Response to Reviewers' Comments

Dear Editor and Reviewer:

We appreciate for your thoughtful and helpful comments. We tried to answer your comments, and our 'Response' is embedded below. All changes we have made in the revised manuscript have not only been mentioned in each response for the reviewer's comments, but also marked with an MS Word tracking option. We hope we have provided the appropriate answers and corresponding modifications. If there are any further questions, please let us know.

Best Regards,

Authors

Authors' response to RC2 from referee # 1:

General Comments

This paper describes a simplified retrieval for NO_2 from solar backscatter measurements, based on wavelength pair ratios (on/off absorption spectral lines)—the Modified Wavelength Pair (MWP) method—designed for use with low-cost hyperspectral sensors that lack the measurement stability of satellite and more-expensive airborne instruments. This technique is applied to the Hyperspectral Imagining Sensor (HIS), which was flown on aircraft over three significant pollution sources in Korea. An analytical uncertainty analysis is included. Results are compared with TROPOMI retrievals, and sub-satellite-grid-scale differences are discussed, in the context of geophysical variations.

The manuscript is well written, thorough, and generally clear. I recommend minor revisions.

Specific Comments

Paragraph starting in L48: Mention TEMPO and other, geostationary spacecraft, which are also achieving fairly good spatial resolution information.

→ Thank you for your suggestion. However, the paragraph is intended to describe the development of Low-Earth Orbit (LEO) satellites and to emphasize that the S5P (Sentinel-5 Precursor) TROPOMI, which has been used in this study for comparison with the airborne HIS measurements, has significantly enhanced spatial resolution than its predecessors, but is still struggling to resolve the spatial inhomogeneity of trace gases such as NO_2. Including additional descriptions about the recent geostationary satellite instruments such as the Geostationary Environment Monitoring Spectrometer (GEMS; Kim et al., 2020) or Tropospheric Emissions: Monitoring of Pollution (TEMPO; Chance et al., 2013; Zoogman et al., 2017) might be more informative, but none of them are currently producing official public-release datasets yet. Moreover, we believe that not mentioning the geostationary satellites in the paragraph would be helpful in terms of the overall fluency of the introduction; hence, we decided to keep it as it was in the original manuscript.

References

Chance, K. V., Liu, X., Suleiman, R. M., Flittner, D. E., Al-Saadi, J. and Janz, S. J.: Tropospheric emissions: monitoring of pollution (TEMPO), In Proc. SPIE 8866, Earth Observing Systems XVIII, doi:10.1117/12.2024479. http://dx.doi.org/10.1117/12.2024479, 2013.

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., Miller, C. 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. kyun, 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.

Zoogman, P., Liu, X., Suleiman, R., Pennington, W., Flittner, D., Al-Saadi, J., Hilton, B., Nicks, D., Newchurch, M., Carr, J., Janz, S., Andraschko, M., Arola, A., Baker, B., Canova, B., Miller, C. C., Cohen, R., Davis, J., Dussault, M., Edwards, D., Fishman, J., Ghulam, A., Abad, G. G., Grutter, M., Herman, J., Houck, J., Jacob, D., Joiner, J., Kerridge, B., Kim, J., Krotkov, N., Lamsal, L., Li, C., Lindfors, A., Martin, R., McElroy, C., McLinden, C., Natraj, V., Neil, D., Nowlan, C., O'Sullivan, E., Palmer, P., Pierce, R., Pippin, M., Saiz-Lopez, A., Spurr, R., Szykman, J., Torres, O., Veefkind, J., Veihelmann, B., Wang, H., Wang, J., and Chance, K.: Tropospheric emissions: Monitoring of pollution (TEMPO), J. Quant. Spectrosc. Ra., 186, 17–39, https://doi.org/10.1016/j.jqsrt.2016.05.008, 2017.

Table 1: Please include some information about the f-number, etendue, and/or SNR of the system (under a given set of circumstances), that would indicate the optical throughput of the system.

→ We have added some additional information (i.e., f-number) of the HIS in Table 1 of the revised manuscript. However, due to our unawareness regarding the optical hardware, we decided not to add the information we are not entirely confident about. Further details about the instruments can be found in "HIS_Headwall_specifications.pdf", which has been added in a supplementary material repository at https://doi.org/10.7910/DVN/YCZ9JU (please note that the HIS has an ADC depth of 16 bits, not 14 bits as stated in the document). Furthermore, discussions about the SNR of the system have been discussed in section 3.3, and no further thorough descriptions are available from the manufacturing perspective.

Section 3.1: Please reference the use of wavelength pairs in other retrievals. For example, wavelength pairs were long used in the retrieval of total ozone.

→ There have been previous studies that used one or more wavelength pairs to retrieve columnar concentrations of atmospheric O_3 and NO_2, which are gases with significant optical depth in the atmosphere in the UV–VIS band attributed to their abundance and strong absorption feature. These studies utilized ground-based instruments such as the Dobson spectrophotometer (Dobson, 1957a; 1957b) or the Brewer spectrophotometer (Brewer, 1973; Brewer and Kerr, 1973; Brewer et al., 1973). We appreciate your suggestion and have added some sentences and corresponding references in lines 130–134 of the revised manuscript.

References

Brewer, A. W.: A Replacement for the Dobson Spectrophotometer?, Pure and Applied Geophysics (PAGEOPH), 106–108, 1973.

Brewer, A. W. and Kerr, J. B.: Total Ozone Measurements in Cloudy Weather, Pure and Applied Geophysics (PAGEOPH), 106–108, 1973.

Brewer, A. W., McElroy, C. T., and Kerr, J. B.: Nitrogen Dioxide Concentrations in the Atmosphere, Nature, 246, 1973.

Dobson, G. M. B.: Observers' handbook for the ozone spectrophotometer, in Annals of the International Geophysical Year, V, Part 1, 46-89, Pergamon Press, 1957a.

Dobson, G. M. B.: Adjustment and calibration of the ozone spectrophotometer, ibid. V, Part I, 90-113, Pergamon Press, 1957b.

Section 3.1: By combining pairs of wavelengths, one with the shorter wavelength having the stronger absorption ("Type A") and the other with longer wavelength having the stronger absorption ("Type B"), you in effect partially cancel the bias from spectrally changing surface reflectivity (since the forward model does not include reflectivity spectral dependence). I suggest you say that explicitly. It would make the mathematical discussion easier to understand.

→ Thank you for the suggestion. The exact reason why we used a set of wavelength pairs comprised of one Type_A wavelength pair and one Type_B wavelength pair, instead of using a randomly matched dual-wavelength pair, was to minimize the effect of random uncertainties on the final analytical solution shown in Eq. (10). When the same type of wavelength pair (i.e., Type_A or Type_B) is used in Eq. (10), both the numerator and the denominator of the first term on the right-hand side are likely to be excessively small in absolute values resulting in amplified uncertainties. Therefore, it is necessary to ensure that a_A and a_B have different signs (as well as b_A and b_B) to ensure sufficiently large values for the denominator and numerator in Eq. (10). We tried to avoid stating too many technical details in the manuscript, which we believe is unnecessary and somewhat veiling the results of the airborne HIS NO_2 VCD observations. Instead, we have added a brief explanatory sentence in line 179 of the revised manuscript as an alternative.

Throughout the entire manuscript, I recommend using r (small r) for the correlation coefficient, to reduce confusion with the radiance ratio R (capital R).

→ Thank you for the suggestion. We tried to differentiate the radiance ratio R with the correlation coefficient R throughout the manuscript, but we agree that following your advice would be helpful for readers to avoid unnecessary confusion. Therefore, we have modified accordingly throughout the revised manuscript.

Paragraphs starting in L182: This section is confusing. In L183, what does "k-fold of the value" mean? The definition of k spectral dependency factor—is not introduced until L195. R is then discussed—I think meaning the radiance ratio—but since reflectivity is also being discussed, it is tempting to think R means reflectivity. In L184, "The same k value can be assumed for the wavelength pairs..."—why can that be assumed? In L193, "relation between the biased VCD estimates"—biased in what way? Why are they biased? Do you mean they include measurement errors?

→ The whole paragraph you are referring to is meant to describe the situation where the spectral dependency of real-world reflectivity differing from assumed model-world reflectivity (i.e., RTM-input surface reflectivity). We assumed arbitrary unspecified constant k to represent the difference in spectral dependency of real-world and model-world reflectivity, and this difference contributes to the discrepancy of model-world radiance ratio (R_{rtm}) and real-world radiance ratio (R_{obs}) as shown in Eq. (2). We thought we first stated "k" by definition in line 183 of the original manuscript, while the statement in line 195 ("spectral dependency factor k") was a simple repetition for the reminder. We never used a symbol or the acronym for reflectivity in Sect. 3.1. and further assured that the "R" only stands for radiance ratio throughout the manuscript by following your advice to change the symbol of correlation coefficients to "r".

The grounds for assuming the same "k" value for the wavelength pairs in their vicinity (i.e., < 5 nm) is because what spectral dependency term k represents is the effect of smoothly varying terms with small high-spectral-frequency variabilities such as surface or aerosol reflectivity. Please refer to the explanations in lines 192–196 of the revised manuscript.

We referred VCD_{obs} as a biased VCD estimate because it is an estimation based on the HIS-observed radiance ratio value (R_{obs}) and the regression coefficients (i.e., a, b in Eq. 1) retrieved from the relation between simulated radiance ratio (R_{rtm}) and input NO_2 VCD to the RTM (Eq. 3). The aforementioned k, or the discrepancy of spectral dependency of reflectivity between the model and the real world (i.e., observations), causes the HIS-observed radiance ratio value (R_{obs}) to differ from the anticipated radiance ratio value (R_{rtm}). What we should retrieve from the radiance ratio affected by the unknown spectral dependency of reflectivity in the real world (i.e., R_{obs}) is the VCD_{true}, or the RTM-input VCD that resulted in the R_{rtm} and the corresponding regression coefficients (i.e., a, b in Eq. 1) representing the $R - VCD$ relationship.

L202, Eq (4): Equating VCD_True with VCD_rtm,A and VCD_rtm,B is confusing. If the modeled VCDs were "true," observations wouldn't be needed. Maybe I don't understand what "True" means in this context?

→ (As an extent to our response on the previous question) The concept of Eq. (4) is that regardless of which radiance ratio was used, the resulting VCD should be the same in principle. The whole point of the MWP method is to cancel out the spectral dependency term of reflectivity so that the NO_2 VCDs can be retrieved from the observed radiance ratios (i.e., $R_{obs,A}$, $R_{obs,B}$) and the VCD-R relation identified from the RTM simulations. VCD_{True} is a value that we would like to retrieve from a set of observed radiance ratios, not affected by the spectral dependency term "k".

L232: "three independent NO_2 VCD"—I understand what you mean, but they aren't really "independent." "Different" may be more accurate.

→ Thank you for the advice, and we have changed it accordingly in the revised manuscript (line241).

L234: "to increase the signal-to-noise ratio" and L236: "spectral binning of ±2 original spectral bins"—It would be good to capture this in Table 1, so the reader doesn't look at the table and this and think you have 5x spectral oversampling. The table would be more useful if it reflected the sampling/binning and SNR that are used in the retrieval.

→ We agree that the phrase "Spectral binning interval" in Table 1 can confuse readers that the spectra we used are effectively 5 times oversampled, which is not true. We appreciate your comment pointing this out, and have modified the corresponding row of Table 1 in the revised manuscript to clarify. The way we have estimated the SNR is from the dark observations; thus, our SNR estimates vary from scene to scene depending on the signal intensity. Therefore, we were unable to provide certain SNR value in a table. However, we have accounted for the SNR per each spectrum considering all the pre-binning (i.e., spectral and spatial binning) while estimating the uncertainties presented in Figs. 5~7.

L254: "highly resolved CTM data": Highly resolved in what way? Horizontal spatial? Vertically? Also, I would not say models produce "data"—they produce "output." Measurements are data.

→ Thank you for the comment. We have modified the corresponding sentence in line 264 of the revised manuscript to avoid referring to model output as "data", and further clarified that the high resolution refers "spatial" resolution. Enhancement of CTM vertical resolution will also help us to obtain a more realistic NO_2 vertical profile, but studies show that the (horizontal) spatial resolution has a greater influence on the accuracy of CTM outputs (Kim et al., 2016; Valin et al., 2011; Zhao et al., 2020).

References

Kim, H. C., Lee, P., Judd, L., Pan, L., and Lefer, B.: OMI NO_2 Column Densities over North American Urban Cities: The Effect of Satellite Footprint Resolution, Geoscientific Model Development, 9 (3), 1111–1123, doi:10.5194/gmd-9-1111-2016, 2016.

Valin, L. C., Russell, A. R., Hudman, R. C., and Cohen, R. C.: Effects of Model Resolution on the Interpretation of Satellite NO_2 Observations, Atmospheric Chemistry and Physics, 11 (22), 11647–11655, doi:10.5194/acp-11-11647-2011, 2011.

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, Atmospheric Measurement Techniques, 13 (4), 2131–2159, doi:10.5194/amt-13-2131-2020, 2020.

Paragraph staring in L366: Q is not clearly defined. What does "transpose the ratio between R_rtm,A and R_rtm,B as Q" (L369) mean? No Q ppears in Eq. (11).

→ The sentence "transpose the ratio between $R_{rtm,A}$ and $R_{obs,A}$ as Q" means that the Q is defined as the $R_{rtm,A}$ over $R_{rtm,B}$ (i.e., $Q = {R_{rtm,A}}/{R_{rtm,B}}$). Q appears in the denominator of the right-hand side of Eq. (11).

L472: "mean value"—How well did the three different VCDs agree? It would be useful to know how similar they are, and if they're different, why.

→ Theoretically, all three VCD values from three different wavelength pair sets should agree well. However, there are cases where three VCD values do not agree well in the actual retrieval. To minimize the unintended biases by simply taking the "mean value" of all three VCD values, we have taken the weighted averaging technique proposed by Xie et al. (2001) to minimize the uncertainty of the "weighted mean" value from the given data points and their respective uncertainties. Therefore, we would like to claim that the "weighted mean value" of the three VCDs, considering their respective uncertainties, has more or less reduced the uncertainty to a certain extent on every occasion, rather than deteriorating the uncertainty range.

Reference

Xie, S. X., Liao, D., and Chinchilli, V. M.: Measurement error reduction using weighted average method for repeated measurements from heterogeneous instruments, Environmetrics, 12(8), 785–790. https://doi.org/10.1002/env.511, 2001.

L484: Footprint is 400 m x 400 m, so this is essentially the entire swath width, correct? It would be helpful to state that.

→ Yes, indeed it is. The entire across-track pixels were post-binned to a single pixel, therefore yielding single-pixel HIS swath data with approximately 400 m swath width (at a flying altitude of 6,000 ft). However, the swath width varied by the observation altitude, becoming wider under higher altitudes and vice versa. We have added an explicit statement emphasizing that the entire across-track pixels were binned to a single post-binned across-track pixel as a final HIS data (please refer to lines 505–506 in the revised manuscript).

Figures 4/5/6/7: How were the TROPOMI data downscaled?

→ Figures 4, 5a, 6a, and 7a are the visualization of downscaled TROPOMI NO_2 VCD_{Total} (entitled "VCD_{summed}" in TROPOMI product) composites, from TROPOMI swaths in October and November from 2019 to 2022. First, the downscaled grid has been pre-determined for each figure domain (i.e., 0.25°×0.25° for Korea, 0.1°×0.1° for the rest of the domains). For every downscaled grid point and TROPOMI swath, TROPOMI pixels located within approximately ± 2 pixel width from the downscaled grid point were selected, and the distances between the downscaled grid point and the selected TROPOMI pixels' center were calculated. Using the inverse of those distances as a weight, the weighted average was taken as a NO_2 VCD value for a single downscaled grid point in a typical TROPOMI swath. Finally, the average of the NO_2 VCD values corresponding to each downscaled grid point were taken and has been visualized as shown in the figures. It is noteworthy that all TROPOMI NO_2 VCD swath data were screened with quality flags (qa_value > 0.75).

The details of the downscaling methodology (as shown above) were omitted in the manuscript because those figures (i.e., Figs. 4, 5a, 6a, and 7a) were intended to show general NO_2 VCD distribution in each respective target domain at a similar time of year, without elaborating with detailed discussions. The figures you have pointed out may appear peculiar, attributed to the limited availability of TROPOMI swath data over the target domains in October and November. We now perceive that some readers might find figures puzzling. Consequently, we have added a few more details on the revised manuscript (Fig. 4 caption).

L575: "Before comparing the collected set of collocated HIS and TROPOMI NO_2 VCDs, bias offsets were incorporated into the HIS NO_2 VCDs"—How well do HIS and TROPOMI agree in general, before bias are removed? I am interested in inherent retrieval bias as well as the representativeness of the TROPOMI footprint. In the Abstract and Summary, it is stated that 0.106 DU is the absolute error of the measurement, in comparison to TROPOMI. Is that true? It seems like some bias has already been removed.

→ The NO_2 VCDs retrieved from the airborne HIS observations and using the MWP method have inherent limitations in estimating the absolute magnitude of NO_2 VCD. To be more specific, in the spectral SF calibration using the clean pixel data, we need to estimate the NO_2 VCD of the clean pixel from external data sources (i.e., CTM, satellite, etc.). That is why we tried to select the clean pixels in the upwind region of the major NOx sources to minimize the discrepancy or the bias attributed to CTM simulated outputs or the satellite (i.e., TROPOMI) data. Nevertheless, the "existing" bias between assumed NO_2 VCD in the clean pixel to the real-world NO_2 VCD at the particular time and scene directly poses a bias to all of the HIS NO_2 VCDs calibrated with particular spectral SF (or the clean pixel data). To minimize this bias, we have further adjusted the NO_2 VCD assumed for the clean pixel for comparison with the TROPOMI data.

After all these efforts to minimize an "universal" bias relevant to the clean pixel NO_2 VCD assumption, we still found that HIS NO_2 VCDs and TROPOMI showing mean bias in different signs by flight-to-flight. We aimed to show the increasing intra-pixel NO_2 VCD variability according to the satellite NO_2 VCD increment in Fig. 8, whereas simply combining all the HIS data containing the mean bias attributed from the clean pixel NO_2 VCD assumption will increase the overall MAE (Mean Absolute Error) or the RMSE (Root Mean Square Error). Therefore, we have removed the mean bias of the HIS NO_2 VCD compared to the collocated TROPOMI data per research flight.

The MAE value of 0.106 DU presented in the abstract and the summary is the value representing the spread of the HIS data compared to the TROPOMI data after removing the mean bias. The bias removal would indeed have reduced the MAE by its value. However, we presented the MAE value not to emphasize the accuracy of HIS-driven NO_2 VCDs, but to show the increasing intra-pixel variability within a satellite pixel in the vicinity of industrial point sources. We are aware that the MAE value, or what Fig. 8 exhibits, does not essentially show the sole effect of the satellite intra-pixel variabilities, and we have tried to explain it to a certain extent considering the possible other uncontrolled factors (i.e., retrieval uncertainties) affecting the HIS-TROPOMI intercomparison.

Technical Corrections

The English usage is generally good. I include several suggestions that I noted when reviewing the paper, along with technical comments, below.

→ Thank you for your valuable advice and suggestions. We sincerely appreciate your comments and have made modifications based on your recommendations, unless specifically mentioned otherwise in the comments below.

Line 12:

"The high spatial"—remove "The"

"(VCDs) were measured"—replace "measured" with "retrieved" (VCDs are not measured, strictly speaking)

"from the airborne"—remove "the"

→ We have changed them in our revised manuscript (line 12).

L24: "The typical"—remove "The"

→ We have changed in our revised manuscript (line 24).

L29: “different observation geometries under complex vertical wind fields”—The winds don’t change the geometry, per se. Better to say something like “different pollution distributions...”

→ Our intention was to highlight that differences in the observation geometries of satellite and airborne instruments can lead to discrepancies in the spatial distribution of retrieved NO_2 VCD, consequently limiting the correlation between spatially collocated satellite and airborne observations. For instance, when the horizontal wind field varies below and above the flying altitude of the aircraft, the plume will be advected in different directions below and above the aircraft according to the respective wind field, yielding different NO_2 VCD distribution patterns observed from the satellite and the aircraft. We have made slight modifications on the revised manuscript to make this statement more straightforward (lines 28–29)

L37: “pollutant”—change to “pollutants”

→ We have changed in our revised manuscript (line 38).

L65:

”calibrations”—change to “calibration”

”retain”—change to “maintain”

→ We have changed in our revised manuscript (line 66).

L75: “East Asia, where the”—remove “the”

→ We have changed in our revised manuscript (line 76).

L78: “calibrations”—change to “calibration”

→ We have changed in our revised manuscript (line 79).

L91: “latest (spectral rows)”—change to “spectral rows at the edges”

→ We have changed in our revised manuscript (line 93).

L92: “grating is used with a concave mirror”—Is it a concave grating, or is there a mirror in addition? Please clarify. Also, no where do you say if HIS uses reflective or transmissive optics. I am assuming reflective.

→ Yes, HIS uses reflective optics as you presume, and has a concave mirror in addition to the diffraction grating. The details about the instrument are available in “HIS_Headwall_specifications.pdf”, which has been added in a supplementary material repository at https://doi.org/10.7910/DVN/YCZ9JU.

L100: “Unlike the”—change to “Despite these”

→ We have changed in our revised manuscript (line 101).

L122: “DOAS fitting”—change to “the DOAS fitting”

→ We have changed in our revised manuscript (line 123).

L127: “compartments”—change to “components”

→ We have changed in our revised manuscript (line 129).

L152: “following the convention”—remove; this is confusing and unnecessary

→ We have changed in our revised manuscript (line 159).

L164: “condition”—change to “conditions”

→ We have changed in our revised manuscript (Fig. 2 caption).

L175: three dots before “VCD” (mathematical symbol meaning “because”)—remove; not necessary

→ We have changed in our revised manuscript (Eq. 1).

L240: “The UVSPEC”—remove “The”

→ We have changed in our revised manuscript (line 250).

Table 3 headings: “Unit”—change to “Units”

→ We have changed in our revised manuscript (Table 3 caption).

Table 3 Solar Zenith Angle and Flight Altitude: Are variable, but would you report the range used? (the column has the other ranges)

→ The SZA and ALT entries for the LUT were adjusted for each research flight to avoid redundant computation, considering the possible range of SZA (i.e., calculated based on local time, latitude, and day of the year) and recorded range of ALT (supplementary Fig. S7). Therefore, SZA and ALT entries, as well as their range, vary by flight-to-flight. We thought elaborating on the details of LUT entries could be unnecessary, but we acknowledge your comment and have tried to add some valuable remarks in Table 3 of the revised manuscript. Explanations of the additional statements that have been added to Table 3 are as follows:

SZA: SZA entries were in 3° intervals for all the flights, and ranged from 42° (for the research flight at 17 October 2020) to 75° (for the research flight at 5 November 2020 and 24 November 2022; observations under SZA > 70° were filtered out for the retrieval). Number of SZA entries varied by flight-to-flight mainly attributed to the total duration of each research flight.

ALT: The number of ALT entries varied from 4 (on 3 November 2020, 5 November 2020, and 25 November 2022) to 9 (on 24 November 2022), primarily influenced by the number of cruising altitudes during each research flight. For instance, research flights with only four ALT entries were those conducted at a single predetermined altitude without any issues during the flight. On the other hand, cruising altitude changed during the flight on 24 November 2022 due to an unexpected order from the air traffic control, resulting in the enlarged number of ALT entries. We tried to make sure ALT entries were in 100 m intervals for ± 100 m range from the cruising altitude, while some additional ALT entries were considered in between the multiple cruising altitudes (i.e., if the cruising altitudes of the research flight were 1.5 km and 3.4 km a.m.s.l., ALT entries were determined as [1.4, 1.5, 1.6, 3.3, 3.4, 3.5 + α] km a.m.s.l.).

L253: "legitimate"—change to "realistic"

→ We have changed in our revised manuscript (line 263).

L258: "23 vertical grid"—23 layers over what altitude/pressure range? 70 hPa top noted in L274, but it should be stated here. Same comment for L266.

→ For the CMAQ model, 23 vertical layers were determined from the terrain height (surface) to 70 hPa, whereas 31 vertical layers for WRF model (apologies that the original manuscript was misleading–made the correction in the revised manuscript line 276) ranged from surface to 50 hPa with hydrostatic pressure coordinates. Moreover, MCIP (Meteorology-Chemistry Interface Processor) module was used to apply WRF-simulated outputs to CMAQ. We have made modifications according to your comment and added some descriptions regarding the vertical range of CMAQ/WRF models in the revised manuscript (lines 268–269 and 276).

L292: "inferring"—change to "implying"

→ We have changed in our revised manuscript (line 303).

L315: "should be calibrated"—change to "was calibrated"

→ We have changed in our revised manuscript (line 326).

L329: "clean pixel"—define what this means when it's first used (here)

→ The description of "clean pixel" has been expatiated on the following paragraph (please refer to lines 346–349 in the revised manuscript), and we believe that the current narrative structure is efficiently conveying the detailed description of the "clean pixel". We kindly ask for your consent to keep the corresponding paragraphs as it was in the original manuscript.

L375: "premises"—change to "circumstances"

→ We have changed in our revised manuscript (line 386).

L400: "Assuming a SNR"—change to "Assuming an SNR"

→ We have changed in our revised manuscript (line 411).

L586: "premises—change to "circumstances"

→ We have changed in our revised manuscript (line 613).

L609: "As an extent"—do you mean "To a certain extent"?

→ What we intended was closer to such expressions as "accordingly" or "moreover", rather than "to a certain extent". Therefore, we have changed the phrase in our revised manuscript (line 637).

L623: "volatile"—change to "unstable"

→ We have changed in our revised manuscript (line 651).

L630: "succeeding"—change to "principal"

→ We intend to express that the uncertainties in the spectral shift calibrations and the instrument noise were a second-major group contributing to the total uncertainties, succeeding the primary source of uncertainty, the uncertainty in spectral scale factor calibration. Therefore, we believe that the word "succeeding" better represents our intention than the “principal”; hence, we decided to keep the sentence as it was in the original manuscript.

Again, we appreciate for your valuable comments and suggestions.

===End of Document===