Response to referee #2

We are thankful to the reviewer for his/her useful comments that will contribute to greatly improve the manuscript. In the following, the reviewer's comments are in black and our response is in red.

This paper presented an assessment of IIR calibration stability using MODIS and a RTM as references. The methods were reasonably well explained; the findings (MODIS29 drift, IIR2 warm bias at cold scene, IIR day/night calibration discrepancy, and seasonal variation) are significant; and the manuscript is well organized and reasonably well written. I recommend its publication with minor revisions.

Major Comments

1. Please consult a native English speaker or professional to improve the English.

Response

The revised manuscript will be proofread by a native English speaker.

2. Section 3.2.3 and Figure 2: Please plot the ISRF for SEVIRI on Meteosat-9/10 (e.g., in blue with different line styles). Since these were replaced with that for SEVIRI on Meteosat-8 in RTM, readers need to know how similar they are to SEVIRI on Meteosat-8, and different they are from MODIS and IIR. Probably expand Table 2 as well.

Response

Figure 2 will be modified as suggested. Table 2 will include central wavenumbers for SEVIRI on Meteosat 8, 9, and 10.

Changes in manuscript

New Fig. 2:

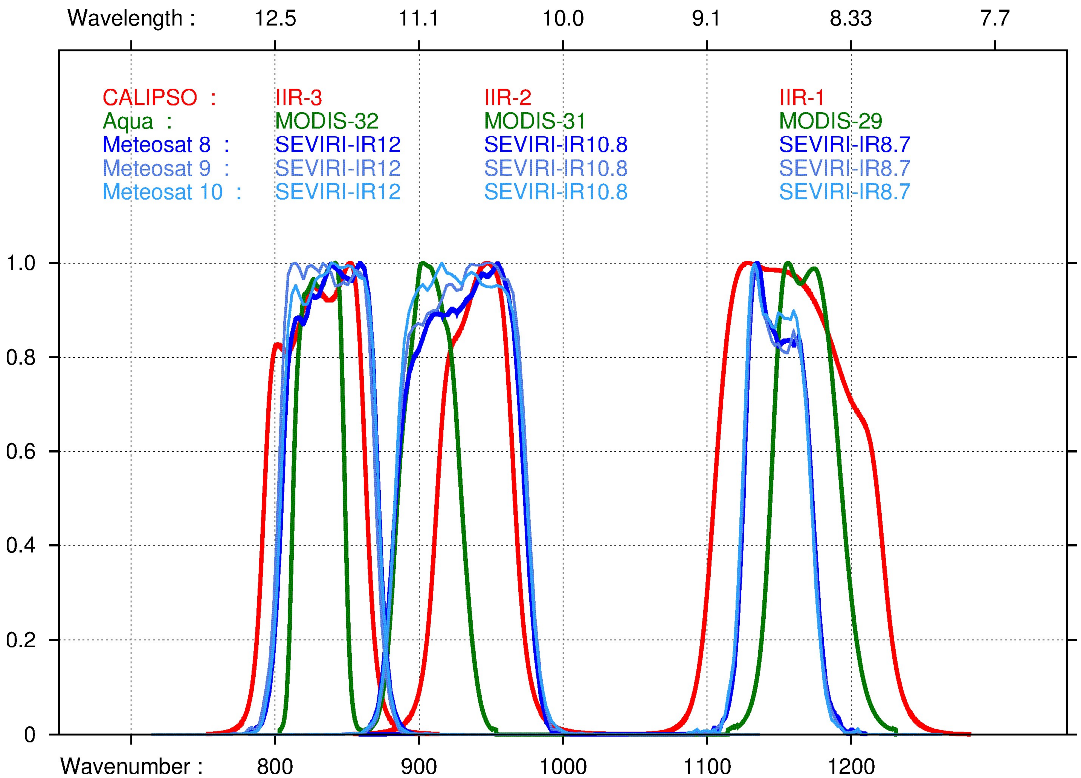

Figure 2: IIR/CALIPSO (red), MODIS/Aqua (green) and SEVIRI/Meteosat 8, 9, 10 (navy blue, medium blue, light blue) Instrument Spectral Response Functions against wavelength in microns (top X-axis) and wavenumber in cm^{-1} (bottom X-axis).

New Table2:

	Channel/wavelength/ wavenumber	Channel/wavelength/ wavenumber	Channel/wavelength/ wavenumber
IIR	#1: 8.635 µm 1158.4 cm^{-1}	#2: 10.644 µm 939.9 cm^{-1}	#3: 12.096 µm 829.1 cm^{-1}
MODIS Aqua	#29: 8.553 µm 1169.3 cm^{-1}	#31:11.025 µm 907.6 cm^{-1}	#32: 12.044 µm 830.8 cm^{-1}
SEVIRI Meteosat 8 Meteosat 9 Meteosat 10	IR8.7 1148.7 cm^{-1} 1148.2 cm^{-1} 1147.7 cm^{-1}	IR10.8 929.3 cm^{-1} 930.1 cm^{-1} 928.7 cm^{-1}	IR12 838.7 cm^{-1} 835.8 cm^{-1} 838.0 cm^{-1}

3. Section 3.2.3: Presumably all types of profile in an air mass are represented in TIGR, but how about abundance? For example, if an air mass has three distinctive types of profile, 80% are Type I, 15% Type II, and 5% Type III. Would your "TIGR" have these three profiles for this air mass, and you compute biases for each of them and come up with an expected bias for this air mass? In that case do you use weighted (by abundance) or arithmetic mean? Or would your "TIGR" have 100 profiles – 80 Type I, 15 Type II, and 5 Type III?

Response

The TIGR climatological library includes 2311 atmospheres, which are sorted out in five air mass types. The tropical air mass type is composed of 872 atmospheres, mid-lat1 and mid-lat2 are composed of 388 and 354 atmospheres, respectively, and polar1 and polar2 are composed of 104 and 593 atmospheres, respectively. For instance, for the air mass type "tropical", which includes 872 atmospheres, the brightness temperatures are simulated for each for these 872 atmospheres. Thus, for the tropical air mass type, the mean TIGR_BTD and the associated standard deviation are then computed using the 872 simulated values and by using the same weight for each of the 872 atmospheres.

Changes in manuscript

Sect. 3.2.2 of the revised manuscript will read as follows. The last sentence has been added to better highlight the number of atmospheres included in each air mass type.

"The simulations have been conducted for the 2311 atmospheres of the TIGR climatological library (Chédin et al., 1985; Chevallier et al., 1998). The 2311 atmospheres are sorted out in five air mass types according to their virtual temperature profiles (Achard, 1991; Chédin et al., 1994), which are namely: 1) tropical, 2) mid-lat1 for temperate conditions, 3) mid-lat2 for cold temperate and summer polar conditions, 4) polar1 for very cold polar conditions, 5) polar2 for winter polar conditions. The tropical air mass type is composed of 872 atmospheres, mid-lat1 and mid-lat2 are composed of 388 and 354 atmospheres, respectively, and polar1 and polar2 are composed of 104 and 593 atmospheres, respectively.

The computation of the mean BTDs and of the associated standard deviations is clarified by modifying the beginning of Sect. 3.2.3 as follows:

"The 4A model fed by the TIGR atmospheres has been used to simulate the brightness temperatures of IIR and of the candidate companion channels. Each of the five TIGR air mass types includes one individual simulation for each individual atmosphere included in the air mass type (i. e. 872 simulations for the tropical type, 388 for mid-lat1, 354 for mid-lat2, 104 for polar-1, and 593 for polar-2). Each TIGR air mass type is then characterized through the mean BTDs between IIR and MODIS or SEVIRI channels and associated standard deviations derived from the individual simulations (hereafter "TIGR_BTD").

4. Section 4.2: Please elaborate on the procedure, for example why σ = 0.7 K? How many of BTDs (%) were rejected as spurious? Does that vary with season and latitude etc.? What are the "uniformity tests"? Do they lead to further rejection? Why does homogeneity matter at all?

Response

A “worst case” standard deviation σ has been computed by taking 0.4 K for TIGR-BTD (see Table 3), the IIR NedT specified before launch (0.5 K), and NedT = 0.1 K for MODIS and SEVIRI (see Table 1), yielding σ = 0.7 K. The “spurious” values are considered “unrealistic” due to the fact that the instruments presumably do not see the same scenes. By testing whether there is an indication that the instruments presumably do not see the same scenes, we want to prevent irrelevant comparisons to enter the statistics. The text was indeed not clear: there are no further uniformity tests and no further rejection.

The fraction of BTDs rejected as spurious is larger than 95% for warm scenes. This will be discussed in Sect. 5.2.1 (see response to comment 7). However, the fraction of BTDs rejected as spurious is smaller for the cold scenes, as low as 30% in the worst case. The reviewer’s question brought to our attention that the distributions are wider than anticipated for the cold scenes, which led us to revisit our assessment for the cold scenes (Sect. 5.2.2). Please see the response to comment 8.

Changes in manuscript

The end of Sect. 4.2 will read as follows:

“A “worst case” standard deviation σ has been computed by taking 0.4 K for TIGR-BTD, the IIR NedT specified before launch (0.5 K), and NedT = 0.1 K for MODIS and SEVIRI, yielding σ = 0.7 K. Using TIGR_BTDs corresponding to each latitude band, BTDs larger or smaller than TIGR_BTD ± 3σ (i. e. ± 2.1 K) are considered unrealistic values due to the fact that the instruments presumably do not see the same scenes. The statistics are computed after rejecting these unrealistic values.”

5. Section 4.3.2: You lost a lot of potential collocations by using only the IIR pixels within 5 km of ERA-I grid, and took risks to compare RTM simulate up to three hours before or after satellite observations. Could you comment on why you don’t interpolate the RTM results in time and space?

Response

We had to find a compromise between number of collocations and accuracy of the description of the atmospheric and surface state related to them. The factors impacting this number and this accuracy are both the distance and the time gap between the IIR pixel and the closest ERA-I grid point. For other previous studies (e.g. at the time of the NOAA/NASA Pathfinder Programme) we have made such time and space interpolations. For this study, which involves three different instruments onboard three different satellites with different optical paths and inherently different pixel sizes, we have made the assumption that the interpolations would inhomogeneously affect the description of the input data to the 4A model. We deliberately chose to use only the IIR pixels within 5 km of the ERA-I grid because the number of points entering the statistics seems sufficient - and no interpolation in time at all.

6. Section 4.3.3: Please explain what you mean by “outliers” and why “residuals found outside the initial monthly mean ± 2 twice the initial standard deviation” are the outliers you defined.

Response

In this section, simulations are compared directly with observations for clear sky scenes. Because the clear sky mask is not perfect, undetected cloudy pixels could be included in the initial statistics. Furthermore, as above (response to comment 4), we want to test whether there is an indication that the instruments presumably do not see the same scenes.

Changes in manuscript

The last sentence of Sect. 4.3.3 will read:

"This procedure is to prevent undetected cloudy pixels to enter the statistics, as well as situations for which the instruments presumably do not see the same scenes".

7. Section 5.2.1: I do not understand the logic behind the statement "Overall, these results demonstrate the good consistency between observed IIR-MODIS BTDs and simulated TIGR_BTDs, which confirms a posteriori that the thresholds chosen for the relative approach (TIGR_BTD ± 2.1 K) are appropriate and that the statistics are not biased." First, we do not know *a priori* "the good consistency" no matter how much we like that. Second, even if that is the right answer, you at least need to show you do not get that without the threshold before making this statement. And finally, how do you know the statistics are unbiased?

Response

We agree with the reviewer and we will revise the text.

Changes in manuscript

The beginning of Sect. 5.2.1 will read as follows:

"The first step of the analysis is to compare the mean BTDs from the relative approach with the simulated TIGR_BTDs (see Sect. 3.2.3). Because the TIGR simulations are for clear sky conditions, the comparisons are conducted for the warmest temperature range at each latitude band. Indeed, the clear sky scenes are a priori the warmest ones, although the warmest scenes could also contain clouds of weak absorption or thicker clouds located near the surface. After application of the (TIGR_BTD ± 2.1 K) thresholding introduced in Sect. 4.2, more than 95% of the pixels contribute to the statistics, and the mean BTDs are changed by less than 0.15 K, which confirms that the thresholding is appropriate."

Later in the same section, the last part of the following sentence will be deleted:

"Overall, these results demonstrate the good consistency between observed IIR-MODIS BTDs and simulated TIGR_BTDs.~~, which confirms a posteriori that the thresholds chosen for the 22 relative approach (TIGR_BTD ± 2.1 K) are appropriate and that the statistics are not biased.~~

8. Section 5.2.2: A few comments here:

a. "Quasi-identical brightness temperatures": That's true only if all channels on all instruments are accurately calibrated. I think this argument was used later, but it is neither explained explicitly nor in proper order such that it creates confusions.

b. Parallax can lead to increased standard deviation of BTDs, but I do not understand why random error in spatial collocation results in Tb bias either way.

c. In clear (warm Tb) tropical region, Tb(10.6μm) is typically a few degrees warmer than Tb(12μm) due to differential absorption (and re-emission) by atmospheric water vapor, and this DTB should decrease to nearly zero for deep convective clouds (cold Tb). So the green diamonds in Fig. 9 [Tb(10.6μm) – Tb(12μm)] seems reasonable to me, with two caveats. (1) It is 1 K instead of 0 K at 205 K, which you explained later. (2) They dipped for 275 K and 285 K mysteriously – could you explain why?

d. "… suggest a possible …": You probably meant "… is consistent with a …". To suggest you need to exclude other possibilities.

e. "Importantly, no issue has been identified at warm temperature for IIR2 when compared to MODIS31": It was stated earlier in this section that "IIR2-MODIS31 BTD varies from 0.51 K to 1 K" – is 0.51 K bias acceptable but 1 K not

Response – Changes in manuscript

Following the question of the reviewer regarding the fraction of BTDs rejected after application of the thresholding (see comment 4), Section 5.2.2 and Fig. 9 have been revised as follows:

"5.2.2 Cold scenes

As the scene temperature decreases, the clouds are denser and colder and the contribution from such absorbing clouds increases while the influence of the surface and near-surface atmosphere to which the IIR and MODIS window channels are the most sensitive for semi-transparent scenes decreases. The fraction of pixels retained after application of the thresholding described in Sect. 4.2 is found to decrease progressively from 95% for the warm scenes to 30%, the smallest fraction found at 200-210 K in the tropics for the IIR3-MODIS32 pair. This is partly due to the fact that the BTDs are distributed over a broader range of values than anticipated, so that the mean BTDs seen in Figs. 3 to 7 are possibly significantly, but systematically, biased for the cold scenes. For a better quantification as temperature decreases, IIR-MODIS BTDs are evaluated using the median values of the whole distributions, without thresholding. The median value is preferred to the mean value to minimize the impact of presumably unrealistic values. Median IIR1-MODIS29 (red), IIR2-MODIS31 (green), and IIR3-MODIS32 (blue) BTDs are shown in Fig. 9a by temperature range from 190-200 K to 280-290 K in the tropics during 2008 for representative months of the four seasons. Mean absolute deviations from the median value are between 2.5 and 5 K. For further evaluation, IIR and MODIS inter-channel BTDs have been analysed. Following the same approach as in Fig. 9a, Fig. 9b shows median IIR1-IIR3 (red, solid), IIR2-IIR3 (green, solid), MODIS29-MODIS32 (red, dashed), and MODIS31-MODIS32 (green, dashed) BTDs. Mean absolute deviations from the median value are here between 0.5 and 3 K. The variations of both IIR and MODIS inter-channel BTDs with temperature are due to the changing optical and microphysical properties of absorbing ice and water clouds located at various altitudes. The analysis of arches as seen in Fig. 9b is the essence of the well-known split-window technique for the retrieval of cloud microphysical properties (Inoue, 1985; Ackerman et al., 1990). IIR and MODIS arches are not of the same amplitude because IIR and MODIS measurements are spectrally different. BTDs at the coldest temperatures (190-200 K) are useful pieces of information regarding the calibration. Indeed, the coldest temperatures (190-200 K) correspond a priori to elevated dense ice-clouds, which, if they behave as blackbody sources, *should lead to quasi-identical brightness temperatures for all channels, assuming a negligible contribution from the atmosphere above the cloud. This is what we observe in Fig. 9b, where the IIR and MODIS inter-channel BTDs are close to zero, showing internal consistency of the calibration within each instrument. However, Fig. 9a shows that the IIR-MODIS BTDs are about 1.6 K on average at 190-200 K for the three pairs of channels. This indicates a warm bias of 1.6 K of IIR with respect to MODIS at 190-200 K. According to Fig. 9a, the warm bias seems to increase progressively as temperature decreases. An increasing IIR calibration bias as temperature decreases could be explained by a drift of the gain with respect to the gain measured in flight at warm temperature. Because IIR has only one sensor, observing such a similar bias for all three channels is conceivable.*

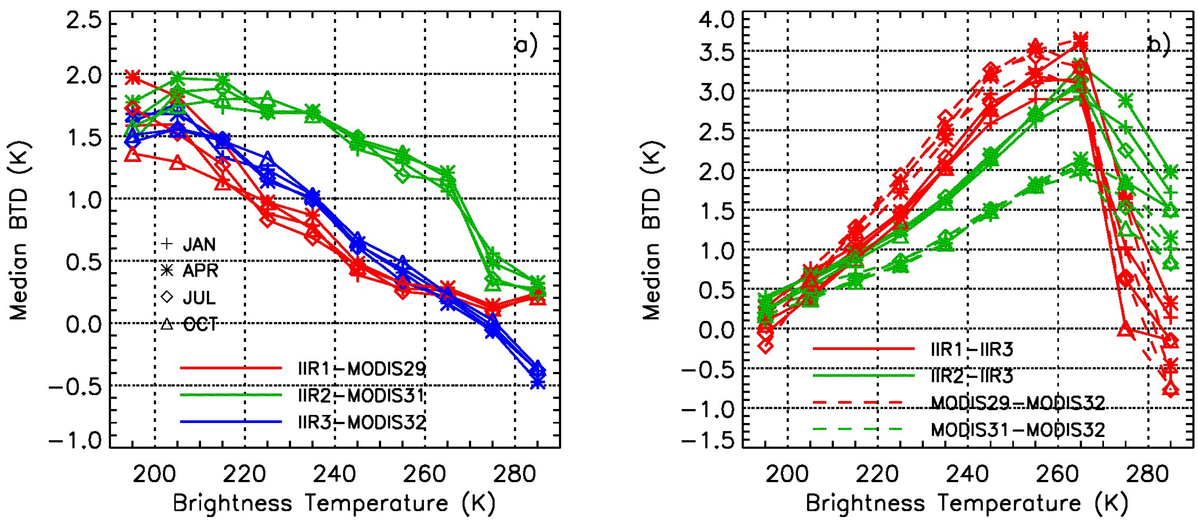

Figure 9: a) Median IIR1-MODIS29 (red), IIR2-MODIS31 (green), and IIR3-MODIS32 (blue) brightness temperature differences against brightness temperature. b) Median IIR1-IIR3 (red, solid), IIR2-IIR3 (green, solid), MODIS29-MODIS32 (red, dashed), and MODIS31-MODIS32 (green, dashed) brightness temperature differences against temperature. Plus sign: January 2008, star: April 2008, diamond: July 2008, triangle: October 2008. Latitude band: 30°S-30°N, ocean.

9. Section 5.2.4:
a. "This phenomenon could be partly explained by the more pronounced seasonal variations of the atmospheric and surface properties in the northern than in the southern latitude bands": Could you explain? This seems a calibration anomaly, of either or both instruments, that has little to do with the objects being observed.

Response

This seems indeed a calibration anomaly, as pointed out by the reviewer. The idea here was to recall that because the IIR-MODIS BTDs are not insensitive to the atmospheric and surface properties, we cannot rule out that the marked seasonal variations observed in the northern hemisphere could be due to the objects being observed. We agree that the statement is not strong. Because this sentence is actually not useful for the demonstration, and we decided to delete it.

Changes in manuscript

The third sentence of Sect. 5.2.4 will be removed.

b. "… for the warmest temperatures … a seasonal variability is clearly seen …": Does Fig. 7 show that the seasonal variation is stronger for colder scenes?

Response

We agree with the reviewer. We should have written "…at any scene temperature, including for the warmest temperatures…"

Changes in manuscript

The sentence at the beginning of Sect. 5.2.4 will read as follows:

"More specifically, it can be noted by comparing Figs. 4 and 5 on one hand (mid-latitudes) and Figs. 6 and 7 on the other hand (polar latitudes) that at any scene temperature, including for the warmest temperatures, with presumably smallest influence from clouds, a seasonal variability is clearly seen in the northern hemisphere but barely in the southern hemisphere."

Minor comments

1. Title: The SEVIRI/Meteosat part is not very important for this paper. It can be eliminated without impairing the main points of the paper, in fact that may enhance the main points. On the other hand, the usage of 4A is an important and integral part of that paper that is neglected from the title.

Response

The title will be modified by adding the notion of simulation.

Changes in manuscript

The title will read:

Long-term assessment of the CALIPSO Imaging Infrared Radiometer (IIR) calibration and stability through simulated and observed comparisons with MODIS/Aqua and SEVIRI/Meteosat

2. p.1, line 15: "…is quantitatively controlled …": perhaps you meant "evaluated", since there is no indications that you have done anything about the deficiencies you identified. Also p.2 line 20.

Response

p.1, line 15: "quantitatively controlled" will be replaced with "quantitatively evaluated" as suggested.

p.2, line 22: "controlled and characterized" will be replaced with "monitored and characterized".

3. P.1, line 19-21: "The pre-launch studies …" implies studies performed before launch, whereas what you described are studies using pre-launch data (ISRF). Similarly, by "… were selected before launch …" you may also mean that these pairs were selected based on pre-launch data.

Response

We confirm that we mean "pre-launch" studies performed before the CALIPSO launch.

4. p.1, line 27: Missing a "." After "since launch".

Response

Will be fixed.

5. P.2, line 14: "artefacts" should be "artifacts".

Response

Will be fixed.

6. P.2, line 15: "As soon as …" should be "As early as …". Also, you said NOAA/NASA Pathfinder started in mid- 1990's here but in line 20 you said Chedin described it ten years before then.

Response

Thank you for this comments.

Indeed, citing the (Chédin et al, 1985) reference page 12 line 25 was a mistake. It will be removed.

Changes in manuscript

a) The sentence page 2 line 15 will start as follows:
   "In the mid- 1990's, the NOAA/NASA…"
b) The sentence page 2 lines 23-25 will read:
   "The monitoring of observational and computational biases or trends over long periods of time started with the NOAA/NASA TOVS (Tiros Operational Vertical Sounder) Pathfinder Program (Scott et al., 1999)".

7.P.4, line 3: "when the solar elevation angle is less than -5°": at nadir earth surface or satellite?

Response- Changes in manuscript

We will write in the manuscript: *"...solar elevation angle at nadir earth surface…".*

8 P.4, line 12: "… in the *flowing* sub-sections."

Response- Changes in manuscript

Sect. 3.1, last sentence before Sect. 3.1.1 will read:

"They are detailed in the following sub-sections"

9 P.4, line 23, "… with nearly simultaneous measurements": why not "… 73 s earlier" that is shorter, simpler, and more precise?

Response

We will delete "with nearly simultaneous measurements", because it is not necessary in this sentence.

10 Fig.1: consider swapping the locations for "30S-0" and "83S-60S" to better show the symmetry between the two hemispheres, and the lack thereof in polar regions.

Response

We will follow this suggestion.

Changes in manuscript

New Fig. 1:

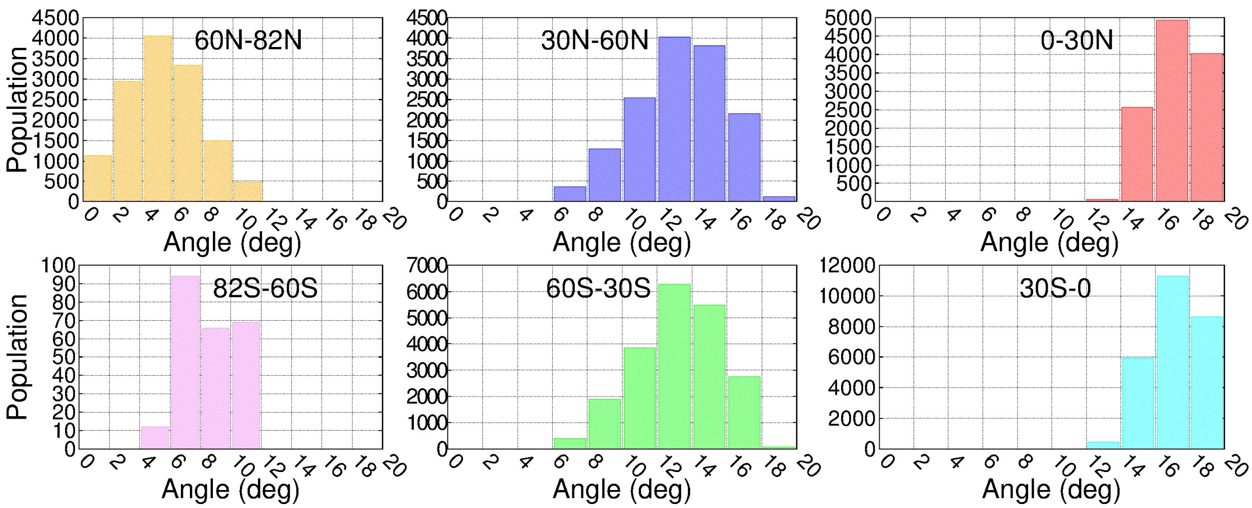

Figure 1: Histograms of MODIS viewing angles for clear sky pixels collocated with the IIR swath over ocean in July 2014: Top row from left to right: 60°N-82°N, 30°N-60°N, 0°-30°N; bottom row from left to right: 82°S-60°S, 60°S-30°S, and 30°S-0.

11 P.6, line 5: "*undesirable*".

Response

Will be fixed.

12 P.7, line 20: The un-explained acronym "REMAP" is confusing since none of the data seem to have been re-mapped.

Response

MODIS and SEVIRI data have been re-mapped on the IIR grid.

Changes in manuscript

To clarify, the second sentence of Sect. 4.1 will read:

"REMAP includes MODIS/Aqua and SEVIRI calibrated radiances collocated with the IIR Level 1B radiances and re-mapped on the IIR 69-km grid."

13. P.9, line 30: … the number of *samples* …

Response

Will be fixed.

14. P.10, line 33, "… to be the most fruitful …": Compared to what?

Response - Changes in manuscript

The first sentence of Sect. 5.2 will now read as follows:

"As seen in Sect. 5.1, because they are both on the A-Train and with no instrumental changes since CALIPSO launch, monitoring differences between IIR and MODIS/Aqua observations turns out to be a more fruitful approach for the assessment of the IIR calibration stability since launch than monitoring differences between IIR and SEVIRI."

15. P.11, line 8: … 0.25 K *for* …

Response

Will be fixed.

16. P.11, line 35: … inherently not *subject* to …

Response

This sentence is not in the revised version of Sect. 5.2.2.

17. P.11, line 36: IIR1-IIR*3* (red) …

Response

Will be fixed.

18. P.12, line 5: … *warm* bias of …

Response

Will be fixed.