This paper entitled "Assessing Earth's skin temperature trends: consistent signals from IASI, MODIS, CCI and ERA5" intends first to evaluate several datasets of skin temperature (over ocean and continents) and then to compare their respective trends.

We thank the reviewer for their remarks. Below is a point-by-point reply in blue. Added text to the manuscript is shown in *Italic*.

Major comments:

- The datasets that are compared are not coherent. Some are clear-sky, other are all sky. Introducing cloudy skin temperatures in the analysis biases the dataset toward colder temperatures. Uncertainties are also higher below the clouds. Comparing datasets so different in nature is not reasonable and I suggest that you do the SAME cloud filtering for each of the datasets so that you can do an analysis on clear-sky only.

We thank the reviewer for this important remark. Following the reviewer's recommendation, we have revised the analysis to ensure a consistent cloud treatment across the datasets.

To achieve this, we adopt the following cloud-filtering strategy as mentioned in the added Section 2.6 as follows. Figures 1 to 3 are added hereafter.

"2.6 Cloud filtering strategy

To ensure a physically consistent intercomparison, all datasets are evaluated under clear-sky conditions. Each dataset is filtered using the most appropriate cloud screening method. For IASI-NN, clear-sky retrievals are already guaranteed by the cloud mask of Whitburn et al. (2022), which is applied at the pixel level prior to retrieval. The same cloud mask is applied to IASI-CDR, retaining only those IASI-CDR retrievals collocated in space and time with IASI-NN clear-sky pixels. Since both products share the same instrument, orbital swath, and overpass times, this constitutes a physically consistent filter that ensures both IASI products are evaluated over identical scenes.

For ERA5, we apply a threshold on the total cloud cover variable (tcc), which is ERA5 own native cloud diagnostic. We retained only grid points and time steps where tcc < 0.01 (10%), consistent with strict clear-sky screening as used in comparable studies (e.g., Rabin et al., 2013). A sensitivity analysis reported by (Ermida et al., 2019; Ermida & Trigo, 2022) demonstrates that varying the cloud fraction threshold between 1% and 30% changes the derived clear-sky bias by less than 0.5 K for 85% of grid points, indicating that our choice of threshold does not materially affect the results; the results of (Wang et al., 2022) confirm that monthly cloud-filtered ERA5 SKT trends are robust across a wide range of cloud cover thresholds.

MODIS Terra LST (v6.1) and ESA LST CCI (v3.00) are clear-sky products by construction, as their retrieval algorithms only produce valid retrievals under cloud-free conditions; no additional cloud filtering is therefore applied to these datasets."

The figures show that the comparison between the different instruments is now slightly better: the cloud filtering does not drastically change our results. For this the discussion of the figures is not changed.

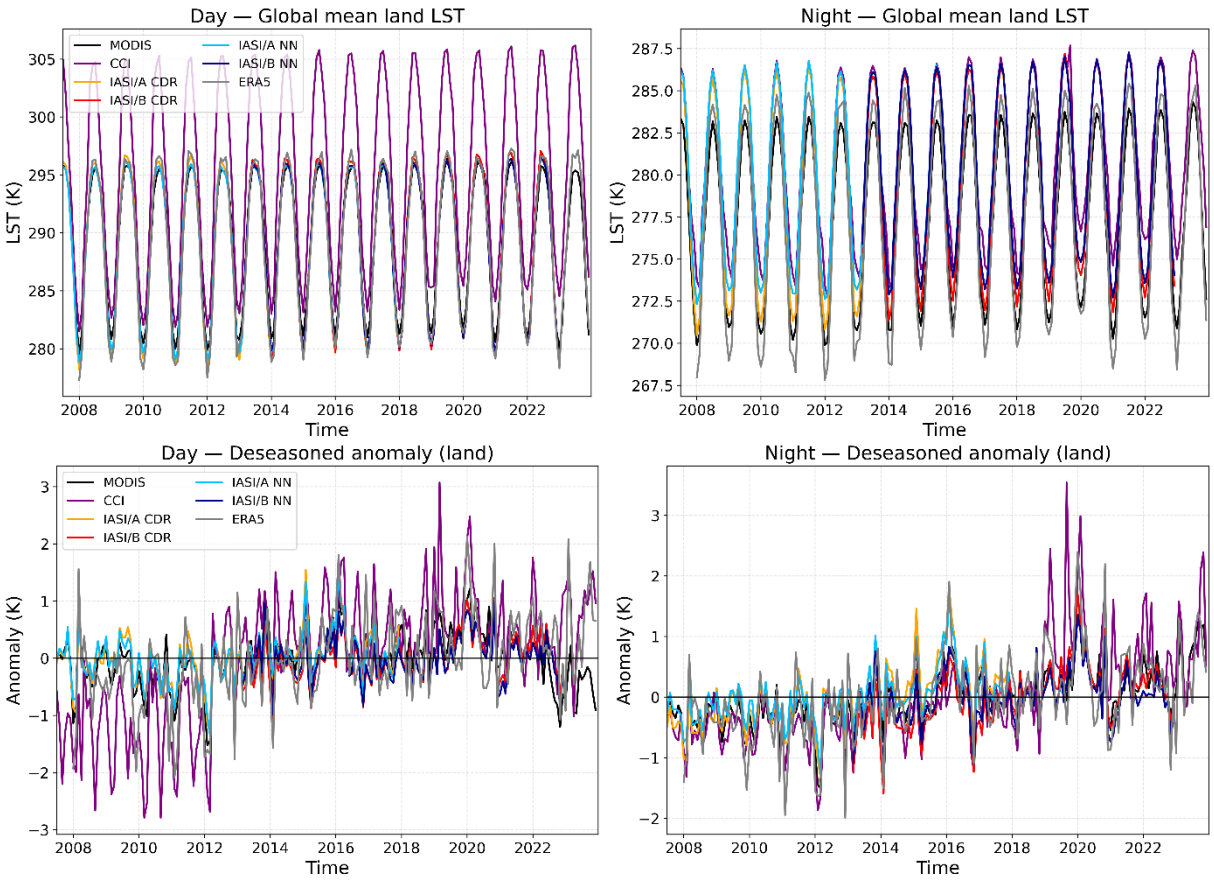

Figure 1. **Global mean land surface temperature (LST) and deseasonalised anomalies for daytime (left) and nighttime (right). IASI Metop-A and Metop-B observations (IASI-A/B) are shown separately for the CDR and NN products. Global means are computed using cosine-latitude weighting.**

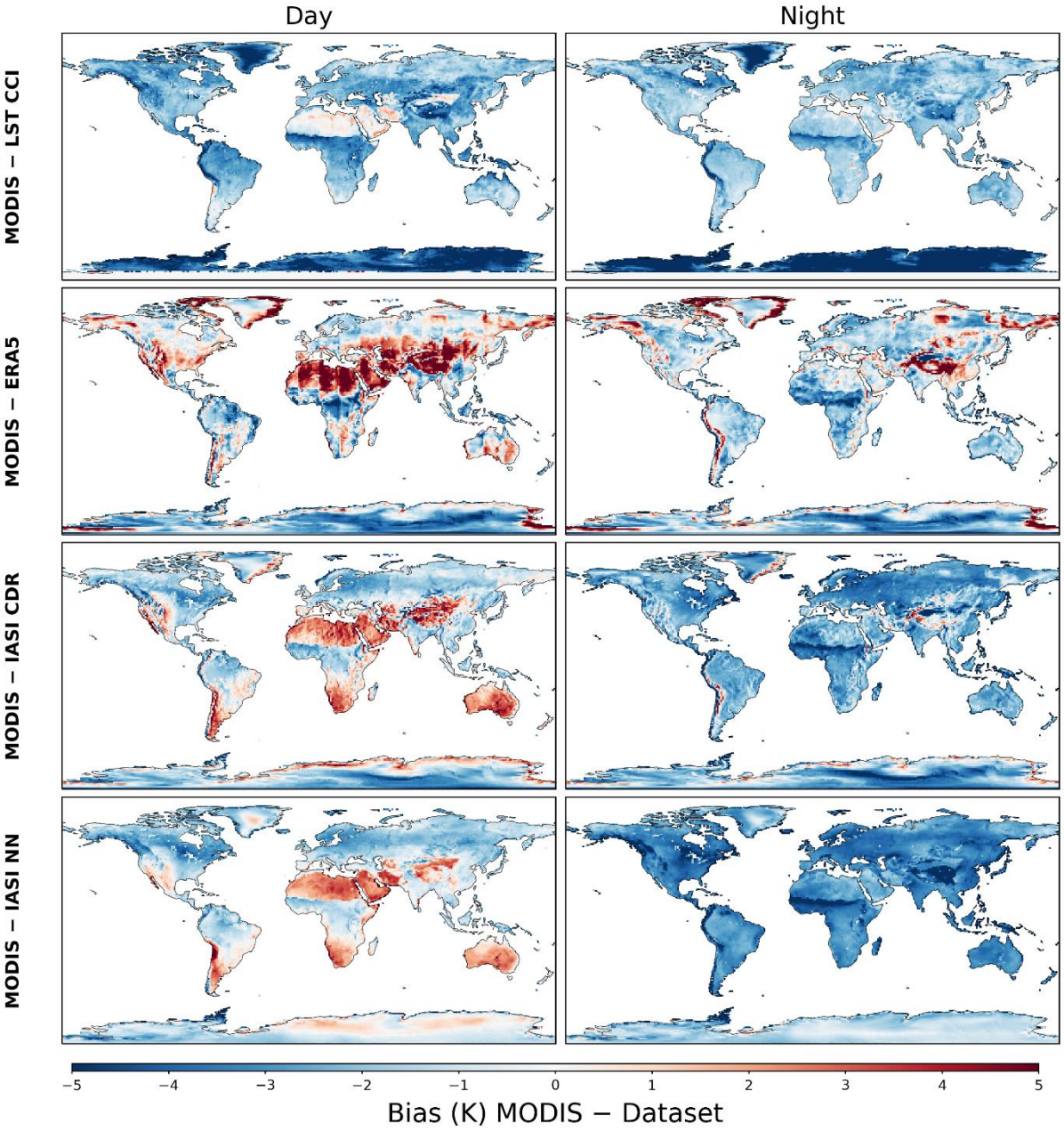

Figure 2. **LST spatial bias averaged over the 2007-2022 period of MODIS during the day (left) and the night (right), with respect to LST CCI, ERA5, IASI-CDR and IASI-NN.**

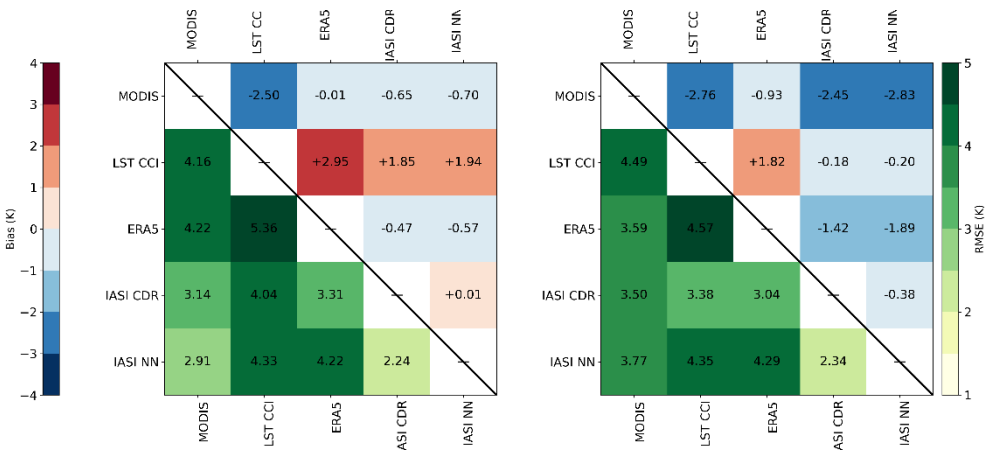

***Figure 3**. Upper triangle (blue to red): global mean bias (row – column, y -x): Lower triangle (in yellow to green): root mean square error (RMSE) between the different pairs of LST datasets, for day (left) and night (right).*

For the bias intercomparison (Sect. 3), datasets are regridded to the common 1°×1° grid after cloud filtering, and statistics are computed over grid cells where all products provide valid clear-sky values within the same month. For the trend analysis (Sect. 4), each dataset is kept on its native grid after cloud filtering to preserve the full spatial and temporal density of each product, avoiding the introduction of IASI's orbital sampling biases into the trend estimates of ERA5, MODIS, and LST CCI.

As per the reviewer suggestion above, all datasets are now cloud free, so that the Figure 2 and the trends in Figure 4 refer to the same cloud free product.

- The diagnostics that are used in the analysis are misleading in my opinion. You perform a lot of averaging on the data and then evaluate the biases on this averaged data. This is not correct. For instance, you can have a dataset with 50% of the samples with -5C of bias, and 50% of the samples with +5C biases, if you average these biases, you will come up with a 0 bias... This is what we can see for instance in Fig2 (column 1 and row 2) where you have very high values and low values, then compute the average bias in Table 3 (if I interpret correctly) at -0.63C. You cannot say that the overal bias between MODIS and ERA is -0.63C... This is highly misleading.

We thank the reviewer for this remark. We acknowledge that the global mean bias, computed by spatially averaging signed differences, can indeed suffer from cancellation between regions of positive and negative bias. To address this, we report the RMSE alongside the mean bias in Figure 3 precisely because these two statistics are complementary: the mean bias captures the systematic offset, while the RMSE reflects the spread of the bias field across grid cells and is not subject to cancellation of opposite-sign contributions. Reporting both together is standard practice in gridded LST intercomparison studies (e.g., Ermida et al., 2019), and we follow this convention. A small mean bias paired with a large RMSE indicates strong spatial heterogeneity in the sign and magnitude of the differences — which is exactly the situation described by the reviewer — while a small RMSE would indicate that the mean bias is spatially representative.

We have added to the supplementary material Figure S2, which shows the spatial distribution of RMSE for the same dataset pairs as Figure 2. The RMSE maps reveal that discrepancies are largest at high latitudes and in regions with highly variable surface emissivities (e.g., arid regions, tropical forests, snow-covered areas). We note that because global averages are computed with cosine-latitude weighting, high-latitude grid cells — where RMSE values are largest — contribute proportionally less to the global mean, which partly explains why the globally-averaged statistics in Table 3 appear modest. The RMSE values are broadly consistent with the spatial bias patterns shown in Figure 2, confirming that the global mean bias does not arise from a near-complete cancellation of large regional signals, but rather reflects a modest systematic offset superimposed on genuine regional variability. We discuss this in the manuscript after the discussion of Figure 2 as follows:

"In Supplementary Material Figure S2, we show the RMSE for the same pairs of datasets. The RMSE is generally between 1 and 3 K over most land surfaces during both day and night, with higher values exceeding 4–5 K over arid regions and at high latitudes. These elevated RMSEs reflect local surface heterogeneity and variable emissivity conditions rather than a systematic global offset. The consistency between the spatial patterns of bias (Figure 2) and RMSE (Figure S2) indicates that the global mean bias values reported in Table 3 are not artefacts of spatial cancellation, but represent a genuine, if spatially non-uniform, systematic difference between the datasets."

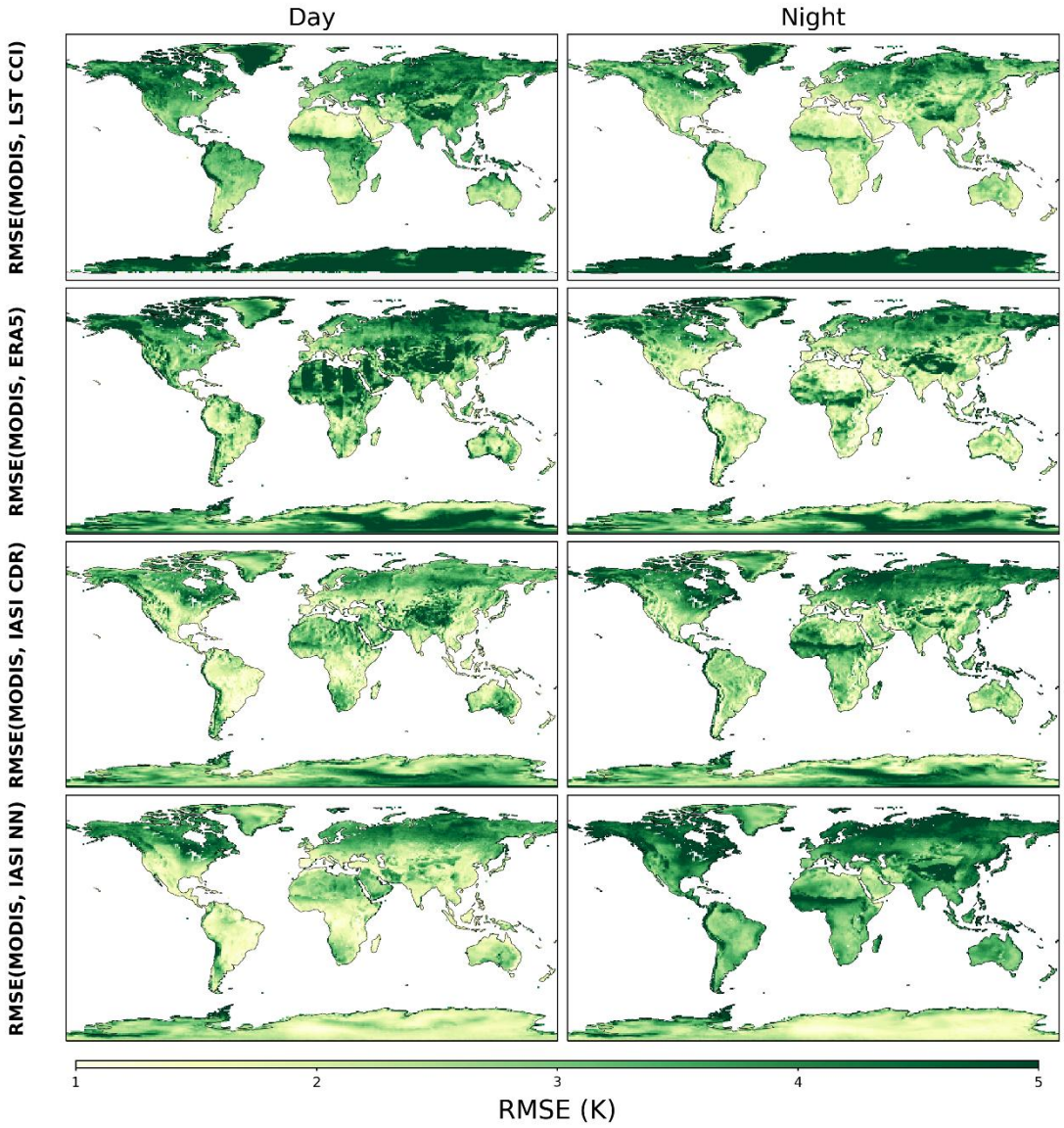

Figure S2. **LST spatial RMSE during the day (left) and the night (right), of MODIS with respect to LST CCI, ERA5, IASI-CDR and IASI-NN.**

- Figure 1, first plot, the ESA CCI is 8°C higher than the other datasets. It means that their global average over land in the whole planet is 8°C higher than the other datasets, at 30°C... ESA CCI LST has been highly evaluated, there are papers and reports on its evaluation, so I believe there are an error in the way the data is represented. With such differences, you need to investigate what is happening, cannot just continue the analysis as if everything is normal.

We thank the reviewer for raising this point and confirm that no error is present in how the data are represented. The offset is a real, physically explicable feature of the merged LST CCI product, and we now address it explicitly.

The key difference lies in the viewing geometry of the two datasets. The NASA MODIS LST product uses the full MODIS swath of ±60° view zenith angle (VZA), whereas the LST CCI IRCDR is restricted to ±22° VZA. This restriction is by design: it ensures consistency across the full LST CCI data record, which also includes the narrower-swath ATSR-2 and AATSR instruments. It has been shown in numerous studies that MODIS LSTs are significantly

depressed at the edges of the swath due to angular emissivity effects and increased atmospheric path length at high viewing angles. Restricting to near-nadir observations in the LST CCI therefore yields systematically warmer global mean values compared to MODIS, which averages over a much wider angular range including colder off-nadir retrievals. This viewing geometry difference is the primary driver of the observed offset between the two products, and it does not reflect a calibration error or a misrepresentation of the data.

We have added a sentence in Section 2 to this effect and updated the discussion of Figure 1 accordingly. We also note that co-author Darren Ghent (PI of the ESA LST CCI project) verified this explanation through direct inspection of matched pixels between the LST CCI IRCDR and the MODIS CCI LST, confirming consistency when viewing geometry is accounted for. The added paragraph is as follows:

"Data are retrieved at a 0.01-degree latitude/longitude grid, that we regridded to 1x1 grid size for discussion and analysis.

An important characteristic of the LST CCI IRCDR that distinguishes it from the NASA MODIS LST product is the restriction of retrievals to view zenith angles (VZA) of ±22°. This restriction is intentional: the full LST CCI data record also includes data from ATSR-2 and AATSR, which have narrower swaths than MODIS or SLSTR, and limiting all sensors to ±22° ensures geometric consistency across the entire multi-decadal record. In contrast, the NASA MODIS LST product uses the full MODIS swath of ±60° VZA. It has been shown that MODIS LSTs are systematically depressed at the edges of the swath due to increased atmospheric path length and angular emissivity effects at high viewing angles (Ermida et al., 2019; Wan, 2014). By averaging over a much wider angular range that includes these colder off-nadir retrievals, the MODIS global mean LST is expected to be lower than the near-nadir-only LST CCI product. This viewing geometry difference is the primary physical explanation for the systematic warm offset of LST CCI relative to MODIS visible in Figure 1, and does not reflect a calibration error. Inspection of matched pixels between the LST CCI IRCDR and the MODIS CCI LST confirms consistency when viewing geometry is accounted for."

- Figure 2, there is a longitudinal structure in the ERA5 data (the vertical bands). You mention this issue but do not solve it.

We acknowledge the longitudinal banding visible in the ERA5 maps. This artefact arises directly from the method used to extract local-time ERA5 data. Rather than downloading and filtering the full hourly ERA5 record — which would require 15 years × 365 days × 24 hourly global grids at 0.25° × 0.25° resolution, an operation that is computationally and storage-wise intractable — we work with monthly-averaged hourly data and convert to local solar time by applying a longitude-based time zone offset (15° = 1 hour). This discrete approximation introduces sharp transitions at timezone boundaries, producing the vertical banding structure visible in Figure 2. The artefact is most pronounced at the edges of each timezone bin, where the rounding error in local time is largest. This limitation and its origin are discussed in Safieddine et al. (2025).

Critically, since the banding pattern is stationary in time — it reflects a fixed property of the longitude grid and the time conversion method, not any physical or instrumental change — it does not bias long-term trend estimates. The trend analysis presented in Section 4 is therefore unaffected. Removing the artefact entirely would require reprocessing the full ERA5 hourly archive at native resolution, which is beyond the scope of the present study.

- You retrieval based on IASI information was trained on the IASU Eumetsat product that is based on IASI data and microwave observations, trained on ERA5 skin temperatures. First issue: training a IASI retrieval, on a IASI retrieval trained on ERA5 targets. I do not see the advantage compared to the retrieval based on ERA5 targets... Second, how can we inter-compare all these products that are not a all independent? And why are such large differences between the datasets?

The reviewer is mixing two distinct EUMETSAT products.

The **EUMETSAT IASI CDR** (used as a comparison dataset here) is a statistical retrieval trained on ERA5 targets, as described in Section 2.4.

The **EUMETSAT NRT Tskin product** (used as the training target for IASI-NN) is entirely different: it is derived via optimal estimation directly from IASI observations, independently of ERA5, and is therefore not "a retrieval trained on ERA5." We use the 2021 NRT product specifically, a year when it is known to be stable (Bouillon et al., 2020). The advantage of this approach over direct ERA5 training is that the IASI-NN target is observationally grounded rather than reanalysis-based, providing an independent perspective on surface temperature.

The two IASI products in this study thus share the same instrument but use completely different retrieval pathways, and their differences are scientifically meaningful. Regarding broader independence: examining differences between products that share same lineage (the radiances) is valuable, and the consistency of warming trends across all datasets in Section 4 — despite their different algorithmic origins — strengthens our conclusions.

We add further explanations into the description of our product based on artificial neural network as follows (new text in shown in italic)

"A second key update is the use of the EUMETSAT Near Real Time (NRT) Tskin product from 2021 as the training target, replacing ERA5, as was done in Safieddine et al. (2020). This allows the NN to emulate an operational optimal-estimation surface temperature directly derived from IASI and ensures consistency over the entire mission. *The NRT product is based on optimal estimation and over the years, EUMETSAT has performed several updates on the real-time processing of both radiances and temperatures, making the time series non-homogeneous, until around 2017 (Bouillon et al., 2020). For this reason, the year 2021 is used as a trustworthy year for the training target."*

Minor comments:

- You use in the title "CCI" for LST ESA CCI. CCI is used by NASA for many variables so this is ambiguous.

We now replace CCI with *ESA CCI* in the title

- Abstract: in general, no paragraphs in an abstract.

Paragraph indentation is now removed in the paragraph

- Line 26: "in in": removed.

- In your NN model, you have the pixel number (in the orbit). What is important in terms of radiative transfer (and then for the retrieval) is the incidence angle. There is a relationship between the pixel number and this incidence angle. I believe giving the angle would really be more adequate than this indicator, discrete, and non-monotonic information.

IASI has 120 pixels per scan line, mapping to view zenith angles of 0–48.3 degrees; after min-max normalization, pixel number and view zenith angle are linearly equivalent as NN inputs, and our sensitivity tests confirmed no improvement in retrieval skill when substituting one for the other (not shown). We note that the pixel is not particularly "monotonic" because the IASI data used is cloud free, so at least 70% of the data, on a daily basis is lost (to give an idea, IASI makes around 1.2 million observations per day, and around 350 000 to 450 000 are cloud free).

- Figure 3 : I already commented on the fact that these numbers are misleading. You are averaging spatially the biases that are positive and negative, this can give you artificially low medium bias. We can see clearly in the maps of Figure 2 that biases are truly non neglectable. To obtain numbers that can synthetize the bias and RMS between the two datasets: 1) you need to do you analysis for each retrieval (don't average on 1° boxes, as this already smooth everything) and everything needs to be done at the native resolution; 2) you need to use the same filters (for clouds, etc.) so that you can compare couple of retrievals for same location and same time; 3) you can represent the maps of bias and RMS at native resolution; 4) if you want to obtain a single number, you can aberage the pixel RMS (as they are positive) but not the biases (I would recommend the average of the pixel biases).

We agree that the global mean bias metric is limited and have improved the analysis accordingly. We address each point in turn.

Regarding points 1 and 3: reprocessing 15 years of MODIS and LST CCI data at native L2 swath resolution is beyond the scope of a revision, as these products are only publicly available as monthly gridded composites and re-downloading and reprocessing the full L2 archive would be a substantial independent undertaking.

Regarding point 2: a consistent cloud-filtering strategy has now been applied across all datasets, as described in the new Section 2.6.

Regarding points 3 and 4: We now report the spatial RMSE in supplementary material because the two are complementary and this was discussed in this review.

References

Bouillon, M., Safieddine, S., Hadji-Lazaro, J., Whitburn, S., Clarisse, L., Doutriaux-Boucher, M., Coppens, D., August, T., Jacquette, E., & Clerbaux, C. (2020). Ten-Year Assessment of IASI Radiance and Temperature. *Remote Sensing*, *12*(15), Article 15. https://doi.org/10.3390/rs12152393

Ermida, S. L., & Trigo, I. F. (2022). A Comprehensive Clear-Sky Database for the Development of Land Surface Temperature Algorithms. *Remote Sensing*, *14*(10), 2329. https://doi.org/10.3390/rs14102329

Ermida, S. L., Trigo, I. F., DaCamara, C. C., Jiménez, C., & Prigent, C. (2019). Quantifying the Clear-Sky Bias of Satellite Land Surface Temperature Using Microwave-Based Estimates. *Journal of Geophysical Research: Atmospheres*, *124*(2), 844–857. https://doi.org/10.1029/2018JD029354

Safieddine, S., Parracho, A. C., George, M., Aires, F., Pellet, V., Clarisse, L., Whitburn, S., Lezeaux, O., Thépaut, J.-N., Hersbach, H., Radnoti, G., Göettsche, F.-M., Martin, M. A., Doutriaux-Boucher, M., Coppens, D., August, T., Zhou, D. K., & Clerbaux, C. (2020). Artificial Neural Networks to Retrieve Land and Sea Skin Temperature from IASI. *Remote Sensing*, *12*(17), Article 17. https://doi.org/10.3390/rs12172777

Wan, Z. (2014). New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment*, *140*, 36–45. https://doi.org/10.1016/j.rse.2013.08.027

Wang, Y.-R., Hessen, D. O., Samset, B. H., & Stordal, F. (2022). Evaluating global and regional land warming trends in the past decades with both MODIS and ERA5-Land land surface temperature data. *Remote Sensing of Environment*, *280*, 113181. https://doi.org/10.1016/j.rse.2022.113181