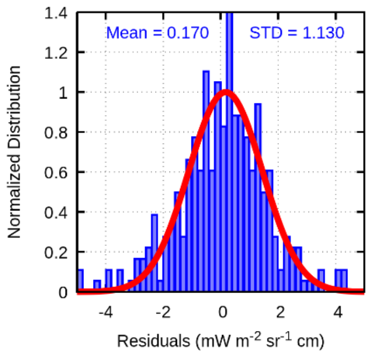
We thank the Referee for their comments and suggestions. In the following we answer the Referee comments point by point:

This paper would greatly benefit from having a comparison with both the E-AERI in radiance space. Due to the differences in the spectral resolutions, I would recommend the authors use the "double difference technique" outlined in Tobin et al. JGR. 2006. As the two instruments are essentially collocated (although vertically offset by 4 m), the spectral differences between 405 and 600 cm-1 should be within the instrument noise (if both systems are well calibrated).

We thank the referee for the suggestion to use the "double difference" technique. The analysis was added to the manuscript (Sect. 4.2.3). As in Tobin et al. we looked at the differences distribution and found an average of 0.17 mW m-2 sr -1 cm and a std-dev 1.13 mW m-2 sr-1 cm. (only the numerical details were added not the following plot)



Line 253: did you assume any cross-level covariance in your a-priori? Were there any cross-correlations between temperature and humidity? There should certainly be cross-level correlations in temperature due to the atmospheric lapse rate, and a long analysis of radiosonde data from the region (or ERA5 data) should indicate if there should be other correlations in the a-priori. If you assume the a-priori is a diagonal matrix, that will essentially increase the information content (DFS) of the retrievals.

To clarify the assumptions on the a-priori covariance matrix the following sentence was added:

"The a-priori covariance matrix was constructed assuming for both parameters a correlation length equal to 2 km between adjacent levels, while no cross-correlation was imposed between temperature and humidity".

Line 260: It is important to note that the gradient in a cloud-free measurement is zero only because it is so dry at the Zugspitze location. If you were in a tropical location, there would be a negative slope. This needs to be stated.

We agree with the Referee, the following was added to the text:

"[…] cloud-free measurement would have a gradient of 0 in the very dry winter conditions at Zugspitze,"

Line 279: In the selection of the 625 cases, did the Raman lidar (or the backscatter lidar, which was briefly mentioned later in the paper) confirm that these were cloud-free?

Yes, the backscattering lidar shows that in case of spectra selected as clear (analysable by KLIMA), no clouds occurred, on the contrary they occurred in case of spectra selected as cloudy (not analysable by KLIMA). In Fig. 1 are shown two FIRMOS spectra selected by the algorithm as clear sky (left panel) and the backscattering lidar signal (right panel) confirms the absence of clouds.

In Fig. B two FIRMOS spectra flagged as cloudy are shown. Here as well, the backscattering lidar signal confirms the occurrence of clouds.

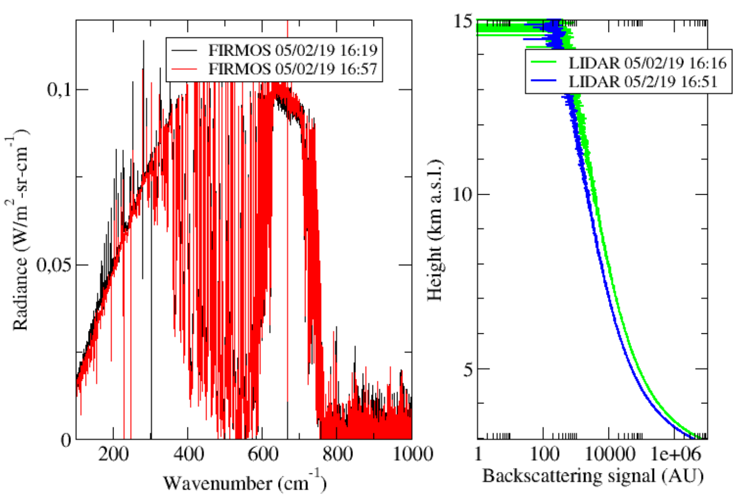


Fig. A. Left panel: FIRMOS spectra selected in the presence of clear sky. Right panel: backscattering lidar signal corresponding to the FIRMOS measurements time.

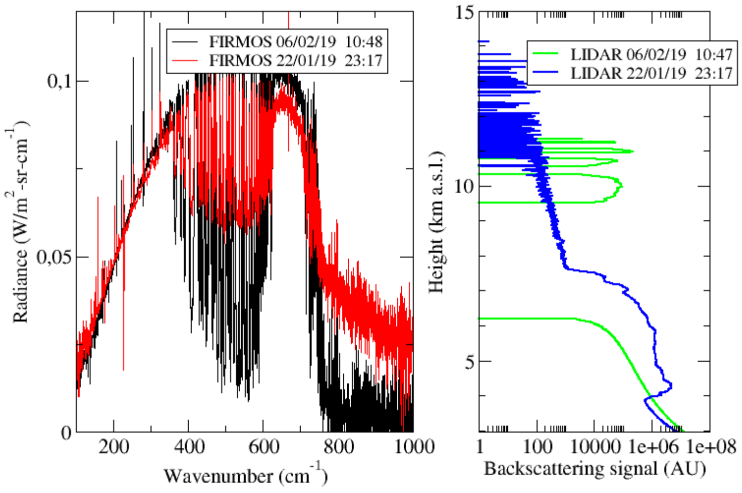


Fig. B. Left panel: FIRMOS spectra selected in the presence of clouds. Right panel: backscattering lidar signal corresponding to the FIRMOS measurements time.

Line 288: it was not clear if the uncertainty used in the retrievals was the NESR or the sum of the NESR and the CalErr. Please clarify this in the text. If the latter, then the chi-squared term being less than 1 could be due to the very conservative estimate of the thermistor error in the blackbodies (stated on line 230).

We, added the following in the text:

“total error (quadratic sum of the NESR and calibration error)”.

In the figure 12 and 13 (now 14 and 15) we demonstrated that the mean of the residuals reproduces well the calibration error, while the SD of the residuals is smaller than the NESR, this is the reason we are convinced NESR and not the calibration error is overestimated.

Moreover, if we set the calibration error, to zero the retrieval cannot reach the chisquare=1 because the calibration error is small with respect to the NESR.

Line 293: the mean residual also will contain any bias error in the forward model (not only instrument calibration error).

That is right, we added in the text:

“or forward model error”

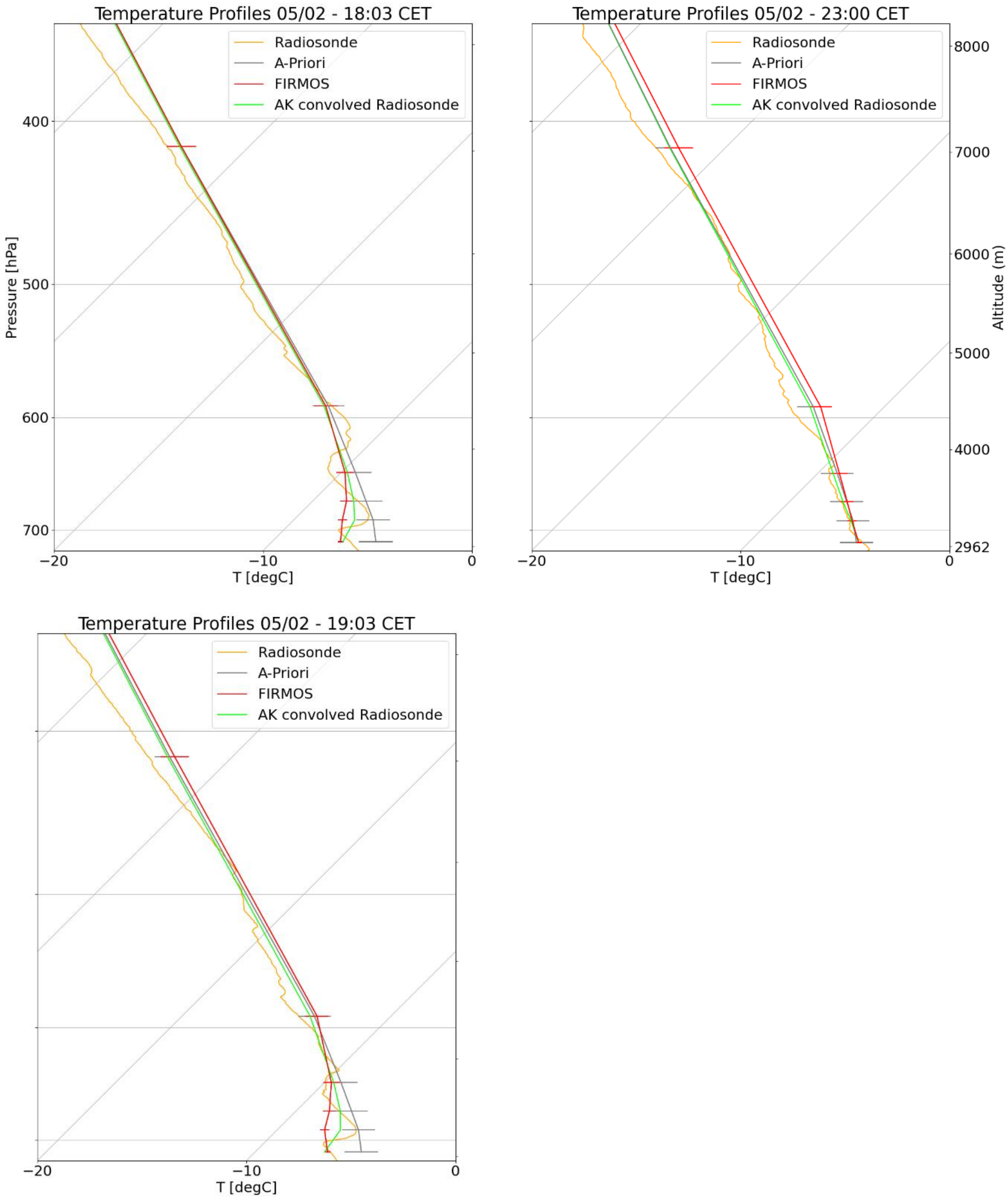
For the two comparisons in Fig 15: it would be nice to include the integrated water vapor (IWV) amount for the two cases. Also, for line 309, the authors suggest that the DOF depends on the surface water vapor content, but it is really dependent on IWV? Turner and Löhnert (JAMC, 2014) showed that the DOF in the water vapor retrieval using AERI observations in the 538-588 cm-1 region depends on IWV.

The value of IWV was added in the figure (now Fig. 17)

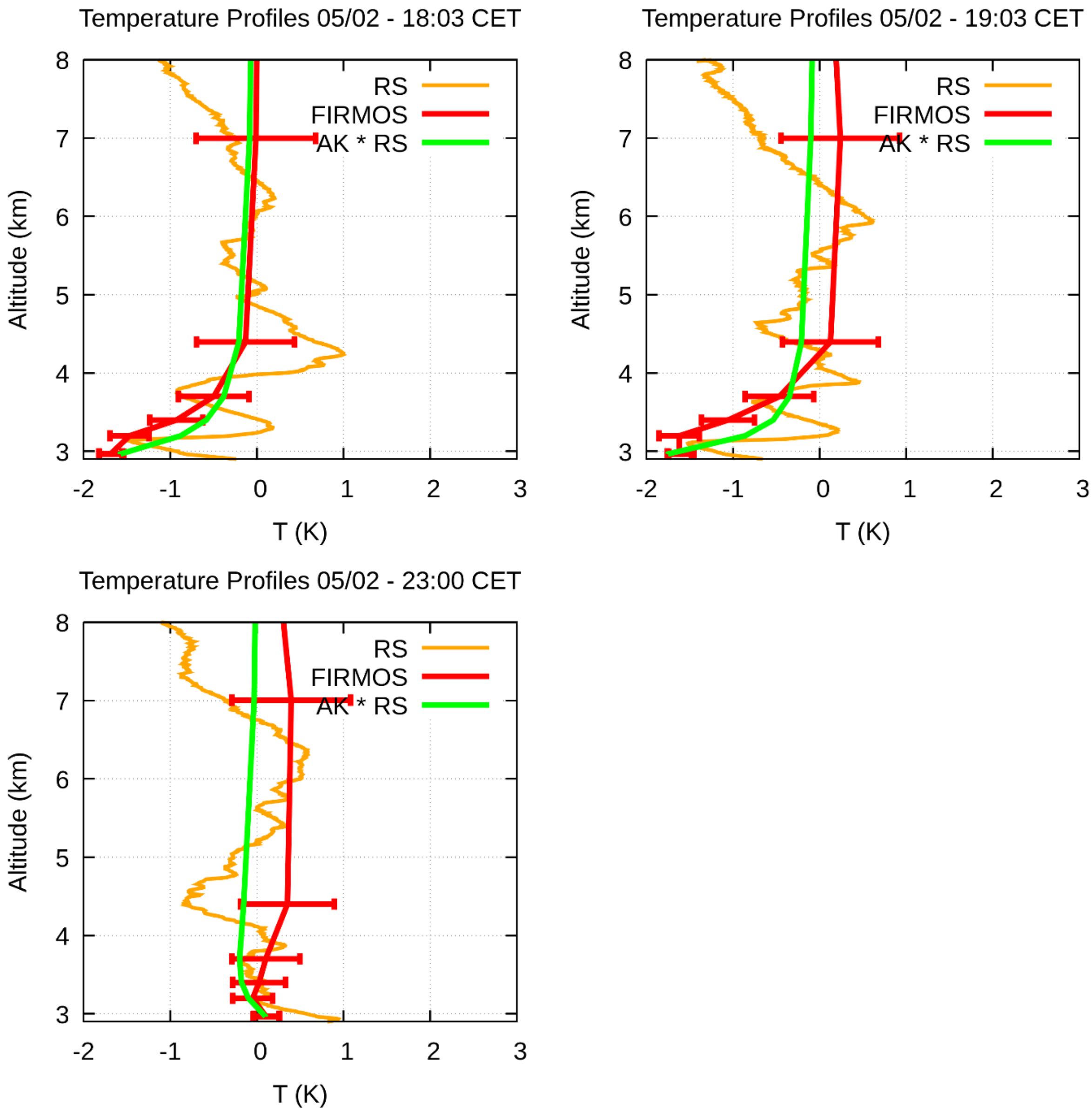
We also added a reference to Turner and and Löhnert 2014.

Figure 20: please replot using a skew-T approach, so that differences of a few degrees can be more easily identified and quantified.

The following are the T profiles using skew-T



Referee2 suggested to plot the T profiles as difference. The following plots show T profiles as difference wrt the a-priori ($\hat{x} - x_a$).



We think the difference plots are more intuitive, and clearly show the FIRMOS T profile and the AK convolved Radiosonde are within the retrieval error.

Fig 23 and resulting analysis: this is pretty unsatisfying. I realize the purpose is to show that the FIRMOS is capturing the evolution of the event well, but the very coarse resolution of the ERA5 data in a mountainous region is totally inadequate to the task. I highly recommend that the comparison be made against higher-resolution NWP output, such as the (order)2-km resolution ICON data from the DWD. And that the figure include a subpanel showing the bias and RMS difference between the NWP model and the FIRMOS.

We contacted DWD enquiring for data for January and February 2019, unfortunately they only provide real-time NWP data. We also followed their recommendation to contact the Fraunhofer Institute IEE for historical data, unfortunately they only have data from 2021.

We removed the section 4.2.4 (comparison with ERA5). As suggested by Ref2 we added the time-series of H2O to the end of section 4.1 (retrieval of geophysical parameters) the time resolution was increased (10 minutes) we deem the plot gives an overview of the dataset in two periods of the campaign when the acquisitions were sufficiently dense and continuous.

Question: Why did the authors not perform KLIMA retrievals using the E-AERI spectra, and then compare the retrievals from the E-AERI with the FIRMOS? This seems like it would be a relatively simple comparison and include a lot more data (e.g., there seems to be hundreds of points in Fig 22), and open up an interesting discussion because their spectral differences between the two instruments.

To share the efforts between the different groups involved in the paper it was preferred to retrieve IWV with IMK algorithm.

In view of the additional analysis suggested by the referee using the Tobin 2006 technique added to Sect 4.2.3 we believe we have shown the two instruments agree very well in the H2O spectral band.