This well written manuscript deals with future spaceborne imaging spectrometers expected to measure water vapour columns with horizontal resolutions of < 100 m. The authors simulate biases in water vapor scaling statistics that will occur at high solar zenith angles due to a solar light path traversing neighboring pixels. To reduce the biases, the authors propose a sampling strategy perpendicular to the solar azimuth angle. This is evident, and the described bias reduction is what one would expect. The merit of this study, which fits very well to AMT, is a quantification of the expected biases in water vapor scaling statistics. The study still lacks details on assumed measurement uncertainties, see specific comments.

Thanks for taking the time to read and think about our paper. We have added the requested details on uncertainty and additional tests on the effect of AOD.

The main text includes new Figure 2(e) and Figure 5(e) panels showing the AOD results, plus text discussing those, and in the discussion & conclusions. We believe these additions are demonstrate our AOD-relevant conclusions without unnecessarily lengthening the paper. We include extra details in this review response to help the reviewer(s) judge our methodology.

A lot of text and 2 figures were added to respond to reviewer 2 – we kept your comments in mind and responded to them where possible in this added content.

Specific Comments:

1. Spatially nonuniform aerosol distributions (as stated in the abstract) are in my opinion not enough addressed. They probably pose the highest challenges to spectroscopy. On the other hand, they may be difficult to assess, and the resulting biases difficult to quantify. It would nevertheless be of high merit to include them in your model framework and to show some related simulation results in section 2.

Tucked away in Supplementary Figure 7 of our last paper we showed that $TCWV_{ret}$ isn't that sensitive to AOD, and our emulators were developed with randomised AOD. We anticipated that with typically small(ish) horizontal gradients in AOD over <1 km there would be minor effects on our results from AOD.

However, this paper really should demonstrate this rather than state it, so we added AOD results in the new Figure 2 and added an AOD-results panel to Figure 5.

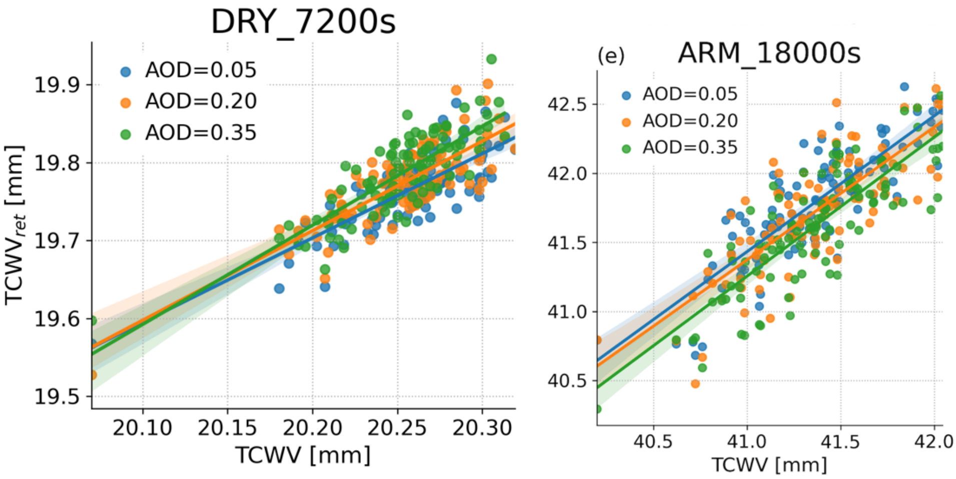
Our approach was to generate a "very very bad case" with large AOD gradients of order 0.3 $km^{-1}$, and show that it has a small effect on estimated $\zeta_2$. The paper shows the ARM_18000s example, it represents the 22 out of 23 cases where the effect on exponents is small. We expand on this here in case the reviewers are interested. If you only care about further summary of main-text changes, see bolded paragraph at the end.

**Additional detail for reviewers**

The original emulators were fit to forward & inverse simulations with randomised true AOD. We re-ran ARM_18000s and DRY_7200s forward and inverse simulations with profiles where AOD is fixed at 0.05, 0.20 or 0.35 and generated new emulators. We used these emulators *only* for the specific AOD sensitivity tests.

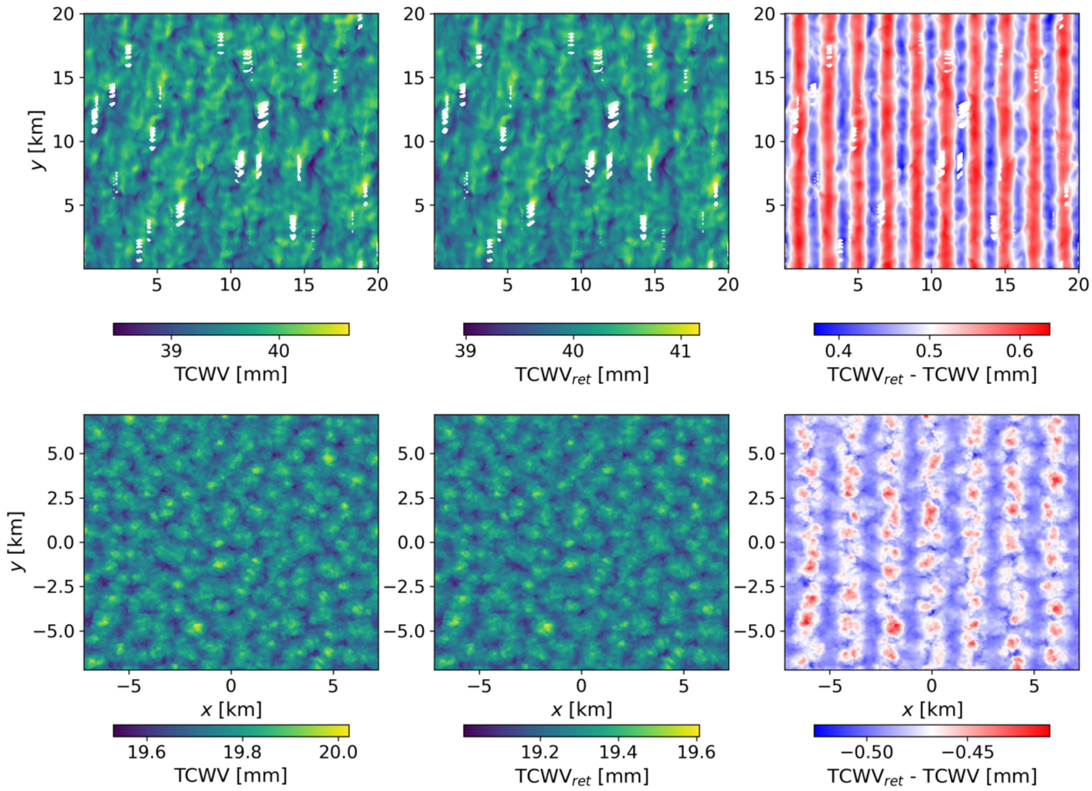
ARM_18000s saw a shifting mean bias of ~0.3 % in $TCWV_{ret}$ when AOD changes from 0.05—0.35. The changes are proportionally smaller for DRY_7200s but include changes in gradient. For our first test we modified our emulators to make the gradient and intercept into functions of AOD and fit them:

$$TCWV_{ret} = a_1(AOD)TCWV_{eff} + a_2(AOD) + \epsilon$$



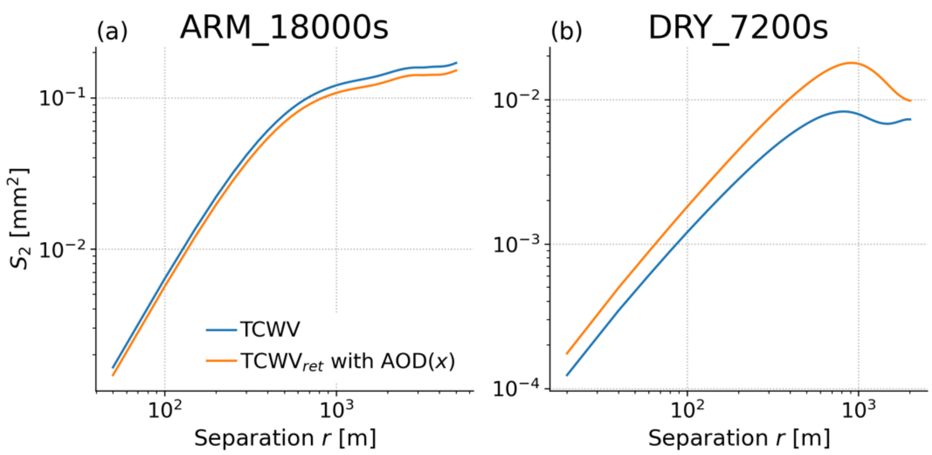

The panels above show the fits that provide sets of $a_1$ and $a_2$ values given AOD in [0.05, 0.20, 0.35 ]. For any AOD we simply linearly interpolate between those to provide our AOD-dependent emulator. The next issue was to decide what sort of spatial variability in AOD to test.

We decided on a rather extreme case: changes from 0.05—0.35 every 1 km, represented by a horizontally-varying sinusoid with period 2 km in either the $x$ or $y$ direction. Below is true TCVW, emulated TCWV and the difference between them for DRY_7200s and ARM_18000s at SZA=0°. The horizontal waviness from the spatial AOD structure is obvious in (c,f). We did not add random error ($\epsilon$) here, but the paper shows that we can identify and remove its effect on $S_2$. (ARM masked values are those that are cloudy or shaded at any of our selected SZAs).




Next, we address how $S_2$ and $\zeta_2$ respond, firstly in just these two snapshots. The figures below show they are offset: this is primarily because of $a_1$ – it scales $S_2$ as shown and discussed in main paper Figure 5(c). Note that panel (a) below differs from the new Figure 5(e) for reasons discussed below.

This DRY case is the outlier value with the lowest $\zeta_2$ of all snapshots, and it's clearly due to the "dip" near ~1 km separation suppressing the gradient. As discussed in the paper, we do not investigate the detailed dynamics of the LES cases here but are only interested in how well we can retrieve the property. In the ARM case the change in $\zeta_2$ is negligible (0.63 vs 0.63), but in DRY_7200s changes greatly from ~0.2 to ~0.4.

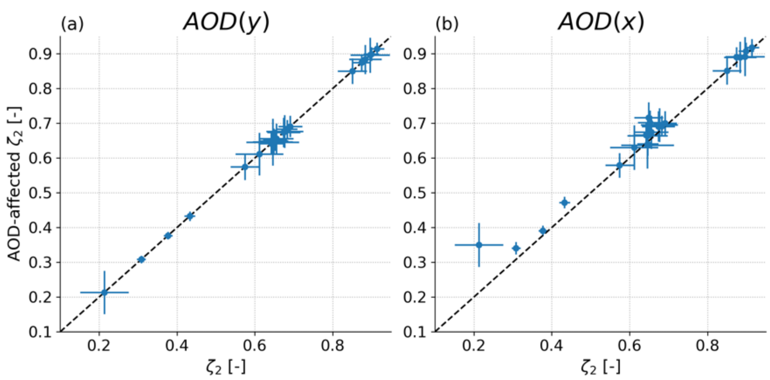


The DRY_7200s difference occurs because the true S2 structure shows a decrease in variance near ~1 km while our AOD-induced variability has a maximum effect at 1 km separation. We wanted to test all snapshots but didn't have time to process the forward and inverse simulations needed to generate individual emulators, so we used another approximation. In percentage terms, the ARM_18000s case shows the largest effect: a 0.1 % change in $TCWV_{ret}$ per 0.1 change in AOD, so we used that as the basis of our next test.

We added a simple treatment of aerosol to all emulated snapshots: a sinusoidal variation in $x$ or $y$ of $TCWV_{ret}$ with an amplitude of ±0.15 % of the field mean and a wavelength of 2 km. This is just like the test above, except there is no change in $a_1$. For Figure 5(e) the test uses $a_1 = 1$, so the $S_2$ lines lie atop each other rather than have the offset seen in the above figure.

This results in a change of 0.3 % in $TCWV_{ret}$ every 1 km to represent a change of ~0.3 in AOD over that distance. The next figure shows how the difference is negligible when calculated perpendicular to the aerosol variation (see panel a) but when calculated parallel to the aerosol variation (panel b) there are changes. The direction dependence is rather like our sensitivity to solar azimuth. From this figure the DRY_7200s case is the outlier of the set: changes in non-DRY snapshots are always <4 % in magnitude, for the DRY snapshots except DRY_7200s it is ~10 %.

One nice thing about ISOFIT is that we would have $TCWV_{ret}$ and retrieved AOD fields, so we could "back calculate" the likely effects of AOD on retrieved $\zeta_2$ and flag cases where it might be important. We have referred to this briefly in the new text.




**In summary: we hope this review response persuades the reviewer that we did our due diligence for AOD. In the main paper we have added the following:**

i. **Figure panel showing how $TCWV_{ret}$(TCWV) changes as a function of AOD in ARM_18000s**
ii. **Figure panel showing how $S_2$ changes with a strong spatial variation in AOD in ARM_18000s**
iii. **Text summarising changes in $\zeta_2$ are generally small even with relatively large AOD gradients, and noting that while it has a large effect in one snapshot, we could use the ISOFIT retrieved fields to flag cases where this is likely in practice. AOD-relevant text is in Sections 2.2, 3.2 and 4.**

**We think these changes strike the balance between being sufficient and brief.**

2. In section 2.2 you define parameters related to assumed measurement uncertainties and biases. Since they are used throughout the study, it would be good to describe them better here, perhaps including a figure which illustrates sensitivity (a1) and bias (a2). In addition, you should be more specific concerning the impact of aerosol layers (comment 1), and concerning probable error correlations between (for example) surface albedo and aerosol concentration variations. Finally, can you assess the impact of simulation idealizations and simplifications which you have likely undertaken?

Our completely restructured Section 2 aims to address these comments. The new Figure 2 shows how changes in individual properties (SZA, surface, AOD) change the response. Our new text emphasises that we propose retrieving only over mixed-vegetation or mixed urban-mineral surfaces, which have similar characteristics. We would also have near-constant SZA in our samples. Since surface type and SZA effects will be near constant if our method is followed, we do not think it is important to include covariance with AOD.

Covariance between the LES-simulated q(z)/T(z) and aerosol is implicitly included in our development of the emulators, so we believe we have now displayed the things that matter for our application.

Technical Comments:

p.5 line 3: “with CWP calculated in the same manner as the TCWV”: also pressure-weighted? Likely not.

Actually yes, given the units we had. This text has now been deleted though: Section 1 new text explains our different path definitions for water vapour and Section 2.2.2 describes the calculation. In hindsight the “pressure weighting” is redundant information – there is only one way to convert our path-traced values into units of mm so we removed that term.

p.7 line 6: “random errors that we estimate”: please give examples (numbers) for these errors, in % of the TCWV.

Added. We slightly rephrased and added: “(0.6 % of TCWV in this case)”. This was the $\sigma_\varepsilon$ discussed in the emulator equation, and the OSSE range was 0.5—0.7 % of mean TCWV depending on the LES case (values in mm are in Richardson et al., 2021, Table 2).

**Citation**: https://doi.org/10.5194/amt-2021-163-RC1