We would like to thank reviewer #2 for his/her reviews. We addressed each comment below and highlighted our answers in red, the referee's comments are black.

The paper presents the first analysis of TROPOMI ALH retrieval. OSSE is conducted to guide the analysis, which is excellent. This reviewer would recommend moderate revision of the paper to shed some light of surface albedo that may affect the ALH accuracy, and to provide somewhat closure between the past theoretical error analysis of ALH and the finding in the paper.

Datasets.
What is the data resolution and range of TROPOMI ALH? In other words, in the retrieval, is ALH data continuous or in discrete values at different pressure level? Can ALH be 0.5 km or lower?

The vertical data resolution is continuous. It ranges from 1050-75 hPa. There is a data field that is fixed to the surface, but the original dataset, which can extend below the surface, is also available in the detailed results field. So, yes, the ALH can be lower than 0.5 km.

We have included the following sentence in the manuscript, Sect. 2.1:
"The vertical data resolution is continuous and ranges from 1050-75 hPa. "

GEM-MATCH A few sentences describing how GEM-MACH estimate the injection height can be insightful. Is the fire radiative power used in CFFEPS?

FRP is not used in CFFEPS, instead fire energy is estimated based on modelled fuel consumed and the estimated heat of combustion of dry fuel. We included the following in the manuscript to describe how the CFFEPS (GEM-MACH) injection height is estimated, in Sect. 2.5 GEM-MACH:

"Fire plume injection height in GEM-MACH is parameterized in the CFFEPS module with hourly modelled meteorology as detailed in Chen et al. (2019). The injection height is determined based on the balance of estimated plume buoyancy and the modelled environmental lapse rate at fire location. Total heat flux from fire is determined from modelled fuel consumed per area and the heat of combustion of dry wood fuel (Byram, 1959). The fraction of energy that enters the plume for convection is further parameterized based on thermodynamic energy balance accounting for heat lost to fuel, moisture, radiation, conduction and incomplete combustion. The hourly plume injection height is determined based on the dry adiabatic equilibrium of the buoyant plume and the modelled environmental lapse rate at fire location. "

OSSE TROPOMI flume heights P8, L20. Using wavelength (instead of wavenumber) is suggested here. In addition, it is noted that TROPOMI uses the spectral fitting to derive ALH, not a simple ratio. In contrast, Xu et al. (2017, 2019, already cited in the manuscript) used the ratio.

As suggested, we changed the reported wavenumber to wavelengths throughout this paragraph.

Further, we have added the following sentence:
"Note that TROPOMI operational algorithm uses spectral fitting to retrieve AER_LH whereas a simple ratio has been used here, similar to Xu et al. (2017, 2019)."

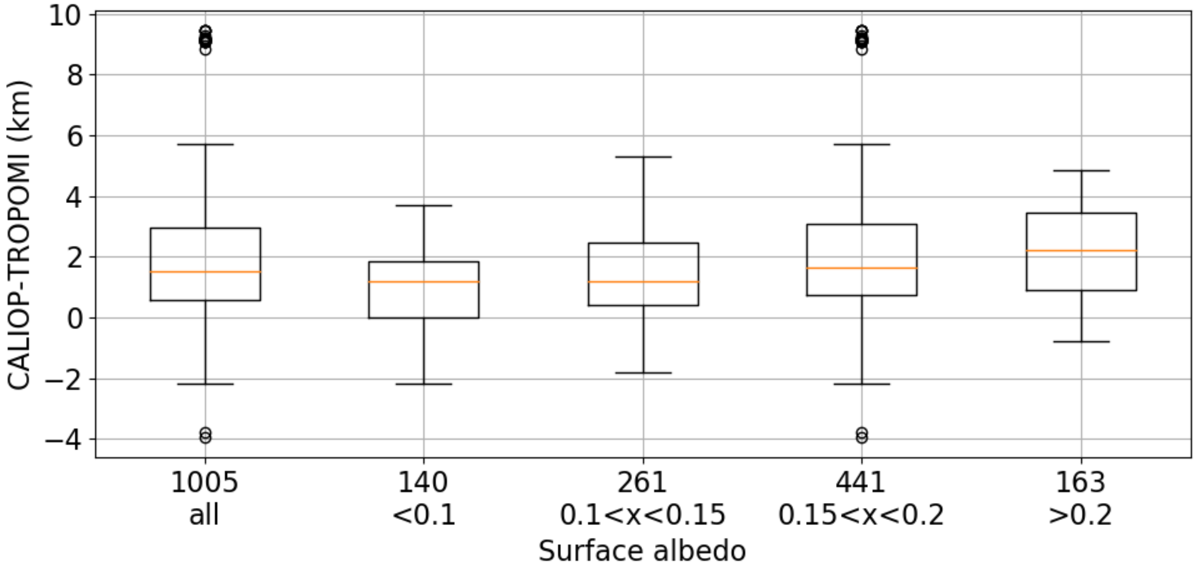
Section 4.1. Some discussion about the reasons for MISR vs. modeled plume height difference can be helpful. Note, most satellite-based fire products provide only pixel based FRP, where the plume rise model should use FRP over the fire area (not pixel area). This paper might be useful here to interpret the difference. Peterson et al., 2014, Quantifying the potential for high-altitude smoke injection in North American boreal forest using the standard MODIS fire products and sub-pixel-based methods, JGR.

We included further description on how the model plume injection height is estimated, FRP is not used to determine the injection height with CFFEPS (inside the GEM-MACH model). See the comment above.

Section 4.2 and results: There are multiple times, 'thin' layer is mentioned. How the thin layer is defined? By optical depth or geometric thickness? Past work has shown that O2-A type of ALH retrieval should be sensitive the high aerosol plumes provided a moderate value of AOD. The appendix in Xu et al. (2019) provides the ALH error estimates for different AOD and different ALH. It shows that retrieval is most sensitive to ALH at 2 km, and should be good to provide ALH from 1 – 8 km with retrieval error of less than .5km for AOD of 0.4 over dark surfaces. Anyhow, these past analyses should be helpful to interpret the physics behind the finding here. Afterall, past work have done several case studies to evaluate the ALH retrieved from O2 band (although not from TROPOMI). It is shown that the retrieval error can be affected not only AOD and ALH, but also by surface albedo. It might be interesting to stratify the ALH differences by surface albedo. As surface albedo increases, the ALH retrieval error can be large. Some comparison and contrasting of the results here with the results in the literature can be more revealing.

Here, we mean geometrical thickness. We have added the word "geometrical" in front of thickness throughout the text. Also, p. 12 l.26 describes how the geometrical thickness is defined.

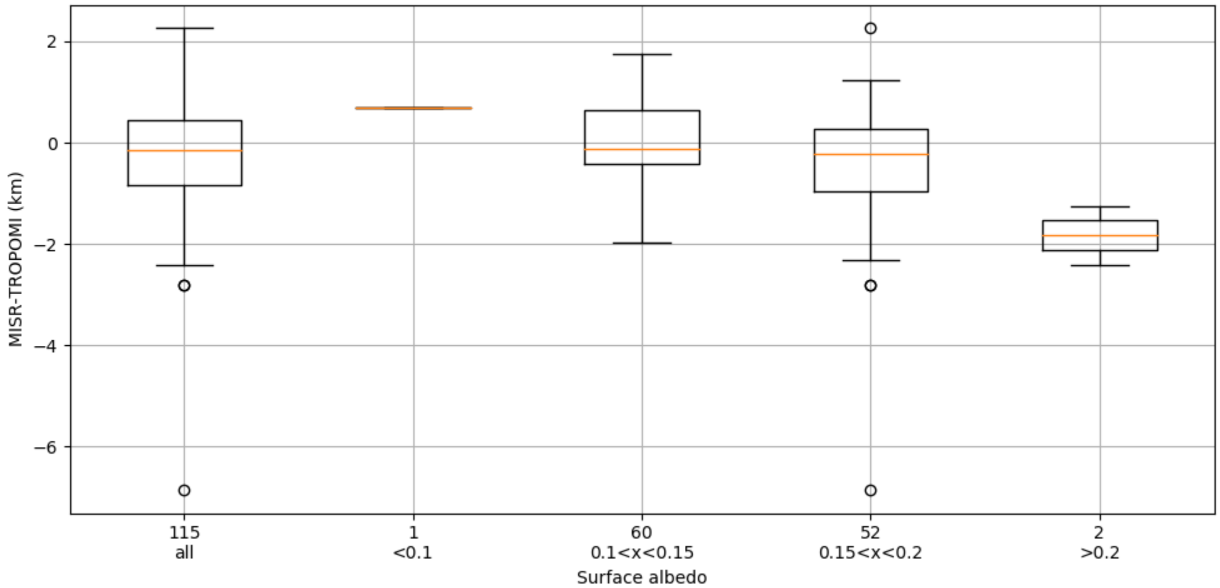
We have looked into the differences based on surface albedo and we have included Fig. 5e to the manuscript.



Showing that the differences between Caliop and TROPOMI increase with an increasing surface albedo.
We have added the following text to the manuscript, p.13, l. 19ff:
"Figure 4e shows that the differences between CALIOP and TROPOMI increase with increasing surface albedo, consistent with the idea that the TROPOMI retrieval algorithm is more sensitive over dark surfaces and possess smaller uncertainties (Sanders and de Haan, 2016; Xu et al., 2019)."

And in the conclusions:
From:
"… TROPOMI aerosol layer heights are more accurate for thicker plumes: the difference between the CALIOP and TROPOMI mid-plume height decreases and the correlation increases with increasing thickness of the plume and for a 3 km thick plume the average difference is only about 50 m."
To:
"…TROPOMI aerosol layer heights are more accurate for thicker plumes and over darker surfaces. As such, the difference between the CALIOP and TROPOMI mid-plume height decreases and the correlation increases with increasing thickness of the plume and for a 3 km thick plume the average difference is only about 50 m. Further, the differences between Caliop and TROPOMI increase with increasing surface albedo."

We can see something similar for MISR vs TROPOMI, however, not as many observations are available for the analysis, and thus we did not add it to the manuscript.

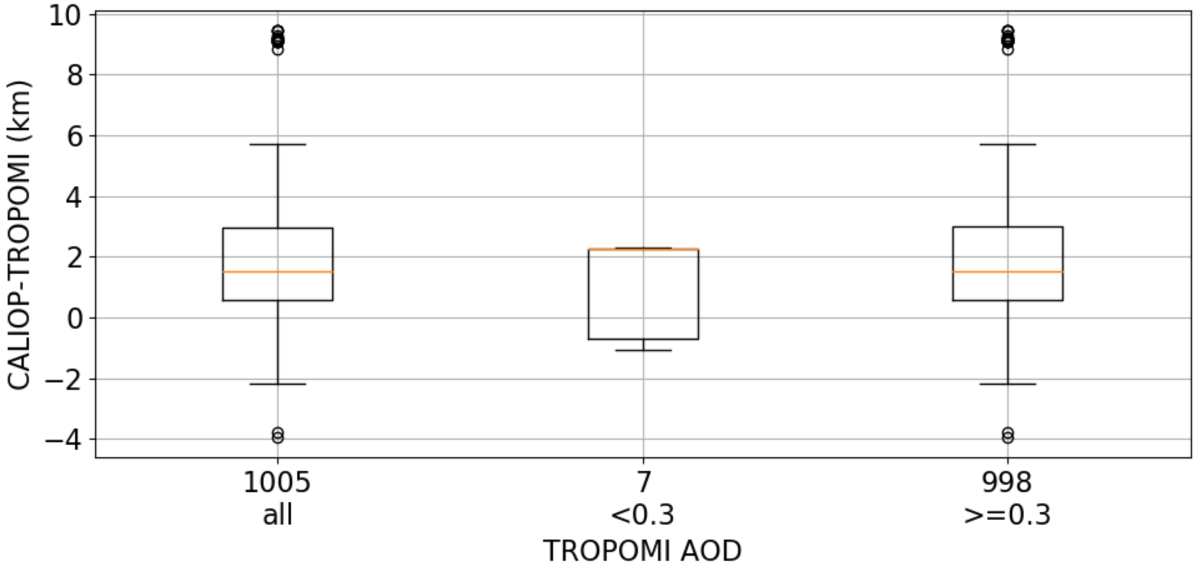


MISR mean plume height vs TROPOMI mean plume height. The differences are increasing for an increasing surface albedo. However, not as many observations are available and only 1-2 plumes have an albedo below 0.1 and above 0.2.

Summary L25-28, P15. Is your finding from the real data more or less consistent with the theoretical error analysis in Xu et al. (2019)?

We see that the differences between the instruments are larger for larger surface albedo, as mentioned in Xu et al., 2019, the error seems to increase with increasing surface albedo. Further, we looked into the AOD and found that we only have 7 cases (out of over 1000) that have AOD < 0.3. The differences for those are higher, however, it is only 7.

In terms of the surface albedo see the previous comment, as suggested by Xu et al., we can see that the TROPOMI plume height is closer to the plume height from MISR or Caliop for darker surfaces.



L5-15, P16. Again, surface albedo is briefly mentioned and discussed here. It might be nice to sort the ALH evaluation by surface albedo. In addition, it is worth mentioning that for thick plumes at the surface, ALH retrieval is expected to have large errors. The analysis presented in the papers show the retrieved ALH is at least 1 km above the surface. There are also cases where TROPOMI ALH is consistent with CALIOP for high and think plumes (Fig. 4b). In other words, in both abstracts and conclusion, it is worth mentioning that the TROPOMI ALH has some success in retrieving high plumes up to 8 km (in addition to that the most accurate retrievals are for thickness plumes from 1-4.5 km).

We have changed the following sentence in the abstract:

From:

"…our results show that the TROPOMI aerosol layer height is more accurate for thicker plumes and plumes below approximately 4.5 km."

To:

"… our results show that the TROPOMI aerosol layer height is more accurate for over dark surfaces, for thicker plumes and plumes between approximately 1-4.5 km."

We have changed the following sentence in the conclusions:

From:

"The TROPOMI plume heights seems more accurate for thicker and lower plumes plumes (<4.5 km altitude)."

To:

“The TROPOMI aerosl layer height seems to be successful in retrieving high plumes up to 8 km, the uncertainties seem reduced for thicker and lower plumes between 1-4.5 km altitude, as well as dark surfaces.”