This manuscript documents an OSSE-type (Observation System Simulation Experiment) study of how the high-spatial resolution spectroscopy observation of total column water vapor from satellite observations should be sampled to understand the horizontal variability and structure of water vapor in the planetary boundary layer (PBL).

Topic of this study is important and suitable for the AMT. However, the manuscript suffers from several major issues and significant flaws as pointed out below. Its methodology (i.e., using a simple emulator instead of full simulator) is not justified and has serious potential problems. No causes and underlying physics are provided for the "solar-smearing bias", which is a key finding of this study. Even though the methodology is problematic, and the results are not explained, the authors still try to propose a universal "new sampling strategy" to the current and future high-resolution spectroscopy sensors. This is overreaching the say the least and could be misleading.

Based on these considerations, I strongly recommend rejection of this manuscript. It this study were published, the "emulator method", the "solar-smearing bias", and "new sampling strategy" could be cited again and again as if they were correct. But they are not, at least not justified by this study.

We thank the reviewer for their rigorous approach, which made it obvious that we had not been explicit enough about important details. Some other readers could clearly be confused by our original submission, so we have added two new figures and text to hopefully avoid this confusion. We believe that all reviewer concerns are now addressed.

Below we reference page and line numbers, which refer to those in the *red lined/track changes* version of the manuscript.

Major problems:

- The first major problem of the manuscript is the lack of important details on the methodology and the discussions are often too short and unsatisfying.
    - Although the concept of "total column water vapour" (TCWV) appears to be simple, the retrieval process can be quite complicated and involves many technical details, especially at high-spatial resolution. For example, when water vapor has both strong horizontal variation and vertical gradient, the solar-viewing geometry will become important because the path-integrated water vapor can be significantly different from the TCWV, depending on how instrument geolocation/collocation is done. In such situation, observations from different angles need to be de-convoluted to re-construct the horizontal and vertical structure of water vapor. The manuscript briefly mentioned this issue in section 2.2 and 3.1 but the discussion is far from clear or satisfying. For example, it is mentioned "TCWVret from input TCWV, which is in fact the integrated water path along the solar path". But how is the "path-integrated water path" converted back to the TCMV (only times a cosine factor?)? Is the definition of TCWV dependent on solar and/or viewing angle? Although Figure 2 provides some information on the vertical variation of water vapor of the cases used in this study, the corresponding discussion in Section 3.1 is so brief (only one sentence) and obscured that it only raises more questions than answers. In particular, it is hard to tell how the author could "confirm that our derived values are indeed representative of bulk PBL statistics" from the

figure, when there seems to be significant vertical variation of epsilon in the PBL.

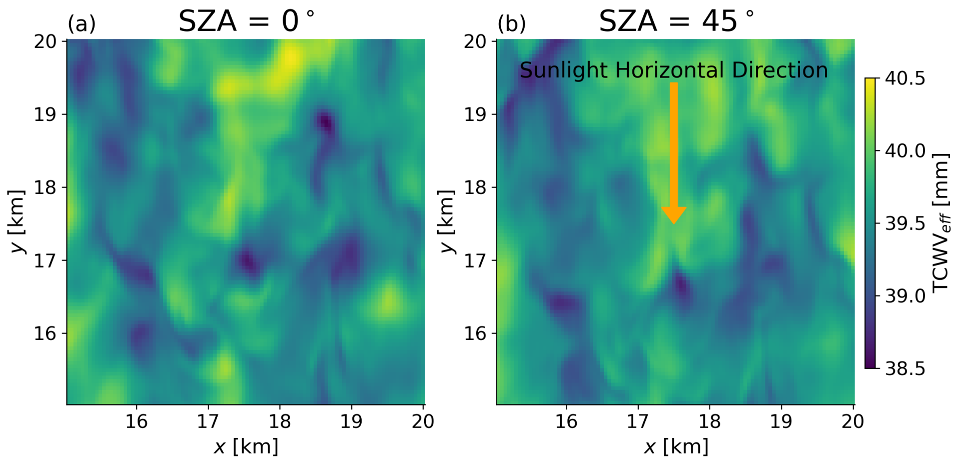
Here is where we realised we explained some of our most important method details very poorly.

We have now added detail in Section 1 on p3L30 onwards and Figure 1. We added equations and description that relate path integrated water vapour (PIWV), real TCWV, and the reported "TCWV" retrieved by VSWIR instrument. We use TCWV terminology for consistency with other VSWIR work, citing 6 papers that call their retrievals "TCWV", although we use subscripts to differentiate properties. In particular, we refer to "effective" TCWV:

$$TCWV_{eff} = \frac{PIWV}{\frac{1}{\mu} + \frac{1}{\mu_0}}$$

Which is the TCWV that would provide the same PIWV given the solar/view geometry. The PIWV is determined from tracing the solar ray through the atmosphere, and our retrieved $TCWV_{ret}$ are estimates of this $TCWV_{eff}$.

Figure 1 compares TCWV and PIWV when SZA=45° in a small part of an LES snapshot domain. An arrow indicates the horizontal component of the solar path; it is visibly obvious that the IWV field is like the TCWV field but "smeared" or "smoothed" in the horizontal. This is simply the result of the solar downward path being diagonal, rather than vertical.



All current imaging spectrometer retrievals of atmospheric water vapor ignore this effect when reporting the TCWV; in other words, they do not perform the tomographic reconstruction that the reviewer rightly calls for. To our knowledge, our study is the first that attempts to account for these effects with this class of instruments.

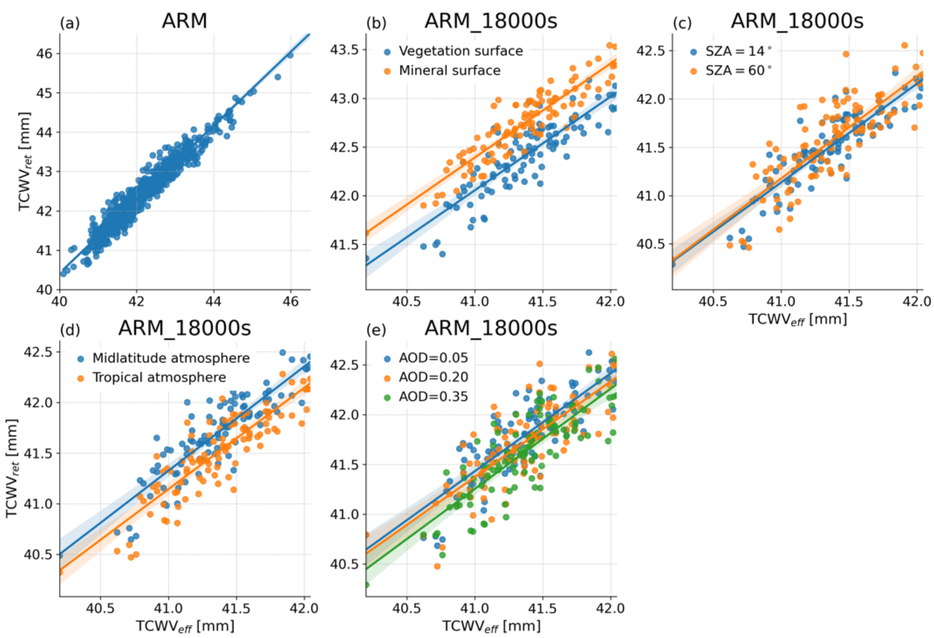
Please see response to next comment for details on how the path-dependence is included in the radiative transfer calculations behind our emulator, and how it is also accounted for in the emulator inputs.

We also changed "confirm that our derived values are indeed representative of bulk PBL statistics" to "…are indeed representative of ζ_2 derived from $PCWV_{PBL}$" since we retrieve the value derived from bulk PBL water vapour, not the average of exponents calculated at higher vertical resolution.

- Some other technical details are also missing. For example, how cloud mask is applied? Is it dependent on the sun-viewing geometry? If cloud mask is independent of sun-viewing geometry then there is apparently an inconsistency between the use of path-integrated TCWV and use of path independent cloud mask. Is the 3-D radiative transfer considered in the simulation or emulation? Previous studies have noted the "halo effects" of cloud in the so-called twilight zone. How are these 3-D effects of cloud treated in the study? Are they simply ignored (i.e., using 1-D RT model), or removed by cloud masking (then how?) or considered in the simulation?

We have split Section 2.2 into two sections. Subsection 2.2.1 is almost entirely new text which describes the OSSE approach: we used 1D RT (MODTRAN) and an optimal estimation inverse method (ISOFIT) to derive the emulator, for which we found Eq. (4) (now Eq. 8) was an adequate representation.

Subsection 2.2.1 also refers to a new Figure 2, which demonstrates the linear relationship we assert for the emulator, and shows how the parameters may change with surface, SZA, retrieval-assumed $q(z)/T(z)$ and AOD.



Subsection 2.2.2 now expands on the ray tracing we used to generate and explicitly states that we use the $TCWV_{eff}$ derived from ray-traced PIWV through the 3-D LES field as input to our emulator. We hope that, in combination with the additional detail in Subsection 2.2.1, this is now clearer. Given this context we believe readers can now understand how our cloud/shadow mask is generated in a completely analogous way: "The same ray-traced calculation is repeated with cloud water q_c to obtain cloud water path (CWP). Footprints are then flagged as cloudy or shaded when $CWP>1\times10^{-3}$ mm…".

We agree with the reviewer that 3-D radiative effects could be very important. These would be implicitly addressed by our suggested airborne experiment but we were remiss in not specifically mentioning it. As noted in the new 2.2.1 text, we did not have the computational resources for 3-D RT across all of our desired cases, especially given our very high spectral resolution requirements. Section 4 now mentions that 3-D RT forward modelling is a good way to improve this: we reference papers behind SHDOM and MYSTIC here.

- The use of a very simple retrieval emulator is not justified and raises many questions.
    - OSSE type of studies often use a "retrieval simulator" consisting of a "forward" RT simulator and an "inverse" retrieval simulator. The simulator should be as "realistic" as possible in comparison with the real retrieval to faithfully capture the influences of various factors on the retrieval. In contrast, this study only uses a seemingly naïve retrieval "emulator" (i.e., equation 4) and the only reason to justify this is "due to computational constraints". This "emulator" skips both the RT simulation process and the retrieval simulation step, and directly connects the retrieval to the input fields in a very simple way (linear). There is no discussion on the accuracy of this emulator in comparison with the "full OSSE simulator" if there is one. As a result, it is unclear if the artifacts in the "retrieval" is meaningful or simply due to the inadequacy of the emulator. It is also hard to imagine what kind of "computational constraints" the authors are referring to. This is a case study based on a handful of LES scenes. Many previous studies have performed full RT simulations, even 3-D RT simulations, based on LES scenes. How and why is the RT or retrieval simulation in this study so computationally expensive?

Please see responses above. The reviewer describes what we believe to be the "correct" way, which is indeed what we did in Richardson et al. (2021). The new 2.2.1 text describes the previous OSSE from which Eq. 8 (originally Eq. 4) is derived, and the new Figure 2 (above) shows some evidence that should persuade readers to provisionally accept the linearity between $TCWV_{eff}$ and $TCWV_{ret}$. More details are in our previous publication.

- The solar-geometry dependent retrieval bias in section 3.3 is interesting. However, I tried hard to find some explanation of the causes and underlying physics but didn't find any. There is neither any reference to previous studies or discussion on whether this phenomenon had been discovered before or completely new. **The authors didn't even bother explaining why this bias is called "solar-smearing" effect.** The word "smear" only occurred twice in the manuscript, one in the title and the other in the conclusion.

Again our failure to sufficiently link back to Richardson et al. (2021) caused confusion. The new text in Section 1 mentioned above describes the physical principle, namely the solar path through a 3-D field and explains why we pick the term:

"The **Error! Reference source not found.**(a) to **Error! Reference source not found.**(b) differences show a smoothing or smearing in the *y* direction, so we refer to these solar-geometry induced changes as the "solar smearing" effect." (figure 1 is the first shown in this review response)

- Event though the “solar-smearing” effect is completely unexplained (and is based on highly questionable methodology), the authors still recommended the “new sampling strategy” to many current and future sensors. This totally unacceptable to me.

The reviewer was right to be cautious given the lack of clarity in our original submission, but we are convinced that the concerns you rightly raised are now addressed. After all, with the exception of 3-D radiative transfer, we actually performed the calculations in ways that you proposed (an OSSE with forward an inverse models) and directly accounted for issues you raised (the complex 3-D structure of the water vapour field). And regarding 3-D radiative transfer, this was a limitation of our available tools and computational resources but it is a very good suggestion which we now mention in the discussion & conclusions.