Review - Towards Best Practices in UAV Thermal Remote Sensing in Complex Environments

Information about review process: interactive review process

Thank you for the helpful feedback and comments and the time spent on the review! We will address the points raised to the best of our abilities in the revised manuscript. Please see below for responses to specific comments (blue, italic text). – *Kathrin Naegeli on behalf of the authors*

RC1: 'Comment on egusphere-2026-787', Anonymous Referee #1, 20 Mar 2026

Review for egusphere-2026-787 by Naegeli et al.

I appreciated the authors' systematic effort to explore and correct biases in UAV-based thermal infrared (TIR) surveys. This is an important and ongoing methodological challenge, and the manuscript addresses a relevant problem.

However, I found the manuscript difficult to follow in its current form. The description of the correction procedures is fragmented across sections (methods vs. results), making it challenging to understand what each experiment is designed to test and how the corrections are derived. In addition, the language throughout the manuscript often lacks precision, which further complicates interpretation.

My main concern relates to the conclusions. Based on the presented analysis, it is not clear that the proposed corrections consistently improve results relative to the raw imagery. In several cases, performance appears comparable—or even degraded—after correction. This raises an important question about the necessity and applicability of the correction workflow, which is not fully addressed. I also expected a more thorough spatial analysis of validation results (e.g., identification of systematic spatial patterns), particularly given that landscape effects are introduced in the introduction but only briefly discussed later.

Despite these concerns, I believe the manuscript has scientific value. However, revisions are needed to improve structure, clarity, and the alignment between results and conclusions.

We thank the referee for their positive comments and suggestions for improvement. We will address the points mentioned above, in particular the clear separation between methods and results. We will also rework the language throughout the manuscript to make it clearer and more precise. We appreciate your overall comment on the effectiveness of the proposed corrections and the conclusions we drew. We will add a statement to the discussion section about the necessity and applicability of the correction workflow, particularly with regard to its performance and importance in representing and including all TIR physics when processing to retrieve absolute LST information. We agree that our spatio-topographic analysis are a bit hidden. However we state our methodology in lines 286-288 and provide the respective figures that are discussed in the supplementary. We will look into expanding the spatial analysis regarding landscape effects by performing a distributed, pixel-based spatial pattern analysis and adding a respective paragraph in the discussion as well as potentially including a figure in the results or the supplementary material.

Specific comments:

Introduction: The introduction is comprehensive, but it does not clearly set up the experimental design. I recommend restructuring it to explicitly introduce the main sources of error (e.g., camera temperature effects, drift, atmospheric/emissivity factors) and link these directly to the three correction approaches tested in the study.

Thank you for your suggestion. We will restructure the introduction to make the main sources of error more clear and better linked to our correction approaches. The revised introduction will be structure as follows (one paragraph each): (i) general introduction about mountain regions, LST as an ECV, remote sensing of LST, (ii) LST research for the cryosphere, (iii) technical impacts i.e. camera temperature effects, drift effects, (iv) environmental impacts i.e. atmospheric and emissivity effects, (v) objectives of the study.

The objectives are currently quite broad. The "challenges and lessons learned" component would be more appropriate as a post-analysis outcome rather than a primary objective. It would help to more clearly distinguish between objectives, experiments, and expected outputs.

Thank you for your comment. We agree that the objectives are not precisely formulated and will therefore rephrase them to align them more closely with the experiments, results and discussion.

L91: Remove "over the Murtèl and Marmugnun rock glacier" in the first sentence, as it is repeated in L95. The paragraph would read more smoothly if it moves from a general description to the specific site.

We will rephrase the first sentences of this paragraph to eliminate duplication in line 91 and 95, as well as move some more general study site information (e.g. meteorological conditions) to the beginning of the paragraph. These adjustments will improve readability of the paragraph.

Figure 1: The ground control point labels in panel (a) are very difficult to read—please increase their size. I also do not find panel (a) particularly informative in its current form. Consider enlarging panel (b) instead, as it better shows surface texture and terrain characteristics.

Thank you for your suggestions. We will rework figure 1 to make in situ observation labels more easy to read. We like to keep the topographic map in the enlarged panel as it shows the complex topography in the greater area surrounding and impacting the survey area, i.e. the mountain face to the south of the survey area (towards Corvatsch and Piz Murtel). We will however show panel b in the enlarged panel (a) overlaying the topographic map to make surface texture and terrain characteristics more easily.

L147: I am not sure what is meant by "GST of the TIR mosaic." Does this refer to the orthomosaic derived from the TIR data and resampled to 20 cm resolution? Please clarify.

This is a typo, we apologise. The sentence should read "GSD of the TIR mosaic", with GSD denoting ground sampling distance.

L150: "PT100 (OMEGA SA1-RTD-4W)" is not clearly described. Please specify that this is a platinum resistance temperature sensor (or equivalent), so the reader understands the measurement type.

We will specify the measurement type of the PT100 by adding "thin-film Resistance Temperature Detector (RTD)" to the sentence.

L170: I do not think this section is necessary. The brief definition of complex terrain in the introduction is sufficient. Instead, this section would be more useful if it provided a short overview of the three correction approaches that will be developed, to help frame the methodology.

We appreciate your comment. The retrieval of accurate LST data from UAV TIR sensors is particularly challenging in complex terrain (in contrast to more homogeneous, flat terrain) and a major focus of the manuscript. We therefore like to keep the detailed definition. However, we agree that a short overview of the three correction approaches after the general introduction of methodologies of the UAV TIR data (e.g. emissivity and atmospheric correction) would be beneficial. We will therefore add a brief paragraph before 3.1 to provide this overview.

L206: I am not clear on what is meant by "camera-target temperature dependency." Since both the blackbody and the camera temperature appear to change together, is this experiment intended to assess errors associated with the operating temperature of the camera–target system? Please clarify exactly what this correction is designed to quantify.

Thank you for your comment. The laboratory analysis investigates the relationship between the target temperature measurement and the camera temperature. Temperature readings from uncooled microbolometers are known to be influenced by the absolute sensor temperature, as well as the Δ-temperature between the sensor and the target. Hence, by using a blackbody facility we were able to make measurements with different *controlled, stable target temperatures (blackbody) while the camera temperature was exposed to artificial wind to have varying camera temperatures alongside these stable blackbody temperatures (and to not overheat). We will add such a clarification of the experiment at the beginning of paragraph 3.1.*

L222: The notation for temperatures is confusing. Please clearly define each variable (e.g., , , blackbody temperature, etc.) and use consistent terminology throughout. It is currently difficult to follow what each term represents and what the goal of the correction is.

We agree that we should introduce our temperature notation more clearly. We will therefore add explanations locally, i.e. T_{raw} *representing the target temperature reading and* T_{camera} *representing the internal camera temperature. Additionally we will add some general naming conventions to the paragraph outlining the different correction just before 3.1 (see also comment above).*

L230: As I understand it, this correction applies when the camera temperature changes while the target temperature remains constant – i.e., during camera warm-up at the start of a flight. Please confirm if this is correct and clarify the conditions under which this correction is relevant.

Thank you for pointing out that this is not clear. As you mentioned, the drift correction aims at correction target temperature readings for unstable camera temperature (i.e. warm-up or cool-down at the beginning of surveys). However, it doesn't imply that the target temperature remains constant but rather uses in-survey knowledge to derive a relationship between the camera temperature once stable and the related target temperature. However, this drift correction clearly only is reliable for homogenous surface (in texture, topography and surface type) that also exhibit a homogeneous surface temperature distribution pattern. Hence, the application is not optimal in our case. Nevertheless, we would like to introduce and apply the correction, and present the results to highlight its inapplicability in complex environments. We will clarify the intended purpose of this drift correction at the beginning of the paragraph and in the discussion.

L242–245: I had difficulty following this section. Please clarify whether the same TCP is used across the "warm," "hot," and "very hot" conditions, the sequence of measurements (e.g., shaded pre-survey, exposed pre-survey, post-survey), and whether the images were taken with the UAV-mounted TIR camera or a separate handheld system (the figure caption suggests a different camera, but this is not clearly stated). If a separate camera is used, please specify this in the data section. Also, why are measurements limited to only three time steps?

Indeed, the same TCP was used for the field calibration. Hence, warm and shaded pre-survey belong together, as well as exposed pre-survey and hot, and post-survey and very hot. All images were taken with the UAV TIR camera, but not while it was flying. Instead, we held the camera in our hands while acquiring the images. We will clarify this by rephrasing the respective paragraph and the caption of figure 3.

L260: Please be consistent in how instruments are referenced. The SI-121-SS is referred to as the Apogee radiometer, but elsewhere as a "thermal infrared radiometer." Use one consistent name. Also, please define NIST.

Thank you. We will use a consistent naming throughout the manuscript for the thermal infrared radiometer and specify the model in the auxiliary data section. We will also introduce NIST as National Institute of Standards and Technology.

L276: Did the radiometer measurements overlap spatially with the GST measurements? It would be valuable to compare these in situ datasets directly, as different measurement approaches (radiometric vs. contact) may yield different results and influence validation.

This is a good suggestion and would be a very valuable comparison. However, unfortunately the radiometers were not placed in the same locations as any of the GST loggers and we can thus not perform such an analysis.

L285: Do you apply each correction individually and then in combination? In other words, are separate mosaics generated for each correction and a final fully corrected mosaic? Please clarify.

This is a very valuable comment. As shown in figure 4 we do not combine the three introduced correction schemes "drift", "lab" and "field", but applied them only singularly. Hence, we have four resulting TIR mosaics, one based on the raw, uncorrected images, and the three resulting form the correction schemes.

L287: Am I correct in understanding that you correlate the spatial index with the error in the corrected TIR data? Please clarify this step.

Yes, we correlate the deviation obtained from the validation per in situ sensor location with the five metrics (elevation, slope, aspect, TRI and irradiance) in order to investigate the possible influence of the latter on validation performance. We state this in lines 286-288 with referred figures being placed in the supplementary.

Figure 4: The figure is difficult to read—please increase text size. Also, the schematic is somewhat unclear: background temperature and emissivity appear to contribute jointly to both emissivity and atmospheric corrections. These processes may need to be represented more clearly as separate pathways. It is also unclear where fits in this framework.

Thank you for your feedback. We will rework the flow chart to make it more reader friendly by increasing the font size, and discriminating the atmospheric and emissivity correction as two separate algorithms applied to all four TIR mosaics. Additionally, we will ensure the different processing levels, i.e. image-based and mosaic-based, are more clearly distinguishable.

L296: As earlier, please clarify whether this refers to the internal camera temperature or the radiometric temperature measured by the camera.

As mentioned in our response to your comment for L206, the camera temperature refers to the internal temperature of the camera. We will rework the entire manuscript to be more consistent in our terminology and introduce these temperature terms in the beginning of the methodology section.

Figure 5: I find this figure difficult to interpret. For example, at 29.9 ℃, is the goal to keep the camera at a constant temperature, or is it varying? The correlation appears to be between camera temperature and blackbody-derived pixel temperature, but the interpretation is not clear. The difference between the blackbody temperature and pixel temperature is attributed to camera effects, but in panel (d), the camera temperature closely matches the blackbody temperature while the pixel temperatures still show substantial variability. This seems inconsistent with the interpretation in L292. Additionally, L301 refers to a "range" in pixel temperature, but the figure shows standard deviation. Across panels (a–e), the range of pixel values appears similar despite large differences in camera temperature variability. Overall, I am missing key information about whether camera temperature is controlled or varying in each experiment, and how this links to the conclusions drawn.

Thank you for your feedback. We agree the figure is not easy to understand. Also our description should be more precise. We will clarify the intended purpose of these experiments more clearly in the respective methodology section. The correlation is not between camera and blackbody temperature, they are both controlled individually and in separate manners. We will clarify this more clearly in the respective text parts.
For the case of panel (d) specifically, we like to mention that the absolute Δ between the camera temperature and the pixel temperature is very large (almost 30 ℃). This highly impacts the stability and accuracy of the pixel temperatures. We will stress this Δ impact more clearly and adjust our statement in L929 accordingly.
To illustrate the absolute temperature values, and in particular the Δ we will add a respective figure to the supplementary material.
You are right, our wording in L301 is not precise, i.e. range should not be linked to the standard deviation. We will rephrase accordingly.

Figure 6: This figure is clearer. It effectively shows that when the camera is warmer, pixel temperatures over a stable blackbody are more variable, whereas lower camera temperatures correspond to more stable measurements.

Yes, this figure is easier to read and your conclusion is correct. It highlights the second important technical impact, namely the camera temperature and its performance related to its own temperature (beside the impact of the Δ-temperature between camera and target as mentioned above). Hence, the camera operates less accurately towards the boundaries of the stated operating temperature range, which in our case of the FLIR Tau 2 is -20 ℃ to +65 ℃. Thus, in the most ideal case a TIR camera on a drone would be temperature controlled and stabilised at around 20 ℃. We will add a respective statement in the discussion.

L304: I assume the internal camera temperature changes (e.g., warms up) during flight. It would help to state this explicitly to guide interpretation of the trends shown.

No, in our surveys the camera temperature in fact cooled down during the flight until it stabilised after about 10-12 minutes (from about 15 to below 10 ℃ in 2022, and above 25 to about 15 ℃ in 2023). See green and purple data points of T_{camera} in panel and b in figure 7. This is due to cooler air temperature at the survey height and the flight wind affecting the camera temperature.

L306: Please clarify what is meant by "highest camera instability." Is this referring to the largest variance, drift, or another metric?

Thank you for your comment. With "the highest camera instability" we refer to the time period of the survey when the camera temperature shows least stability, i.e. in the first 10-12 minutes of each survey. We will clarify this by rephrasing the respective sentence.

Figure 7: I am not fully clear on what is being shown. Are these the raw data used to derive the drift correction described in the methods? If so, this should be stated explicitly. More broadly, the manuscript structure is confusing: corrections are described in the methods, but the data used to derive them appear in the results.

You are absolutely right, the separation between methodology and results it not entirely clean. We will work towards including all relevant information and data needed to decsribe the individual corrections in the methodology section and keep the results as results-based as possible. Hence, we will move figure 7 (and likely 5) to the methodology section.

Figure 8: "Raw" is not a correction scheme—please revise the caption accordingly. It would help to clearly distinguish between raw data, individual corrections, and the combined correction.

Absolutely right, we will modify the caption accordingly. However, there are no combined corrections, hence we do have the raw mosaic, i.e. based on raw images obtained by the UAV TIR camera, the drift mosaic, i.e. based on the drift corrected images, the lab mosaic, i.e. based on the lab corrected images, and the field mosaic, i.e. based on the images field corrected iamges.

L320: Replace "overflown" with "flown over."

Thanks, we will change it.

L330: The comparison of validation approaches is valuable. However, the distinction between contact-based measurements (e.g., GST) and radiometric measurements should be introduced earlier in the manuscript, as it is central to interpreting these results.

We wil add a respective statement in the section "Auxiliary data" to point out that GST sensors are measuring contact skin temperature, while the TIR radiometers measure radiometric skin temperature.

L350: I find this statement somewhat misleading. While one case shows a difference of 0.03 °C versus 0.04 °C, Figure 10 suggests that, overall, the corrections do not consistently improve performance relative to raw imagery. This should be discussed more cautiously.

Thank you for this comment. We agree that the performance at all validation points is not significant, and this should be made clearer. However, we would prefer to include a more extensive discussion of this issue at the end of the first paragraph of the 'Discussion' section rather than here in the 'Results' section. Additionally, we would like to point out that, regardless of the limited impact in our use case, retrieving absolute LST requires consideration of all the involved physics. Applying these corrections is a step towards taking these physics into account in the best possible way.

Figure 10: Is there a reason results are not grouped by sensor type (e.g., radiometer vs. GST)? Grouping them may help reveal patterns. The variability across validation approaches is striking. It also appears that corrections do not consistently reduce error. For example, in 2023, raw and field comparisons are very similar, and drift correction may even worsen agreement. At the same time, Figure 9 suggests that drift correction improves spatial patterns (e.g., removing the cold corner), indicating that different metrics (bias vs. spatial structure) may lead to different conclusions. This is worth discussing explicitly.

Thank you for your feedback. We listed the in situ validation sensors on the y-axis according to their elevation, i.e. the lowest is listed at the top and the highest one at the bottom, to be most aligned with figure 1. However, your argument of grouping by sensor type is absolutely valuable and will change the figure accordingly.
Additionally, will provide average validation performance per correction approach and sensor type, as well as elaborate on these findings towards the end of the first paragraph of the discussion (see also our answer to your comment just above).

L365: Based on Figure 10, I do not think this statement is supported. Raw imagery appears to perform similarly to corrected outputs in many cases, which weakens the argument for applying all corrections.

This is a very valid comment which we appreciate. However, we like to stress the importance of accounting for all physics that impact TIR measurements in the best possible way, regardless of the overall impact on absolute values. Moreover, the correct consideration of the physics is of greater value and importance than our validation, which clearly has limitations. We will rephrase and elaborate more on the validation results and their limitations in the revised manuscript. Please see also our answers to your two comments just above.

L366: This section feels like a mix of recommendations and reflections on what did not work. Consider restructuring it under a clearer heading (e.g., "Recommendations") and separating it from the uncertainty discussion.

We want to highlight that the three paragraphs in this section give a set of recommendations to the community to be conducted prior to a UAV TIR survey based on our experience: i) proper characterisation of camera-target temperature dependency. Knowledge about camera stability and behaviour in a controlled setting is key to trust data from field surveys; ii) knowledge of environmental variables and understanding of their possible impact on observed temperature observations; and iii) the flight planning that can already substantially minimise impact effects such as directionality. To account for these recommendations and the discussion therefore more clearly in the title, we will adjust the headings of 5.1 (prior), 5.2 (during) and 5.3 (post) accordingly.

L372: I do not understand this sentence. Please clarify what is meant by "inevitable" and how this relates to assessing bias. If correlations between metrics were an objective, they should be clearly presented rather than only mentioned in discussion.

Thank you for your comment. Generally we like to stress that if thermal data in complex environments shall be acquired, knowledge about the topographic characteristics is highly important (i.e. inevitable). It does not directly relate to assessing biases but is crucial for deciding what kind of corrections should be applied and how the validation should be set up as stated in that paragraph. We agree that so far this was only discussed *but not supported by quantitative findings except for figure A5 in the supplementary material. We will look into including a more detailed analysis on spatial patterns and the influence of the topographic characteristics within the scope of this study in order to support our existing qualitative statements.*

L385: This is the first time it is mentioned that TCPs were intended for direct calibration of the TIR orthomosaic but were not successful. This should be introduced earlier, especially given their role in pre- and post-flight correction.

Indeed, we had one TCP in the survey area which, as it turned out, was far too small (even though it was about 3.75 times the GSD). Main reason for it to be unusable for calibration purpose was the thermal contrast between the TCP and its surroundings, which was far too large. In fact we mention the TCPs in the auxiliary data section. We will add information highlighting the use of the one in the survey area as well as the one used for pre- and post-flight thermal calibration for clarification.

L432: Micro-slope effects and emissivity changes (particularly at angles $>30°$) are important. It remains unclear whether averaging over these effects produces behaviour equivalent to a flat surface, or whether biases remain.

Thanks for raising the point about directionality and the impact of micro-topography on surface temperature and emissivity. We agree that large angle introduce uncertainties and state this in our discussion. We will expand on this point by including a statement about the impact of micro-topography/sloped terrain and resulting possible directionality effects in out uncertainty and limitations section. Additionally, as mentioned in other answers, we will expand our analysis on the impact of the spatio-topographic analysis, where slope is one metric. In regard to emissivity changes, we can clearly state that the predominant lithology in the study area are granodiorite. A more detailed characterisation based on mineral-specific absorption features with the broadband TIR and the RGB cameras is unfortunately not possible. Including the impact of surface roughness in the emissivity characterisation would indeed be an interesting topic, which however is beyond the scope of this study. We nevertheless like to include these aspects and will thus add some discussion and respective literature in the limitations and outlook parts.

L441: TRI is discussed but not shown in the manuscript. If it is important to the analysis, it should be included rather than only referenced.

In fact, the TRI is plotted in figure 1g. We add the reference to the figure to this sentence. Please note also our other answers on expanding the spatio-topographic analysis where TRI is a metric.

Figure 11: This appears to be a standard field workflow and does not seem specific or novel to this study. Its added value to the manuscript is limited.

Thank you for your feedback. We agree that it outlines the standard field workflow, which, to our knowledge, does not exist for UAV TIR surveys to retrieve accurate absolute LST, particularly in complex environments. Additionally, we would like to mention that we have received very positive feedback from other TIR UAV users who have expressed their high appreciation of the set of recommendations and the schematic overview figure (figure 11), as opposed to equation-heavy publications with scattered information on the workflow. We would therefore like to keep it.

Citation: https://doi.org/10.5194/egusphere-2026-787-RC1