted NO_2 columns observed by Pandora are likely occurring over spatial scales smaller than the satellite resolutions, as shown in the preceding example. This theoretical threshold decreases as the spatial resolution is coarsened. The decreasing linear slope with increasing pixel size simply results from this flattening of the satellite-to-Pandora relationship at the higher NO_2 values. This threshold does not imply that the satellite cannot detect such large amounts of pollution, but instead reflects that in these regions NO_2 enhancements of these large magnitudes are not sustained over spatial scales that are representative of current/future space-based retrievals. Such behavior makes sense if the high NO_2 values are associated with localized heterogeneous features rather than more mixed regional-scale enhancements, which is a reasonable conceptual model for urban areas such as these. The value of the threshold likely varies by region and may not be generalizable, but these results suggest that there is a limit to the magnitude of pollution that can be validated using the localized measurements of Pandora (i.e. Pandora is only representative of each satellite product up to a certain pollution scale). As longer-term Pandora data records become available, it is possible that examining temporal variance statistics at each site can allow development of screening criteria to help identify such situations.
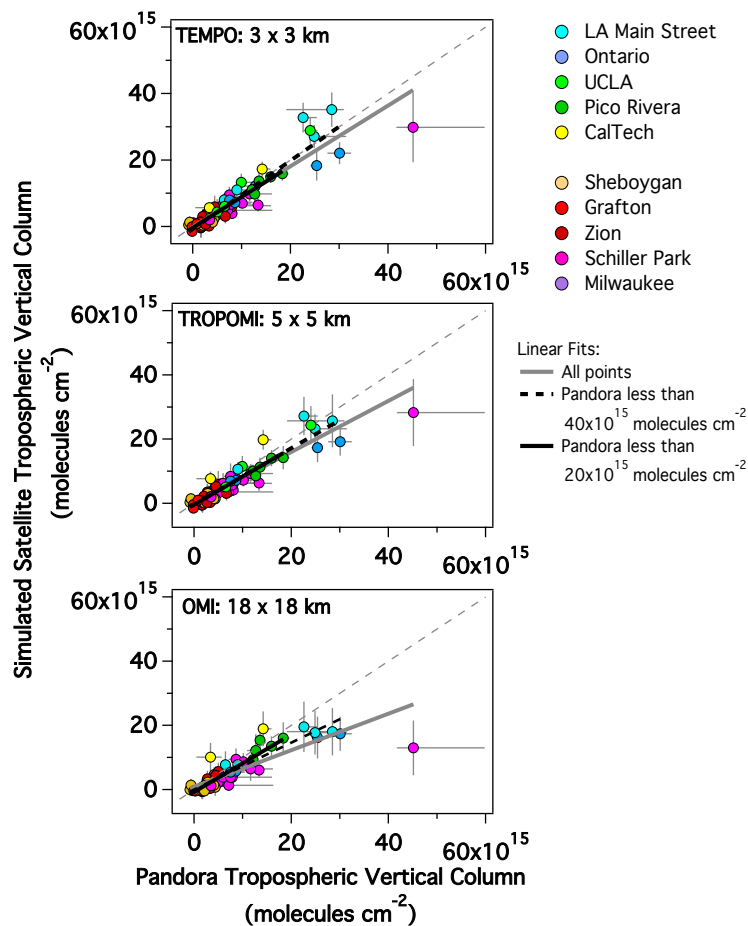
18

Figure 9: Scatter plot of GeoTASO TropVCs scaled to the nadir areal resolution of TEMPO (a), TROPOMI (b), and OMI (c) TropVCs vs. Pandora TropVCs colored by site and their associated linear fits (statistics listed in Table 3). Vertical bars show standard deviation of GeoTASO TropVCs within the upscaled pixel. Horizontal bars show the minimum and maximum of Pandora data within the ± 5-minute coincidence window. The grey dashed line indicates the 1:1 line.
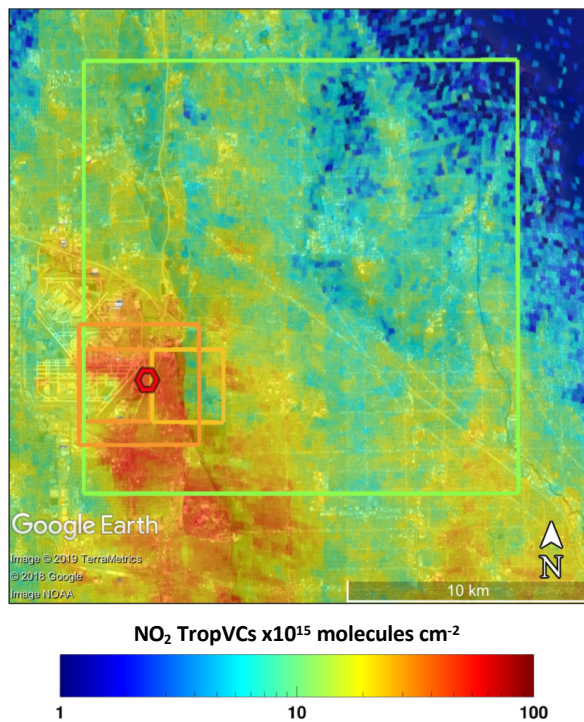
Figure 10: Map of GeoTASO TropVCs on a log10 color scale at the 250 x 250 m spatial resolution during the morning raster of June 1st, 2017 over Schiller Park (same as Figure 4) with the overlaid simulated pixels for TEMPO, TROPOMI, and OMI encompassing the Schiller Park ground site colored by the magnitude of their upscaled TropVC. The hexagon indicates the location of the Pandora and is colored by the observed NO_2 from the Pandora.

The data from these campaigns are not sufficient for confidently quantifying such thresholds but do allow the sensitivity to different thresholds to be explored by examining how the statistics vary as pollution thresholds for Pandora coincidences are decreased. To do this, each regression is repeated by excluding points for which Pandora TropVCs exceed 40×10^{15} molecules cm^{-2} (dashed regression line in Fig. 9) and exceed 20×10^{15} molecules cm^{-2} (solid black regression line in Fig. 9). Only the most polluted Schiller Park coincidence is excluded when filtering by Pandora TropVCs less than 40×10^{15} molecules cm^{-2}. The example discussed in association with Fig. 10 shows that this is a case in which the Pandora and GeoTASO are observing a highly localized feature. With the removal of this point, the slopes increase (improve) and the correlations remain similar, with the exception of a notable improvement at OMI resolution. The TEMPO scale comparison resembles the results at the nominal spatial scale, but slope still decreases as data are scaled to the coarser TROPOMI and OMI resolutions. When considering coincidences limited to Pandora TropVCs less than 20×10^{15} molecules cm^{-2}, the

20

correlations all degrade by 5-10% due to the smaller dynamic range with similar noise but are still on order of 0.8 or better. Pixel size has less influence on calculated slope, with only a 7% decrease in slope between the nominal pixel size and the OMI scale resolutions. Even below these thresholds, individual coincidences are often associated with influence of spatial heterogeneity as shown by the vertical whiskers. In fact, no single Pandora site was completely homogeneous for all coincidences (determined by the standard deviation of the GeoTASO measurements in the area surrounding each site as well as visual examination of Pandora time-series near the times of coincidences). However, if this type of pollution environment is sampled enough, the variability could be statistically characterized sufficiently enough to allow identification of retrieval biases or errors.

This analysis shows that, if a priori information at high enough spatial resolution is applied in the satellite retrieval, Pandora and satellite scale products can compare well statistically within these regions up to the nadir pixel size of OMI for Pandora columns less than 20×10^{15} molecules cm^{-2} (although individual points can still deviate largely from the 1:1 relation due to sub-pixel variability). The threshold pollution scale for TROPOMI comparisons to Pandora is higher, somewhere between $20\text{-}40 \times 10^{15}$ molecules cm^{-2}, and for TEMPO is higher still, up to 40×10^{15} molecules cm^{-2}. This analysis would benefit from more coincidences in the range of $20\text{-}60 \times 10^{15}$ molecules cm^{-2} to more confidently define pollution scale thresholds acceptable for Pandora and satellite product comparisons. Acceptable pollution ranges for data comparison may even vary from region-to-region as each urban area could have distinct spatial variability patterns.

These results suggest that Pandora spectrometer direct-sun observations are very useful for validating satellite NO_2 products at the spatial resolutions of TEMPO and TROPOMI, as the products are highly correlated with points tightly spanning both sides of the 1:1 line for values of up to approximately 30×10^{15} molecules cm^{-2} (and perhaps higher, but this dataset lacks sufficient observations above that level). However, even at the finest satellite product resolution (TEMPO), comparisons with Pandora observations are not exempt from outliers caused by extreme degrees of subpixel heterogeneity. When choosing validation locations, areas known to have consistently strong spatiotemporal variability at spatial scales finer than the satellite retrieval (e.g., Schiller Park) should be avoided. When possible, additional measurements (i.e., high resolution airborne mapping) would help investigate uncertainty associated with spatial representativeness in the analysis of retrieval bias.

3.3 Comparison of OMI satellite products and ground-based Pandora measurements

During the LMOS and LA Basin flights, OMI was the highest resolution space-based platform in orbit, as TROPOMI was not launched until October of that year. Since 2004, OMI has been observing global NO_2 columns at a regional spatial scale on daily timescales. The simulated OMI comparisons in Sect. 3.2 are considered an idealized 'best case' scenario for data comparison due to the near-nadir spatial resolution and the high resolution a priori used for GeoTASO AMFs. In actuality, OMI's pixel area increases over 1200% from nadir to the swath edge and most of the a priori inputs to the AMF calculations have coarser spatial resolutions than those used in these GeoTASO retrievals. This section demonstrates how actual OMI observations compare with the Pandora measurements in the LMOS and LA Basin domains during May-July 2017 to cover the time period during which OMI BEHR data were available and the Pandoras were operating in the two regions.

Figure 11 shows the 444 coincidences between the OMI BEHR TropVC and the Pandora TropVC (Pandora total column minus the OMI stratospheric column) with the best-fit linear regression for all points shown as a dashed black line. The points are further subset by considering only pixel areas less than 1000 km^2 (black points and solid black regression line). Coincidence and comparison criteria include OMI pixels unaffected by the row anomaly encompassing the location of a Pandora with a MODIS cloud fraction less than 20%. Pandora data are averaged within a $\pm$ 5 min window of the OMI overpass time to be consistent with the temporal window used to assess variability in Sect 3.1.
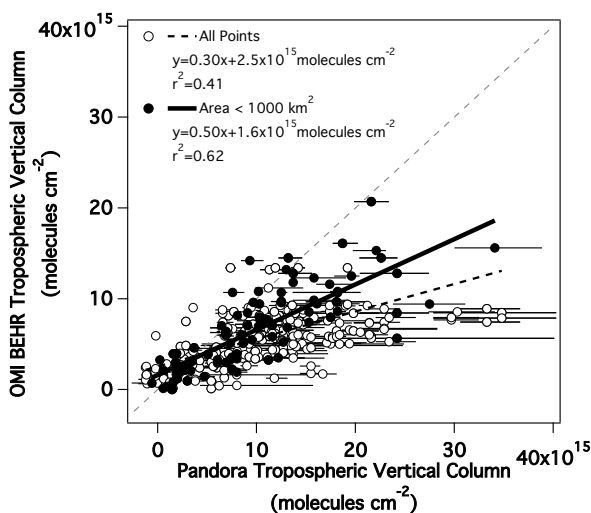
21

Figure 11: OMI Native Pixel BEHR NO₂ tropospheric vertical column data vs. Pandora NO₂ tropospheric column. The filled circles indicate pixel areas less than 1000 km². The linear fit to all points is shown in dashed black and the linear fit for pixels less than 1000 km² is shown in solid black. Horizontal whiskers show the maximum and minimum of Pandora data ± 5 minutes of the OMI overpass. The thin dashed line is the 1:1 line.

Considering all pixel areas, the BEHR TropVCs are moderately correlated with Pandora TropVCs ($r^2 = 0.41$) with a low slope (0.30) and are on average biased 50% lower than Pandora. Considering only pixels with areas less than 1000 km², correlation and slope improve by at least 50% ($r^2 = 0.62$, slope=0.5), again demonstrating the sensitivity of these direct comparisons to areal representativeness of the satellite pixel. Restricting OMI pixels to even smaller sizes (not shown) does not result in further statistical improvement, probably because the number of coincidences quickly decreases as the areal threshold decreases (40 coincidences when including only pixel areas below 500 km² and 23 coincidences for below 400 km²). The slopes shown in Figure 11 are notably lower than those calculated from the idealized GeoTASO-simulated idealized nadir OMI pixels discussed in Sect. 3.2, by 58% for all size pixels and 30% for pixel area < 1000 km². These differences are indicative of improvements associated with using higher spatial resolution a priori information, as used in the GeoTASO retrievals.

The horizontal whiskers in Figure 11 show the maximum and minimum in the Pandora measurement within the 10-minute temporal window of the OMI overpass. Some of these coincidences are impacted by very large variations in NO₂ measured by Pandora within a small window of the OMI overpass. This suggests that in addition to the magnitude of NO₂ as a potential heterogeneity screen discussed in Sect. 3.2, coincidence criteria based on the temporal variability of the fast measurements from Pandora could also help identify coincidences that may be impacted by small-scale highly heterogeneous environments.

To further explore sensitivity to AMF, Table 4 lists the linear regression statistics of OMI retrievals against Pandora observations for three products: the BEHR vertical column data shown in Fig. 11, the NASA Standard Product V3 vertical column (SP TropVC), and tropospheric slant column products from OMI and Pandora (TropSC) (Recall that BEHR also uses the SP slant column). Consistently comparing these slant and vertical column statistics with respect to Pandora observations provides an assessment of the influence of AMF a priori assumptions. For these Pandora coincidences, SP
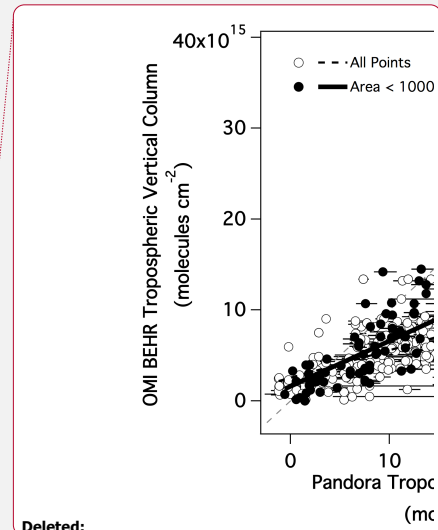
22

TropVCs are on average 51% lower than BEHR TropVCs, resulting in a 40% lower slope for SP TropVCs and Pandora. These comparisons demonstrate improvement due to incorporating the higher resolution a priori inputs in the BEHR product. Improvements to the spatial resolution of a priori input for OMI AMF calculations have previously been shown to help reduce biases (Russell et al., 2011; Lin et al., 2014). There is a minimal difference in r^2 when comparing Pandora to BEHR or SP TropVCs for OMI pixels less than 1000 km^2. But there is a dramatic loss in correlation between SP TropVCs and Pandora as pixel size increases, demonstrating that the impact of coarse a priori input worsens with increasing pixel size.

OMI tropospheric slant column (TropSC) comparisons with Pandora are slightly better than SP vertical column (SP TropVC) comparisons, indicating that the SP AMF is introducing additional variance in the SP TropVC product. The loss in skill associated with applying the SP AMF is evidenced by the 30% decrease in the slope for all coincidences and 22% decrease for pixels less than 1000 km^2 in comparison to TropSC comparisons. The application of the finer-resolution BEHR AMF largely retains the amount of skill in the TropSC with a slight improvement in slope.

Previous comparisons of OMI with Pandora direct-sun column measurements in urban environments have also shown low to moderate correlations for vertical-to-vertical column comparisons with low biases found in areas subject to spatial heterogeneity of NO_2 (Lamsal et al., 2014; Reed et al., 2015; Kim et al., 2016; Goldberg et al., 2017). Similar results have also been obtained from comparisons with ground-based MAX-DOAS retrievals (Lin et al., 2014; Chan et al., 2015). Monthly averaged OMI NASA SP V3 observations were also found to be biased low using Pandora observations in Korea (Herman et al., 2018) and MAX-DOAS observations in Hong Kong (Krotkov et al., 2017), but both studies found consistent seasonal trends in the satellite and surface-based column measurements. Not all comparison studies result in such a strong low bias associated with OMI. Boersma et al. (2018) found no low bias in OMI observations, but care was taken to limit influence by spatial variability and there were only 31 coincidences ranging from 5-10x10^{15} molecules cm^{-2}, which are pollution levels that are likely less influenced by extreme spatial heterogeneity. Lamsal et al. (2014) noted that MAX-DOAS/OMI bias (-16.3%) was not quite as low as reported by the Pandora comparisons in a study in the polluted (up to 40x10^{15} molecules cm^{-2}) Tsukubu region of Japan, which may be related to either the larger spatial footprint of MAX-DOAS observations (~10 km from Boersma et al., 2018) or different spatiotemporal thresholds appropriate for that specific region. For validation purposes, there should be value in combining direct-sun and all-sky (MAX-DOAS) measurements at the same locations. The MAX-DOAS measurements, with their ability to measure omni-directionally, can be more spatially representative albeit at the expense of coarser temporal resolution and higher measurement uncertainties, while the direct-sun observations are more frequent with very low measurement uncertainties but potentially larger influences of local scale variability.

Theoretically, satellite retrieval uncertainty could be better quantified with measurements from airborne simulators like GeoTASO that are capable of mapping entire satellite pixels. During LMOS/LA Basin flights, GeoTASO sampled inside 177 OMI pixels within ± 1 hour of the OMI overpass. However, during these 2017 flights, no OMI pixels were entirely mapped within a ± 1-hour period from the OMI overpass; the maximum OMI pixel area covered was approximately 80% in one coincidence and only 7 pixels were mapped by at least 50%. Therefore, GeoTASO data from these campaigns do not provide enough data to independently assess OMI bias. However, with the finer spatial resolutions of new platforms (e.g. TROPOMI and TEMPO), mapping satellite pixels within an acceptable temporal threshold from the overpass time should be more easily achieved, allowing an additional point of view between the space-based observations and ground-based platforms such as Pandora.

Table 4: Statistics between Pandora and OMI BEHR and SP TropVC and TropSC based on OMI pixel area.

OMI Pixel Area	Slope			Intercept (x10^{15} molecules cm^{-2})			r^2			N
	BEHR TropVC	SP TropVC	TropSC	BEHR TropVC	SP TropVC	TropSC	BEHR TropVC	SP TropVC	TropSC	
< 1000 km^2	0.50	0.34	0.44	1.6	1.2	1.3	0.62	0.61	0.65	125
All	0.30	0.18	0.26	2.5	1.9	2.4	0.41	0.32	0.39	444

4 Conclusions

During May-June 2017, an observing strategy was executed to build a number of airborne high spatial resolution gapless maps with GeoTASO over networks of ground-based instruments along the western shore of Lake Michigan and in the LA Basin. Each region had a network of five Pandora spectrometers (10 total) operating in direct-sun mode providing accurate NO$_2$ vertical column measurements at high temporal resolution with which to compare GeoTASO NO$_2$ TropVC retrievals. The sub-kilometer airborne NO$_2$ retrievals from GeoTASO are highly correlated with Pandora spectrometer observations with a slope near 1:1. Most of the apparent discrepancies between GeoTASO and Pandora TropVCs are associated with high variability and are therefore sensitive to assumptions made for identifying coincidences: e.g., the uni-directional Pandora viewing footprint vs. the omni-directional 750 m radius used to subset GeoTASO. At least one incidence was from a likely GeoTASO TropVC bias induced by inaccurate a priori in a complex environment that is difficult to simulate (a sea breeze frontal convergence zone). Future considerations for such comparisons should include assessments based on temporal and spatial criteria, as well as different environmental conditions. Despite these individual discrepancies, the GeoTASO retrievals accurately capture the spatiotemporal variability of NO$_2$ in these regions.

The raster mapping strategy at sub-kilometer resolution allows assessment of the impacts of sub-pixel heterogeneity within the larger areal pixel resolutions of past, present, and planned satellite instruments. Data from each raster are binned to fixed grids to simulate the pixel areas of nadir observations from TEMPO (3 x 3 km), TROPOMI (5 x 5 km) and OMI (18 x 18 km). Distinct spatial features (sea breeze front, industrial areas, etc.) within urban areas of Los Angeles and Chicago can be distinguished at TEMPO, and perhaps even TROPOMI, scale but cannot be resolved at the nadir area scale of OMI. The coarseness of the OMI scale can even alias the apparent spatial location of NO$_2$ maxima within urban regions. As pixel size increases, the linear statistics with Pandora degrade in these urban regions due to non-resolvable spatial heterogeneity that is most often associated with highly-polluted features that lie within only a portion of the larger pixels. This suggests that for validation purposes, there may be a practical limit on the magnitude of Pandora measurements that can be compared to satellite retrievals as occurrences of exceptionally high concentration NO$_2$ features appear to occur at spatial scales smaller than the satellite resolutions. In the regions of this study, this resolvable pollution scale appears to be near 20x10^{15} molecules cm^{-2} at OMI spatial resolution. TROPOMI and TEMPO resolution satellite products have the capability to compare well with Pandora at higher pollution scales, although more coincidences above 30x10^{15} molecules cm^{-2} would help refine this conclusion further. These thresholds provide only general guidance, as individual coincidences can still be influenced by spatial heterogeneity at any magnitude of pollution. However, if such environments are sampled enough, the variability could be statistically characterized sufficiently enough to allow identification of retrieval biases or errors.

Actual OMI observations show poorer agreement with Pandora direct-sun observations than is indicated by the idealized simulated satellite product. This is partially attributable to the decrease in OMI spatial resolution off-nadir, with some of the coincidences having pixel areas up to 1200% larger than the nadir pixel areas. When BEHR TropVC pixels are filtered to include only areas less than 1000 km^2, the Pandora linear regression slope improves by 66% and r^2 by 51%. The degraded real-world performance is also due to coarse information in the AMF calculation. An indication of this sensitivity is shown by the vertical column regression statistics against Pandora coincidences, in which the BEHR product has a better

24

slope than the NASA Standard Product. Further, the regression of slant column products is better than the SP vertical column, showing that uncertainties introduced by coarse AMF assumptions can even degrade the satellite NO$_2$ products.

These results reiterate the past challenges in using OMI observations to evaluate the magnitude of NO$_2$ pollution within urban areas, as the large footprint spatially averages NO$_2$ over an area that in reality has more fine-scale features. With the launch of TROPOMI in October 2017 and the geostationary sensors arriving in the next decade, their unprecedented spatial and temporal resolutions offer significantly improved capability for accurately assessing pollution within urban areas. Future validation work of satellite retrievals should consider using networks collecting frequent direct-sun Pandora observations spread throughout a wide-range of pollution environments that, through time, can collect enough data to statistically characterize effects of small-scale heterogeneity in order to evaluate satellite retrievals. Assessing temporal behavior of the Pandora observations (and the Geostationary observations when available) surrounding coincidence times can add additional information about whether or not any resultant difference between the measurements can be associated with spatiotemporal heterogeneity (e.g., the sea breeze front at LA Main Street in Fig. 6 and Fig. 7). Future small-scale campaigns similar to those reported here, combining high spatial resolution airborne mapping observations and high temporal resolution ground-based reference measurements with high-resolution modeling studies, provide a framework for validation of geostationary air quality measurements that will greatly enhance their usage for science and applications.

Author Contributions

LMJ processed the GeoTASO NO$_2$ retrievals and led writing of the manuscript. All co-authors provided feedback on the methodology and contributed to the final manuscript. JAA, RBP, and LMJ led flight planning activities for LMOS and/or the LA Basin Flights. SJJ, MGK, and LMJ collected the GeoTASO dataset during LMOS and/or the LA Basin flights, and SJJ and MGK calibrated and processed the L1b spectra for GeoTASO. RBP provided the NAM-CMAQ analysis used in the vertical column retrieval. JJS, LCV, DW set up and monitored the Pandora spectrometers during LMOS. RS and NA provided the Pandora spectrometers for the LA Basin and NA set up the LA Basin Pandora instruments. AC, MM, and MT processed the Pandora NO$_2$ retrievals.

Data Availability

All datasets are publicly available online: GeoTASO data are archived at https://www.air.larc.nasa.gov/cgi-bin/ArcView/lmos, Pandora data are found at data.pandonia.net, and OMI data are from Laughner et al. (2018b).

Acknowledgements

We would like to acknowledge NASA Earth Science Division's GEO-CAPE Mission Study for funding GeoTASO flights, the LMOS Science Teams, Caroline Nowlan and Gonzalo Gonzalez Abad at Harvard SAO for providing knowledge and support in the use of the SAO AMF calculation tool, the South Coast Air Quality and Monitoring District (SCAQMD) and colleagues at UCLA and CalTech for providing accommodations for the Pandora instruments in the LA Basin, our EPA colleagues and site hosts for maintaining Pandora instruments in the LMOS domain, NASA SARP 2017 and NSRC, Barry Lefer and the NASA Tropospheric Composition Program for inviting us to fly as part of the Student Airborne Research Program in the LA Basin, and the NASA Langley Research Center pilots and flight crew during both field missions.

References

Beirle, S., Boersma, K. F., Platt, U., Lawrence, M. G. and Wagner, T.: Megacity Emissions and Lifetimes of Nitrogen Oxides Probed from Space, Science, 333(6050), 1737–1739, doi:10.1126/science.1207824, 2011.

25

Deleted: , demonstrating the benefit of higher resolution a priori inputs to the AMF calculation in the two regions of this study.

Deleted: actually

Deleted: illustrate

Deleted: NO$_2$

Deleted: the

Deleted: Instruments such as GeoTASO will be helpful for simulating satellite datasets over urban regions, providing a transfer standard between the satellite observations and ground-based reference measurements from validation standards such as Pandora.

Deleted: is

Deleted: 2018

Boersma, K. F., Eskes, H. J., Dirksen, R. J., van der A, R. J., Veefkind, J. P., Stammes, P., Huijnen, V., Kleipool, Q. L., Sneep, M., Claas, J., Leitão, J., Richter, A., Zhou, Y. and Brunner, D.: An improved tropospheric NO_2 column retrieval algorithm for the Ozone Monitoring Instrument, Atmos. Meas. Tech., 4(9), 1905–1928, doi:10.5194/amt-4-1905-2011, 2011.

Boersma, K. F., Eskes, H. J., Richter, A., De Smedt, I., Lorente, A., Beirle, S., van Geffen, J. H. G. M., Zara, M., Peters, E., Van Roozendael, M., Wagner, T., Maasakkers, J. D., van der A, R. J., Nightingale, J., De Rudder, A., Irie, H., Pinardi, G., Lambert, J.-C. and Compernolle, S. C.: Improving algorithms and uncertainty estimates for satellite NO_2 retrievals: results from the quality assurance for the essential climate variables (QA4ECV) project, Atmos. Meas. Tech., 11(12), 6651–6678, doi:10.5194/amt-11-6651-2018, 2018.

Bourassa, A. E., McLinden, C. A., Sioris, C. E., Brohede, S., Bathgate, A. F., Llewellyn, E. J., and Degenstein, D. A.: Fast NO_2 retrievals from Odin-OSIRIS limb scatter measurements, Atmos. Meas. Tech., 4, 965-972, doi:10.5194/amt-4-965-2011, 2011.

Broccardo, S., Heue, K.-P., Walter, D., Meyer, C., Kokhanovsky, A., van der A, R., Piketh, S., Langerman, K. and Platt, U.: Intra-pixel variability in satellite tropospheric NO_2 column densities derived from simultaneous space-borne and airborne observations over the South African Highveld, Atmos. Meas. Tech., 11(5), 2797–2819, doi:10.5194/amt-11-2797-2018, 2018.

Bucsela, E. J., Krotkov, N. A., Celarier, E. A., Lamsal, L. N., Swartz, W. H., Bhartia, P. K., Boersma, K. F., Veefkind, J. P., Gleason, J. F. and Pickering, K. E.: A new stratospheric and tropospheric NO_2 retrieval algorithm for nadir-viewing satellite instruments: applications to OMI, Atmos. Meas. Tech., 6(10), 2607–2626, doi:10.5194/amt-6-2607-2013, 2013.

Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weißenmayer, A., Richter, A., DeBeek, R., Hoogen, R., Bramstedt, K. and Eichmann, K.-U.: The global ozone monitoring experiment (GOME): Mission concept and first scientific results, Journal of the Atmospheric Sciences, 56(2), 151–175, 1999.

Cede, A.: Manual for Blick Software Suite 1.3, version 7, available at: http://pandonia.net/media/documents/BlickSoftwareSuite_Manual_v7.pdf (last access: 26 March 2019), 2017.

CEOS: A Geostationary Satellite Constellation for Observing Global Air Quality: An International Path Forward, [online] Available from: http://ceos.org/document_management/Virtual_Constellations/ACC/Documents/AC-VC_Geostationary-Cx-for-Global-AQ-final_Apr2011.pdf (Accessed 6 March 2018), 2011.

Chan, K. L., Hartl, A., Lam, Y. F., Xie, P. H., Liu, W. Q., Cheung, H. M., Lampel, J., Pöhler, D., Li, A., Xu, J., Zhou, H. J., Ning, Z. and Wenig, M. O.: Observations of tropospheric NO_2 using ground based MAX-DOAS and OMI measurements during the Shanghai World Expo 2010, Atmospheric Environment, 119, 45–58, doi:10.1016/j.atmosenv.2015.08.041, 2015.

Chance, K. and Kurucz, R. L.: An improved high-resolution solar reference spectrum for earth's atmosphere measurements in the ultraviolet, visible, and near infrared, Journal of Quantitative Spectroscopy and Radiative Transfer, 111(9), 1289–1295, doi:10.1016/j.jqsrt.2010.01.036, 2010.

EPA: Technical Note – Guidance for Developing Enhanced Monitoring Plans [online] Available from: https://www3.epa.gov/ttnamti1/files/ambient/pams/PAMS%20EMP%20Guidance.pdf (Accessed 1 April 2019).

EPA: Air Quality Statistics Report. [online] Available from: https://www.epa.gov/outdoor-air-quality-data/air-quality-statistics-report (Accessed 18 March 2019).

5

10

15

20

25

30

35

40

45

Deleted: https://

Deleted: .org/

van Geffen, J. H. G., Eskes, H. J., Boersma, K. F., Maasakkers, J. D. and Veefkind, J. P.: TROPOMI ATBD of the total and tropospheric NO_2 data products, KNMI. [online] Available from: https://sentinel.esa.int/documents/247904/2476257/Sentinel-5P-TROPOMI-ATBD-NO2-data-products (Accessed 16 April 2019), 2019.

5 Goldberg, D. L., Lamsal, L. N., Loughner, C. P., Swartz, W. H., Lu, Z. and Streets, D. G.: A high-resolution and observationally constrained OMI NO_2 satellite retrieval, Atmos. Chem. Phys., 17(18), 11403–11421, doi:10.5194/acp-17-11403-2017, 2017.

10 Hastings, D. and Dunbar, P.: Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Documentation, Volume 1.0, National Oceanic and Atmospheric Administration, National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80303, USA, 1999.

Heckel, A., Kim, S.-W., Frost, G. J., Richter, A., Trainer, M. and Burrows, J. P.: Influence of low spatial resolution a priori data on tropospheric NO_2 satellite retrievals, Atmos. Meas. Tech., 4(9), 1805–1820, doi:10.5194/amt-4-1805-2011, 2011.

15 Herman, J., Cede, A., Spinei, E., Mount, G., Tzortziou, M. and Abuhassan, N.: NO_2 column amounts from ground-based Pandora and MFDOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation, Journal of Geophysical Research, 114(D13), doi:10.1029/2009JD011848, 2009.

20 Herman, J., Evans, R., Cede, A., Abuhassan, N., Petropavlovskikh, I. and McConville, G.: Comparison of ozone retrievals from the Pandora spectrometer system and Dobson spectrophotometer in Boulder, Colorado, Atmos. Meas. Tech., 8(8), 3407–3418, doi:10.5194/amt-8-3407-2015, 2015.

25 Herman, J., Spinei, E., Fried, A., Kim, J., Kim, J., Kim, W., Cede, A., Abuhassan, N., and Segal-Rozenhaimer, M.: NO_2 and HCHO measurements in Korea from 2012 to 2016 from Pandora spectrometer instruments compared with OMI retrievals and with aircraft measurements during the KORUS-AQ campaign, Atmos. Meas. Tech., 11, 4583–4603, https://doi.org/10.5194/amt-11-4583-2018, 2018.

30 IGACO: International Global Atmospheric Chemistry Observations Strategy Theme Report. (No. ESA SP-1282, GW No. 159, WMO TD No. 1235)., [online] Available from: http://www.fao.org/gtos/igos/docs/IGACO-Theme-Report-2004-4.pdf (Accessed 6 March 2018), 2004.

35 Ingmann, P., Veihelmann, B., Langen, J., Lamarre, D., Stark, H. and Courrèges-Lacoste, G. B.: Requirements for the GMES Atmosphere Service and ESA's implementation concept: Sentinels-4/-5 and -5p, Remote Sensing of Environment, 120, 58–69, doi:10.1016/j.rse.2012.01.023, 2012.

Judd, L. M., Al-Saadi, J. A., Valin, L. C., Pierce, R. B., Yang, K., Janz, S. J., Kowalewski, M. G., Szykman, J. J., Tiefengraber, M. and Mueller, M.: The Dawn of Geostationary Air Quality Monitoring: Case Studies From Seoul and Los Angeles, Front. Environ. Sci., 6, doi:10.3389/fenvs.2018.00085, 2018.

40 Kim, H. C., Lee, P., Judd, L., Pan, L. and Lefer, B.: OMI NO_2 column densities over North American urban cities: the effect of satellite footprint resolution, Geoscientific Model Development, 9(3), 1111–1123, doi:10.5194/gmd-9-1111-2016, 2016.

Kim, J., Kim, M. and Choi, M.: Monitoring Aerosol Properties in East Asia from Geostationary Orbit: GOCI, MI and GEMS, in Air Pollution in Eastern Asia: An Integrated Perspective, pp. 323–333, Springer, Cham., 2017.

45

Kleipool, Q. L., Dobber, M. R., Haan, J. F. de and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, Journal of Geophysical Research: Atmospheres, 113(D18), doi:10.1029/2008JD010290, 2008.

Deleted: Atmospheric Measurement Techniques,

Kowalewski, M. G. and Janz, S. J.: Remote sensing capabilities of the GEO-CAPE airborne simulator, Proc. SPIE, 9218, 92181I, doi:10.1117/12.2062058, 2014.

Krotkov, N. A., Lamsal, L. N., Celarier, E. A., Swartz, W. H., Marchenko, S. V., Bucsela, E. J., Chan, K. L., Wenig, M. and
5 Zara, M.: The version 3 OMI NO$_2$ standard product, Atmos. Meas. Tech., 10(9), 3133–3149, doi:10.5194/amt-10-3133-2017, 2017.

Lamsal, L. N., Krotkov, N. A., Celarier, E. A., Swartz, W. H., Pickering, K. E., Bucsela, E. J., Gleason, J. F., Martin, R. V., Philip, S., Irie, H., Cede, A., Herman, J., Weinheimer, A., Szykman, J. J. and Knepp, T. N.: Evaluation of OMI operational
10 standard NO$_2$ column retrievals using in situ and surface-based NO$_2$ observations, Atmos. Chem. Phys., 14(21), 11587–11609, doi:10.5194/acp-14-11587-2014, 2014.

Lamsal, L. N., Janz, S. J., Krotkov, N. A., Pickering, K. E., Spurr, R. J. D., Kowalewski, M. G., Loughner, C. P., Crawford, J. H., Swartz, W. H. and Herman, J. R.: High-resolution NO$_2$ observations from the Airborne Compact Atmospheric Mapper:
15 Retrieval and validation: High-Resolution NO$_2$ Observations, Journal of Geophysical Research: Atmospheres, 122(3), 1953–1970, doi:10.1002/2016JD025483, 2017.

Laughner, J. L., Zhu, Q., and Cohen, R. C.: The Berkeley High Resolution Tropospheric NO$_2$ product, Earth Syst. Sci. Data, 10, 2069–2095, doi:10.5194/essd-10-2069-2018, 2018a.
20

Laughner, J., Zhu, Q., and Cohen, R.: Berkeley High Resolution (BEHR) OMI NO$_2$ - Native pixels, monthly profiles, UC Berkeley Dash, Dataset, doi:10.6078/D1N086, 2018b.

Laughner, J. L., Zhu, Q., and Cohen, R. C.: Evaluation of version 3.0B of the BEHR OMI NO$_2$ product, Atmos. Meas.
25 Tech., 12, 129-146, doi:10.5194/amt-12-129-2019, 2019.

Lawrence, J. P., Anand, J. S., Vande Hey, J. D., White, J., Leigh, R. R., Monks, P. S. and Leigh, R. J.: High-resolution measurements from the airborne Atmospheric Nitrogen Dioxide Imager (ANDI), Atmospheric Measurement Techniques, 8(11), 4735–4754, doi:10.5194/amt-8-4735-2015, 2015.
30

Leitao, J., Richter, A., Vrekoussis, M., Kokhanovsky, A., Zhang, Q. J., Beekmann, M. and Burrows, J. P.: On the improvement of NO$_2$ satellite retrievals – aerosol impact on the airmass factors, Atmos. Meas. Tech., 3, 475–493, doi:10.5194/amt-3-475-2010, 2010.

35 Leitch, J. W., Delker, T., Good, W., Ruppert, L., Murcray, F., Chance, K., Liu, X., Nowlan, C., Janz, S. J., Krotkov, N. A., Pickering, K. E., Kowalewski, M., and Wang, J.: The GeoTASO airborne spectrometer project, Earth Observing Systems XIX, Proc. SPIE, 9218, 92181H–92181H–9, doi:10.1117/12.2063763, 2014.

Levelt, P. F., Oord, G. H. J. van den, Dobber, M. R., Malkki, A., Visser, H., Vries, J. de, Stammes, P., Lundell, J. O. V. and
40 Saari, H.: The ozone monitoring instrument, IEEE Transactions on Geoscience and Remote Sensing, 44(5), 1093–1101, doi:10.1109/TGRS.2006.872333, 2006.

Li, J., Chen, S., Qin, W., Murefu, M., Wang, Y., Yu, Y. and Zhen, Z.: Analysis of Accuracy of MODIS BRDF Product (MCD43 C6) Based on MISR Land Surface BRF Product – A Case Study of the Central Part of Northeast Asia, ISPRS -
45 International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII–3, 819–823, doi:10.5194/isprs-archives-XLII-3-819-2018, 2018.

Liang, J., Horowitz, L. W., Jacob, D. J., Wang, Y., Fiore, A. M., Logan, J. A., Gardner, G. M. and Munger, J. W.: Seasonal budgets of reactive nitrogen species and ozone over the United States, and export fluxes to the global atmosphere, J.
50 Geophys. Res., 103(D11), 13435–13450, doi:10.1029/97JD03126, 1998.

28

Deleted: https://

Deleted: .org/

Deleted: 2018

Deleted: https://

Deleted: .org/

Deleted: 19

Deleted: A., Pickering, K. E., Kowalewski, M. and Wang, J.: The GeoTASO airborne spectrometer project, edited by J. J. Butler, X. (Jack) Xiong, and X. Gu, p. 92181H., 2014.

Lin, J.-T., Martin, R. V., Boersma, K. F., Sneep, M., Stammes, P., Spurr, R., Wang, P., Van Roozendael, M., Clémer, K. and Irie, H.: Retrieving tropospheric nitrogen dioxide from the Ozone Monitoring Instrument: effects of aerosols, surface reflectance anisotropy, and vertical profile of nitrogen dioxide, Atmospheric Chemistry and Physics, 14(3), 1441–1461, doi:10.5194/acp-14-1441-2014, 2014.

Liu, F., Beirle, S., Zhang, Q., Dörner, S., He, K. and Wagner, T.: NOₓ lifetimes and emissions of cities and power plants in polluted background estimated by satellite observations, Atmos. Chem. Phys., 16(8), 5283–5298, doi:10.5194/acp-16-5283-2016, 2016.

Liu, X., Bhartia, P. K., Chance, K., Spurr, R. J. D. and Kurosu, T. P.: Ozone profile retrievals from the Ozone Monitoring Instrument, Atmos. Chem. Phys., 10, 2521-2537, doi: 10.5194/acp-10-2521-2010, 2010.

Lucht, W., Schaaf, C. B. and Strahler, A. H.: An algorithm for the retrieval of albedo from space using semiempirical BRDF models, IEEE Transactions on Geoscience and Remote Sensing, 38(2), 977–998, doi:10.1109/36.841980, 2000.

McLinden, C. A., Olsen, S. C., Hannegan, B., Wild, O., Prather, M. J. and Sundet, J.: Stratospheric ozone in 3-D models: A simple chemistry and the cross-tropopause flux, Journal of Geophysical Research: Atmospheres, 105(D11), 14653–14665, doi:10.1029/2000JD900124, 2000.

Meier, A. C., Schönhardt, A., Bösch, T., Richter, A., Seyler, A., Ruhtz, T., Constantin, D.-E., Shaiganfar, R., Wagner, T., Merlaud, A., Roozendael, M. V., Beleganté, L., Nicolae, D., Georgescu, L. and Burrows, J. P.: High-resolution airborne imaging DOAS measurements of NO₂ above Bucharest during AROMAT, Atmos. Meas. Tech., 10, 1831–1857, doi:10.5194/amt-10-1831-2017, 2017.

Noguchi, K., Richter, A., Rozanov, V., Rozanov, A., Burrows, J. P., Irie, H. and Kita, K.: Effect of surface BRDF of various land cover types on geostationary observations of tropospheric NO₂ Atmos. Meas. Tech., 7(10), 3497–3508, doi:10.5194/amt-7-3497-2014, 2014.

Nowlan, C. R., Liu, X., Leitch, J. W., Chance, K., González Abad, G., Liu, C., Zoogman, P., Cole, J., Delker, T., Good, W., Murcray, F., Ruppert, L., Soo, D., Follette-Cook, M. B., Janz, S. J., Kowalewski, M. G., Loughner, C. P., Pickering, K. E., Herman, J. R., Beaver, M. R., Long, R. W., Szykman, J. J., Judd, L. M., Kelley, P., Luke, W. T., Ren, X. and Al-Saadi, J. A.: Nitrogen dioxide observations from the Geostationary Trace gas and Aerosol Sensor Optimization (GeoTASO) airborne instrument: Retrieval algorithm and measurements during DISCOVER-AQ Texas 2013, Atmos. Meas. Tech., 9(6), 2647–2668, doi:10.5194/amt-9-2647-2016, 2016.

Nowlan, C. R., Liu, X., Janz, S. J., Kowalewski, M. G., Chance, K., Follette-Cook, M. B., Fried, A., González Abad, G., Herman, J. R., Judd, L. M., Kwon, H.-A., Loughner, C. P., Pickering, K. E., Richter, D., Spinei, E., Walega, J., Weibring, P., and Weinheimer, A. J.: Nitrogen dioxide and formaldehyde measurements from the GEOstationary Coastal and Air Pollution Events (GEO-CAPE) Airborne Simulator over Houston, Texas, Atmos. Meas. Tech., 11, 5941-5964, doi:10.5194/amt-11-5941-2018, 2018.

Palmer, P. I., Jacob, D. J., Chance, K., Martin, R. V., Spurr, R. J. D., Kurosu, T. P., Bey, I., Yantosca, R., Fiore, A. and Li, Q.: Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde retrievals from the Global Ozone Monitoring Experiment, J. Geophys. Res., 106(D13), 14539–14550, doi:10.1029/2000JD900772, 2001.

Pierce, R. B., T. Schaack, J. A. Al-Saadi, T. D. Fairlie, C. Kittaka, G. Lingenfelser, M. Natarajan, J. Olson, A. Soja, T. Zapotocny, A. Lenzen, J. Stobie, D. Johnson, M. A. Avery, G. W. Sachse, A. Thompson, R. Cohen, J. E. Dibb, J. Crawford, D. Rault, R. Martin, J. Szykman, and J. Fishman: Impacts of background ozone production on Houston and Dallas, Texas,

air quality during the Second Texas Air Quality Study field mission, J. Geophys. Res., 114, D00F09, doi:10.1029/2008JD011337, 2009.

Popp, C., Brunner, D., Damm, A., Van Roozendael, M., Fayt, C. and Buchmann, B.: High-resolution NO$_2$ remote sensing from the Airborne Prism EXperiment (APEX) imaging spectrometer, Atmospheric Measurement Techniques, 5(9), 2211–2225, doi:10.5194/amt-5-2211-2012, 2012.

Prather, M.: Catastrophic loss of stratospheric ozone in dense volcanic clouds, Journal of Geophysical Research: Atmospheres, 97(D9), 10187–10191, doi:10.1029/92JD00845, 1992.

Reed, A. J., Thompson, A. M., Kollonige, D. E., Martins, D. K., Tzortziou, M. A., Herman, J. R., Berkoff, T. A., Abuhassan, N. K. and Cede, A.: Effects of local meteorology and aerosols on ozone and nitrogen dioxide retrievals from OMI and pandora spectrometers in Maryland, USA during DISCOVER-AQ 2011, J Atmos. Chem., 72(3–4), 455–482, doi:10.1007/s10874-013-9254-9, 2015.

Rothman, L. S., Gordon, I. E., Barbe, A., Benner, D. C., Bernath, P. F., Birk, M., Boudon, V., Brown, L. R., Campargue, A., Champion, J.-P., Chance, K., Coudert, L. H., Dana, V., Devi, V. M., Fally, S., Flaud, J.-M., Gamache, R. R., Goldman, A., Jacquemart, D., Kleiner, I., Lacome, N., Lafferty, W. J., Mandin, J.-Y., Massie, S. T., Mikhailenko, S. N., Miller, C. E., Moazzen-Ahmadi, N., Naumenko, O. V., Nikitin, A. V., Orphal, J., Perevalov, V. I., Perrin, A., Predoi-Cross, A., Rinsland, C. P., Rotger, M., Šimečková, M., Smith, M. A. H., Sung, K., Tashkun, S. A., Tennyson, J., Toth, R. A., Vandaele, A. C. and Vander Auwera, J.: The HITRAN 2008 molecular spectroscopic database, Journal of Quantitative Spectroscopy and Radiative Transfer, 110(9–10), 533–572, doi:10.1016/j.jqsrt.2009.02.013, 2009.

Russell, A. R., Perring, A. E., Valin, L. C., Bucsela, E. J., Browne, E. C., Wooldridge, P. J. and Cohen, R. C.: A high spatial resolution retrieval of NO$_2$ column densities from OMI: method and evaluation, Atmos. Chem. Phys., 11(16), 8543–8554, doi:10.5194/acp-11-8543-2011, 2011.

Schaaf, C., Z. Wang: MCD43A1 MODIS/Terra+Aqua BRDF/Albedo Model Parameters Daily L3 Global - 500m V006, doi:10.5067/modis/mcd43a1.006, 2015.

Schönhardt, A., Altube, P., Gerilowski, K., Krautwurst, S., Hartmann, J., Meier, A. C., Richter, A. and Burrows, J. P.: A wide field-of-view imaging DOAS instrument for two-dimensional trace gas mapping from aircraft, Atmos. Meas. Tech., 8(12), 5113–5131, doi:10.5194/amt-8-5113-2015, 2015.

Souri, A. H., Choi, Y., Pan, S., Curci, G., Nowlan, C. R., Janz, S. J., Kowalewski, M. G., Liu, J., Herman, J. R. and Weinheimer, A. J.: First Top-Down Estimates of Anthropogenic NO$_x$ Emissions Using High-Resolution Airborne Remote Sensing Observations, Journal of Geophysical Research: Atmospheres, 123(6), 3269–3284, doi:10.1002/2017JD028009, 2018.

Spurr, R. J. D.: VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media, Journal of Quantitative Spectroscopy and Radiative Transfer, 102(2), 316–342, doi:10.1016/j.jqsrt.2006.05.005, 2006.

Stajner, I., Davidson, P., Byun, D., McQueen, J., Draxler, R., Dickerson, P. and Meagher, J.: US National Air Quality Forecast Capability: Expanding Coverage to Include Particulate Matter, in Air Pollution Modeling and its Application XXI, edited by D. G. Steyn and S. Trini Castelli, pp. 379–384, Springer Netherlands., 2011.

Tack, F., Merlaud, A., Iordache, M.-D., Danckaert, T., Yu, H., Fayt, C., Meuleman, K., Deutsch, F., Fierens, F. and Van Roozendael, M.: High-resolution mapping of the NO$_2$ spatial distribution over Belgian urban areas based on airborne APEX remote sensing, Atmos. Meas. Tech., 10, 1665–1688, doi:10.5194/amt-10-1665-2017, 2017.

Deleted: C.

Deleted: W.:

Deleted:

Tack, F., Merlaud, A., Meier, A. C., Vlemmix, T., Ruhtz, T., Iordache, M.-D., Ge, X., van der Wal, L., Schuettemeyer, D., Ardelean, M., Calcan, A., Constantin, D., Schönhardt, A., Meuleman, K., Richter, A. and Van Roozendael, M.: Intercomparison of four airborne imaging DOAS systems for tropospheric NO2 mapping – the AROMAPEX campaign, Atmos. Meas. Tech., 12(1), 211–236, doi:10.5194/amt-12-211-2019, 2019.

Thalman, R. and Volkamer, R.: Temperature dependent absorption cross-sections of O2–O2 collision pairs between 340 and 630 nm and at atmospherically relevant pressure, Physical Chemistry Chemical Physics, 15(37), 15371, doi:10.1039/c3cp50968k, 2013.

Valin, L. C., Russell, A. R., Bucsela, E. J., Veefkind, J. P., and Cohen, R. C.: Observation of slant column NO2 using the super-zoom mode of AURA-OMI, Atmos. Meas. Tech., 4, 1929-1935, doi:10.5194/amt-4-1929-2011, 2011a.

Valin, L. C., Russell, A. R., Hudman, R. C., and Cohen, R. C.: Effects of model resolution on the interpretation of satellite NO2 observations, Atmos. Chem. Phys., 11, 11647-11655, doi:10.5194/acp-11-11647-2011, 2011b.

Vandaele, A. C., Hermans, C., Simon, P. C., Carleer, M., Colin, R., Fally, S., Mérienne, M. F., Jenouvrier, A. and Coquart, B.: Measurements of the NO2 absorption cross-section from 42 000 cm−1 to 10 000 cm−1 (238–1000 nm) at 220 K and 294 K, Journal of Quantitative Spectroscopy and Radiative Transfer, 59(3–5), 171–184, doi:10.1016/S0022-4073(97)00168-4, 1998.

Volkamer, R., Spietz, P., Burrows, J. and Platt, U.: High-resolution absorption cross-section of glyoxal in the UV–vis and IR spectral ranges, Journal of Photochemistry and Photobiology A: Chemistry, 172(1), 35–46, doi:10.1016/j.jphotochem.2004.11.011, 2005.

Zoogman, P., Liu, X., Suleiman, R. M., Pennington, W. F., Flittner, D. E., Al-Saadi, J. A., Hilton, B. B., Nicks, D. K., Newchurch, M. J., Carr, J. L., Janz, S. J., Andraschko, M. R., Arola, A., Baker, B. D., Canova, B. P., Chan Miller, C., Cohen, R. C., Davis, J. E., Dussault, M. E., Edwards, D. P., Fishman, J., Ghulam, A., González Abad, G., Grutter, M., Herman, J. R., Houck, J., Jacob, D. J., Joiner, J., Kerridge, B. J., Kim, J., Krotkov, N. A., Lamsal, L., Li, C., Lindfors, A., Martin, R. V., McElroy, C. T., McLinden, C., Natraj, V., Neil, D. O., Nowlan, C. R., O'Sullivan, E. J., Palmer, P. I., Pierce, R. B., Pippin, M. R., Saiz-Lopez, A., Spurr, R. J. D., Szykman, J. J., Torres, O., Veefkind, J. P., Veihelmann, B., Wang, H., Wang, J. and Chance, K.: Tropospheric emissions: Monitoring of pollution (TEMPO), Journal of Quantitative Spectroscopy and Radiative Transfer, 186, 17–39, doi:10.1016/j.jqsrt.2016.05.008, 2017.

Deleted: https://

Deleted: .org/

Deleted: https://

Deleted: .org/