Satellite remote-sensing capability to assess tropospheric column ratios of formaldehyde and nitrogen dioxide: case study during the LISTOS 2018 field campaign

Matthew S. Johnson[1], Amir H. Souri[2], Sajeev ~~Philip[2]~~Philip[3], Rajesh ~~Kumar[3]~~Kumar[4], Aaron ~~Naeger[4]~~Naeger[5], ~~Amir H. Souri[5],~~ Jeffrey Geddes[6], Laura Judd[7], Scott Janz[8], Heesung Chong[2], John Sullivan[8]

[1]Earth Science Division, NASA Ames Research Center, Moffett Field, CA 94035, USA.
[2]Atomic and Molecular Physics (AMP) Division, Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA, USA.
~~[2]Centre~~ [3]Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, Jia Sarai, Hauz Khas, New Delhi, Delhi 110016, India.
~~[3]Research~~ [4]Research Applications Laboratory, National Center for Atmospheric Research, Boulder, CO 80305, USA.
~~[4]Short~~[5]Short-term Prediction Research and Transition Center, University of Alabama in Huntsville, Huntsville, AL 35805, USA.
~~[5]Atomic and Molecular Physics (AMP) Division, Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA, USA.~~
[6]Earth and Environment Department, Boston University, Boston, MA, 02215, USA.
[7]NASA Langley Research Center, Hampton, VA 23681, USA.
[8]NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

Correspondence to: Matthew S. Johnson (matthew.s.johnson@nasa.gov)

Abstract. Satellite retrievals of tropospheric column formaldehyde (HCHO) and nitrogen dioxide (NO_2) are frequently used to investigate the sensitivity of ozone (O_3) production to concentrations and emissions of nitrogen oxides (NO_x) and volatile organic carbon compounds (VOCs). ~~Space-based remote sensing information of chemical proxies for NO_x (i.e., NO_2) and VOCs (i.e., HCHO), in particular the ratios of tropospheric column HCHO and NO_2 (FNRs), provide insight into the non-linear relationship of O_3 formation in the lower troposphere. Ultraviolet–visible (UV/VIS) satellite spectrometers such as the Ozone Monitoring Instrument (OMI) and TROPOspheric Monitoring Instrument (TROPOMI) are capable of providing FNR information with high spatiotemporal coverage, yet a recent study suggested that the biases and noise of satellite retrievals are the largest source of uncertainty for applying satellite-derived FNRs to better understand O_3 production sensitivities. To~~This study quantified~~y~~, and inter-compared, the systematic biases and uncertainties in retrievals of NO_2 and HCHO tropospheric columns, and resulting HCHO to NO_2 ratios (FNRs), from two of the most commonly-applied satellite sensors to investigate O_3 production sensitivities (Ozone Monitoring Instrument (OMI) and TROPOspheric Monitoring Instrument (TROPOMI)) ~~, we evaluated OMI and TROPOMI retrievals of NO_2 and HCHO tropospheric columns, and resulting FNRs,~~ using Geostationary Trace gas and Aerosol Sensor Optimization (GeoTASO) and GEO-CAPE Airborne Simulator (GCAS) airborne remote-sensing data taken during the Long Island Sound Tropospheric Ozone Study 2018 (LISTOS 2018) between June 25 to September 6, 2018. Compared to ~~suborbital~~ aircraft-based remote-sensing observations of tropospheric column NO_2 and HCHO, the accuracy of OMI (using both the National Aeronautics and Space Administration (NASA) version 4 and the Quality Assurance for Essential Climate Variables (QA4ECV) retrieval algorithms) and TROPOMI were magnitude-dependent with high biases ~~(i.e., satellite tropospheric columns > suborbital tropospheric columns)~~ in clean/background environments and a tendency towards more accurate comparisons to even ~~a low bias~~low biases ~~(i.e., satellite tropospheric columns < suborbital tropospheric columns)~~ in moderately- to polluted-regions~~moderate to polluted regions~~. Campaign-averaged NO_2 ~~median~~ systematic biases for OMI, using both the NASA and QA4ECV algorithms, and TROPOMI were similar in magnitude ~~at~~ (~~±~~0.3-0.4~~±4.1~~ × 10^{15} molecules cm^{-2}) ~~(6.3%) and 0.4±4.5 × 10^{15} molecules cm^{-2} (6.8%), respectively~~with OMI displaying a slight high median bias and TROPOMI resulting in small low biases. Uncertainties in the three satellite retrieval products for NO_2 were similar (~4.0-4.5 × 10^{15} molecules cm^{-2}) with TROPOMI retrievals having slightly less spread in the data compared to OMI products. ~~TROPOMI retrievals of NO_2 had a campaign-averaged median bias of -0.3±3.7 × 10^{15} molecules cm^{-2} (-4.8%) and 0.3±3.3 × 10^{15} molecules cm^{-2} (5.8%) when averaged at finer (0.05° × 0.05°) and coarser (0.15° × 0.15°) spatial resolution.~~ The three satellite products (NASA OMI, QA4ECV OMI, and TROPOMI) differed more when evaluating tropospheric column HCHO retrievals. Noise in the HCHO retrievals, likely due to low signal-to-noise ratios and the fact the ultraviolet–visible wavelength~~UV/VIS~~ measurement sensitivity at shorter wavelengths used in HCHO retrievals are low in the troposphere, resulted in low correlations and high ~~oscillation/variability in bias (bias standard deviation)~~ root mean squared error (RMSE) values in all three satellite products~~.~~. C~~with c~~ampaign-averaged tropospheric HCHO median biases were~~of~~ 5.1±7.8 × 10^{15} ~~molecules cm^{-2} (38.7%)~~, 2.3±8.9 × 10^{15} ~~molecules cm^{-2} (17.3%)~~, 1.9±6.7 × 10^{15} ~~molecules cm^{-2} (12.9%)~~, and 2.9±4.9 × 10^{15} molecules cm^{-2} ~~(23.1%)~~ for NASA OMI, QA4ECV OMI, and TROPOMI at finer (0.05° × 0.05°) and coarser (0.15° × 0.15°) spatial resolution, respectively. Uncertainty values in NASA and QA4ECV OMI HCHO retrievals were ~9.0 × 10^{15} molecules cm^{-2} and the higher spatial resolution and sufficient

signal-to-noise retrievals from TROPOMI resulted in HCHO RMSE values ~30% lower compared to OMI. Spatially-averaging TROPOMI tropospheric column HCHO, along with NO_2 and FNRs, to coarser resolutions similar to the OMI native pixel size proved to reduce the bias standard deviation and uncertainty of the retrieval data. Systematic biases in OMI and TROPOMI NO_2 and HCHO retrievals tended to cancel out resulting in all three satellite products comparing well to observed FNRs (campaign-averaged median biases <0.4). However, while satellite-derived FNRs had minimal campaign-averaged median biases, unresolved errors in the indicator species did not cancel out in FNR calculations resulting in large RMSE values compared to observations. Uncertainties in HCHO retrievals were determined to drive the unresolved biases in FNR retrievals. Given that TROPOMI has less noise in the HCHO retrievals compared to OMI, the higher spatial resolution sensor had less bias standard deviation and RMSE in FNR values. Finally, this study discusses a) the capabilities of OMI and TROPOMI to retrieve spatiotemporal variability of FNRs and potential sources for the systematic errors and uncertainty of these sensor's retrievals, b) importance of a priori vertical profiles of NO_2 and HCHO used in trace gas retrievals, and c) different aspects of retrieval algorithms which impact OMI and TROPOM FNR retrieval errors.

~~While large median biases, and enhanced variability in bias, were derived for HCHO, errors in both NO_2 and HCHO tropospheric columns tended to offset as all three satellite products compared well to observed FNRs with campaign-averaged median biases from NASA OMI, QA4ECV OMI, and TROPOMI of 0.4±3.8 (11.0%), -0.2±3.3 (5.4%), and 0.4±2.3 (13.0%), respectively. While satellite derived FNRs had minimal campaign averaged median biases, the statistical analysis shows that all satellite FNR values still had large bias standard deviation due to unresolved errors in satellite retrievals of HCHO. This result is important as accurate retrievals (minimal median biases) of FNRs from satellites do not suggest the accuracy of the underlying proxy species. The reduction in noise in satellite retrievals of HCHO with additional calibration and improved sensor design and/or improved a priori information of the vertical profiles of HCHO in the troposphere to avoid the impact of the low measurement sensitivity in the shorter UV/VIS wavelengths used to retrieve HCHO is critical for reducing unresolved biases in satellite retrievals of FNRs. Furthermore, this work demonstrates the large impact of a) a priori vertical profiles of NO_2 and HCHO for calculations of Air Mass Factors in tropospheric column trace gas retrievals in both OMI and TROPOMI, b) spatiotemporal averaging to increase signal-to-noise, and c) different retrieval algorithms on retrieval errors. Finally, the novel diurnal information of tropospheric FNRs that is expected to be provided by the upcoming NASA geostationary sensor Tropospheric Emissions: Monitoring of Pollution (TEMPO) is investigated and compared to low earth orbiting sensors currently applied to investigate tropospheric FNRs.~~

1 Introduction

Tropospheric ozone (O_3) is a harmful pollutant and near-surface concentrations of this species have detrimental impacts on human- and environmental-health (Kampa and Castanas, 2008; Van Dingenen et al., 2009). The production and destruction rates of tropospheric O_3 are controlled by complex chemical reactions involving the primary precursor species of nitrogen oxides (NO_x = nitric oxide and nitrogen dioxide (NO + NO_2)) and volatile organic compounds (VOCs) (Sillman, 1999; Lelieveld and Dentener, 2000). It is critical to understand precursor species emissions and subsequent atmospheric chemistry controlling surface-level O_3 production rates since the United States (US) Environmental Protection Agency (EPA) designs and enforces concentration limits of criteria pollutants (e.g., O_3, NO_2, carbon monoxide, particulate matter, and sulfur dioxide) under the National Ambient Air Quality Standards (NAAQS)~~. The current NAAQS for O_3 requires that 3-year averaged annual fourth highest daily maximum 8-hour mean concentrations be $\leq$ 70 ppb~~ (US EPA, 2015). To reduce and maintain surface-level O_3 concentrations below NAAQS thresholds, many regions have designed and implemented emission control strategies of precursor species. To design effective emission reduction strategies, knowledge about the non-linear sensitivity of O_3 formation to NO_x and VOCs is critical (Crutzen, 1973; Sillman, 1999). Based on the relative concentrations of NO_x and VOCs, the formation of O_3 is sensitive to perturbations of either NO_x (NO_x-limited regimes) or VOC emissions (NO_x-saturated or VOC/radical-limited regimes). These O_3 sensitivity regimes are separated by a transitional regime where O_3 formation is sensitive to changes in both NO_x and VOC emissions.

To understand the non-linear relationship of O_3 formation to NO_x and VOC emissions in complex chemical environments (e.g., polluted regions and areas of heterogeneous~~heterogenous~~ concentrations/emissions of NO_x and VOCs), spatiotemporally dense in situ measurements or airborne remote-sensing observations of precursor species concentrations and chemical reactivity are desired (e.g., Souri et al., 2020). Since these measurements are often spatiotemporally sparse, to supplement the time and space void of these observations, thoroughly evaluated model simulations can be applied. However, the accuracy of chemical transport models (CTMs) is highly dependent on inputs such as emission inventories, simulated meteorology, chemistry mechanisms, and removal processes all of which have varying levels of uncertainty. These model uncertainties can directly impact the understanding of the non-linear relationship of O_3 formation when using these simulated data (e.g., Choi and Souri, 2015). In the absence of accurate in situ measurements or high spatiotemporal aircraft-based~~suborbital~~ remote-sensing information of chemical proxies for NO_x (i.e., NO_2) and VOCs (i.e., formaldehyde (HCHO)), satellite retrievals of these species have also been demonstrated to provide insight into the O_3-NO_x-VOC relationship (Tonnesen and Dennis, 2000; Martin et al., 2004; Duncan et al., 2010; Souri et al., 2017; Jin et al., 2017, 2020). The ratio of HCHO to NO_2 concentrations (hereinafter FNR) has been demonstrated to provide information to monitor the local sensitivity of O_3 production from the chemical loss of HO_2+RO_2 (LRO_x) and chemical loss of NO_x (LNO_x) controlling O_3-NO_x-VOC chemistry (Tonnesen and Dennis, 2000; Kleinman et al., 2001).

Multiple past and current space-based spectrometers have the capability to retrieve simultaneous NO_2 and HCHO tropospheric columns to calculate FNRs for studying O_3 production sensitivity regimes including Global Ozone Monitoring Experiment (GOME, Martin et al., 2004), GOME-2 (Choi et al., 2012), Ozone Monitoring Instrument (OMI, Duncan et al., 2010), SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY

(SCIAMACHY, Jin et al., 2020), and TROPOspheric Monitoring Instrument (TROPOMI, Chan et al., 2020, Souri et al., 2021). In addition to these low earth orbiting (LEO) satellites, Tropospheric Emissions: Monitoring of Pollution (TEMPO) is an upcoming National Aeronautics and Space Administration (NASA) geostationary satellite mission which will retrieve hourly NO_2 and HCHO tropospheric columns over North America (Zoogman et al., 2017; Chance et al., 2019). This geostationary sensor over North America is part of a constellation of air quality spaceborne sensors including the Geostationary Environment Monitoring Spectrometer (GEMS) instrument onboard the Korean Aerospace Research Institute GEO-KOMPSAT-2B satellite (Kim et al., 2020) and the European Space Agency (ESA) Sentinel-4 mission (ESA, 2017). Satellite retrievals of NO_2 and HCHO have been applied to determine the sensitivity of O_3 formation to NO_x and VOC emissions at coarse spatial and temporal scales (e.g., Martin et al., 2004; Duncan et al., 2010) to finer spatiotemporal scales and focusing on long-term trends (e.g., Choi et al., 2012; Jin and Holloway, 2015; Choi and Souri, 2015; Schroeder et al., 2017; Souri et al., 2017; Jin et al., 2017, 2020). However, uncertainties remain in how accurately satellites can retrieve information needed to study surface-level or planetary boundary layer (PBL) O_3-NO_x-VOC relationships. These uncertainties stem from a) the exact thresholds of FNRs that separate NO_x-limited, transition, and VOC-limited regimes, b) the ability of tropospheric column retrievals to represent PBL chemical composition for air quality purposes due to variability in the vertical structure of NO_2 and HCHO concentrations and satellite sensitivity throughout the entire troposphere, c) whether HCHO is an effective proxy for total VOC reactivity, d) satellite spatial representation errors, and e) the accuracy/uncertainty of satellite retrievals of tropospheric column HCHO and NO_2. Of all these sources of uncertainty, mean/median and random biases due to noise in satellite retrievals of tropospheric column HCHO and NO_2 may be the largest source of error for retrieving FNRs using satellite sensors (Souri et al., 2022a). Therefore, it is vital to accurately define the level of errors/biases associated with satellite sensors to understand the capability of using this spatiotemporally-dense data source for investigating the impact of NO_x and VOC emission perturbations on O_3 chemistry.

This study is designed to demonstrate the effectiveness of two frequently applied satellites for evaluating O_3-NO_x-VOC relationships (i.e., OMI and TROPOMI) to accurately retrieve tropospheric HCHO and NO_2 column concentrations and ~~the subsequent tropospheric column~~ FNRs. OMI and TROPOMI retrievals have been evaluated in numerous studies (e.g., Judd et al., 2020; Vigouroux et al., 2020; Zhu et al., 2020; Lamsal et al., 2021), typically focusing on a specific sensor and species (e.g., evaluating OMI or TROPOMI and NO_2 or HCHO separately); however, not for the accuracy to retrieve tropospheric column FNRs. Here we validate OMI and TROPOMI retrievals of HCHO and NO_2, and subsequent FNRs, with airborne spectrometer data obtained during the Long Island Sound Tropospheric Ozone Study 2018 (LISTOS 2018) field campaign conducted during the summer of 2018 in the northeast region of the US. Furthermore, this work ~~demonstrates the additional information of tropospheric FNRs that is expected to be provided by the upcoming NASA geostationary sensor TEMPO~~investigates the capability of OMI and TROPOMI to capture the spatiotemporal variability of observed FNRs and discusses the possible causes of systematic error and uncertainties in these retrievals. The manuscript is designed as follows. Section 2 presents the satellite, airborne remote-sensing, model data~~.~~, and evaluation techniques ~~applied~~ used in this study. The results and discussion are reported in Sect. 3 and the final conclusions are presented in Sect. 4.

2 Methods

This study focuses on the spatial domain and time period (June 25 to September 6, 2018) of the LISTOS 2018 field campaign (https://www.nescaum.org/documents/listos~~; https://www-air.larc.nasa.gov/missions/listos/index.html~~) ~~field campaign~~. This campaign was chosen due to the overlap of the TROPOMI and OMI missions, the availability of airborne spectrometer retrievals (i.e., Geostationary Trace gas and Aerosol Sensor Optimization (GeoTASO) and GEO-CAPE Airborne Simulator (GCAS)) of tropospheric column HCHO and NO_2 which are effective satellite validation data (e.g., Judd et al., 2020), and the large spatiotemporal coverage of the airborne spectrometer data. ~~Many S~~studies have applied stationary sources of ground-based remote-sensing data to validate OMI and TROPOMI (e.g., MAX-DOAS, FTIR, Pandora); however, using the airborne GeoTASO and GCAS products allows for the evaluation of the satellite retrievals in variable environments (i.e., clean/background to heterogeneous~~heterogenous~~/polluted regions) in the same day. The rest of this section describes the remote-sensing and model data applied in this study for evaluation of tropospheric column HCHO and NO_2 from OMI and TROPOMI.

2.1 OMI remote-sensing products

The Dutch-Finnish nadir viewing spectrometer OMI, onboard the polar-orbiting NASA Aura satellite, which was launched in 2004, is an ultraviolet–visible (UV/Vis) spectrometer (Levelt et al., 2006). Retrievals are made from three wavelength channels between 260 to 510 nm (UV-1: 264 to 311 nm, UV-2: 307 to 383 nm, Vis: 349 to 504 nm). Aura-OMI has a local equatorial overpass time of ~13:45 with nearly-complete daily global surface coverage due to the large ~2,600 km swath width. Level-2 (L2) tropospheric vertical column density (VCD) OMI NO_2 retrievals from the NASA version 4 standard product (OMNO2; Lamsal et al., 2021) and the NASA operational OMI HCHO version 3 product using the Smithsonian Astrophysical Observatory (SAO) retrieval algorithm (OMHCHO; González Abad et al., 2015, 2016) were applied in this study. To investigate the impact of different retrieval algorithms, we also apply tropospheric column OMI NO_2 and HCHO data derived in the Quality Assurance for Essential Climate Variables (QA4ECV) project (see Sect. 2.1.2).

Starting in 2007, OMI experienced a field-of-view blockage known as the "row anomaly" which affects the data quality at all retrieval wavelengths for some rows (Dobber et al., 2008; Schenkeveld et al., 2017). The row anomaly in NO_2 and HCHO retrievals ~~is~~ was avoided in this study ~~by~~ using data quality flags to ~~filtering~~filter out rows/pixels flagged by the row anomaly detection algorithm. ~~The postprocessing bias correction approach using the reference sector method for OMI HCHO is applied here and corrects for the row anomaly in HCHO data (De Smedt et al., 2015).~~ OMI data also has systematically~~systematic~~ biased retrievals in a striped pattern running in 60 cross-track field-of-views. A "de-striping" correction is already applied to the NO_2 data (Boersma et al., 2011) and the reference sector method corrects for these artifacts in the HCHO data (De Smedt et al., 2015; González Abad et al., 2015; Zara et al., 2018).

2.1.1 OMI – NASA OMNO2 and OMHCHO

The primary OMI data applied in this study are the L2 tropospheric VCD OMNO2 and OMHCHO retrievals provided at ~13 km × 24 km near nadir to ~24 km × 160 km towards the edge of the swath. Lamsal et al. (2021) describes the

OMNO2 retrieval algorithm in detail and is explained here only briefly (referred to as NASA OMI NO_2 throughout). The NASA OMI NO_2 retrieval uses a differential optical absorption spectroscopy (DOAS) approach, with a fitting window between 405 and 465 nm, to derive slant column densities (SCD) of NO_2. Tropospheric NO_2 columns are separated from the entire atmospheric column using an observation-based stratosphere–troposphere separation scheme described in Bucsela et al. (2013). Tropospheric SCDs are then converted to tropospheric VCDs using an Air Mass Factor (AMF) calculated with a radiative transfer model and simulated atmospheres from a CTM. The AMF is an integrated product of scattering weights ~~(SWs)~~ and trace gas profile shapes (Palmer et al., 2001). Specifics for the basic data used in AMF calculations for NASA OMI NO_2 are presented in Table S1. Tropospheric AMFs are calculated in NASA OMI NO_2 retrievals using monthly-averaged a priori profiles from the NASA Global Modelling Initiative (GMI) model at 1.0° × 1.25° spatial resolution, clouds from the OMI O_2–O_2 algorithm (Vasilkov et al., 2018), and surface albedo from geometry-dependent surface Lambertian equivalent reflectivity (GLER) data (Vasilkov et al., 2017; Qin et al., 2019; Fasnacht et al., 2019). The uncertainty of the tropospheric NASA OMI NO_2 product has been shown to vary with cloudiness and pollution concentrations and is in the range of ~20% to ~60% (Bucsela et al., 2013), with contributions from errors in spectral fitting, stratospheric correction, and AMF calculations.

González Abad et al. (2015, 2016) describes the OMHCHO retrieval algorithm in detail (referred to as NASA OMI HCHO throughout). Briefly, retrievals of HCHO SCDs are obtained by spectrally fitting OMI radiances using the basic optical absorption spectroscopy (BOAS) method (Chance, 1998) with a fitting window between 328.5 and ~~346~~356.~~0~~ 5 nm. ~~Then, like NASA OMI NO_2 retrievals,~~HCHO SCDs are converted to VCDs applying derived AMFs using GEOS-Chem a priori profiles at 2.0° × 2.5° spatial resolution, cloud information (Martin et al., 2002; Acarreta et al., 2004), and surface albedo data (Kleipool et al., 2008~~Vasilkov et al., 2014~~). Information about the basic input data for AMF calculations for the NASA OMI HCHO retrieval are presented in Table S1. Finally, postprocessing across-track bias corrections are applied by comparing daily HCHO VCDs with background VCDs simulated with the GEOS-Chem CTM over a clean region (known as the reference sector). The uncertainty of the NASA OMI HCHO product has been shown to vary with pollution concentration ranging from ~45% to ~105% with largest contributions from the spectral fitting and AMF calculations (González Abad et al., 2015, 2016).

2.1.2 OMI – QA4ECV NO_2 and HCHO

For comparison to the NASA OMI retrieval products, we inter-compared and evaluated OMI NO_2 and HCHO retrievals from the QA4ECV project (www.qa4ecv.eu). Retrievals from the QA4ECV NO_2 version 1.1 and QA4ECV HCHO version 1.2 data products are applied in this study and are provided daily at the same spatial resolution as the NASA OMI products (~13 km × 24 km near nadir to ~24 km × 160 km towards the edge of the swath). Zara et al. (2018) describes the QA4ECV OMI NO_2 and HCHO slant column retrievals and Boersma et al. (2018) and De Smedt et al. (2018) describe the entire QA4ECV OMI NO_2 and HCHO retrieval algorithms, respectively, in detail. They are summarized here briefly.

QA4ECV retrievals of NO_2 SCDs are obtained by linear fits of optical depths to the observed optical depth using the DOAS technique with a fitting window between 405 and 465 nm (Boersma et al., 2018). While the QA4ECV NO_2 retrieval is based on DOAS methods, it differs from the NASA OMI NO_2 retrieval in ~~many~~ some of the retrieval

steps (Compernolle et al., 2020). ~~For instance, the OMNO2 retrieval algorithm uses non linear fits of modelled reflectance to the observed reflectance. Furthermore, NASA OMI NO_2 uses an iterative fitting procedure compared to a simultaneous fitting applied in QA4ECV.~~ To calculate tropospheric AMFs, the QA4ECV NO_2 retrieval algorithm uses the same surface albedo (Kleipool et al., 2008) and cloud products (Veefkind et al., 2016) as the previous NASA OMI NO_2 version 3 data (see Lamsal et al., 2021); however, uses daily a priori profiles from the TM5 CTM at 1.0° × 1.0° spatial resolution. Tropospheric VCDs of NO_2 are separated from the entire column using output from the global TM5 assimilation model in the QA4ECV NO_2 retrieval. For detailed information on the differences in spectral fitting between the NASA OMI NO_2 and QA4ECV NO_2 retrieval algorithms we refer you to Zara et al. (2018). For details about differences between AMF calculations in the NASA and QA4ECV OMI algorithms see Lorente et al. (2017). QA4ECV NO_2 data have been shown to perform relatively well in clean to moderately polluted regions and have a low bias in highly polluted regions (Compernolle et al., 2020). Retrievals of QA4ECV HCHO SCDs are conducted in a similar manner to QA4ECV NO_2 using the DOAS technique and optical depths with a fitting window between 328.5 and ~~346~~359.0 nm (Zara et al., 2018; De Smedt et al., 2018). For information about the inputs used in AMF calculations for QA4ECV OMI NO_2 and HCHO retrievals see Table S1. QA4ECV HCHO retrievals show minimal bias in clean to moderately polluted regions and low biases in polluted locations (e.g., De Smedt et al., 2021).

2.2 TROPOMI remote-sensing products

The TROPOMI hyperspectral spectrometer (including eight bands in the UV, VIS, near-infrared, and shortwave infrared wavelengths) is onboard the Sentinel-5 Precursor (S5P) satellite developed by the ESA which was launched in October 2017. TROPOMI is in orbit with a similar local equatorial overpass time (local time ~13:30) as OMI. TROPOMI has a swath width of ~2,600 km and a ground pixel size of 3.5 km × 7.0 km at nadir during the time of this study (since August 6, 2019 TROPOMI data is available at 3.5 km × 5.5 km) which is >12 times finer than OMI. TROPOMI retrievals have been used in numerous recent studies investigating processes controlling NO_2 concentrations and trends (e.g., Goldberg et al., 2021) and FNRs (Wu et al., 2022), taking advantage of the high spatiotemporal resolution of the sensor, along with being validated thoroughly (e.g., Judd et al., 2020; De Smedt et al., 2021). The high spatial resolution information provided by TROPOMI, compared to past UV/VIS spaceborne sensors, reduces the representation error of each retrieved NO_2 and HCHO pixel (Souri et al., 2022b). In this study, we apply daily TROPOMI tropospheric column NO_2 v2.3.1 (van Geffen et al., 2022) and HCHO v1.1.5 retrievals (De Smedt et al., 2018). For TROPOMI NO_2 data we used the product provided by the Product Algorithm Laboratory (PAL) which applies the NO_2 v2.3.1 algorithm but for the time period between April 2018 - September 2021. The retrievals of both species use QA4ECV methods described above applying the DOAS methods with spectral fitting windows between 405.0 and 465.0 nm for NO_2 (Boersma et al., 2018) and 328.5 and ~~346~~359.0 nm for HCHO (De Smedt et al., 2018). TROPOMI retrievals are similar to those from the QA4ECV OMI product as it applies the same a priori profiles from the TM5 model, albedo data, and cloud fraction information. TROPOMI NO_2 v2.3.1 retrievals do differ from QA4ECV OMI NO_2 products as it uses cloud pressure input from the O_2-A band following the FRESCO-wide approach (van Geffen et al., 2022) instead of O_2–O_2 absorption. Similarly, TROPOMI HCHO v1.1.5 retrievals differ from the QA4ECV OMI HCHO data through applying the S5P ROCINN algorithm which uses the

O_2-A for cloud pressures (Loyola et al., 2018) instead of O_2–O_2 absorption. For more information about the input data sets used in AMF calculations for TROPOMI NO_2 and HCHO retrievals see Table S1.

~~2.3 TEMPO synthetic retrieval product~~

~~One component of the pre-launch activities of the geostationary TEMPO satellite mission is to generate synthetic retrieval data for end user communities, which closely represents the planned operational products of the mission planned for launch in early 2023 (Naeger et al., 2021). The synthetic data products are provided daily and at the expected 2.0 km × 4.75 km (at nadir) spatial resolution of TEMPO. Synthetic TEMPO data is applied in this study to demonstrate the additional FNR information, which will be provided by the high spatiotemporal resolution (including up to hourly information during the daylight hours) of this geostationary sensor, compared to existing low earth orbit sensors (i.e., OMI and TROPOMI).~~

~~Hourly model output of NO_2 and HCHO vertical profiles from the NASA Global Modeling and Assimilation Office (GMAO) GEOS Composition Forecasting (GEOS-CF) model were used as input into the TEMPO proxy development methodology and sampled at the TEMPO footprint (2.0 km × 4.75 km at the center of the Field of Regard) to represent the "true" state of the atmosphere. GEOS-CF simulates meteorology and aerosol and trace gas concentrations in both the troposphere and stratosphere at a high global spatiotemporal resolution (0.25° × 0.25°) and 72 vertical layers (Knowland et al., 2020, 2022; Keller et al., 2020). The GEOS-CF model is a reliable source for the "true" atmosphere as it has been shown to produce realistic concentrations of aerosols and trace gases in comparison to remote sensing observations and in situ measurements (e.g., Keller et al., 2020; Johnson et al., 2021; Knowland et al., 2022). Scattering weights in clear and cloudy conditions are derived from pre-computed lookup tables of radiances and NO_2 and HCHO scattering weights at 440 and 340 nm, respectively, as a function of the TEMPO viewing geometry centered at 91°W, surface reflectance, cloud fraction, cloud pressure, and absorption by O_3. The Geostationary Coastal and Air Pollution Events (GEOCAPE) Radiative Transfer Tool (based on the Vector Linearized Discrete Ordinate Radiative Transfer (VLIDORT) model (Spurr, 2006)) was used to create the lookup table. A fast optical centroid pressure simulator (Joiner et al., 2012), a simple mixed Lambertian model where clouds are parameterized as opaque reflective surfaces, was used to account for the effects of clouds on the SWs. Surface model reflectance is based on Moderate Resolution Imaging Spectroradiometer (MODIS) Blue Sky Albedo calculations, and the Cox-Munk Glitter kernel + whitecap parameterization + water leaving radiances. Climatology albedo is based on the OMI Lambertian equivalent reflectance. After deriving the SWs from the lookup table, "true" SCDs are calculated from the summation of the target trace gas (i.e., NO_2, HCHO) concentration profile multiplied by the scattering weights in each model layer.~~

~~To convert the true SCD to the final proxy VCD, we i) applied a statistical random noise model developed in Zoogman et al. (2017) for TEMPO proxy data, primarily as a function of spectral signal-to-noise ratios and column abundance of NO_2 and HCHO, to the "true" SCDs, ii) applied climatological AMFs derived from NO_2 and HCHO scattering weights based on the same lookup table approach as discussed above, but using NO_2 and HCHO profiles from hourly model output data from the Goddard Earth Observing System Model version 5 with GEOS-Chem as a chemical model at ~12 km grid spacing (G5NR-Chem; Hu et al., 2018).~~

2.~~4~~ 3 Airborne spectrometers

The primary evaluation data set used in this study is from the UV/VIS airborne remote-sensing data product from GeoTASO and GCAS flown during the LISTOS 2018 field campaign ~~(16 flights between June 18 and October 19, 2018)~~. Due to the fact that no bias-corrected tropospheric column HCHO data is available during LISTOS 2018 from the Pandora network, this ground-based remote-sensing network is not applied here. Both the GeoTASO and GCAS instruments and retrievals are very similar and together provide a consistent evaluation data set (see specific details on the instruments and NO_2 and HCHO retrievals in Kowalewski and Janz (2014), Leitch et al. (2014), Nowlan et al. (2016, 2018), and Judd et al. (2020)). GeoTASO and GCAS NO_2 and HCHO data were obtained from a nominal flight altitude of 9 km above ground level (agl) covering the majority of the troposphere. The airborne data from 13 flight days between June 25 and September 6, 2018 (see Table 1) are provided with a native spatial resolution of 250 m × 250 m. To reduce noise in the raw GeoTASO and GCAS retrievals, the data were averaged to a 1 km × 1 km spatial resolution. In total, measurements from 8 and 12 flight days were spatiotemporally co-located with OMI and TROPOMI overpasses, respectively. A detailed explanation of the measurements and flights conducted during LISTOS 2018 is provided in Judd et al. (2020).

The airborne GeoTASO and GCAS retrievals are used here as the reference data set for validating all satellite data. However, the airborne remote-sensing data is not without error. A nearly identical GeoTASO and GCAS tropospheric column NO_2 data set used in this work was applied in Judd et al. (2020) and was evaluated with a network of Pandora systems. Judd et al. (2020) demonstrated that the airborne NO_2 retrievals had a median bias of ~1% with uncertainty within ±25% with no magnitude dependent biases. Due to minimal availability of ground-based remote-sensing Pandora data of HCHO, airborne GeoTASO and GCAS retrievals of this species has had limited evaluation. Nowlan et al. (2018) did evaluate GCAS tropospheric HCHO retrievals using P-3B airborne in situ measurements and determined GCAS had generally good performance with a < 10% bias (minimal magnitude ~~dependance~~ dependence in bias) and high correlation with observations. Overall, the satisfactory comparison of airborne GeoTASO and GCAS tropospheric column NO_2 and HCHO with independent observations provides confidence that this data can be applied as a reference data set to validate OMI and TROPOMI retrievals. However, it should be kept in mind that there is some level of error/bias associated with the GeoTASO and GCAS data used in this study (e.g., Nowlan et al., 2016; 2018; Judd et al., 2020).

The GeoTASO and GCAS data taken during the LISTOS 2018 campaign provided a unique opportunity to use airborne remote-sensing observations of NO_2 and HCHO to validate both OMI and TROPOMI coincidently. The overlap of data from both OMI and TROPOMI spaceborne sensors afforded by the timing of LISTOS 2018 is a novel aspect of this validation data set. This airborne data differs from many of the recent satellite validation studies which use longer term information from networks of point-source measurements (e.g., Pandora, MAX-DOAS) (e.g., Compernolle et al., 2020; Vigouroux et al., 2020; Verhoelst et al., 2021; Lamsal et al., 2021; Souri et al., 2022a). The airborne sensors used here allowed for evaluation of OMI and TROPOMI over large areas on multiple different days which equates to having tens to hundreds of clustered ground-based systems on each flight day. Having long-term observations for robust temporal validation of satellite sensors is ideal; however, this case study is unique in that it provides information about the performance of coincident retrievals from OMI and TROPOMI over variable emission

source regions (urban to rural areas) and scenes with differing geophysical characteristics (e.g., surface albedo, tropospheric compositions, clouds, aerosol amounts/elevation, etc.) during a single flight which is a novel aspect of this work. Even though there are limited observations available from the flights in LISTOS (Table 1), all correlation statistics presented in this study (Table 2 and 3) are significant to a 95% confidence interval (p-value < .05).

Table 1. Airborne (GeoTASO and GCAS) flight information (date, flight times, number of co-located satellite and airborne FNR grids) used in this study from the LISTOS 2018 field campaign.

Flight Day Number	Date	Time (Hours in UTC)	OMI FNR co-locations[1]	TROPOMI FNR co-locations[2]
1	June 25, 2018	Morning: 12.5–15.7 Afternoon: 16.8–20.3	12	201
2	June 30, 2018	Morning: 12.2–15.6 Afternoon: 16.7–20.4	37	251
3	July 2, 2018	Morning: 11.4–16.6 Afternoon: 17.9–21.5	6	66
4	July 19, 2018	Morning: 11.4–15.3 Afternoon: 16.9–20.9	0	155
5	July 20, 2018	Morning: 11.4–15.3 Afternoon: 17.1–21.1	5	136
6	August 5, 2018	Morning: 12.5–16.5 Afternoon: 17.8–22.3	5	0
7	August 6, 2018	Morning: 11.7–16.0 Afternoon: 17.2–21.5	0	67
8	August 15, 2018	Morning: 11.2–15.5 Afternoon: 17.0–21.6	0	150
9	August 16, 2018	Morning: 11.3–15.3 Afternoon: 17.3–21.5	0	108
10	August 24, 2018	Morning: 10.9–15.3 Afternoon: 16.6–21.0	20	147
11	August 28, 2018	Morning: 11.3–15.3 Afternoon: 16.6–20.3	8	150
12	August 29, 2018	Morning: 11.2–15.1 Afternoon: 16.6–20.8	0	166
13	September 6, 2018	Morning: 11.9–15.8 Afternoon: 17.2–21.4	8	96

[1]OMI FNR co-locations for the near-native 0.15° × 0.15° spatial resolution gridded data.
[2]TROPOMI FNR co-locations for the near-native 0.05° × 0.05° spatial resolution gridded data.

2.~~5~~ 4 CMAQ model simulation

The prior vertical profiles play a major role in satellite retrievals of chemical constituents in the troposphere (e.g., Palmer et al., 2001; Boersma et al., 2007; Johnson et al., 2018). Furthermore, past research has demonstrated that using a well-constrained, high spatial resolution, air quality model or CTM as the a priori profile source for satellite retrievals can improve VCD results (e.g., Laughner et al., 2019). To compare NASA OMI and TROPOMI tropospheric NO_2, HCHO, and FNR retrievals using a common a priori profile data set, we conduct sensitivity tests applying model

simulated vertical profiles of NO_2 and HCHO produced by the Community Multiscale Air Quality Model (CMAQ) to reprocess the NASA OMI and TROPOMI retrievals. Reprocessing OMI and TROPOMI NO_2 and HCHO retrievals with a common, high spatial resolution (4 km × 4 km^2), model data product removes differences in the satellite products due to using different coarse spatial resolution model data sources as a priori vertical profiles.

We used CMAQ version 5.3 for air quality simulations during the LISTOS 2018 campaign. The CMAQ simulations were driven offline using the meteorological fields simulated by the Weather Research and Forecasting (WRF) model version 4.1. The WRF-CMAQ spatial domain set-up is shown in Fig. S1. The outer WRF domain covers the contiguous United States (CONUS) at a horizontal grid spacing of 12 km × 12 km^2 ~~(481 × 369 grid points)~~ and the inner WRF domain covers the northeastern US, encompassing the entire LISTOS 2018 campaign domain, at a horizontal grid spacing of 4 km × 4 km^2 ~~(237 × 189 grid points)~~. Both the outer and inner model domains use 35 vertical levels between the surface and 50 hPa. The WRF configuration follows Appel et al. (2017), which includes improved representation of the land-surface processes and vertical mixing, and employs four-dimensional data assimilation ~~(also called grid nudging)~~ every 6 hours to limit the growth of meteorological errors in the simulations (WRF configuration details in Table ~~S1~~S2). A 15-day spin up period was used for the WRF-CMAQ simulations to minimize the impacts of errors ~~in~~ from initial conditions. Anthropogenic emissions of trace gases and aerosols are based on the National Emissions Inventory (NEI) representative of 2014 because that was the latest available inventory from EPA at the time of emission preparation. NEI 2014 emissions were processed using the Sparse Matrix Operator Kernel Emissions (SMOKE) model with the same configuration as adopted in the EPA 2014 emissions modeling platform (https://www.epa.gov/air-emissions-modeling/2014-version-71-platform). The same WRF simulations described above were used to drive SMOKE for generating meteorology-dependent anthropogenic emissions. Biogenic emissions of trace gases and aerosols are calculated online within the model using the Biogenic Emissions Inventory System (BEIS). The gas-phase chemistry and aerosol processes are represented using Carbon bond 6 (CB06) version r3 with AERO7 treatment of the secondary organic aerosols. Chemical lateral boundary conditions for the outer domain were based on the idealized profiles available in CMAQ but are dynamically provided to the inner domain every hour based on the outer domain simulations.

2.~~6~~ 5 Evaluation techniques

In order to perform a systematic, direct comparison of daily satellite products to airborne retrievals, OMI and GeoTASO/GCAS data were spatially-averaged to 0.15° × 0.15° (~15 km × 15 km^2, similar to OMI nadir spatial resolution) for evaluating OMI. TROPOMI and airborne observations were spatially-averaged at 0.05° × 0.05° (~5 km × 5 km^2, similar to TROPOMI nadir spatial resolution) for evaluating TROPOMI data. To investigate the impact of the higher spatial resolution of TROPOMI, NO_2, HCHO, and FNR retrievals from this sensor were also averaged to the 0.15° × 0.15° for inter-comparison with OMI evaluation statistics. In order to smooth and reduce the noise of satellite data, we apply a point oversampling technique (e.g., McLinden et al., 2012) when spatially averaging the retrievals. This method uses a larger grid box radius, compared to the averaging resolution, to bin individual retrievals. When averaging satellite data to the 0.15° × 0.15° spatial resolution (standard radius of 0.075°), we employed a radius twice the standard size equal to 0.15°. Similarly, when averaging satellite data to the 0.05° × 0.05° spatial resolution

(standard radius of 0.025°) we applied a radius of 0.05°. By spatially-averaging the tropospheric column NO_2 and HCHO GeoTASO/GCAS data we minimized the spatial representation error between OMI and TROPOMI satellite retrieved pixels with those of GeoTASO/GCAS.

Given that the nominal flight altitude for GeoTASO and GCAS observations was 9 km agl, in order to directly compare to satellite tropospheric column retrievals, we scaled airborne tropospheric column NO_2 values by multiplying the observed values by the ratio of the total tropospheric NO_2 column abundance over the tropospheric column NO_2 abundance below 9 km agl (i.e., $\frac{\int Tropospheric\ NO_2\ (surface\ to\ tropopause)}{\int Tropospheric\ NO_2\ (surface\ to\ 9\ km\ agl)}$). This scaling factor for NO_2, which showed that typically 60% to 99% of tropospheric NO_2 is below 9 km agl, was derived for each co-located GeoTASO and GCAS retrieval, using WRF-CMAQ simulations~~using the WRF-CMAQ simulations described in Sect. 2.5~~. Tropospheric column HCHO data from GeoTASO and GCAS were not scaled due to the fact that typically >95% of the total column HCHO is below the nominal aircraft flight altitude.

For comparison to satellite retrievals, GeoTASO and GCAS data were co-located to OMI and TROPOMI data using a temporal threshold of ± 60 minutes. Before GeoTASO and GCAS HCHO and NO_2 data were co-located with satellite data they were filtered to remove airborne retrievals where the radiance flag was > 0.5 as they are considered to be influenced by clouds or glint. We initially applied a temporal threshold of ± 30 minutes; however, this resulted in < 50 total co-locations with OMI retrievals throughout the study time period. Therefore, the longer temporal threshold criteria was necessary to achieve enough co-locations for statistical evaluation. The longer temporal threshold of ± 60 minutes resulted in only slightly larger median biases compared to when applying the ± 30 minute threshold. The similar bias statistics using temporal offsets of 30 and 60 minutes agrees with other studies which show minimal dependence~~dependance~~ on temporal offsets between 0 and 60 minutes (e.g., Tack et al., 2021). It should be noted that the temporal threshold of ± 60 minutes, and spatial gridding/averaging methods applied in this study, resulted in slightly larger spread in TROPOMI NO_2 data when evaluated to GeoTASO and GCAS data compared to the results in Judd et al. (2020) which used a ± 30 minute co-location threshold.

Satellite retrievals with high quality were filtered for use by removing individual retrievals that did not have quality flags (qa) = 0 for HCHO and NO_2 when applying OMI data. This qa value is suggested in the OMI data user's manuals for the application of the highest quality science data and for the removal of OMI pixels impacted by the row anomaly. For TROPOMI, individual retrievals of NO_2 and HCHO that had qa < 0.75 and qa < 0.5 were removed prior to spatial averaging, respectively, as recommended by the TROPOMI data user manuals for each species. ~~These qa values were selected based on guidance from OMI and TROPOMI user's guides to remove data with large uncertainty to produce high quality science data products.~~ Furthermore, to avoid anomalous OMI and TROPOMI retrieval values of HCHO, we remove VCDs with lower and upper bounds of -8.0×10^{15} and 7.6×10^{16} molecules cm^{-2}, respectively. These bounds were determined from typical HCHO VCD values and a threshold of 3 times the fitting uncertainty of OMI retrievals presented by~~following~~ Zhu et al. (2020). Similarly, to avoid anomalous OMI and TROPOMI retrieval values of NO_2, we remove VCDs with lower and upper bounds of -1.08×10^{15} and 8.07×10^{16} molecules cm^{-2}, respectively (personal communication with OMI NO_2 algorithm team). Both OMI and TROPOMI retrievals with solar zenith angles > 70° and effective cloud fractions > 30% and > 50%, respectively were also removed. These additional thresholds were chosen based on guidance from the OMI and TROPOMI data user's guides. Finally, only co-located

spatially-averaged grids that had 75% spatial coverage by GeoTASO/GCAS data and airborne remote-sensing NO_2 VCDs > 1.0×10^{15} molecules cm^{-2} were used in the evaluation.

The statistical evaluation of daily and campaign-averaged (includes all flights displayed in Table 1) OMI and TROPOMI retrievals with co-located GeoTASO and GCAS spatially-averaged data was primarily done using bias (median), oscillation/variability in bias represented by the standard deviation of bias (referred to as bias standard deviation throughout), normalized median bias (NMB) which are normalized to the magnitude of observed data, root mean squared error (RMSE), and simple linear regression statistics (slope, y-intercept, coefficient of determination (R^2)) based on ordinary least-squares.

3 Results

In this section we evaluate the capability of NASA OMI ~~products (hereinafter referred to as NASA OMI)~~, QA4ECV OMI ~~retrievals (hereinafter referred to as QA4ECV)~~, and TROPOMI to retrieve tropospheric columns of NO_2, HCHO, and FNRs during ~~the~~ LISTOS 2018 (temporally-averaged values from all flights hereinafter referred to as campaign-averaged). ~~We further evaluate these retrievals on a day characterized by large NO_2 pollution focusing on NASA OMI and TROPOMI.~~ We also present results of a sensitivity test using common a priori vertical profiles of NO_2 and HCHO from WRF-CMAQ to reprocess NASA OMI and TROPOMI retrievals. Finally, we ~~present information on the expected additional FNR information that will be provided from the future NASA geostationary TEMPO satellite~~discuss the relative error of FNR calculations from uncertainty in HCHO and NO_2 column retrievals, the capability of OMI and TROPOMI to capture observed spatial and temporal patterns of FNRs during the campaign, and potential sources of systematic bias and uncertainty (this term is used throughout to describe all unresolved errors beyond systematic biases such as random errors and relative biases) of OMI and TROPOMI FNR retrievals..

3.1 Campaign-averaged tropospheric FNRs

Airborne observations during the summer of 2018 suggest that during the mid-day hours large ~~regions~~ areas of FNRs ≤ 1.0 occurred over the urban regions surrounding New York City (NYC). The term "urban" here is used qualitatively as the region close in proximity to the center of NYC where elevated tropospheric column NO_2 values over NO_x emission regions ~~are~~ were frequently observed. The opposite is true for the usage of "rural" ~~hereinafter~~. Figure 1 shows the campaign-averaged FNRs from OMI (NASA and QA4ECV) and TROPOMI retrievals, averaged to spatial resolutions of 0.15° × 0.15° and 0.05° × 0.05°, ~~respectively~~ compared to co-located airborne remote-sensing products. These regions of FNRs ≤ 1.0 likely have O_3 production which is limited by VOC emissions. Outside of the VOC/radical-limited region around NYC, airborne observations show a clear transition zone of FNRs between 1.0 and 2.0 and NO_x-limited regimes (FNR > 2.0) in the rural regions of the northeast US. It should be noted these FNR thresholds being discussed follow the assumptions of Duncan et al. (2010); however, there are uncertainties in the exact thresholds separating O_3 sensitivity production regimes and they can be spatiotemporally variable (e.g., Lu and Chang, 1998; Schroeder et al., 2017; Souri et al., 2020; Ren and Xie, 2022). For example, a recent study by Jin et al. (2020) suggests that VOC/radical-limited regimes around NYC transition to NO_x-limited regimes for FNRs between

2.9 and 3.8. For simplicity, we use the constant FNR ratio thresholds defined by Duncan et al. (2010) for discussion throughout the ~~rest of this~~ study.

Satellite retrievals during the summer of 2018 also displayed the same general regional patterns of FNRs in the northeast US that were observed by airborne remote-sensing (see Fig. 1). NASA OMI, QA4ECV OMI, and TROPOMI retrieved lower FNRs in the urban region of NYC and a transition to NO_x-limited regimes in the rural regions. However, all satellite products show higher FNRs (between 1.0 and 3.0) in the areas where airborne observations clearly observed NO_x-saturated regimes. In general, TROPOMI FNRs at the 0.05° × 0.05° spatial resolution have the lowest values over NYC in better agreement with airborne observations. The higher spatial resolution satellite data provided by TROPOMI also has a smaller spatial extent of a transition zone and VOC/radical-limited regimes in comparison to the two OMI products. TROPOMI FNR retrievals and airborne observations display a clear urban/rural interface; however, OMI products result in noisier spatial patterns. Between the two OMI retrieval products, QA4ECV FNR values are lower in the observed VOC/radical-limited region in comparison to NASA OMI and appear to compare more favorably to airborne observations.

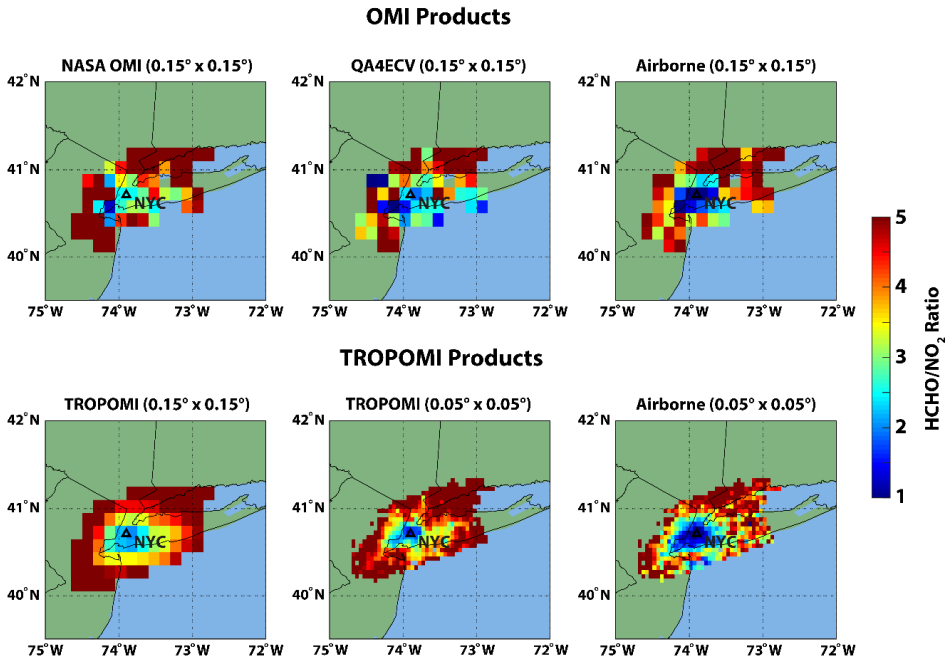


Figure 1: NASA OMI, QA4ECV OMI, TROPOMI, and airborne tropospheric column FNR retrievals averaged for all flights conducted during the LISTOS 2018 field campaign. All co-located OMI and airborne remote-sensing tropospheric column FNR values are averaged at 0.15° × 0.15° and TROPOMI co-locations are averaged at both 0.05° × 0.05° and 0.15° × 0.15° spatial resolutions. The black triangle indicates the location of the city of NYC.

Figure 1 illustrates the impact of retrieval spatial resolution on the ability of satellite-derived FNRs to reproduce observed O_3 sensitivity production regimes. TROPOMI retrieval data better captures the spatial pattern and urban/rural interface of observed O_3 sensitivity production regimes compared to OMI data. TROPOMI results when gridded near the native resolution of the sensor (0.05° × 0.05°), while still higher compared to observed FNRs around NYC, were able to retrieve FNRs < 2.0. However, when averaged to a resolution similar to the native resolution of

OMI (0.15° × 0.15°), TROPOMI data suggests higher FNRs ≥ 2.0 in the vicinity of NYC, in line with OMI retrieval products.

It should be noted that satellite- and airborne-retrieved FNRs are dependent on both tropospheric NO_2 and HCHO ~~values~~data. Median/mean and unresolved biases in FNRs can then be driven by errors in either retrievals of NO_2 and/or HCHO. Therefore, the following sections of this work investigate the statistical ~~comparison~~ evaluation of NASA OMI, QA4ECV OMI, and TROPOMI tropospheric NO_2, HCHO, and resulting FNRs ~~compared to airborne observations~~.

3.2 Statistical evaluation of OMI and TROPOMI retrievals

3.2.1 Tropospheric column NO_2 ~~retrievals~~ systematic bias

The spatial pattern of campaign-averaged tropospheric column NO_2 retrieved by the satellites and airborne sensors highlight the large pollution region around the urban ~~region~~ areas of NYC ~~during the summer of 2018~~ (see Fig. S2). ~~Tropospheric column NO_2 concentrations~~Tropospheric NO_2 columns over NYC from both satellite and airborne observations frequently exceed 1.0×10^{16} molecules cm^{-2} within 60 minutes of the OMI and TROPOMI overpass times. However, while airborne tropospheric column NO_2 values in the rural regions surrounding NYC were frequently observed to be $< 2.0 \times 10^{15}$ molecules cm^{-2}, satellite retrievals have larger background ~~tropospheric column NO_2 concentrations~~tropospheric NO_2 columns between 2.0×10^{15} and $> 4.0 \times 10^{15}$ molecules cm^{-2}. This suggests OMI and TROPOMI retrievals have a high bias in background ~~tropospheric column NO_2 concentrations~~tropospheric NO_2 columns (spatial distribution of NO_2 bias in NASA OMI, QA4ECV OMI, and TROPOMI shown in Fig. S4). This high bias in satellite background tropospheric column NO_2 values can possibly be linked to underestimated abundance of free tropospheric NO_2 in CTMs used as a priori profile data sets for OMI and TROPOMI retrievals resulting in AMFs which are too low (e.g., Silvern et al., 2019). Furthermore, studies have shown that the coarse spatial resolution of the CTMs used to derive a priori NO_2 profiles for OMI and TROPOMI cannot resolve the sharp gradients of NO_2 at the urban/rural interface and lead to the overestimate of satellite retrievals in low pollution regions (Lamsal et al., 2014; Tack et al., 2021). Finally, other aspects of the satellite retrievals such as biases in stratospheric NO_2 concentrations and separation from the troposphere, aerosol interference, and surface albedo could contribute to these overestimations in background, low pollution regions (e.g., Lamsal et al., 2021).

Figure 2 shows the comparison of co-located NASA OMI, QA4ECV OMI, and TROPOMI retrievals of ~~tropospheric column NO_2 concentrations~~tropospheric NO_2 columns with observed data from all flights (statistical evaluation shown in Table 2). ~~This~~ The ~~figure and Table 2 further emphasize the~~ high bias of background ~~tropospheric column NO_2 concentrations~~tropospheric NO_2 columns retrieved by NASA OMI, QA4ECV OMI, and TROPOMI~~. All satellite products typically have a high bias~~ compared to the small ~~tropospheric column NO_2 concentrations~~tropospheric NO_2 columns ($< 5.0 \times 10^{15}$ molecules cm^{-2}) measured outside the urban regions of NYC~~,~~ result in linear regression slopes < 0.65~~,~~ and positive y-intercepts when compared to the airborne observations. Some of this high bias in background ~~tropospheric column NO_2 concentrations~~tropospheric NO_2 columns is offset in the campaign-averaged median biases by the fact that the satellite retrievals have a low bias compared to NO_2 values observed over polluted regions ($> 1.0 \times 10^{16}$ molecules cm^{-2}). TROPOMI at its near native spatial resolution has the

least high bias of background tropospheric NO_2 columns demonstrated by the lower y-axis intercept compared to all OMI and TROPOMI data products at the coarser spatial resolution. Overall, NASA OMI displayseed a small campaign-averaged median bias (NMB %) of 0.4±4.1 × 10^{15} molecules cm^{-2} (6.3%) in comparison to tropospheric column NO_2 observations. QA4ECV OMI data resultsed in a campaign-averaged median bias of 0.4±4.5 × 10^{15} molecules cm^{-2} (6.8%). Finally, TROPOMI retrievals havved a campaign-averaged median bias of -0.3±3.7 × 10^{15} molecules cm^{-2} (-4.8%) and 0.3±3.3 × 10^{15} molecules cm^{-2} (5.8%) when averaged at 0.05° × 0.05° and 0.15° × 0.15° spatial resolution, respectively. It should be noted that the TROPOMI low bias in tropospheric column NO_2 is improved with the newer retrieval algorithm used in this study compared to early versions of the data product (e.g., v1.2.2 had a campaign-averaged median low bias of -1.3±4.0 × 10^{15} molecules cm^{-2}). The results here suggest that OMI, and to a lesser extent TROPOMI, tropospheric column NO_2 retrievals errors have a magnitude dependence and tend to have some high bias in rural/background regions and a low bias in moderately to highly polluted regions which agrees with past validation studies (e.g., Zhao et al., 2020; Lamsal et al., 2021; Verhoelst et al., 2021).

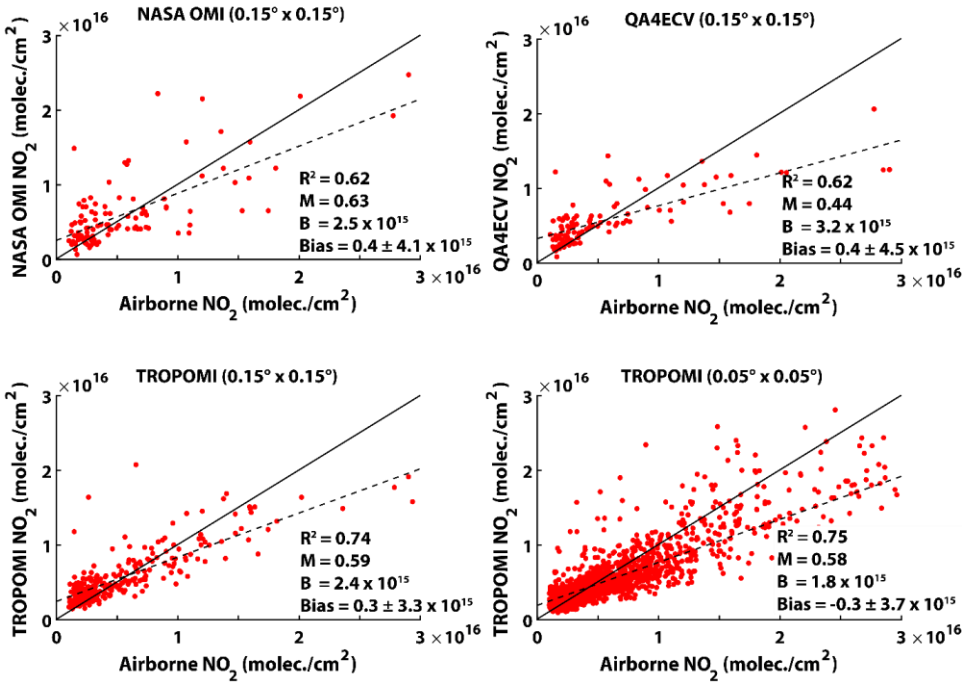


Figure 2: Scatter plots illustrating the comparison of satellite- (NASA OMI, QA4ECV OMI, and TROPOMI) and airborne-retrieved tropospheric NO_2 (molecule cm^{-2}) for each co-located measurement taken during the field campaign. All co-located OMI and airborne remote-sensing tropospheric column NO_2 values are averaged at the 0.15° × 0.15° resolution and TROPOMI co-located data are averaged at 0.15° × 0.15° and 0.05° × 0.05° spatial resolution. The solid black line shows the 1:1 comparison and the dashed line shows the linear regression fit of the comparison. The figure inset shows the main statistics (coefficient of determination (R^2), slope (M), y-intercept (B), and median bias and bias standard deviation) of the comparison of satellite and airborne tropospheric column NO_2 data.

3.2.2 Tropospheric column NO_2 uncertainty

~~NASA OMI displays a small campaign-averaged median bias (NMB %) of 0.4±4.1 × 10^{15} molecules cm^{-2} (6.3%) in comparison to tropospheric column NO_2 observations. QA4ECV OMI data results in a campaign-averaged median bias of 0.4±4.5 × 10^{15} molecules cm^{-2} (6.8%). Finally, TROPOMI retrievals have a campaign-averaged median bias~~

~~of 0.3±3.7 × 10^{15} molecules cm^{-2} (4.8%) and 0.3±3.3 × 10^{15} molecules cm^{-2} (5.8%) when averaged at 0.05° × 0.05° and 0.15° × 0.15° spatial resolution, respectively. It should be noted that the TROPOMI low bias in tropospheric column NO_2 is improved with the newer retrieval algorithm used in this study compared to early versions of the data product (e.g., v1.2.2 had a campaign averaged median low bias of -1.3±4.0 × 10^{15} molecules cm^{-2}).~~ In addition to ~~mean/median biases~~systematic biases discussed above, ~~bias standard deviation~~RMSE, which is indicative of noise in the satellite retrievals resulting in unresolved errors, is very important for accurate retrievals of the spatial-resolved daily tropospheric column NO_2, HCHO, and FNRs. At the near-native spatial resolution of the three satellite retrievals, ~~the standard deviation in bias of the data were similar~~RMSE values were similar (~~≈~~(≥4.0 × 10^{15} molecules cm^{-2}) with QA4ECV OMI data having the largest bias standard deviation and RMSE values and TROPOMI having ~~had~~ the least noise in the data (RMSE = 3.9 × 10^{15} molecules cm^{-2}) (see Table 2). TROPOMI data averaged to match OMI spatial resolution displayed lower ~~bias standard deviation~~RMSE values of ~~≈~~3~~.0~~ 3 × 10^{15} molecules cm^{-2}. At both spatial resolutions, TROPOMI tropospheric NO_2 data has ~~slightly~~ less spread in the data compared to OMI products. ~~The results here suggest that OMI and TROPOMI tropospheric column NO_2 retrievals errors have a magnitude dependence and tend to have a high bias in rural/background regions and a low bias in moderately to highly polluted regions which agrees with past validation studies (e.g., Judd et al., 2020; Compernolle et al., 2020; Lamsal et al., 2021).~~ The larger noise in OMI tropospheric NO_2 SCDs compared to TROPOMI NO_2 SCDs has been shown in recent studies (van Geffen et al., 2020, 2022) and has been attributed to reduced noise in TROPOMI due to its higher spatial resolution (factor of >12 better) and similar, to even better, signal-to-noise ratios.

To determine if the higher spatial resolution and lesser noise of TROPOMI retrievals resulted in more favorable comparisons to observations, we further compared TROPOMI tropospheric column NO_2 values to OMI results. TROPOMI ~~tropospheric column NO_2 concentrations~~tropospheric NO_2 columns at 0.05° × 0.05° displayed the lowest campaign-averaged median bias of all satellite products, and the higher spatial resolution data better reproduces the spatial patterns of observed tropospheric column NO_2. This is emphasized by the higher correlation and lower RMSE values when evaluating TROPOMI ~~tropospheric column NO_2 concentrations~~tropospheric NO_2 columns with observations in comparison to the other satellite products and visually more clearly separating the urban/rural interface seen in tropospheric NO_2 (see Fig. S2). Finally, TROPOMI NO_2 data averaged to the coarser spatial resolution of OMI has a similar campaign-averaged high median bias as both OMI retrieval algorithms; however, displayed RMSE values nearly twice as small as NASA and QA4ECV OMI, further emphasizing the importance of spatial resolution for retrieving tropospheric NO_2 columns.

Table 2. Statistical evaluation of NASA OMI, QA4ECV OMI, and TROPOMI retrievals of tropospheric column NO_2 and HCHO and resulting FNRs. Statistics presented are the number of co-located grids (N), mean concentration ± standard deviation from satellite (Sat Conc.) and airborne (Air. Conc.) retrievals, median bias ± bias standard deviation, NMB (%), RMSE, coefficient of determination (R^2 ~~,~~), and linear regression slope (Slope).

~~NASA OMI (0.15° × 0.15°)~~				~~QA4ECV (0.15° × 0.15°)~~			
	~~FNR~~	~~HCHO*~~	~~NO_2*~~		~~FNR~~	~~HCHO*~~	~~NO_2*~~
~~N~~	~~101~~	~~101~~	~~116~~	~~N~~	~~82~~	~~85~~	~~106~~
~~Bias~~	~~0.4±3.8~~	~~5.1±7.8~~	~~0.4±4.1~~	~~Bias~~	~~-0.2±3.3~~	~~2.3±8.9~~	~~0.4±4.5~~

~~NMB~~	~~11.0~~	~~38.7~~	~~6.3~~	~~NMB~~	~~-5.4~~	~~17.3~~	~~6.8~~
~~R^2~~	~~0.23~~	~~0.19~~	~~0.62~~	~~R^2~~	~~0.17~~	~~0.19~~	~~0.62~~
~~Slope~~	~~1.0~~	~~0.46~~	~~0.63~~	~~Slope~~	~~0.67~~	~~0.54~~	~~0.44~~
~~TROPOMI (0.15° × 0.15°)~~				~~TROPOMI (0.05° × 0.05°)~~			
	~~FNR~~	~~HCHO*~~	~~NO_2*~~		~~FNR~~	~~HCHO*~~	~~NO_2*~~
~~N~~	~~261~~	~~261~~	~~261~~	~~N~~	~~1693~~	~~1741~~	~~1802~~
~~Bias~~	~~0.3±1.4~~	~~2.9±4.9~~	~~0.3±3.3~~	~~Bias~~	~~0.4±2.3~~	~~1.9±6.7~~	~~-0.3±3.7~~
~~NMB~~	~~9.3~~	~~23.1~~	~~5.8~~	~~NMB~~	~~13.0~~	~~12.9~~	~~-4.8~~
~~R^2~~	~~0.48~~	~~0.40~~	~~0.74~~	~~R^2~~	~~0.29~~	~~0.28~~	~~0.75~~
~~Slope~~	~~0.75~~	~~0.47~~	~~0.59~~	~~Slope~~	~~0.70~~	~~0.55~~	~~0.58~~
NASA OMI (0.15° × 0.15°)				QA4ECV OMI (0.15° × 0.15°)			
	FNR	HCHO*	NO_2*		FNR	HCHO*	NO_2*
N	101	101	116	N	82	85	106
Sat. Conc.	4.4±4.3	17.5±7.5	6.3±5.3	Sat. Conc.	3.7±3.5	16.5±9.1	5.9±3.9
Air. Conc.	3.6±2.1	13.2±7.1	6.1±6.6	Air. Conc.	3.4±2.2	13.3±7.4	6.1±6.9
Bias	0.4±3.8	5.1±7.8	0.4±4.1	Bias	-0.2±3.3	2.3±8.9	0.4±4.5
NMB	11.0	38.7	6.3	NMB	-5.4	17.3	6.8
RMSE	3.8	8.9	4.1	RMSE	3.3	9.4	4.5
R^2	*0.23*	*0.19*	*0.62*	R^2	*0.17*	*0.19*	*0.62*
Slope	1.0	0.46	0.63	Slope	0.67	0.54	0.44
TROPOMI (0.15° × 0.15°)				TROPOMI (0.05° × 0.05°)			
	FNR	HCHO*	NO_2*		FNR	HCHO*	NO_2*
N	261	261	261	N	1693	1741	1802
Sat. Conc.	3.6±1.8	15.9±4.7	5.9±4.2	Sat. Conc.	4.0±2.6	16.2±7.0	5.7±4.6
Air. Conc.	3.2±1.7	12.8±6.3	6.0±6.1	Air. Conc.	3.4±2.0	14.6±6.7	6.6±6.9
Bias	0.3±1.4	2.9±4.9	0.3±3.3	Bias	0.4±2.3	1.9±6.7	-0.3±3.7
NMB	9.3	23.1	5.8	NMB	13.0	12.9	-4.8
RMSE	1.4	5.6	3.3	RMSE	2.3	6.7	3.9
R^2	*0.48*	*0.40*	*0.74*	R^2	*0.29*	*0.28*	*0.75*
Slope	0.75	0.47	0.59	Slope	0.70	0.55	0.58

*concentration, bias, and RMSE units are $\times 10^{15}$ molecules cm^{-2}.

[c]correlation values which are presented in italics are statistically significant to a 95% confidence interval.

3.~~2~~2.~~2~~ 3 Tropospheric column HCHO ~~retrievals~~ systematic bias

The spatial pattern of campaign-averaged tropospheric column HCHO retrieved by the satellites and airborne sensors ~~highlight~~ demonstrate the large HCHO concentrations in both urban and rural regions during the summer of 2018 (see Fig. S3). This differs from tropospheric column NO_2, which is primarily emitted from anthropogenic sources, due to the fact HCHO has both anthropogenic and natural precursor emission sources and precursors with longer atmospheric lifetimes. The longer lifetime of precursor species producing HCHO result in less heterogeneity and gradients in

HCHO concentrations throughout the domain. Airborne observations of tropospheric column HCHO concentrations show that over NYC the concentrations are on average ~1.5 × 10^{16} molecules cm^{-2}, and can exceed 2.5 × 10^{16} molecules cm^{-2} during the afternoon hours (see Fig. S3). Both OMI and TROPOMI retrieval products have smaller gradients between HCHO concentrations in the urban and rural regions in comparison to airborne observations.

Figure 3 shows the scatter plot comparison of co-located NASA OMI, QA4ECV OMI, and TROPOMI retrievals of ~~tropospheric column HCHO concentrations~~tropospheric HCHO columns compared to observed data (statistical evaluation shown in Table 2). This figure and Table 2 illustrate the high bias of background ~~tropospheric column HCHO concentrations~~tropospheric HCHO columns retrieved by NASA OMI, QA4ECV OMI, and TROPOMI compared to airborne observations (spatial distribution of HCHO bias in OMI and TROPOMI shown in Fig. S5). All satellite products have a high bias when tropospheric columns HCHO are ≤ 1.5 × 10^{16} molecules cm^{-2}, linear regression slopes < 0.60, and positive y-intercepts when compared to observations ~~(in agreement with Vigouroux et al. (2020))~~. Both OMI retrieval products and TROPOMI data better replicate the larger HCHO concentrations (between 1.5 × 10^{16} and 3.0 × 10^{16} molecules cm^{-2}) with some small low bias in more polluted regions (> 3.0 × 10^{16} molecules cm^{-2}). On average, NASA OMI had the largest campaign-averaged median high bias of 5.1±7.8 × 10^{15} molecules cm^{-2} (38.7%). QA4ECV OMI data results in a lower campaign-averaged median high bias of 2.3±8.9 × 10^{15} molecules cm^{-2} (17.3%). Finally, TROPOMI retrievals had the lowest campaign-averaged median high bias of 1.9±6.7 × 10^{15} molecules cm^{-2} (12.9%) at 0.05° × 0.05° spatial resolution and 2.9±4.9 × 10^{15} molecules cm^{-2} (23.1%) when averaged at 0.15° × 0.15°. ~~Spatially averaging TROPOMI tropospheric column HCHO, along with tropospheric column NO_2 and FNRs, to coarser grids in order to increase signal to noise aided in reducing the bias standard deviation in HCHO retrieval products (see Table 2).~~

The results of the validation shown in Fig. 3 and Table 2 are consistent with recent validation studies such as the work of Vigouroux et al. (2020) and De Smedt et al. (2021) which also show that in regions of high ~~tropospheric column HCHO concentrations~~tropospheric HCHO columns, OMI and TROPOMI retrievals are generally consistent with some moderate low bias. However, in regions of lower background ~~tropospheric column HCHO concentrations~~tropospheric HCHO columns, both OMI and TROPOMI HCHO retrievals are biased high and OMI products tend to display a larger high bias compared to TROPOMI. ~~Furthermore, these two studies agree with our analysis that TROPOMI HCHO has lower bias standard deviation, and higher correlations with observations, compared to both OMI products evaluated here. The larger spread in tropospheric HCHO from OMI compared to TROPOMI is likely due to the weaker signal to noise in OMI and potentially the fewer co-located data points for statistical analysis. This is further demonstrated by the TROPOMI bias standard deviation being nearly a factor of two smaller compared to NASA OMI and QA4ECV when averaged to the OMI spatial resolution. TROPOMI HCHO retrievals have the smallest median bias and bias standard deviation compared to observations, and highest correlation with airborne observations, suggesting this newer sensor can better retrieve HCHO compared to OMI during this time period.~~ In order to provide more of a quantitative comparison with these recent validation studies of OMI and TROPOMI HCHO (Vigouroux et al., 2020; De Smedt et al., 2021), we separated our collocated satellite/airborne data points into clean (<5.0×10^{15} molecules cm^{-2}) and polluted (≥8.0×10^{15} molecules cm^{-2}) scenes. We chose a slightly higher threshold for separating clean HCHO columns to optimize the number of colocations for statistics and to be

similar to Vigouroux et al. (2020). We also added a highly polluted threshold (>16.0×10^{15} molecules cm^{-2}) to further emphasize our results. Table S3 summarizes the median bias ± bias standard deviation and NMB results for NASA OMI, QA4ECV OMI, and TROPOMI at coarser/fine spatial resolution for the different HCHO column magnitudes. While the positive tropospheric HCHO column biases derived in our study are higher compared to the recent studies of Vigouroux et al. (2020) and De Smedt et al. (2021), the magnitude dependence is the same. We show here that clean/background satellite HCHO columns are larger than observations for all satellite products and transition to a low bias in highly polluted regions.

3.2.4 Tropospheric column HCHO uncertainty

The NASA and QA4ECV OMI HCHO retrievals had RMSE values ~9.0 × 10^{15} molecules cm^{-2} with QA4ECV data having slightly larger data spread. The higher spatial resolution and sufficient signal-to-noise of TROPOMI resulted in HCHO RMSE values (6.7 × 10^{15} molecules cm^{-2}) ~25-30% lower compared to OMI. Spatially averaging TROPOMI tropospheric column HCHO to coarser grids in order to increase signal-to-noise aided in reducing RMSE values in HCHO retrieval products to 5.6 × 10^{15} molecules cm^{-2} (see Table 2). While both TROPOMI and OMI tropospheric HCHO retrievals display large noise, TROPOMI has correlation values better compared to OMI with R^2 values being a factor of 2 higher at the same spatial resolution. Vigouroux et al. (2020) and De Smedt et al. (2021) agree with our analysis that TROPOMI HCHO has lower RMSE values, and higher correlations with observations, compared to both OMI products evaluated here. The larger spread in tropospheric HCHO from OMI compared to TROPOMI is likely due to the weaker signal-to-noise in OMI and potentially the fewer co-located data points for statistical analysis. Overall, TROPOMI HCHO retrievals have the smallest median bias and RMSE values compared to observations, and highest correlation with airborne observations, suggesting this newer sensor can better retrieve HCHO compared to OMI during this time period.

All three satellite HCHO products have larger ~~bias standard deviations~~RMSE values and low correlations, when compared to the statistical evaluation of satellite NO_2 retrievals, when evaluated with observed tropospheric HCHO data. This highlights the large noise in these ~~retrieval~~ products likely driven by low signal-to-noise in HCHO retrievals. TROPOMI SCD retrievals of HCHO have been shown in recent work (e.g., De Smedt et al., 2021) to have less noise compared to OMI due to the higher spatial resolution and at least the same signal-to-noise. Furthermore, UV/VIS retrievals at shorter wavelengths (~340 nm) have much smaller sensitivity to HCHO compared to longer wavelengths (~440 nm) employed for NO_2 retrievals (Lorente et al., 2017). The sensitivity of UV/VIS retrievals to HCHO is lower throughout the middle and lower troposphere compared to NO_2, due to stronger Rayleigh scattering at shorter wavelengths, approaching twice as low near the surface (Lorente et al., 2017). The higher sensitivity of NO_2 retrievals in the middle to lower troposphere, compared to HCHO, is important as the highest concentrations, and largest spatiotemporal variability, of both NO_2 and HCHO occur lower in the troposphere near the PBL likely leading to the higher correlation and lower ~~bias standard deviation~~RMSE values in the tropospheric column NO_2 statistical evaluation.

~~Spatially averaging TROPOMI tropospheric column HCHO, along with tropospheric column NO_2 and FNRs, to coarser grids in order to increase signal-to-noise aided in reducing the bias standard deviation in HCHO retrieval~~

~~products (see Table 2).Furthermore, these two studies agree with our analysis that TROPOMI HCHO has lower bias standard deviation, and higher correlations with observations, compared to both OMI products evaluated here. The larger spread in tropospheric HCHO from OMI compared to TROPOMI is likely due to the weaker signal-to-noise in OMI and potentially the fewer co-located data points for statistical analysis. This is further demonstrated by the TROPOMI bias standard deviation being nearly a factor of two smaller compared to NASA OMI and QA4ECV when averaged to the OMI spatial resolution. TROPOMI HCHO retrievals have the smallest median bias and bias standard deviation compared to observations, and highest correlation with airborne observations, suggesting this newer sensor can better retrieve HCHO compared to OMI during this time period.~~

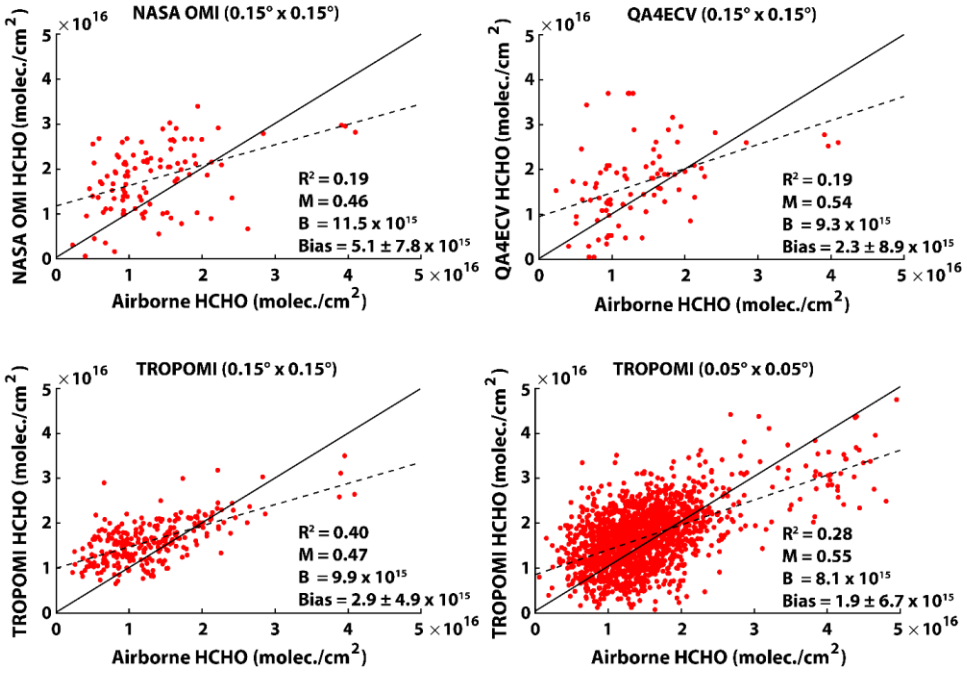


Figure 3: Scatter plots illustrating the comparison of satellite- (NASA OMI, QA4ECV OMI, and TROPOMI) and airborne-retrieved tropospheric HCHO (molecule cm^{-2}) for each co-located measurement taken during the field campaign. All co-located OMI and airborne remote-sensing tropospheric column HCHO values are averaged at the 0.15° × 0.15° resolution and TROPOMI co-located data are averaged at 0.15° × 0.15° and 0.05° × 0.05° spatial resolution. The solid black line shows the 1:1 comparison and the dashed line shows the linear regression fit of the comparison. The figure inset shows the main statistics (coefficient of determination (R^2), slope (M), y-intercept (B), and median bias and bias standard deviation) of the comparison of satellite and airborne tropospheric column HCHO data.

3.~~2~~2.~~3~~ 5 Tropospheric column FNR ~~retrievals~~ systematic bias

The spatial distribution of tropospheric FNRs observed by aircraft measurements during LISTOS 2018 was discussed previously (see Sect. 3.1). Here we evaluated the accuracy of NASA OMI, QA4ECV OMI, and TROPOMI retrieved FNRs compared to observations. Figure 4 shows the scatter plot comparison of co-located NASA OMI, QA4ECV OMI, and TROPOMI retrievals of tropospheric column FNRs compared to observed data (statistical evaluation shown in Table 2). NASA OMI displays a campaign-averaged median bias of 0.4±3.8 (11.0%) and QA4ECV OMI data resulted in a campaign-averaged median bias of -0.2±3.3 (-5.4%). TROPOMI retrievals had a campaign-averaged median bias of 0.4±2.3 (13.0%) and 0.3±1.4 (9.3%) when averaged at 0.05° × 0.05° and 0.15° × 0.15° spatial

resolution, respectively. NASA OMI, QA4ECV OMI, and TROPOMI FNR retrievals had similar median biases compared to observations when averaged at coarser spatial resolutions (see Table 2). Regardless of how tropospheric column NO_2 and HCHO compared to observations, all satellite products evaluated here resulted in campaign-averaged median biases ≤ 0.4 suggesting that the ~~mean of~~systematic/median biases in the individual proxy species ~~can~~ for NASA and QA4ECV OMI and TROPOMI offset to result in accurate median campaign-averaged FNR values. Visual inspection of TROPOMI and QA4ECV OMI retrievals suggests that these two products have the best ability to replicate the lowest observed FNRs over NYC during the field campaign (see Fig. 1). However, besides NASA OMI retrievals, the satellite products have linear regression slopes < 1.0 indicating a high bias for lower FNR values and some small low bias for higher observed FNRs. NASA OMI had a constant offset (slope = 1.0) of 0.8 for all values of observed FNRs.

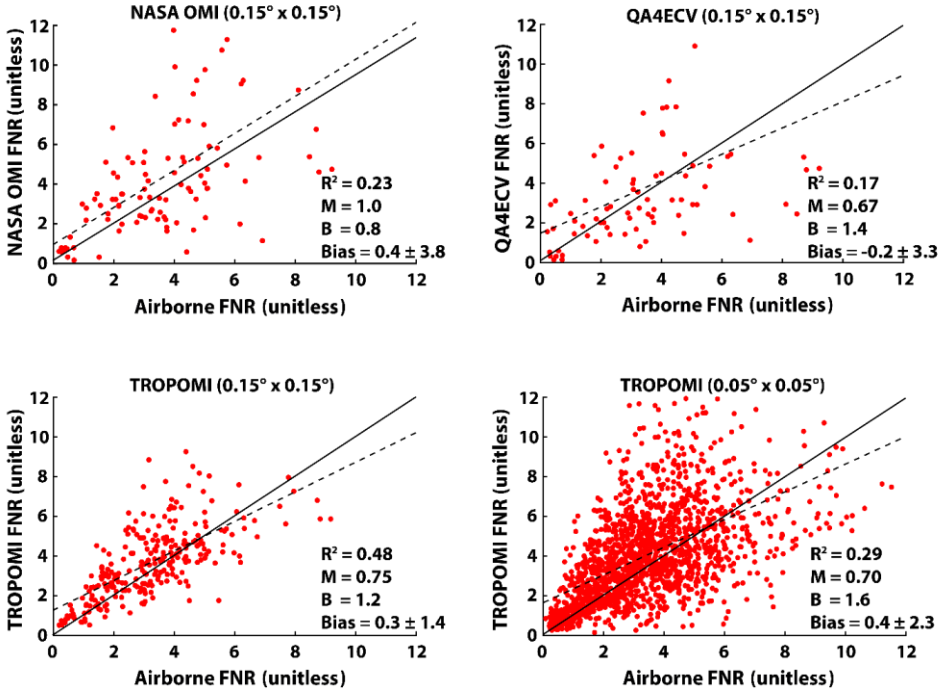


Figure 4: Scatter plots illustrating the comparison of satellite- (NASA OMI, QA4ECV OMI, and TROPOMI) and airborne-retrieved tropospheric FNR (unitless) for each co-located measurement taken during the field campaign. All co-located OMI and airborne remote-sensing tropospheric column FNR values are averaged at the 0.15° × 0.15° resolution and TROPOMI co-located data are averaged at 0.15° × 0.15° and 0.05° × 0.05° spatial resolution. The solid black line shows the 1:1 comparison and the dashed line shows the linear regression slope of the comparison. The figure inset shows the main statistics (coefficient of determination (R^2), slope (M), y-intercept (B), and median bias and bias standard deviation) of the comparison of satellite and airborne tropospheric column FNR data.

~~Visual inspection of TROPOMI and QA4ECV OMI retrievals suggests that these two products have the best ability to replicate the lowest observed FNRs over NYC during the field campaign (see Fig. 1). However, besides NASA OMI retrievals, the satellite products have linear regression slopes < 1.0 indicating a high bias for lower FNR values and some small low bias for higher observed FNRs. NASA OMI had a constant offset (slope = 1.0) of 0.8 for all values of observed FNRs.~~

3.2.6 Tropospheric column FNR uncertainty

The results of this study emphasize that the ability of satellites to accurately observe spatiotemporal patterns of daily FNRs is dependent on retrievals of both tropospheric column HCHO and NO_2. All three satellite products displayed high correlation with tropospheric column NO_2 observations, suggesting these spaceborne sensors can accurately assess the spatial and temporal patterns of this species. However, all the satellite products had very low correlation and high RMSE values when compared with observations of tropospheric HCHO~~, directly resulting in the low correlation of satellite FNR values compared to observations~~. In fact, the rank in correlation levels of all four FNR satellite products evaluated here directly matches the rank in correlation levels of tropospheric HCHO. This leads to the conclusion that given bias variability and RMSE in satellite tropospheric HCHO are large due to noise in the retrieval and low measurement sensitivity of shorter wavelengths in the troposphere, and they directly drive~~s~~ the ~~bias variability~~uncertainty ~~-~~ in FNR retrievals, satellite HCHO observations are the limiting factor of using spaceborne retrievals to accurately assess daily FNRs for investigating O_3 chemistry and sensitivity regimes.

An interesting finding of this study is that the systematic/median bias of OMI and TROPOMI HCHO and NO_2 tropospheric columns tend to cancel out resulting in low median biases for FNRs during LISTOS. However, the unresolved biases in HCHO and NO_2 retrievals when compared to airborne observations do not cancel out. This is clear as the RMSE values for FNRs are still large. Furthermore, biases for HCHO and NO_2 retrievals from NASA and QA4ECV OMI and TROPOMI are not correlated with R^2 values <0.05 for all three satellite products. The uncertainty in HCHO and NO_2 retrievals resulted in FNR RMSE values for NASA OMI, QA4ECV OMI, and TROPOMI (at near native spatial resolutions) of 3.8, 3.3, and 2.3, respectively. Spatially-averaging TROPOMI tropospheric column HCHO data was shown to reduce the noise in the data, resulting in FNR RMSE values for TROPOMI at the coarser spatial resolution to be 1.4 which is nearly a factor of two lower compared to OMI data at the same spatial resolution. Overall, the large noise and unresolved error in tropospheric HCHO retrievals directly result in the uncertainty in FNR retrievals. It should be noted that the HCHO validation data from GeoTASO and GCAS are also hindered by weak absorption signatures in the shorter UV/VIS wavelengths and could add to the bias variability and RMSE values derived in this study. However, the level of ~~bias variability~~uncertainty ~~-~~of tropospheric column HCHO data from OMI and TROPOMI derived in this study ~~agrees~~ are generally consistent with other recent studies (e.g., Vigouroux et al., 2020; De Smedt et al., 2021)~~ which used other sources of evaluation data~~; therefore, we feel the conclusions drawn here are robust.

It is expected that HCHO retrievals will be inherently nosier compared to NO_2. There are two main reasons for this: a) optical depths for HCHO peak in the UV range (<380 nm) at the same wavelengths coinciding with large Rayleigh scattering and optical depths of O_3 leading to a weak/noisy signal, and b) the stronger NO_2 optical depths in the visible wavelength range (400-500 nm), where there are higher signal-to-noise ratios, permits retrievals with less noise. To further evaluate the comparison of uncertainty in remote-sensing of NO_2 and HCHO we compared GCAS/GeoTASO precision levels for the two species. Nowlan et al. (2018) derived the precision of the airborne remote-sensing systems used for NO_2 and HCHO retrievals in this study. Nowlan et al. (2018) quantified precisions of 1.0×10^{15} molecules cm^{-2} and 1.9×10^{16} molecules cm^{-2} at a fine spatial resolution of 250 m × 500 m for NO_2 and HCHO, respectively. Averaging the data to the spatial resolution of 0.05° × 0.05° improves these precision levels to

6.4×10^{13} molecules cm^{-2} and 1.2×10^{15} molecules cm^{-2} for NO_2 and HCHO, respectively. The campaign-averaged column NO_2 and HCHO abundances from GCAS/GeoTASO at 0.05° × 0.05° were 6.6×10^{15} molecules cm^{-2} and 1.5×10^{16} molecules cm^{-2}, respectively. Comparing the precision values of Nowlan et al. (2018) to the mean abundances during LISTOS at the same spatial resolution results in mean precision levels of 1% and 8% for NO_2 and HCHO, respectively. Overall, from this analysis it is expected that the HCHO retrievals should have a factor of 5-10 more noise compared to NO_2.

3.3 High pollution case study

During the LISTOS 2018 campaign there were large tropospheric column NO_2 values retrieved on August 24, 2018 by both NASA OMI and TROPOMI. This day was also identified as a day of high NO_2 pollution concentrations, albeit not an O_3 exceedance day, during the campaign by Judd et al. (2020). Figure 5 illustrates the values of tropospheric FNRs retrieved by NASA OMI and TROPOMI and measured by airborne observations on this day. Figure S4 displays the spatial distribution of tropospheric column NO_2 and HCHO from NASA OMI and TROPOMI and the scatter plot comparison to airborne observations. Figure 5 demonstrates that both satellite retrievals and airborne observation data observed large areas of VOC/radical limited O_3 regimes (FNR < 1.0) in the NYC region. On this day, tropospheric column NO_2 values measured by GCAS reached values > 2.0×10^{16} molecules cm^{-2} in large portions of the VOC/radical limited regions. Furthermore, when comparing airborne tropospheric column HCHO values on August 24, 2018 (Fig. S4) to campaign averaged values (Fig. S3), it is clear that HCHO concentrations were lower on this day compared to other days throughout the summer. In combination with the large NO_2 concentration, this further increased the VOC/radical limitation on this day. The low HCHO/VOC concentrations measured by airborne and space based remote sensing products throughout the extensive VOC limited regime could be the reason why a large-scale O_3 exceedance event was not experienced on August 24, 2018 in proximity to NYC.

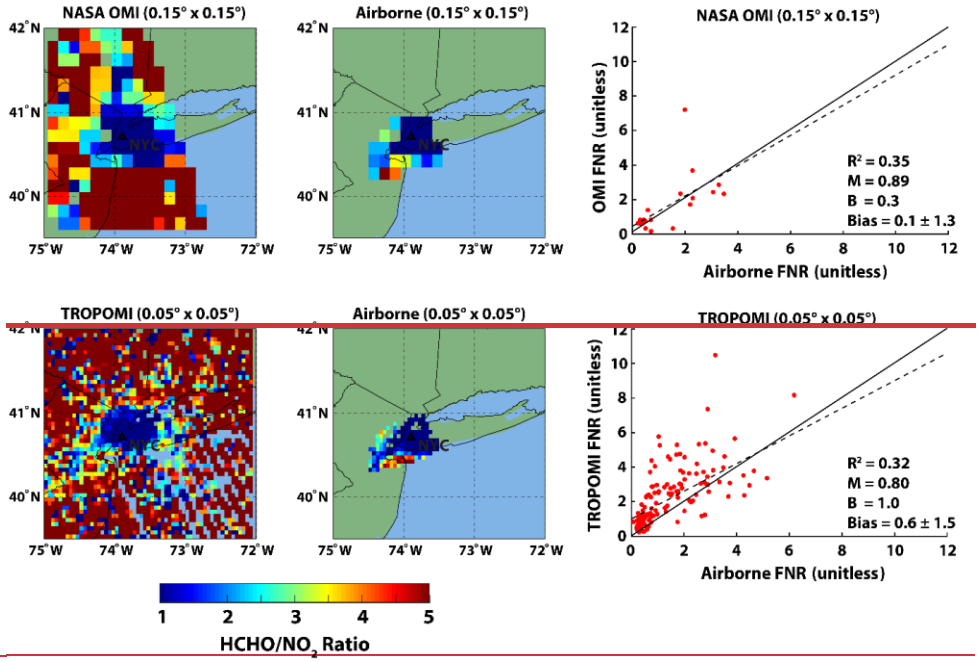
Table 3. Statistical evaluation of NASA OMI and TROPOMI retrievals of tropospheric column NO_2 and HCHO, and resulting FNRs, on August 24, 2018. Statistics presented are number of co-located grids (N), median bias ± bias standard deviation, normalized median bias (NMB, %), coefficient of determination (R^2), and linear regression slope (Slope).

NASA OMI (0.15° × 0.15°)				TROPOMI (0.05° × 0.05°)			
	FNR	HCHO*	NO_2*		FNR	HCHO*	NO_2*
N	20	20	21	N	147	147	154
Bias	0.1±1.3	4.8±4.8	3.5±7.2	Bias	0.6±1.5	4.7±6.3	-0.6±8.8
NMB	9.6	66.1	28.5	NMB	40.9	56.3	-4.3
R^2	0.35	0.25	0.65	R^2	0.32	0.03	0.73
Slope	0.89	1.17	0.49	Slope	0.80	0.33	0.42

*bias units are $\times 10^{15}$ molecules cm^{-2}.

NASA OMI retrievals in the region of lowest FNRs (40.5°N – 41.0°N, 74.0°W – 73.5°W) compared well to observations. In this region, average NASA OMI tropospheric FNRs (0.84) and GCAS observations at 0.15° × 0.15° spatial resolution (1.00) were both ≤ 1.0. TROPOMI retrievals resulted in slightly larger average FNRs (1.15) in this

~~area of VOC/radical limited regions compared to GCAS observations at 0.05° × 0.05° spatial resolution (0.76). Both NASA OMI and TROPOMI had similar median biases (~5.0 × 10^{15} molecules cm^{-2}) when compared to observed tropospheric HCHO; however, TROPOMI had a smaller median bias (0.6±8.8 × 10^{15} molecules cm^{-2}) compared to NASA OMI (3.5±7.2 × 10^{15} molecules cm^{-2}) when evaluated with measured tropospheric NO_2 data (see Table 3). This lower median bias in TROPOMI retrieved tropospheric column NO_2 compared to the same sensor's HCHO statistics led to the slight high bias compared to observed FNRs (0.6±1.5). The similar high biases in NASA OMI tropospheric column NO_2 and HCHO resulted in FNRs which compared well to the observed values (median bias = 0.1±1.3). This further emphasizes the important result of this study that when investigating satellite retrievals of FNRs, that accurate FNRs do not necessarily mean that the particular satellite sensor accurately retrieves both NO_2 and HCHO. For instance, NASA OMI has a NMB of < 10% when compared to GCAS FNRs; however, both tropospheric column NO_2 and HCHO have NMB values of 28.5% and 66.1%, respectively. Furthermore, both NASA OMI and TROPOMI have FNR linear regression slopes ≥ 0.8, suggesting accurate retrievals; however, tropospheric column NO_2 and HCHO data from both satellite products have linear regression slopes which largely deviate from unity (see Fig. S4). Therefore, one should have caution when assuming satellite retrieved accuracy of FNRs as offsetting biases in NO_2 and HCHO can mask errors in both, or individual, retrieved products.~~



~~Figure 5: NASA OMI, TROPOMI, and airborne tropospheric column FNR retrievals on August 24, 2018. All co-located satellite and airborne remote sensing tropospheric column FNR values are averaged at 0.15° × 0.15° for the OMI inter-comparison and 0.05° × 0.05° spatial resolution for TROPOMI. The black triangle indicates the location of the city of NYC. The direct comparison of co-located NASA OMI and TROPOMI FNR data to airborne observations is shown in the scatter plots (right column). The solid black line shows the 1:1 comparison and the dashed line shows the linear regression fit of the comparison. The figure inset shows the main statistics (coefficient of determination (R^2), slope (M), y-intercept (B), and median bias and bias standard deviation) of the comparison of satellite and airborne tropospheric column FNR data.~~

~~A major challenge for accurately retrieving tropospheric FNRs with satellite sensors to evaluated O_3 sensitivity production regimes is the noise in daily retrievals of HCHO due to low signal-to-noise ratios and low measurement sensitivity of shorter UV/VIS wavelengths to HCHO in the troposphere. The noise in both NASA OMI~~

~~and TROPOMI tropospheric column HCHO data on August 24, 2018 can be seen in Fig. S4. Both NASA OMI and TROPOMI HCHO retrievals display low correlation values when compared to observations (see Table 3). Despite TROPOMI having higher correlation with observed NO_2 compared to NASA OMI, the very low correlation of TROPOMI with observed HCHO results in lower correlation and higher bias standard deviations of FNRs compared to NASA OMI. This further emphasizes that the large bias variability, due to noisy data, in retrievals of tropospheric column HCHO are the limiting factor in using spaceborne observations of daily FNRs.~~

3.~~4~~ 3 Common a priori sensitivity test

This section analyzes the impact of using common, high spatial resolution (4 km × 4 km^2), WRF-CMAQ-predicted NO_2 and HCHO vertical profiles as a priori information in NASA OMI and TROPOMI retrievals. GeoTASO and GCAS retrievals were not reprocessed in order to have a consistent reference data set for satellite evaluation. Figure ~~6~~ 5 shows the campaign-averaged FNRs from NASA OMI and TROPOMI retrievals, when reprocessed with WRF-CMAQ NO_2 and HCHO a priori vertical profiles, compared to co-located airborne remote-sensing products (scatter plot comparison displayed in Fig. ~~S5~~S6; statistical evaluation shown in Table ~~3~~3). Comparing NASA OMI FNRs from this figure to Fig. 1, it is evident that using high spatial resolution WRF-CMAQ-predicted NO_2 and HCHO vertical profiles as a priori information resulted in FNR retrievals that are better able to capture the low FNR values (FNR ≤ 1.0) observed around NYC. Reprocessed TROPOMI FNRs also have lower values around NYC; however, were reduced less compared to OMI retrievals. Furthermore, ~~when~~ comparing the results in Fig. ~~S5~~ S6 to Fig. 4 further demonstrates how the reprocessed satellite retrievals better capture the lower FNR values (FNR < 2.0).

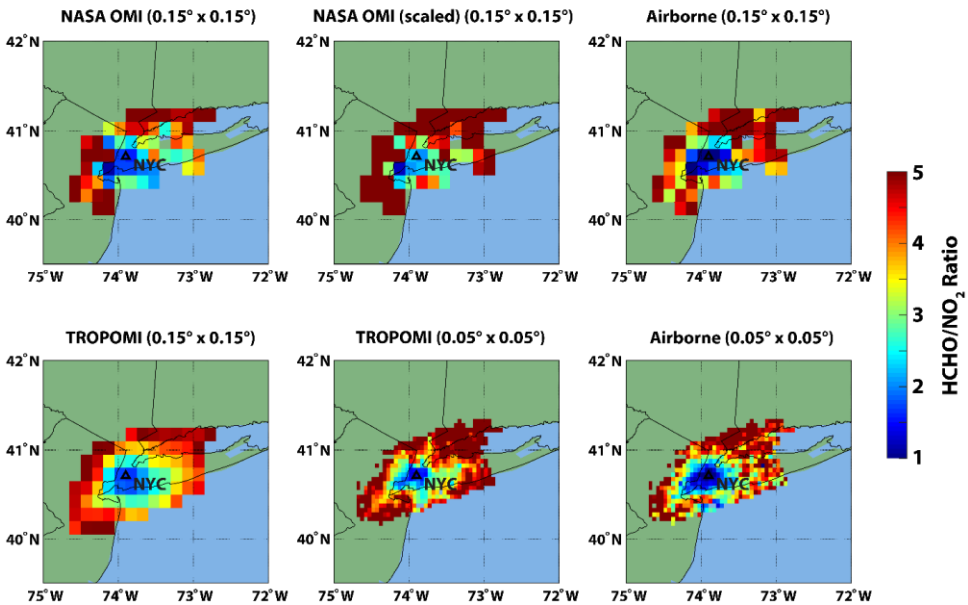


Figure ~~6~~5: NASA OMI and TROPOMI reprocessed tropospheric column FNR retrievals compared to airborne FNR observations averaged for all flights. All co-located OMI and airborne remote-sensing tropospheric column FNR values are averaged at 0.15° × 0.15° and TROPOMI co-locations are averaged at both 0.15° × 0.15° and 0.05° × 0.05° spatial resolution. The OMI FNR retrievals calculated with the scaled WRF-CMAQ profiles are identified as "scaled" in the figure panel titles. The black triangle indicates the location of the city of NYC.

Comparing standard retrieval products from NASA OMI (see Fig. S2 for NO_2 and Fig. S3 for HCHO) to reprocessed retrievals using WRF-CMAQ a priori profiles (see Fig. ~~S6~~ S7 for NO_2 and Fig. ~~S7~~ S8 for HCHO), it is clear that in general the higher spatial resolution model data resulted in larger tropospheric column NO_2 and slightly larger tropospheric column HCHO values. For TROPOMI, reprocessing the retrievals with WRF-CMAQ a priori information caused increases in tropospheric column NO_2 over polluted regions, but small decreases over rural areas characterized by background concentrations. Tropospheric column HCHO data for the reprocessed TROPOMI data were slightly lower in more polluted urban regions near NYC and much lower in the rural areas dominated by background concentrations compared to standard retrievals.

The increases in NASA OMI ~~tropospheric column NO_2 concentrations~~tropospheric NO_2 columns resulted in a small negative bias in FNR retrievals (-0.3±3.9), compared to a small positive bias in the standard products (0.4±3.8). When compared to airborne observations the reprocessed NASA OMI tropospheric column NO_2 data displays a large positive median bias (3.1±5.1 × 10^{15} molecules cm^{-2}) which was not evident in the standard retrieval products. Similarly, for evaluation of the reprocessed NASA OMI tropospheric column HCHO data, a higher positive bias (8.6±7.8 × 10^{15} molecules cm^{-2}) was calculated compared to observations. It should be noted, as previously discussed, that ~~offsetting high~~ systematic/median biases in both reprocessed NASA OMI tropospheric column NO_2 and HCHO retrievals offset ~~resulted~~ resulting in ~~mean~~ median FNR values that compared relatively well to observations. However, the uncertainty in reprocessed satellite HCHO and NO_2 retrievals did not cancel out resulting in FNR RMSE values which were still large for NASA OMI (3.9) and TROPOMI (3.5).

Table ~~4~~3. Statistical evaluation of NASA OMI and TROPOMI retrievals of tropospheric column NO_2 and HCHO, and resulting FNRs, when reprocessed with high spatial resolution WRF-CMAQ a prior information. Statistics presented are the number of co-located grids (N), mean concentration ± standard deviation from satellite (Sat Conc.), median bias ± bias standard deviation, NMB (%), RMSE, coefficient of determination (R^2), and linear regression slope (Slope).~~Statistics presented are number of co-located grids (N), median bias ± bias standard deviation, normalized median bias (NMB, %), coefficient of determination (R^2), and linear regression slope (Slope).~~

~~NASA OMI (0.15° × 0.15°)~~				~~Scaled NASA OMI (0.15° × 0.15°)[1]~~			
	~~FNR~~	~~HCHO*~~	~~NO_2*~~		~~FNR~~	~~HCHO*~~	~~NO_2*~~
~~N~~	~~101~~	~~101~~	~~116~~	~~N~~	~~101~~	~~101~~	~~116~~
~~Bias~~	~~-0.3±3.9~~	~~8.6±7.8~~	~~3.1±5.1~~	~~Bias~~	~~0.5±3.2~~	~~4.4±7.1~~	~~-0.3±3.9~~
~~NMB~~	~~-9.4~~	~~65.7~~	~~50.0~~	~~NMB~~	~~16.7~~	~~35.6~~	~~-4.2~~
~~R^2~~	~~0.17~~	~~0.30~~	~~0.65~~	~~R^2~~	~~0.21~~	~~0.25~~	~~0.67~~
~~Slope~~	~~0.85~~	~~0.70~~	~~1.03~~	~~Slope~~	~~1.05~~	~~0.50~~	~~0.76~~
~~TROPOMI (0.15° × 0.15°)~~				~~TROPOMI (0.05° × 0.05°)~~			
	~~FNR~~	~~HCHO*~~	~~NO_2*~~		~~FNR~~	~~HCHO*~~	~~NO_2*~~
~~N~~	~~261~~	~~261~~	~~261~~	~~N~~	~~1693~~	~~1741~~	~~1802~~
~~Bias~~	~~-0.3±1.4~~	~~-1.2±5.1~~	~~0.1±3.8~~	~~Bias~~	~~0.2±2.2~~	~~-0.1±6.3~~	~~-0.4±4.1~~
~~NMB~~	~~-9.1~~	~~-9.4~~	~~2.0~~	~~NMB~~	~~4.7~~	~~-0.3~~	~~-6.4~~
~~R^2~~	~~0.43~~	~~0.35~~	~~0.61~~	~~R^2~~	~~0.32~~	~~0.32~~	~~0.67~~
~~Slope~~	~~0.67~~	~~0.41~~	~~0.55~~	~~Slope~~	~~0.74~~	~~0.58~~	~~0.61~~

NASA OMI (0.15° × 0.15°)				Scaled NASA OMI (0.15° × 0.15°)[1]			
	FNR	HCHO*	NO_2*		FNR	HCHO*	NO_2*
N	101	101	116	N	101	101	116
Sat. Conc.	3.5±4.3	20.4±8.9	10.5±8.5	Sat. Conc.	3.1±3.6	15.8±7.1	6.3±6.2
Bias	-0.3±3.9	8.6±7.8	3.1±5.1	Bias	0.5±3.2	4.4±7.1	-0.3±3.9
NMB	-9.4	65.7	50.0	NMB	16.7	35.6	-4.2
RMSE	3.9	10.6	6.7	RMSE	3.5	7.8	3.9
R^2	0.17	0.30	0.65	R^2	0.21	0.25	0.67
Slope	0.85	0.70	1.03	Slope	1.05	0.50	0.76
TROPOMI (0.15° × 0.15°)				TROPOMI (0.05° × 0.05°)			
	FNR	HCHO*	NO_2*		FNR	HCHO*	NO_2*
N	261	261	261	N	1693	1741	1802
Sat. Conc.	3.2±1.7	12.8±4.4	6.0±4.3	Sat. Conc.	3.4±2.6	14.6±6.8	6.6±5.2
Bias	-0.3±1.4	-1.2±5.1	0.1±3.8	Bias	0.2±2.2	-0.1±6.3	-0.4±4.1
NMB	-9.1	-9.4	2.0	NMB	4.7	-0.3	-6.4
RMSE	1.4	5.2	3.8	RMSE	2.2	6.3	4.1
R^2	0.43	0.35	0.61	R^2	0.32	0.32	0.67
Slope	0.67	0.41	0.55	Slope	0.74	0.58	0.61

*concentration, bias, and RMSE~~bias~~ units are $\times 10^{15}$ molecules cm^{-2}.

[1]reprocessed with "scaled" CMAQ a priori profiles.

The larger ~~tropospheric column NO_2 concentrations~~tropospheric NO_2 columns in reprocessed NASA OMI data using high spatial resolution model data as a priori information was also shown in past studies (e.g., Souri et al., 2016; Goldberg et al., 2017). Both our study and the work by Goldberg et al. (2017) show that high spatial resolution CMAQ-predicted NO_2 a priori profiles results in OMI ~~tropospheric column NO_2 concentrations~~tropospheric NO_2 columns that are as high as a factor of 2 larger than the standard retrievals. This high bias is caused by smaller AMFs calculated due to the shape factor of high spatial resolution CMAQ-predicted NO_2 concentrations having a too steep NO_2 gradient. The steeper shape factor is caused by higher NO_2 concentrations in the PBL and lower values in the free troposphere compared to the a priori profiles used in standard NASA OMI retrievals. The change in HCHO shape factors when using WRF-CMAQ a priori profiles resulted in slightly higher ~~tropospheric column HCHO concentrations~~tropospheric HCHO columns when compared to standard products for the same reason as tropospheric column NO_2. Similar to Goldberg et al. (2017), we used airborne in situ observations of NO_2 and HCHO from LISTOS 2018 and the Ozone Water-Land Environmental Transition Study 2 (OWLETS-2, https://www-air.larc.nasa.gov/missions/owlets/) field campaigns, OWLETS-2 took place just prior to LISTOS-2018 during the summer of 2018 in the Baltimore, MD region, to correct the model-predicted a priori profiles for use in NASA OMI retrievals and is discussed later in this section.

TROPOMI reprocessed retrievals at 0.05° × 0.05° spatial resolution displayed improved performance when compared to all standard retrieval products of HCHO and FNR. ~~Tropospheric column NO_2 concentrations~~Tropospheric NO_2 columns in reprocessed TROPOMI retrievals resulted in a slightly lower median

biases (-0.4±4.1 × 10^{15} molecules cm^{-2}) compared to the standard products (-0.3±3.7 × 10^{15} molecules cm^{-2}) with slightly larger RMSE values in the reprocessed NO_2 retrievals. Reprocessing TROPOMI retrievals of tropospheric column HCHO resulted in smaller concentrations and ~~much~~ improved median biases ~~and bias standard deviation~~ (-0.1±6.3 × 10^{15} molecules cm^{-2}) and RMSE values (6.3 × 10^{15} molecules cm^{-2}) compared to the median bias (1.9±6.7 × 10^{15} molecules cm^{-2}) and RMSE (6.7 × 10^{15} molecules cm^{-2}) in the standard products ~~(1.9±6.7 × 10^{15} molecules cm^{-2})~~. The good performance of both reprocessed TROPOMI NO_2 and HCHO resulted in FNR values with a smaller median bias when evaluated with observations (0.2±2.2) compared to standard products (0.4±2.3) and slightly lower RMSE values.

As mentioned earlier, when WRF-CMAQ-predicted a priori profiles were used in NASA OMI retrievals it resulted in smaller AMF calculations compared to standard products, resulting in larger tropospheric column NO_2 and HCHO ~~concentrations~~ and higher biases when evaluated with observations. Following methods similar to Goldberg et al. (2017) we used the University of Maryland (UMD) Cessna 402B airborne observations to apply in situ data observational constraints on the NO_2 and HCHO a priori profiles applied in NASA OMI retrievals. The evaluation of WRF-CMAQ-predicted NO_2 (14 flights during LISTOS 2018 and OWLETS-2) and HCHO (7 flights during LISTOS 2018) vertical profiles using airborne data is displayed in Fig. ~~S8~~S9. The comparison of WRF-CMAQ-predicted NO_2 concentrations to airborne in situ observations emphasizes how the a priori profile vertical gradients from the model runs are too steep. Compared to measured NO_2 values, the model displays a high bias below 1 km agl of ~0.4 ppb which was often > 50% larger than observations. This is in stark contrast to the model performance above 2 km agl where the model has a low bias of -0.2 to -0.4 ppb often approaching 100% lower than observations. For the WRF-CMAQ comparison to airborne in situ HCHO data, the model has a low bias throughout the lower troposphere, with larger low biases near the surface (-3.0 ppb between 0-1 km agl) and smaller low biases in the free troposphere (~-1.3 ppb above 2 km agl). These low biases range between -50 to -100% lower compared to measured values. In addition to biases in emission inventories, chemical mechanisms, and other physiochemical parameterizations applied in CTMs, meteorological predictions by WRF, such as wind speed and direction, must have limited errors in order to accurately predict the horizontal and vertical distribution of NO_2 and HCHO concentration (e.g., Laughner et al., 2016; Liu et al., 2021). Compared to the airborne in situ observations taken during LISTOS 2018 and OWLETS-2, WRF wind speed and direction predictions during this study performed relatively well with median correlation (R) and bias values of 0.70 and 0.63 and ≤1.0 m s^{-1} in the u- and v-wind components, respectively. While the WRF simulations applied in this study capture the spatiotemporal variability and general magnitude of observed wind speed and direction, this does not mean that simulated meteorology did not partially contribute to the errors in vertical NO_2 and HCHO profiles simulated by WRF-CMAQ.

Using this model evaluation, we applied approximated scaling factors to the a priori profiles to reprocess NASA OMI data (hereinafter referred to as "scaled"). Separate scaling factors were applied above and below the PBL, approximated to be at 1.5 km agl, where noticeable differences in model performance were evident. For NO_2, the model displays a high bias in the PBL and a low bias in the free troposphere and we apply a scaling factor of 0.5 to WRF-CMAQ a priori NO_2 profiles in the PBL and 5.0 above the PBL. For HCHO, WRF-CMAQ predictions displayed low biases throughout the lower troposphere, and we applied a scaling factor of 2.0 to WRF-CMAQ a priori profiles

in the PBL and 5.0 above the PBL. These scaling factors are approximations of the model performance and are simply applied to determine the impact of "raw" and "scaled" WRF-CMAQ-simulated a priori profiles in NASA OMI NO_2 and HCHO retrievals. Since the UMD Cessna 402B in situ data have limited spatiotemporal coverage of the LISTOS-2018 and OWLETS-2 domains, we did not want to apply overly specific scaling factors to represent all locations/times studied in this work.

The spatial distribution of FNRs derived from the scaled NASA OMI reprocessed NO_2 and HCHO retrievals is shown in Fig. ~~6~~ 5 (scatter plot comparison displayed in Fig. ~~S5~~S6; statistical evaluation shown in Table ~~4~~3). From Table ~~4~~ 3 and Fig. ~~6~~ 5 it can be seen that the scaled WRF-CMAQ a priori profiles result in higher FNR values and improved tropospheric column NASA OMI NO_2 and HCHO retrievals compared to reprocessed products using the raw model output (see Fig. ~~S6~~ S7 and ~~S7~~S8). Scaled NASA OMI tropospheric column NO_2 and HCHO retrievals had ~~much~~ smaller median biases ~~and bias standard deviations~~ of -0.3±3.9 × 10^{15} molecules cm^{-2} and 4.4±7.1 × 10^{15} molecules cm^{-2} and much lower RMSE values of 3.9 × 10^{15} molecules cm^{-2} and 7.8 × 10^{15} molecules cm^{-2}, respectively, compared to the retrievals with raw WRF-CMAQ predictions. This result demonstrates the need for accurate shape factors (i.e., vertical distribution of trace gases) to be used as a priori information in NASA OMI retrievals. Finally, the improved accuracy of tropospheric column NO_2 and HCHO retrievals using scaled WRF-CMAQ model predictions resulted in a slightly higher magnitude of FNR median bias (0.5±3.2); however, with lower RMSE values, compared to reprocessed data using raw CMAQ predictions. In comparison to standard NASA OMI products, the reprocessed satellite data using scaled WRF-CMAQ a priori information had similar median biases in FNR values and lower median biases for HCHO (4.4±7.1 × 10^{15} molecules cm^{-2}) and NO_2 (-0.3±3.9 × 10^{15} molecules cm^{-2}). All reprocessed data variables using scaled model simulated shape factors, due to the reduction in uncertainty in retrieve HCHO and NO_2 data, had lower RMSE values, higher correlation (except for FNR), and similar to better linear regression slopes compared standard satellite retrievals.

3.4 Discussion of satellite errors and capabilities

3.4.1 Relative error of FNR calculations

There are numerous sources of error when using satellite retrievals of tropospheric column HCHO and NO_2 for investigating surface-level or PBL O_3 production sensitivity regimes. The primary uncertainty sources are using indicator species to infer the complex chemistry driving O_3 production and destruction, horizontal spatial representation error, uncertainty in converting tropospheric columns to PBL and surface-level values, and satellite retrieval unresolved biases (Souri et al., 2022a). As for the impact of satellite retrieval errors on the interpretation of O_3 sensitivity, the recent study by Souri et al. (2022a) shows that satellite retrievals errors, in particular the unresolved error in HCHO products, are the largest source of uncertainty in using satellite FNRs to investigate O_3 sensitivity. Here we propagate the uncertainty (RMSE) calculated from NASA OMI, QA4ECV OMI, and TROPOMI to FNR calculations during LISTOS 2018 using Eq. (15) from Souri et al. (2022a) and created maps of the relative error (see Fig. 6). From this figure it can be seen that satellite retrieval errors in HCHO and NO_2 contribute significantly to satellite-derived FNR relative errors. In the largest NO_x emission source regions of NYC, where combined column abundances of HCHO and NO_2 are largest, is where the lowest relative errors of FNRs occur. For TROPOMI, which

has the smallest values of uncertainty/RMSE compared to both NASA and QA4ECV OMI algorithms for HCHO and NO_2, relative errors are as low as ~40%. Away from the emission region of NYC these relative error values reach as high as ~80%. Similar patterns of relative error in FNRs from NASA and QA4ECV OMI retrievals are derived; however, the lowest relative error values over NYC are ~50% and reach values up to 100%. The largest relative errors are seen outside the source region of NYC in QA4ECV OMI retrievals due to having the largest uncertainty in HCHO and lower column abundances of this species in the rural regions of the domain. In addition to the fact that the less noisy retrievals from TROPOMI result in lower relative errors in FNR data, Fig. 6 further demonstrates the larger uncertainty in OMI as the relative error patterns are more heterogeneous. The spatial averaging of TROPOMI data results in the lowest relative errors of all four satellite products discussed in this study. TROPOMI at the coarser (0.15° × 0.15°) spatial resolution had relative errors as low as 35% and only increase to ~60% outside of the source location of NYC.

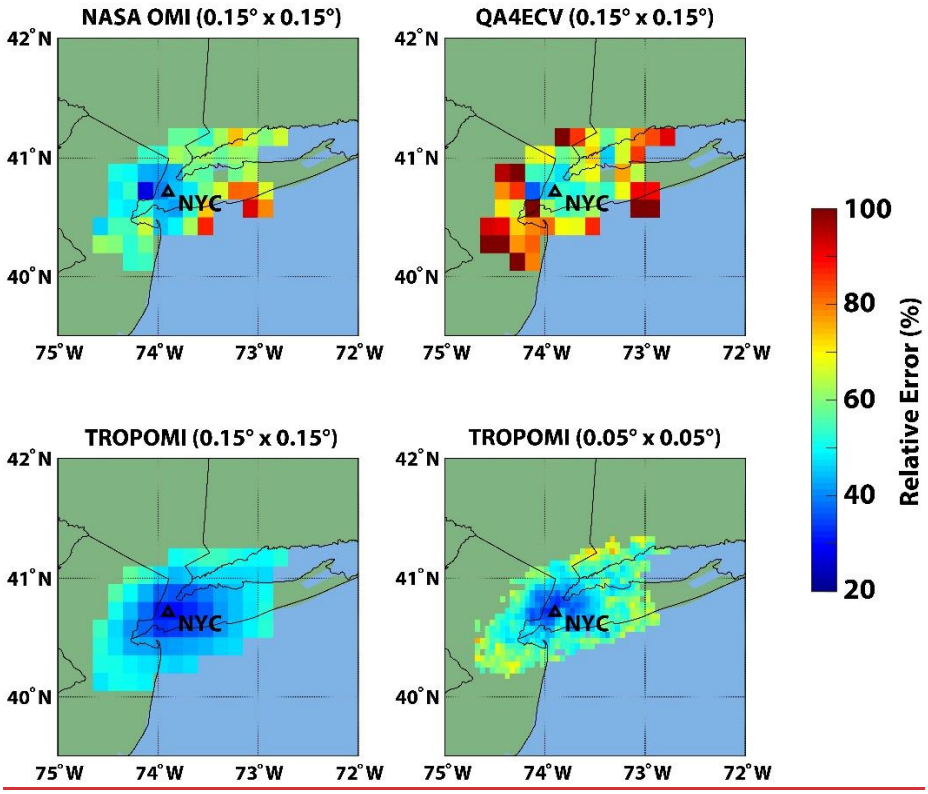


Figure 6: Campaign-averaged relative error in FNR products from NASA OMI, QA4ECV OMI, and TROPOMI due to uncertainty in HCHO and NO_2 retrievals. All co-located OMI and airborne remote-sensing tropospheric column FNR values are averaged at 0.15° × 0.15° and TROPOMI co-locations are averaged at both 0.05° × 0.05° and 0.15° × 0.15° spatial resolutions. The black triangle indicates the location of the city of NYC.

3.4.2 Spatial and temporal capabilities of satellite FNR retrievals

Given the limited spatiotemporal data coverage provided by the LISTOS campaign, a robust understanding of the temporal capabilities of OMI and TROPOMI to retrieve FNRs is not possible. LEO satellites obtain, at best, a single snapshot of both HCHO and NO_2 each day, so one could only hope to obtain daily variability of FNRs from these

spaceborne systems. To determine whether OMI and TROPOMI could capture the variability of the daily mean tropospheric column quantities of NO_2, HCHO, and FNRs over the entire LISTOS domain from airborne data, we compared these daily mean values from NASA OMI, QA4ECV OMI, and TROPOMI to the airborne observations. For NASA OMI, daily correlation (R^2) values were 0.85 (p = 0.001), 0.58 (p = 0.03), and 0.26 (p = 0.20) for NO_2, HCHO, and FNRs, respectively. For QA4ECV OMI, daily correlation values were 0.85 (p = 0.001), 0.80 (p = 0.002), and 0.47 (p = 0.06) for NO_2, HCHO, and FNRs, respectively. For TROPOMI, daily correlation values were 0.92 (p = <0.001), 0.85 (p = <0.001), and 0.41 (p = 0.03) for NO_2, HCHO, and FNRs, respectively. All daily correlation statistics for HCHO and NO_2 were significant to a 95% confidence interval and suggest that both OMI and TROPOMI can capture the overall inter-daily magnitudes of FNR indicator species. However, only TROPOMI could observe the daily variability of domain-wide FNRs within a 95% confidence interval. This suggests that unresolved errors in either HCHO or NO_2 retrievals (the analysis from this study suggests uncertainty in HCHO are driving FNR bias variability) from OMI, using both the NASA and QA4ECV algorithms, are too large to confidently capture the inter-daily variability in FNRs.

The same analysis was conducted for NASA and QA4ECV OMI except just for retrievals near the large anthropogenic source regions in NYC (within 0.35 degrees of the city center) where relative errors due to satellite retrievals for FNR calculations were the lowest (see Fig. 6). Daily correlation (R^2) values for FNR retrievals near the source region of NYC for NASA OMI (0.13; p-value = 0.39) were reduced compared to domain-wide means and QA4ECV OMI (0.66; p-value = 0.01) correlations were improved near the source region of NYC. Indicator species correlation values from NASA OMI were degraded compared to the domain-wide analysis suggesting that this satellite product may not be able to capture inter-daily variability of FNRs even in large source regions. However, this analysis suggests that QA4ECV OMI data has the capability to retrieve daily variability of FNRs in large emission regions such as NYC to a statistically significant level. Overall, TROPOMI retrievals at both fine and coarse spatial resolutions evaluated in this study are able to capture daily variability of tropospheric FNRs over the entire domain and emission source regions better compared to OMI products.

Recent studies have shown that averaging OMI data (especially HCHO retrievals) for longer temporal periods can reduce the noise and uncertainty in this data product. For example, in the recent paper by Souri et al. (2022a), it was shown that unresolved errors in OMI HCHO can be reduced in monthly-averages compared to daily retrievals by ~33% while there was little improvement in uncertainty statistics of NO_2 retrievals from OMI. However, recent studies (e.g., Schroeder et al., 2017) have also shown that for trend studies, monthly-averaging column FNR data can mask FNR temporal gradients that exist within that period. This could hinder the results of trend studies of pollution conditions on O_3 exceedance days, and days of lower pollution, which is a primary purpose of using satellite column FNR data.

To understand the extent to which OMI and TROPOMI retrieval products lose spatial information (variance) compared to airborne data during the LISTOS campaign, we applied the algorithm named SpaTial Representation Error EstimaTor (STREET) (Souri, 2022) using NASA OMI and TROPOMI retrieval data. This method creates semivariograms determining the changes in spatial variability with distance for a defined variable (for this case we used tropospheric column HCHO and NO_2). The maximum variance at which the modeled semivariogram levels off

is defined as a sill and data sets with larger sill values possesses richer spatial information. Figure S10 shows semivariograms, and the fitted stable Gaussian function described in Souri et al. (2022a), applied to TROPOMI and NASA OMI compared to airborne NO_2 columns. Concerning the comparison of TROPOMI and airborne data at 0.05° × 0.05° resolution, we observe airborne semivariogram as high as 20×10^{15} molecules cm^{-2}, a factor of two larger than what TROPOMI achieves. At a ~20 km length scale, TROPOMI can only observe ~40% of the airborne spatial variance, indicating that the spatial representation error in TROPOMI is ~60% at this scale. Similarly, NASA OMI fails to recreate >50% of the maximum variance observed in airborne data at 0.15° × 0.15° resolution. At ~20 km length scale, the spatial loss of OMI is >70%.

Figure S10 depicts the semivariograms and fitted exponential curves applied to TROPOMI and airborne HCHO columns. Immediately evident is that both semivariograms level off at longer distances compared to the analysis of NO_2. This stems from the fact that HCHO columns tend to be spatially more homogeneous in the region of the LISTOS domain. For most length scales, TROPOMI can relatively well replicate the spatial variance observed in airborne data (~70%), which is explainable by the fact that HCHO concentrations are not highly heterogeneous in this region. We do not present the semivariogram for NASA OMI HCHO columns as the underlying unresolved biases in OMI are very large, introducing artifacts that cannot be solely attributable to unresolved spatial scales. Overall, TROPOMI and OMI capture spatial variance of NO_2 similarly, TROPOMI performs slightly better; however, OMI is unable to capture the spatial variability of observed HCHO due to unresolved biases in this retrieval product. Since TROPOMI is able to capture the observed HCHO variability to a sufficient degree, combing these two facts suggest that TROPOMI has better capability to retrieve FNR spatial variability compared to OMI products.

3.4.3 Reasons for systematic bias and uncertainty in FNRs

As demonstrated in this study, median biases of OMI and TROPOMI HCHO and NO_2 retrievals tend to cancel out when calculating tropospheric column FNRs. Figures S4 and S5 show that the median bias spatial distribution of all satellite HCHO and NO_2 retrievals are similar with a small low median bias in column abundances near the source region of NYC and high biases in the background regions. Table S1 shows that AMF calculations from NASA OMI, QA4ECV OMI, and TROPOMI use many of the same input data sets for geophysical variables (e.g., surface albedo, cloud fraction, cloud radiance, etc.) resulting in campaign-averaged AMFs of HCHO, NO_2, and the ratios of these products (AMF FNRs) which are relatively similar across the LISTOS domain (see Fig. S11). For all satellite products, HCHO and NO_2 AMFs have much less variability compared to AMFs derived for airborne data which along with SCD biases may contribute to the median high biases in background HCHO and NO_2 retrievals. A primary reason for the inability of satellites to capture AMF variability over the LISTOS domain is likely the shape factors being used for these calculations having spatial resolutions of 1.0° × 1.0° to even coarser grids (Table S1). Furthermore, while TROPOMI and QA4ECV OMI retrievals used daily model data for shape factor calculations, NASA OMI uses monthly products which will be challenged to capture the large spatiotemporal variability of tropospheric HCHO and NO_2 vertical profiles in urban and rural regions occurring in reality. Finally, coarse geophysical input data sets used in AMF calculations (see Table S1) will not capture the spatial distribution of these variables in reality. Airborne AMF calculations use much higher spatial resolution input data sets (e.g., 500 m surface albedo data (Judd et al., 2020)

compared to 0.5° × 0.5° or coarser surface reflectivity products used in OM and TROPOMI) and shape factors are calculated with 12 km × 12 km CMAQ model simulations which both aid in the much larger spatial variability of AMFs not captured in satellite retrievals.

The more interesting aspect found in this study is that unresolved errors in HCHO and NO_2 retrievals don't cancel out in FNR calculations as do the systematic/median biases. While there are some reasons why uncertainty in HCHO and NO_2 retrievals could stem from opposite impacts of geophysical parameters in AMF calculations, such as AMF uncertainties in HCHO and NO_2 having opposite trends with increasing surface reflectance (comparing Fig. 10 from De Smedt et al. (2018) and Fig. 20 from Liu et al. (2021)), these differences are minor and overall AMF calculations for both species in NASA OMI, and QA4ECV OMI, and TROPOMI have similar input data sets. A portion of the uncertainty of HCHO and NO_2 retrievals not canceling out stems from the AMF calculations shown in Fig. S11. In order for HCHO and NO_2 AMFs to have no impact on VCD uncertainty cancelations, AMF FNRs would be a constant or similar value at all locations. However, from Fig. S11 it is shown that AMF FNRs, while having smooth spatial variability, are not a constant value. Therefore, some of the unresolved error residual in the FNR calculations will be due to differences in HCHO and NO_2 AMF calculations. This is emphasized in NASA OMI AMF FNR plots in Fig. S11 where different CTMs, at different spatial resolutions (see Table S1), are used to derive HCHO and NO_2 shape factors leading to noticeable differences in the respective AMF calculations. This likely is one of the reasons that NASA OMI FNRs have the largest uncertainty (highest bias standard deviation and RMSE values) compared to airborne data (see Table 2) of all OMI and TROPOMI satellite products. Finally, the airborne AMFs are more variable compared to satellite products due to the finer-scale shape factors and geophysical parameter input data used in AMF calculations which satellites inherently are not able to capture, contributing to the satellite uncertainty.

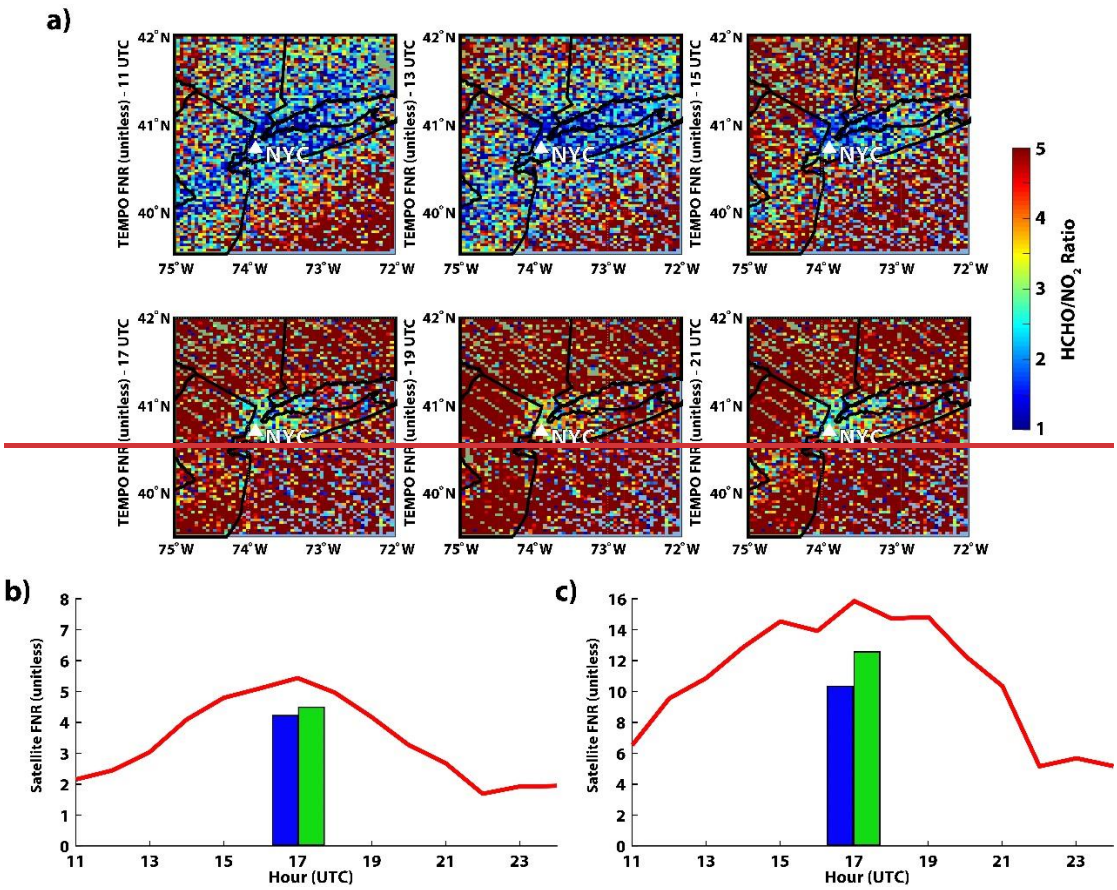
The rest of the remaining unresolved error in FNR calculations is likely due to the SCD retrievals from OMI and TROPOMI sensors. As demonstrated in this study the uncertainty in both OMI and TROPOMI retrievals of HCHO is large. The SCD retrievals of HCHO from TROPOMI have been shown in the past to have less noise compared to OMI due to the higher spatial resolution and at least the same signal-to-noise (De Smedt et al., 2021). The larger uncertainty in OMI retrievals of HCHO compared to TROPOMI directly leads to the higher bias standard deviation and RMSE values for derived FNRs in OMI compared to TROPOMI (see Table 2). This is further emphasized in the spatially-averaged TROPOMI data (at 0.15° × 0.15° to match OMI data) where HCHO and FNR retrievals have a factor of 2-3 lower RMSE compared to NASA OMI and QA4ECV OMI. TROPOMI NO_2 SCDs have also been shown to have less noise compared to OMI retrievals due to the higher spatial resolution and similar signal-to-noise (van Geffen et al., 2020, 2022). This is also shown in Table 2 when averaging TROPOMI data to match the OMI spatial resolution. Overall, HCHO and NO_2 SCD noise contributes to uncertainty in OMI and TROPOMI VCDs and are not cancelled out in FNR calculations; however, the reduced noise in TROPOMI SCD retrievals leads to improved VCDs of HCHO and NO_2 abundances and the ratios of these products.

~~3.5 Expected FNR information from TEMPO~~

~~TEMPO is expected to provide revolutionary information about air quality in North America (Zoogman et al., 2017; Chance et al., 2019). This geostationary sensor will provide tropospheric column NO_2 and HCHO data, and resulting~~

FNR products, up to every 1-2 hours during daylight hours. Here we demonstrate the expected improvement in the information content of tropospheric FNRs due to the diurnal retrievals of TEMPO compared to low earth orbit sensors (e.g., OMI and TROPOMI) which retrieve a single snapshot at ~13:30 local time. Synthetic OMI and TROPOMI FNRs are derived here by averaging the synthetic TEMPO data at 0.13° × 0.25° and 0.07° × 0.05°, respectively (representative of these sensor's native spatial resolution). This was done to provide synthetic FNR data from the three sensors which only differ based on the spatiotemporal sampling frequency and not retrieval specifics. Once TEMPO is launched, studies should be conducted to determine the difference in tropospheric FNR values due to the actual retrieval products between this geostationary sensor and other low earth orbit satellites.

One of the main improvements in tropospheric FNR information that is expected from TEMPO compared to low earth orbit sensors is the additional data throughout the diurnal cycle. Figure 7a shows the spatial distribution of 2-hour averaged synthetic TEMPO FNRs averaged at 0.03° × 0.05° spatial resolution (representative of the native spatial resolution of TEMPO) on July 12, 2020. This day was chosen due to data availability and the limited cloud coverage simulated for the synthetic product on this day. This figure also shows the time series of FNRs retrieved by TEMPO, and 13:30 retrieved OMI and TROPOMI FNRs, averaged for an urban region within 0.25° of NYC (Fig. 7b) and a rural region 1° north and west of NYC. It is clear that significant information about tropospheric column NO_2 and HCHO, and resulting FNRs, which will be used to investigate O_3 production sensitivity regimes will be gained when TEMPO is launched compared to OMI and TROPOMI. Due to emissions and chemical production/destruction of NO_2 and HCHO, tropospheric FNRs vary significantly throughout the day (by around a factor of 3 in the vicinity of NYC on this day, see Fig. 7b) with large swaths of VOC/radical-limited regions in the northeast US in the morning hours which transition to NO_x-limited regimes during the afternoon. Rural regions also display a strong diurnal pattern of FNRs; however, with much higher values compared to urban regions due to the lack of significant NO_x emission sources (see Fig. 7c). During the afternoon hours when overpasses of OMI and TROPOMI occur, FNRs are larger compared to morning and evening values which will be retrieved by TEMPO. In addition to the improved temporal resolution of TEMPO, the increased spatial resolution of these retrievals compared to OMI and TROPOMI will also provide improved information on spatial distributions of these proxy species and FNRs.



~~**Figure 7: a) Synthetic TEMPO 2-hour-averaged tropospheric column FNR retrievals between 11 and 21 UTC on July 12, 2020. All synthetic TEMPO tropospheric column FNR data are averaged to a 0.03° × 0.05° spatial resolution. The white triangle indicates the location of the city of NYC. This figure also shows the timeseries of hourly TEMPO FNR values (red line between 11 on July 12, 2020 and 00 UTC on July 13, 2020) averaged b) within 0.25° of NYC and c) a rural location 1° north and west of NYC. The value of tropospheric column FNR retrieved by synthetic OMI (blue bar) and TROPOMI (green bar) at ~13:30 local time (represented by 17 UTC data) are presented in panel b) and c).**~~

4 Conclusions

This study presents a statistical evaluation and inter-comparison of tropospheric FNR retrievals from two commonly applied ~~low-earth-orbit~~LEO sensors for investigating O_3 production sensitivity regimes (i.e., OMI and TROPOMI). The evaluation of NASA OMI, QA4ECV OMI, and TROPOMI retrievals of tropospheric NO_2 and HCHO, and resulting FNRs, was conducted with airborne remote-sensing observations (GeoTASO and GCAS) during LISTOS 2018. Past studies have focused on the evaluation of satellite retrievals of tropospheric column NO_2 and HCHO, individually; however, this is the first study to validate and inter-compare multiple satellite platform's and retrieval algorithm's ability to retrieve tropospheric FNRs and also quantify the impact of horizontal spatial resolution, a priori vertical profile information, and different retrieval algorithms. The quantification of satellite-retrieved tropospheric FNRs biases/errors is currently an important, but relatively unknown, uncertainty when applying spaceborne remote-sensing products to investigate O_3 production regimes.

NASA OMI, QA4ECV OMI, and TROPOMI retrievals reproduce ~~this~~ the general spatial pattern of observed tropospheric FNRs; however, displayed higher FNRs (between 1.0 and 3.0) in the urban regions of NYC where observations suggest NO_x-saturated regimes (FNR < 1.0). The statistical evaluation of these satellite products illustrated that all three retrievals have a high bias of background-level tropospheric column NO_2 and HCHO concentrations. The satellite retrievals compare more accurately to larger tropospheric column NO_2 and HCHO values observed in the moderately polluted areas with a tendency towards a low bias in the more polluted areas. The magnitude-dependent biases for OMI and TROPOMI NO_2 and HCHO derived in this study agrees with other recent validation projects (e.g., Judd et al., 2020; Vigouroux et al., 2020; Zhao et al., 2020; Compernolle et al., 2020; Lamsal et al., 2021; De Smedt et al., 2021; Verhoelst et al., 2021). Both OMI and TROPOMI retrievals compared well to observed NO_2 throughout the campaign; ~~with NMB values < 10%. The~~ however, the statistical comparison with observed HCHO data resulted in larger and more variable biases between the three satellite products. Overall, daily- and campaign-averaged comparisons of the satellite HCHO data to observations displayed large ~~bias standard deviations~~RMSE values emphasizing the large noise in these retrieval products which hinders the accuracy of FNRs from spaceborne sensors. Averaging TROPOMI HCHO data to coarser spatial resolutions, in order to improve signal-to-noise, proved capable to reduce ~~bias standard deviations~~ RMSE values ~~compared~~when compared to observations. While all three satellite products at the near native spatial resolutions had low systematic campaign-averaged FNR median biases < 0.5, suggesting median/systematic biases in HCHO and NO_2 data cancel out, the ~~bias standard deviations were high~~RMSE values for FNRs remained large (> 2.0), primarily due to ~~noise in the~~ uncertainty in HCHO and NO_2 retrievals not offsetting. Given the limited measurement sensitivity of shorter UV/VIS wavelengths to HCHO in the middle to lower troposphere, improved information (in situ, remote-sensing, or models) of the vertical profiles of HCHO to be used as a priori information would benefit satellite remote-sensing capabilities for observing HCHO and FNRs.

The higher spatial resolution of TROPOMI, along with a good signal-to-noise ratio, allows this sensor to better capture the spatiotemporal variability and urban/rural interface of tropospheric column NO_2 and HCHO values and resulting FNRs. This satellite data had the highest correlations with observed NO_2, HCHO, and FNRs throughout the campaign, along with lowest ~~bias standard deviation~~RMSEs of all three satellite products. The added benefit of TROPOMI spatial resolution is important as this sensor has now been operational for 5+ years and can be applied in trend analysis along with case studies. Future studies of FNR trends should include both OMI and TROPOMI retrievals and determine best practices to fuse/link the two data sets.

~~NASA OMI retrievals of tropospheric FNRs had lower median biases and bias standard deviations compared to TROPOMI on a day identified as having high NO_2 pollution (August 24, 2018). However, this sensor did not provide more accurate retrievals of HCHO and NO_2 compared to TROPOMI. This fact, along with results from the campaign-averaged analysis, demonstrates that biases in tropospheric NO_2 and HCHO can offset resulting in accurate FNR retrievals. While accurate FNR retrievals are informative, and necessary to studying regimes of O_3 production sensitivity, the actual magnitudes of tropospheric NO_2 and HCHO concentrations are vital for calculating/investigating quantitative O_3 production rates (Souri et al., 2022a). Therefore, it is important to understand the accuracy of not only a satellite's FNR retrievals, but also the ability to retrieve the magnitudes of both chemical proxy species. Another~~

~~interesting finding during this high NO_2 event was that all satellite and airborne observations measured a large region where O_3 production was likely VOC limited (FNRs < 1.0) in the vicinity of NYC; however, no large scale O_3 exceedance events occurred on this day. Interestingly, all satellite and airborne observations on this day also measured lower than average HCHO concentrations in the vicinity of NYC and these low concentrations of HCHO, a proxy for VOC abundances, in the VOC limited O_3 production regime around NYC could have been the reason why O_3 formation was not elevated on this day.~~

Applying multiple retrieval algorithms to the radiances of a single satellite sensor is of interest in order to determine how input variables (e.g., information on a priori vertical profiles, clouds, surface albedo, etc.) impact the retrieval performance to identify the most accurate data products. In this study, we evaluated results of OMI retrievals applying two well-known retrieval algorithms (i.e., NASA version 4 product and output from the QA4ECV project). Results from the two retrievals were similar for NO_2 but differed primarily in tropospheric column HCHO, where NASA OMI data had a median bias a factor of two larger than QA4ECV data. Both retrieval algorithms resulted in ~~high bias standard deviation~~large RMSE values indicative of the noise ~~of~~ in tropospheric HCHO retrievals. While NASA OMI data displayed less accurate retrievals in HCHO, and similar performance for NO_2, compared to QA4ECV data, NASA OMI data resulted in FNR values with similar median bias and slightly higher ~~bias standard deviations~~RMSEs. Given that both the NASA and QA4ECV retrievals of tropospheric HCHO resulted in noisy data products from OMI ~~(illustrated by large bias standard deviations)~~, this emphasizes the need for improved signal-to-noise and calibration and improved a priori vertical profile information of HCHO to negate the low measurement sensitivity of HCHO in the middle to lower troposphere for future satellite sensors and/or improved retrieval algorithms of HCHO. ~~Addressing this issue, a new SAO OMI collection 4 HCHO retrieval product is planned to be released by the end of 2022 (personal communication with the SAO algorithm team). The new retrieval could represent a step forward in the quality of the OMI HCHO product with improvements in OMI radiance calibration and quality control translating to a more stable and less noisy HCHO retrievals from OMI. Future studies should apply this potentially improved OMI HCHO retrieval product to evaluate the improvement in satellite-derived FNRs.~~

Our study investigated the impact of high spatial resolution WRF-CMAQ-predicted NO_2 and HCHO a priori profiles on OMI and TROPOMI retrievals of FNRs. Using the WRF-CMAQ-predicted a priori information resulted in relatively~~highly~~ accurate retrievals of FNRs with median biases ≤ 0.5 over the entire campaign similar to the accuracy of the operational OMI and TROPOMI retrievals. However, while reprocessed NASA OMI data had only a small low median bias in FNR, the high spatial resolution model data resulted in large high biases in both tropospheric NO_2 and HCHO. These high biases are caused by errors in the shape factor imposed by the model data. We scaled WRF-CMAQ-predicted vertical profiles of a priori NO_2 and HCHO using airborne in situ observations which resulted in improved tropospheric column NASA OMI NO_2 and HCHO retrievals compared to reprocessed products using the raw model output~~in smaller biases in the traces gas retrievals~~. In comparison to standard NASA OMI products the reprocessed satellite data using scaled WRF-CMAQ a priori information had similar median biases in FNR values and lower median biases in both indicator species. All reprocessed data variables using scaled model simulated shape factors due to the reduction in unresolved error in retrieved HCHO and NO_2 data had lower RMSE, higher correlation (except for FNR), and similar to better linear regression slopes compared to observations. TROPOMI reprocessed data

on the other hand had improved performance when using the higher spatial resolution WRF-CMAQ data as the a priori product compared to standard retrievals which apply coarser resolution TM5 output. ~~This demonstrates the need for accurate shape factors (i.e., vertical distribution of trace gases) to be used as a priori information in OMI retrievals. Furthermore, while high spatial resolution CTM simulations likely better capture spatial heterogeneity of trace gases such as NO_2 and HCHO, shape factors imposed by this specific WRF-CMAQ analysis degraded OMI retrieval performance compared to standard data products which use the coarser resolution GMI model as a the a priori. TROPOMI reprocessed data on the other hand had improved performance when using the higher spatial resolution WRF-CMAQ data as the a priori product compared to standard retrievals which apply coarser resolution TM5 output.~~ The fact that TROPOMI native spatial resolution is similar to the WRF-CMAQ resolution used in this study, could have resulted in the better results when reprocessing TROPOMI data compared to OMI. Future studies should investigate the impact of various spatial resolution a priori profile data sets, ranging from the ~1° × 1° GMI and TM5 model data used for OMI and TROPOMI, respectively, to much higher resolution air quality model simulations, on the results of reprocessed satellite NO_2 and HCHO retrievals~~data~~.

Overall, the systematic biases, uncertainty (i.e., RMSE), and correlations presented in this study can be used in future studies when interpreting the accuracy of OMI and TROPOMI retrievals of FNRs, and the two indicator species, used for investigating O_3 sensitivity regimes applying satellite products. A main take away from this study is that it is necessary to statistically evaluate both the tropospheric FNRs, and the NO_2 and HCHO products, individually, as large median biases in both NO_2 and HCHO satellite products can offset resulting in accurate median/mean FNR values. However, this study emphasizes that uncertainty in NO_2 and HCHO satellite retrievals do not offset in OMI or TROPOMI products greatly hindering the accuracy of daily scenes of FNRs from these sensors. The large unresolved biases in tropospheric column HCHO retrievals appear to be the controlling and limiting factor of daily FNR accuracy. The relative error of FNRs from OMI and TROPOMI are ~40-60% in regions of large NO_x emissions; however, approach error levels of 100% or more in regions characterized by background HCHO and NO_2 concentrations. While both TROPOMI and OMI captured some of the spatiotemporal variability of observed NO_2 within the LISTOS domain, only TROPOMI is able to capture spatiotemporal HCHO variability with uncertainty low enough for potentially capturing daily FNR variability. The unresolved error in HCHO retrievals from OMI is too large and likely limits the application of this data on a daily basis near the native spatial resolution of the sensor. Overall, the individual satellite products display varying degrees of capability to retrieve tropospheric FNRs and it is necessary to further validate OMI and TROPOMI retrievals using other field campaign or stationary network data in different regions of the world to determine regional biases, and identify the primary controlling factors of systematic biases and uncertainty.~~Overall, the biases, bias standard deviations, and correlations presented in this study can be used in future studies when interpreting the accuracy of OMI and TROPOMI retrievals of FNRs used for investigating O_3 sensitivity regimes applying satellite products. The individual satellite products display varying degrees of capability to retrieve tropospheric FNRs and it is necessary to further validate OMI and TROPOMI retrievals using other field campaign data in different regions of the world to determine regional biases, and identify the primary controlling factors of systematic and random errors (e.g., cloud fraction, surface albedo, spatial resolution, signal to noise ratios, a priori information, etc.). A main take away from this study is that it is necessary to statistically evaluate

~~both the tropospheric FNRs, and the NO_2 and HCHO products, individually, as large biases in both NO_2 and HCHO satellite products can offset resulting in accurate FNR values.~~ The large bias variability/noise in tropospheric column HCHO retrievals appear to be the controlling and limiting factor of daily FNR accuracy. ~~Our study goes beyond investigating median biases, as the noise in satellite retrievals of HCHO result in large bias standard deviations when compared to observations. The large bias variability/noise in tropospheric column HCHO retrievals appear to be the controlling and limiting factor of daily FNR accuracy.~~

Acknowledgements

Matthew Johnson, Sajeev Philip (grant number: 80NSSC20K1182), Rajesh Kumar (grant number: 80NSSC20K1234), Amir Souri (grant number: 80NSSC21K1333), and Jeffrey Geddes (grant number: 80NSSC20K1033) were funded for this work through NASA's Aura Science Team (NNH19ZDA001N-AURAST). Laura Judd and Scott Janz are collaborators on the NASA Aura Science Team project which funded the majority of this work and their contribution to this study was through in-kind efforts. Sajeev Philip acknowledge support from the NASA Academic Mission Services by Universities Space Research Association at NASA Ames Research Center during the initial stages of this study. Finally, Aaron Naeger is funded through the NASA TEMPO project and his efforts for this study was through in-kind efforts. The authors perceive no financial, or other affiliations, which are conflicts of interest. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at NASA Ames Research Center. The National Center for Atmospheric Research is sponsored by the National Science Foundation. Finally, the views, opinions, and findings contained in this report are those of the authors and should not be construed as an official NASA or United States Government position, policy, or decision.

Data Availability

The primary data sources used in this study were the NASA OMI NO_2 and SAO HCHO (https://earthdata.nasa.gov/; last access: 4/27/2020), QA4ECV OMI NO_2 and HCHO (http://www.qa4ecv.eu/ecvs; last access: 3/3/2020), and TROPOMI PAL NO_2 (https://data-portal.s5p-pal.com/; last access: 12/20/2020) and operational HCHO (https://earthdata.nasa.gov/; last access: 4/27/2020) satellite data. For evaluating these satellite products we use airborne remote sensing data from GeoTASO and GCAS which were downloaded from the LISTOS-2018 campaign data repository (https://www-air.larc.nasa.gov/missions/listos/index.html; last access: 4/21/2020). Finally, the synthetic TEMPO data product was downloaded from: https://weather.msfc.nasa.gov/tempo/data.html; last access: 4/15/2021.

References

Acarreta, J. R., De Haan, J. F., and Stammes, P.: Cloud pressure retrieval using the O_2–O_2 absorption band at 477 nm, J. Geophys. Res.-Atmos., 109, D05204, doi:10.1029/2003JD003915, 2004.

Appel, K. W., Napelenok, S. L., Foley, K. M., Pye, H. O. T., Hogrefe, C., Luecken, D. J., Bash, J. O., Roselle, S. J., Pleim, J. E., Foroutan, H., Hutzell, W. T., Pouliot, G. A., Sarwar, G., Fahey, K. M., Gantt, B., Gilliam, R. C., Heath, N. K., Kang, D., Mathur, R., Schwede, D. B., Spero, T. L., Wong, D. C., and Young, J. O.: Description and evaluation of the Community Multiscale Air Quality (CMAQ) modeling system version 5.1, Geosci. Model Dev., 10, 1703–1732, https://doi.org/10.5194/gmd-10-1703-2017, 2017.

Boersma, K. F., Eskes, H. J., Veefkind, J. P., Brinksma, E. J., van der A, R. J., Sneep, M., van den Oord, G. H. J., Levelt, P. F., Stammes, P., Gleason, J. F., and Bucsela, E. J.: Near-real time retrieval of tropospheric NO_2 from OMI, Atmos. Chem. Phys., 7, 2103–2118, https://doi.org/10.5194/acp-7-2103-2007, 2007.

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, 1905–1928, https://doi.org/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, 6651–6678, https://doi.org/10.5194/amt-11-6651-2018, 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, 2607–2626, https://doi.org/10.5194/amt-6-2607-2013, 2013.

Chan, K. L., Wiegner, M., van Geffen, J., De Smedt, I., Alberti, C., Cheng, Z., Ye, S., and Wenig, M.: MAX-DOAS measurements of tropospheric NO_2 and HCHO in Munich and the comparison to OMI and TROPOMI satellite observations, Atmos. Meas. Tech., 13, 4499–4520, https://doi.org/10.5194/amt-13-4499-2020, 2020.

Chance, K.: Analysis of BrO measurements from the Global Ozone Monitoring Experiment, Geophys. Res. Lett., 25, 3335–3338, doi:10.1029/98GL52359, 1998.

Chance, K., Liu, X., Miller, C. C., González Abad, G., Huang, G., Nowlan, C., Souri, A., Suleiman, R., Sun, K., Wang, H., Zhu, L., Zoogman, P., Al-Saadi, J., Antuña-Marrero, J. C., Carr, J., Chatfield, R., Chin, M., Cohen, R., Edwards, D., Fishman, J., Flittner, D., Geddes, J., Grutter, M., Herman, J. R., Jacob, D. J., Janz, S., Joiner, J., Kim, J., Krotkov, N. A., Lefer, B., Martin, R. V., Mayol-Bracero, O. L., Naeger, A., Newchurch, M., Pfister, G. G., Pickering, K., Pierce, R. B., Rivera Cárdenas, C., Saiz-Lopez, A., Simpson, W., Spinei, E., Spurr, R. J. D., Szykman, J. J., Torres, O., and Wang, J.: TEMPO Green Paper: Chemistry, Physics, and Meteorology Experiments with the Tropospheric Emissions: Monitoring of Pollution Instrument, in: Sensors, Systems, and

Next-Generation Satellites XXIII, edited by: Neeck, S. P., Kimura, T., and Martimort, P., p. 10, SPIE, Strasbourg, France, https://doi.org/10.1117/12.2534883, 2019.

Choi, Y. and Souri, A.: Chemical condition and surface ozone in large cities of Texas during the last decade: observational evidence from OMI, CAMS, and Model Analysis, Remote Sens. Environ., 168, 90–101, doi:10.1016/j.rse.2015.06.026, 2015.

Choi, Y., Kim, H., Tong, D., and Lee, P.: Summertime weekly cycles of observed and modeled NO_x and O_3 concentrations as a function of satellite-derived ozone production sensitivity and land use types over the Continental United States, Atmos. Chem. Phys., 12, 6291–6307, doi:10.5194/acp-12-6291-2012, 2012.

Compernolle, S., Verhoelst, T., Pinardi, G., Granville, J., Hubert, D., Keppens, A., Niemeijer, S., Rino, B., Bais, A., Beirle, S., Boersma, F., Burrows, J. P., De Smedt, I., Eskes, H., Goutail, F., Hendrick, F., Lorente, A., Pazmino, A., Piters, A., Peters, E., Pommereau, J.-P., Remmers, J., Richter, A., van Geffen, J., Van Roozendael, M., Wagner, T., and Lambert, J.-C.: Validation of Aura-OMI QA4ECV NO_2 climate data records with ground-based DOAS networks: the role of measurement and comparison uncertainties, Atmos. Chem. Phys., 20, 8017–8045, https://doi.org/10.5194/acp-20-8017-2020, 2020.

Crutzen, P. J.: Gas-phase nitrogen and methane chemistry in the atmosphere. In Physics and Chemistry of the Upper Atmosphere, Proceedings of a Symposium Organized by the Summer Advanced Study Institute, B.M. McCormac, ed. Dordrecht, Holland: D. Reidel Publishing Co., 110-124, 1973.

De Smedt, I., Stavrakou, T., Hendrick, F., Danckaert, T., Vlemmix, T., Pinardi, G., Theys, N., Lerot, C., Gielen, C., Vigouroux, C., Hermans, C., Fayt, C., Veefkind, P., Müller, J.-F., and Van Roozendael, M.: Diurnal, seasonal and long-term variations of global formaldehyde columns inferred from combined OMI and GOME-2 observations, Atmos. Chem. Phys., 15, 12519–12545, https://doi.org/10.5194/acp-15-12519-2015, 2015.

De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernolle, S., Van Roozendael, M., Richter, A., Hilboll, A., Peters, E., Pedergnana, M., Loyola, D., Beirle, S., Wagner, T., Eskes, H., van Geffen, J., Boersma, K. F., and Veefkind, P.: Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project, Atmos. Meas. Tech., 11, 2395–2426, https://doi.org/10.5194/amt-11-2395-2018, 2018.

De Smedt, I., Pinardi, G., Vigouroux, C., Compernolle, S., Bais, A., Benavent, N., Boersma, F., Chan, K.-L., Donner, S., Eichmann, K.-U., Hedelt, P., Hendrick, F., Irie, H., Kumar, V., Lambert, J.-C., Langerock, B., Lerot, C., Liu, C., Loyola, D., Piters, A., Richter, A., Rivera Cárdenas, C., Romahn, F., Ryan, R. G., Sinha, V., Theys, N., Vlietinck, J., Wagner, T., Wang, T., Yu, H., and Van Roozendael, M.: Comparative assessment of TROPOMI and OMI formaldehyde observations and validation against MAX-DOAS network column measurements, Atmos. Chem. Phys., 21, 12561–12593, https://doi.org/10.5194/acp-21-12561-2021, 2021.

Duncan, B. N., Yoshida, Y., Olson, J. R., Sillman, S., Martin, R. V., Lamsal, L., Hu, Y., Pickering, K. E., Retscher, C., Allen, D. J., and Crawford, J. H.: Application of OMI observations to a space-based indicator of NO_x and VOC controls on surface ozone formation, Atmos. Environ., 44, 2213–2223, doi:10.1016/j.atmosenv.2010.03.010, 2010.

Dobber, M., Kleipool, Q., Dirksen, R., Levelt, P., Jaross, G., Taylor, S., Kelly, T., Flynn, L., Leppelmeier, G., and Rozemeijer, N.: Validation of Ozone Monitoring Instrument level 1b data products, J. Geophys. Res., 113, D15S06, https://doi.org/10.1029/2007JD008665, 2008.

European Space Agency: Sentinel-4: ESA's geostationary atmospheric mission for Copernicus operational services, ESA Rep. SP-1334, 92 pp., http://esamultimedia.esa.int/multimedia/publications/SP-1334/SP-1334.pdf, 2017.

Fasnacht, Z., Vasilkov, A., Haffner, D., Qin, W., Joiner, J., Krotkov, N., Sayer, A. M., and Spurr, R.: A geometry-dependent surface Lambertian-equivalent reflectivity product for UV–Vis retrievals – Part 2: Evaluation over open ocean, Atmos. Meas. Tech., 12, 6749–6769, https://doi.org/10.5194/amt-12-6749-2019, 2019.

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, 11403–11421, https://doi.org/10.5194/acp-17-11403-2017, 2017.

Goldberg, D. L., Anenberg, S., Mohegh, A., Lu, Z., and Streets, D. G.: TROPOMI NO_2 in the United States: A detailed look at the annual averages, weekly cycles, effects of temperature, and correlation with surface NO_2 concentrations, Earth's Future, 9, e2020EF001665, https://doi.org/10.1029/2020EF001665, 2021.

González Abad, G., Liu, X., Chance, K., Wang, H., Kurosu, T. P., and Suleiman, R.: Updated Smithsonian Astrophysical Observatory Ozone Monitoring Instrument (SAO OMI) formaldehyde retrieval, Atmos. Meas. Tech., 8, 19–32, https://doi.org/10.5194/amt-8-19-2015, 2015.

González Abad, G., Vasilkov, A., Seftor, C., Liu, X., and Chance, K.: Smithsonian Astrophysical Observatory Ozone Mapping and Profiler Suite (SAO OMPS) formaldehyde retrieval, Atmos. Meas. Tech., 9, 2797–2812, https://doi.org/10.5194/amt-9-2797-2016, 2016.

Hu, L., Keller, C. A., Long, M. S., Sherwen, T., Auer, B., Da Silva, A., Nielsen, J. E., Pawson, S., Thompson, M. A., Trayanov, A. L., Travis, K. R., Grange, S. K., Evans, M. J., and Jacob, D. J.: Global simulation of tropospheric chemistry at 12.5 km resolution: performance and evaluation of the GEOS-Chem chemical module (v10-1) within the NASA GEOS Earth system model (GEOS-5 ESM), Geosci. Model Dev., 11, 4603–4620, https://doi.org/10.5194/gmd-11-4603-2018, 2018.

Jin, X. and Holloway, T.: Spatial and temporal variability of ozone sensitivity over China observed from the Ozone Monitoring Instrument, J. Geophys. Res., 120, 7229–7246, https://doi.org/10.1002/2015JD023250, 2015.

Jin, X., Fiore, A. M., Murray, L. T., Valin, L. C., Lamsal, L. N., Duncan, B., Folkert, B., De, S., Abad, G. G., Chance, K., and Tonnesen, G. S.: Evaluating a Space-Based Indicator of Surface Ozone-NO_x-VOC Sensitivity Over Midlatitude Source Regions and Application to Decadal Trends, J. Geophys. Res.-Atmos., 122, 10439–10461, https://doi.org/10.1002/2017JD026720, 2017.

Jin, X., Fiore, A., Boersma, K. F., De Smedt, I., and Valin, L.: Inferring Changes in Summertime Surface Ozone–NO_x–VOC Chemistry over U.S. Urban Areas from Two Decades of Satellite and Ground-Based Observations, Environ. Sci. Technol., 54, 6518–6529, https://doi.org/10.1021/acs.est.9b07785, 2020.

Johnson, M. S., Liu, X., Zoogman, P., Sullivan, J., Newchurch, M. J., Kuang, S., Leblanc, T., and McGee, T.: Evaluation of potential sources of a priori ozone profiles for TEMPO tropospheric ozone retrievals, Atmos. Meas. Tech., 11, 3457–3477, https://doi.org/10.5194/amt-11-3457-2018, 2018.

Johnson, M. S., Strawbridge, K., Knowland, K. E., Keller, C., and Travis, M.: Long range transport of Siberian biomass burning emissions to North America during FIREX-AQ, Atmos. Environ., 252, 118241, https://doi.org/10.1016/j.atmosenv.2021.118241, 2021.

Joiner, J., Vasilkov, A. P., Gupta, P., Bhartia, P. K., Veefkind, P., Sneep, M., de Haan, J., Polonsky, I., and Spurr, R.: Fast simulators for satellite cloud optical centroid pressure retrievals; evaluation of OMI cloud retrievals, Atmos. Meas. Tech., 5, 529–545, https://doi.org/10.5194/amt-5-529-2012, 2012.

Judd, L. M., Al-Saadi, J. A., Szykman, J. J., Valin, L. C., Janz, S. J., Kowalewski, M. G., Eskes, H. J., Veefkind, J. P., Cede, A., Mueller, M., Gebetsberger, M., Swap, R., Pierce, R. B., Nowlan, C. R., Abad, G. G., Nehrir, A., and Williams, D.: Evaluating Sentinel-5P TROPOMI tropospheric NO_2 column densities with airborne and Pandora spectrometers near New York City and Long Island Sound, Atmos. Meas. Tech., 13, 6113–6140, https://doi.org/10.5194/amt-13-6113-2020, 2020.

Kampa, M. and Castanas, E.: Human health effects of air pollution, Environ. Pollut., 151, 362–367, https://doi.org/10.1016/j.envpol.2007.06.012, 2008.

Keller, C. A., Knowland, K. E., Duncan, B. N., Liu, J., Anderson, D. C., Das, S., Lucchesi, R. A., Lundgren, E. W., Nicely, J. M., Nielsen, J. E., Ott, L. E., Saunders, E., Strode, S. A., Wales, P. A., Jacob, D. J., and Pawson, S.: Description of the NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0, Earth and Space Science Open Archive, p. 38, https://doi.org/10.1002/essoar.10505287.1, 2020.

Kim, J., Jeong, U., Ahn, M.-H., Kim, J. H., Park, R. J., Lee, H., Song, C. H., Choi, Y.-S., Lee, K.-H., Yoo, J.-M., Jeong, M.-J., Park, S. K., Lee, K.-M., Song, C.-K., Kim, S.-W., Kim, Y. J., Kim, S.-W., Kim, M., Go, S., Liu, X., Chance, K., Chan Miller, C., Al-Saadi, J., Veihelmann, B., Bhartia, P. K., Torres, O., Abad, G. G., Haffner, D. P., Ko, D. H., Lee, S. H., Woo, J.-H., Chong, H., Park, S. S., Nicks, D., Choi, W. J., Moon, K.-J., Cho, A., Yoon, J., Kim, S.-k., Hong, H., Lee, K., Lee, H., Lee, S., Choi, M., Veefkind, P., Levelt, P. F., Edwards, D. P., Kang, M., Eo, M., Bak, J., Baek, K., Kwon, H.-A., Yang, J., Park, J., Han, K. M., Kim, B.-R., Shin, H.-W., Choi, H., Lee, E., Chong, J., Cha, Y., Koo, J.-H., Irie, H., Hayashida, S., Kasai, Y., Kanaya, Y., Liu, C., Lin, J., Crawford, J. H., Carmichael, G. R., Newchurch, M. J., Lefer, B. L., Herman, J. R., Swap, R. J., Lau, A. K. H., Kurosu, T. P., Jaross, G., Ahlers, B., Dobber, M., McElroy, C. T., and Choi, Y.: New Era of Air Quality Monitoring from Space: Geostationary Environment Monitoring Spectrometer (GEMS), B. Am. Meteorol. Soc., 101, E1–E22, https://doi.org/10.1175/bams-d-18-0013.1, 2020.

Kleipool, Q. L., Dobber, M. R., de Haan, J. F., and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, J. Geophys. Res., 113, D18308, https://doi.org/10.1029/2008jd010290, 2008.

Kleinman, L. I., Daum, P. H., Lee, Y.-N., Nunnermacker, L. J., Springston, S. R., Weinstein-Lloyd, J., and Rudolph, J.: Sensitivity of ozone production rate to ozone precursors, Geophys. Res. Lett., 28, 2903–2906, https://doi.org/10.1029/2000GL012597, 2001.

Knowland, K. E., Keller, C. A., and Lucchesi, R.: File Specification for GEOS-CF Products, GMAO Office Note No. 17 (Version 1.1), 37 pp., http://gmao.gsfc.nasa.gov/pubs/office_notes, 2020.

Knowland, K. E., Keller, C. A. Wales, P. A., Wargan, K., Coy L., Johnson, M. S., Lui, J., Lucchesi, R. A., Eastham, S. D., Fleming, E., Liang, Q., Leblanc, T., Livesey, N. J., Walker, K. A., Ott, L. E., and Pawson, S.: NASA GEOS

~~Composition Forecast Modeling System GEOS-CF v1.0: Stratospheric composition, J. Adv. Model. Earth Syst., 14, e2021MS002852. https://doi.org/10.1029/2021MS002852, 2022.~~

Kowalewski, M. G. and Janz, S. J.: Remote sensing capabilities of the GEO-CAPE airborne simulator, SPIE Conference Proceedings, San Diego, California, United States, https://doi.org/10.1117/12.2062058, 2014.

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 standard NO_2 column retrievals using in situ and surface-based NO_2 observations, Atmos. Chem. Phys., 14, 11587–11609, https://doi.org/10.5194/acp-14-11587-2014, 2014.

Lamsal, L. N., Krotkov, N. A., Vasilkov, A., Marchenko, S., Qin, W., Yang, E.-S., Fasnacht, Z., Joiner, J., Choi, S., Haffner, D., Swartz, W. H., Fisher, B., and Bucsela, E.: Ozone Monitoring Instrument (OMI) Aura nitrogen dioxide standard product version 4.0 with improved surface and cloud treatments, Atmos. Meas. Tech., 14, 455–479, https://doi.org/10.5194/amt-14-455-2021, 2021.

Laughner, J. L., Zare, A., and Cohen, R. C.: Effects of daily meteorology on the interpretation of space-based remote sensing of NO_2, Atmos. Chem. Phys., 16, 15247–15264, https://doi.org/10.5194/acp-16-15247-2016, 2016.

Laughner, J. L., Zhu, Q., and Cohen, R. C.: Evaluation of version 3.0B of the BEHR OMI NO_2 product, Atmos. Meas. Tech., 12, 129–146, https://doi.org/10.5194/amt-12-129-2019, 2019.

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.

Lelieveld, J. and Dentener, F. J.: What controls tropospheric ozone?, J. Geophys. Res., 105, 3531-3551, doi:10.1029/1999JD901011, 2000.

Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Dirksen, R. J., Malkki, A., Visser, H., de Vries, J., and Stammes, P.: The ozone monitoring instrument, IEEE Trans. Geosci. Remote Sens., 44, 1093–1101, https://doi.org/10.1109/TGRS.2006.872333, 2006.

Lorente, A., Folkert Boersma, K., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., Barkley, M., De Smedt, I., Van Roozendael, M., Wang, Y., Wagner, T., Beirle, S., Lin, J.-T., Krotkov, N., Stammes, P., Wang, P., Eskes, H. J., and Krol, M.: Structural uncertainty in air mass factor calculation for NO_2 and HCHO satellite retrievals, Atmos. Meas. Tech., 10, 759-782, doi:10.5194/amt-10-759-2017, 2017.

Loyola, D. G., Gimeno García, S., Lutz, R., Argyrouli, A., Romahn, F., Spurr, R. J. D., Pedergnana, M., Doicu, A., Molina García, V., and Schüssler, O.: The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor, Atmos. Meas. Tech., 11, 409–427, https://doi.org/10.5194/amt-11-409-2018, 2018.

Lu, C. H., and Chang, J. S.: On the indicator-based approach to assess ozone sensitivities and emissions features, J. Geophys. Res., 103, 3453–3462, doi:10.1029/97JD03128, 1998.

Liu, S., Valks, P., Pinardi, G., Xu, J., Chan, K. L., Argyrouli, A., Lutz, R., Beirle, S., Khorsandi, E., Baier, F., Huijnen, V., Bais, A., Donner, S., Dörner, S., Gratsea, M., Hendrick, F., Karagkiozidis, D., Lange, K., Piters, A. J. M., Remmers, J., Richter, A., Van Roozendael, M., Wagner, T., Wenig, M., and Loyola, D. G.: An improved

TROPOMI tropospheric NO2 research product over Europe, Atmos. Meas. Tech., 14, 7297–7327, https://doi.org/10.5194/amt-14-7297-2021, 2021.

Liu, X., Mizzi, A. P., Anderson, J. L., Fung, I., and Cohen, R. C.: The potential for geostationary remote sensing of NO_2 to improve weather prediction, Atmos. Chem. Phys., 21, 9573–9583, https://doi.org/10.5194/acp-21-9573-2021, 2021.

Martin, R. V., Chance, K., Jacob, D. J., Kurosu, T. P., Spurr, R. J. D., Bucsela, E., Gleason, J. F., Palmer, P. I., Bey, I., Fiore, A. M., Li, Q., Yantosca, R. M., and Koelemeijer, R. B. A.: An improved retrieval of tropospheric nitrogen dioxide from GOME, J. Geophys. Res.-Atmos., 107, ACH 9-1–ACH 9-21, doi:10.1029/2001JD001027, 2002.

Martin, R. V., Fiore, A. M., and Donkelaar, A. V.: Spase-based diagnosis of surface ozone sensitivity to anthropogenic emissions, Geophys. Res. Lett., 31, L06120, doi:10.1029/2004GL01941, 2004.

McLinden, C. A., Fioletov, V., Boersma, K. F., Krotkov, N., Sioris, C. E., Veefkind, J. P., and Yang, K.: Air quality over the Canadian oil sands: A first assessment using satellite observations, Geophys. Res. Lett., 39, 4, https://doi.org/10.1029/2011GL050273, 2012.

~~Naeger, A. R., Newchurch, M. J., Moore, T., Chance, K., Liu, X., Alexander, S., Murphy, K., and Wang, B. A.: Revolutionary Air Pollution Applications from Future Tropospheric Emissions: Monitoring of Pollution (TEMPO) Observations, Bull. Amer. Meteor. Soc., 102, 9, E1735–E1741, https://doi.org/10.1175/BAMS-D-21-0050.1, 2021.~~

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, 2647–2668, https://doi.org/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, https://doi.org/10.5194/amt-11-5941-2018, 2018.

Palmer, P. I., Jacob, D. J., Fiore, A. M., and Martin, R. V., Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde retrievals from the Global Ozone Monitoring Experiment, J. Geophys. Res., 106, 14539–514550, https://doi.org/10.1029/2000JD900772, 2001.

Qin, W., Fasnacht, Z., Haffner, D., Vasilkov, A., Joiner, J., Krotkov, N., Fisher, B., and Spurr, R.: A geometry-dependent surface Lambertian-equivalent reflectivity product for UV–Vis retrievals – Part 1: Evaluation over land surfaces using measurements from OMI at 466 nm, Atmos. Meas. Tech., 12, 3997–4017, https://doi.org/10.5194/amt-12-3997-2019, 2019.

Ren, J. and Xie, S.: Diagnosing ozone-NO_x-VOC sensitivity and revealing causes of ozone increases in China based on 2013–2021 satellite retrievals, Atmos. Chem. Phys. Discuss. [preprint], https://doi.org/10.5194/acp-2022-347, in review, 2022.

Schroeder, J. R., Crawford, J. H., Fried, A., Walega, J., Weinheimer, A., Wisthaler, A., Wisthaler, A., Muller, M., Mikoviny, T., Chen, G., Shook, M., Blake, D., and Tonesen, G. S.: New insights into the column CH_2O/NO_2 ratio as an indicator of near-surface ozone sensitivity, J. Geophys. Res. Atmos., 122, 8885–8907, https://doi.org/10.1002/2017JD026781, 2017.

Schenkeveld, V. M. E., Jaross, G., Marchenko, S., Haffner, D., Kleipool, Q. L., Rozemeijer, N. C., Veefkind, J. P., and Levelt, P. F.: In-flight performance of the Ozone Monitoring Instrument, Atmos. Meas. Tech., 10, 1957–1986, https://doi.org/10.5194/amt-10-1957-2017, 2017.

Sillman, S.: The relation between ozone, NO_x and hydrocarbons in urban and polluted rural environments, Atmos. Environ., 33, 1821–1845, 1999.

Silvern, R. F., Jacob, D. J., Mickley, L. J., Sulprizio, M. P., Travis, K. R., Marais, E. A., Cohen, R. C., Laughner, J. L., Choi, S., Joiner, J., and Lamsal, L. N.: Using satellite observations of tropospheric NO_2 columns to infer long-term trends in US NO_x emissions: the importance of accounting for the free tropospheric NO_2 background, Atmos. Chem. Phys., 19, 8863–8878, https://doi.org/10.5194/acp-19-8863-2019, 2019.

Souri, A. H., Choi, Y., Jeon, W., Li, X., Pan, S., Diao, L. and Westenbarger, D. A.: Constraining NO_x emissions using satellite NO_2 measurements during 2013 DISCOVER-AQ Texas campaign, Atmos. Environ., 131(2), 371–381, doi:10.1016/j.atmosenv.2016.02.020, 2016.

Souri, A. H.: Characterization of Errors in Satellite-based $HCHO/NO_2$, Tropospheric Column Ratios with Respect to Chemistry, Column to PBL Translation, Spatial Representation, and Retrieval Uncertainties, https://github.com/ahsouri/STREET, 2022.

Souri, A. H., Choi, Y., Jeon, W., Woo, J.-H., Zhang, Q., and Kurokawa, J.-i.: Remote sensing evidence of decadal changes in major tropospheric ozone precursors over East Asia, J. Geophys. Res., 122, 2474–2492, https://doi.org/10.1002/2016JD025663, 2017.

Souri, A. H., Nowlan, C. R., Wolfe, G. M., Lamsal, L. N., Chan Miller, C. E., Abad, G. G., Janz, S. J., Fried, A., Blake, D. R., Weinheimer, A. J., Diskin, G. S., Liu, X., and Chance, K.: Revisiting the effectiveness of $HCHO/NO_2$ ratios for inferring ozone sensitivity to its precursors using high resolution airborne remote sensing observations in a high ozone episode during the KORUS-AQ campaign, Atmos. Environ., 224, 117341, https://doi.org/10.1016/j.atmosenv.2020.117341, 2020.

Souri, A. H., Chance, K., Bak, J., Nowlan, C. R., González Abad, G., Jung, Y., Wong, D. C., Mao, J., and Liu, X.: Unraveling pathways of elevated ozone induced by the 2020 lockdown in Europe by an observationally constrained regional model using TROPOMI, Atmos. Chem. Phys., 21, 18227–18245, https://doi.org/10.5194/acp-21-18227-2021, 2021.

Souri, A. H., Johnson, M. S., Wolfe, G. M., Crawford, J. H., Fried, A., Wisthaler, A., Brune, W. H., Blake, D. R., Weinheimer, A. J., Verhoelst, T., Compernolle, S., Pinardi, G., Vigouroux, C., Langerock, B., Choi, S., Lamsal, L., Zhu, L., Sun, S., Cohen, R. C., Min, K.-E., Cho, C., Philip, S., Liu, X., and Chance, K.: Characterization of

Errors in Satellite-based HCHO / NO_2 Tropospheric Column Ratios with Respect to Chemistry, Column to PBL Translation, Spatial Representation, and Retrieval Uncertainties, Atmos. Chem. Phys. Discuss. [preprint], https://doi.org/10.5194/acp-2022-410, in review, 2022a.

Souri, A. H., Chance, K., Sun, K., Liu, X., and Johnson, M. S.: Dealing with spatial heterogeneity in pointwise-to-gridded- data comparisons, Atmos. Meas. Tech., 15, 41–59, https://doi.org/10.5194/amt-15-41-2022, 2022b.

~~Spurr, R.: VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media, J. Quant. Spectrosc. Ra., 102, 316–342, 2006.~~

Tack, F., Merlaud, A., Iordache, M.-D., Pinardi, G., Dimitropoulou, E., Eskes, H., Bomans, B., Veefkind, P., and Van Roozendael, M.: Assessment of the TROPOMI tropospheric NO_2 product based on airborne APEX observations, Atmos. Meas. Tech., 14, 615–646, https://doi.org/10.5194/amt-14-615-2021, 2021.

Tonnesen, G. S. and Dennis, R. L.: Analysis of radical propagation efficiency to assess ozone sensitivity to hydrocarbons and NO_x 2. Long-lived species as indicators of ozone concentration sensitivity, J. Geophys. Res., 105, 9227–9241, 2000.

U.S. Environmental Protection Agency: National Ambient Air Quality Standards for Ozone - Final Rule, Federal Register, 80, 65292-65468, https://www.gpo.gov/fdsys/pkg/FR-2015-10-26/pdf/2015-26594.pdf, 2015.

Van Dingenen, R., Dentener, F. J., Raes, F., Krol, M. C., Emberson, L., and Cofala, J.: The global impact of ozone on agricultural crop yields under current and future air quality legislation, Atmos. Environ., 43, 604–618, https://doi.org/10.1016/j.atmosenv.2008.10.033, 2009.

van Geffen, J., Boersma, K. F., Eskes, H., Sneep, M., ter Linden, M., Zara, M., and Veefkind, J. P.: S5P TROPOMI NO_2 slant column retrieval: method, stability, uncertainties and comparisons with OMI, Atmos. Meas. Tech., 13, 1315–1335, https://doi.org/10.5194/amt-13-1315-2020, 2020.

van Geffen, J., Eskes, H., Compernolle, S., Pinardi, G., Verhoelst, T., Lambert, J.-C., Sneep, M., ter Linden, M., Ludewig, A., Boersma, K. F., and Veefkind, J. P.: Sentinel-5P TROPOMI NO_2 retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data, Atmos. Meas. Tech., 15, 2037–2060, https://doi.org/10.5194/amt-15-2037-2022, 2022.

Vasilkov, A., Joiner, J., and Seftor, C.: First results from a rotational Raman scattering cloud algorithm applied to the Suomi National Polar-orbiting Partnership (NPP) Ozone Mapping and Profiler Suite (OMPS) Nadir Mapper, Atmos. Meas. Tech., 7, 2897–2906, doi:10.5194/amt-7-2897-2014, 2014.

Vasilkov, A., Qin, W., Krotkov, N., Lamsal, L., Spurr, R., Haffner, D., Joiner, J., Yang, E.-S., and Marchenko, S.: Accounting for the effects of surface BRDF on satellite cloud and trace-gas retrievals: a new approach based on geometry-dependent Lambertian equivalent reflectivity applied to OMI algorithms, Atmos. Meas. Tech., 10, 333–349, https://doi.org/10.5194/amt-10-333-2017, 2017.

Vasilkov, A., Yang, E.-S., Marchenko, S., Qin, W., Lamsal, L., Joiner, J., Krotkov, N., Haffner, D., Bhartia, P. K., and Spurr, R.: A cloud algorithm based on the O_2-O_2 477 nm absorption band featuring an advanced spectral fitting method and the use of surface geometry-dependent Lambertian-equivalent reflectivity, Atmos. Meas. Tech., 11, 4093–4107, https://doi.org/10.5194/amt-11-4093-2018, 2018.

Veefkind, J. P., de Haan, J. F., Sneep, M., and Levelt, P. F.: Improvements to the OMI O2–O2 operational cloud algorithm and comparisons with ground-based radar–lidar observations, Atmos. Meas. Tech., 9, 6035–6049, https://doi.org/10.5194/amt-9-6035-2016, 2016.

Verhoelst, T., Compernolle, S., Pinardi, G., Lambert, J.-C., Eskes, H. J., Eichmann, K.-U., Fjæraa, A. M., Granville, J., Niemeijer, S., Cede, A., Tiefengraber, M., Hendrick, F., Pazmiño, A., Bais, A., Bazureau, A., Boersma, K. F., Bognar, K., Dehn, A., Donner, S., Elokhov, A., Gebetsberger, M., Goutail, F., Grutter de la Mora, M., Gruzdev, A., Gratsea, M., Hansen, G. H., Irie, H., Jepsen, N., Kanaya, Y., Karagkiozidis, D., Kivi, R., Kreher, K., Levelt, P. F., Liu, C., Müller, M., Navarro Comas, M., Piters, A. J. M., Pommereau, J.-P., Portafaix, T., Prados-Roman, C., Puentedura, O., Querel, R., Remmers, J., Richter, A., Rimmer, J., Rivera Cárdenas, C., Saavedra de Miguel, L., Sinyakov, V. P., Stremme, W., Strong, K., Van Roozendael, M., Veefkind, J. P., Wagner, T., Wittrock, F., Yela González, M., and Zehner, C.: Ground-based validation of the Copernicus Sentinel-5P TROPOMI NO_2 measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks, Atmos. Meas. Tech., 14, 481–510, https://doi.org/10.5194/amt-14-481-2021, 2021.

Vigouroux, C., Langerock, B., Bauer Aquino, C. A., Blumenstock, T., Cheng, Z., De Mazière, M., De Smedt, I., Grutter, M., Hannigan, J. W., Jones, N., Kivi, R., Loyola, D., Lutsch, E., Mahieu, E., Makarova, M., Metzger, J.-M., Morino, I., Murata, I., Nagahama, T., Notholt, J., Ortega, I., Palm, M., Pinardi, G., Röhling, A., Smale, D., Stremme, W., Strong, K., Sussmann, R., Té, Y., van Roozendael, M., Wang, P., and Winkler, H.: TROPOMI–Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations, Atmos. Meas. Tech., 13, 3751–3767, https://doi.org/10.5194/amt-13-3751-2020, 2020.

Wu, S., Lee, H. J., Anderson, A., Liu, S., Kuwayama, T., Seinfeld, J. H., and Kleeman, M. J.: Direct measurements of ozone response to emissions perturbations in California, Atmos. Chem. Phys., 22, 4929–4949, https://doi.org/10.5194/acp-22-4929-2022, 2022.

Zara, M., Boersma, K. F., De Smedt, I., Richter, A., Peters, E., van Geffen, J. H. G. M., Beirle, S., Wagner, T., Van Roozendael, M., Marchenko, S., Lamsal, L. N., and Eskes, H. J.: Improved slant column density retrieval of nitrogen dioxide and formaldehyde for OMI and GOME-2A from QA4ECV: intercomparison, uncertainty characterisation, and trends, Atmos. Meas. Tech., 11, 4033–4058, https://doi.org/10.5194/amt-11-4033-2018, 2018.

Zhao, X., Griffin, D., Fioletov, V., McLinden, C., Cede, A., Tiefengraber, M., Müller, M., Bognar, K., Strong, K., Boersma, F., Eskes, H., Davies, J., Ogyu, A., and Lee, S. C.: Assessment of the quality of TROPOMI high-spatial-resolution NO_2 data products in the Greater Toronto Area, Atmos. Meas. Tech., 13, 2131–2159, https://doi.org/10.5194/amt-13-2131-2020, 2020.

Zhu, L., González Abad, G., Nowlan, C. R., Chan Miller, C., Chance, K., Apel, E. C., DiGangi, J. P., Fried, A., Hanisco, T. F., Hornbrook, R. S., Hu, L., Kaiser, J., Keutsch, F. N., Permar, W., St. Clair, J. M., and Wolfe, G. M.: Validation of satellite formaldehyde (HCHO) retrievals using observations from 12 aircraft campaigns, Atmos. Chem. Phys., 20, 12329–12345, https://doi.org/10.5194/acp-20-12329-2020, 2020.

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), J. Quant. Spectrosc. Ra., 17–39, https://doi.org/10.1016/j.jqsrt.2016.05.008, 2017.