**Review AMT-2023-263:**
**Sensitivity of thermodynamic profiles retrieved from ground-based microwave and infrared observations to additional input data from active remote sensing instruments and numerical weather prediction models**

**General comments :**

This paper evaluates the sensitivity of temperature and humidity retrievals to different combinations of ground-based remote sensing instruments as well as NWP model information. It uses the state of the art optimal estimation retrieval algorithm TROPOe that has been used in several scientific publications to demonstrate the improved accuracy of thermodynamic retrievals when combining together active remote sensing (water vapor lidar, RASS, ceilometers) and passive instruments (ground-based microwaver radiometers (MWR) and infrared spectrometers (IRS)). This new study is relevant for the scientific community due to the uniqueness of co-located remote sensing instruments at the same site (MWR, IRS, RASS, ceilometers and tower measurements) over a long time period (fall 2021-winter 2022). It also contributes to a new perspective of improvement for TROPOe by including NWP model information above 4 km altitude within the algorithm. The results show that RASS contributes to improved temperature retrievals within the boundary layer while the inclusion of NWP information significantly improves the temperature and humidity retrievals mainly above 3 km. Overall the manuscript is very well written, figures are well presented and explained. However, I am concerned by the conclusions of this study that are quite limited due to the small dataset of co-located radiosoundings available (only 15 RS). As clearly stated by the authors, conclusions taking into account cloudy-sky conditions could be different for the synergy between MWR and IRS observations for example. I am also wondering if degradations in the retrievals could happen when including the NWP profile above 4 km directly in the observation vector (if I understood well) that might not been observed in this study due to the limited dataset of validation. In fact, the retrieval algorithm could try to minimize the distance between the atmospheric state and the a priori as well as observations with potentially non-consistent observations (the NWP profile could potentially try to push the algorithm in a direction while the observation in another direction). The author should clarify this point and better justify the methodology used.

Additionally, I think a discussion on the differences in TROPOe results (DFS, vertical resolution, mean uncertainties, retrieved profiles) for clear-sky versus cloudy-sky days (as the observations are all available during one year) with respect to the different configurations could be beneficial to the paper. Even without available radio-soundings, DFS, vertical resolution and uncertainties could have been discussed as well statistical distributions of retrieval differences. If the TROPOe retrievals are available over a long time period and the authors can lead this analysis, I think it could improve the current manuscript.

Finally, the manuscript does not clearly state if MWRs are used with off-zenith scans. Several publications have demonstrated the significant increase in DFS by including off-zenith observations. If MWR observations have been used with zenith only observations, the comparison with the AERI instrument is not really fair and we could except more accurate temperature and lapse rate retrievals with the inclusion of off-zenith channels. This will not drastically change the

conclusion with RASS and RAP but it would definitely affect the conclusion comparing single passive instruments.

Afer taking into account the scientific points listed, I would recommend the publication of this manuscript in AMT.

**Major comments :**

- **Section 2.2** : As shown in Djalalova 2022 , even after nitrogen calibration, significant biases can be observed in MWR observations. Optimal estimation is very sensitive to biases in observations. Did you implement any bias correction or quality control of the brightness temperatures before applying TROPOe ? Even if the IRS is self-calibrating, was there any check on potential biases in IRS also ?

In line with the general evaluation : did you use off-zenith observations for the MWR ? If only zenith observations have been used, it should be clearly stated through all the manuscript that all conclusions comparing the MWR and IRS retrievals are underestimating the capability of current MWRs that are generally used with boundary layer scans to improve the vertical resolution of temperature profile. If another configuration with zenith and off-zenith observations could be included in the manuscript, it would be beneficial for the discussion.

**Section 2.4** :It is not clear to me if the RAP model is used within the a priori profile or within the observation vector. If it is used within the observation vector how was defined the corresponding observation error covariance matrix ? As mentioned in the overal evaluation, I am concerned that this methodology could degrade the retrievals in case of larger errors in the NWP profile non consistent with the other observations that might not been observed due to the limited number of radiosounding observations. Could you clearify and justify the methodology ? Several publications using an alternative approach with NWP model used directly within the a priori profile should also be cited (Hewison 2007, Cimini et al 2015, Martinet et al 2020) to discuss the difference with your methodology.

**Figures 4 and 5:**

- When the information content from observations is small, the inclusion of external information from NWP models has a significant impact on the retrievals. This is demonstrated in this study in figures 4 and 5 both for temperature above 4 km agl and to a larger extent for humidity above 1.5 km. Could the improvement on water vapor be larger by using NWP profiles from the surface up to the top of the atmosphere ? Both MWR and IRS have lower information on humidity compared to temperature so we could expect a larger benefit when using NWP information even below 4 km. Did you perform a sensitivity study using the whole NWP profile and not only the profile above 4 km ? Could you justify this choice to start at 4 km even for humidity ?

**Figure 5 :**

It seems that the configuration MWR + RAP degrades the vertical resolution of the configuration #1 with MWR only below 1.5 km : could you comment this result ? Do you have any explanation on

this slight degradation (which is overall pretty small compared to the large improvement that you get above 2 km) ?

**Figure 6 :**

The degradation due to the inclusion of the RAP data is significant for configuration #1 even below 4 km (from ~ 0.3 g/kg to 0.5 g/kg at 500m). This degradation is not really observed for the configuration #5 (IRS only). Do you have any hypothesis to explain this degradation when only the MWR is used ? I assume that this degradation could be due to the prior state covariance matrix used to spread the information from the observation level to adjacent vertical levels : considering that IRS has more information content in humidity than the MWR alone, the retrievals below 4 km might be better constrained by the observation while, in the MWR configuration, most of the prior state modification below 4 km is driven by the vertical correlations specified in the prior state covariance matrix. Did you test different prior state covariance matrices to evaluate the sensivity of the retrievals to this matrix ? Did you try to use the whole RAP profiles with its corresponding error covariance matrices to evaluate if this degradation is still observed ?

**Figure 7 :**

The temperature bias is signifiantly increased in the configuration MWR + IRS compared to MWR or IRS only when averaged over 5km, I am puzzled by this result : could you include a discussion ?

**- Figure 8 :**

To be fair on your comment, I think the averaged bias and MAE of mixing ratio over 3 km should be presented as well as the inclusion of RAP significantly degrades the statistics of mixing ratio compared to MWR only below 3 km (which might give different results than your current conclusion that the impact of RAP only degrades slightly the bias and MAE).

**Minor comments:**

- line 41 : aren’t => are not.

- Table 1 : can you check the unit of mixing ratio (g / km ?)

- line 108 : isn’t => is not

- line 368 : 0.5 g /km => 0.5 g/kg.