To the editor,

We would like to thank you for your consideration of this work for publication in Atmospheric Measurement Techniques, and to thank both reviewers for their kind and constructive comments. We particularly appreciate the reviewers' positive remarks on the manuscript, this being the firs

Below we set out a point-by-point response to the reviewers, indicating the reviewer comment **(RC)**, author response **(AR)**, and manuscript change **(MC)** where relevant. We abbreviate the originally submitted version as 'SV', the newly revised version as 'RV', and the version with tracked changes as 'TCV'.

Reviewer 1

RC: *Main Comment:*
The Methods section describes a great deal of effort to constrain the input data for LBLRTM, but then in section 4.1 and again in section 5, the study second-guesses this work and proposes an alternate preparation of data. In section 4.1 we learn that the surface skin temperature could be better constrained and in section 5 evidence is provided suggesting that the radiosonde humidities were biased dry, which is not a surprising error. There is little independent evidence of the skin temperature bias, but the biases in the CO2 band (~650-750 cm-1) in Figs. 6b and 6c suggest a positive bias also in the near-surface air temperature supplied to LBLRTM of up to +1 K, since FINESSE is most likely better calibrated than this. I think that the authors should constrain the input data as best as they can up-front in Methods (including better near-surface air temperature, their best estimate of skin temperature, and the bias-corrected humidities) and provide a single best-estimate simulation for comparison in Results. Then, in Discussion provide the transparency by reminding the reader of some of this data preparation work and then commenting on how it affected the results.

AR: This is a useful suggestion that can help to focus the paper. In the RV, we focus the study more clearly on PREFIRE validation, using FINESSE to constrain the input atmospheric profile as much as possible.

We have thus taken the following approach to the points addressed here. We begin by investigating FINESSE radiances using the in-situ data as is. We adjust the profile first to amend the water vapour bias (scaling radiosonde water vapour to match the ERA5 column). We then address the biases in the centre of the 15 µm CO_2 band by reducing the temperature of the whole profile by 0.86K; it was found that a whole-profile adjustment was helpful to not only reduce the bias in band centre but also reduce residuals in the band wings. An independent analysis of ERA5 simulated versus FINESSE radiances under clear conditions sampled throughout the WHAFFFERS campaign shows no persistent bias in one direction or another across the CO_2 band centre. Therefore, we do not think the feature here is instrument related and consider it justified to scale the atmospheric temperature profile in the RV.

When turning to PREFIRE, we include the GOES surface temperature in Methods as our primary skin temperature source rather than ERA5, and use our 'best-case' profile from FINESSE, thus only having one profile to compare to PREFIRE observations instead of the current 4. We also add an uncertainty estimate from our humidity and temperature constraints.

MC: We have revised §2.3 (detailing sources of ancillary data) to reflect our atmospheric profile now using primarily in-situ data, having more limited use of ERA5, and using the GOES ABI temperature

retrieval for the surface temperature input. We have removed discussion about 'ERA5' vs 'In-situ' profiles since this comparison is no longer relevant, and Table 1 (SV) has also been removed. Figure 3 and its caption have been updated to show the atmospheric profiles 'A' (= original 'In-situ' profile) and 'C' (= fully optimised profile). In §3.3, discussion detailing the choice of *h* has been slightly revised (L311–318 TCV) now that the ERA5 profile is no longer used.

The 'Results' section from SV has been split into two sections. The new §4 in RV (L338–397 TCV) focuses on optimising ground-based closure with FINESSE, and this is where we discuss adjustments to the profile's temperature and humidity. We keep some of the factual discussion of the residuals from SV but comparison between ERA5 and In-situ profiles has again been removed. The plot of FINESSE residuals (Fig. 5 RV) has been updated to remove the ERA5/In-situ profile comparison and instead reflect profiles labelled A, B, and C which reflect the sequential improvements to the profile.

The new §5 (L398–444 TCV) then presents the results from PREFIRE, again with modifications to reflect that we are no longer comparing two profiles. The subsection in SV about modifying with the GOES ABI surface temperature has been removed, with much of this content moved to §2.3. Comparisons between ERA5 and In-situ profiles have been removed from the Discussion section. The abstract and conclusions have been updated to include the optimisation procedure carried out using FINESSE.

RC: *Title: The title as it stands is a little overstated for comparison of a single measurement. Maybe "Validation of PREFIRE mid- and far-infrared radiances during the WHAFFERS campaign" or similar.*
AR: We acknowledge that it is possible to read the title in its first revision as overstated and accept this change for the RV.
MC: Title changed to "Validation of PREFIRE mid- and far-infrared radiances during the WHAFFFERS campaign".

RC: *L73-74: Could you provide some specifics on the meteorology? The FIR would be semi-transparent for precipitable water vapor < ~ 1 cm. What was the PWV?*
AR: The PWV values for the profiles used are recorded as text in Fig. 3, and we have added a pointer to this in the main body of the text in RV (L177 SV). We have highlighted in the RV that each radiosonde launched in the campaign measured PWV < 0.6cm, meaning semi-transparent conditions in all cases even after accounting for the presumed dry bias of the radiosonde.
MC: Added note on the radiosondes measuring low TCWV (L76 TCV). Mentioned the TCWV values of the profiles being shown in the figure in L208 TCV.

RC: *L90: There are a lot of sensors given only as their acronyms in the paper; AERI, CrIS, IASI, as well as models (PCRTM), some of WHAFFFERS. I know it's a lot of words, but maybe a table would help: full name, spectral resolution, absolute accuracy, reference.*
AR: We have expanded the acronyms and provided references where relevant in the RV.
MC: Several of the above acronyms expanded and referenced in L116–119 (TCV). Discussion of AERI has been removed as we felt that such detail of upgrades to FINESSE was no longer relevant (L89–96 TCV), so the AERI has not been expanded or cited. It would not be relevant to expand and cite each element of the 'WHAFFFERS' acronym.

RC: *L117: What are the criteria for valid spectra? You mean clear skies?*
AR: The word "valid" intended to indicate PREFIRE spectra that have passed the quality control checks; this has been clarified in RV.

MC: Change of wording in L123 (TCV).

RC: *L118: Why is nighttime a problem? Sounds optimal to me.*
AR: Measurements were not generally taken at nighttime during the campaign due to the working hours of the group; FINESSE requires manual supervision during operation. We have noted this in RV.
MC: Added 'outside of working hours' in L125 (TCV).

RC: *L123-124: "good stability". Can you quantify this, perhaps for several hours around the overpass? It seems like a useful data point for assessing uncertainty.*
AR: We find that the variability (standard deviation) in spectra over the averaging window is equal to the NESR (see Fig. RC1-1, top left panel), indicating that no statistically significant change in the scene is observed over this time. Over broader time windows, the effect of changes in temperature, humidity, and cloud cover becomes visible (Fig. RC1-1). A clarification that variability matches the NESR has been added to RV.

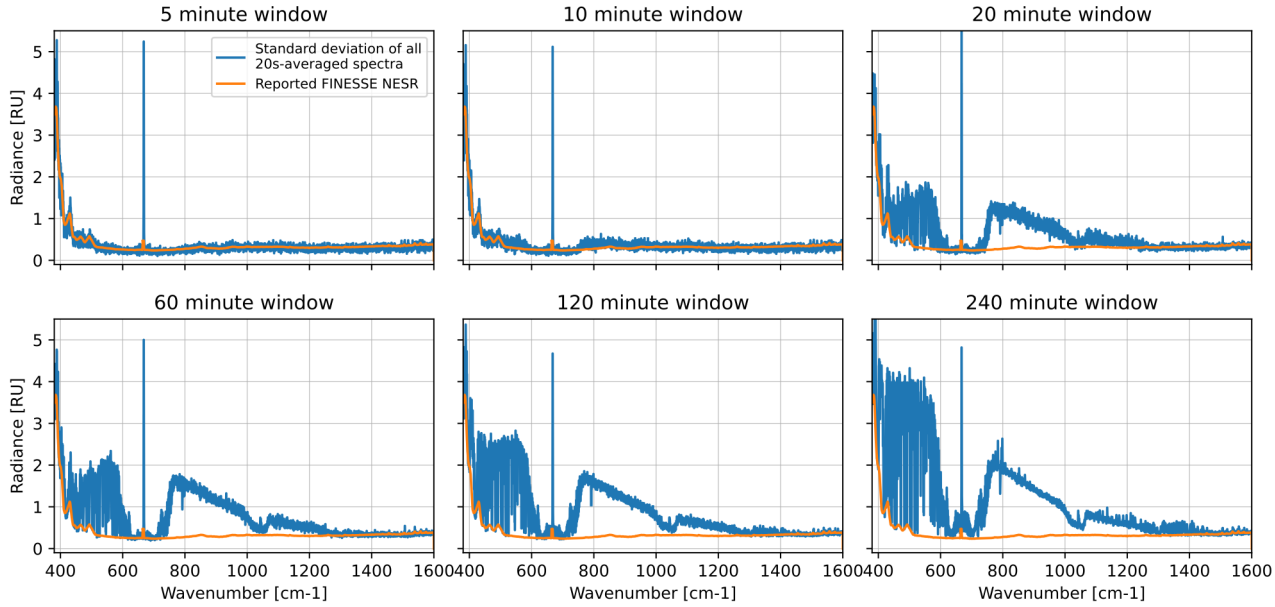


Figure RC1-1: Standard deviation in observed FINESSE radiances, calculated over increasing time windows centred on the PREFIRE overpass. The FINESSE NESR is shown for comparison. 1 RU = 1 mW $(m^2\ sr\ cm^{-1})^{-1}$.
MC: The wording of this point has been made more precise, see L131–132 (TCV).

RC: *L149-150: Can you elaborate on this bias?*
AR: Bias in the meteorological sensors at the Gault site is not of key importance to our study and was mentioned mainly to alert other potential users of the data. Across many measurement days, the 1.5m and 3.5m sensors were consistently warmer than those at other heights (higher, lower, and in between) which we believe to be unphysical. However, to simplify the paper, we have removed this point of discussion in the RV.
MC: See L155–158 (TCV) where we have created a simpler exposition of the sensors from the stationary weather station, now just mentioning the four probes used, rather than those that we did not.

RC: *L151-154: FIR surface emissivity seems like a really large source of uncertainty. It would neither be well-known nor easily matched to the surface cover in the actual FOV, which is at least partly*

mixed and sloped (FIR surface emissivity of a partially snow-covered forested hill, yikes!). I would like a better sense of what sort of uncertainty poorly constrained surface emissivity causes.
AR: We also recognise the importance of the surface emissivity in this study. The inclusion of MODIS imagery in Fig. 2 is intended to show the surface observed by the satellite and demonstrates the extent of the snow cover over what is largely flat land. We used a 'medium snow' surface type from the Huang et al. (2016) dataset as a reasonable mid-range estimate and included the spread in snow types in our uncertainty estimate.

For modelled snow surfaces, the effect of varying viewing angles up to 15° (the maximum PREFIRE viewing zenith angle across all scenes) on the surface emissivity was found to be minimal, causing a maximum change in emissivity of only 0.0005. A surface containing some more 'sheet-like' ice (e.g. if some snow had thawed and refrozen) could have a different emissivity but there is insufficient data to investigate the actual nature of the surface cover at that time. In Figure RC1-2 (solid lines), we demonstrate that even a 20% change to ice would have only a small impact on observed radiances, or equally a change to a different size of snow.

We note that the current (SV) estimate of forward-modelling uncertainty due to snow crystal size-dependent spread in emissivity is lower than that due to surface temperature by a factor of about 4, and that emissivity uncertainty only impacts (semi-)transparent atmospheric regions (dashed lines, Fig. RC1-2). It appears that the surface emissivity uncertainty is underestimated for some parts of the spectrum, so the estimate of this uncertainty has been modified in RV to reflect the red dot-dash line.

There is a lack of modelled or observational data for the FIR emissivity of the woodland or urban areas included in the PREFIRE FOVs. The Huang et al 2016 dataset indicates that the emissivity of woodland surfaces is not drastically different to snow where the surfaces are observed in the MIR. Unfortunately, we also do not have PREFIRE retrieved emissivity to compare to for this scene since it is out of the range of operational retrievals (polewards of 60° latitude). We highlight this FIR emissivity data gap in RV.

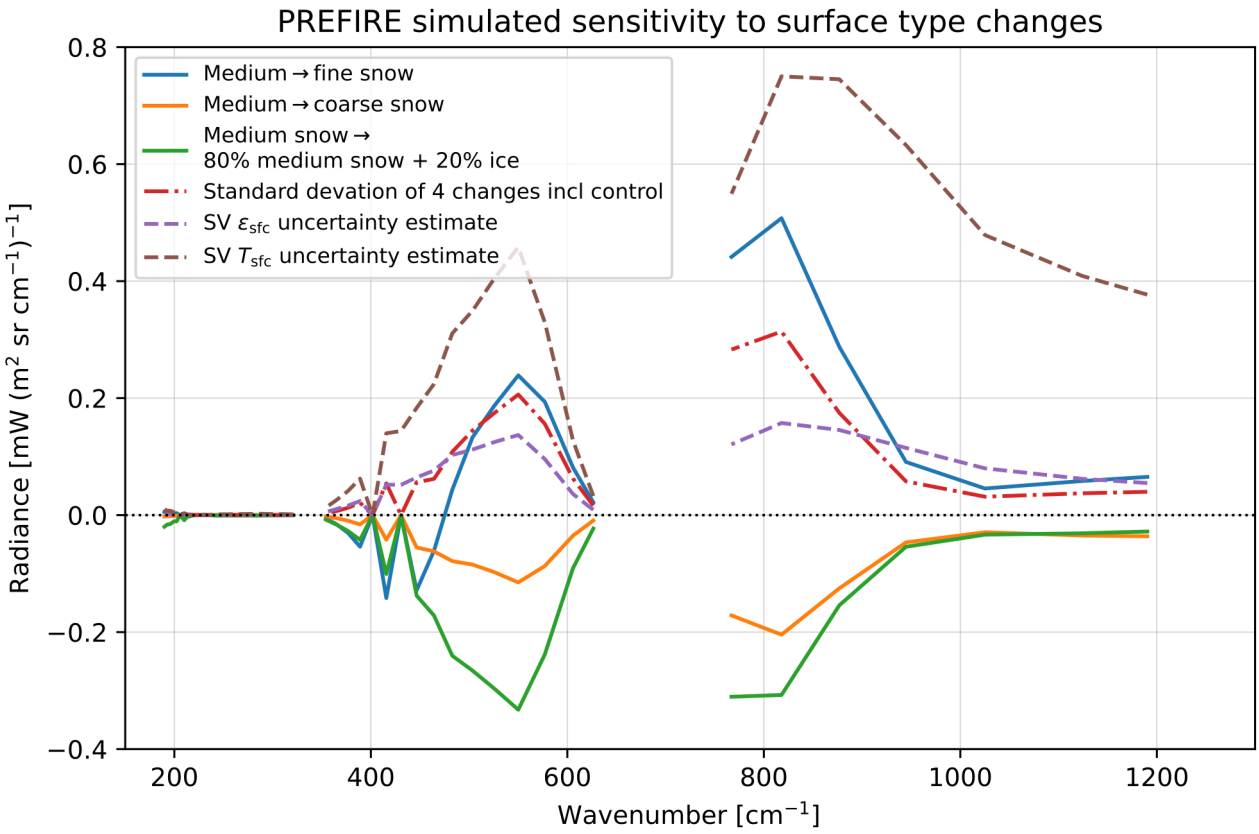


Figure RC1-2: Solid lines: sensitivity of PREFIRE to changes in simulated surface type, represented as offsets from the 'medium snow' type used as default in simulations. Dash-dot line: standard deviation of these offsets, including the control (i.e. zero offset). Dashed lines: uncertainty estimates in radiances from the SV, resulting from uncertain surface emissivity and surface temperature.

MC: We have updated the emissivity uncertainty estimate in the RV, such that its magnitude is now greater at ~550 cm^{-1} and ~800 cm^{-1} (dash-dot red line). We have added a mention of the lack of modelled and observed emissivity datasets (L511–517, TCV).

RC: *L155: I don't see ozone listed here, but it was clearly included based on Figure 6.*
AR: Ozone concentrations in the study have been taken from ERA5 (see L137, SV).
MC: Ozone is now highlighted separately, see L158–160 (TCV).

RC: *L205: Did ERA5 assimilate this radiosonde?*
AR: This was a research radiosonde launch not forming part of any operational observation network and as such it was not assimilated into any reanalysis product such as ERA5.
MC: No change needed.

RC: *Figure 5: For uncertainties that are not NESR, what are we looking at, mean bias, RMSE?*
AR: Interpretation of the uncertainties in the figure should be guided by the text, but some clarification in the figure keys and caption can help and so we have added this in the RV.

- In (a), calibration uncertainty is the instrument 'absolute radiometric accuracy'.
- The uncertainties in (b) are reported solely as 'uncertainty' in the PREFIRE product, we believe this is equivalent to NESR.
- In (c-f) SV, uncertainties are the dy described in the text §3.3. We believe that the approach taken approximates a 1-sigma deviation in spectral radiance due to a 1-sigma uncertainty in profile input quantities.

- The uncertainties in (g,h) SV indicate the expected range of the radiative uncertainty due to uncertainty in H2O spectral parameters. Spectroscopic uncertainty values represent only an estimate of the effect due to the inexact nature of the source uncertainty values (HITRAN and Mlawer et al 2019) and so we cannot apply an exact statistical label to our derived uncertainty estimates. In the RV we include more in Discussion about this uncertainty and note that there is a need for the spectroscopic community to share experimental uncertainty results or for uncertainties that are shared to be properly utilised and documented in spectroscopic databases.

MC: The uncertainties section (§3) has been rearranged and modified to clarify and consolidate discussion. We have swapped the 'Instrument performance' and 'Collocation uncertainty' subsections to match the order shown in the uncertainty figure. The plots for collocation uncertainty have been included in the "main" uncertainty figure (Fig. 4 in RV), thus eliminating Fig. 4 (separate collocation uncertainty figure) in SV and putting all uncertainties considered in one plot. Legends in the figure have had "NESR" and "accuracy" added (a, b), and we have added in the caption and in L317 (TCV) that the profile uncertainties (e, f) are at a 1-σ level. We have added discussion of the lack of spectroscopic uncertainty data in L486–492 (TCV).

RC: *Figure 6: One thing that really stands out in this comparison is the difference in the CO2 rotational band from ~650-750 inverse cm. Looks like the surface temperature input to LBLRTM is biased high by an amount I'm guessing is close to 1 K and it seems like the error in that value should be smaller. … Ah, I see the text explanation now (L305): I don't think you want to hand-wave this away. The near-surface can likely be constrained better and it might be important to the UW data for PREFIRE.*

AR: We have addressed this point in the response to the "Main Comment" above.

MC: See above.

RC: *L350-351: I don't understand "instrument effects". Maybe be more specific here. And what evidence is there to "believe" this?*

AR: The wording was perhaps unclear here and has been revised in RV. The evidence to believe this is that all other known sources of uncertainty have been accounted for in the forward modelling part of the closure experiment such that the residuals must, by elimination, be due to deficiencies of the observing system.

MC: See L449–450 (TCV) for the clarification including a remark of the uncertainty accounting.

Reviewer 2

RC: *The first [concern] is that the entire analysis is only a single case, and thus there is a question about how representative these results are overall. I would like to have the authors add more discussion in the final section of the paper about this.*

AR: This is an important point to highlight and so we have added more discussion about this in RV. We also acknowledge that there is clearly a greater range of PREFIRE observation conditions to test: observations from different sensors in the cross-track array, observations from Satellite 2, and different in-orbit environments to which there may be some sensitivity (daytime/nighttime causing thermal effects in the satellite bus).

We believe that focusing on a single case has allowed us to scrutinise many details of the atmospheric profile and uncertainties at all points in the closure tests. We hope that the rigorous

methodology of this study can be employed to strengthen future longer-term monitoring and validation experiments.
MC: We have added a note to this effect in the discussion, see L456–462 (TCV).

RC: *Second: there were other ground-based spectrally resolved instruments at the site also. Why were these not included in the analysis? For example, the AERI covers a good fraction of the same spectral bands as the FINESSE, and showing that either the AERI and FINESSE are in good agreement, or similarly that the AERI agrees well with the ERA5-DW profile, would provide significantly more support that the FINESSE results are solid.*
AR: This is a salient point and so we have investigated the potential for including AERI in the analysis. Direct comparison of AERI and FINESSE observations at anything close to native resolution is highly non-trivial due to differences in the instruments' spectral sampling and line shape.

A potentially simpler approach is using our atmospheric profile to simulate AERI radiances. We show the agreement obtained using the ERA5-DW profile in Figure RC2-1 (middle panel) with the equivalent residuals for FINESSE shown in the bottom panel. It is immediately evident that the AERI residuals show improved agreement with the simulations in the centre of the 15µm CO_2 band. However, at the same time, the residuals in the band wings are markedly higher (exceeding the expected uncertainty levels) and there is a noticeable negative bias across the main atmospheric window which is not seen in the FINESSE observations. Work is ongoing to explore why these enhanced residuals in the AERI comparison are present (using observations from across the campaign) and this will form the basis of a separate study led by the McGill team. We note that the AERI and FINESSE instruments were displaced by c.50m at the site which may lead to some differences in near-surface atmospheric conditions sensed by the two. Sensitivity tests show that PREFIRE would be insensitive to these changes, but they may add confusion in interpreting ground-based results. For these reasons we have not included the comparison in the RV.

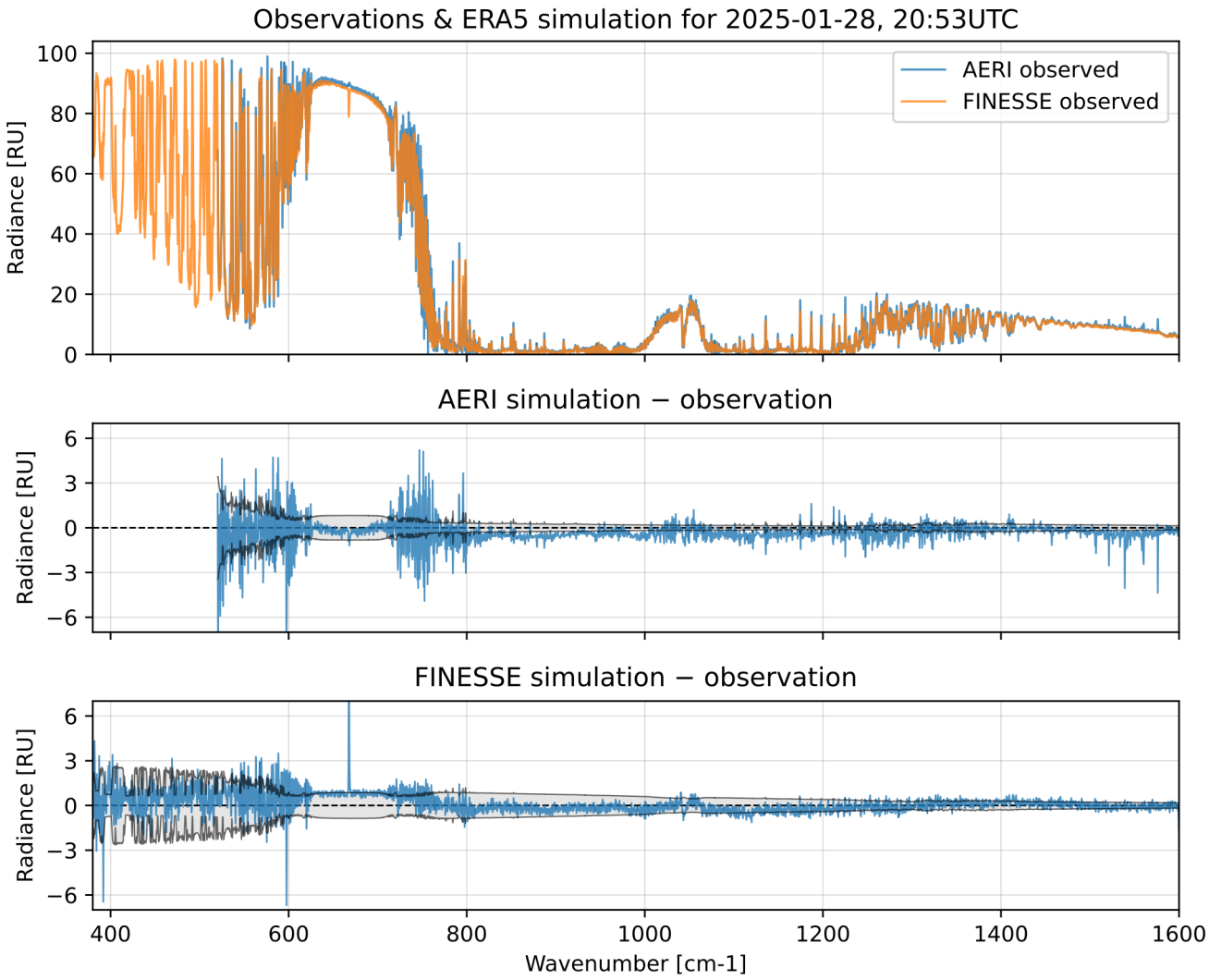


Figure RC2-1: Comparison of AERI (middle panel) and FINESSE (bottom) residuals for the ERA5-DW profile used in SV. The shading and black lines in the bottom two panels indicate the uncertainty ranges due to NESR, instrument calibration, and uncertainty in input the profile quantities (that is, all sources considered in the SV except spectroscopic uncertainty).

MC: No changes for this point.

RC: *Third, and this is a recommendation not a requirement: The authors show that there is a near-surface temperature error in the ERA5-DW and InSitu-DW results, as well as some water vapor challenges in the latter. The Mlawer et al. 2019 and Mlawer et al. 2024 (JGR) papers use thermodynamic profiles retrieved from the AERI assuming spectroscopy from RHUBC-I is well-validated. Would your results and interpretation change if you used a retrieved thermodynamic profile in your closure analysis?*

AR: We believe that performing a full retrieval of thermodynamic profiles from FINESSE is unlikely to alter the results significantly from what we present in the RV. As can be seen from Fig. 5 (in RV), the residuals predominantly lie within the combined observational/simulation uncertainty estimates. By using independently gathered thermodynamic profiles we can exploit the high spectral resolution and accuracy of the FINESSE instrument to assess, insofar as is possible given other uncertainties, the agreement of current spectroscopy with the observations. As hinted at by the reviewer, most retrievals assume the spectroscopy as a given. We argue that this is not true for

the far-infrared in particular, given the relative lack of field or laboratory spectroscopic measurements across this range.

Considering the comments from both reviewers, we have slightly reframed the paper to focus more on validation of the PREFIRE L1 observations (please see AC1 for details). As such, we have used FINESSE to constrain these temperature and humidity errors seen in the SV, in effect a 'manual' retrieval of these quantities. PREFIRE has very limited sensitivity to the H_2O spectroscopy (see Fig. 4g, h in RV) and so we do not lose an opportunity to assess this accuracy by following this approach.
MC: These changes align with those detailed above in response to the main comments from the first reviewer.

Additional minor edits

- Small changes in §3.3 have been made to improve readability of the description of the use of Jacobians to evaluate uncertainty (L289–310, TCV) including one correction on L299 (TCV).
- We edited the description of the radiosonde dry bias in the conclusions (L546–550, TCV) to add more concrete detail.
- We have made the description of the uncertainties for PREFIRE in the conclusions more precise, by highlighting the three bands studied (L558–576, TCV).

We believe that, with the changes to the manuscript detailed above, we have satisfactorily addressed the reviewers' comments in the revised version, and that the manuscript is ready to progress to the next stage in publication.

With best wishes,

Benedict Pery and co-authors.