



1     **Assessing sub-grid variability within satellite pixels using airborne mapping**  
2                                   **spectrometer measurements**

3

4     **Wenfu Tang<sup>1,2</sup>, David P. Edwards<sup>2</sup>, Louisa K. Emmons<sup>2</sup>, Helen M. Worden<sup>2</sup>, Laura M.**  
5     **Judd<sup>3</sup>, Lok N. Lamsal<sup>4,5</sup>, Jassim A. Al-Saadi<sup>3</sup>, Scott J. Janz<sup>4</sup>, James H. Crawford<sup>3</sup>, Merritt**  
6     **N. Deeter<sup>2</sup>, Gabriele Pfister<sup>2</sup>, Rebecca R. Buchholz<sup>2</sup>, Benjamin Gaubert<sup>2</sup>, Caroline R.**  
7     **Nowlan<sup>6</sup>**

8     <sup>1</sup>Advanced Study Program, National Center for Atmospheric Research, Boulder, CO, USA

9     <sup>2</sup>Atmospheric Chemistry Observations and Modeling, National Center for Atmospheric Research,  
10     Boulder, CO, USA

11     <sup>3</sup>NASA Langley Research Center, Hampton, VA 23681, USA

12     <sup>4</sup>NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

13     <sup>5</sup>Universities Space Research Association, Columbia, MD 21046, USA

14     <sup>6</sup>Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA

15

16     **Abstract**

17             Sub-grid variability (SGV) of atmospheric trace gases within satellite pixels is a key issue  
18     in satellite design, and interpretation and validation of retrieval products. However, characterizing  
19     this variability is challenging due to the lack of independent high-resolution measurements. Here  
20     we use tropospheric NO<sub>2</sub> vertical column (VC) measurements from the Geostationary Trace gas  
21     and Aerosol Sensor Optimization (GeoTASO) airborne instrument with a spatial resolution of  
22     about 250 m × 250 m to quantify the normalized SGV (i.e., the standard deviation of the sub-grid  
23     GeoTASO values within the sampled satellite pixel divided by their mean of the sub-grid  
24     GeoTASO values within the sampled satellite pixel) for different satellite pixel sizes. We use the  
25     GeoTASO measurements over the Seoul Metropolitan Area (SMA) and Busan region of South  
26     Korea during the 2016 KORUS-AQ field campaign, and over the Los Angeles Basin, USA during  
27     the 2017 SARP field campaign. We find that the normalized SGV of NO<sub>2</sub> VC increases with  
28     increasing satellite pixel sizes (from ~10% for 0.5 km × 0.5 km pixel size to ~35% for 25 km × 25  
29     km pixel size), and this relationship holds for the three study regions, which are also within the  
30     domains of upcoming geostationary satellite air quality missions. We also quantify the temporal  
31     variability of the retrieved NO<sub>2</sub> VC within the same satellite pixels (represented by the difference  
32     of retrieved values at two different times of a day). For a given satellite pixel size, the temporal  
33     variability within the same satellite pixels increases with the sampling time difference over SMA.  
34     For a given small (e.g., ≤4 hours) sampling time difference within the same satellite pixels, the  
35     temporal variability of the retrieved NO<sub>2</sub> VC increases with the increasing spatial resolution over  
36     the SMA, Busan region, and the Los Angeles basin.

37             The results of this study have implications for future satellite design and retrieval  
38     interpretation, and validation when comparing pixel data with local observations. In addition, the



39 analyses presented in this study are equally applicable in model evaluation when comparing model  
40 grid values to local observations. Results from the Weather Research and Forecasting model  
41 coupled with Chemistry (WRF-Chem) model indicate that the normalized satellite SGV of  
42 tropospheric NO<sub>2</sub> VC calculated in this study could serve as an upper bound to the satellite SGV  
43 of other species (e.g., CO and SO<sub>2</sub>) that share common source(s) with NO<sub>2</sub> but have relatively  
44 longer lifetime.  
45

## 46 1. Introduction

47 Characterizing sub-grid variability (SGV) of atmospheric chemical constituent fields is  
48 important in both satellite retrievals and atmospheric chemical-transport modeling. The inability  
49 to resolve sub-grid details is one of the fundamental limitations of grid-based models (Qian et al.,  
50 2010) and has been studied extensively (e.g., Boersma et al., 2016; Ching et al., 2006; Denby et  
51 al., 2011; Pillai et al., 2010; Qian et al., 2010). Pillai et al. (2010) found that the SGV of column-  
52 averaged carbon dioxide (CO<sub>2</sub>) can reach up to 1.2 ppm in global models that have a horizontal  
53 resolution of 100 km. This is an order of magnitude larger than sampling errors that include both  
54 limitations in instrument precision and uncertainty of unresolved atmospheric CO<sub>2</sub> variability  
55 within the mixed layer (Gerbig et al., 2003). Denby et al. (2011) suggested that the average  
56 European urban background exposure for nitrogen dioxide (NO<sub>2</sub>) using a model of 50-km  
57 resolution is underestimated by ~44% due to SGV.

58 In contrast, much less attention has been paid to the sub-grid variability within satellite  
59 pixels (e.g., Broccardo et al., 2018; Judd et al., 2019; Tack et al., 2020). Indeed, some previous  
60 studies (e.g., Kim et al., 2016; Song et al., 2018; Zhang et al., 2019; Choi et al., 2020) used satellite  
61 retrievals to study SGV in models, and calculated representativeness errors of model results with  
62 respect to satellite measurements (e.g., Pillai et al., 2010). Even though satellite retrievals of  
63 atmospheric composition often have smaller uncertainties than model results, it has not been until  
64 recently that the typical spatial resolution of atmospheric composition satellite products has  
65 reached scales comparable to regional atmospheric chemistry models (< ~10 km).

66 Until recently, accurate in-situ measurements with sufficient spatiotemporal coverage have  
67 not been available. As a result, it has been challenging to quantify satellite SGV, even though this  
68 is a key issue in designing, understanding and correctly interpreting satellite observations. This is  
69 especially important in the satellite instrument develop process, during which the required  
70 measurement precision and retrieval resolution need to be defined in order to meet the science  
71 goals. In addition, when validating and evaluating relatively coarse-scale satellite retrievals by  
72 comparing with in situ observations, SGV introduces large uncertainties. This work is partly  
73 motivated by validation requirements and considerations for the upcoming geostationary orbit  
74 (GEO) satellite constellation for atmospheric composition that includes the Tropospheric  
75 Emissions: Monitoring Pollution (TEMPO) mission over North America (Chance et al., 2013;  
76 Zogman et al., 2017), the Geostationary Environment Monitoring Spectrometer (GEMS) over  
77 Asia (Kim et al., 2020), and the Sentinel-4 mission over Europe (Courrèges-Lacoste et al., 2017).

78 The measurements of the Geostationary Trace gas and Aerosol Sensor Optimization  
79 (GeoTASO) airborne instrument provide a unique dataset for quantifying satellite SGV.  
80 GeoTASO is an airborne remote sensing instrument capable of high spatial resolution retrieval of



81 UV-VIS absorbing species like NO<sub>2</sub>, formaldehyde (HCHO; Nowlan et al., 2018) and sulfur  
82 dioxide (SO<sub>2</sub>; Chong et al., 2020), and with measurement characteristics similar to the GEMS and  
83 TEMPO GEO satellite instruments. The GeoTASO data used here were taken in gapless, grid-like  
84 patterns – or “rasters” – over the regions of interest, providing essentially continuous spatial  
85 coverage that was repeated up to four times a day in some cases. As such, the GeoTASO data  
86 provide a preview of the type of sampling that is expected from the GEO satellite sensors, making  
87 the data particularly suitable for our study. We focus on the GeoTASO measurements made during  
88 the Korea United States Air Quality (KORUS-AQ) field experiment in 2016. The measurements  
89 from KORUS-AQ have been widely used by researchers for various air quality topics, including  
90 quantification of emissions and model and satellite evaluation (e.g., Deeter et al., 2019; Huang et  
91 al., 2018; Kim et al., 2018; Miyazaki et al., 2019; Spinei et al., 2018; Tang et al., 2018, 2019; Souri  
92 et al., 2020, Gaubert et al., 2020). We further compare our findings from KORUS-AQ with flights  
93 conducted during the NASA Student Airborne Research Program (SARP) in 2017 over the Los  
94 Angeles (LA) Basin to test the general applicability of our findings. The KORUS-AQ mission took  
95 place within the GEMS domain, while the SARP in 2017 is within the domain of TEMPO. Given  
96 the similarity between the TEMPO and GEMS instruments in terms of spectral ranges, spectral  
97 and spatial resolution, and retrieval algorithms (Al-Saadi et al., 2014), such comparison is  
98 reasonable and useful in facilitating the generalization of the results from the study.

99 We use the tropospheric NO<sub>2</sub> vertical column (VC) retrieved by GeoTASO as a tool to  
100 assess satellite SGV. NO<sub>2</sub> is an important air pollutant that is primarily generated from  
101 anthropogenic sources such as emissions from the energy, transportation, and industry sectors  
102 (Hoesly et al., 2018). NO<sub>2</sub> is a reactive gas with a typical lifetime of a few hours in the planetary  
103 boundary layer (PBL), although it can also be transported over long distance in the form of  
104 peroxyacetyl nitrate (PAN) and nitric acid. NO<sub>2</sub> is a precursor of tropospheric ozone and secondary  
105 aerosols, and has a negative impact on human health and the environment (Finlayson-Pitts et al.,  
106 1997). The results from this paper’s analysis of NO<sub>2</sub> also have implications for other air pollutants  
107 that share common source(s) with NO<sub>2</sub>, but that have somewhat longer lifetimes, for example,  
108 carbon monoxide (CO) and SO<sub>2</sub>.

109 In this study, we apply a satellite pixel random sampling technique and the spatial structure  
110 function analysis to GeoTASO data (described in Section 2) to quantify the SGV of satellite pixel  
111 NO<sub>2</sub> VC at a variety of spatial resolutions. We analyze the relationship between satellite pixel size  
112 and satellite SGV, and the relationship between satellite pixel size and the temporal variability of  
113 NO<sub>2</sub> observations (Section 3). We then discuss the implications for satellite design, satellite  
114 retrieval interpretation, satellite validation and evaluation, and satellite–in situ data comparisons  
115 (Section 4). Implications for general local observations and grid data comparisons are also  
116 discussed. Section 5 presents our conclusions.

## 117 **2. Data and methods**

118 In this section, we describe the GeoTASO instrument, campaign flights and the different  
119 analysis techniques used to characterize the satellite pixel SGV. We outline two approaches:  
120 satellite pixel random sampling to investigate separately both spatial variability and temporal  
121 variability, and the construction of spatial structure functions for an alternative measure of spatial  
122 variability.



## 123 **2.1 GeoTASO instrument**

124 In this study, we focus on GeoTASO retrievals of tropospheric NO<sub>2</sub> Vertical Column (VC).  
125 GeoTASO is a hyperspectral instrument (Leitch et al., 2014) that measures nadir backscattered  
126 light in the ultraviolet (UV; 290–400 nm) and visible (VIS; 415–695 nm). As one of NASA’s  
127 airborne UV–VIS mapping instruments, it was designed to support the upcoming GEO satellite  
128 missions by acquiring high temporal and spatial resolution measurements with dense sampling for  
129 optimizing and experimenting with new retrieval algorithms (Leitch et al., 2014; Nowlan et al.,  
130 2016; Lamsal et al., 2017; Judd et al., 2019). GeoTASO has a cross-track field of view of 45° (+/  
131 22.5° from nadir), and the retrieval pixel size at nadir is approximately 250 m×250 m from typical  
132 flight altitudes of 24,000–28,000 feet (7.3–8.5 km). The dense sampling of the GeoTASO datasets  
133 is a unique feature and provides the opportunity to study the expected spatial and temporal  
134 variability within the satellite NO<sub>2</sub> retrieval pixels at high resolution. The GeoTASO data used in  
135 this study are mostly cloud-free. Validation of GeoTASO NO<sub>2</sub> retrievals during KORUS-AQ with  
136 Pandora shows ~10% difference on average. The uncertainty estimate is lower than that reported  
137 by Nowlan et al. [2016].

## 138 **2.2 The 2016 KORUS-AQ field campaign**

139 The KORUS-AQ field measurement campaign (Al-Saadi et al., 2014), took place in May–  
140 June 2016, to help understand the factors controlling air quality over South Korea. One of the goals  
141 of KORUS-AQ was the testing and improvement of remote sensing algorithms in advance of the  
142 launches of GEMS, TEMPO, and Sentinel-4 satellite missions. It is hoped that the high-quality  
143 initial data products from the GEO missions will facilitate their rapid uptake in air quality  
144 applications after launch (Al-Saadi et al., 2014; Kim et al., 2020). During KORUS-AQ, GeoTASO  
145 flew onboard the NASA LaRC B200 aircraft. We focus on the data taken over the Seoul  
146 Metropolitan Area (SMA) that is highly urbanized and polluted, and the greater Busan region, that  
147 is somewhat less urbanized and less polluted (Figure 1). Figure 2 shows the 12 GeoTASO data  
148 rasters (i.e., gapless maps) acquired over SMA. Figure S1 shows the 2 GeoTASO rasters acquired  
149 over the Busan region.

## 150 **2.3 The 2017 SARP field campaign**

151 During the NASA Student Airborne Research Program (SARP) flights in June 2017,  
152 ([https://airbornescience.nasa.gov/content/Student\\_Airborne\\_Research\\_Program](https://airbornescience.nasa.gov/content/Student_Airborne_Research_Program)), GeoTASO was  
153 flown onboard the NASA LaRC UC-12B aircraft over the LA Basin (Figure S2, which also shows  
154 the landcover). A detailed description and analysis of these data can be found in Judd et al. (2018;  
155 2019). In this study, we compare our analyses and findings from KORUS-AQ with those using the  
156 GeoTASO data over the LA Basin to test the general applicability of our findings.

## 157 **2.4 Satellite pixel random sampling for spatial variability**

158 GeoTASO provides continuous measurements in a gapless map pattern at high spatial  
159 resolution (Figures 2, S1, and S2). This dataset allows us to sample and study the SGV of coarser  
160 spatial resolution hypothetical satellite pixels sampling the same domain. To mimic satellite  
161 observations and quantify the satellite SGV, we randomly sample the GeoTASO data with  
162 hypothetical satellite pixels spanning 27 different pixel sizes (0.5 km×0.5 km, 0.75 km×0.75 km,



163 1 km×1 km, 2 km×2 km, up to 25 km×25 km). Because of the move to smaller pixel sizes in the  
164 future satellite missions, and the limitation in the maximum hypothetical satellite pixel size  
165 sampled using the random sampling method, the analysis of SGV only goes up to 25 km × 25 km.  
166 This sampling process is conducted for each hour of each selected flight over the regions of interest  
167 during the KORUS-AQ and SARP campaigns. For every sampled satellite pixel, the mean  
168 ( $MEAN_{pixel}$ ) and standard deviation ( $SD_{pixel}$ ) of the GeoTASO tropospheric NO<sub>2</sub> VC data within  
169 the pixel are calculated to represent the satellite SGV. Normalized satellite SGV is calculated by  
170 the standard deviation of the GeoTASO data within the sampled satellite pixel divided by the mean  
171 of the GeoTASO data within the sampled satellite pixel ( $SD_{pixel}/MEAN_{pixel}$ ).

172 We use a set of 10,000 hypothetical satellite pixels at each size to include all of the  
173 GeoTASO data in the analysis and to cover as many locations as possible. Our sensitivity test  
174 indicates that the results do not change by halving the sample size. Because the data are located  
175 closely in space but may be sampled at slightly different times for the same flight, we separate  
176 GeoTASO data into hourly bins for each flight before pixel sampling in order to reduce the impact  
177 of temporal variability of the GeoTASO data within a single satellite pixel sample.

178 As an illustration, we describe the procedure below for the May 17<sup>th</sup> afternoon flight  
179 (Figure 3) that was conducted from 13:00 to 17:00 local time: (1) the GeoTASO data during this  
180 flight were divided into four hourly groups according to the measurement time, i.e., 13:00-14:00,  
181 14:00-15:00, 15:00-16:00, and 16:00-17:00; (2) for each of the 27 hypothetical satellite pixel sizes,  
182 we randomly generate 10,000 satellite pixel locations within each hourly group. Therefore, for  
183 each hour, we sample 270,000 satellite pixels (27 different satellite pixel sizes and 10,000 samples  
184 for each size), and for this example flight, we have a total of up to 1,080,000 possible satellite  
185 pixels in each of 4 hourly groups. Note that the actual samples used in the analysis are less than  
186 1,080,000 because we discarded a sampled satellite pixel if it is not covered by GeoTASO data for  
187 at least 75% of its area.

188 We tested other choices of the coverage threshold in addition to 75% over SMA (not shown  
189 here). The results are similar for small pixels ( $< \sim 10 \text{ km}^2$ ), as they are more likely to be covered  
190 by GeoTASO data regardless of the threshold value. For larger pixels ( $> \sim 15 \text{ km}^2$ ), the satellite  
191 SGV is slightly lower when using 30% or 50% as the area coverage threshold, because larger  
192 pixels act like smaller pixels when only partially covered. The threshold of 75% was chosen as a  
193 trade-off between sample size and representation.

## 194 **2.5 Satellite pixel random sampling for temporal variability**

195 We also quantify the temporal variability of the retrieved NO<sub>2</sub> VC within the same satellite  
196 pixels for different satellite pixel sizes. To calculate temporal variability within a hypothetical  
197 satellite pixel, we need GeoTASO data to cover the hypothetical satellite pixel at different times  
198 during the day. During the KORUS-AQ and 2017 SARP campaigns, rasters were treated as single  
199 units (Judd et al., 2019). Each raster produces a contiguous map of data that we consider as roughly  
200 representative of the mid-time of the raster. Unlike the calculation of SGV, which is based on data  
201 separated into hourly bins (section 2.4) to reduce the impact of temporal variability in the  
202 calculated spatial variability, the satellite pixel random sampling to assess temporal variability is  
203 based on rasters, and only conducted for days with multiple rasters. This is to ensure that the



204 sampled hypothetical satellite pixels have multiple values at different times of the day, and hence  
205 maximize the sample size.

206 To assess temporal variability within the hypothetical satellite pixels, we randomly select  
207 50,000 pixel locations for each of the 27 hypothetical satellite pixel sizes, and use this same set of  
208 pixel locations to sample the GeoTASO data for each raster across all flights for a given day. This  
209 process is repeated for all days with multiple rasters, and the 75% of area coverage threshold is  
210 also applied. When there are two or more raster values of  $MEAN_{pixel}$  for a given pixel location  
211 separated by time  $Dt$ , the temporal mean difference (TeMD) within the satellite pixel is calculated  
212 as:

$$213 \quad TeMD(Dt) = average( |MEAN_{pixel}(t) - MEAN_{pixel}(t + Dt)| ) \quad (1)$$

214 This procedure is repeated for each satellite pixel size.

## 215 **2.6 Spatial structure function**

216 Structure functions have been applied to in situ measurements and model-generated  
217 tropospheric trace gases to analyze their spatial and temporal variability in previous studies (Harris  
218 et al., 2001). The Spatial Structure Function (SSF) (Fishman et al., 2011; Follette-Cook et al., 2015)  
219 is an alternative measure to the satellite pixel random sampling described above for quantifying  
220 spatial variability, and in this work, we apply the SSF to GeoTASO data to assist our analysis of  
221 satellite SGV. The main difference between the two measures is that the SSF is based on individual  
222 GeoTASO data points, while the results from satellite pixel random sampling are based on sampled  
223 satellite pixels. The SSF is defined here follows Follette-Cook et al. (2015):

$$224 \quad f(NO_{2,VC}, D) = average( |NO_{2,VC}(x + D) - NO_{2,VC}(x)| ) \quad (2)$$

225 where  $NO_{2,VC}$  is tropospheric  $NO_2$  VC.  $f(NO_{2,VC}, Distance)$  calculates the average of the  
226 absolute value of  $NO_{2,VC}$  differences across all data pairs (measured in the same hourly bin) that  
227 are separated by a distance  $D$ . To calculate SSF, the first step is the same as the first step of the  
228 satellite pixel random sampling: we group GeoTASO data hourly for each flight to reduce the  
229 impact of temporal variability of the GeoTASO data, and we only pair each GeoTASO data point  
230 with all the other GeoTASO data in the same hourly bin. More details on structure functions can  
231 be found in Follette-Cook et al. (2015).

## 232 **2.7 WRF-Chem simulation**

233 To briefly demonstrate the application of this technique on model evaluation and other  
234 species, we show results of a WRF-Chem simulation (Weather Research and Forecasting model  
235 coupled to Chemistry) with a resolution of  $3 \text{ km} \times 3 \text{ km}$  over SMA in the Discussion section. The  
236 simulation used NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids  
237 as initial and boundary conditions, and the model meteorological fields above the PBL were  
238 nudged 6-hourly. KORUS version 3 anthropogenic emissions and FINN version 1.5 fire emissions  
239 (Wiedinmyer et al., 2011) were used.

240



## 241 3. Results

242 In this section, we discuss the results for SGV over the different regions considered. Results  
243 are presented for the hypothetical satellite pixel random sampling for spatial variability and  
244 temporal variability, and for the spatial structure function analysis for spatial variability.

### 245 3.1 Sub-grid variability (SGV) within satellite pixels

246 SMA, the Busan region, and the LA Basin have different levels of pollution – the average  
247 values of the GeoTASO NO<sub>2</sub> VC data over the SMA, the Busan region, and the LA Basin are  
248  $2.3 \times 10^{16}$  molecules cm<sup>-2</sup>,  $1.1 \times 10^{16}$  molecules cm<sup>-2</sup>, and  $1.3 \times 10^{16}$  molecules cm<sup>-2</sup>, respectively.  
249 Over the three regions, the mean values (MEAN<sub>pixel</sub>) and absolute values of standard deviation  
250 (SD<sub>pixel</sub>) of the hypothetical satellite pixels sampled over GeoTASO NO<sub>2</sub> VC data are different  
251 (Figure S3). This is consistent with previous studies suggesting absolute values of SGV can vary  
252 regionally (Judd et al., 2019; Broccardo et al., 2018). However, we find that the normalized  
253 satellite SGV (calculated as the ratio of SD<sub>pixel</sub> to MEAN<sub>pixel</sub> for a sampled pixel) is similar over  
254 each of the areas, regardless of the absolute level of pollution as represented by MEAN<sub>pixel</sub> (Figure  
255 4). Over SMA (Figure 4a), the mean normalized satellite SGV of tropospheric NO<sub>2</sub> VC increases  
256 smoothly from ~10% for the pixel size of 0.5 km × 0.5 km, to ~35% for the pixel size of 25 km ×  
257 25 km. The interquartile variation of the satellite SGV also increases with satellite pixel sizes. The  
258 patterns of the sampled satellite pixels over the Busan region (Figure 4b) and LA Basin (Figure  
259 4c) are also found to be similar to those over SMA. Furthermore, Figures S4 and S5 show that  
260 even the individual flights over the three domains generally follow the same pattern, except in the  
261 case of the June 9 PM flight that is discussed below.

262 We also compare normalized satellite SGV for different levels of pollution, regardless of  
263 their regions (Figure S6). The normalized satellite SGV for the less polluted pixels (MEAN<sub>pixel</sub>  
264 being lower than the average value of all pixels, i.e.,  $2 \times 10^{16}$  molecules cm<sup>-2</sup>) also shows an overall  
265 similar pattern as for the more polluted pixels (MEAN<sub>pixel</sub> being higher than the average value of  
266 all pixels). We notice that at small pixel sizes, less polluted pixels have higher normalized satellite  
267 SGV, possibly contributed by relatively higher retrieval noise at lower pollution levels.

268 In addition to the comparison between different domains and pollution levels, we also  
269 compare this relationship in the morning and afternoon. The variation of normalized SGV and  
270 pixel size in the morning and afternoon are generally similar for the three regions (Figure S7),  
271 except for the large size pixels over SMA, where the normalized SGV is larger in the afternoon  
272 than in the morning. This difference is driven by the GeoTASO data from June 9 PM (Figure S4),  
273 as the normalized SGV pattern for the afternoon agrees well with the normalized SGV pattern for  
274 the morning when the June 9 PM data are excluded. Figure S1 shows that the June 9 PM NO<sub>2</sub>  
275 pollution level is higher than other days under meteorological conditions of light winds and  
276 moderate temperatures. The MEAN<sub>pixel</sub> values increases ~60% going from 1 km × 1 km to 25 km  
277 × 25 km pixel size, while SD<sub>pixel</sub> dramatically increases ~7 times from 1 km × 1 km to 25 km × 25  
278 km. This is higher than any other day, and results in the highest SGV encountered over SMA at  
279 the large pixel sizes. We also notice that the normalized SGV does not generally change  
280 significantly in the range of 20 km × 20 km to 25 km × 25 km. However, in the case of SMA for  
281 June 9 PM, the normalized SGV (as well as SD<sub>pixel</sub>) increases significantly and monotonously with  
282 pixel size in the range of 20 km × 20 km to 25 km × 25 km.



283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

We show the normalized SGV for individual rasters over SMA (Figure 5) to indicate the uncertainty range of the normalized SGV shown in Figure 4. The spread of SGV across different individual rasters represents the uncertainties of using the averaged normalized SGV for a specific case. Note that the variation of normalized SGV with pixel size for individual rasters generally follows the same pattern (i.e., increases with satellite pixel size), especially when the pixel size is small ( $\leq 10 \text{ km} \times 10 \text{ km}$ ). The normalized SGV increases from  $\sim 10\%$  to  $\sim 25\%$ , with the uncertainty range consistently being  $\pm 5\%$  when the pixel size is smaller than  $10 \text{ km} \times 10 \text{ km}$ . When the pixel size is larger than  $10 \text{ km} \times 10 \text{ km}$ , the uncertainty range broadens with pixel sizes from  $\pm 5\%$  ( $10 \text{ km} \times 10 \text{ km}$ ) to  $\pm 15\%$  ( $25 \text{ km} \times 25 \text{ km}$ ). This means that when the satellite pixel size is large, using the mean normalized SGV in Figure 4 to represent specific cases may lead to larger uncertainties. Therefore, our analysis reveals a threshold for spatial resolution at about  $10 \text{ km} \times 10 \text{ km}$ . Below this resolution, SGV can be characterized by the mean value with relatively smaller uncertainty ( $\pm 5\%$ ) and hence high confidence, even with large diurnal or day-to-day variations. The spatial resolutions of TEMPO, GEMS, and TROPOMI (TROPOspheric Monitoring Instrument, Veeffkind et al., 2012; Griffin et al., 2019; van Geffen et al., 2019) are within this  $\leq 10 \text{ km} \times 10 \text{ km}$  range, while the resolution of OMI (Ozone Monitoring Instrument, Levelt et al., 2006; 2018) is not. This means that applying this study (e.g., Figure 4) to OMI for a specific case study (e.g., a specific day) requires extra caution.

302

303

304

305

306

307

308

We tested the sensitivity of the results over SMA to sampling GeoTASO data with hypothetical satellite pixels grouped by complete flight, rather than grouping the data by time in hourly bins. The resulting patterns and relationships are similar, except that the normalized satellite SGV increases  $\sim 5\%$  for pixels of small sizes due to the inclusion of temporal variability (Figure S8a). We also tested the results for sampling satellite pixels by raster instead of within hourly bins. The results are again similar to Figure 4, except that the normalized satellite SGV increases  $\sim 1\%$  for pixels of small sizes due to the inclusion of temporal variability (Figure S8b).

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

The three regions investigated in this work have different levels of urbanization and air pollution (Figures 1 and S2). PBL conditions are also different in the morning and afternoon (Figure S9). The similarity of the relationships between the satellite pixel size and the normalized satellite SGV over these different regions (Figure 4) suggests that this relationship may be generalizable to  $\text{NO}_2$  VC over regions with different levels of urbanization and air pollution, and different PBL conditions. Moreover, Figures 4 and 5 point to the possibility of developing a generalized look-up table for the expected normalized satellite SGV for  $\text{NO}_2$  VC at different satellite pixel sizes, especially for small pixel sizes (e.g., TEMPO, GEMS, and TROPOMI). This would be useful in satellite design, satellite retrieval evaluation and interpretation, and satellite-in situ data comparisons. For example, the satellite pixel size of tropospheric  $\text{NO}_2$  VC retrievals from GEMS, TEMPO, TROPOMI, and OMI are highlighted in Figure 4. Following Judd et al. (2019), we choose  $3 \text{ km} \times 3 \text{ km}$ ,  $5 \text{ km} \times 5 \text{ km}$ ,  $7 \text{ km} \times 8 \text{ km}$ , and  $18 \text{ km} \times 18 \text{ km}$  pixels to represent the expected area of the satellite pixels for TEMPO ( $2.1 \text{ km} \times 4.4 \text{ km}$ ), TROPOMI ( $3.5 \text{ km} \times 7 \text{ km}$ ), GEMS ( $7 \text{ km} \times 8 \text{ km}$ ), and OMI ( $18 \text{ km} \times 18 \text{ km}$ ), respectively. The expected normalized satellite SGV for TEMPO, TROPOMI, GEMS, and OMI are 15–20%,  $\sim 20\%$ , 20–25%, and  $\sim 30\%$ , respectively. Taking the TEMPO example, this implies that the satellite SGV could potentially lead to uncertainties of 15–20% in a validation exercise comparing a satellite retrieval with sub-satellite local ground measurements of  $\text{NO}_2$  VC as might be obtained from a Pandora spectrometer.





327 As a result, we should caution that calculating a pixel mean bias when evaluating against local  
328 measurements within the pixel sometimes may be optimistic due to the cancellation of sub-grid  
329 positive and negative biases.

### 330 **3.2 Temporal variability (TeMD) within the same satellite pixels**

331 In addition to satellite spatial SGV, we also analyze the temporal variability (i.e., TeMD)  
332 within the same hypothetical satellite pixels. Figure 6 shows TeMD of satellite retrieved  
333 tropospheric NO<sub>2</sub> VC over SMA as a function of hypothetical satellite pixel size and the separation  
334 time Dt between flight rasters as described in section 2.5. The results for 27 satellite pixel sizes  
335 analyzed are shown by different colors, while results for selected satellite pixel sizes are  
336 highlighted by thicker lines. For all the pixel sizes, TeMD increases monotonically with the time  
337 difference Dt between two sampled raster values within the same pixel. The TeMD of tropospheric  
338 NO<sub>2</sub> VC is around  $0.75 \times 10^{16}$  molecules cm<sup>-2</sup> for a Dt of 2 hours over SMA for all the sampled  
339 satellite pixel sizes, and increases to  $\sim 2 \times 10^{16}$  molecules cm<sup>-2</sup> for Dt of 8 hours. This indicates that,  
340 along with improvements in the satellite retrieval spatial resolution with smaller pixels, improving  
341 the satellite retrieval temporal resolution with higher frequency measurements is also an effective  
342 way to enhance capability in resolving variabilities of NO<sub>2</sub>. This is expected because of NO<sub>2</sub>'s  
343 relatively short lifetime ( $\sim$  a few hours) and strong diurnal cycle due to emission activities,  
344 chemistry and photolysis rate (Fishman et al., 2011; Follette-Cook et al., 2015). The diurnal cycle  
345 of the PBL also plays a large role because horizontal dispersion occurs as the PBL thickens during  
346 the day. Early in the morning, the PBL is low ( $\sim 1400$  m during 9:00-11:00 in SMA) and strong  
347 sources are evident such as traffic on major highways, etc. As the day progresses, the PBL height  
348 increases ( $\sim 1800$  m during 15:00-17:00; Figure S9) allowing for greater horizontal mixing to take  
349 place. By early afternoon, emissions from all the major sources in the central region have mixed  
350 together to form a wide area of high pollution over the urban center. Judd et al. (2018) point out  
351 that the topography over SMA also plays a role in the ability to mix horizontally as the PBL grows.  
352 Therefore, the TeMD can be large between morning and afternoon (i.e., for Dt larger than 6 hours).

353 For a small Dt (2 or 4 hours), TeMD increases when increasing the satellite retrieval spatial  
354 resolution (i.e., smaller pixel size). This is especially true for short time periods (e.g., 2 hours and  
355 4 hours), which is more important for the GEO satellite measurements. For example, for Dt of 2  
356 hours, TeMD for satellite pixels of 1 km  $\times$  1 km is about  $0.80 \times 10^{16}$  molecules cm<sup>-2</sup>, while TeMD  
357 for satellite pixels of 25 km  $\times$  25 km is about  $0.73 \times 10^{16}$  molecules/cm<sup>2</sup> ( $\sim 9\%$  lower); when Dt is  
358 4 hours, TeMD for satellite pixels of 1 km  $\times$  1 km is about  $1.3 \times 10^{16}$  molecules cm<sup>-2</sup>, while TeMD  
359 for satellite pixels of 25 km  $\times$  25 km is about  $1.1 \times 10^{16}$  molecules/cm<sup>2</sup> ( $\sim 15\%$  lower). This indicates  
360 that when increasing the satellite retrieval spatial resolution (decreasing pixel size), the temporal  
361 variability of the retrieved values will increase, even though the normalized satellite spatial SGV  
362 decreases. Thus, temporal resolution should be increased in conjunction with the increase in spatial  
363 resolution in order to enhance the accuracy of the satellite products. This is expected because  
364 averaging over a larger region smooths out temporal variability so producing smaller hourly  
365 differences. Our finding here is consistent with that of Fishman et al. (2011).

366 GeoTASO data over the Busan region is limited. Given the fewer flights, we are not able  
367 to show how TeMD changes with Dt over the Busan region in this study. However, we are able to  
368 show the relationship between TeMD and satellite pixel sizes for a limited range of Dt. During  
369 KORUS-AQ, there were only two rasters sampled over Busan with a Dt of 2 hours (Figure S10).



370 For this Dt of 2 hours, TeMD increases slightly when increasing the satellite retrieval spatial  
371 resolution (smaller pixel size). More data over the Busan region would help significantly for this  
372 analysis. As for sampled hypothetical satellite pixels over the LA Basin, for a given Dt, TeMD  
373 increases when increasing the satellite retrieval spatial resolution (smaller pixel size) for Dt equal  
374 to 4 and 8 hours (Figure S11). We note that with only 2 flight days of flight data, the GeoTASO  
375 data over LA is also limited. Despite the limited sample sizes, TeMD increases when increasing  
376 the satellite retrieval spatial resolution over both the Busan region and the LA Basin, which is  
377 consistent with the relationships over the SMA for a small Dt.

### 378 3.3 Results from Spatial Structure Function (SSF)

379 In this section, we show the analysis of SSF over SMA (Figure 7) as a complement to our  
380 analysis in Section 3.1. As mentioned before, SSF and SGV are different measures of spatial  
381 variability and are not directly comparable. This is because SSF is calculated based on differences  
382 between a single GeoTASO measurement and all the other GeoTASO measurements on the map,  
383 while SGV is derived based on variation among all the GeoTASO measurements within a  
384 hypothetical satellite pixel unit. SSF measures the averaged spatial difference at a given distance,  
385 while SGV directly quantifies the expected spatial variability within a satellite pixel at a given size.  
386 As both SSF and SGV are related to spatial variability, we include SSF in this study as an extension  
387 to SGV.

388 Figure 7a shows that the SSF in SMA initially increases with the distance between data  
389 points, peaks at around 40–60 km during most flights, and then decreases with distance between  
390 60 and 140 km. The number of paired GeoTASO data points when the distance is larger than 100  
391 km is relatively small (Figure S12) therefore conclusions beyond this distance are not included in  
392 this analysis. The increases in SSF for distances in the range of 1–25 km (Figure 7b) are consistent  
393 with the relationship between pixel sizes and the normalized satellite SGV shown in Figure 4. For  
394 example, over the 1–25 km range, Fig 4a shows the median increases from around 8% to around  
395 28%, an increase by a factor of 3.5, and the black line in Figure 7 shows an approximately similar  
396 factor (from  $0.33 \times 10^{16}$  molecules/cm<sup>2</sup> for 1 km to  $1.5 \times 10^{16}$  molecules/cm<sup>2</sup> for 25 km). This  
397 increase of SSF between 1–25 km is also seen over the Busan region and the LA Basin (Figure  
398 S13). We also notice that SSF shows a relatively strong dependence on the particular GeoTASO  
399 flight, while SGV is less sensitive, especially for small pixel sizes.

400 The shapes of the SSF are generally consistent with previous studies for modeled or in situ  
401 observations of NO<sub>2</sub> (Fishman et al., 2011; Follette-Cook et al., 2015). Previous studies also  
402 suggest that different aircraft campaigns may share the common shape of SSF but different  
403 magnitudes, which is strongly related to the fraction of polluted samples versus samples of  
404 background air in the campaign (Crawford et al., 2009; Fishman et al., 2011). Differences in the  
405 shape and size of particular cities also contribute to the differences in the SSF. For example, at a  
406 certain distance SSF may compare polluted areas within the same urban region, while over a  
407 different smaller city, the comparison at the same distance reveals the gradient between the  
408 polluted city and cleaner surrounding background air, so resulting in different peak values. Valin  
409 et al. (2011) found that the maximum in OH feedback in a NO<sub>x</sub>-OH steady-state relationship  
410 corresponds to a NO<sub>2</sub> e-folding decay length of 54 km in 5m/s winds. This may partially explain  
411 the peak between 40–60 km in SSF. As shown in Figures 2 and S7, the overall spatial variability  
412 over SMA is higher in the afternoon. Over SMA, the SSF in the morning is generally smaller than



413 in the afternoon, indicating higher spatial variability of tropospheric NO<sub>2</sub> VC in the afternoon (see  
414 also Judd et al., 2018). As described in Section 2.6, SSF discussed here (Figure 7) is calculated  
415 based on hourly bin. We also include SSF that is calculated within rasters in the supplement (Figure  
416 S14). The overall shapes of SSF (Figure S14) calculated on raster basis are similar to SSF  
417 calculated on hourly basis (Figure 7).

418 Previous studies (Fishman et al., 2011; Follette-Cook et al., 2015) used SSF values at a  
419 particular distance to indicate the satellite precision requirement at a corresponding resolution in  
420 order to resolve spatial structure over the pixel scale. For GEMS, the expected spatial differences  
421 over the scale of its pixel for the SMA and Busan regions are  $\sim 7.5 \times 10^{15}$  molecules cm<sup>-2</sup> and  
422  $\sim 3.5 \times 10^{15}$  molecules cm<sup>-2</sup>, respectively, taking the SSF values at 5 km to be representative. For  
423 TEMPO, the spatial difference is  $\sim 2.8 \times 10^{15}$  molecules cm<sup>-2</sup> over LA Basin taking the SSF value  
424 at 3 km. Assuming the NO<sub>2</sub> measurement precision requirement to be  $1 \times 10^{15}$  molecules cm<sup>-2</sup> for  
425 both TEMPO and GEMS (Chance et al., 2013; Kim et al., 2020), the expected spatial differences  
426 over the three regions are considerably higher than the precision requirement and should be easily  
427 characterized by both the GEMS and TEMPO missions.

#### 428 **4. Discussions and implications**

429 The relationship between satellite pixel sizes and the normalized satellite SGV is fairly  
430 robust over the different regions studied here, and Figure 4 points to the possibility of developing  
431 a generalized look-up table if more data were available in other regions. A generalized relationship  
432 between satellite pixel sizes and the temporal variability (Figure 6) is not as evident as the  
433 relationship between satellite pixel sizes and the normalized satellite SGV due to limited data.  
434 However, it is still useful for satellite observations over SMA, which is in the GEMS domain and  
435 should be helpful in satellite retrieval interpretation.

436 This study also has implications for satellite validation and evaluation, and satellite-in situ  
437 data comparisons of other trace gas species. Our initial motivation to study satellite SGV arose  
438 from our previous work on validation of MOPITT (Measurements of Pollution in the Troposphere)  
439 CO retrievals over urban regions (Tang et al., 2020). In that study, we compared the satellite  
440 retrievals with aircraft profiles, and realized that satellite SGV and representativeness error of  
441 aircraft profiles in the comparisons to MOPITT retrievals introduced uncertainties in the validation  
442 results. Previous studies have noticed the same issue for NO<sub>2</sub> (e.g., Nowlan et al., 2016, 2018;  
443 Judd et al., 2019; Pinardi et al., 2020; Tack et al., 2020), but this issue is difficult to address and  
444 quantify due to the limited spatial coverage of most aircraft observations. Even though only a few  
445 trace gas species are routinely retrieved, the gapless raster datasets of GeoTASO are a possible  
446 way to address this problem. The normalized SGV of the GeoTASO tropospheric NO<sub>2</sub> VC might  
447 serve as an upper bound to the SGV of CO, SO<sub>2</sub> and other species that share common source(s)  
448 with NO<sub>2</sub> but have relatively longer lifetimes, even if their spatial distributions may have different  
449 patterns (e.g., Chong et al., 2020). For example, at the resolution of 22 km × 22 km (resolution of  
450 MOPITT CO retrievals), the expected normalized satellite SGV of tropospheric NO<sub>2</sub> VC is  $\sim 30\%$ .  
451 Therefore, we might expect the normalized satellite SGV for tropospheric CO VC to be lower than  
452 this value.

453 To demonstrate this idea, we use the WRF-Chem regional model at an intermediary step.  
454 At the model resolution, if the SVG of the WRF-Chem model and GeoTASO NO<sub>2</sub> VC agree



455 reasonably well, then the model can be used to predict the SVG of other species that are chemically  
456 constrained with NO<sub>2</sub> at the model resolution and at coarser resolutions. This is shown in Figure 8  
457 which illustrates how SGV varies with satellite pixel size for NO<sub>2</sub> VC, CO VC, SO<sub>2</sub> VC, and  
458 formaldehyde (HCHO) VC calculated from a WRF-Chem simulation. The modeled NO<sub>2</sub>, CO, SO<sub>2</sub>,  
459 and HCHO concentrations are converted to VC, and are filtered to match the rasters of GeoTASO  
460 measurements (Figure S15). As expected, SGV of modeled NO<sub>2</sub> VC is higher than SGV of  
461 modeled CO VC, SO<sub>2</sub> VC, and HCHO VC. We also notice that SGV for modeled NO<sub>2</sub> VC, CO  
462 VC, SO<sub>2</sub> VC, and HCHO VC increases with pixel size, which is similar to that for GeoTASO  
463 measurements. The SGV for GeoTASO NO<sub>2</sub> shown in this figure (black lines) is calculated based  
464 on GeoTASO data that are regridded to the WRF-Chem grid (3 km × 3 km), making it slightly  
465 different from that in Figure 4. Note that a more comprehensive comparison requires further work  
466 and ideally actual dense GeoTASO-type measurements of CO and other species to address  
467 differences due to local sources on the background concentrations.

468 This study is also relevant to model comparison and evaluation with local observations.  
469 Whenever local observations are compared to grid data (e.g., comparisons between satellite  
470 retrievals and local observations, comparisons between grid-based model and local observations,  
471 and data assimilation), SGV will introduce uncertainties that need to be quantified to better  
472 interpret and understand the comparison results. For example, we note that at the resolution of 14  
473 km×14 km (a typical resolution for the forward-looking Multi-Scale Infrastructure for Chemistry  
474 and Aerosols Version 0; MUSICA-V0, <https://www2.acom.ucar.edu/sections/multi-scale-chemistry-modeling-musica>; Pfister et al. [2020]), the expected normalized satellite SGV of  
475 tropospheric NO<sub>2</sub> VC is ~25-30%. When comparing model simulations at a coarser resolution with  
476 local observations for tropospheric NO<sub>2</sub> VC, a normalized SGV larger than ~25-30% may be  
477 expected. If comparing for a specific vertical layer instead of vertical column, an even larger  
478 normalized SGV may occur.  
479

## 480 5. Conclusions

481 Satellite SGV is a key issue in interpreting satellite retrieval results. Quantifying studies  
482 have been lacking due to limited high-resolution observations. In this study, we quantified likely  
483 GEO satellite SGV by using GeoTASO measurements of tropospheric NO<sub>2</sub> VC over the urbanized  
484 and polluted Seoul Metropolitan Area (SMA) and the less-polluted Busan region during KORUS-  
485 AQ, and the Los Angeles (LA) Basin during the 2017 SARP campaigns. The main findings of this  
486 work are the following:

- 487 (1) The normalized satellite SGV increases with hypothetical satellite pixel sizes based on satellite  
488 pixel random sampling of hourly GeoTASO data, from ~10% (±5% for specific cases such as  
489 an individual day/time of day) for a pixel size of 0.5 km × 0.5 km to ~35% (±10% for specific  
490 cases such as an individual day/time of day) for the pixel size of 25 km × 25 km. This  
491 conclusion holds for all the three study regions, despite their different levels of urbanization  
492 and pollution, and for time of day, morning or afternoon.
- 493 (2) The normalized satellite SGV of tropospheric NO<sub>2</sub> VC could serve as an upper bound to  
494 satellite SGV of CO, SO<sub>2</sub> and other species that share common source(s) with NO<sub>2</sub> but have  
495 relatively longer lifetime, as supported by the high-resolution WRF-Chem simulation.
- 496 (3) The temporal variability (TeMD) within the same hypothetical satellite pixels increases with  
497 sampling time differences (Dt) over SMA. TeMD ranges from ~0.75×10<sup>16</sup> molecules cm<sup>-2</sup> at



498 Dt of 2 hours to  $\sim 2 \times 10^{16}$  molecules  $\text{cm}^{-2}$  (about three times higher) at Dt of 8 hours. TeMD is  
499 likely impacted by the short lifetime and diurnal cycle of  $\text{NO}_2$  due to emission activities and  
500 photolysis rate, and the meteorology and PBL evolution during the day. Improving the satellite  
501 retrieval temporal resolution is an effective way to enhance the capability of satellite products  
502 in resolving variabilities of  $\text{NO}_2$ .

503 (4) Temporal variability (TeMD) increases when increasing the satellite retrieval spatial resolution  
504 (i.e., smaller pixel size) in SMA. For example, when Dt is 2 hours, TeMD for satellite pixels  
505 with the size of  $25 \text{ km} \times 25 \text{ km}$  is about 20% lower compared to TeMD for satellite pixels with  
506 the size of  $1 \text{ km} \times 1 \text{ km}$ . Thus, temporal resolution should be increased along with any increase  
507 in spatial resolution in order to enhance the accuracy of satellite products.

508 (5) The spatial structure function (SSF) firstly increases with the distance between data points,  
509 peaks at around 40-60 km during most flight days, and then decreases with distance. This is  
510 generally consistent with previous studies.

511 (6) SSF analyses suggest that GEMS will encounter  $\text{NO}_2$  VC pixel scale spatial differences of  
512  $\sim 7.5 \times 10^{15}$  and  $\sim 3.5 \times 10^{15}$  molecules  $\text{cm}^{-2}$  over the SMA and Busan regions, respectively.  
513 TEMPO will encounter  $\text{NO}_2$  VC spatial differences at its pixel scale of  $\sim 2.8 \times 10^{15}$  molecules  
514  $\text{cm}^{-2}$  over the LA Basin. These differences should be easily resolved at the stated measurement  
515 precision requirement of  $1 \times 10^{15}$  molecules  $\text{cm}^{-2}$ .

516 (7) These findings are relevant to future satellite design and satellite retrieval interpretation,  
517 especially now with the deployment of the high-resolution GEO air quality satellite  
518 constellation, GEMS, TEMPO, and Sentinel-4. This study also has implication for satellite  
519 product validation and evaluation, satellite-in situ data comparisons, and more general point-  
520 grid data comparisons. These share similar issues of sub-grid variability and the need for  
521 quantification of representativeness error.

522 We note that this study has some uncertainties and limitations. (1) The variability at a  
523 resolution finer than  $250 \text{ m} \times 250 \text{ m}$  (i.e., GeoTASO's resolution) may introduce uncertainties to  
524 the analysis here, although this is beyond the scope of this study. (2) Even though a large number  
525 of GeoTASO retrievals have been analyzed in this study, we would still benefit from more high-  
526 resolution measurements with a broader spatiotemporal coverage, particularly over the Busan  
527 region. More GeoTASO-type data over the Busan region will help testing the consistence in TeMD  
528 over different regions. (3) The KORUS-AQ campaign was conducted in Spring (May and June),  
529 and the 2017 SARP campaign was also conducted in June. More GeoTASO-type measurements  
530 over South Korea during different season(s) would be particularly helpful to understand and  
531 generalize the findings in this study.

532 This work demonstrates the value of continued flights of GeoTASO-type instruments obtaining  
533 continuous, high spatial resolution data several times a day, particularly for the upcoming  
534 validation exercises for the GEO air quality satellite constellation.

535

### 536 **Acknowledgement**

537 The authors thank the GeoTASO team for providing the GeoTASO measurements. The authors  
538 thank the KORUS-AQ and SARP team for the campaign data. We thank the DIAL-HSRL team



539 for the mixing layer height data (available at [https://www-air.larc.nasa.gov/cgi-](https://www-air.larc.nasa.gov/cgi-bin/ArcView/korusaq)  
540 [bin/ArcView/korusaq](https://www-air.larc.nasa.gov/cgi-bin/ArcView/korusaq)). Tang was supported by a NCAR Advanced Study Program Postdoctoral  
541 Fellowship. Edwards was partially supported by the TEMPO Science Team under Smithsonian  
542 Astrophysical Observatory Subcontract SV3-83021. The National Center for Atmospheric  
543 Research (NCAR) is sponsored by the National Science Foundation. The authors thank Ivan  
544 Ortega and Sara-Eva Martinez-Alonso for helpful comments on the paper.

545

#### 546 **Data availability**

547 The KORUS-AQ and SARP data are available at [https://www-air.larc.nasa.gov/cgi-](https://www-air.larc.nasa.gov/cgi-bin/ArcView/korusaq)  
548 [bin/ArcView/korusaq](https://www-air.larc.nasa.gov/cgi-bin/ArcView/korusaq) and <https://www-air.larc.nasa.gov/cgi-bin/ArcView/lmos>, respectively.

549

#### 550 **Reference**

551 Al-Saadi, Jassim, Gregory Carmichael, James Crawford, Louisa Emmons, Saewung Kim, Chang-  
552 Keun Song, et al. (2014). KORUS-AQ: An international cooperative air quality field study in  
553 Korea, the KORUS-AQ white paper, 2014 ([https://espo.nasa.gov/korus-aq/content/](https://espo.nasa.gov/korus-aq/content/KORUS-AQ_White_Paper) KORUS-  
554 AQ\_White\_Paper).

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

558 Broccardo, S., Heue, K.-P., Walter, D., Meyer, C., Kokhanovsky, A., van der A, R., Piketh, S.,  
559 Langerman, K., and Platt, U.: Intra-pixel variability in satellite tropospheric NO<sub>2</sub> column densities  
560 derived from simultaneous space-borne and airborne observations over the South African  
561 Highveld, *Atmos. Meas. Tech.*, 11, 2797–2819, <https://doi.org/10.5194/amt-11-2797-2018>, 2018.

562 Chance, K., Liu, X., Suleiman, R. M., Flittner, D. E., Al-Saadi, J., & Janz, S. J. (2013, September).  
563 Tropospheric emissions: monitoring of pollution (TEMPO). In *Earth Observing Systems XVIII*  
564 (Vol. 8866, p. 88660D). International Society for Optics and Photonics.

565 Ching, J., Herwehe, J., and Swall, J.: On joint deterministic grid modeling and sub-grid variability  
566 conceptual framework for model evaluation, *Atmos. Environ.*, 40, 4935–4945, 2006.

567 Choi, S., Lamsal, L. N., Follette-Cook, M., Joiner, J., Krotkov, N. A., Swartz, W. H., Pickering,  
568 K. E., Loughner, C. P., Appel, W., Pfister, G., Saide, P. E., Cohen, R. C., Weinheimer, A. J., and  
569 Herman, J. R.: Assessment of NO<sub>2</sub> observations during DISCOVER-AQ and KORUS-AQ field  
570 campaigns, *Atmos. Meas. Tech.*, 13, 2523–2546, <https://doi.org/10.5194/amt-13-2523-2020>, 2020.

571 Chong, H., Lee, S., Kim, J., Jeong, U., Li, C., Krotkov, N. A., ... & Ahn, M. H. (2020). High-  
572 resolution mapping of SO<sub>2</sub> using airborne observations from the GeoTASO instrument during the  
573 KORUS-AQ field study: PCA-based vertical column retrievals. *Remote Sensing of Environment*,  
574 241, 111725.

575 Courrèges-Lacoste, G. B., Sallusti, M., Bulsa, G., Bagnasco, G., Veihelmann, B., Riedl, S., ... &  
576 Maurer, R. (2017, September). The copernicus sentinel 4 mission: a geostationary imaging UVN



- 577 spectrometer for air quality monitoring. In *Sensors, Systems, and Next-Generation Satellites XXI*  
578 (Vol. 10423, p. 1042307). International Society for Optics and Photonics.
- 579 Crawford, J.H., et al., 2009. Assessing scales of variability for constituents relevant to future  
580 geostationary satellite observations and models of air quality. *AGU*, 90(52), Fall Meet. Suppl.,  
581 Abstract A53Ae0237.
- 582 Denby, B., Cassiani, M., de Smet, P., de Leeuw, F., & Horálek, J. (2011). Sub-grid variability and  
583 its impact on European wide air quality exposure assessment. *Atmospheric environment*, 45(25),  
584 4220-4229.
- 585 Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Mao, D., Martínez-Alonso, S., Worden,  
586 H. M., Ziskin, D., and Andreae, M. O.: Radiance-based retrieval bias mitigation for the MOPITT  
587 instrument: the version 8 product, *Atmos. Meas. Tech.*, 12, 4561–4580,  
588 <https://doi.org/10.5194/amt-12-4561-2019>, 2019.
- 589 Finlayson-Pitts, B. J., & Pitts, J. N. (1997). Tropospheric air pollution: ozone, airborne toxics,  
590 polycyclic aromatic hydrocarbons, and particles. *Science*, 276(5315), 1045-1051.
- 591 Fishman, J., Silverman, M. L., Crawford, J. H., & Creilson, J. K. (2011). A study of regional-scale  
592 variability of in situ and model-generated tropospheric trace gases: Insights into observational  
593 requirements for a satellite in geostationary orbit. *Atmospheric environment*, 45(27), 4682-4694.
- 594 Follette-Cook, M. B., Pickering, K. E., Crawford, J. H., Duncan, B. N., Loughner, C. P., Diskin,  
595 G. S., ... & Weinheimer, A. J. (2015). Spatial and temporal variability of trace gas columns derived  
596 from WRF/Chem regional model output: Planning for geostationary observations of atmospheric  
597 composition. *Atmospheric Environment*, 118, 28-44.
- 598 Friedl, M., Sulla-Menashe, D. (2015). MCD12C1 MODIS/Terra+Aqua Land Cover Type Yearly  
599 L3 Global 0.05Deg CMG V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed  
600 2019-08-12 from <https://doi.org/10.5067/MODIS/MCD12C1.006>.
- 601 Gaubert, B., L. K. Emmons, K. Raeder, S. Tilmes, K. Miyazaki, A. F. Arellano Jr, N. Elguindi, C.  
602 Granier, W. Tang, J. Barré, H. M. Worden, R. R. Buchholz, D. P. Edwards, P. Franke, J. L.  
603 Anderson, M. Saunio, J. Schroeder, J.-H. Woo, I. J. Simpson, D. R. Blake, S. Meinardi, P. O.  
604 Wennberg, J. Crouse, A. Teng, M. Kim, R. R. Dickerson, H. He, X. Ren, S. E. Pusede, and G. S.  
605 Diskin (2020), Correcting model biases of CO in East Asia: impact on oxidant distributions during  
606 KORUS-AQ, *Atmos. Chem. Phys.*, 20, 14617–14647,  
607 <https://doi.org/10.5194/acp-20-14617-2020>.
- 608 Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S.,  
609 and Grainger, C. A.: Toward constraining regional-scale fluxes of CO<sub>2</sub> with atmospheric  
610 observations over a continent: 1. Observed spatial variability from airborne platforms, *J. Geophys.*  
611 *Res.-Atmos.*, 108, 4756, doi:4710.1029/2002JD003018, 2003.
- 612 Griffin, D., Zhao, X., McLinden, C. A., Boersma, F., Bourassa, A., Dammers, E., ... & Hayden, K.  
613 (2019). High-resolution mapping of nitrogen dioxide with TROPOMI: First results and validation  
614 over the Canadian oil sands. *Geophysical Research Letters*, 46(2), 1049-1060.



- 615 Harris, D., Foufoula-Georgiou, E., Droegemeier, K.K., Levit, J.J., 2001. Multiscale statistical  
616 properties of a high-resolution precipitation forecast. *J. Hydrometeor.* 2, 406e418.
- 617 Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert,  
618 J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-  
619 I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014)  
620 anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data  
621 System (CEDS), *Geosci. Model Dev.*, 11, 369–408, <https://doi.org/10.5194/gmd-11-369-2018>,  
622 2018.
- 623 Huang, M., Crawford, J. H., Diskin, G. S., Santanello, J. A., Kumar, S. V., Pusede, S. E., ... &  
624 Carmichael, G. R. (2018). Modeling Regional Pollution Transport Events During KORUS-AQ:  
625 Progress and Challenges in Improving Representation of Land-Atmosphere Feedbacks. *Journal of*  
626 *Geophysical Research: Atmospheres*, 123(18), 10-732.
- 627 Judd, L. M., Al-Saadi, J. A., Janz, S. J., Kowalewski, M. G., Pierce, R. B., Szykman, J. J., Valin,  
628 L. C., Swap, R., Cede, A., Mueller, M., Tiefengraber, M., Abuhassan, N., and Williams, D.:  
629 Evaluating the impact of spatial resolution on tropospheric NO<sub>2</sub> column comparisons within urban  
630 areas using high-resolution airborne data, *Atmos. Meas. Tech.*, 12, 6091–6111,  
631 <https://doi.org/10.5194/amt-12-6091-2019>, 2019.
- 632 Kim, H. C., Lee, P., Judd, L., Pan, L., and Lefer, B.: OMI NO<sub>2</sub> column densities over North  
633 American urban cities: the effect of satellite footprint resolution, *Geosci. Model Dev.*, 9, 1111–  
634 1123, <https://doi.org/10.5194/gmd-9-1111-2016>, 2016.
- 635 Kim, H., Zhang, Q., & Heo, J. (2018). Influence of intense secondary aerosol formation and long-  
636 range transport on aerosol chemistry and properties in the Seoul Metropolitan Area during spring  
637 time: results from KORUS-AQ. *Atmospheric Chemistry and Physics*, 18(10), 7149-7168.
- 638 Kim, J., Jeong, U., Ahn, M. H., Kim, J. H., Park, R. J., Lee, H., ... & Jeong, M. J. (2020). New era  
639 of air quality monitoring from space: Geostationary Environment Monitoring Spectrometer  
640 (GEMS). *Bulletin of the American Meteorological Society*, (2020), 00.
- 641 Lamsal, L.N., Janz, S.J., Krotkov, N.A., Pickering, K.E., Spurr, R.J.D., Kowalewski, M.G.,  
642 Loughner, C.P., Crawford, J.H., Swartz, W.H., Herman, J.R., 2017. High-resolution NO<sub>2</sub>  
643 observations from the airborne Compact atmospheric mapper: retrieval and validation. *J. Geophys.*  
644 *Res.: Atmospheres* 122 (3), 1953–1970.
- 645 Leitch, J. W., Delker, T., Good, W., Ruppert, L., Murcray, F., Chance, K., Liu, X., Nowlan, C.,  
646 Janz, S. J., Krotkov, N. A., Pickering, K. E., Kowalewski, M., and Wang, J.: The GeoTASO  
647 airborne spectrometer project, Earth Observing Systems XIX, 17–21 August 2014, San Diego,  
648 California, United States, *Proc. SPIE*, 9218, 92181H-9, <https://doi.org/10.1117/12.2063763>, 2014.
- 649 Levelt, P. F., Oord, G. H. J. van den, Dobber, M. R., Malkki, A., Visser, H., Vries, J. de, Stammes,  
650 P., Lundell, J. O. V., and Saari, H.: The ozone monitoring instrument, *IEEE T. Geosci. Remote*,  
651 44, 1093–1101, <https://doi.org/10.1109/TGRS.2006.872333>, 2006.





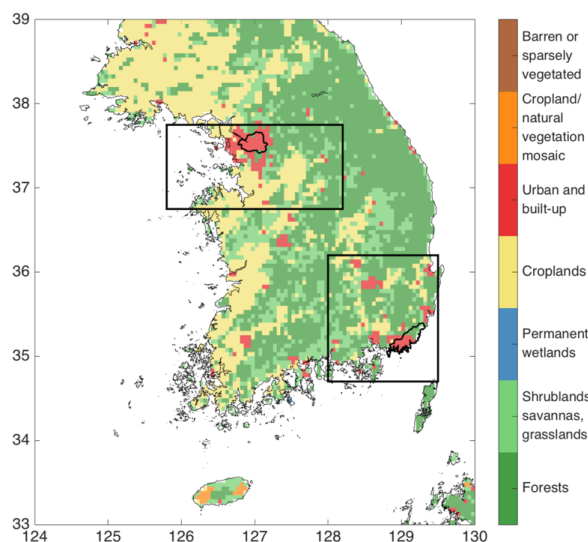
- 652 Levelt, P. F., Joiner, J., Tamminen, J., Veefkind, J. P., Bhartia, P. K., & Carn, S. (2018). The Ozone  
653 Monitoring Instrument: overview of 14 years in space. *Atmospheric Chemistry and Physics*, 18,  
654 5699.
- 655 Miyazaki, K., Sekiya, T., Fu, D., Bowman, K. W., Kulawik, S. S., Sudo, K., ... & Eskes, H. (2019).  
656 Balance of Emission and Dynamical Controls on Ozone During the Korea-United States Air  
657 Quality Campaign From Multiconstituent Satellite Data Assimilation. *Journal of Geophysical  
658 Research: Atmospheres*, 124(1), 387-413.
- 659 National Centers for Environmental Prediction/National Weather Service/NOAA/U.S.  
660 Department of Commerce. 2015, updated daily. NCEP GDAS/FNL 0.25 Degree Global  
661 Tropospheric Analyses and Forecast Grids. Research Data Archive at the National Center for  
662 Atmospheric Research, Computational and Information Systems Laboratory.  
663 <https://doi.org/10.5065/D65Q4T4Z>.
- 664 Nowlan, C. R., Liu, X., Leitch, J. W., Chance, K., González Abad, G., Liu, C., Zoogman, P., Cole,  
665 J., Delker, T., Good, W., Murcray, F., Ruppert, L., Soo, D., Follette-Cook, M. B., Janz, S. J.,  
666 Kowalewski, M. G., Loughner, C. P., Pickering, K. E., Herman, J. R., Beaver, M. R., Long, R. W.,  
667 Szykman, J. J., Judd, L. M., Kelley, P., Luke, W. T., Ren, X., and Al-Saadi, J. A.: Nitrogen dioxide  
668 observations from the Geostationary Trace gas and Aerosol Sensor Optimization (GeoTASO)  
669 airborne instrument: Retrieval algorithm and measurements during DISCOVER-AQ Texas 2013,  
670 *Atmos. Meas. Tech.*, 9, 2647–2668, <https://doi.org/10.5194/amt-9-2647-2016>, 2016.
- 671 Nowlan, C. R., Liu, X., Janz, S. J., Kowalewski, M. G., Chance, K., Follette-Cook, M. B., Fried,  
672 A., González Abad, G., Herman, J. R., Judd, L. M., Kwon, H.-A., Loughner, C. P., Pickering, K.  
673 E., Richter, D., Spinei, E., Walega, J., Weibring, P., and Weinheimer, A. J.: Nitrogen dioxide and  
674 formaldehyde measurements from the GEOstationary Coastal and Air Pollution Events (GEO-  
675 CAPE) Airborne Simulator over Houston, Texas, *Atmos. Meas. Tech.*, 11, 5941–5964,  
676 <https://doi.org/10.5194/amt11-5941-2018>, 2018.
- 677 Pfister, G. G., Eastham, S. D., Arellano, A. F., Aumont, B., Barsanti, K. C., Barth, M. C., ... &  
678 Fiore, A. M. (2020). The Multi-Scale Infrastructure for Chemistry and Aerosols (MUSICA).  
679 *Bulletin of the American Meteorological Society*, 101(10), E1743-E1760.
- 680 Pillai, D., Gerbig, C., Marshall, J., Ahmadov, R., Kretschmer, R., Koch, T., and Karstens, U.: High  
681 resolution modeling of CO<sub>2</sub> over Europe: implications for representation errors of satellite  
682 retrievals, *Atmos. Chem. Phys.*, 10, 83–94, <https://doi.org/10.5194/acp-10-83-2010>, 2010.
- 683 Pinardi, G., Van Roozendaal, M., Hendrick, F., Theys, N., Abuhassan, N., Bais, A., Boersma, F.,  
684 Cede, A., Chong, J., Donner, S., Drosoglou, T., Dzhola, A., Eskes, H., Frieß, U., Granville, J.,  
685 Herman, J. R., Holla, R., Hovila, J., Irie, H., Kanaya, Y., Karagkiozidis, D., Kouremeti, N.,  
686 Lambert, J.-C., Ma, J., Peters, E., Piters, A., Postlyakov, O., Richter, A., Remmers, J., Takashima,  
687 H., Tiefengraber, M., Valks, P., Vlemmix, T., Wagner, T., and Wittrock, F.: Validation of  
688 tropospheric NO<sub>2</sub> column measurements of GOME-2A and OMI using MAX-DOAS and direct  
689 sun network observations, *Atmos. Meas. Tech.*, 13, 6141–6174, [https://doi.org/10.5194/amt-13-  
690 6141-2020](https://doi.org/10.5194/amt-13-6141-2020), 2020.



- 691 Qian, Y., Gustafson Jr., W. I., and Fast, J. D.: An investigation of the sub-grid variability of trace  
692 gases and aerosols for global climate modeling, *Atmos. Chem. Phys.*, 10, 6917–6946,  
693 <https://doi.org/10.5194/acp-10-6917-2010>, 2010.
- 694 Song, H., Zhang, Z., Ma, P.-L., Ghan, S., and Wang, M.: The importance of considering sub-grid  
695 cloud variability when using satellite observations to evaluate the cloud and precipitation  
696 simulations in climate models, *Geosci. Model Dev.*, 11, 3147–3158, <https://doi.org/10.5194/gmd-11-3147-2018>, 2018.
- 698 Spinei, E., Whitehill, A., Fried, A., Tiefengraber, M., Knepp, T. N., Herndon, S., ... & Richter, D.  
699 (2018). The first evaluation of formaldehyde column observations by improved Pandora  
700 spectrometers during the KORUS-AQ field study. *Atmospheric Measurement Techniques*, 11(9),  
701 4943-4961.
- 702 Souri, A. H., Nowlan, C. R., Wolfe, G. M., Lamsal, L. N., Miller, C. E. C., Abad, G. G., ... &  
703 Diskin, G. S. (2020). Revisiting the effectiveness of HCHO/NO<sub>2</sub> ratios for inferring ozone  
704 sensitivity to its precursors using high resolution airborne remote sensing observations in a high  
705 ozone episode during the KORUS-AQ campaign. *Atmospheric Environment*, 117341.
- 706 Tack, F., Merlaud, A., Iordache, M.-D., Pinaridi, G., Dimitropoulou, E., Eskes, H., Bomans, B.,  
707 Veeffkind, P., and Van Roozendaal, M.: Assessment of the TROPOMI tropospheric NO<sub>2</sub> product  
708 based on airborne APEX observations, *Atmos. Meas. Tech. Discuss.* [preprint],  
709 <https://doi.org/10.5194/amt-2020-148>, in review, 2020.
- 710 Tang, W., Arellano, A. F., DiGangi, J. P., Choi, Y., Diskin, G. S., Agustí-Panareda, A., Parrington,  
711 M., Massart, S., Gaubert, B., Lee, Y., Kim, D., Jung, J., Hong, J., Hong, J.-W., Kanaya, Y., Lee,  
712 M., Stauffer, R. M., Thompson, A. M., Flynn, J. H., and Woo, J.-H.: Evaluating high-resolution  
713 forecasts of atmospheric CO and CO<sub>2</sub> from a global prediction system during KORUS-AQ field  
714 campaign, *Atmos. Chem. Phys.*, 18, 11007–11030, <https://doi.org/10.5194/acp-18-11007-2018>,  
715 2018.
- 716 Tang, W., Emmons, L. K., Arellano Jr, A. F., Gaubert, B., Knote, C., Tilmes, S., ... & Blake, N. J.  
717 (2019). Source Contributions to Carbon Monoxide Concentrations During KORUS-AQ Based on  
718 CAM-chem Model Applications. *Journal of Geophysical Research: Atmospheres*, 124(5), 2796-  
719 2822.
- 720 Tang, W., Worden, H. M., Deeter, M. N., Edwards, D. P., Emmons, L. K., Martínez-Alonso, S.,  
721 Gaubert, B., Buchholz, R. R., Diskin, G. S., Dickerson, R. R., Ren, X., He, H., and Kondo, Y.:  
722 Assessing Measurements of Pollution in the Troposphere (MOPITT) carbon monoxide retrievals  
723 over urban versus non-urban regions, *Atmos. Meas. Tech.*, 13, 1337–1356,  
724 <https://doi.org/10.5194/amt-13-1337-2020>, 2020.
- 725 Valin, L. C., Russell, A. R., Hudman, R. C., and Cohen, R. C.: Effects of model resolution on the  
726 interpretation of satellite NO<sub>2</sub> observations, *Atmos. Chem. Phys.*, 11, 11647–11655,  
727 <https://doi.org/10.5194/acp-11-11647-2011>, 2011.
- 728 Veeffkind, J. P., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G., ... & Van Weele, M.  
729 (2012). TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of



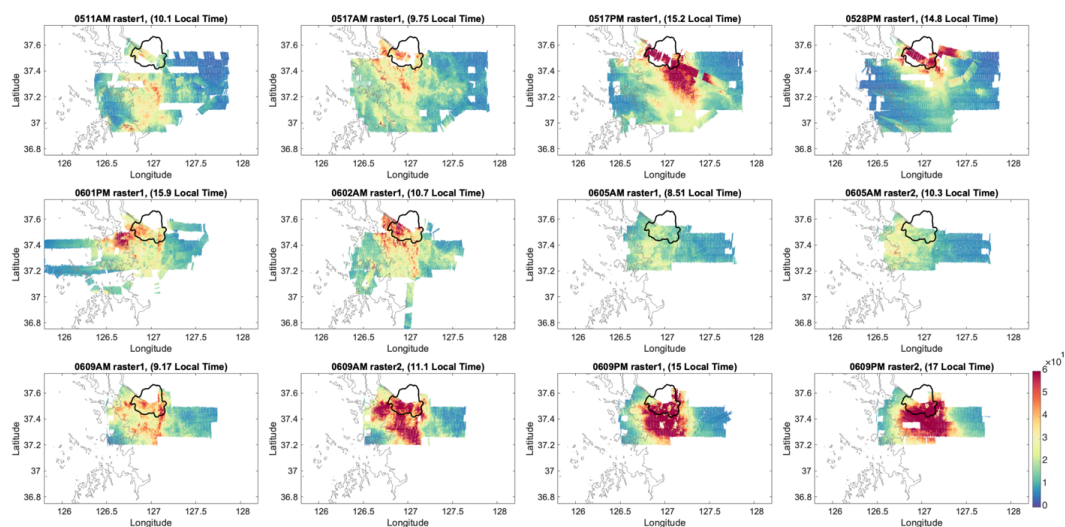
- 730 the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sensing*  
731 of Environment, 120, 70-83.
- 732 van Geffen, J., Boersma, K. F., Eskes, H., Sneep, M., ter Linden, M., Zara, M., and Veeffkind, J.  
733 P.: S5P TROPOMI NO<sub>2</sub> slant column retrieval: method, stability, uncertainties and comparisons  
734 with OMI, *Atmos. Meas. Tech.*, 13, 1315–1335, <https://doi.org/10.5194/amt-13-1315-2020>, 2020.
- 735 Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., &  
736 Soja, A. J. (2011). The Fire INventory from NCAR (FINN): A high resolution global model to  
737 estimate the emissions from open burning. *Geoscientific Model Development*, 4(3), 625.
- 738 Zhang, Z., Song, H., Ma, P.-L., Larson, V. E., Wang, M., Dong, X., and Wang, J.: Subgrid  
739 variations of the cloud water and droplet number concentration over the tropical ocean: satellite  
740 observations and implications for warm rain simulations in climate models, *Atmos. Chem. Phys.*,  
741 19, 1077–1096, <https://doi.org/10.5194/acp-19-1077-2019>, 2019.
- 742 Zoogman, P., Liu, X., Suleiman, R. M., Pennington, W. F., Flittner, D. E., Al-Saadi, J. A., ... &  
743 Janz, S. J. (2017). Tropospheric emissions: Monitoring of pollution (TEMPO). *Journal of*  
744 *Quantitative Spectroscopy and Radiative Transfer*, 186, 17-39.  
745



746 **Figure 1.** Domain of the study over South Korea and the land cover. Boxes indicate location of  
747 the SMA (upper left) and the Busan region (lower right) domains. Land cover data are from  
748 MODIS Terra and Aqua MCD12C1 L3 product, version V006, annual mean at 0.05° resolution;  
749 Friedl et al., 2015.  
750  
751



752

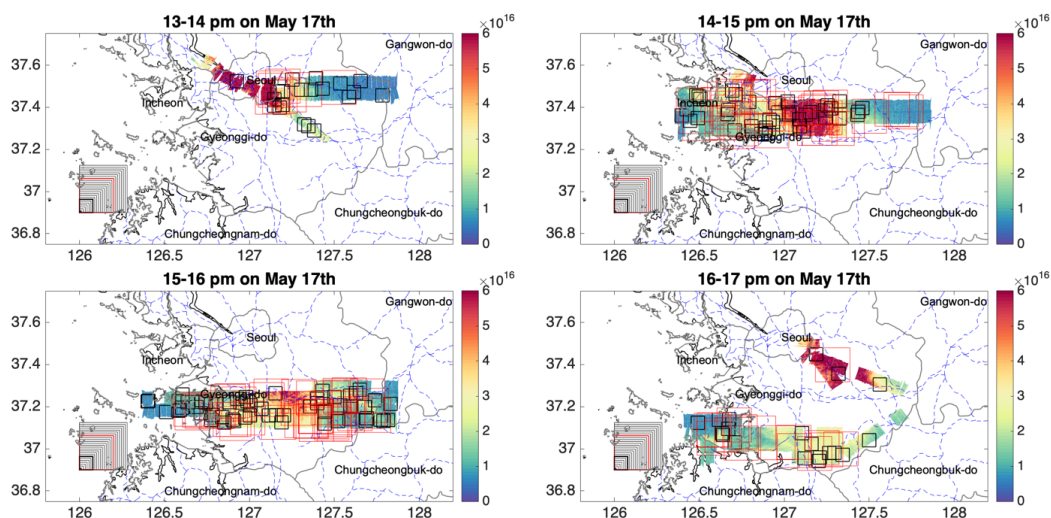


753  
754  
755  
756  
757  
758  
759  
760  
761  
762

**Figure 2.** GeoTASO data of tropospheric NO<sub>2</sub> vertical column (molecules cm<sup>-2</sup>) measured during KORUS-AQ over the Seoul region. Each panel shows a separate raster. Panel titles show month, day, AM/PM, raster number on that date, and mean time of raster acquisition. There were nine flights sampling rasters over Seoul. The May 01 AM, May 17 AM, May 17 PM, May 28 PM, June 01 PM, and June 02 AM flights each sampled one raster. The June 05 AM, June 09 AM, and June 09 PM flights each sampled two rasters. As a result, there were two flights and two rasters on May 17th, one flight and two rasters on June 5th, and two flights and four rasters on June 9th.



763

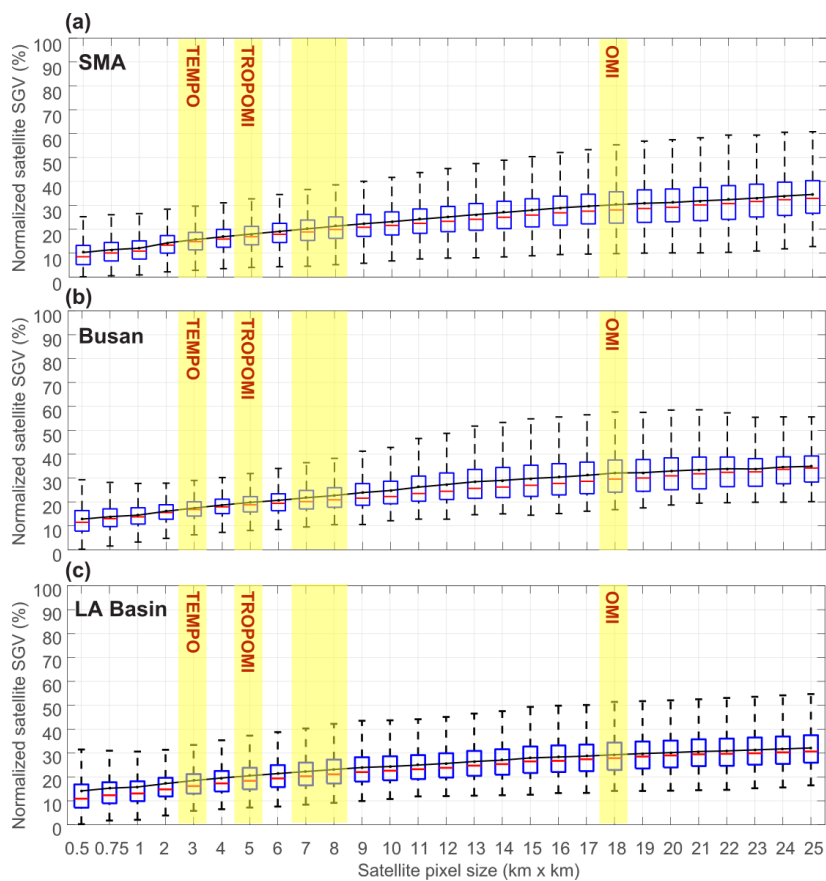


764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776

**Figure 3.** Demonstration of the hypothetical satellite pixel random sampling method. Each subplot is an hour during May 17<sup>th</sup> PM flight. For each hour, we randomly sample 10000 hypothetical satellite pixels at each different pixel sizes (i.e., 0.5 km×0.5 km, 0.75 km×0.75 km, 1 km×1 km, 2 km×2 km, ... , 25 km×25 km) over the GeoTASO data of tropospheric NO<sub>2</sub> vertical column (molecules cm<sup>-2</sup>) every hour. The sampled pixel size (from 0.5 km×0.5 km to 25 km×25 km) are shown in the lower-left corner of each sub-plot. Only 100 samples for pixel size of 7 km×7 km (thick black box) and 100 samples for 18 km × 18 km are shown for demonstration purposes. Samples that fail to pass the 75% coverage threshold are not shown. Coastlines, Province/Metropolitan City boundaries are shown by gray solid lines. Main roads are shown by blue dashed lines (data are from <http://www.diva-gis.org/gdata>).

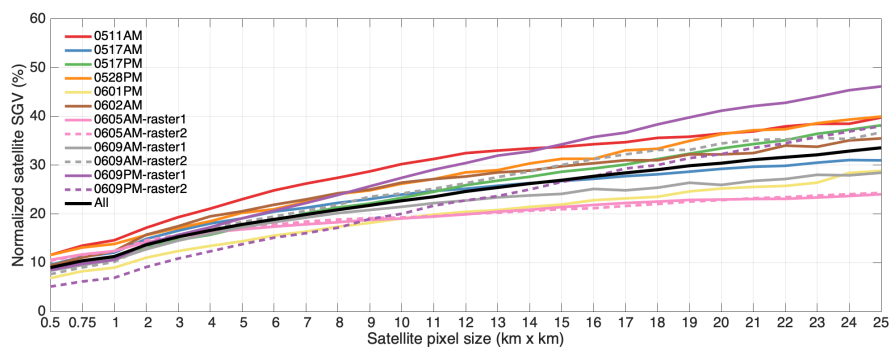


777



778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789

**Figure 4.** Boxplot (with medians represented by red bars, interquartile ranges between 25th and 75th percentiles represented by blue boxes, and the most extreme data points not considered outliers represented by whiskers) for the normalized satellite sub-grid variability (SGV) over the Seoul Metropolitan Area (a), the Busan region (b), and Los Angeles Basin (c). Normalized satellite SGV is calculated as the standard deviation of the GeoTASO data within the sampled satellite pixel divided by the mean of the GeoTASO data within the sampled satellite pixel. The black lines represent the mean of the normalized satellite SGV at a given size. The resolutions of TEMPO, TROPOMI, GEMS, and OMI are highlighted by the yellow shade in the Figure.

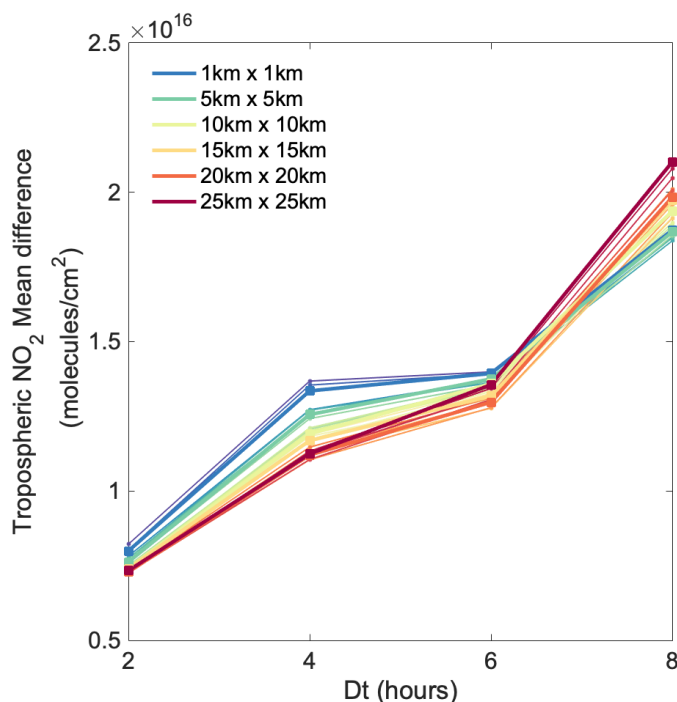


790  
791  
792  
793  
794  
795  
796  
797  
798

**Figure 5.** Average of the normalized satellite sub-grid variability (SGV) sampled individually from the twelve rasters (represented by the colored lines), and sampled from all the twelve rasters together (represented by the black line) over the Seoul Metropolitan Area during KORUS-AQ. Normalized satellite SGV is calculated by the standard deviation of the GeoTASO data within the sampled satellite pixel divided by the mean of the GeoTASO data within the sampled satellite pixel.



799



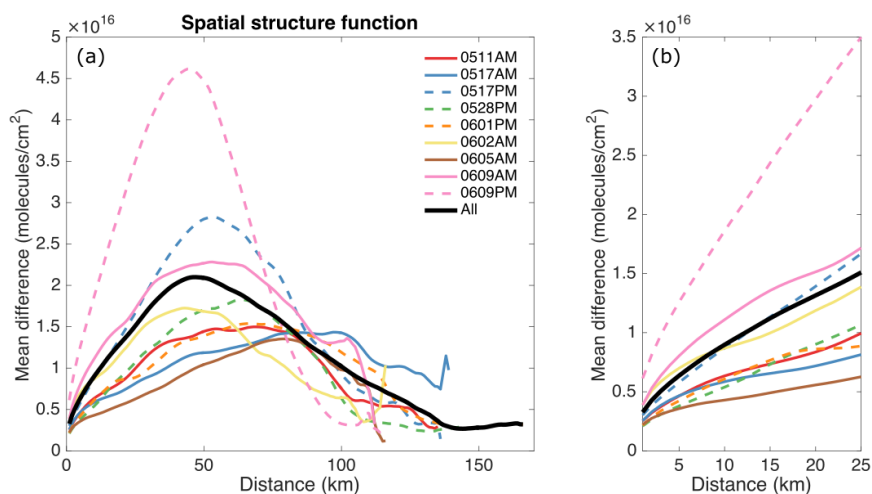
800  
801  
802  
803  
804  
805  
806  
807  
808  
809

**Figure 6.** Temporal mean differences (TeMD) of hypothetical satellite pixel retrieved tropospheric  $\text{NO}_2$  vertical column ( $\text{molecules cm}^{-2}$ ) over the Seoul Metropolitan Area (y-axis) as a function of satellite pixel size time difference (Dt). Mean differences for the time difference of Dt are calculated by averaging absolute value of the differences across all sampled satellite pixels that have two values with time difference of Dt. Results for each pixel size are color-coded, with selected sizes shown with thicker lines for reference.



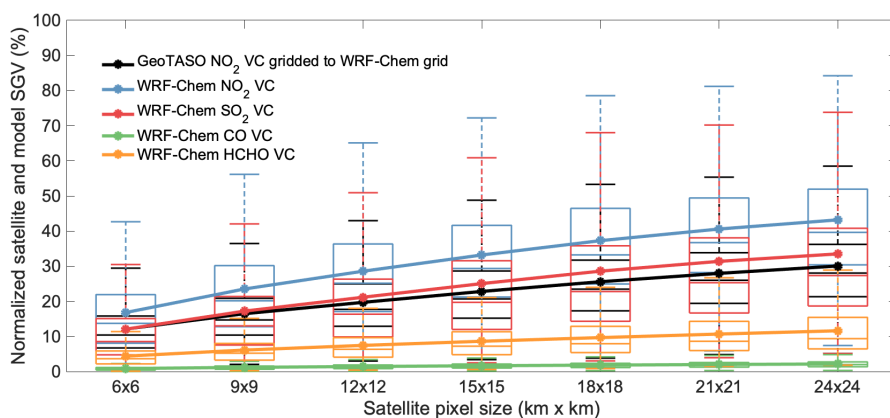


810  
811  
812



813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829

**Figure 7.** (a) Spatial Structure Function (SSF) for GeoTASO data of tropospheric  $\text{NO}_2$  vertical column molecules  $\text{cm}^{-2}$  over the Seoul Metropolitan Area (SMA) during KORUS-AQ and (b) the zoom-in version of panel (a) for distance range of 1-25 km. The SSF calculates average of absolute value of  $\text{NO}_{2,\text{VC}}$  differences (i.e., mean difference; y-axis) across all data pairs (measured in the same hourly bin) that are separated by different distance (x-axis). The SSF based on GeoTASO data measured during morning flights are in solid colored lines while the SSF based on GeoTASO data measured during afternoon flights are in dashed colored lines. The SSF based on all the data is in the black solid line.



830

831

832 **Figure 8.** Boxplot of hypothetical satellite normalized SGV of NO<sub>2</sub> vertical column (VC), SO<sub>2</sub>  
833 VC, CO VC, and formaldehyde (HCHO) VC derived from the WRF-Chem simulation with a  
834 resolution of 3 km × 3 km (colored lines), and GeoTASO NO<sub>2</sub> VC that gridded to the WRF-Chem  
835 grid (black lines) over the Seoul Metropolitan Area. Medians are represented by red bars,  
836 interquartile ranges between 25th and 75th percentiles by blue boxes, and the most extreme data  
837 points not considered outliers by whiskers. The modeled NO<sub>2</sub>, CO, SO<sub>2</sub>, and HCHO are filtered to  
838 match the rasters of GeoTASO measurements.

839