Responses to Reviewer 2

TC-2022-113: "*High-resolution debris cover mapping using UAV-derived thermal imagery: limits and opportunities*" by Gök, et al.

We would like to thank reviewer #2 for the comments and suggestions that helped us to prepare an improved revised version of our manuscript. Please find our responses in blue below your comments.

General comments:

Gök et al. present a comprehensive study investigating the derivation of debris thickness from thermal imagery at different times throughout a day using two approaches: solving of a SEBM using reanalysis data, and a least squares regression method that utilises in-situ debris thickness measurements. This study advances knowledge by investigating the diurnal changes in debris thickness estimations and addresses the impact of thermal drift due to the use of uncooled microbolometers by applying for a manually calibrated drift compensation. My line specific minor comments to improve clarity in the manuscript are below and any major issues to be addressed are in bold.

Specific comments:

ABSTRACT

L10: Specify what LST you use (i.e. LST over debris layer)

> Good point, changed to: "In this study, we present land surface temperatures (LST) of supraglacial debris cover and its diurnal variability …"

L15: Can you specify which method is the most appropriate.

> Added sentence: "Although the rational curve approach requires in-situ field measurements, the approach is less sensitive to uncertainties in LST measurements compared to the SEBM approach."

L18: Please state recommendation for which time of day best represents the most accurate debris thickness estimations.

> We added: "We find the rational curve approach with LST measured in the evening hours to yield the best result."

INTRODUCTION

L22: I would disagree with the statement that debris thickness is generally rather thin, debris thickness is highly variable! – please rephrase and/or add a supporting reference.

We added a reference to Rounce et al., 2021

L26: Change heavily to extensively or similar word.

Changed to: "Consequently, glaciers with widespread and thick debris cover can persist longer at lower elevations than debris-free glaciers …"

L37: Provide references for these processes – e.g. Kirkbride and Deline, 2013; Hartmeyer et al. 2022a,b.

Done.

L38: Provide reference for 'debris thickness varies with time' and 'recent studies document changes…'

We changed the first sentence to "As all these processes vary with time, supraglacial debris cover ought to change in time, too. " We did not add references to the sentence "Indeed, recent studies document changes in debris cover thickness in various mountain ranges on Earth.", because these are provided in the following sentence: "Most studies, however, focus on changes in the extent of debris cover (Shukla et al., 2009; Bhambri et al., 2011; Glasser et al., 2016; Tielidze et al., 2020, Kaushik et al., 2022), whereas studies documenting changes in thickness are relatively rare (Stewart et al., 2021; Gibson et al., 2017)."

L43: Reword sentence for clarity.

Very true, we changed the sentence to: "In particular, the abundance of supraglacial streams, ponds and ice cliffs can increase or decrease rapidly across the glacier surface (Anderson et al., 2021)."

L52: Add Gibson et al. 2017 as reference for empirical estimation of debris thickness using in situ debris thickness measurements.

Done.

L61: Westoby et al. 2020 use UAV optical imagery to estimate debris thickness and changes, so perhaps rephrase to emphasize that debris thickness estimation using UAV thermal imagery has remained elusive.

Good point. Changed to: "Debris thickness was recently mapped using oblique LST (Herreid, 2021), but the quantification of debris thickness from UAV thermal imagery has remained elusive."

STUDY AREA

L70: Units for coordinates.

Changed to: "46.01 °N, 7.46 °E"

L76: Detail the specific dimensions of the study area here and replace 'numerous' with the specific number of flights.

> Changed to: A relatively small study area of 60000 m² allowed for 8 UAV flights covering the entire study area throughout the day.

MATERIALS AND METHODS

L90: Change pre-define to pre-defined.

> Done.

L92: Can you state the resolution of your imagery here too? I.e. … has a resolution of 640x512 pixels, equating to a thermal image resolution of XXm x XXm over the study site.

> Changed to: "… has a resolution of 640x512 pixels, equating roughly to a thermal image resolution of 0.17 m x 0.16 m and measures longwave …"
>
> The UAV flight path was terrain adjusted using a SRTM DEM to maintain the same distance to the ground. However, as the SRTM DEM has a coarser resolution, the elevation cannot be maintained exactly. This results in variation of the raw thermal image resolution. During the photogrammetry processing, differences in spatial resolution are canceled out.

L103: Add a statement that reflects the inherent bias of making in-situ manual hole measurements (i.e. the presence of a large boulder/ thick debris makes it less likely that it will be chosen it as a spot to dig).

> Changed to: "The debris cover close to lateral moraines consists of very large boulders (>0.5 m) that rendered measurements impractical and thus introduced a bias on the point measurements."

L108: Can you provide a bit more detail about what unbalanced thermal conditions are? Spell it out for the reader.

> Changed to: "Unbalanced thermal conditions between the inner parts of the sensor and the ambient temperature (e.g. the sensor cools down after UAV take-off, changing wind conditions, or heats up by direct incident shortwave radiation) introduce a temperature bias."

L119: Why is there a reduced framerate?

> We reduced the frame rate to facilitate data management as the default setting records thermal images in a framerate of 9Hz. We reduced the framerate to 1 image per second, which we considered to be enough. We were not aware of the consequences during the post processing using the Teax Thermal Capture 2.0 and the accompanying software ThermoViewer 3.0.7. The tau2 in-flight calibration is documented in the image metadata after each calibration event. By reducing the

framerate, many of these frames get lost. We informed the manufacturer that this is relevant information that should be included in the manual.

L130: Define FPA acronym used in figure in caption.

Added to caption: “The temperatures of the focal plane array (FPA) and housing case returned from the sensor are shown in orange and bold black dashed lines.”

L137: Change ‘asserting’ to ‘assuming’ – this is an assumption that you have made for the final calibration, the ice surface may be 0°C but ice surface temperatures can be colder than 0°C and this should be acknowledged.

Done.

It is also apparent that after your correction, some ice surfaces are now pushing 10°C (frame number ~250 (c)) where the spline interpolation was not super effective, this should be explained and acknowledged too.

Good point. We added: “However, for large ice temperature variation, the spline interpolation may not capture the temperature offset as shown in Fig. 2c (frame ~250).”

L145: State number of thermal and optical images used in this sentence.

We understand that this section needs more information and changed the text accordingly, including the number of images used.

L155: State size of test sight so comparison with the footprint of the reanalysis data is possible.

We added “(~150 m × 350 m)”.

L166: Do you mean ‘by accounting for the water vapor content…’?

Thanks for pointing out, that this is unclear. We do not account for water vapor in the atmosphere between the surface the and sensor when calculating the LST. We assume water vapor to have a minor influence on the LST measurement and would be compensated by our bulk correction approach. To be clearer we changed the sentence to: “We think, radiation attenuated by water vapor in the atmosphere between sensor and ground would be spatially uniform and thus compensated by our calibration procedure. “

L168: Why/how are they the best classification results?

Good point! By comparison with the optical imagery and according to the algorithms’ Out-of-bag (OOB) error, ice surface faces were best discriminated from surrounding debris when the temperature difference was largest, at 15 h. To be clearer on this point we changed the sentence to: “By comparison with the optical imagery and according to the algorithms mean prediction error, we found best classification results when the temperature differences between ice and debris were the largest, at 15 h. Data from this flight were used to classify the thermal imagery.”

L169: Add statement to end of sentence along the lines of 'thus data from this time stamp were used to classify the thermal imagery'.

Done.

L172: I'm not sure I understand how ΔS is a rate of change if the right-hand side of Eq. 2 is fluxes in W m-2?

ΔS is the rate of change of heat stored in a layer of debris (see e.g. Brock et al, 2010). To balance surface energy fluxes at sub-daily time intervals, the debris layer must be treated as a volume with variable heat storage. It requires knowledge of the average rate of mean debris temperature change, which we obtain by fitting a sine function to the diurnal LST variation in each pixel of the map and forming the first derivative for each time of flight (described in L220-L225 and examples shown in Fig. 5).

L198: Can you comment on the accuracy of wind speed data taken from reanalysis products and the impact this will have on your subsequent debris thickness estimates (see Schauwecker et al 2015; Stewart et al 2021)?

Good point. There is no doubt that the accuracy of wind speed from reanalysis data is low and not well suited for a small valley glacier with a complex wind regime. We address that point in section 5.1 with our sensitivity test. Figure 12 shows the impact of wind speed (and other parameters) on the debris thickness estimates at the different times of the day by the coverage panel. In short, the wind speed has a strong impact on the sensible heat flux and therefore on b (Eq. 8). In the first half of the day this leads to unphysical solutions of Eq. 7. We discuss that point in detail in section 5.1. Interestingly, in the afternoon hours, wind speed has no big influence on the SEBM-derived results.

L203: State what your definition of 'thin' debris thickness is.

Good point. We added: "(<10 cm)" to the text, based on results shown in Conway and Rasmussen (2000), Nicholson and Benn, 2006; Rounce and McKinney, 2013, which we cite in the text.

L229: Can you please justify why the debris thickness was estimated by solving a quadratic rather than previously documented methods such as that in Rounce and McKinney, 2014? The exclusion of sites for which there is not a real or positive solution to the quadratic equation means that a large proportion of your data is excluded from further analysis – which poses quite a large problem when your study area is already quite small.

In our study, we extended the SEBM by the component ΔS compared to Rounce and McKinney, 2014. Using remotely sensed LST data this has not been done before as most remote sensing products do not provide LST data on sub-daily time intervals. ΔS (the change in heat storage) itself is a function of debris thickness. Consequently, the SEBM now contains G(d) and ΔS(d) which then results in a quadratic equation when solving for debris thickness.

L230: Change testing to training.

Done.

RESULTS

L236: I think it is the estimation of debris thickness that changes rather than the relationship between LST and debris thickness, no? i.e. when hotter LSTs are observed, thicker debris will be estimated?

True. We changed the sentence to: "The LST changes over the day in a cyclic manner (early morning cool – afternoon hot – evening cool) and consequently the ability to estimate debris thickness using LST changes accordingly (Fig. 3)."

L240: A key problem for me with this figure is the lack of consistency between the areas of ice and debris in each subplot – theoretically, these areas should remain consistent throughout the day, the time scale of this study is not large enough to observe actual change in the cliff geometries. This then throws into question the accuracy of the data in each time stamp if cliffs are not consistently detected. Can you provide an explanation / justification for this?

Yes, we agree. Based on a similar comment of reviewer 1, we paste here our response:

Thanks for pointing out the inconsistency in ice cliffs – that's an important point! The inconsistency in the bare ice geometry is due to the residual uncertainties of the ice surface temperature. Figure 2 shows that the ice surfaces can vary by several degrees (~6 K). This means that our threshold of 0.5 K in figure 3 does not capture all ice surfaces. We added to the caption of figure 3: "Due to the residual uncertainties of the LST, ice surface geometries appear inconsistent in time."

L243: Can you show a linear regression line and an R2 value on each subplot to support this statement?

Good idea. Done.

L270: Are these LST ranges using raw LST or offset corrected LST?

These are corrected LST.

L295: In caption (or next to the color bar in the figure) state what aspect degrees refer to (i.e. N S E W). Also, is the debris thickness manually measured debris thickness? Make this clear.

Changed to: "… colorized by debris thickness measured in the field." AND "… with lines colorized by terrain aspect with 0/360 ° facing north."

L310: To support this statement, include a histogram of manually measured debris thicknesses for comparison with the modelled debris thicknesses in Fig. 7.

Great idea. We added a histogram to figure 1.

L312: I am concerned about how much data is not valid in Fig7a-d, and I'm not convinced that 'no valid solution' is a sufficient explanation for the lack of data. A surface energy balance model should not be unsolvable. To compare data from different time stamps, the data needs to be (and should be) spatially consistent.

Thanks for pointing that out. Reviewer 1 had similar concerns; therefore, we copy our answer here:

Yes, we agree with the reviewer that these unphysical results are relevant, especially in the morning flights. We also agree that these uncertainties may raise concern for areas with existing solutions. We devoted much of chapter 5.1 to discussing and explaining the cause of this issue. In short, such issues arise specifically when the debris is heating up from morning to mid-day when uncertainties in LST or reanalysis-derived quantities can cause no solution to the equation. In addition, we address the impact of several parameters on the model and the unphysical results (sensitivity test Figure 12).

We expanded the discussion section in 5.1 on these issues in the revised version and changed the sentence in the caption text of Figure 7 to: "White regions show regions where the surface energy balance model has no valid solution for debris thickness due to high uncertainties in the land surface temperature and reanalysis data".

L318: Quantify 'pattern of thin debris predictions'.

Changed to: "Predictions of thicker (>10 cm) debris are primarily found in the afternoon and evening hours (17 h to 22 h) and the pattern of thin debris (<10 cm) predictions, primarily in the central part of the glacier, is relatively consistent in time."

L343: Sentence beginning 'Predicted debris thicknesses…' does not make sense.

Changed to: "The modelled debris thicknesses range varies between 0 cm and 30 cm, but with early flights at 9 h and 11 h lacking predictions greater than 10 cm (Fig. 10a, b)."

DISCUSSION

L374: underestimates compared to what?

Changed to: "The SEBM approach mostly underestimates debris thickness at all flight times compared to field measurements."

L406: Figure units! X axis and also mean debris thickness in top right.

Done.

L417-421**:** My takeaway from this is that 1) wind speed is not modelled well with ERA-5 data, 2) if wind speed is increased to 'realistic' values then the amount of 'valid' debris thickness pixels decrease significantly? Can you discuss what implications this has in terms of the methodology? I.e. would you recommend that SEBM are not used in conjunction with thermal data to estimate debris thickness?

> Correct, we added a sentence to the discussion: "The high sensitivity of the SEBM approach to uncertainties in LST and the reanalysis data, reduce the suitability to reliably estimate debris thickness."

L434: Can you quantify 'in parts more accurate'?

> To be more precise we changed the sentence to: "Debris thickness predictions using least squares regression of a rational curve yield RMSE values between 6 and 8 cm, similar to the results of the SEBM approach"

L451-452: Westoby et al 2020 do this with optical imagery and a geodetic based debris thickness estimation.

> Good point. We changed the text to: "High resolution studies can, improve our understanding of processes that move and distribute debris on glacier surfaces (Westoby et al., 2020). Spatiotemporal debris thickness estimates using UAV derived thermal images has the potential to serve as a new approach to quantify how debris is mobilised across the surface of the glacier over short times scales."

L455**:** It would be good to see a debris thickness difference map (i.e. rational curve – sebm debris thickness) to quantify the differences between the two methods as panels c and d visually look very different. Are the differences between the two methods significant on a pixel-by-pixel basis?

> Great idea. We will add a panel showing the difference between the two approaches and a second panel showing a 2d histogram between both model results in the revised version.

L512-520: I would like to see a recommendation of 1) which method is the better predictor of debris thickness, and 2) at which time of day the method appears to be the most accurate. This seems to be missing from the discussion and the paper in general.

> True, we add: "However, as this approach is less sensitive to uncertainties in LST, we recommend the rational curve approach to estimate debris thickness as long as enough debris thickness measurements are available."

Technical corrections:

Ensure LST and not LST's throughout the paper.

> Done.

Rounce and McKinney, 2013 – should this be Rounce and McKinney, 2014, see reference below?

> Thanks, done.

L150: Check section numbers here and throughout (i.e. where they're referenced, such as in L187).

Thanks, done.

L362: Figureb?

Done.

L364: that thin debris should be than thin debris.

Done.