Response to Reviewer comments on *"IASI spectral radiance performance validation: case study assessment from the JAIVEx field campaign"* by A. M. Larar et al.

**Reviewer #1:**

General comments

The variety of comparisons and related presentations the authors provide is, in many parts, unnecessary long, especially as far as the number of figures is concerned. There are too many panels in single figures, with a strong abuse of colour and space.

[We agree that some of the comparisons discussed in the manuscript could be presented with less text and figures for readers more familiar with infrared spectral radiance measurements, instrumentation, and validation; however, we feel that the more elaborate coverage approach implemented herein will be beneficial to a significant number of ACP readers.]

In general, the paper gives much less than it promises. In fact, although the title says *IASI spectral radiance performance validation..*, the validation is quantitatively performed only in terms of band-averaged portion of the spectrum, and, therefore, not for spectrally resolved radiances.

[This is not exactly true. In fact, comparisons are done (and shown in the plots) at the highest possible, as-measured spectral resolution. The metric for quantitative assessment in reported differences is mean difference of spectral radiance across the bands, and therefore distinctly different from "band-averaged" comparisons.]

… in consideration of the fact that nothing is said for the CO2 v2 band, which is fundamental for temperature sounding and for which we paid both for AIRS and IASI. I guess that this is so because NAST-I was flown at 15-17 km, therefore missing all the intense emission from the stratosphere to the top, which, in turn, makes it meaningless any attempt of direct comparison for the CO2 and Ozone band, as well. These limitations should be explicitly stated in the introduction and conclusion sections, where it should be stressed that a direct comparison is only possible for those portion of the Earth spectrum, which are driven from tropospheric emission: viz., atmospheric windows and the H2O v2 band (for this last case because water vapour is mostly confined to the troposphere).

[It is correct that atmospheric contributions between the aircraft and satellite are not included in nadir-looking airborne sensors used in this study and, yes, that is why comparisons shown in the manuscript are limited to those spectral regions having insignificant absorption/emission above the aircraft altitude (i.e. $H_2O$ band and window regions). Comparisons in other spectral regions can also be performed using aircraft data, however, one must also include a modeling component which introduces its own uncertainties and was therefore not included within analysis presented herein. The unique contribution obtainable from airborne sensors is discussed in the introduction, conclusion, as well as other portions of the manuscript.]

Specific Comments:

1) Introduction section. After having established the strength of their high-altitude aircraft validation approach, the authors should provide a fair discussion on the possible drawbacks, including different potions of the atmosphere sensed with the airplane and satellite instruments, different radiometric and spectral characteristics of the instruments, time and space co-location, different Field of View geometry and so on.
[This is already addressed in the manuscript and in the previous response above. Note that the different spectral characteristics of the instruments are accounted for within the comparisons presented.]

2) Introduction section, page 10195, line 12. The Blumstein's reference to IASI is not the most appropriate here. IASI has a long history: the activities on IASI began around 1992. In 1993 Cayla presented the first general overview of the instrument (*Cayla, F.-R., 1993: IASI infrared interferometer for operations and research, in: Chedin, A., Chahine, M.T., Scott, N.A. (Eds.), High Spectral Resolution Infrared Remote Sensing for Earth's Weather and Climate Studies. NATO ASI Series, **I 9**, Springer Verlag, Berlin-Heidelberg, 9-19*). Further details about contribution through the years to IASI can be found at the web site http://smsc.cnes.fr/IASI/Fr/A_publications.htm . I do not understand why the authors make reference to conference papers when there are appropriate IASI presentations published in peer reviewed journals. If the problem is that authors have to acknowledge CNES and EUMETSAT this can be done (as indeed they did!) in the acknowledgment section.

[More peer-reviewed references pertaining to IASI instrument heritage have been added as suggested by this reviewer.]

3) Section 3, page 10200, line 12. It is important to be clear about which LBLRTM version the authors have used, including the version for the continuum absorption of H2O.

[We agree. The specific version of LBLRTM, along with a reference, has been added to the manuscript text as suggested.]

4) Section 4.1 and Fig. 4 page 10202. Apparently this case is only shown just to make the point that a comparison with simulations is not accurate enough for the purpose of radiance validation. This is stated in a way which I found a bit naive. Our ability to make a proper use of the IASI radiance ultimately rest on our ability to produce accurate synthetic IASI spectral radiance. Should the authors be right, we have to conclude that it has a been a tremendous waste of money to fly IASI. I know that this is not the real feeling of the authors, since they have a quite different attitude when discussing their contribution to IASI *retrieval* capability in other papers in this same special issue dedicated to IASI. Science should be objective and should not depend on the specific (subjective) context. Furthermore, the authors use just one spectrum in Fig. 4, which cannot be considered as a *significant statistics*. The comparison would be much more informative by including IASI error bars (radiometric noise). Please revise bias and rms figures provided in the body of the Fig. 4b. I do not believe that the RMS difference is

9.1 K, in the case of the retrieval. This is inconsistent with the curves shown in figure and the fact that IASI NEDT in this spectral interval is of order 0.1 to 0.2 K at 280 K. Even for the case of a standard atmosphere, the RMS difference of 210.1 K (sic!) is unbelievable. If the authors want to insist on this comparison, they should show the spectral residual (IASI-Calculation) together with the ± σ interval. Then, the comparison IASI vs. retrieval would be enough. Finally, if the authors did well the calculations shown in Fig. 4b (and I insist that I have problems with the RMS difference), then a mean difference of 0.21 K across the band is not so much different from the equivalent values for (NAST-I-IASI) and (NAST-I-AIRS), namely 0.08 K and 0.11 K, respectively, they quote in Fig. 13. The order of magnitude is the same; therefore the claim of the authors that they need to fly an interferometer for a better validation of spectral radiance is not sustained from their calculations themselves. To fly an aircraft at 15 Km with a series of expensive instrumentations and gain only a factor of about 2 in bias seems to me really a waste of technology. Finally, figure 4a is not informative and can be removed. Why so much color to indicate a point on a map!

[The point being made by the referenced figure and corresponding text is that spectral radiance can be verified most accurately by direct comparison with similar measurements rather than simulations. This is demonstrated with the select example shown in Figure 4 without requiring a statistically-significant representation. The "forward model uncertainties" which produce errors in the mean differences across the band shown are systematic in nature making comparison with IASI (random) noise uninformative. The reviewer is very correct in pointing out the unbelievable "RMS differences." This is a typo and really should be "RSS differences" as was actually calculated, and has been corrected in this figure.

It is certainly agreed that accurate simulation capability is essential, and doable, to fully exploit the IASI data for NWP and other geophysical applications. The comparison with retrieved profile shown is only included for illustrative purposes, i.e. even to point out that an independent simulation (i.e. different forward model) using best estimate for atmospheric "truth" can still produce errors larger than directly-measured radiance alternatives due to modeling uncertainties and differences. Usage of the same forward model that produced the retrieval would have produced mean radiance differences on the order of hundredths of a degree, but would not represent a comparison to independent "truth". Figure 4a is included to show location of compared spectra within IASI scene, i.e. relative to nadir track, clear air and clouds, land and water surfaces, to illustrate spatial and geophysical context.]

5) Section 4.2, page. 10203 (Intra-platform comparison). Please remove the two figures 5 and 6 and related discussion. IASI is a high spectral resolution infrared spectrometer. These figures and related elaborations are much more suited for a report. Here, they only delay the most important comparisons: NAST-I vs. IASI, IASI vs. AIRS and AIRS vs. NAST-I.

[While we agree that figures subsequent to 5 and 6 are most important since they are the ones illustrating comparisons of high spectral resolution radiance, Figures 5 & 6 also serve an important purpose as noted in the text. Specifically, they are included as a demonstration of intra-platform radiometric consistency; this not only provides higher confidence in subsequent inter-platform comparison results but, also, serves as a reference which could be useful for extrapolating performance assessments beyond the airborne field campaign domain.]

6) Section 4.2 (NAST-I vs. S-HIS). The qualitative comparison shown in Fig. 7 is really non informative. First, how many spectra are you averaging? Second, for this case you are not limited by altitude considerations, since I assume that S-HIS and NAST-I were flown at the same altitude. Then please show also a comparison for the spectral interval 640 to 800 cm-1. Please, show spectral residuals (NAST-I – S-HIS) together with the ± σ interval, properly scaled in case more spectra are averaged.

[The comparison of NAST-I vs S-HIS shown in Figure 7 serves as an intra-platform comparison to establish internal consistency of NAST-I prior to subsequent usage of NAST-I for the inter-platform comparisons. Averaging is performed for collocated scenes throughout this flight to illustrate stability over the entire flight and provide the best comparison of radiometric accuracy. These sensors are on the same aircraft and could be compared for their entire spectral extent, however, 640 to 800 $cm^{-1}$ (and other spectral regions not shown) have intentionally not been included since they are not being used in subsequent comparisons between IASI and AIRS and therefore serve no purpose in the present study.]

7) Fig. 8 and related discussion. This figure has a poor meaning without a discussion on the absolute accuracy of NAST-I and S-HIS. Which is validating which here? Which is more accurate and stable? The authors need here to explain why does the bias change sign by moving from long to short waves? This could be a clue for a miscalibration of NAST-I or S-HIS. Finally I do not find informative to inter-compare NAST-I and S-HIS applying a so heavy smoothing such as that applied by the authors, which is a box car of 10 cm-1.

[Figure 8 is included as an intra-platform comparison of NAST-I and S-HIS illustrating internal consistency of NAST-I radiometric accuracy with S-HIS prior to subsequent comparisons of NAST-I and IASI and AIRS. As noted in the text, S-HIS has undergone extensive radiometric accuracy traceability testing with NIST (National Institute of Science and Technology) standards so this comparison does provide some confidence in NAST-I radiometric performance in an absolute sense. Since a concrete measure of "truth" in flight is not known with full certainty, it is the consistency of two independently-calibrated sensors that is most significant and informative for subsequent comparisons. The subtle differences (on the order of hundredths of a degree) between NAST-I and S-HIS shown in Figure 8 could very well be indicative of calibration differences between these airborne sensors, as noted by this reviewer, however these are insignificant relative to differences noted between NAST-I and the spacecraft sensors (i.e. an order of magnitude smaller). Note that the longwave and shortwave bands are detected on different detectors so a sign change in bias differences is not beyond expectation. As noted in the text, the relatively large 10 $cm^{-1}$ boxcar smoothing is applied in Figure 8 to better facilitate radiometric calibration inter-comparison between these sensors (i.e. to make completely independent from any possible spectral calibration / sampling differences).]

8) Figures 9 and 10. This case is left to the reader visual interpretation; so that I think it is unnecessary. Furthermore, it could be also dangerous, since figure 10c shows a marked sinc-beat (lower corner on the right-hand-side), which could be the result of a less than accurate IASI calibration. Unless you are able to provide a valid explanation for this spurious behavior, please refrain from presenting it.

[Figures 9 and 10 serve to demonstrate the importance of spatial and temporal coincidence when performing spectral radiance comparisons. Specifically, Figure 9 sets the stage by illustrating the temporal variation of the geophysical field being compared along with cross-section logistics. Figure 10 then shows the resultant water vapor band spectral radiance cross-sections along the Figure 9 – defined cross-sections. These figures exemplify that time coincidence becomes more important than space coincidence for a time-varying geophysical field sampled ~ 3 hours apart, except for regions having high spatial variability (i.e. over land)—an important point for this case study. The images in Figure 10 are formulated using the IASI and AIRS data with as-measured spectral and spatial characteristics that are then uniformly sampled to the same spectral and spatial image grid. To confirm that spectral and spatial sampling differences were not contributing to interpretation errors, these same radiance cross-sections were heavily smoothed over their sampling domains to ensure no sampling-induced artifacts would remain; while this experiment produced different looking data products, the same spatial and temporal coincidence conclusions were evident. This can also be demonstrated via cross-sections of band averages as illustrated in the figure included below this paragraph (Figure A). This response addresses the concern expressed by the reviewer regarding a feature in Figure 10-c, which is the result of a narrow geophysical water vapor feature observed with relatively sparse IASI spatial sampling which can cause artifacts upon interpolation for imaging. Note that this same feature is not present in the IASA cross-section in Figure 10-a, having the same IASA spectral calibration, due to different spatial features in the geophysical field—therefore it is not indicative of an IASI calibration issue.]

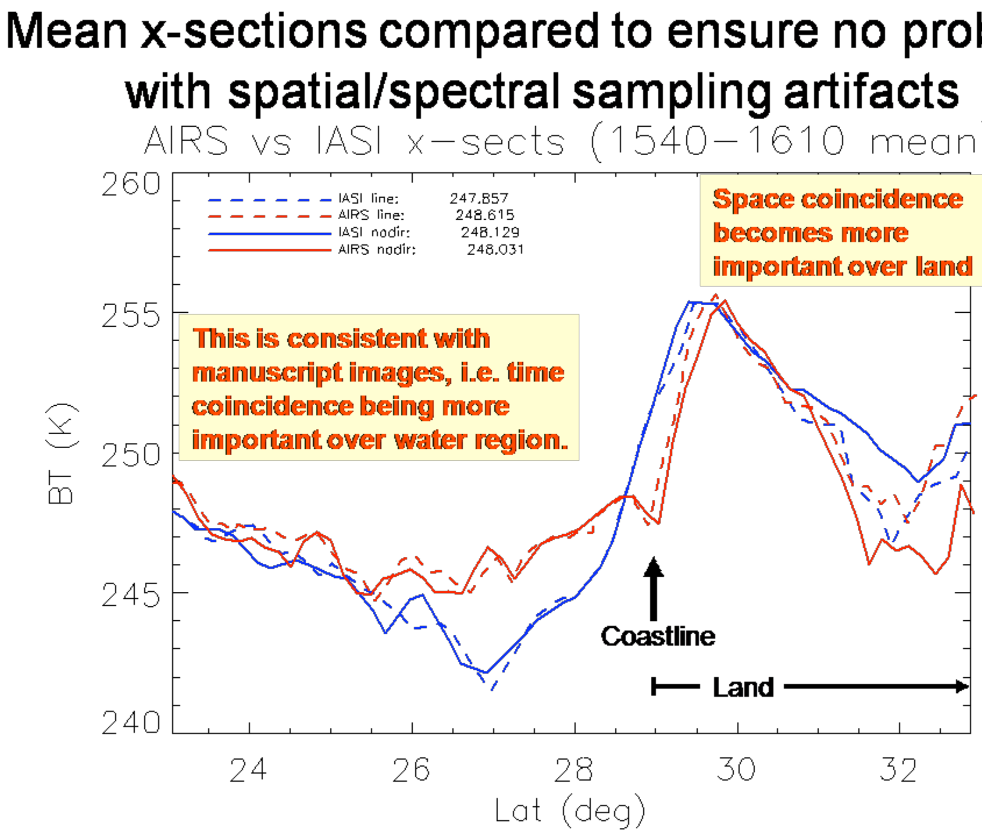


**Figure A. Figure in support of above comments.**

9) Figure 11. The case made in this figure is quite obvious to me. It is quite obvious that scene variability is the most critical issue when comparing satellite vs. satellite. Therefore Fig. 11 could be removed and save space to explain and discuss the most important section 4.3.b

[The purpose of this figure is to show that direct comparison of spatially-coincident but not temporally-coincident measurements is not good enough for this case study day (i.e. having significant scene variability over the ~ 3 hours between satellite overpass events). So it is not included to say that scene variability is in general the most critical issue when comparing satellite versus satellite but, rather, to quantify the scene evolution for this particular case and show it is too large for such direct comparisons (i.e. much larger than potential instrument differences being examined, and could be distinctly different from other cases closer in time and/or having less scene evolution). Since this is the first and only occurrence of this type of comparison in the manuscript, we feel it does serve a purpose and is not obvious to the average reader.]

10) Section 4.3.b (Aircraft vs. Spacecraft) on page 10205. First, please explain what you did (if you did something) to match the different IFOV of NAST-I, IASI and AIRS. Did you consider any averaging along the horizontal? What are you showing in Figs. 12 to 14 is a single spectrum or are you averaging more spectra? If yes, how many? Furthermore, I

am not pleased with a simple band-averaged consistency. A more quantitative approach should show spectral residuals and related error bars.

[Near-nadir NAST-I IFOVs (~ 2 km) are spatially-collocated within the larger IASI and AIRS IFOVs (12 and 14 km, respectively). As discussed in the text, spectra shown in Figures 12-14 are for single spacecraft sensor IFOVs relative to combined (i.e. averaged) near-nadir NAST-I observations coincident in space in time. Since only near-nadir NAST-I spectra are combined, the average is primarily along the aircraft flight track direction. The exact number of NAST-I spectra averaged varies depending upon sampling specifics for each collocation. Spectral residuals are purposely not included since these figures include spectra of different spectral resolutions and sampling positions. Alternatively, mean differences over the bands are given for the cases where NAST-I is degraded in resolution and sampling to be consistent to what it is being compared (i.e. IASI or AIRS). Also, since radiometric accuracy is the focus, such mean differences are sufficient to show any existing systematic offsets.]

11) Section 4.3.c. twelve figures to explain a simple linear fit are really impressive! Please shorten the number of figures in this section.

[Section 4.3-c contains 5 figures not twelve, i.e. Figures 15-19. Without knowing which Figure(s) the reviewer is suggesting to remove it is hard to respond, however, we can say that we think all Figures presented serve a purpose sufficient to justify their inclusion. Furthermore, we feel it is a good thing that a simple linear fit is sufficient for this case (which, by the way, is not valid for a similar analysis done for the water vapor region) and believe it in no way lessens the need for properly presenting analysis methodology and results.]

12) Summary and Conclusion section. The second paragraph of this section contains bold statements that need to be under-emphasized. To me, the best and cheap mean to have SI-traceable measurements is to put from now on, onboard satellites, common-based-technology calibration black bodies. All in all this paper shows that Europe and USA share the same state-of-art black-body technology (or more likely the same seller). The same conclusion of the authors could have been arrived at by a direct comparison of NAST-I, AIRS and IASI black-bodies. In fact, the paper does not say much about the spectral consistency and quality among the various instruments, since it limits itself to consider only band-averaged quantities. In the end, the methodology set up by the authors is a very expensive way to say that the IASI black-body does work. The authors should fairly state that their method has pros and cons and that at moment the spectral consistency is better analyzed by direct comparison with simulations, while because of possible bias in spectroscopy and forward modeling the overall radiometric consistency is better assessed trough a direct comparison with aircraft instrumentation. This is a fair compromise and I hope it may help.

[Statements in second paragraph have been underemphasized as suggested by the reviewer.

Future missions aiming to improve climate quality measurements and inter-calibration capability of satellite sensors are seeking to fly SI-traceable blackbodies with in-flight verification capability. At present, however, it is very difficult to make convincing arguments of how in-flight systems, e.g., NAST-I, AIRS, and IASI, inter-compare post-launch based only on ground-based blackbody characterizations. An independent in-flight-coincident measure, such as that provided by the airborne NAST-I, is needed to assess system-level post-launch performance. Spectral performance is addressed in the paper but not as explicitly as radiometric calibration performance. As the reviewer point out, simulations are a very good and inexpensive way to quickly assess spectral calibration performance, especially in spectral regions where line parameters are known quite well. Airborne sensors can also be helpful in this area and become particularly useful when investigating anomalous performance, e.g., the impact of sub-pixel scene non-uniformities on the instrument line shape (ILS). The benefits obtained from including airborne sensors in satellite system cal/val are addressed in the manuscript.]

Technical corrections:

1. Figure 4. *IASI measured spectra* should read *IASI measured spectrum*, since just one single spectrum is shown here.

[Suggested correction has been made.]

**Reviewer #2:**

- I found some of the plots are not easy to understand (see earlier comments, eg I found Fig 16 quite unreadable) ;

[Figure 16 has been made larger.]

- Definition of spectral resolution is not rigorous. Page 6 : the spectral resolution quoted here is .25 cm-1 for the NAST-I interferometer. Apozided, or unapodized? 1/OPD or HWHM? Page 7: For IASI a value of 0.25 cm-1 is quoted for spectral sampling. The spectral resolution is 0.5, Gaussian apodized. Because the paper deals with radiance validation I think that the same definition should be used for all instruments (and should be explicit in the paper). A table comparing all spectral resolutions for the instruments quoted in this paper, along with proper definition, would be welcome.

[The NAST-I spectral resolution (1/2*OPD) is 0.25 $cm^{-1}$ (unapodized). The IASI spectral resolution is 0.5 cm-1 (Gaussian apodized) and utilized on the 0.25 cm-1 (unapodized spectral resolution) spectral scale as reported in the L1C data files. The spectra plots in the Figures include both as-measured and spectrally-equivalent spectral resolution and sampling, whereas the noted quantitative mean differences across bands are reported uniquely for spectrally-same comparisons. This has been clarified in the manuscript text ensuring such spectral resolution information is explicit for all included instruments.]

- It would be very helpful to compare the findings of this paper in terms of radiance agreement, as compared to the initial instrumental radiance requirements provided in the science plans for both IASI and AIRS instruments, eg for radiometric stability. Does it match with the expected values?

[IASI and AIRS have exceeded their originally-specified requirements for radiometric accuracy and precision, as demonstrated by the closeness of comparisons presented within this manuscript, and are therefore raising the bar of expectations for such future sensors. A similar statement is added within the concluding remarks.]

- The study relies only on a few cases. Can the conclusion be extended to a longer period/ larger scale?

[Results presented in this paper are for a single case study day within a single field campaign. Correspondingly, some measure or assumption regarding satellite sensor radiometric stability is needed to properly extrapolate results to be representative over longer time periods and larger spatial scales. This is the subject of further analysis and reporting and can be accomplished through increasing the temporal and geographical diversity of inter-comparisons by including more case study days within an individual campaign, data from more campaigns, and incorporating data from other measurement and model systems (e.g. NWP analysis fields); this approach can exploit the long time series and global overlap of model fields and other coincident satellite observations while utilizing how all systems inter-compare with the aircraft sensor reference measurements obtained within the limited number of field campaign observations. Similar statements have been included within the manuscript concluding remarks.]