We thank Review#1 for their comments. In red italic are our responses to each of the comments

1- This paper presents a methodology to retrieve skin temperature from IASI observations using a neural network approach. The channels and retrieval methods seem scientifically correct; however, I'm concerned with the calibration procedure. The authors have chosen skin temperature datasets from EUMETSAT and ERA5, which is acceptable. However, if I understood correctly, the NN is then trained using direct IASI observations. This is not an appropriate procedure since it will result in a NN that is biased towards the datasets used for the retrieval (EUMETSAT and ERA5). This is clear in Fig.5, where the comparison with the ERA5 show the lowest biases. The common procedure is to use a database of atmospheric profiles (from ERA5, for instance) together with a Radiative Transfer Model in order to obtain the best estimates of the relationship between top-of-atmosphere brightness temperatures and skin temperature. This is the procedure generally used in all currently available operational products. The calibration database is of very high importance in statistical retrieval methods. As such, although the methods are sound, the calibration database is not and I believe it will significantly impact the quality of your retrievals.

Yes, the calibration database is very important for a statistical procedure. Using a radiative transfer model to build this database is one of the ways we can construct it. We call this procedure a « physical » database. Another approach is to use an « empirical » database where real satellite observations are put in coincidence with direct observations (radiosondes, buoys, etc…). The authors of this paper have proposed long time ago a calibration dataset based on reanalysis outputs such as in the work done by Aires et al. 2005; Kolassa et al., 2013; and Rodriguez-Fernandez et al., 2015. They have shown that when doing this, it is possible to obtain a satellite retrieval that has no global bias with the reanalysis, but can have strong regional biases with it. The retrieval, even if trained with the reanalysis, does not reproduce the reanalysis, the time and spatial variations are driven by the satellite observations.

Such approach is also very interesting if we want to assimilate the retrieved parameter into the reanalysis. This approach has recently been implemented and tested at ECMWF (Rodriguez-Fernandez et al., 2019). The authors have shown that this procedure was an improvement over the classical inversion that used a radiative transfer. This is obviously not always the case but shows the pertinence of the approach. Please note that our procedure is not operationally used at ECMWF.

We addressed the Reviewer's comment, by adding the following sentence when discussing ANN in section 2.3 as follows:

*"*The feasibility of using ANN to T_{skin} retrieval has been shown for instance by Aires et al. (2002) for IASI, and has also been performed to tackle various problems in atmospheric remote sensing (Blackwell and Chen, 2009; Hadji-Lazaro et al., 1999; Whitburn et al., 2016; Van Damme et al., 2017). ***The retrieval, even if trained with the reanalysis, does not reproduce the reanalysis; the time and spatial variations are driven by the satellite observations (Aires et al. 2005; Kolassa et al., 2013; Rodriguez-Fernandez et al., 2015****).*"

2- There is also no reference to how the authors deal with emissivity. If I understood correctly, you simply disregard it, which means that there will possibly be strong discrepancies between different land covers. Please the discuss the implications of this simplification.

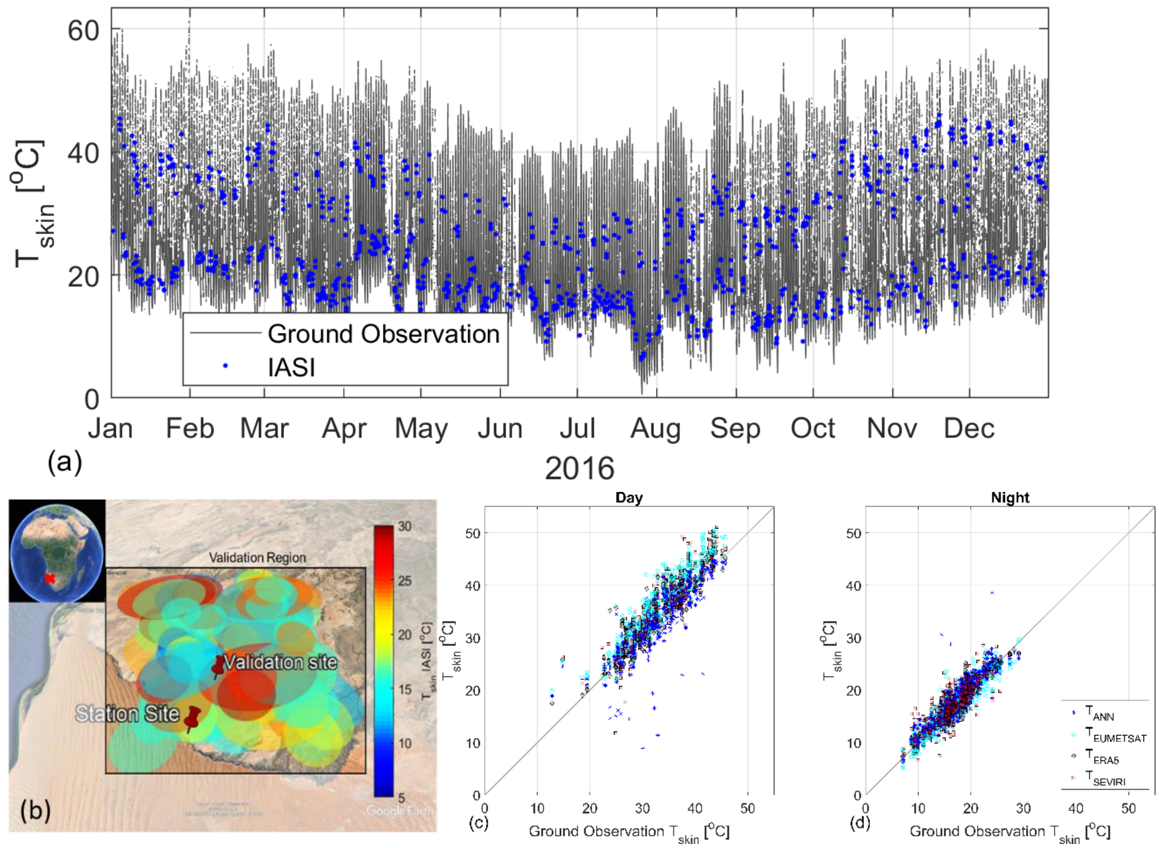
The authors of this paper are well aware of the complexity to deal with emissivity. They have been working on surface emissivities (MW, IR over land and ocean) for the last two decades (Aires et al., 2011; Paul et al., 2012). The latter study (Paul et al., 2012) shows a complex physical scheme to simultaneously retrieve Ts and surface emissivities. In this study, as the goal is to look at temperature changes trends for specific locations, we intend to present a simpler approach where information on emissivity is assumed to be included in the radiance spectra since it is quite exhaustive (see the conclusions of the study). The results and comparison on a global scale are very encouraging, although we agree that it probably affects validation locally. We will look into including information on emissivities in the future.

3- Regarding the inter-comparison and validation exercises, please provide more details on how the spatial matching is performed? Is SEVIRI resampled to the other products resolution or do you use the closest pixel?

The nearest pixel is chosen. It is mentioned in section 2.4.3: "IASI and SEVIRI data are spatially co-located when observations from each instrument are less than 5 minutes apart, and within 0.25 degrees in longitude and latitude."

4- For the in situ validation a single month does not seem enough to properly validate the products. At least different times of year should be considered. The differences you found for SEVIRI are significantly higher than what was previously reported by Gottshe et al. (2016), how do you justify this? You could have also used SEVIRI to access the spatial variability of the site, e.g. through the std of all SEVIRI pixels within an ISASI observation.

Generally speaking, it is hard to validate satellite measurements with ground LST given that the footprint of the satellite instrument will have various land surface types and the LST will therefore be an effective measure of this surface inhomogeneity. Gobabab is therefore uniquely suitable for validating IASI LST because of the large homogenous areas around it, as figure 7 (hereafter) panel (b) shows. To extend our analysis and to address the Reviewer concerns (also a concern for the second Reviewer), we performed a validation over the whole year (instead of just one month). The results and discussion show similar results to the one-month validation, as the figure hereafter shows:



*New **Figure 7. Comparison of IASI T_{ANN} with ground observations at Gobabeb: (a) Diurnal and seasonal variation of T_{skin}; (b) station and validation site location with a one-month example of IASI-coincident observations; (c) T_{ANN} versus in-situ T_{skin} during the day; and (d) during the night for all coincident observations in 2016.***

The discussion of the figures has been updated in different locations in the main text, and we point out that the conclusions are not very different from the original ones.

5- Also, in the validation report of EUMETSAT product (EUM/TSS/REP/13/684650), they found that because they were using an area quite far from the station (as you are) sometimes the station area as clouded while the satellite footprint was clear. You might want to use SEVIRI to remove observations when the station is under clouds.

It won't be possible to use clear-sky SEVIRI measurements to choose clear-sky measurements at the site, because the validation is done at the crossing-time of IASI, which might not correspond to the crossing time of SEVIRI at the site.

We think that the Reviewer brings a good point, so we discuss this in the text at the end of section 3: **"Differences between the different products and ground measurements can be due to cases where the sky at the in-situ measurement site is at least partly**

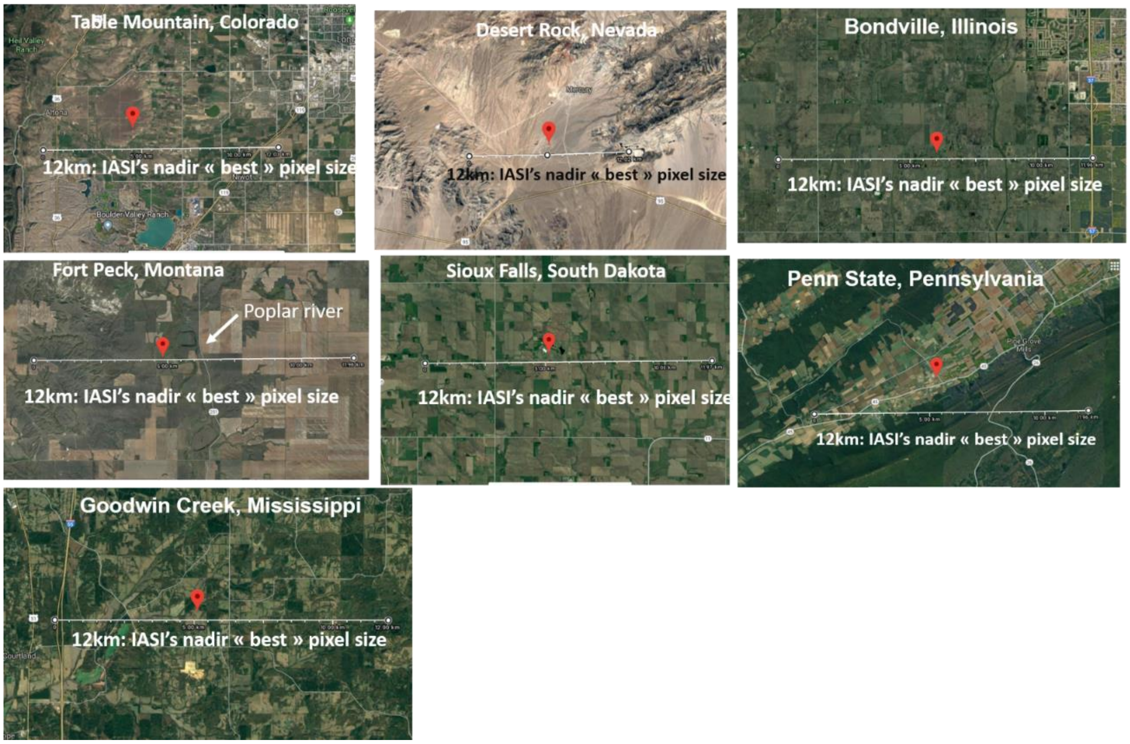
cloudy/clear, while being clear/partly cloudy at the validation site (EUMETSAT, 2013).”

6- Despite the constrains related the spatial resolution, the authors could also have performed station comparisons with other KIT and e.g. SURFRAD stations as they could provide further information on the quality of the retrievals. Jimenez et al. (2017), for instance, used these to validate retrievals from the AMSR-E, which has approximately the same spatial resolution.

We extend the validation around Gobabeb to a whole year, as discussed in point #4 in this review. However, to answer the Reviewer specific comment, we'd like to point out that AMSR-E Tskin retrieval is placed in the 14×8 km^2 (=112 km^2) swath grid of the 36.5 GHz channel. IASI's pixel area is at best a circle of π x 6 x 6 km^2 at nadir (=113 km^2) and an ellipse with an area up to ~ π x 10 x 20 km^2 at its outermost viewing angle of 48° (off-nadir).

The SUFRAD stations shown hereafter are all around inhomogeneous land surface types. A 12 km ruler is placed over each of the SUFRAD locations to show the minimal IASI pixel (which would be at best 4 pixels out of the 120 pixels in one swath, without cloud filtering which usually takes out 2/3 of measurements). Clearly, from the pictures, many land types exist around the validation sites, which complicates the validation.

All SURFRAD stations deliver long-term measurements of the surface radiation budget. This is done by measuring downwelling and upwelling broadband solar and thermal infrared (TIR) irradiance. Skin temperature has to be derived from incoming and outgoing IR radiance measurements, and by estimating emissivities. We therefore believe that this would introduce many sources of error into the comparison with IASI since the emissivity is a function of land type that changes over the IASI's pixel. We therefore only use in our discussion/validation the Gobabeb station. Upon discussing with the KIT's stations PI, we realized again that the horizontal resolution is an issue. Gobabab is the only of KIT's site that is suitable for validating IASI LST: the homogenous areas around the other sites are just too small.



The location of the different SUFRAD stations. The white arrows are the best/Nadir IASI observations. It is clear from these pictures that validation around these sites is challenging. Source: Google Maps.

References added

Aires, F., Prigent, C., and Rossow, W.B.: Soil moisture at a global scale. II – Global statistical relationships. JGR, 110, D11, D11103, 10.1029/2004JD005094, 2005.

Aires, F., Prigent, C., Bernardo, F., Jimenez, C., Sounders, R., and Brunel, P.: A Tool to Estimate Land Surface Emissivities in the Microwaves (TELSEM) for use in numerical weather prediction schemes. QJRMS, 137: 690-699, 10.1002/qj.803, 2011.

EUMETSAT: IASI L2 Metop-B Validation Report, v3A, 19 July 2013, EUM/TSS/REP/13/684650, https://www.eumetsat.int/website/home/Data/TechnicalDocuments/index.html, 2013.

Kolassa, J., Aires, F., Polcher J., Prigent, C., Jimenez, C., and Pereira, J.M.: Soil moisture retrieval from multi-instrument observations: Part I – Information content analysis and retrieval methodology, JGR, 118, 10, 4847-4859, 10.1029/2012JD018150, 2013.

Paul, M., Aires, F., and Prigent, C.:An innovative physical scheme to retrieve simultaneously surface temperature and emissivities based on a high-resolution infrared emissivity interpolator, JGR, 117, D11302, 10.1029/2011JD017296, 2012.

Rodriguez-Fernandez, N.J., Aires, F., Richaume, P., Cabot, F., Jimenez, C., Kerr, J., Kolassa, J., Mahmoodi, A., Prigent, C., and Drush, M.: Soil moisture retrieval from SMOS observations using neural networks, IEEE TGRS, 53, 11, 10.1109/TGRS.2015.2430845, 2015.

Rodríguez-Fernández, N., de Rosnay, P., Albergel, C., Richaume, P., Aires, F., Prigent, C., Kerr, Y.: SMOS Neural Network Soil Moisture Data Assimilation in a Land Surface Model and Atmospheric Impact, Remote Sens. 11*(11), 1334; https://doi.org/10.3390/rs11111334, 2019.*