

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.
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.
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?
4. Section 4.2: Please elaborate on the procedure, for example why  $\sigma = 0.7$  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?
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?
6. Section 4.3.3: Please explain what you mean by “outliers” and why “residuals found outside the initial monthly mean  $\pm 2$  twice the initial standard deviation” are the outliers you defined.
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\_BTDS  $\pm 2.1$  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?
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  $T_b$  bias either way.
  - c. In clear (warm  $T_b$ ) tropical region,  $T_b(10.6\mu\text{m})$  is typically a few degrees warmer than  $T_b(12\mu\text{m})$  due to differential absorption (and re-emission) by atmospheric water vapor, and this DTB should decrease to nearly zero for deep convective clouds (cold  $T_b$ ). So the green diamonds in Fig. 9 [ $T_b(10.6\mu\text{m}) - T_b(12\mu\text{m})$ ] 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
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.
  - b. “... for the warmest temperatures ... a seasonal variability is clearly seen ...”: Does Fig. 7 show that the seasonal variation is stronger for colder scenes?

#### 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.
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.
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.
4. p.1, line 27: Missing a “.” After “since launch”.
5. P.2, line 14: “artefacts” should be “artifacts”.

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.
7. P.4, line 3: “when the solar elevation angle is less than  $-5^{\circ}$ ”: at nadir earth surface or satellite?
8. P.4, line 12: “... in the *flowing* sub-sections.”
9. P.4, line 23, “... with nearly simultaneous measurements”: why not “... 73 s earlier” that is shorter, simpler, and more precise?
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.
11. P.6, line 5: “*undesirable*”.
12. P.7, line 20: The un-explained acronym “REMAP” is confusing since none of the data seem to have been re-mapped.
13. P.9, line 30: ... the number of *samples* ...
14. P.10, line 33, “... to be the most fruitful ...”: Compared to what?
15. P.11, line 8: ... 0.25 K *for* ...
16. P.11, line 35: ... inherently not *subject* to ...
17. P.11, line 36: IIR1-IIR3 (red) ...
18. P.12, line 5: ... *warm* bias of ...