Responses to Reviewer 2

egusphere-2024-1228: "Land surface temperature trends derived from Landsat imagery in the Swiss Alps" by Gök, et al. 2024

Thank you kindly for taking the time to carefully read our manuscript and for the constructive comments which helped us to improve it. In the following, we will address all comments point by point and suggest respective corrections.

This study used Landsat-derived LST to analyze long-term LST trends in the Swiss Alps, referenced against ground observations from the IMIS network. Overall, this work is nicely done and offers valuable insights into identifying the origin of potential biases in Landsat-derived LST trends in mountainous terrain. I also find this paper is generally well written and structured. Below are some comments, primarily regarding the clarification of methodologies, which I hope the authors can address:

L55-L60: Please describe the spatial resolution of the thermal bands of AVHRR and Landsat.

We will add information on the spatial resolution and change the text to:

"As the robustness of trends increases with longer time series, LST records from the Advanced Very-High-Resolution Radiometer (AVHRR) **with 1000 m spatial resolution** and the Landsat Program (**60 – 120 m spatial resolution**), are particularly useful for this purpose (Prata, 1994; Gutman and Masek, 2012)."

L103: Change "artefacts" to "artifacts".

Thanks, done.

L104: Clarify what "section 0" refers to, similarly for L116.

Thanks. The cross-reference was incorrect. L104 refers to section 2.3 and L116 to section 2.4. We will change that accordingly.

L134: Ensure temperature units are consistent throughout the text.

Throughout the manuscript, we use degrees Celsius (°C) for absolute temperature values and Kelvin (K) for relative temperatures such as differences and rates. Figure 3 provides a good example for why we think it is useful to do so. Panel (a) shows LST in °C, which most people are used. Panel (b) shows residuals between modeled and observed LST in units K. Here, the unit of °C would erroneously raise the impression of reference to the absolute Celsius temperature scale.

To be clear about our usage, we will add to section 2.1:

"Throughout the study, we use unit degrees Celsius (°C) for absolute temperatures and Kelvin (K) for temperature differences and rates."

L156: Please briefly explain the filters used for masking clouds and duplicates.

Thanks for pointing out that more explanations are needed. As this point and the next are related to each other, we answer both comments here and suggest respective changes for a revised version of the manuscript.

We used two filters to prepare the LST time series for robust trend calculation (1) the removal of clouds and subsequently outliers from LST time series and (2) removal of multiple observations from the same day.

(1) The first order cloud removal is based on the Landsat Pixel QA band. We masked pixels that are cloud-flagged with at least a medium confidence tag in the QA band. Undetected residual cloudy pixels were further removed by applying a threshold on the residuals between modelled and observed LST. We chose a large threshold of +/- 30 K to not interfere with the high natural variability of LST and refer to previous work of Fu et al. (2014). Although most outliers in the LST time series can be attributed to undetected clouds (very low temperatures), we can't exclude very high-temperature outliers, by for example wildfires, that cause anomalous LST observations. Therefore, we applied the threshold for both positive and negative deviations. We will add a remark that the positive threshold allows masking of wildfires, for example.

(2) Landsat scenes are overlapping along- and across orbit track with the degree of overlap depending on the latitude. The across track overlap increases the number of observations from different days and is thus beneficial for LST trend calculation (see supplement figure C1). The along track overlap causes multiple LST observations of the same day which potentially can introduce a bias in the LST trends. To find along track overlap regions we implemented a filter for images in the same path, differing by less than 100 seconds. This approach allows to find each image's next temporal neighbor. From each select image pair we masked the overlap region in the neighbor the image.

We will change the respective section in the text to:

"Prior to fitting Eq. 2 to the Landsat LST data, we implemented filters to mask (1) duplicate LST observations with the same date that result from along-track overlapping Landsat scenes, and (2) cloud-contaminated pixels. **The along-track duplicates were removed by creating image pairs of each Landsat scene and its temporal neighbour in the same path and masking the overlapping region of the adjacent scene. The pairs of subsequent Landsat scenes were identified by a difference in acquisition time of less than 100 seconds which is small enough to only select the directly following scene.** Cloud masking was done using the Landsat C2 Pixel Quality Assessment Band (QA) cloud flag **with at least medium confidence** (Dwyer et al., 2018; Zhu and Woodcock, 2012)."

L165: Explain how the specific threshold is determined. It is unclear if applying an upper threshold of +30 K makes sense when trying to find cold extremes caused by undetected clouds.

See the previous answer.

L228-L229: The values of metrics do not match those shown in Figure 4 and Table 1. Please verify and correct them.

Thanks for pointing that out. We corrected the numbers in the text and checked the entire text to make sure no other issues exist.

Figure 5: You mention a total of 119 stations providing surface temperature observations, but only 115 are included in Figure 5d. Does this mean the remaining four stations had short time series and were excluded from the trend analysis? However, Figure 1 suggests all stations used should have consistent records for at least five years, please clarify. Additionally, the overall trends across stations derived from Landsat and IMIS LST should be given and compared.

Indeed, for the LST trend comparison we used LST data from only 115 IMIS stations that have sufficient long-time series (Figure 5). We added this to the caption of the figure. However, for the direct

comparison of IMIS and Landsat LST, the record length is not important, and we can make use of all available LST data. Therefore Figure 4 contains LST data from all 119 stations. We note that we do provide all LST trends from IMIS stations and the LST trends from the corresponding Landsat observations in supplement file A. A direct comparison between the trend values is possible, but complicated from the fact that they cover different and only partly overlapping time periods, the Landsat trends are affected by the bias we report, and the Landsat data is only from clear-sky days. Disentangling these different factors is what we attempt throughout our study.

L244-L245: While suggesting that LE07 is the most robust, it would be useful to see its distribution when restricting the record length to be comparable to LT05/LC08. This will help understand the impact of temporal overlap on the residuals.

We agree that further analysis on the residuals could reveal more insights on the impact of each sensor on the LST and LST trend measurements. However, here we focus on the impact of record length on LST trends. Differences in accuracy and precision for the individual sensors are given by the LST comparison in Figure 4.

L298-303: Does the ΔLST here represent the trend fitted by IMIS LST at 9:29 minus the trend at 10:16? If so, I assume you are explaining the LST trend bias caused by different acquisition times. Please ensure clarity. Also, explain how the average trend bias is calculated from the ΔLST.

Thank you for highlighting that the previous explanations were insufficient for understanding the estimation of the LST trend bias value. We will adjust the text to:

"We pointed out in section 2.1 that Landsat acquisition times have changed between 1984 and 2022. Approximating this change by a linear model for the acquisition time, yields a time difference of 47 minutes over a period of 38.5 years (from 9:29 in 1984 to 10:16 in 2022; Figure 2). To estimate how much LST difference we would expect to result purely from this 47-minute delay in image acquisition, we exploit the high temporal resolution IMIS data, by calculating for every day and every IMIS station the LST difference between 10:16 and 9.29. The daily LST differences (ΔLST) show a bimodal distribution (Figure 8), which we separated using bimodal Gaussian regression. During melting periods, snow surfaces remain locked at the melting point and ΔLST values are essentially zero (blue curve). The remaining ΔLST values are normally distributed (red curve) with a mean ΔLST of 1.72 K and a standard deviation of 0.93 K. Over a 38.5-year period, this suggests an average LST trend bias of 0.045 K yr^{-1}. However, the IMIS stations are located on flat to gently sloping terrain and the LST trend bias varies with topography"

L318-320: This statement is unclear; please elaborate on its significance.

Thank you for pointing out the unclarity. We reconsidered the statement and decided to remove the sentence in the revised version.

L349-355: Although mentioned in the conclusion, it is helpful to emphasize that the Landsat trends can be less reliable when the data period being examined is relatively short (e.g., less than 10-15 years).

Yes, good point, we will emphasize the effect of a shorter time series on LST trend variability and change the text accordingly:

"Trends with such large temporal overlap are aligned well about the 1:1 line with a mean accuracy of -0.02 Kyr^{-1}, while record lengths < 15 years show significantly more variability. However, long record comparison is dominated by LE07, which has the most overlap in the observation period (**Fehler! Verweisquelle konnte nicht gefunden werden.**)."

L413: It appears the LST trend peaks for an aspect of ~255° (Figure 9c?). Please verify it.

Yes, correct. We were referring to minimum and maximum values of LST trends and ΔSin found at ~75° and ~255°C. However, we think at this point it is sufficient to mention only the maximum value and will change the sentence to:

“The additional radiation flux received during the 47-minute time window peaks for surfaces that are oriented orthogonal to the sun position, at an aspect value of approximately 130°, whereas the LST trend and ΔSin peaks at approximately 255° (Figure 9).”