We thank the Referee for the thoughtful and detailed comments. We hope we have addressed all of the Referee's concerns and we think that our manuscript did benefit from the constructive comments made by all Referees. In the following text, the Referee's comments are in black and our answers are in red.

==

This is a well-written paper that investigates an important issue in ground-based remote sensing: what improvements arise when additional datastreams are included in the retrieval? For years now, TROPoe and its antecedent AERIoe have included RAP or RUC profiles in the calculations, but the improvement in the retrieval has not been fully quantified. This paper is of the appropriate scope and novelty for inclusion in AMT. I have a few corrections and suggestions that will help improve the readability and utility of this paper, but these should be easy to address. None of these issues rise beyond the level of minor corrections.

We thank the Referee for the overall positive comments.

Most significant issue:

Sound propagates differently at different times of day, and the uncertainties introduced into the RASS observations by horizontal winds are also going to have a strong diurnal cycle. Because of that, one might assume that the biases and MAEs exhibited here are not constant throughout the day, but instead have a noticeable diurnal cycle. It may be that there's an insufficient number of radiosondes, especially at non-daytime hours, to fully investigate this. Regardless, the possible diurnal impact should be discussed, even if it's not anticipated to be an important issue and can easily be dismissed.

We agree with the Referee that sounds propagate at different speeds as a function of the temperature. This is indeed the principle on which the RASS is based. We also agree that the horizontal winds have a diurnal cycle, which affects the possibility of the sound wave to be advected out of the measuring volume of the RASS, impacting the height coverage of the instrument. This was already specified in Section 2.2 *"Moreover, the maximum height reached by the RASS is variable and limited by the advection of the propagating sound wave out of the radar's field of view and by sound attenuation (a function of both radar frequency and atmospheric conditions such as temperature, humidity; May and Wilczak, 1993)"*. This is now reworded to *"Moreover, the maximum height reached by the RASS is variable and limited by the advection of the propagating sound wave out of the radar's field of view (which can be different at different times of the day, as horizontal winds can have a strong diurnal cycle) and by sound attenuation (a function of both radar frequency and atmospheric conditions such as temperature, humidity; May and Wilczak, 1993)"*. We are however not aware of a reference on RASS errors having a diurnal cycle and, as the Referee mentioned, our dataset would not allow for such investigation as we have a limited number of radiosondes and only at daytime hours.

Minor comments:

1. One small issue I had while reading this paper was understanding how the RASS was integrated into the retrieval. The TROPoe retrieval works in T and q space, and it's relatively easy to understand how the RAP profiles would be able to be included as they exist (or can easily be converted) into those variables. However, RASS is measuring Tv which is a unique variable for the retrieval. How does the retrieval address this? Is it just calculated at the end of each iteration from the interim T/q profile and compared to the RASS observations?

Yes, the Referee is correct and this is how it is done: before computing the Jacobian, the virtual temperature is computed from the state vector and compared to the RASS measured virtual temperature. This is now clarified in Section 3: *"We note that since the RASS measures virtual temperature, when this is included as input, the virtual temperature is computed at the end of each TROPoe iteration from the state vector and compared to the RASS measured virtual temperature."*

2. When I read line 76, noting that the radiosondes were interpolated to the TROPoe grid, I wondered why the sondes weren't smoothed with the retrieval averaging kernal instead. I see that was addressed later in Line 195. Still, I wonder if this is the best approach. Would it be better to smooth the sondes, calculate the differences as a function of height, and then interpolate the vertical profile of the differences to a common grid to facilitate the analysis between different combinations of instruments?

We agree with the Referee that smoothed radiosonde observed profiles can be computed using the averaging kernel. However, the averaging kernel depends on the retrieval parameters (e.g., which datasets are used as input in the TROPoe runs), so for our 12 configurations we would have 12 different averaging kernels. For each of these the smoothed radiosonde profile can be quite different from each other and also from the original unsmoothed radiosonde profile. Consequently, while comparison of the retrievals to the relative averaging kernels radiosonde profiles can be used to minimize the vertical representativeness effects due to the different vertical resolutions of these profiles, we were not convinced that a statistical comparison between the 12 TROPoe configurations would be fair if each of their retrieved profiles is compared to a different averaging kernel-smoothed radiosonde profile. Therefore, we decided to present the statistical analysis in the manuscript comparing the various TROPoe retrieval configurations to the unsmoothed radiosonde profiles, just interpolated to the same vertical levels of the retrieved profiles. Additionally, we decided that we want to ultimately present statistical values relative to real radiosonde observations.

3. In certain instances, the additional observation vector entries degrade the profiles, either by increased bias or MAE. This is an interesting finding that ought to be discussed more. What is causing this, how consistent is it (are a couple of retrievals dragging everything down, or are most of the retrievals behaving similarly)?

When both passive instruments have enough information the retrieval of the combination of the 2 has to balance between both inputs, which might not be optimal. Additionally, physical retrievals might struggle to overcome systematic biases in the observations, and although we are bias correcting the MWR observations there may still be some residual bias that isn't accounted for correctly. Nevertheless, from Fig. 6 we see that the combination of the 2 passive instruments is beneficial in the lower part of the atmosphere (< 2 km), where the MWR tends to struggle identifying the correct height and shape of inversions (configuration #9 better than configuration #1). As a matter of fact, even in Fig. 7a we see that for the first point on the x axis the MAE of temperature is improved compared to the MWR only configuration, both when averaging over the lowest 5 km and lowest 3 km of the atmosphere.

Technical corrections:

Line 42: this sentence is somewhat awkwardly phrased and would be easier to interpret if rewritten.

The above-mentioned sentence has been re-worded in the revised version of the manuscript from: *"Several ground-based sensors are nowadays available and active in many geographical locations including in-situ or remote, and active or passive sensors."*, to: *"Several ground-based sensors (including in-situ or remote, and active or passive sensors) are currently available and operational in many geographical locations."*

Line 43: I don't think it's quite correct to state that in situ sensors only provide point measurements. Surface meteorology stations like ASOS only provide observations at a single point fixed in 3D space, that is true. However, aircraft-based observations like AMDAR are in situ but producing moderately-dense vertical profiles and the sheer density of these observations next to major airports produces a multidimensional web of observations.

We agree with the Referees, and mention of the aircraft-based observations has been included in the revised version of the manuscript, in the Introduction: *"In-situ sensors only provide point measurements (except for aircraft-based observations that can produce moderately-dense vertical profiles, although only sporadically in time)"*.

Line 54 and elsewhere: the name of the radiosonde is the Vaisala RS-41, not Vaisala-41.

Corrected, thanks for catching the mistake.

Line 96: it is somewhat unclear what the upper range of the IRS is: why does it range from 3000-5000 cm^-1? Did the two IRSes each have a different spectral range?

The IRS used in our study goes out to 5,000 cm^{-1}, but the AERI (which reference we used) measures out to 3,000 cm^{-1}.

Line 440: necessarily, not necessary

Corrected, thanks for catching the mistake.

Line 444: cloudy conditions, not cloudy-conditions.

Corrected, thanks for catching the mistake.

Line 454: Based on my experience as an AMT author, making data available by request may be insufficient for their standards. It may be good to identify an archive where these data may be stored and given a DOI.

Following the Referee’s suggestion, we uploaded all radio soundings and TROPoe output files in the Zenodo data archive: 10.5281/zenodo.10815373.