Review of: Assessing sub-grid variability within satellite pixels using airborne mapping spectrometer measurements (Tang et al., 2021)

The manuscript discusses sub-grid variability (SGV) within satellite pixels using high resolution airborne GeoTASO observations, acquired over three urban areas. A quantitative way to assess SGV for different hypothetical satellite pixel sizes is presented based on two methods: random sampling and spatial structure function. Also temporal variability in satellite pixels has been studied for different sizes. The scientific content of the paper fits well within the scope of AMT. However, major revisions (detailed below) need to be conducted in the paper before publication.

General comments:

-I repeat the comment from the quick access report that the whole document should be carefully scanned by the main author and some co-authors to remove unclarities and confusion. It is not always written in a well-thought concise way, hiding sometimes the key messages. Also, the authors refer often to "this error", "that method", etc. while it is not always clear to what is referred exactly. This effort would certainly bring the manuscript to a higher level.

-SGV is obviously strongly linked to effective variability/heterogeneity in the NO_2 field. It is shown that SGV is similar in three different studied regions. It is also stated that the method is generally applicable to quantify SGV and that for example a LUT could be used (e.g. p.3, l.94 and l.98). I doubt the latter. Even if the study is applied on three regions and even if you take 10.000 pixels randomly in this area, all of them are strongly urbanized which is reflected in the high mean NO_2 VC for all three areas. It should be clearly stated (in introduction and conclusion, maybe also by adapting the title) that results are valid for urban regions and are not proven for background areas that are characterized by much less heterogeneity. For such areas we can expect a lower impact of SGV. You could also specify that the SGV studied here can be seen as an upper limit, in the same way as you discussed for other species when compared to NO_2.

-The introduction doesn't mention anything about your study of temporal variability, while this is an important part in the further discussion and conclusion. Please write a few lines in the introduction and clarify why it is also relevant or an added-value to study the temporal variability in addition to SGV.

-p.2 l.66: "Until recently, accurate in-situ measurements with sufficient spatiotemporal coverage have not been available"→ I don't agree with this statement. For example, in the US and Europe there is a dense network of in-situ stations for quite a long time. The key message in this paragraph should be the difficulty to use in-situ observations for satellite validation and study of SGV or difficulty to compare a point measurement with gridded data in general and associated representativeness error. Please elaborate on this.

-p.2 l.78: Linked to the previous remark, start this paragraph by discussing the added-value of airborne mapping observations <u>in general</u> for SGV and satellite validation. There are other data sets used in other studies for assessing SGV (note the references made a bit earlier). Then, continue discussing the performed GEOTASO measurements used in this study.

-p.7 l. 268: Please reconsider if this paragraph is really needed here. It is rather confusing. Especially as temporal variability is discussed in the next section.

-Sect.4, paragraph 2: This paragraph is very confusing. It should be rewritten in a well thought and concise way. Comparison of satellite retrievals with aircraft CO vertical profiles are mentioned as motivation of

this work. Than it is mentioned that a same problem arises for NO_2 while you are mainly referring to satellite validation papers based on airborne data. Airborne mapping observations, able to cover full satellite pixels, are exactly a type of measurement that minimize the spatial representativeness error, this in contrast to in-situ, ground-based observations, vertical soundings, etc. Then, the next sentence starts to discuss other species and that GEOTASO is able to address 'this' problem (which problem?). GEOTASO was also used in the Nowlan et al. and Judd et al. references that were just mentioned and identified as: difficult to address 'this issue'.

-Sect. 3.2: Nothing is mentioned for conditions where Dt is larger than 4 hours. The case seems to be inverted from Dt 4 to 8 hours according to Figure 6 (TeMD increases with decreasing spatial resolution). Please add an explanation to the paper.

-Sect. 4: Please consider a discussion on the impact of your findings (both spatial and temporal variability) on top-down emission estimations from satellite observations at higher spatial resolution (e.g. TROPOMI, GEMS) and higher temporal resolution (e.g. GEMS, TEMPO, S-4), when compared to emission estimations from e.g. OMI?

Minor comments:

-p.3 l.85: 'As such, the GeoTASO data…' → Mention the GeoTASO spatial resolution already here.

-p.4 l.132: 'The dense sampling of the GeoTASO datasets is a unique feature' → There are quite a lot of other imaging systems that can obtain a similar or better spatial resolution. Please generalize this statement to overall airborne mapping observations.

-p.4 l.136: Which uncertainty estimate are you referring to?

-Sect. 2.1: It would help the reader to indicate a typical flight time to acquire one grid map already here in Sect. 2.1 in order to understand the temporal variability issue mentioned later on.

-p.6 l.222: In principle you still scan an area on the ground, so specify that you take the center lat/long of the pixels…or do I misunderstand this approach?

-p.8 l.297: You focused on GEO missions so far, so please mention as well S-4.

-p.8 l. 306: Not clear what you mean with "We also tested the results for sampling satellite pixels by raster instead of within hourly bins." …and what the difference is with the previous statements.

-Figure 6: Please rewrite caption in a more clear and concise way.

Technical corrections:

-p.6 l.225: Adapt "Distance" to "D"

-p.9 l.342: "…NO_2's relatively short lifetime…" → …the relative short lifetime of NO_2…

-p.9 l.364: Correct the sentence.

-p.12 l. 458: HCHO has already been defined earlier.

-p.12 l. 468: Define ‘local observations’. Do you mean ground-based in-situ?

-p.13 l. 503: This statement is not valid for TeMD > 4 hours.