Response to comments of reviewer 2

We thank the anonymous reviewer for the positive comments and the suggestions, which helped to improve the manuscript. In the following we provide a point-to-point response to all reviewer comments. The reviewer's comments are printed in italic and our response in roman font type. We indicate the line numbers of the revised manuscript where larger revisions have been made. For the reviewer's convenience we also copied larger changes we made to the manuscript to this response and enclosed them with quotation marks.

This paper outlines three key improvements to the TROPoe retrieval algorithm, a software package that is seeing increasing use in both operational and research milieus. These three improvements, namely the addition of a water vapor band to the retrieval, the integration of the t–1 retrieval to help improve temporal consistency and reduce striping, and the impact of radiance noise inflation, are all discussed. Overall, this is a well-written and compelling manuscript that fits with the scope of AMT and is suitable for publication after a number of small issues are addressed. These are mostly issues associated with explanations and justifications.

Response: Thank you for this positive evaluation.

The most significant issue I see is in the conversation about inflating the radiance noise to account for the fact that the model error is not expressly addressed. This invokes a somewhat lengthly list of questions, but in its current form the manuscript could do more to justify why this approach is proper and valid. Are uncertainties really fungible like that? Can one inflate one set of uncertainties and assume that it encompasses a different set of uncertainties that exist for an entirely different set of reasons? What are the limitations on including model uncertainties in the retrieval (i.e. just how expensive is it to do it explicitly, and by what factor is the outlined approach better)? Has there ever been an attempt to treat the TROPoe model errors explicitly, and if so, how do those results compare to the noise inflation approach? What is the purpose of reducing the noise with the PCA filter if one is just going to inflate it right back up again? Why is the number of converged retrievals the appropriate measure to determine if the proper inflation factor has been reached?

Response: The observational uncertainty (Se) needs to include contributions from the actual observations (Sy) and uncertainties in the forward model used to create the simulated observations (Sb'), using the notations in Maahn et al. (2020). Specifying Sb' has been done for microwave profiling systems (i.e., Cimini et al. 2018); however, there are only a few dozen lines to consider in the microwave. There are approximately 1,000 water vapor absorption lines in the spectral region used for the TROPoe retrievals from infrared sounders. We would have to estimate the uncertainty in the strength, width, and temperature dependence of each (and how these uncertainties are correlated) to compute the uncertainty in the forward model properly. Thus, the dimension of the b parameter vector is approximately 3,000 elements, if we only consider water vapor. But there are nearly 20,000 CO_2 lines between 500 and 960 cm^{-1} which would need to be considered also. Even if only the strongest absorption lines were considered, the total number of absorption lines (from H_2O and CO_2) would still be more than 2,000 in total. We are working to do this in the infrared, but it is a big project and is work-in-progress.

And thus, we needed a way to account for the forward model uncertainty so that we don't overfit the data. Turner and Blumberg (2019) first used this approach of applying the noise filter to reduce the random error in the observations, but using the original observational uncertainty for the sum of the two components. We found that this approach is still insufficient for some instruments (the observational uncertainty is instrument dependent) and hence proposed a minimum noise level to be used. To determine the appropriate noise, we not only looked at the number of converged retrievals, but also considered cDFS and MAE compared to radiosondes. Our choice of the noise level was a compromise between solution availability, information content, and error.

We added a footnote to the introduction, describing the challenge with including the uncertainty of the forward model:

'There are approximately 1,000 water vapor lines and nearly 20,000 CO_2 lines in the spectral region used for TROPoe retrievals from IRS, and we would have to estimate the uncertainty in the strength, width, and temperature dependence of each (and how their uncertainties are correlated) to compute the uncertainty in the forward model. Even if only the strongest lines were included, the number of lines would still exceed 2,000 in total. For MWR, the uncertainty of the forward model has been specified by Cimini et al. (2018); however, there are only a few dozen lines to consider in the microwave.'

1 Some other smaller issues:

In many cases, the IRSes and MWRs used in this study are not at the same location, but instead are located within the same general climate regime. Does that have any impact, i.e.can we compare the *moisture variability for the tropical IRS to that of the MWR? My guess is that it's fine, but it probably should be discussed.*

Response: In Fig. 2, we show near-surface temperature over mixing ratio for all sites to illustrate the differences in climatological conditions. Mixing ratio values at the two tropical sites (MAO and SAV) and at the two mid-latitude sites (SGP in April and LIN) are in a similar range. However, the variability is different. To better illustrate this, we added mean and standard deviation for each site to Fig. 2. For example, moisture variability was larger at SGP in April than LIN and larger at MAO than SAV, indicated by the larger errorbars. Because TROPOEIN aims to improve the temporal consistency in the high-pass filtered TROPoe retrievals (using a cut-off time of 3 h), we compared standard deviation for water vapor and temperature for high-pass and low-pass filtered data separately (Figs. 1,2 in this response). Most of the difference in variability between the sites in the same climatological regime are found in the low-pass filtered data (blue bars). The variability in the high-pass filtered data is much more similar, which means that the improvements of temporal consistency in the TROPOEIN experiments are done for similar conditions and that the findings for IRS- and MWR-based retrievals are comparable.

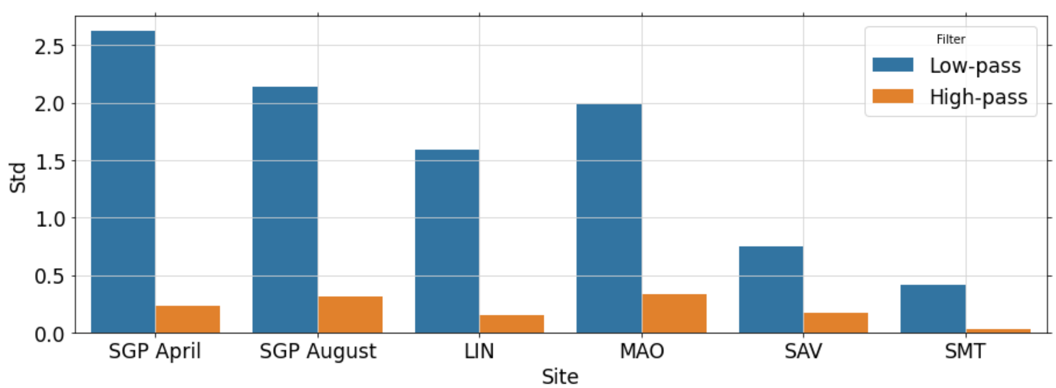

Figure 1: Standard deviation of near-surface measured water vapor mixing ratio for high and low pass-filtered data.

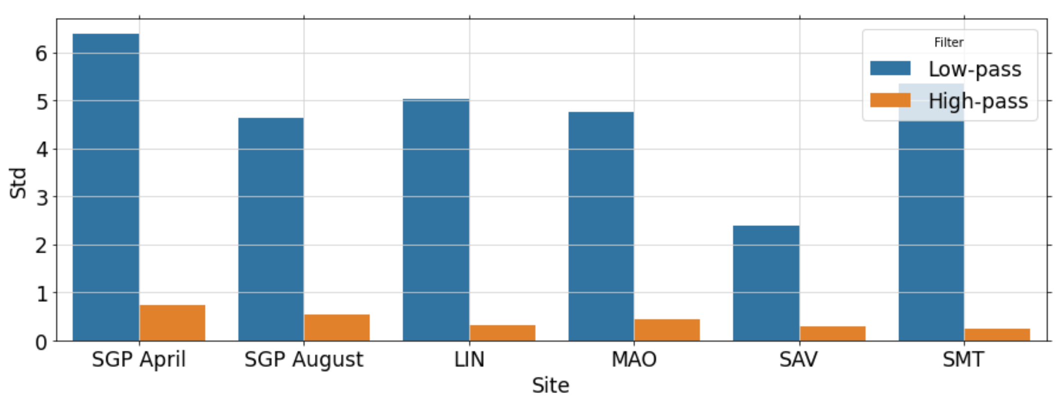

Figure 2: Standard deviation of near-surface measured temperature for high and low pass-filtered data.

We added this sentence to the description of thermodynamic conditions at the sites (l. 114-119):

"While the mean values are very similar for sites in the same climatological regime, i.e. MAO and SAV and SGP in April and LIN, the standard deviations vary, which may have implications for our experiment to improve the temporal consistency. However, the differences in standard deviation are mostly related to variations on time scales of several hours and more. Since we evaluate the improvements to temporal consistency on a shorter time scale, we are confident that the results for IRS- and MWR-based retrievals in the same climatological regime are comparable."

Line 119: is it "fore optics," "fore-optics," or "foreoptics?" I've seen all three, but I think I've seen the last one the most.

Response: Changed to foreoptics.

Line 129, Table 1: The geographic column would benefit by also adding some place names ("Oklahoma USA," "Brazil," "Greenland," etc.) Also, the parentheses in the Number of Radiosondes column are mismatched.

Response: Added and changed.

Line 158: Specify that the 55 levels in thermodynamic profile retrievals are for TROPoe; as it is written, it sounds like it's the case for all thermodynamic retrievals regardless of instrument.

Response: Changed.

Lines 216-226: Adding the WV band may help the number of converged retrievals, but is there an impact on their accuracy? Moreover, is there a discernable impact on the performance of the T retrievals in addition to the WV retrievals?

Response: The accuracy of the retrievals is investigated by comparing the retrieved profiles to radiosonde profiles in Sect. 4.3. Mean absolute and relative errors are shown in Fig. 15. Adding the additional water vapor band improved the retrieval errors for both water vapor and temperature. The improvements in temperature retrievals were smaller than for water vapor retrievals. This is described in Sect. 4.3

Line 251: The temporal consistency between the atmosphere is going to vary based on the diurnal cycle. Are there plans to vary the noise inflation uncertainty of the previous retrieval based on time of day?

Response: This is a great thought. We currently have no plans to implement a time dependent noise inflation uncertainty, but may consider it in the future.

Lines 255-259: how were the specific values for N decided?

Response: We chose the values for N empirically. These values need to be large enough to not suppress any real variability in the boundary layer between the 10-min consecutive profiles. The chosen values for N increase the uncertainty of water vapor in the boundary layer by a factor of up to 5 and increase the uncertainty of temperature by up 3 deg. The typical uncertainty in the boundary layer for water vapor is between 0.5 and 1 g/kg (except for the very dry environment at SMT) which means that water vapor mixing ratio in the boundary layer is allowed to change by more than 2.5 g/kg in a 10 min period without constraining the solution by the previous profile. We decided to go with these rather high values for N to be on the conservative side.

We added an explanation to the text (l.273-275):

"The values of N were determined empirically and the rather high values in the boundary layer allow water vapor mixing ratio to change by more than 2.5 g kg^{-1} (with σ_{WVMR} typically larger than 0.5 g kg^{-1}) and temperature by more than 3 °C close to the surface within a 10 min period, without suppressing the change by the previous profile. "

Line 276: If the processing is typically executed independently for each day, then when looking at continuous time series that span the 0000 UTC hour, there will be an artifact of increased variability for some discernible time period every day. Can the algorithm be modified to take into account retrievals from the previous day?

Response: The reviewer is right that the profiles shortly after 0 UTC are impacted by running the retrieval for individual days and will always have a lower information content. Because TROPoe is computationally expensive (especially for the IRS-based retrievals), several days are usually processed in parallel when historical data are being processed. This means that conditions of the previous day before midnight are not necessarily available when the retrieval is run for a specific day. But in a real-time processing mode, the code could be modified to read in the output from the previous day so that this artifact won't exist. We added this information to the text (l. 292-296):

'Note that this independent processing of individual days may lead to an artifact of increased variability shortly after 00:00 UTC. The independent processing is done because TROPoe is computationally expensive (especially for the IRS-based retrievals) and the retrieval is usually run for several days in parallel when historical data are being processed. In real-time processing this artifact could be avoided by reading in the output from the previous day. '

Line 368: fewer, not less

Response: Changed.

Line 470: Remove the comma between noise and decreases.

Response: Changed.

References

Cimini, D., Rosenkranz, P.W., Tretyakov, M.Y., Koshelev, M.A. and Romano, F., 2018. Uncertainty of atmospheric microwave absorption model: impact on ground-based radiometer simulations and retrievals. *Atmospheric Chemistry and Physics*, *18*(20), pp.15231-15259.

Maahn, M., Turner, D.D., Löhnert, U., Posselt, D.J., Ebell, K., Mace, G.G. and Comstock, J.M., 2020. Optimal estimation retrievals and their uncertainties: What every atmospheric scientist should know. *Bulletin of the American Meteorological Society*, *101*(9), pp.E1512-E1523.