Reviewer 1

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 2.8 as follows. Figures 1 to 3 are added hereafter.

"2.8 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 (tcc) variable, ERA5's native cloud diagnostic, retaining only grid points and time steps where tcc < 0.2 (20%). This clear-sky criterion is consistent with the approach used in comparable studies (Ermida et al., 2019; Ermida & Trigo, 2022), which also demonstrate through sensitivity analysis 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. Moreover, the ERA5 Tskin trends derived from monthly cloud-filtered data are shown to be robust across a wide range of cloud cover thresholds (Wang et al., 2022).

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 ESA SST CCI/C3S SST and IST product combines infrared clear-sky retrievals with all-sky passive microwave observations through optimal interpolation, and is therefore not cloud-filtered. Applying a cloud mask, a posteriori, using ERA5 monthly cloud cover fraction, for example, would be the only feasible approach to approximate clear-sky conditions, but this would introduce an ERA5-dependent filtering that is inconsistent with the cloud screening applied to IASI and MODIS, which is based on their own instrument-level detection. We therefore present the ESA CCI/C3S SST and IST product as an all-sky all-day reference."

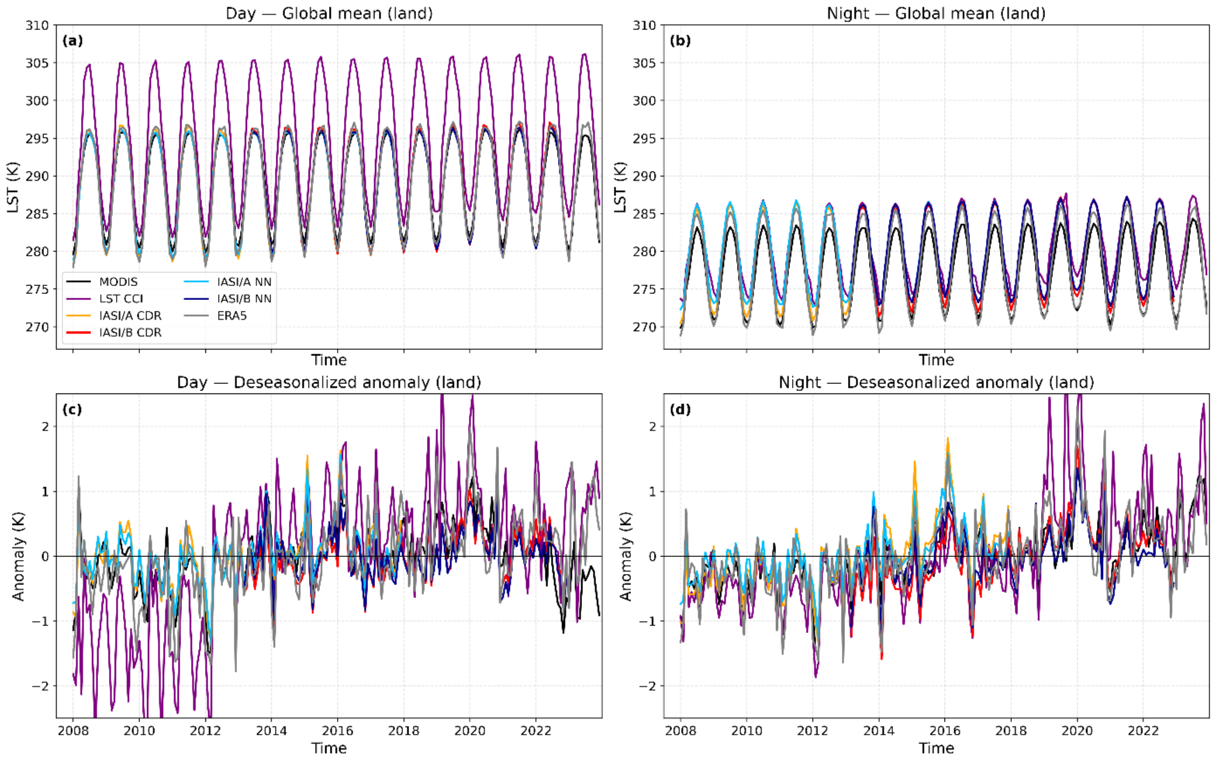


Figure 1. Global mean land surface temperature (LST, panels a and b) and deseasonalised anomalies (panels c and d) for daytime (left) and nighttime (right). IASI Metop-A and Metop-B observations (IASI-A/B) are shown separately for the CDR and NN products. Global means are computed using cosine-latitude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

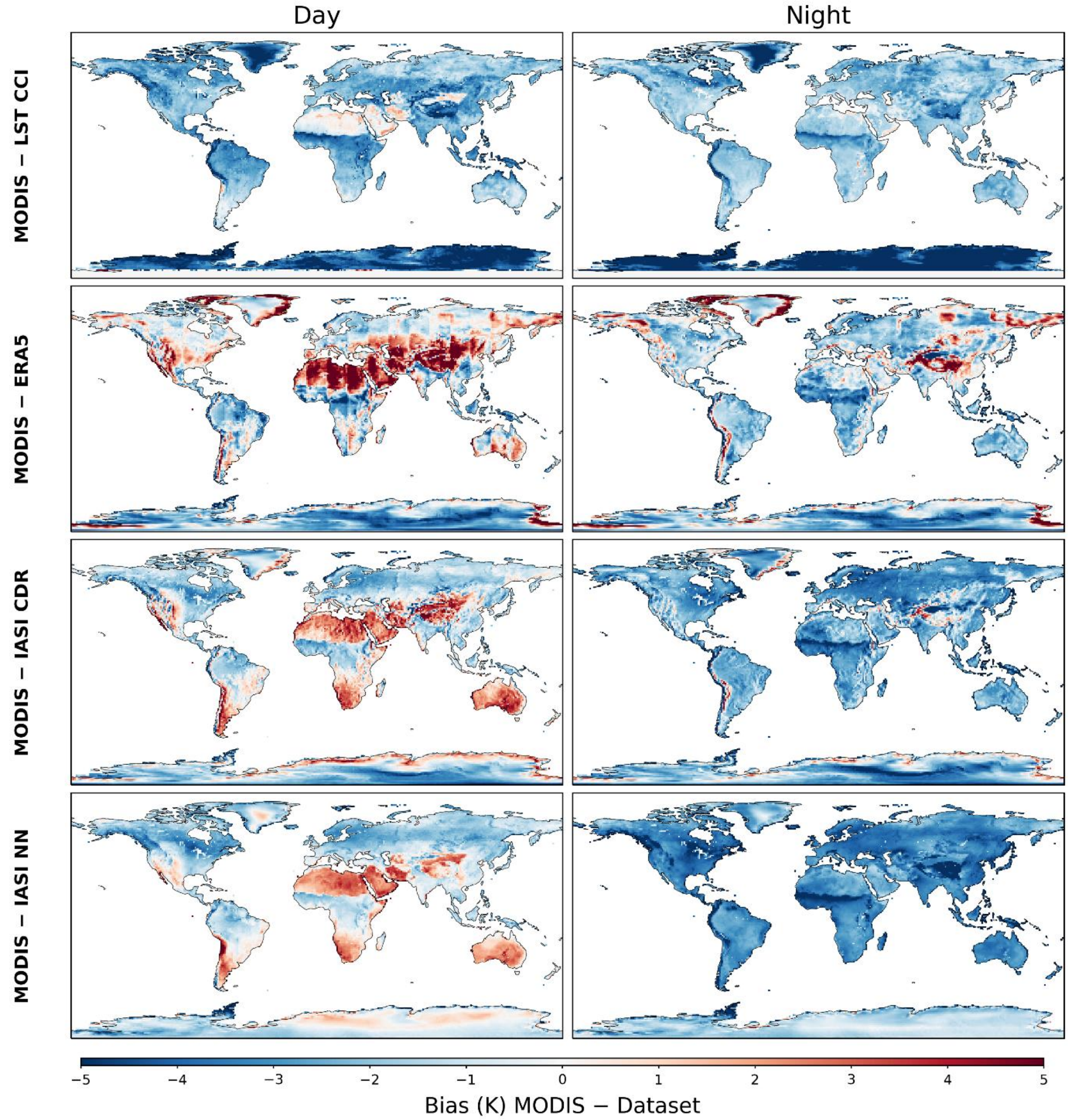


Figure 2. LST spatial bias averaged over the 2007-2022 period of MODIS during the day (left) and the night (right), with respect to ESA LST CCI, ERA5, IASI-CDR and IASI-NN. A *cold nighttime bias with respect to MODIS is evident across all datasets. ERA5 land/sea mask is used to mask IASI and ERA5 sea grid points.*

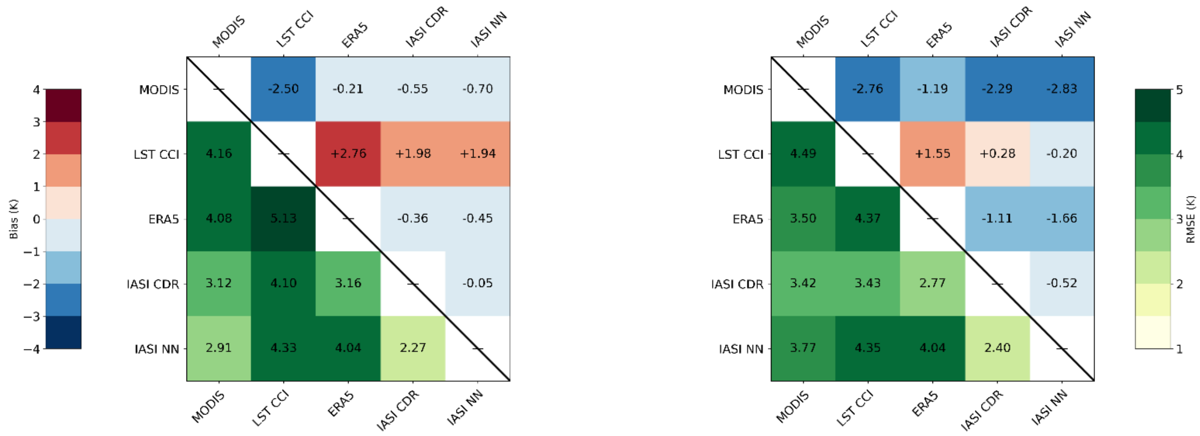


Figure 3. Upper triangle (blue to red): global mean bias (row – column, y -x): Lower triangle (in yellow to green): root mean square error (RMSE) between the different pairs of LST datasets, for day (left) and night (right). Statistics are computed over land grid cells only (using the ERA5 land-sea mask), weighted by cosine-latitude.

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."

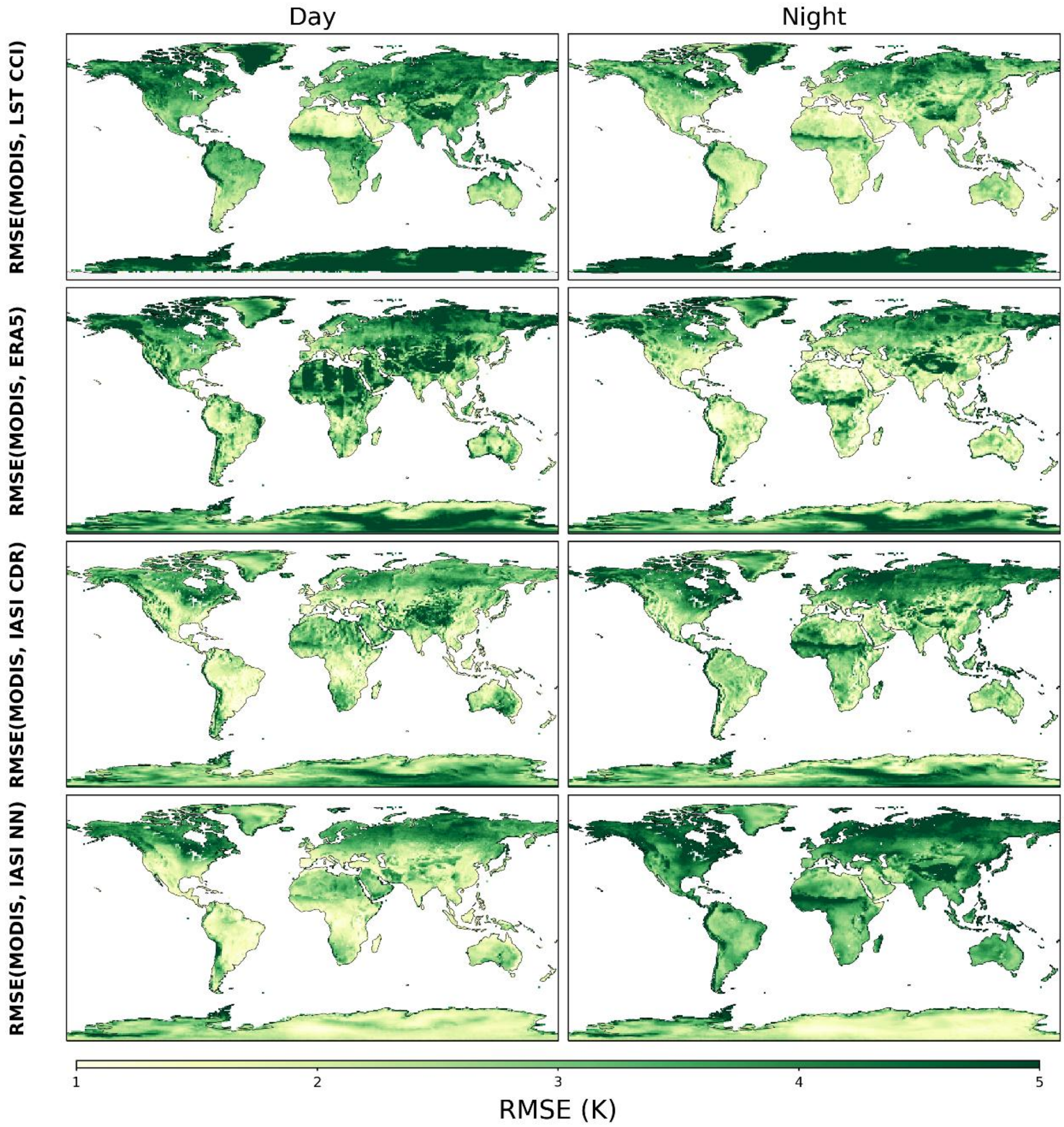


Figure S2. **LST spatial RMSE during the day (left) and the night (right), of MODIS with respect to LST CCI, ERA5, IASI-CDR and IASI-NN.**

- 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; Z. 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.

Reviewer 2

The topic of this paper is of high interest. Indeed, multiple skin temperature datasets are currently available, yet this variable has not received much attention from the climate community. There are three main reasons for this: 1) Clouds introduce substantial uncertainties and errors in LST/SST retrievals. 2) Multi-sensor datasets often suffer from inhomogeneities during sensor transitions, making the calculation of trends extremely challenging and uncertain, ad 3) Even the highest-quality datasets are still not long enough to allow the derivation of robust and meaningful trends.

We thank the reviewer for their comments. Our responses are listed in blue for each of the comments raised. New text is shown in *italic.*

Therefore, in a paper whose main motivation is to encourage the use of these datasets, I would expect these issues to be addressed with particular care. Unfortunately, this does not appear to be the case here. All-sky datasets are compared with clear-sky datasets, and sea ice is not treated appropriately. The LST CCI dataset exhibits clear problems when compared with the remaining datasets (large biases and apparent inhomogeneities). The manuscript even states that "This version of the CCI product has not yet been fully validated," which raises the question of why it is used in the analysis at all.

In the updated version of the paper, we cloud-filtered all datasets and the figures/discussion have been updated accordingly. As for the LST CCI product, the idea behind presenting it, is a first step validation, because very little literature review exist that evaluates this dataset. We discuss this further in the review.

Regarding the selection of datasets, one may also question why ERA5-Land was not used over land, given that surface processes are represented much more accurately there. ERA5 appears to be treated as an absolute "truth," despite the fact that it has several well-documented issues. Over the ocean, similarly to the approach taken over land, it is unclear why a CCI-like product was not included in the comparison (see minor comments below). Additionally, the IASI-NN product does not appear to have been validated (at least no reference to a validation study is provided).

The different points are discussed in this review in much detail, to sum up:

- ERA5 land: since we are now using the total cloud cover to clear cloudy scenes, and it is not available in ERA5 Land, we choose to keep ERA5.
- ERA5 as ground truth: we added a discussion on the ERA5 limitations. Indeed, we do believe that Tskin ERA5 has caveats, as it is a model output, and not an assimilated product.
- CCI like products: we added SST CCI ESA product as the Reviewer suggests.
- IASI-NN: the initial goal of this paper was to validate this product and the trends associated. Then it became a more general paper, with several

products available to compare. We find that this paper could serve as a reference for many LST/SST products. We hope that this version is suitable for publication.

Overall, I feel that there are too many unresolved issues, and that both the discussion and the data treatment remain too superficial. Therefore, I recommend that the paper undergo **major revision** before it can be considered for final publication.

We hope that with this new version, the issues that the Reviewer raises are now solved.

List of Issues:

- **L14 –** "Earth's skin temperature (Tskin), i.e. land and sea surface temperature (LST and SST)" – perhaps ice should also be mentioned.

Correct. The phrase now reads: *"Earth's skin temperature (Tskin), i.e. land, ocean/sea, and ice surface temperatures (LST, SST, and IST)."*

The first phrase of the introduction is now also updated to: *"In contrast, land and, ocean/sea and ice surface temperatures (LST, and SST and IST), collectively referred to as Earth's skin temperature (Tskin), [...]"*

- **L16** – "infrared sounders" should be **"sounder"**, since only IASI is used.

"Infrared sounders" is now replaced with "infrared instruments" to account for IASI, MODIS, AATSR and SLSTR-B (used in the LST CCI).

- **L39** – "Owing to its direct sensitivity to surface–atmosphere interactions, Tskin serves as a reliable tracer of climate variability."

  This statement appears to be more of a hypothesis that the paper aims to demonstrate rather than an established premise. Many researchers would disagree with this claim, since Tskin datasets are affected by several issues that limit their reliability. These challenges are precisely why they remain underused in climate studies.

This phrase is now rephraseded to: "Owing to its direct sensitivity to surface–atmosphere interactions, Tskin *is expected to* serve as a reliable tracer of climate variability, *provided dataset consistency and quality are ensured.*"

- **L44–45** – Please format variable symbols in italic.

Done.

- **Section 2.3** – For a full understanding of the results, the manuscript should provide a broader explanation of how ERA5 Tskin is derived in the model for both land and ocean. The caveats associated with ERA5 Tskin are not sufficiently emphasized and may create the misleading impression that it can be used as an absolute "truth." It would also be useful to comment on the independence of ERA5 relative to the other datasets.

Section 2.3 is now Section 2.4 (since we added a new ESA CCI SST section 2.3). ERA5 is considered by many as the reference for temperatures as it assimilates a lot of data. We also think it might have caveats, in particular for Tskin as not many instruments can measure it. We added the following at the end of section 2.4 in order to make it clear, and not mislead the reader with this impression. We also added a section 2.7 called "Interdependency between datasets" that addresses the Reviewer's second remark.

We added the following at the end of (now) Section 2.4:

"ERA5 should not be treated as an independent ground truth, but rather as a physically consistent reference that is subject to its own model-dependent uncertainties. Over land, ERA5 Tskin is a prognostic variable computed within the HTESSEL (Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land) land surface scheme (Balsamo et al., 2009), as part of the Integrated Forecast System (IFS). It is integrated forward in time by solving the surface energy balance at each model timestep, and is therefore sensitive to model parameterisations of soil thermal properties, vegetation, and snow. Over ocean, ERA5 Tskin is not prognostic but is instead derived from a prescribed SST analysis with a diagnostic cool-skin correction applied (Hirahara et al., 2016; Zeng & Beljaars, 2005). ERA5 does not directly assimilate Tskin, but does assimilate SST analyses over ocean and infrared radiances over both land and sea (Hersbach et al., 2020), which carry indirect surface temperature information."

And a new section 2.7 is added:

"***2.7. Interdependency between datasets***

The datasets used in this study are not fully independent of one another, which is relevant when interpreting intercomparison results. ERA5 assimilates IASI radiances for atmospheric profile retrieval (Hersbach et al., 2020), which partially constrains its land surface scheme and thus its Tskin. IASI-CDR in turn uses ERA5 atmospheric profiles as training input, introducing a dependency on ERA5. IASI-NN uses the EUMETSAT NRT product as its training target, making it more independent; however, it incorporates MODIS emissivities as auxiliary input, linking it partially to the MODIS instrument which retrieves LST directly from infrared radiances (and is effectively independent of ERA5 Tskin). Finally, LST CCI, although also derived from infrared radiances, was intercalibrated against IASI (Ghent et al., 2025), introducing a degree of dependency between these two datasets. The ESA CCI/C3S SST and IST product is derived from an independent suite of satellite radiometers and is anchored to in-situ drifting-buoy measurements (Embury et al., 2024), making it largely independent of both the IASI and ERA5 datasets used here. Over sea-ice-covered regions, the definition and measurement of Tskin becomes more complex. Infrared satellite retrievals such as IASI and MODIS are sensitive to the radiometric skin temperature of the ice surface, but their accuracy depends on cloud screening and surface emissivity characterisation, both of which are particularly challenging in polar regions. ERA5 treats sea ice explicitly through a dedicated surface scheme. In this study, sea-ice-covered areas are not analysed separately and are included within SST, and the inter-

dataset biases observed at high latitudes and discussed later are partly attributable to these differences in sea-ice treatment across datasets."

- **L104** – "making a reference instrument" → **"making it a reference instrument."**

Done.

- **Table 1** – The table is not rendered correctly in the manuscript version I am reading.

This is Table 1. We will work eventually with the production team so that the Table is rendered correctly.

Table 1. Inputs and target variables for neural network training. Total features/inputs: 90 over sea and 94 over land. Target variable: EUMETSAT Tskin.

Sea	Land
Longitudes	Longitudes
Latitudes	Latitudes
87 radiance channels (scaled ×1e^5)	87 radiance channels (scaled ×1e^5)
Pixel number	Pixel number
–	Four CAMEL emissivity channels
Target: EUMETSAT Tskin	Target: EUMETSAT Tskin

- **Section 2.5.1** – It would be useful to explicitly state that this variability is completely missed in ERA5, which relies on static ancillary information.

We thank the reviewer for this comment. The following sentence *in italic* is now added:

"Across all four TIR channels, CAMEL exhibits predominantly negative emissivity trends, with spatially coherent patterns. Although the magnitude of these trends is small (>-0.1%/year), their persistence across channels suggests a systematic signal rather than random noise. The trends in emissivities therefore affect the Tskin trends and are a necessary input to the Tskin retrievals. *We note that this monthly and interannual emissivity variability is not accounted for in ERA5, which relies on static, time-invariant emissivity climatology*".

- **Figure S1** – This figure raises several questions. For example, what is the source of the statistically significant positive trends over desert regions?

  Also correct the typo in the caption: **"emissvities"** → **"emissivities."**

The typo is now corrected. The statistically significant positive trends over desert regions, in particular the Sahara, the Arabian Peninsula and parts of Central Asia are indeed intriguing, and difficult to explain. The following phrase is now added to the legend in Figure S1.

"Positive trends over arid and hyper-arid regions are likely associated with known limitations of the CAMEL dataset over deserts, where ASTER GED sampling is sparse and surface emissivity estimates are sensitive to episodic wetting and dust events."

- **L172** – "January 2018 onward" – onward until when? Please specify the last month included in the analysis.

This line is now changed to: January 2018 – *December 2022*.

- **L185** – It may be fairer to use ERA5-Land in this comparison, as the representation of land surface processes is significantly improved relative to ERA5.

Thank you for this comment. We do agree with you, however since we are now using the total cloud cover to clear ERA5 cloudy scenes, and it is not available in ERA5 Land, we choose to keep ERA5.

- **L195** – "The shift from ASTER to MODIS/Terra in April 2012" – this likely refers to a shift from **AATSR**, not ASTER.

This typo is now corrected, thanks.

- **L197** – "This version of the CCI product has not yet been fully validated."

  What aspects are still missing from the validation? Why is the dataset published and used here if it has not been fully validated?

The phrasing has been corrected to "not yet been validated" to avoid ambiguity. By this we mean that LST CCI v3.00 is a relatively recent product (August 2025) for which a comprehensive, peer-reviewed validation against independent datasets has not yet been published. Rather than relying on external validation, we use the intercomparison itself as an internal consistency check, and we explicitly flag the sensor-transition discontinuities visible in the anomaly time series as evidence of residual inhomogeneities that users should be aware of. The inclusion of an unvalidated product is therefore deliberate — identifying such issues across datasets is precisely one of the contributions of this intercomparison.

- **L230** – "More generally, the IASI-NN product is the only one that is clear-sky."

  This was not clear earlier in the manuscript. In addition, this statement is not entirely correct, as MODIS and LST CCI are also clear-sky products. Please highlight this earlier in Section 2, since this distinction is crucial for understanding differences among the datasets.

Indeed, as per the different reviewers' comments, all datasets are now cloud filtered. And a new section 2.8 is added as follows that addresses the Reviewer comment. Last part of this paragraph is discussed in a later comment related to the added L4 ESA CCI C3S SST.

"2.8 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 (tcc) variable, ERA5's native cloud diagnostic, retaining only grid points and time steps where tcc < 0.2 (20%). This clear-sky criterion is consistent with the approach used in comparable studies (Ermida et al., 2019; Ermida & Trigo, 2022), which also demonstrate through sensitivity analysis 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. Moreover, the ERA5 Tskin trends derived from monthly cloud-filtered data are shown to be robust across a wide range of cloud cover thresholds (Wang et al., 2022).

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 ESA SST CCI/C3S SST and IST product combines infrared clear-sky retrievals with all-sky passive microwave observations through optimal interpolation, and is therefore not cloud-filtered. Applying a cloud mask, a posteriori, using ERA5 monthly cloud cover fraction, for example, would be the only feasible approach to approximate clear-sky conditions, but this would introduce an ERA5-dependent filtering that is inconsistent with the cloud screening applied to IASI and MODIS, which is based on their own instrument-level detection. We therefore present the ESA CCI/C3S SST and IST product as an all-sky all-day reference."

The figures show that the comparison between the different instruments is now better.

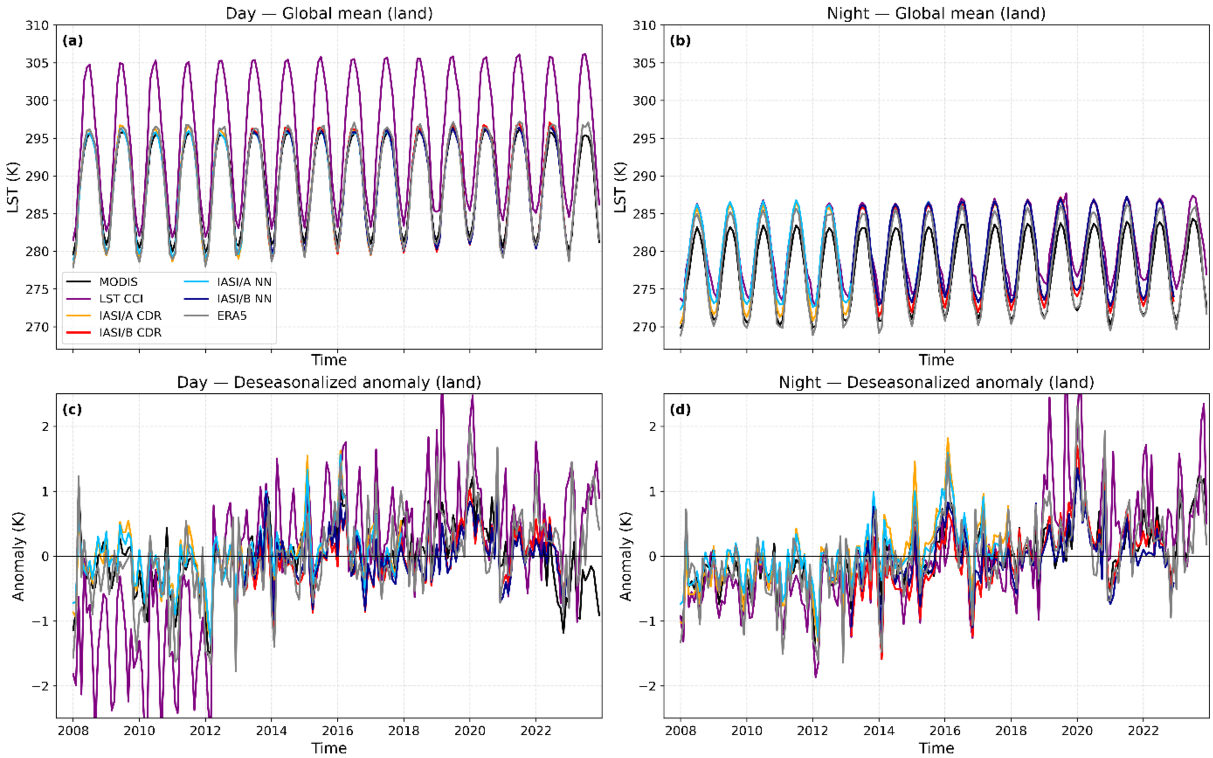


Figure 1. Global mean land surface temperature (LST, panels a and b) and deseasonalised anomalies (panels c and d) for daytime (left) and nighttime (right). IASI Metop-A and Metop-B observations (IASI-A/B) are shown separately for the CDR and NN products. Global means are computed using cosine-latitude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

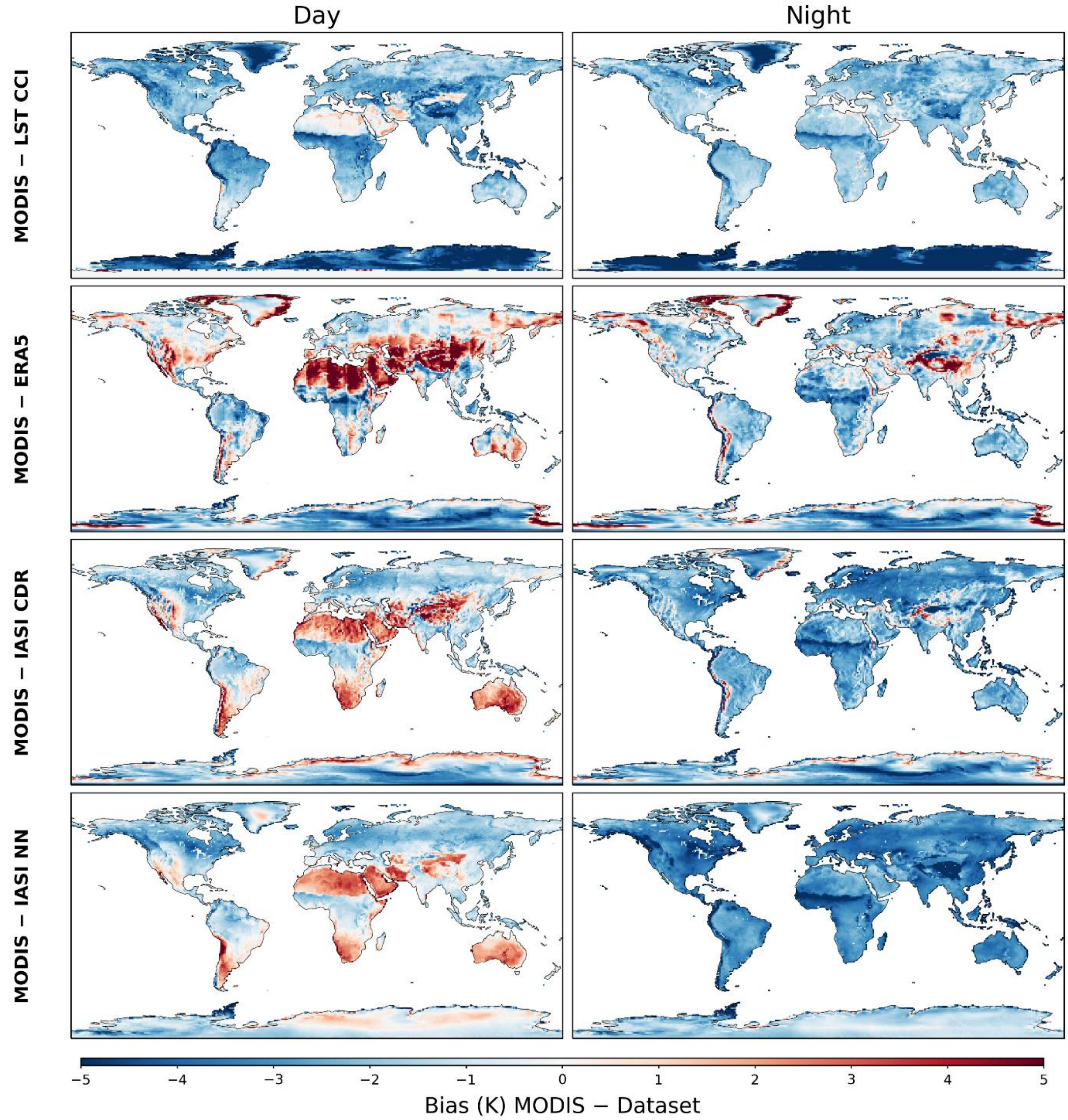


Figure 2. LST spatial bias averaged over the 2007-2022 period of MODIS during the day (left) and the night (right), with respect to ESA LST CCI, ERA5, IASI-CDR and IASI-NN. A *cold nighttime bias with respect to MODIS is evident across all datasets. ERA5 land/sea mask is used to mask IASI and ERA5 sea grid points.*

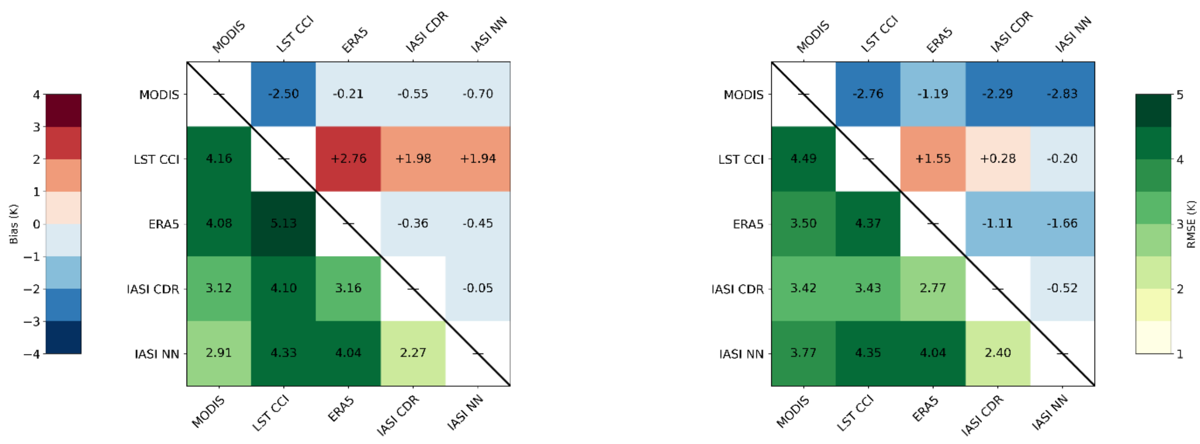


Figure 3. Upper triangle (blue to red): global mean bias (row – column, y -x): Lower triangle (in yellow to green): root mean square error (RMSE) between the different pairs of LST datasets, for day (left) and night (right). Statistics are computed over land grid cells only (using the ERA5 land-sea mask), weighted by cosine-latitude.

Figure 3 – The top and bottom axis labels are partially cut off. Please correct.

Figure 3 is now in the correct format.

- **Section 3.2** – A fair comparison would include SST from CCI. Please consider including it in a revised manuscript, as it would be valuable to assess how an independent satellite-based dataset represents these signals. Based on the results shown in Figure 5, I suggest using the SST/IST product:

  https://data.marine.copernicus.eu/product/SST_GLO_SST_L4_REP_OBSERVATIONS_010_024/description

  which properly accounts for temperatures over sea ice.

We thank the reviewer for this suggestion. As this dataset is independent from the other products used, it makes sense to use it too for our validation. But it is worth noting that is it not straightforward as in the new version of the manuscript, all datasets are clear-sky. This L4 product, is not. Moreover, this is a joined (day and night datasets), and our analysis separates day and night. However, according to the ESA website, only this L4 dataset covers (homogeneously) our study period over 2008 to 2022 as shown in Figure Review 1, so indeed this is the right reference dataset to use.

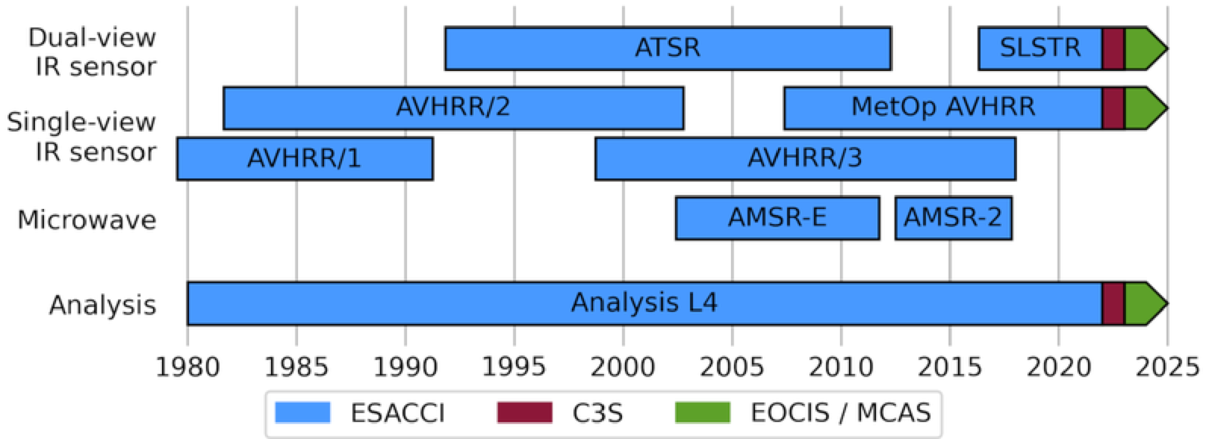


Figure Review 1. Overview of products included in SST-CCI CDR v3.0. Figure from https://climate.esa.int/en/projects/sea-surface-temperature/data/

This section 2.3 has been added, and Figures 4, 5 and 6 have been updated:

"2.3 ESA CCI and Copernicus climate change service (C3S) Reprocessed SST/IST

To extend the ocean surface temperature comparison with an independent satellite-based dataset, we include the ESA SST CCI and C3S Reprocessed Sea Surface and Sea Ice Temperature product, distributed through the Copernicus Marine Service (Høyer et al., 2014; Høyer & She, 2007; Nielsen-Englyst et al., 2023). We use the variable "st", which corresponds to the analysed sea (at 20 cm depth) and ice surface (surface skin) temperature. This Level-4 (L4) product provides gap-free daily analyses at a native horizontal resolution of 0.05° × 0.05°.

The product is generated using the Danish Meteorological Institute Optimal Interpolation (DMIOI) system, which combines input from multiple satellite sources (Embury et al., 2024); and IST retrievals from the AASTI (Arctic and Antarctic Surface Temperatures from thermal Infrared satellite radiometers) and C3S IST CDR/Interim CDR v.1. The inclusion of IST, constrained by a multi-source sea-ice concentration composite, makes this product particularly suited for representing surface temperatures in polar regions, the sea ice, and the marginal ice zone — areas where other SST products are typically undefined or unreliable. The product is extracted over the study period of 2008–2022, which we regrid to a 1° × 1° grid for consistency with the other datasets used in this study. In this work, this dataset is the only one that is all-sky, and not separated between day and night. In this study, we call it ESA CCI/C3S SST and it serves as such as a reference to the other SST datasets."

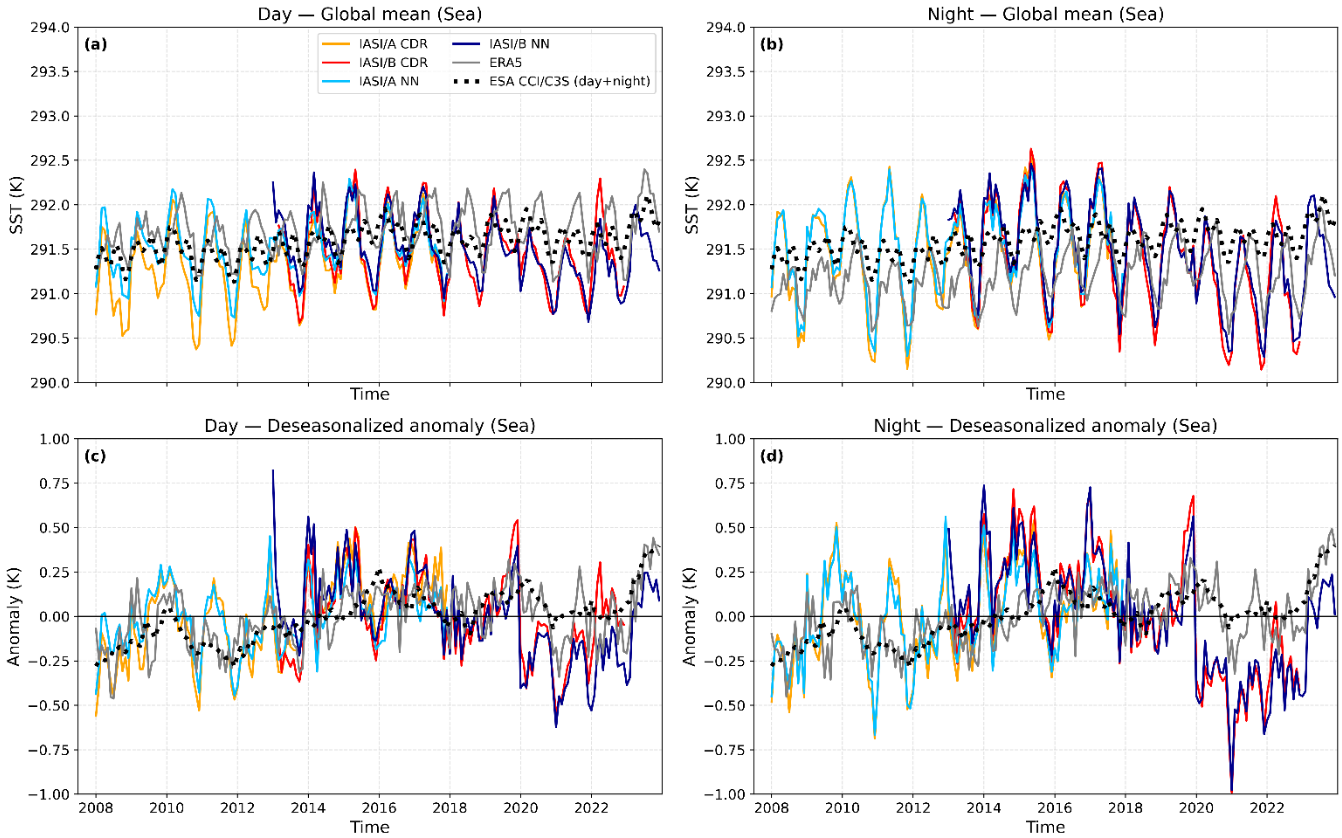


Figure 4. Upper panels: SST global mean during the day (panel a) and the night (panel b) for ERA5, IASI-CDR and IASI-NN**,** and *ERA5 CCI and C3S SST all-sky product which is shown in dotted black line. It is the same for the day and night panels as it is a combined product.* IASI data are separated between Metop A and Metop B. Lower panels: the deseasonalized anomaly of SST for the different products, during the day (panel c) and night (panel d). Global means are computed using cosine-latitude weighting.

The following sentence have been added to the discussion of Figure 4, as well as a sentence to explain the differences seen in 2020 and 2021 for IASI-B (both CDR and NN).

"The deseasonalized anomaly time series (panels c and d) show strong consistency across all datasets in both timing and amplitude of interannual variability, including the 2015–2016 El Niño warming and the subsequent 2020–2022 La Niña cooling. Larger inter-dataset differences, in particular post-2020 for IASI (CDR and NN), reflect systematic offsets attributable to differences in clear-sky sampling strategy rather than spurious trends or variability. The stronger negative anomalies in IASI during 2020–2022 are consistent with the La Niña period, during which increased cloudiness in the tropics reduces the availability of clear-sky retrievals, causing the IASI global mean to be cooler. The ESA CCI/C3S SST product shows systematically damped anomaly peaks in both the positive and negative directions relative to the clear-sky datasets, consistent with the spatial and temporal smoothing inherent to the optimal interpolation scheme used to produce gap-free, all sky, all day coverage, which reduces the amplitude of extreme anomalies by construction."

New figure 5: Since now we have 4 SST datasets (and to not have a large figure of 6x2 panels), we use the IASI NN product as a reference (as it is ours, and we want to indirectly validate it). The new Figure 5 looks as follows:

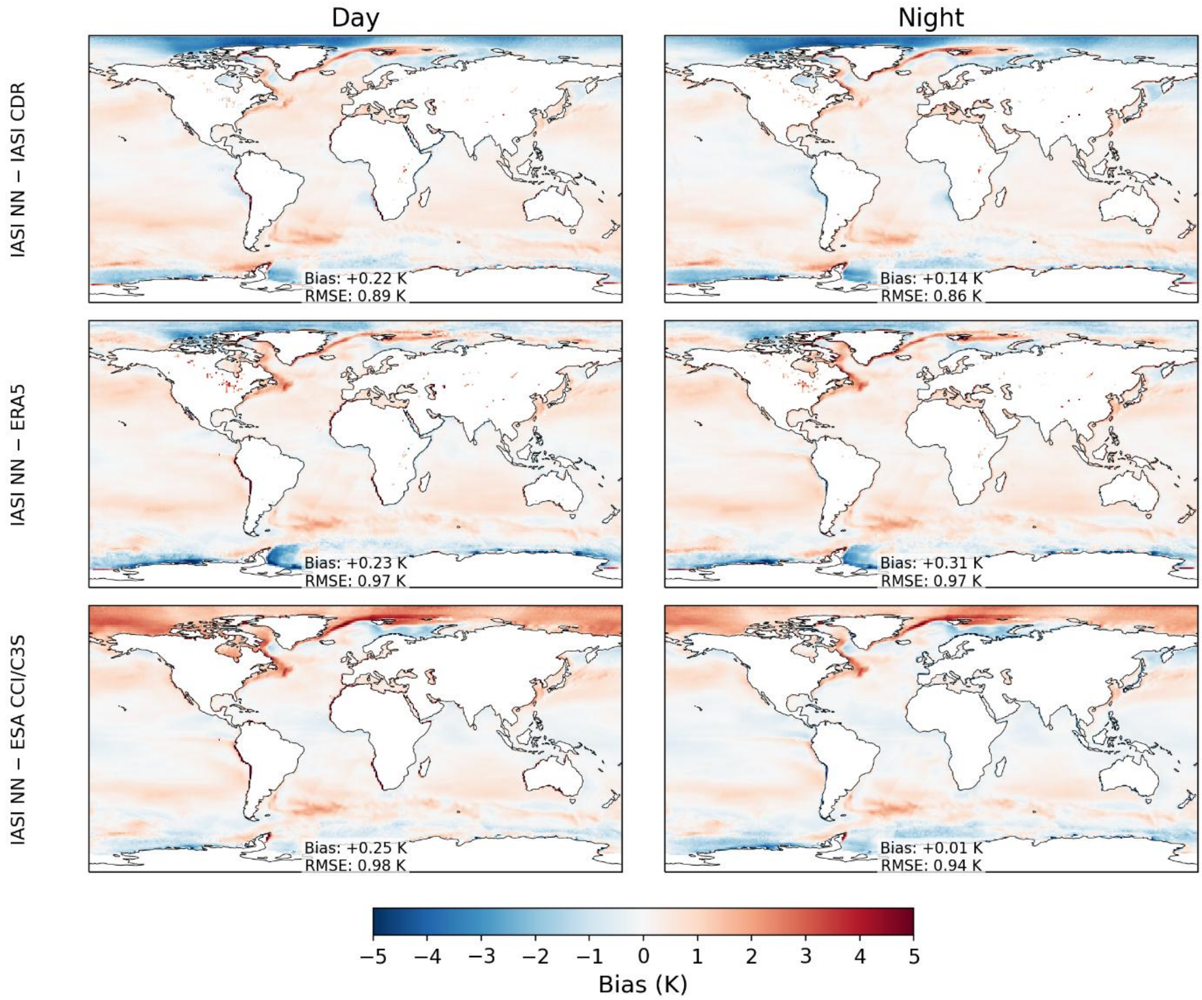


Figure 5. Spatial bias of IASI-NN ocean skin temperature relative to IASI-CDR, ERA5, and the ESA CCI/C3S SST analysis. Maps show daytime (left column) and nighttime (right column) biases, computed as the time mean of the monthly-mean differences over the datasets' common period (2008–2022). The ESA CCI/C3S product is a gap-filled, all-sky, daily-mean analysis with no day/night separation, so the same daily field is differenced against both the daytime and nighttime IASI-NN fields; IASI and ERA5 are restricted to clear-sky scenes. All datasets were regridded to a common 1°×1° grid and land points were masked using the ERA5 land–sea mask. Global mean bias and RMSE, area-weighted by the cosine of latitude, are indicated in each panel.

The updated discussion is now:

"Over the open ocean, IASI-NN agrees closely with all three products: global mean biases are small (+0.01 to +0.31 K), RMSE stays below 1 K in every panel, and differences across most of the ice-free ocean lie within ±1 K. Thin warm filaments along the western boundary currents (Gulf Stream, Kuroshio) mark sharp SST fronts that are smoothed in the gap-filled ESA CCI/C3S

analysis. The largest discrepancies are confined to high latitudes and differ in sign between products. Sharing the same instrument and the same clear-sky sampling, IASI-NN and IASI-CDR isolate the effect of the retrieval method: they show the closest agreement of all pairs (RMSE < 0.9 K), with IASI-NN colder than IASI-CDR over the Arctic, consistent with IASI-CDR being trained on, and therefore not independent of, ERA5 (Sect. 2.7). Relative to ERA5, IASI-NN is colder over the Southern Ocean, in line with ERA5's documented warm bias over sea ice (Graham et al., 2019; Herrmannsdörfer et al., 2023) ; relative to the all-sky ESA CCI/C3S analysis it is instead warmer across the Arctic. These contrasting signs indicate that the high-latitude spread reflects the differing definition, retrieval and sea-ice treatment of skin temperature across the datasets (Sect. 2.7) rather than a single common cause:"

L280 – Given the large daytime inhomogeneity for CCI in 2012, a cautionary note regarding the interpretation of trends should appear at the beginning of Section 4. Currently, this is only mentioned later in the discussion.

This cautionary note is now added at the beginning of Section 4:

"We first note that LST CCI daytime trend results should be interpreted with caution. As shown in Section 3, LST CCI exhibits a marked daytime inhomogeneity around 2012, associated with the sensor transition from AATSR to MODIS Terra within the CCI processing chain. This discontinuity affects the daytime time series over land and may introduce spurious signals in the derived trends."

- **L310 –** "except IASI-NN ones over sea." This statement appears somewhat overstated. Even over land, the areas showing significant trends are quite limited.

To avoid overstatement, this phrase has been removed.

- Although significant trends are shown in the emissivity maps in Figure S1, no interpretation is provided regarding how these trends may affect the LST trends in datasets that rely on these emissivity inputs.

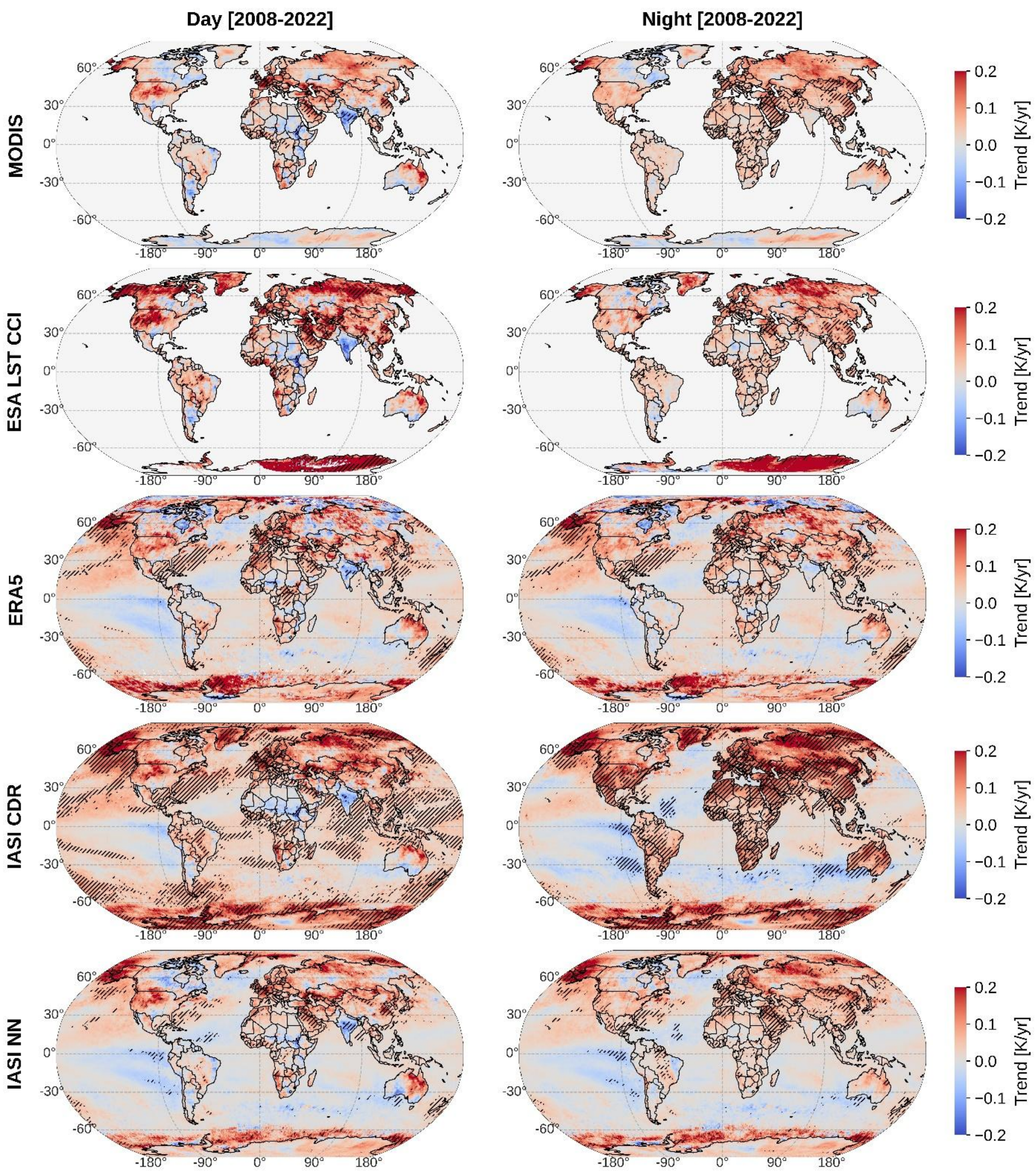
We'd like to note that little literature review exists on the subject as LST and emissivity are interdependent. However, it is indeed interesting to mention the added value of using monthly emissivities. The following discussion is now added in order to highlight the effect of using a monthly variable emissivity trend:

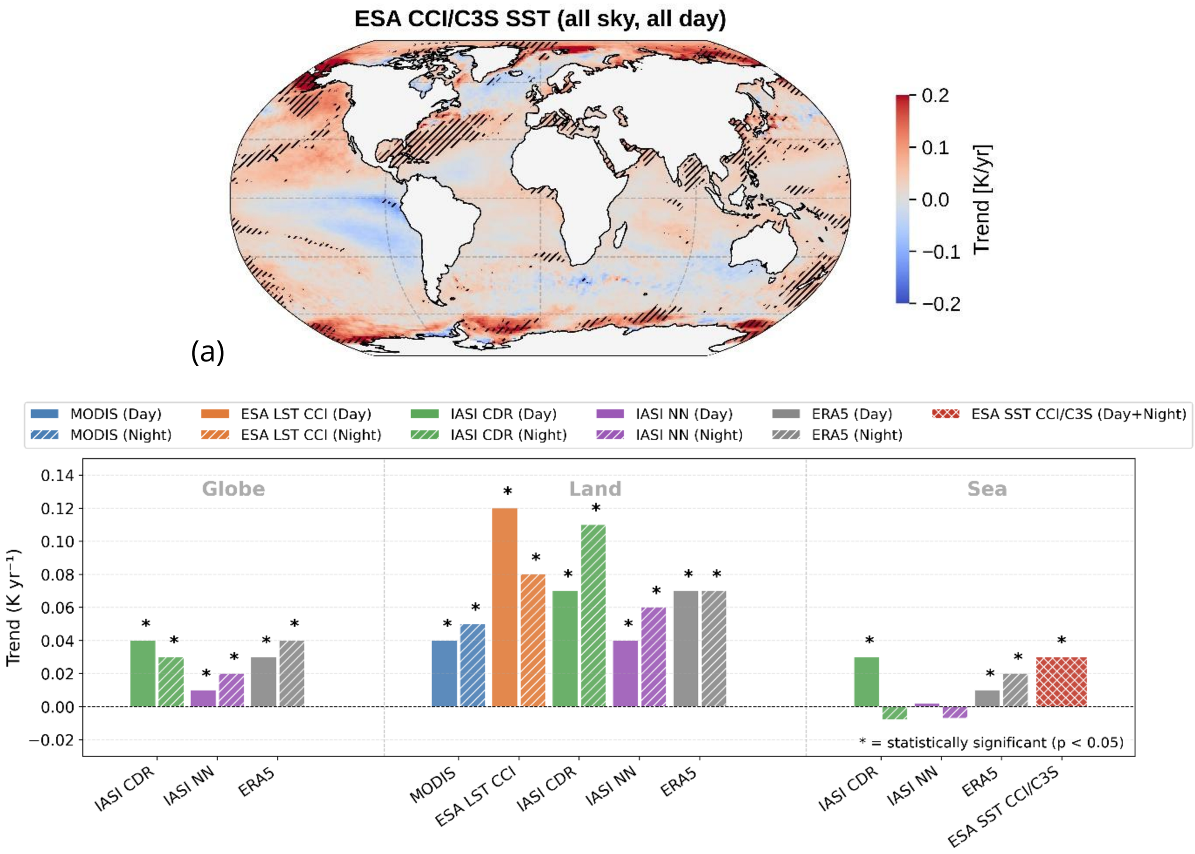
"*The inclusion of time-varying CAMEL emissivities in the IASI-NN retrieval has implications for the derived LST trends. For example, Zhou et al. (2021) showed that neglecting emissivity trends leads to an overestimation of the global mean Tskin trend by approximately 4.9 × 10 $^{-3}$ K/yr (over the period 2008 to 2020). By explicitly incorporating monthly CAMEL emissivity fields and their interannual variability, the IASI-NN is therefore expected to yield more physically realistic LST trends than datasets relying on static emissivity climatology, such as ERA5."*

- **Figure 6 –** It is confusing that LST CCI provides values over sea ice. These were masked in the comparisons shown in Figure 2 but not in Figure 6. Please ensure consistency (also in Figure S2).

Additionally, correct the occurrences of **"°K"** in the caption.

Indeed. This is now corrected for the LST CCI, and the sea ice values are masked using the ERA5 land sea mask. The figure caption is corrected as well for K/yr instead of °K/year. We added the SST trends as suggested by the reviewer in a previous comment. This is the updated Figure 6, panels (a) and (b). The SST trends agree well with our results.





(a)

(b)

Figure 6. (a) Spatial trends [K/yr] during day (left column) and night (right column) for MODIS, ESACCI LST, ERA5, IASI-CDR, IASI-NN, and ESA CCI/C3S SST (all sky, all day,) over the period January 2008-December 2022. Stipples correspond to regions where trends are significant according to the standard Mann-Kendall test (p-value < 0.05). (b) Global (Tskin), sea (SST), land (LST) trends (K/yr) from MODIS, ESA CCI LST, IASI-CDR, IASI-NN, ERA5, and ESA CCI/C3S SST over the period [January 2008–December 2022]. Trends are computed with the Theil–Sen estimator on annual latitude-weighted and deseasonalised averages. Significant trends are marked with an Asterix (*) above the bars.

Reviewer 3

This study compares several skin temperature products: four remote sensing products and one reanalysis (ERA5) product. It considers their spatial biases as well as biases in global-mean trends. In principle, if carefully done, this is a valuable exercise and could be very useful to the community. In its current form, however, large ambiguities in make the study's results difficult to interpret. I recommend major revisions.

We thank the reviewer for their comments. We revised the manuscript. Our responses are listed in blue hereafter. Added text is shown in italic.

General Comments

Please address what "skin temperature" actually means. For the ocean surface, does it refer to the upper micrometers? Millimeters? Does the quantity measured depend on the remote sensing method? There can be significant thermal contrasts in just the very top of the ocean surface, so this is important to address.

The following sentence was added:

"Over the ocean, the term "skin temperature" specifically refers to the radiometric temperature of the uppermost ~10–20 µm of the water column — the so-called cool skin layer — which is typically 0.1–0.5 K cooler than the water just below, owing to net heat loss at the air–sea interface (Donlon et al., 2002; Minnett et al., 2019)."

Sea ice: How do these different datasets measure the skin temperature in areas with sea ice, and how might this affect the overall results of your study?

Indeed. We now acknowledge the limitation of infrared sounders over sea ice. We add the following in the added section 2.7:

"Over sea-ice-covered regions, the definition and measurement of Tskin becomes more complex. Infrared satellite retrievals such as IASI and MODIS are sensitive to the radiometric skin temperature of the ice surface, but their accuracy depends on cloud screening and surface emissivity characterisation, both of which are particularly challenging in polar regions. ERA5 treats sea ice explicitly through a dedicated surface scheme. In this study, sea-ice-covered areas are not analysed separately, and the inter-dataset biases observed at high latitudes and discussed later are partly attributable to these differences in sea-ice treatment across datasets."

Clouds: it doesn't make sense to directly compare temperatures in clear-sky and all-sky datasets.

All datasets are now cloud-filtered (based on the comments of Reviewers 1 and 2). The following section has been added, and the figures and the analysis updated.

"2.8 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 (tcc) variable, ERA5's native cloud diagnostic, retaining only grid points and time steps where tcc < 0.2 (20%). This clear-sky criterion is consistent with the approach used in comparable studies (Ermida et al., 2019; Ermida & Trigo, 2022), which also demonstrate through sensitivity analysis 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. Moreover, the ERA5 Tskin trends derived from monthly cloud-filtered data are shown to be robust across a wide range of cloud cover thresholds (Wang et al., 2022).

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 ESA SST CCI/C3S SST and IST product combines infrared clear-sky retrievals with all-sky passive microwave observations through optimal interpolation, and is therefore not cloud-filtered. Applying a cloud mask, a posteriori, using ERA5 monthly cloud cover fraction, for example, would be the only feasible approach to approximate clear-sky conditions, but this would introduce an ERA5-dependent filtering that is inconsistent with the cloud screening applied to IASI and MODIS, which is based on their own instrument-level detection. We therefore present the ESA CCI/C3S SST and IST product as an all-sky all-day reference."

It's unclear to me how discontinuities in satellite observations in space and time are accounted for in this analysis. In spatially and temporally complete SST or near-surface land temperature products, algorithmic adjustments are made to fill in spatial and temporal gaps, such that averaged values can be interpreted to represent a best estimate of a global mean or a daily/monthly mean. To what extent are these observational products homogenized? See comments below.

Temporal coverage: Are "daytime" temperatures specifically referring to 9:30am or 10am temperatures? Please be more specific about how the correction to the temperatures based on local crossing times are made.

Yes. Daytime and nighttime temperatures in this study refer to the morning (~09:30–10:30 local solar time) and evening (~21:30–22:30 local solar time) overpasses respectively. We make this clear in the beginning of Section 3:

"In terms of temporal sampling, all datasets are evaluated at comparable local solar times (except for the ESA CCI/C3S SST), spanning a window of approximately 09:30 to 10:30 for daytime and 21:30 to 22:30 for nighttime. Specifically, IASI crosses the equator at 09:30 local

solar time, MODIS Terra at approximately 10:30, and LST CCI is provided at 10:00. ERA5 hourly data are sampled at 10:00 local solar time by applying a longitude-based offset (15° longitude = 1 hour), without correction for daylight saving time, consistent with the fixed solar crossing times of polar-orbiting satellites. No further correction for the residual 30–60 minute spread between instruments is applied beyond what is already performed internally within the LST CCI product (Ghent et al., 2025). At the monthly timescale used throughout this study, this residual temporal offset is not expected to introduce biases of climatological significance."

Spatial coverage: Is the spatial coverage clear sky or all sky? Are we considering a true global mean? If so, how are areas not "seen" by the satellite filled in? These are all major ambiguities that must be accounted for and described before making an "apples to apples" comparison between data products.

The current products are now cloud free as per the Reviewers' suggestion. Indeed, this means that there will be areas that are not "seen" by the satellite, typically around the Equator. And these will depend on each of the products' cloud coverage algorithms. This applies to all products, including ERA5, when we use the cloud filtering approach (see added Section 2.8).

Generally, these issues must be treated with more specificity, rather than with broad allusion, for the comparisons made to be useful and interpretable.

In order to make the analysis concise and clear, the following paragraph is now added as an introduction to section 3 Results. It serves as a summary of our approach to dataset harmonization and comparability.

"In terms of temporal sampling, all datasets are evaluated at comparable local solar times (except for the ESA CCI/C3S SST), spanning a window of approximately 09:30 to 10:30 for daytime and 21:30 to 22:30 for nighttime. Specifically, IASI crosses the equator at 09:30 local solar time, MODIS Terra at approximately 10:30, and LST CCI is provided at 10:00. ERA5 hourly data are sampled at 10:00 local solar time by applying a longitude-based offset (15° longitude = 1 hour), without correction for daylight saving time, consistent with the fixed solar crossing times of polar-orbiting satellites. No further correction for the residual 30–60 minute spread between instruments is applied beyond what is already performed internally within the LST CCI product (Ghent et al., 2025). At the monthly timescale used throughout this study, this residual temporal offset is not expected to introduce biases of climatological significance. Furthermore, all datasets are evaluated under clear-sky conditions, as described in detail in Section 2.8. It should be noted that the cloud screening method differs necessarily across datasets, as each product applies its own retrieval-specific approach. While this means the clear-sky subsets are not strictly identical across products, no universally applicable cloud mask exists that could be applied consistently across such diverse observing systems and reanalysis products, and the approach adopted here represents the most physically appropriate filtering strategy for each dataset. One additional asymmetry in the cloud filtering strategy concerns ERA5. Unlike the satellite products, for which cloud screening is applied at the individual retrieval or overpass level before monthly aggregation, ERA5 cloud filtering is applied to monthly-mean total cloud cover. This means that months with episodic cloudiness but low mean cloud cover may be retained, resulting in a less strict effective clear-sky selection than for the satellite datasets."

On the use of reanalysis: in this study, reanalysis is treated as if it were another observational product. However, reanalysis is a fundamentally different type of data product, as it assimilates observations into a model, and prognostically predicts skin temperature. It is interesting to include reanalysis but these distinctions should be maintained throughout the text.

Considering the other Reviewers comments the ERA5 section has been rewritten to clarify Tskin is a prognostic variable. Here is the new section 2.3

"ERA5 Tskin

We use the Tskin variable from the ERA5 reanalysis (Hersbach et al., 2020), produced by ECMWF, which is framed within the Copernicus Climate Change Service (C3S) of the European Commission. ERA5 datasets are at 0.25° × 0.25° resolution (native horizontal resolution of ERA5 is ~31km), that are regridded to a 1°x1° grid size for the study.

Hourly averages are computed for each month of the year and are then converted to local time by calculating the time zone offset based on the longitude (15° = 1 hour). Our analysis does not account for daylight saving time. This is consistent with satellite local-solar-time definitions, as polar-orbiting instruments such as IASI maintain fixed solar crossing times (e.g., 09:30/21:30), independent of civil time zones.

"ERA5 should not be treated as an independent ground truth, but rather as a physically consistent reference that is subject to its own model-dependent uncertainties. Over land, ERA5 Tskin is a prognostic variable computed within the HTESSEL (Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land) land surface scheme (Balsamo et al., 2009)*, as part of the Integrated Forecast System (IFS). It is integrated forward in time by solving the surface energy balance at each model timestep, and is therefore sensitive to model parameterisations of soil thermal properties, vegetation, and snow. Over ocean, ERA5 Tskin is not prognostic but is instead derived from a prescribed SST analysis with a diagnostic cool-skin correction applied* (Hirahara et al., 2016; Zeng & Beljaars, 2005)*. ERA5 does not directly assimilate Tskin, but does assimilate SST analyses over ocean and infrared radiances over both land and sea* (Hersbach et al., 2020)*, which carry indirect surface temperature information."*

Also, have there been comparisons between skin temperature values in other reanalysis products?

We thank the reviewer for raising this point. Skin temperature has been intercompared across reanalysis products in few previous studies, although such comparisons remain less common than those for 2 m air temperature and have largely focused on the marine domain (SST) rather than land. Over the ocean, SST in MERRA-2 and ERA-Interim has been evaluated against independent shipborne M-AERI measurements from the AEROSE cruises, with mean differences on the order of 0.1 K (Luo et al., 2020), and ERA5 SSTskin has been similarly assessed against M-AERI data (Luo & Minnett, 2020). At the global scale, surface skin temperature trends over 2003–2021 have been compared between ERA5 and JRA-55 and found to be in good mutual agreement and consistent with AIRS satellite observations (N. Wan et al., 2024). These studies indicate broad consistency in skin temperature between modern reanalyses, particularly in trend patterns, while also showing that systematic differences arise from each product's cool-skin parameterization and assimilated observations. Our analysis complements this body of work by extending the comparison to satellite-derived skin/land surface temperature products (IASI, MODIS, ESA CCI) alongside ERA5, with explicit separation of day and night orbital phases.

Since there are little literature review on LST, we choose not to add anything to the main manuscript.

Some figure captions are not adequately descriptive.

Most captions have been updated with more description (see for example the captions for Figures 3 and 5 hereafter).

Data availability: can you make the regridded and harmonized (to local time) data available? Or the code used to perform the harmonization?

The code is now available here (added to the section data availability statement):

https://github.com/sarahsafieddine/utc-to-local-time/tree/main

Specific comments

Introduction: As written, the introduction doesn't provide adequate context for the reader, instead jumping immediately into the details of each dataset. The paper should be accessible to researchers who may be interested in using skin temperature for scientific analysis, but may not be already familiar with each of the individual data products. In the introduction, therefore, I would suggest the following. Around line 40, instead of or in addition to presenting a general statement of Planck's law, some broad discussion of remote sensing methods capable of measuring skin temperature and the strengths and weaknesses of these methods would be warranted. I.e., infrared sensing offers a good measure of physical temperatures as surfaces tend to have emissivities close to unity at thermal IR wavelengths, but infrared sensing of surface temperature is only possible in cloud-free areas, and there is some sensitivity to water vapor as well. Microwave sensing is less sensitive to clouds, but emissivity vary and the resolution is less sharp… etc. Satellites have incomplete spatial coverage, moreover… Basically, before you discuss each of the individual datasets, it would be helpful to discuss: what are some of the key sources of uncertainty that could lead to variation across these products?

We thank the reviewer for this recommendation. We now added the following paragraph after Planck's Equation:

"Satellite-based Tskin retrievals rely primarily on thermal infrared (TIR) radiometry in atmospheric window channels (8–13 μm), where most natural surfaces have emissivities close to unity, making TIR-derived temperatures a close approximation of the physical skin temperature (Guillevic et al., 2018; Hulley et al., 2019). However, TIR retrievals are only possible under cloud-free conditions, and all TIR observations are affected by atmospheric water vapour absorption. Passive microwave radiometry is less sensitive to cloud cover but surface emissivity in the microwave is highly variable, particularly over land and sea ice, and spatial resolution is considerably coarser (Prigent et al., 2016). Reanalysis products such as ERA5 provide spatially complete, all-sky fields but Tskin is a prognostic model variable constrained by data assimilation rather than a direct observation, making it sensitive to model parameterisations of surface energy balance and land surface properties (Hersbach et al., 2020). These fundamental differences in measurement principle, spatial sampling, and retrieval methodology mean that no single dataset can be considered a universal reference, and systematic intercomparison across products is essential to assess the robustness of Tskin-based climate signals (Ermida et al., 2019; Hulley et al., 2019). »

Line 70: you don't need to repeat that you regridded to 1x1 for every dataset. Also to clarify, did you regrid for all the analysis, or just the spatial comparison?

This line was removed. The regridding is done for all the analysis as it is now clear in the introductory paragraph of Section 3.

Line 86: Please add a little more detail about how this adjustment was made.

The phrase "The MODIS brightness temperature data was also intercalibrated against IASI." Is now changed to:

"In addition, an adjustment is made to account for the half-hour difference in the satellites' equatorial crossing times; this correction is computed in brightness temperature space using radiative transfer simulations."

Line 90: ERA5 is reanalysis, not an observational product. How is surface temperature data assimilated into ERA5?

This paragraph is now added (also added as a response to a previous comment):

"*ERA5 should not be treated as an independent ground truth, but rather as a physically consistent reference that is subject to its own model-dependent uncertainties. Over land, ERA5 Tskin is a prognostic variable computed within the HTESSEL (Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land) land surface scheme* (Balsamo et al., 2009)*, as part of the Integrated Forecast System (IFS). It is integrated forward in time by solving the surface energy balance at each model timestep, and is therefore sensitive to model parameterisations of soil thermal properties, vegetation, and snow. Over ocean, ERA5 Tskin is not prognostic but is instead derived from a prescribed SST analysis with a diagnostic cool-skin correction applied* (Hirahara et al., 2016; Zeng & Beljaars, 2005)*. ERA5 does not directly assimilate Tskin, but* *does assimilate SST analyses over ocean and infrared radiances over both land and sea* (Hersbach et al., 2020)*, which carry indirect surface temperature information."*

Section 2: In general, what criteria did you use to select which datasets to analyze?

The datasets were selected based on three main criteria: availability, overpass time consistency, and complementarity. Overpass time is a particularly critical constraint — products with significantly different equatorial crossing times, such as AIRS, cannot be meaningfully compared with IASI without introducing substantial diurnal sampling biases. A second motivation was the validation of our newly developed IASI-NN product, for which the EUMETSAT IASI-CDR constitutes a natural comparator, as both are derived from the same instrument. To broaden the intercomparison, we additionally included a well-established heritage satellite product (MODIS Terra), a multi-sensor climate data record (ESA LST CCI v3.00), and a reanalysis product (ERA5) as a physically consistent reference. Together, these five datasets span a range of retrieval methodologies, temporal homogenisation strategies, and spatial sampling approaches, making them well suited for a systematic intercomparison of Tskin-based climate signals.

We choose to not add anything to the manuscript since everything we mentioned in our answer can be now implied in this new version of the manuscript.

Line 197: This version of the CCI product "has not been fully validated" – so it has been partly validated already? If so, how? What are the implications of using unvalidated data in your analysis?

We thank the reviewer for this comment. The phrasing has been corrected to "not yet been validated" to avoid ambiguity. By this we mean that LST CCI v3.00 is a recently released product (August 2025) for which a comprehensive, peer-reviewed validation against independent datasets has not yet been published. Rather than relying on external validation, we use the intercomparison itself as an internal consistency check, and we explicitly flag the sensor-transition discontinuities visible in the anomaly time series as evidence of residual inhomogeneities that users should be aware of. The inclusion of an unvalidated product is therefore deliberate — identifying such issues across datasets is precisely one of the contributions of this intercomparison.

Line 213: Please describe the local time conversion method.

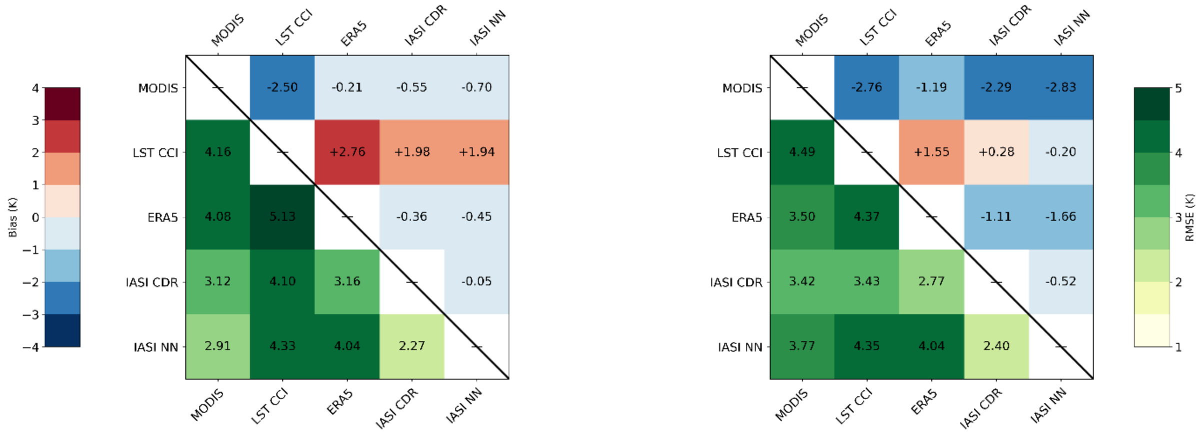
The local time conversion method is described in Section 2.4 as follows:

"Hourly averages are computed for each month of the year and are then converted to local time by calculating the time zone offset based on the longitude (15° = 1 hour). Our analysis does not account for daylight saving time. This is consistent with satellite local-solar-time definitions, as polar-orbiting instruments such as IASI maintain fixed solar crossing times (e.g., 09:30/21:30), independent of civil time zones."

… and has now been explicitly referenced at this point in the text to guide the reader directly to the relevant explanation.

Figure 3: I'm confused about what these global mean values are meant to capture. What happens over sea ice, what happens over

The ERA5 land-sea mask, which is used to separate land and ocean grid cells in Figure 3 (which has been introduced in this new version of the manuscript), classifies sea ice as ocean rather than land. Sea-ice-covered grid cells are therefore excluded from the land statistics shown in Figure 3. The figure caption has been updated to make this explicit as such:



"Figure 3. *Upper triangle (blue to red): global mean bias (row – column): Lower triangle (in yellow to green): root mean square error (RMSE) between the different pairs of LST datasets, for daytime (left) and nighttime (right). Statistics are computed over land grid cells only (using the ERA5 land-sea mask) and weighted by cosine-latitude."*

Figure 4: how do you treat sea ice?

Ocean grid cells are defined using the ERA5 land-sea mask, which classifies sea ice as ocean; sea-ice-covered areas are therefore included in the global mean SST and spatial bias maps. The three datasets treat sea ice differently: ERA5 computes Tskin over sea ice through a dedicated sea-ice scheme, while IASI-NN and IASI-CDR retrieve the radiometric skin temperature of the ice surface where cloud-free conditions allow, with reduced accuracy due to variable emissivity and low thermal contrast. These differences in sea-ice treatment contribute to the larger inter-dataset biases observed at high latitudes in Figure 5, as already noted in Section 2.7.

Moreover, we have now added the L4 ESA CCI and C3S combined SST product that provides temperatures over sea and ice surfaces, based on Reviewer 2 comments. The discussion of the SST is updated throughout the manuscript.

In order to make this clear, we have added the following into the legend of Figure 5:

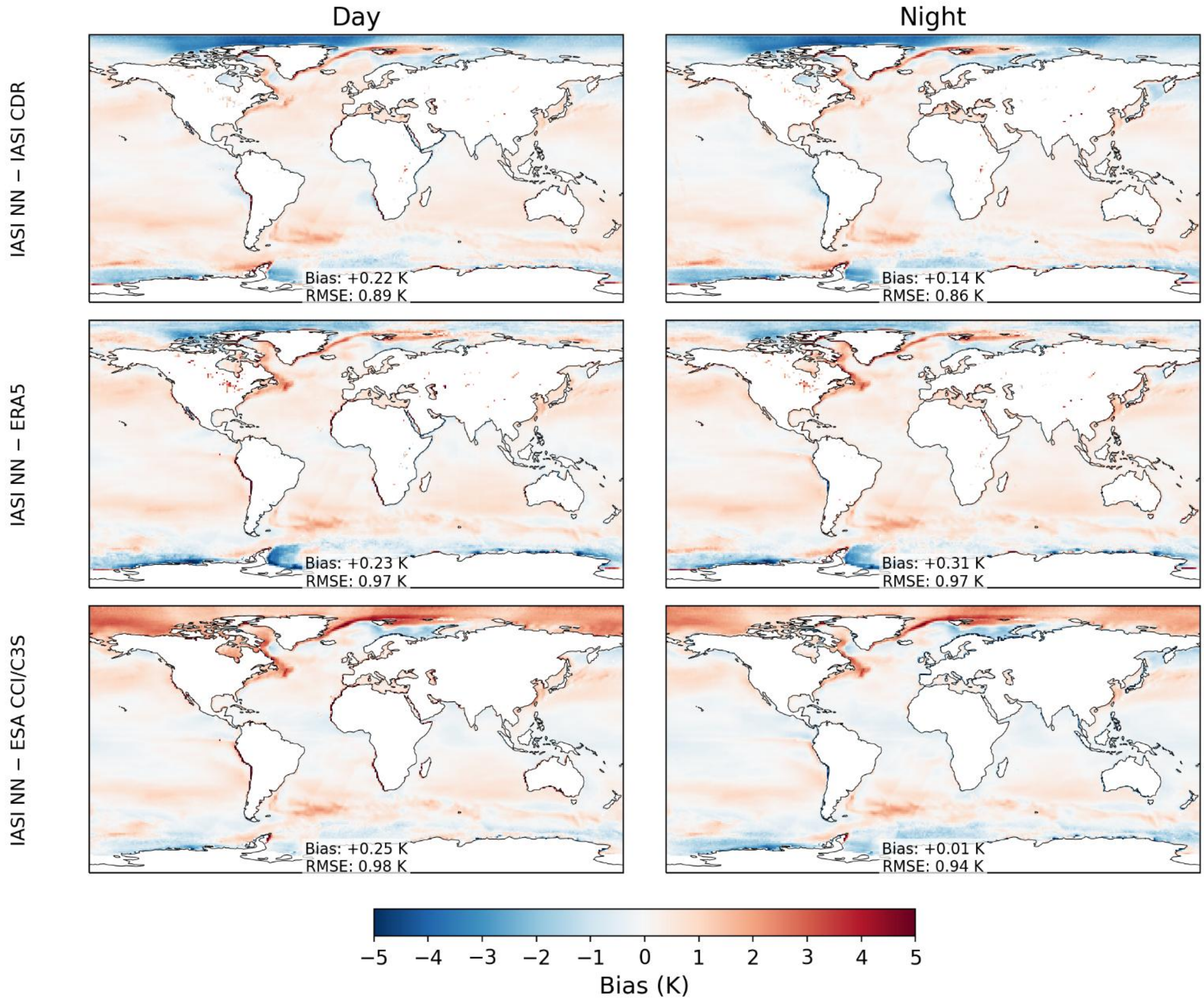


Figure 5. Spatial bias of IASI-NN ocean skin temperature relative to IASI-CDR, ERA5, and the ESA CCI/C3S SST analysis. Maps show daytime (left column) and nighttime (right column) biases, computed as the time mean of the monthly-mean differences over the datasets' common period (2008–2022). The ESA CCI/C3S product is a gap-filled, all-sky, daily-mean analysis with no day/night separation, so the same daily field is differenced against both the daytime and nighttime IASI-NN fields; IASI and ERA5 are restricted to clear-sky scenes. All datasets were regridded to a common 1°×1° grid and land points were masked using the ERA5 land–sea mask. **Sea-ice-covered grid cells are classified as ocean by this mask and are therefore included.** Global mean bias and RMSE, area-weighted by the cosine of latitude, are indicated in each panel.

Line 251: Biases between the SST datasets?

Yes, we rephrased the sentence to "biases between the datasets over the sea".

Line 284: historical trends – over what period?

We believe that the Reviewer is referring to this line/sentence:

“Both observations and general circulation model projections suggest current and future significant temperature increases in Siberia consistent with the effects of Arctic amplification, with an increase mostly pronounced in central Siberia reaching 4°C over the past few decades (Masson-Delmotte, et al., 2021; Tchebakova et al., 2009).”

We add more information on the period and rephrase the sentence as such:

“Both observations and general circulation model projections suggest significant temperature increases consistent with Arctic amplification*, with the Arctic warming nearly four times faster than the global average over 1979–2021* (Masson-Delmotte, et al., 2021; Rantanen et al., 2022)*. This broad Arctic warming pattern is consistent with the strong positive Tskin trends observed in Alaska and Siberia in our 2008–2022 dataset.”*

Line 295: Regarding extreme rainfall events, how is this related to heating and cooling trends?

This sentence suggests reasons to explain the cooling trends. We rephrase it to make this clear:

“This cooling trend is also seen in central and east Africa. In warm tropical environments, *increased cloud cover and latent heat release associated with more frequent and intense extreme rainfall events can suppress surface temperatures (Hegerl et al., 2007), providing a physical mechanism consistent with the observed cooling signal.”*

Figure 6: What is the color coding?

This whole figure has now been updated as follows:

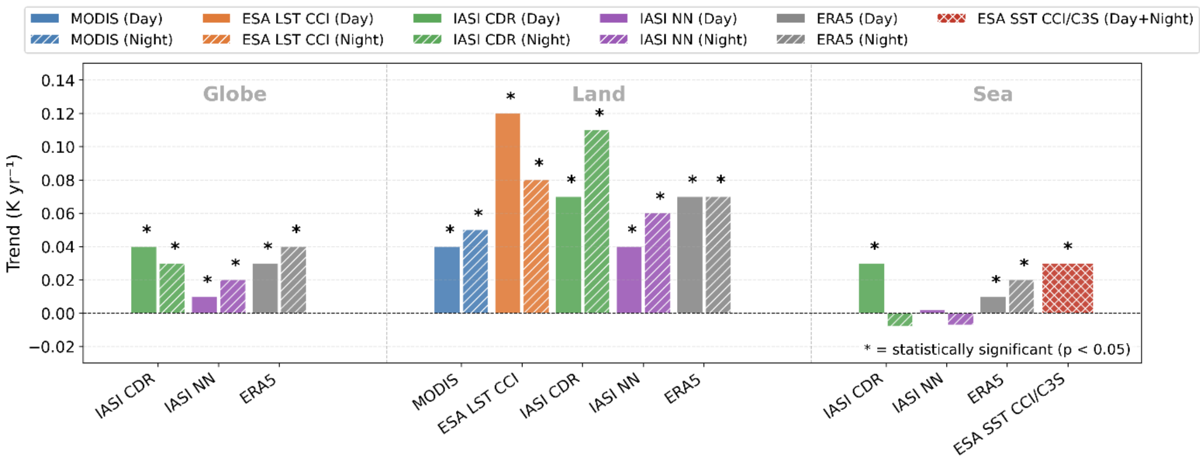


Figure 6. (b) Global (Tskin), sea (SST), land (LST) trends (K/yr) from MODIS, ESA CCI LST, IASI-CDR, IASI-NN, ERA5, and ESA CCI/C3S SST over the period [January 2008–December 2022].

Trends are computed with the Theil–Sen estimator on annual latitude-weighted and deseasonalised averages. Significant trends are marked with an Asterix (*) above the bars.

Technical corrections

Line 66: should read “between 9:30am and 10:30pm”

This line has been rephrased to:

“For consistency between datasets, the satellites orbits are all between 9:30am/pm and 10:30 am/pm. Reanalysis data is chosen at 10 am/pm.”

Line 81: “is” à “has been” :

Corrected

References

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., Hurk, B. van den, Hirschi, M., & Betts, A. K. (2009). A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. *Journal of Hydrometeorology*, *10*(3), 623–643. https://doi.org/10.1175/2008JHM1068.1

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

Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J., Ward, B., & Murray, M. J. (2002). Toward Improved Validation of Satellite Sea Surface Skin Temperature Measurements for Climate Research. *Journal of Climate*, *15*(4), 353–369. https://doi.org/10.1175/1520-0442(2002)015%3C0353:TIVOSS%3E2.0.CO;2

Embury, O., Merchant, C. J., Good, S. A., Rayner, N. A., Høyer, J. L., Atkinson, C., Block, T., Alerskans, E., Pearson, K. J., Worsfold, M., McCarroll, N., & Donlon, C. (2024). Satellite-based time-series of sea-surface temperature since 1980 for climate applications. *Scientific Data*, *11*(1), 326. https://doi.org/10.1038/s41597-024-03147-w

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

Graham, R. M., Hudson, S. R., & Maturilli, M. (2019). Improved Performance of ERA5 in Arctic Gateway Relative to Four Global Atmospheric Reanalyses. *Geophysical Research Letters*, *46*(11), 6138–6147. https://doi.org/10.1029/2019GL082781

Guillevic, P., Göttsche, F., Nickeson, J., Hulley, G., Ghent, D., Yu, Y., Trigo, I., Hook, S., Sobrino, J. A., Remedios, J., Román, M., & Camacho, F. (2018). *Land Surface Temperature Product Validation Best Practice Protocol Version 1.1—January, 2018*. https://doi.org/10.5067/doc/ceoswgcv/lpv/lst.001

Herrmannsdörfer, L., Müller, M., Shupe, M. D., & Rostosky, P. (2023). Surface temperature comparison of the Arctic winter MOSAiC observations, ERA5 reanalysis, and MODIS

satellite retrieval. *Elementa: Science of the Anthropocene*, *11*(1), 00085. https://doi.org/10.1525/elementa.2022.00085
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(n/a), 1999–2049. https://doi.org/10.1002/qj.3803
Hirahara, S., Alonso-Balmaseda, M., de Boisseson, E., & Hersbach, H. (2016). *Sea Surface Temperature and Sea Ice Concentration for ERA5*. ECMWF. https://www.ecmwf.int/en/elibrary/16555-sea-surface-temperature-and-sea-ice-concentration-era5
Høyer, J. L., Le Borgne, P., & Eastwood, S. (2014). A bias correction method for Arctic satellite sea surface temperature observations. *Remote Sensing of Environment, Liege Colloquium Special Issue: Remote Sensing of Ocean Colour, Temperature and Salinity*, *146*, 201–213. https://doi.org/10.1016/j.rse.2013.04.020
Høyer, J. L., & She, J. (2007). Optimal interpolation of sea surface temperature for the North Sea and Baltic Sea. *Journal of Marine Systems, Marine Environmental Monitoring and Prediction*, *65*(1), 176–189. https://doi.org/10.1016/j.jmarsys.2005.03.008
Hulley, G. C., Ghent, D., Göttsche, F. M., Guillevic, P. C., Mildrexler, D. J., & Coll, C. (2019). 3—Land Surface Temperature. In G. C. Hulley & D. Ghent (Eds.), *Taking the Temperature of the Earth* (pp. 57–127). Elsevier. https://doi.org/10.1016/B978-0-12-814458-9.00003-4
Luo, B., & Minnett, P. J. (2020). Evaluation of the ERA5 Sea Surface Skin Temperature with Remotely-Sensed Shipborne Marine-Atmospheric Emitted Radiance Interferometer Data. *Remote Sensing*, *12*(11), 1873. https://doi.org/10.3390/rs12111873
Luo, B., Minnett, P. J., Szczodrak, M., Nalli, N. R., & Morris, V. R. (2020). Accuracy Assessment of MERRA-2 and ERA-Interim Sea Surface Temperature, Air Temperature, and Humidity Profiles over the Atlantic Ocean Using AEROSE Measurements. *Journal of Climate*, *33*(16), 6889–6909. https://doi.org/10.1175/JCLI-D-19-0955.1
Masson-Delmotte, V., P. Zhai, A. Pirani, S.L., Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R., & Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou. (2021). *IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. In Press. https://www.ipcc.ch/assessment-report/ar6/
Minnett, P. J., Alvera-Azcárate, A., Chin, T. M., Corlett, G. K., Gentemann, C. L., Karagali, I., Li, X., Marsouin, A., Marullo, S., Maturi, E., Santoleri, R., Saux Picart, S., Steele, M., & Vazquez-Cuervo, J. (2019). Half a century of satellite remote sensing of sea-surface temperature. *Remote Sensing of Environment*, *233*, 111366. https://doi.org/10.1016/j.rse.2019.111366
Nielsen-Englyst, P., Høyer, J. L., Kolbe, W. M., Dybkjær, G., Lavergne, T., Tonboe, R. T., Skarpalezos, S., & Karagali, I. (2023). A combined sea and sea-ice surface temperature climate dataset of the Arctic, 1982–2021. *Remote Sensing of Environment*, *284*, 113331. https://doi.org/10.1016/j.rse.2022.113331
Prigent, C., Jimenez, C., & Aires, F. (2016). Toward "all weather," long record, and real-time land surface temperature retrievals from microwave satellite observations. *Journal of Geophysical Research: Atmospheres*, *121*(10), 5699–5717. https://doi.org/10.1002/2015JD024402
Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., & Laaksonen, A. (2022). The Arctic has warmed nearly four times faster than the

globe since 1979. *Communications Earth & Environment*, *3*(1), 1–10. https://doi.org/10.1038/s43247-022-00498-3

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, N., Lin, X., Pielke Sr., R. A., Zeng, X., & Nelson, A. M. (2024). Global total precipitable water variations and trends over the period 1958–2021. *Hydrology and Earth System Sciences*, *28*(9), 2123–2137. https://doi.org/10.5194/hess-28-2123-2024

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

Zeng, X., & Beljaars, A. (2005). A prognostic scheme of sea surface skin temperature for modeling and data assimilation. *Geophysical Research Letters*, *32*(14). https://doi.org/10.1029/2005GL023030

Zhou, D. K., Larar, A. M., & Liu, X. (2021). Surface Skin Temperature and Its Trend Observations From IASI on Board MetOp Satellites. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 1665–1675. https://doi.org/10.1109/JSTARS.2020.3046421