Dear Dr. Lindsey Nicholson,

Thank you very much for your insightful and constructive comments. We hope that the following response will address all the comments and we will make appropriate changes in the revised paper to strengthen its content.

**Points to be addressed:**

1. SfM DEMs: more specific detail on the error assessment of the derived DEMs is required as detailed below in specific comments.

- Please see the response to the specific comments addressed below.

2. Newly proposed method for determining microtopographic $z_0$: While I support what was trying to be achieved through this effort, I am not convinced of the value of introducing this new method without validation of it against an aerodynamic roughness length derived from meteorological instrumentation. The relationships of Lettau and others are usually validated against wind profile determinations of $z_0$ carried out in the field or in a wind tunnel. As stated above, I think the paper would be better and more focused without this section.

- The lack of aerodynamic measurements is a limitation of this as well as other similar studies and is a very important area of future work. Previous work, e.g., Rees and Arnold (2006), has relied upon surface roughness estimates from other studies to assess the reasonableness of their results when aerodynamic data were not collected. In this study, we originally intended to use the Lettau-Munro method, but found that estimations using this method greatly underestimated the values of surface roughness of debris-covered glaciers in the field (Section 4.2, p3516, L18-27). As aerodynamic data were not collected in this study, we rely upon the results of Inoue and Yoshida (1980), which estimated surface roughness using wind speed profiles at two sites on the Khumbu glacier. Based on the description of these two sites, these sites seemed to be similar to the debris cover on Imja-Lhotse Shar. Specifically, Sites B and C in our study have similar debris cover to Areas III and IV from Inoue and Yoshida (1980), respectively. This provides a means of assessing the reasonableness of the methods developed in this paper.

  Additionally, the various sites assessed in this study provide a range of debris cover that is typical of debris-covered glaciers in this region. Therefore, it is important that the method developed in this paper captures this inter-site variability (Section 4.2, p3517, L1-24). We also believe that the robustness of the method developed in this paper was greatly improved in response to the discussion comment from Evan Miles. The use of an obstacle density of 30% provides an objective approach for the selection of the obstacle threshold height that is supported scientifically (Smith, 2014). This 30% obstacle density threshold was also shown to hold for various resolutions of the DEM further supporting its use.

Therefore, while we acknowledge its limitations, we believe that the method developed in this study presents a novel approach for estimating the surface roughness that yields reasonable results and is a significant step forward in the current state of knowledge of surface roughness techniques.

Changes to the manuscript: A discussion regarding the lack of aerodynamic roughness measurements has been added to Section 4.1:

"One of the main limitations of this study is the lack of aerodynamic roughness measurements to validate the developed methods. Previous work, e.g., Rees and Arnold (2006), has relied upon surface roughness estimates from other studies to assess the reasonableness of their results when aerodynamic data were not collected. This study relies upon the results of Inoue and Yoshida (1980), which estimated surface roughness using wind speed profiles at two sites on the Khumbu glacier. Specifically, Sites B and C in this study have similar debris cover to Areas III and IV from Inoue and Yoshida (1980), respectively. This provides a means of assessing the reasonableness of the methods developed in this paper. Additionally, the four sites selected in this study provide a range of debris cover that is typical of debris-covered glaciers in this region. Therefore, it is important that the method developed in this paper captures this inter-site variability.

Site B had the highest value of $z_0$ (0.043 m), which was expected since the debris cover includes larger boulders up to 1 m in size (Figure 2). Furthermore, this value of 0.043 m is similar to the higher value of 0.060 m for $z_0$ derived from a region on the Khumbu glacier that consisted of large granitic boulders of 1-2 m in size lying on top of schistose rocks with a grain size varying from a few centimeters to 0.5 m (Inoue and Yoshida, 1980). Site C, which comprised the smallest grain sizes of the four sites in this study, agrees well with the smaller value of $z_0$ (0.0035 m) derived by Inoue and Yoshida (1980) for an area where the supraglacial debris was deposited as dispersed boulders ranging in size of 0.01 – 0.05 m. The few boulders ranging in size of up to 0.15 m may be the reason for Site C's slightly larger value of $z_0$ (0.006 m). Sites A and D were composed of boulders and grains that varied in size between those found in Sites B and C; therefore, we deem the value of $z_0$ of 0.016 and 0.014 m for Sites A and D, respectively, to be reasonable. These values also agree fairly well with the $z_0$ of 0.016 m measured by Brock et al. (2010) on a debris-covered glacier in Italy that comprised a mixture of granites and schists of predominantly cobble size, with occasional boulders of < 1 m size."

3. Model calibration: It would be advantageous to additionally perform a multisite optimization to obtain single optimized values for albedo, k, and roughness, rather than a value for each stake. These values could then be applied to all ablation stake sites to give an idea of how useful the model will be when applied to sites for which no specific optimization is available.

- As the reviewer notes, the use of a single optimized parameter set for all sites would be beneficial in estimating the ablation at other sites. However, in order to equally weight all the sites, this set of parameters was derived as the mean value from the model optimization at the 10 sites. The resulting values were a $z_0$ of 0.015 m, an albedo of 0.32, and a thermal conductivity of 1.52 W $m^{-1}$ $K^{-1}$. These parameters will be applied to the other sites for which an individual optimization was not performed to estimate ablation rates.

  Changes to the manuscript: The following has been added to Section 5.3 to explain the use of the average calibrated parameters: "For sites that only had an ablation stake, the average calibrated parameters for that particular latent heat flux model were used. Additionally, a comparison was performed for the $LE_{Rain}$ model using both the average calibrated parameters for all the sites and the calibrated parameters for each individual site (Figure 7)."

4. Model results: (a) Including scatter plots as well as the line plots in Figure 4 could be more helpful for visualizing the prevalence and nature of model biases. As you observe a positive model bias for the nightly minimum temperature I would suggest color coding the scatter plot according to the hour of the day to show the timing of any biases. This approach might also be useful for discriminating more detail about the relative performance of differing methods of modeling LE, by highlighting scatter points for which LE was being modelled in a different color. (b) I became a bit confused as to exactly what data were being used for model validation, so this needs [to be] cleared up and state more explicitly regarding the use of $R^2$ values between modeled and measured total ablation. I'd also like a small table or explicit listing of values compared the 3 model time resolutions to available stake data.

- Figure R1-1 shows the positive bias of the nightly minimum, i.e., the nightly low was typically overestimated by the model. One possible explanation for this is the use of NCEP/NCAR data for the incoming longwave radiation, which dominates the energy flux at night.

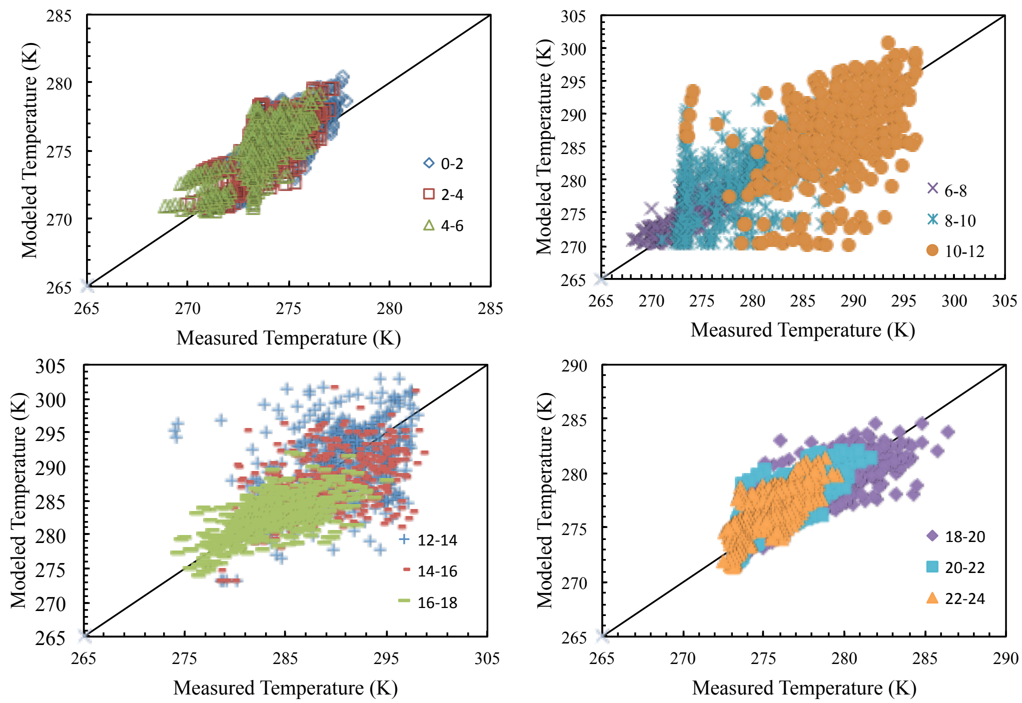


Figure R1-1. Scatterplots of measured and modeled temperature for Site 11 at the surface for the $LE_{Rain}$ model.

The surface temperature sensors for the months of June and July were used to calibrate the model, while the surface temperature sensors for the months of August, September, and October were used to validate the model. Ideally, two separate years of data would be available such that one year could be used for calibration and the latter for validation, but these were not available.

As requested by the reviewer, Table R1-1 shows a comparison of the total ablation as a function of the different temporal resolutions for the three sites where ablation measurements were made. As expected and detailed in the paper, the 6 hrly time step underestimates the modeled ablation with the 30 min time step. The differences between measured and modeled ablation rates are likely due to site-specific properties and not the temporal resolution.

Table R1-1. Modeled and measured ablation rates (m) using the $LE_{Rain}$ model for the three sites where the ablation stakes did not completely melt out of the ice.

| | Site 8 | Site 13 | Site 15 |
|---|---|---|---|
| Measured | 0.92 | 0.85 | 0.89 |
| 30 min | 1.76 | 0.76 | 1.22 |
| 6 hrly | 1.47 | 0.69 | 0.98 |
| Daily | 1.76 | 0.74 | 0.97 |

Changes to the manuscript: The following figure has been added to the paper to show the nightly bias:

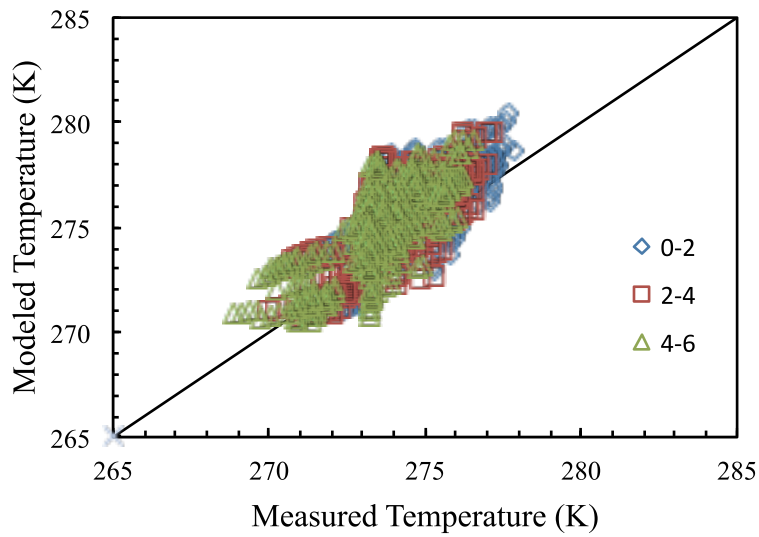


Figure 6. Scatterplot of measured and modeled temperature for Site 11 at the surface for the $LE_{Rain}$ model showing the positive temperature bias overnight.

This has also been accompanied by the following sentence in Section 5.1: "The positive bias of the nightly minimum is apparent between the hours of 0:00 and 6:00 (Figure 6)."

**Specific comments:**

P3507/L23: Replace 'laying' with 'overlying'

- The manuscript has been changed accordingly.

P3507/L26: What was the reason for the data loss? Failed loggers? Inaccessible? Sensors becoming exposed? Might be useful information for others.

- Large changes in the topography, i.e., some sites on a collapsed slope burying the sensors such that they could not be found, and a couple of the loggers on the sensors failed. For future work, we recommend tying the sensors to a string such that all the sensors remain together. A sentence has been added to the manuscript regarding the reason for data loss as follows: "… as sensors were lost due to large changes in the topography and some of the loggers failed."

P 3508/L8: Why was site 4 treated differently?

- Site 4 had a debris thickness greater than 1 m, which was the limit of our manual excavation (p3508, line 11). The sentence has been revised to read as follows: "…

with the exception of Site 4, where the debris thickness was greater than 1.0 m and therefore was estimated assuming a linear temperature profile from the mean temperatures over the study period…"

P3508/23: How did you compute k from the temperatures? I assume you used the method of Conway and Rasmussen (2000), but you need to state the method and reference.

- Yes, Conway and Rasmussen (2000) was used and has been added to the manuscript.

P3509/L9: What do you mean by unvalidated here?

- Pyramid Station provided us with raw meteorological data prior to their quality control processing. This detail has been added to the paper as follows: "… unvalidated, i.e., prior to their quality control processing, …"

P3509/L15: When density did you assume for your snowfall rate to get SWE? Was it a constant value?

- We assumed a density of 150 kg $m^{-3}$. This detail has been added to the paper as follows: "… to derive a snowfall rate assuming a density of snow of 150 kg $m^{-3}$."

P3509/L16: Perhaps it's useful to add a % of missing data?

- Missing 11.9% of data from May 31 to October 12. This has been added to the paper as follows: "… with a few short gaps (missing 11.9% data)."

P3509/L20: Was this comparison done on a month by month basis or on the average of all 4 months?

- The NCEP/NCAR downward longwave radiation data from 2003 to 2010 between the months of June and September were resampled using a linear interpolation such that the temporal resolution of the NCEP/NCAR data would agree with the hourly meteorological measurements from Pyramid Station. The comparison was performed on all of these data and it was found that the NCEP/NCAR data overestimated the incoming longwave radiation by an average of 29 W $m^{-2}$. It is important to note that the comparison was not done for any time period where data were missing from Pyramid Station.

  Changes to the manuscript: The language has been adjusted as follows to clarify: "... the incoming longwave radiation flux at Pyramid Station from 2003 to 2010 (neglecting any data gaps) between the months of June and September revealed that NCEP/NCAR overestimated the incoming longwave radiation by an average of 29 W $m^{-2}$ (results not shown)."

P3510/L26: How well did the linearly interpolated diurnal LWI cycle represent that measured at the Pyramid Station 2003 – 2010?

- Figure R1-2 and R1-3 show that the linearly interpolated NCEP/NCAR values capture the diurnal cycle of incoming longwave radiation, but consistently overestimates the values at Pyramid Station. Furthermore, R2 shows there is a low correlation ($R^2$ = 0.21) between Pyramid Station and NCEP/NCAR. This poor correlation resulting from the use of reanalysis data as opposed to in-situ meteorological data is likely a major source of error in the model.

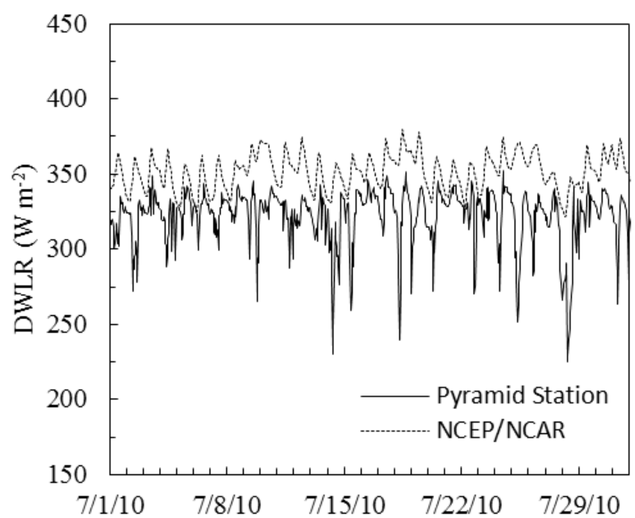


Figure R1-2. Sample time series of Pyramid Station and NCEP/NCAR incoming longwave radiation (W $m^{-2}$) for July 2010.

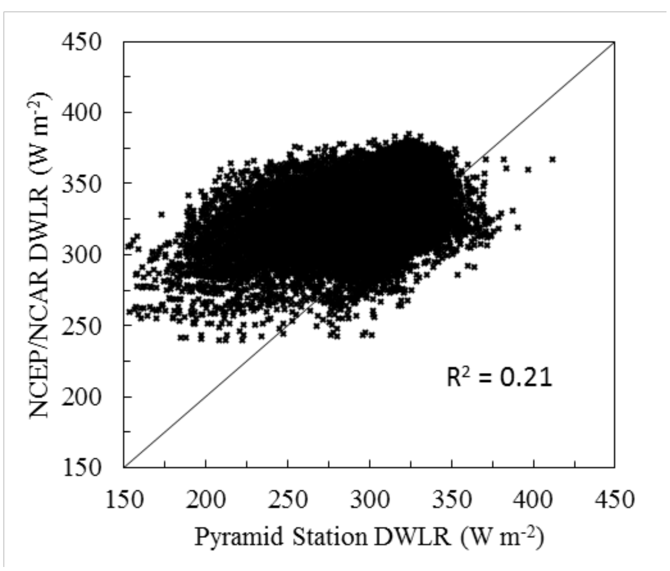


Figure R1-3. Comparison of incoming longwave radiation for Pyramid Station and NCEP/NCAR for all time steps between June and September for the years 2003 – 2010.

P3510/L21: Can you provide information on the accuracy of your GCPs as it affects the resultant scale of the SfM model as far as I understand it.

- The error associated with the total station is very small (0.4 mm) due to the inherent accuracy of the total station and the short distances between the total station and the measurement of the GCPs. For our analysis, the use of cones made it difficult to identify the exact location on the top of the cone in each photo, and this is a source of uncertainty. However, the absolute accuracy of the DEMs is not crucial for our work; it is the relative accuracy that is most important for our roughness calculations, and this will only be affected by GCP placement to a very small extent.

  Changes to the manuscript: the sentence has been revised as follows: "Ground control points (GCPs), collected using a total station with an error less than 0.4 mm, are then used to…"

P3510/L12: State here that your GCPs were obtained using a total station with an accuracy better than 1 mm.

- This change was made in response to the comment above.

P3512/L7: I am not clear what you mean by unit width in this part of your method.

- The original relationship between surface roughness and obstacles derived by Lettau (1969) uses information regarding the obstacle's height, depth, and width (Equation 1). However, the methods used in this paper are based on a transect approach. Therefore, the concept of a unit width (which ultimately cancels out with itself) is used to show the relationship between the original method by Lettau (1969) and the method developed in this paper using a transect approach.

P3513/L5: This is a little unclear to me, but I think you mean that LE = 0 unless the RH in the overlying air (at 2m height at the Pyramid AWS) is at 100%, at which point you also set the surface RH to be 100%? Can you express this more precisely in the text please.

- Yes, your understanding is correct. This has been edited in the text "(2) assuming it is dry unless the relative humidity is 100%, at which point the surface relative humidity is assumed to also be 100% based on the assumption that the water vapor above the surface is well mixed."

P5314/L16: In addition to the reference, please add a sentence describing the nature of the simple snowmelt model used.

- Details of the snowmelt model have been added: "In the event of snow, a simple snowmelt model was used (Fujita and Sakai, 2014), which applies an energy balance over the snow surface that includes net radiation, turbulent heat fluxes, and conductive heat flux with the debris layer in addition to a variable surface albedo of the snow based on the number of days since fresh snow and the air temperature."

P3515/L11: What was the reason for not computing k at site 14 for the 0.05 m depth?

- Site 14 had sensors at 1 cm, 5 cm, and 24 cm. Since the sensors surrounding the 0.05 m depth were not approximately equidistant, a thermal conductivity was not computed.

P3515/L22: You found previously that k varied with depth, or it was dependent on total debris thickness?

- Rounce and McKinney (2014) found the thermal conductivity of the upper 10 cm to be 0.60 W $m^{-1}$ $K^{-1}$ compared to those below 10 cm, which were 1.20 W $m^{-1}$ $K^{-1}$, i.e., the thermal conductivity varied with depth. No relationships between thermal conductivity and total debris thickness were developed.

P3515/L25: Reiterate here the duration of the measurements used in Rounce and McKinney (2014). Also, was there any observable trend in k over time – that would also indicate temperature dependency?

- The duration of measurements used in Rounce and McKinney (2014) was 13-24 Sept 2013.

  Figure R1-4 shows the trends in monthly thermal conductivity (solid lines) for all three sites and the respective depths at which thermal conductivity was measured. Once again, there is no apparent trend between depth and thermal conductivity. On the other hand, there does appear to be a seasonal trend in thermal conductivity such that the thermal conductivity is typically higher in July and August (and June in a few cases) and is smaller in September and October. This appears to closely follow trends in air temperature, where higher air temperatures are observed in June, July, and August. In addition to air temperature, it is likely that rainfall is also contributing to differences in thermal conductivity as moisture in the debris will greatly impact the thermal conductivity. Developing an understanding of the moisture in the debris cover is an important area of future work as it likely will have a great influence on the thermal conductivity in addition to the latent heat fluxes.

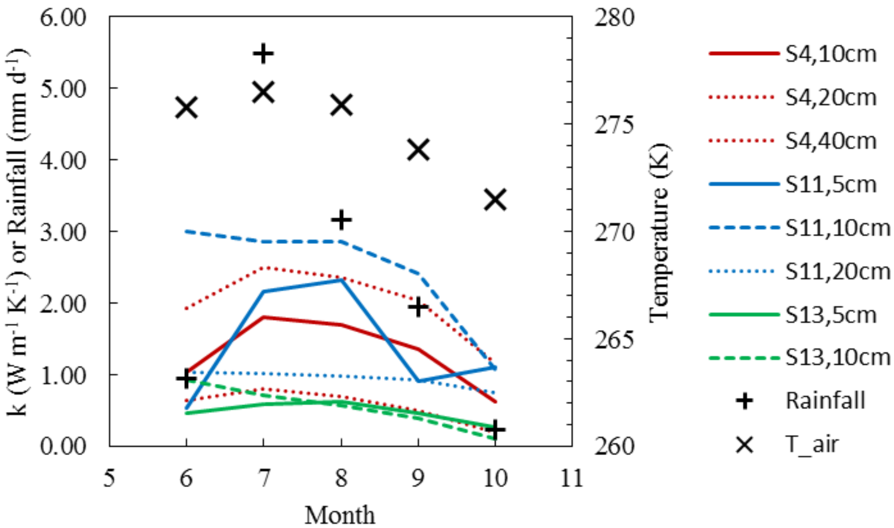


Figure R1-4. Trends in monthly thermal conductivity compared to rainfall and air temperature (T_air).

<u>Changes to the manuscript:</u> The duration of Rounce and McKinney (2014) was added to the text as follows: "… it is likely that the temporally-limited data (13-24 Sept 2013) presented in Rounce and McKinney (2014)…".

A sentence regarding the monthly trends associated with thermal conductivity was added to Section 4.1 as follows: "The thermal conductivities did appear to have a trend over the monsoon season where the highest thermal conductivities were typically observed in July and August, which coincides with higher average air temperature and increased precipitation compared to the other months."

P3516/L7: State explicitly how the error is computed – is it the RMSE computed between the total station location of the GCP and the SfM DEM location of all 4 corner markers? Presumably before you remove the planar slope? Are the values in the table the average or the maximum of these 4 corner marker errors? What is the error on the DEM produced by Agisoft? Did you need to reject any images from the analysis?

- The error for each marker is calculated as the difference between the source (measured) coordinates and the coordinates estimated by PhotoScan Pro and are reported as the root mean square error. In other words, it is the root mean square error between the measured GCPs (all 4 corner markers) from the total station and their estimated position in the SfM DEM. We use this value of root mean square error as the error of the DEM. No images were rejected from the analysis.

  <u>Changes to the manuscript:</u> The following sentence has been added to the manuscript: "The error of the DEM was computed as the root mean square error based on the differences between the measured GCPs from the total station and the modeled position of the GCPs from the software."

P3516/L15: Do you really think 4 GCPs per plot is too sparse? Why?

- Given the quality and distribution of the GCPs within our plots, four points are probably sufficient to generate an accurate transformation. We took care to ensure the points were not clustered, which has been shown in previous work to be a source of error (e.g., James and Robson, 2012; Javernick et al., 2014). The internal consistency of the model is therefore likely to be robust. Adding further GCPs would be a trade-off between improving the distribution of the ground control data and introducing additional sources of error into the process. On reflection, therefore, we suggest that a low number of well-surveyed points is sufficient for plot-scale analyses such as this provided they are evenly spaced throughout the plot.

  Changes to the manuscript: "as well as the sparse coverage of GCPs in our plots" has been removed from the text.

P3516/L25: How was this value for $z_0$ determined. Why do you consider it more accurate than Munro's Method?

- Inoue and Yoshida (1980) estimated $z_0$ from wind speed profiles of near neutral runs at two sites on the Khumbu glacier. The range of values that we estimated from the Lettau-Munro method was 0.0022 – 0.0091 m compared to those reported on the Khumbu glacier, which ranged from 0.0035 – 0.060 m (Inoue and Yoshida, 1980; Takeuchi et al., 2000). Based on the description of the debris cover from Inoue and Yoshida (1980) and our observations of debris cover on Imja-Lhotse Shar glacier, our Site B seemed comparable to their Area III, which had a larger value of $z_0$. Furthermore, the smallest roughness estimation made by Inoue and Yoshida of 0.0035 m was performed on an area of small schist with bare ice, which is further described as the "uppermost part of the ablation zone where debris first appears on the glacier surface. The upper part of this area is the ogive zone which is essentially debris-free, and the lower part is the ice pinnacle zone where supraglacial debris deposits as a dispersed bouldery veneer around ice pinnacles." One would expect this area of mixed debris and bare ice to have a smaller surface roughness than Sites A, C, and D from our study; however, they all have similar values. These comparisons help show that the Lettau-Munro appears to be underestimating surface roughness estimates. Hence, we believe the modified method developed in this paper is an important step forward as it provides reasonable estimates of debris-covered surface roughness that is based on the original methods developed by Lettau (1969), but utilizes the high accuracy of the SfM DEM. Furthermore, the use of the 30% threshold, which is also rooted in the methods of Lettau (1969), provides an objective approach for selecting the minimum obstacle height.

  Changes to the manuscript: The method of measurement has been added to Section 4.2 in the sentence regarding the previous studies as follows: "These values are towards the lower end of those previously reported in literature, which were estimated from wind speed profiles and range from 0.0035 to 0.060 m (Inoue and Yoshida, 1980; Takeuchi et al., 2000; Brock et al., 2010)."

P3518/L24: Specifically which field data? Just temperature?

- Yes, model calibration was performed only using debris temperatures. This has been revised in the text to read "… how well each method models the measured debris temperatures."

P3519/L2: Remove this sentence. It is tautological as the optimization must achieve reasonable values for these parameters as you constrain their possible range according to values from the literature.

- This sentence has been removed.

P3519/L11: $R^2$ between what variables? It seems you performed this on all the available temperature records, correct? Perhaps you'd be better off just doing it for the surface temperatures as (a) you have few measurements at depth and one might be poorly located and (b) other things might be going on within the debris and affecting individual temperature readings at depth are in some way taken into account by using a single optimized k value for each site.

- The $R^2$ was calculated for all the available temperature records, i.e., those on the surface and at depth. The reviewer makes a strong point that optimizing the model using only the surface temperatures would be consistent across all sites and the lack of knowledge of the moisture and thermal conductivity within the debris make it difficult to effectively model the subsurface temperatures. The subsurface temperatures will be better used for estimating the thermal conductivity. Furthermore, the potentially poorly located sensor should not influence the thermal conductivity calculations as shown by Conway and Rasmussen (2000).

  Additionally, Table R1-2 shows that the differences between an optimization performed using all the temperature sensors and only the surface sensors is fairly minimal with the only large discrepancy being Site 13 likely a result of compensating for the poorly located sensor. These results were also reported for the unbounded case to highlight the differences in thermal conductivity, which is the parameter that one would expect to change the most through the incorporation of sensors within the debris. However, these optimized thermal conductivities are once again higher than those previously measured in the field and close to the thermal conductivity of solid granite gneiss (Robertson, 1988). Therefore, with the exception of Site 13, the thermal conductivities would be equal for the bounded condition, which shows there is little difference between including all the sensors or only using the surface sensors. For consistency between the sites and the other reasons discussed above, all the optimizations in the revised paper only use the surface temperatures.

Table R1-2. $LE_{Rain}$ model optimization with k unbounded using all the temperature sensors and only the surface sensors.

| Site | $LE_{Rain}$ - Surface Only | | | $LE_{Rain}$ - All Sensors | | |
|---|---|---|---|---|---|---|
| | α | k[1] | $z_0$[2] | α | k[1] | $z_0$[2] |
| 4 | 0.13 | 2.51 | 0.012 | 0.13 | 2.55 | 0.016 |
| 11 | 0.20 | 3.40 | 0.006 | 0.19 | 3.47 | 0.009 |
| 13 | 0.10 | 0.92 | 0.025 | 0.10 | 1.88 | 0.020 |
| 14 | 0.31 | 2.95 | 0.009 | 0.37 | 2.41 | 0.010 |
| Avg | 0.19 | 2.45 | 0.013 | 0.20 | 2.58 | 0.014 |
| Std | 0.09 | 1.08 | 0.008 | 0.12 | 0.66 | 0.005 |

[1]units of W m-1 K-1; [2]units of m

Changes to manuscript: The calibration was performed using only the surface temperatures as recommended. This has been made clear in Section 3.2: "The calibration was performed by minimizing the total sum of squares of the measured versus modeled surface temperature for each site and was done independently for the three methods used to estimate the latent heat flux."

And also in Section 5.1: "The albedo, thermal conductivity, and surface roughness for each of the three methods were optimized by minimizing the sum of squares of the surface temperature for each site (Table 4)."

P3519/L13: What physical field evidence leads you to believe the sensor moved down over time? I'm not sure how it could do so? Could it just have been poorly located at the outset? If the sensor at 20 cm depth in site 13 was actually at a greater depth, this would also affect the calculation of k at that site and be a reason for your anomalously low k value for this site.

- Unfortunately, there is no field evidence that led us to believe the sensor moved down over time, since the sensor depths were unable to be re-measured prior to retrieval. It is very possible that the sensor was simply poorly located during installation. Another possibility is the estimate of thermal conductivity was too small. We have removed the sentence.

P3520/L5: Might it be clearer to use an alternate data format given that much of the English speaking world does not use the US month/day convention?

- Yes, the manuscript has been changed to (16-18 June and 25-27 July).

P3520/L4: Consider using '… There are a few days for which a positive bias in temperature can be seen during the daily high and nightly low', as overestimating the 'low' might imply modeled temperatures lower than those measured during the nightly low.

- Good point. The manuscript has been changed to read, "there are a few days for which a positive bias in temperature can be seen during the daily high and nightly low (see for example 16-18 June and 25-27 July)."

P3520/L8: Typo – do you mean daily high here?

- Yes, it was a typo. It was supposed to be referring to the nightly low having a positive temperature bias. The manuscript has been changed to read: "one possible explanation for the positive bias in temperature in the nightly low is an overestimation of the incoming longwave radiation…"

P3521/L15: I'd suggest removing this last sentence, as it's not really necessary.

- Agreed. The sentence has been deleted.

P3522/L1: Why not compute ablation for all 14 ablation stake sites? It might provide a more useful model test, as in reality researchers will likely be applying the model to sites for which optimized inputs are not available. See my point above about a single multi-site optimization.

- The reviewer makes a great point here. The sites that had ablation stakes, but lacked temperature sensors to calibrate the parameters, were assumed to have the mean values of thermal conductivity, albedo, and surface roughness as those from the $LE_{Rain}$ model such that ablation was computed at all the sites that had a temperature sensor or an ablation stake.

  Changes in the manuscript: The results are reported in Figure 5 and discussed in further detail in Section 5.3 as follows: "For sites that only had an ablation stake, the average calibrated parameters for that particular latent heat flux model were used."

P3522/L21: The time periods of the ablation stake measurements and the modeled ablation do not match. Did you run the optimized model for the whole period of the ablation stake measurements to provide a comparison of these data? It becomes clear from later text that you did this, but make it clear here as well.

- The inconsistency between the time periods for the ablation stake measurements and the modeled ablation estimates is an important factor that needs to be addressed. Meteorological data was only available from 31 May to 12 October. The meteorological data was required to run the model; therefore, the model was only run during this time period. The temperature sensors and ablation stakes were installed on 17-18 May, but the first 48 hours of the temperature sensor data was discarded to allow the sensors to equilibrate with the debris. Figure R1-5A shows the debris temperature at various depths at Site 4 throughout the entire duration of the study. Figure R1-5B shows the time series from when the ablation stakes and temperature sensors were installed until meteorological data were available. The temperature profiles indicate that there was snow on top of the debris from 26 May to 1 June.

During this time, we assume that no melting was occurring within the debris. However, from 18 May to 26 May, the debris profiles indicate that melting was occurring. We assume that the daily melt rate over this time period is equivalent to the average daily melt rate that was modeled for the first week of June, i.e., 02 – 09 June. Figure R1-5C shows a heavy snow event, consistent with reports in the field, on 13 October, such that the debris remained snow covered until 20 October. The lower thermistors, located at a depth of 40 and 83 cm in the debris (S4-40 and S4-83, respectively), show that the temperature in the debris remained around freezing until the temperature sensors were removed from the debris on 09 November. Therefore, we assume that no melting occurred after 12 October. This is supported by field observations during the retrieval of sensors where the debris was completely frozen and an ice axe was needed to remove the temperature sensors at select sites. Furthermore, the melt during the transition seasons is much smaller than the melt during July and August, so this assumption should not significantly impact the ablation rates over the entire melt season. Nonetheless, these assumptions are required such that the modeled ablation rates and the measured rates from the ablation stakes are temporally consistent.

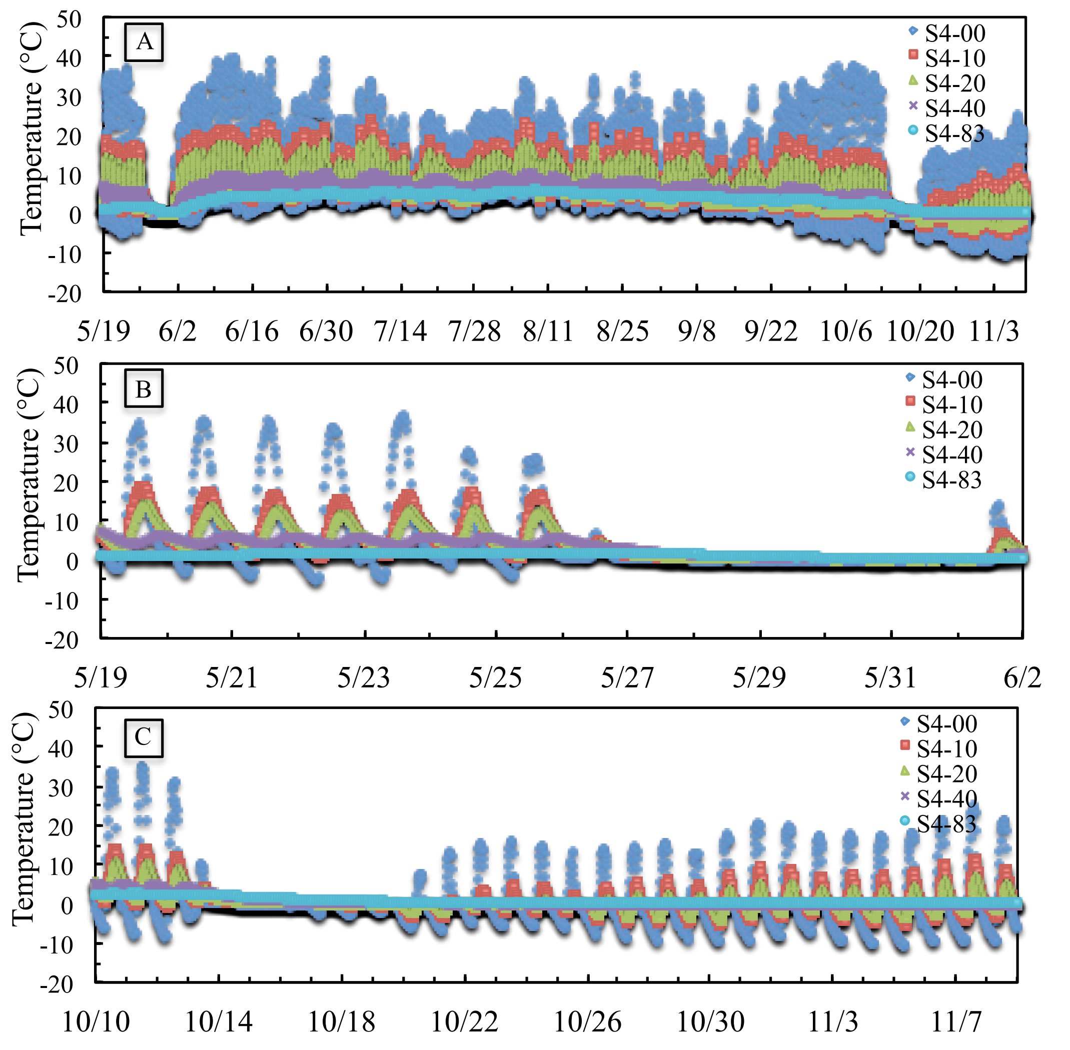

Figure R1-5. Debris temperature measurements from Site 4 for (A) entire study period, (B) prior to modeling time period, and (C) post modeling time period.

Changes in the manuscript: The time period and the assumptions used to have the modeled ablation be consistent with the measured ablation stakes are described in detail in Section 5.3 as follows:

"Ablation rates were computed for all 15 sites that had a temperature sensor or an ablation stake. For sites that only had an ablation stake, the average calibrated parameters for that particular latent heat flux model were used. Additionally, a comparison was performed for the $LE_{Rain}$ model using both the average calibrated parameters for all the sites and the calibrated parameters for each individual site (Figure 7). Ablation rates were modeled over the same time period as the ablation stakes (18 May to 09 November). For days where no meteorological data was available, i.e., the data gaps, the ablation for that day was assumed to be equal to the daily ablation rate for that specific month. As the available meteorological data began on 31 May, the daily ablation rate for the month of May was assumed to be equal to the daily ablation rate of the first week of June. Temperature sensors reveal the debris was snow covered from 26 May to 01 June, so the melting during these days was assumed to be zero. Temperature profiles also show the debris was snow covered from 13-20 October and deeper thermistors reveal the temperature remained around freezing until the sensors were removed in November. Therefore, the melt rates after the 12 October were assumed to be zero."

P3524/L5: Your sensitivities to k are similar to those found in Nicholson and Benn (2006) for debris cover of ca. 30 cm, but in that work we found that the sensitivity of modeled melt (as a % of melt) to k is dependent on the thickness of the debris cover (see Table 2). Your results show higher k sensitivity at sites with lower ablation, coinciding with thicker debris. I think this is worth mentioning in the context of our results where we explicitly explored the thickness dependency of the sensitivities. Nicholson and Benn (2006) found slightly lower sensitivity to albedo than you, and our sensitivity to z0 was variable but similar to your findings. It might be worth comparing these explicitly as the model in Nicholson and Benn (2006) uses only daily averages, but the sensitivities as quoted appear to be similar.

- As the reviewer comments, Nicholson and Benn (2006) assessed the sensitivity of modeled melt compared to $z_0$ and also assessed the uncertainty associated with depth averaged thermal conductivity, k, for both wet and dry debris. Interestingly, the uncertainty associated with the wet debris at Larsbreen (1.67±0.35) was very similar to the uncertainty approximated by our results (±0.40). The uncertainty associated with the thermal conductivity from our results showed that thicker debris is more sensitive than thinner debris. Nicholson and Benn (2006) show the maximum absolute percent change in melt rate in response to a 1% change in the individual input parameters. For thermal conductivity, the maximum absolute percent change in melt rate increases with increasing debris thickness, which is consistent with the results of our sensitivity analysis.

Similarly, the sensitivity analysis shows that the sensitivity of the total melt with respect to surface roughness is also more sensitive for thicker debris. This trend was also observed by Nicholson and Benn (2006) for the wet debris, but not for the dry debris. Interestingly, the initial sensitivity analysis was performed by varying $z_0$ by ±10%, which resulted in an average ±1.5% change in total ablation. This result is highly consistent with the sensitivity found by Nicholson and Benn (2006) of 1-2% caused by a 10% change for surface roughness. However, the revised sensitivity analysis (in response to a comment from the other review) shows that the total ablation is quite sensitive to changes in surface roughness, especially for decreases in surface roughness.

Lastly, the thickness of the debris does not appear to have any effect with respect to variations in albedo, which is also consistent with Nicholson and Benn (2006).

Changes in the manuscript:
The discussion regarding the sensitivity analysis has been changed to incorporate these comparison to Nicholson and Benn (2012) and the uncertainties regarding each parameter (in response to a comment from the other reviewer). Section 5.4 has been completely revised as detailed in the response to the other reviewer.

P3525/L8: So here for the temporal resolution you model ablation for all 14 sites, but in the evaluation of your standard model only for 10 sites?

- The discussion paper only compared the 10 sites where ablation was modeled with their respective ablation stake measurements. However, the use of the average optimized parameters allows the other 5 sites to be modeled such that all 14 ablation stake measurements may be compared to the modeled ablation estimates. For the assessment of temporal resolution, the $R^2$ values of the temperature sensors were used for comparison and the modeled ablation estimates for all 15 sites were also used to compare differences in melt rates between the different time steps.

  Changes to the manuscript: The text has been revised to read "The $R^2$ correlation coefficients for the sites with temperature sensors and the modeled total melt for all 15 sites were used to assess the effect of temporal resolution on model performance."

P3525/L25: So here you are comparing the total ablation from your stake data – only available at sites 8, 13, and 15? Or are you comparing to your higher temporal resolution model? I am unclear as to which temporal resolution is performing best as compared to the stake data at these 3(2) sites. I say 2 sites because site 13 is clearly complicated, and by optimizing the model on a poorly located temperature sensor at depth it makes a poor test of the model.

- The ablation comparisons are simply assessing modeled total ablation for various temporal resolutions such that only the effects of the temporal resolution are being assessed. The comparison of the modeled ablation and the measured ablation rates reveals that 2 of the 3 sites overestimated ablation, while the other underestimated ablation. A comparison of the modeled ablations for the different temporal resolutions at these three sites was not included in the text, as it does not yield any useful information. The response to the previous comment should clarify what is being compared in the text to assess the effect of temporal resolution.

P3526/L4: In addition to your comment about snow, which is at least partly dependent on the manner in which snow is treated in the model, a daily average model is likely to perform poorly during seasonal transitions even if the surface is not snow-covered as temperature profiles through the surface debris that is either heating up or cooling down through the seasonally transition are often not linear, as shown by the measurements in Nicholson and Benn (2012).

- Good point. The sentence has been adjusted to include this as well: "… the model will perform poorly towards the transition seasons due to changes in the temperature profiles resulting from freeze/thaw conditions and higher amounts of snow fall. Additionally, caution should be taken with respect to ablation estimates as these will likely be slightly underestimated."

**References:**

Rees, W.G. and Arnold, N.S.: Scale-dependent roughness of a glacier surface: implications for radar backscatter and aerodynamic roughness modelling, J. Glaciol., 52, 214-222, 2006.

James, M.R. and Robson, S.: Straightforward reconstruction of 3D surfaces and topography with a camera: Accuracy and geoscience application, Journal of Geophysical Research, 117, F03017, doi:10.1029/2011JF002289, 2012.

Javernick, L., Brasington, J., and Caruso, B.: Modeling the topography of shallow braided rivers using Structure-from-Motion photogrammetry, Geomorphology, 213, 166-182, 2014.

Dear Reviewer,

Thank you very much for your insightful and constructive comments. We have addressed these below and have revised text in the manuscript accordingly.

**Specific Comments**

*Sections 3.1 and 4.2: $z_0$ estimation using Structure from Motion*

How was the accuracy of the DEM assessed? You analyze variations at the 0.01 m scale in estimation of $z_0$; is that resolution justified? The maximum area of the DEM according to the values given at line 20 on p. 3510, is 4 $m^2$, but how much of this area is useable for assessing microtopography?

- The accuracy of the DEM is assessed using the errors estimated by Agisoft in the SfM software. This is the root mean square error based on the differences between the measured GCPs from the total station and the modeled position of the GCPs from the software.

  The total RMSE values associated with each DEM are reported in Table 2 and ranged from 0.008 m – 0.024 m. We chose a pixel spacing of 0.01 m such that the resolution of the DEM was on the same order as the DEM errors (3 of the 4 sites had errors less than 0.01 m), which justifies the choice of the resolution.

  At each site, the area inside the cones was clipped such that the cones would not influence the surface roughness calculations. The actual area of these clipped plots that were used for analysis were 3.98, 3.23, 5.29, and 3.72 $m^2$ for Sites A-D, respectively.

  <u>Changes to the manuscript:</u>

  "The error of the DEM was computed as the root mean square error based on the differences between the measured GCPs from the total station and the modeled position of the GCPs from the software" has been added to the paragraph on the surface roughness methods (p3510).

  "The DEMs were resampled to a resolution of 0.01 m such that their resolution was on the same order as their respective errors (3 of the 4 sites had a total RMSE less than 0.01 m)" has been added to the field results of surface roughness (Section 4.2, p3516).

SfM is not my area of expertise, so I cannot comment on the validity of the methodology, but I like the fact that your approach attempts to generalize the characteristic roughness over the area of the DEM and generate a probable range of $z_0$ rather than specific $z_0$ values on individual profiles. Something that is missing from the discussion in Section 4.2, however, is an appreciation that the atmospheric boundary layer and its corresponding $z_0$ value develops over

100s of m of fetch. It is therefore questionable whether 4 $m^2$ 'samples' of the glacier microtopography are sufficient to characterize the surface roughness over such large areas, and it is doubtful whether the wind profile would be able to adjust to order of magnitude roughness changes (Table 2) over the short distances in your study area. Consequently it is physically meaningless to have $z_0$ as a tuneable parameter at individual stakes. It would be more realistic if the 4 SfM sites were combined to generate a single $z_0$ value and its range.

- As the reviewer notes, the scale at which the atmospheric boundary layer develops is a critical parameter in determining how $z_0$ will vary over the course of the debris-covered glacier. Lettau (1969) conducted his experiments over fetches of 50 m in length. In regions of relatively homogenous terrain, this is not an issue; however, on debris-covered glaciers, the terrain is highly heterogeneous over a 50 m span as the terrain is both hummocky and the grain sizes vary greatly. Therefore, it is a crucial area of future work to understand the scale (fetch length) at which $z_0$ varies over the glacier.

  Brock et al. (2006) compared microtopographic and aerodynamic roughness over snow, slush, and ice and found close agreement. Interestingly, they found no significant difference between the use of a 3 m and 15 m transect; however, they did state that a shorter pole would be unlikely to capture a sufficient sample of roughness elements and a longer pole should be used if vertical changes are greater than 1 m. The use of hundreds of transects over a ~4 $m^2$ grid has the benefit of expanding the number of surface roughness elements that can be captured compared to a single traditional cross section. In this regard, while the approach is likely limited to roughness elements < 1 m, the approach should be able to capture a sufficient number of elements. Additionally, the method developed in this paper using an obstacle density of 30% allows one to take advantage of the high resolution DEM that can help to capture the irregularly shaped elements as opposed to using the method of Munro (1989), which as Brock et al. (2006) states was "a necessary generalization… due to the difficulty of measuring and converting irregularly shaped elements in a $z_0$ value."

  Unfortunately, there is no comparison between microtopographic and aerodynamic $z_0$ values on debris-covered glaciers. The $z_0$ values reported on debris cover (Inoue and Yoshida, 1980; Takeuchi et al., 2000; Brock et al., 2010) only measured aerodynamic roughness. Therefore, the scale (or fetch length) at which the aerodynamic roughness may vary over the terrain is unknown. This is an important area of future work. Nonetheless, the techniques developed in this study show a range of $z_0$ values for various grain sizes found on Imja-Lhotse Shar glacier. The average value of these four sites is 0.018 m (standard deviation of 0.013), which is similar to the value of 0.016 m found by Brock et al. (2010) on a debris-covered glacier in Italy "comprising a mixture of granites and schists of predominantly cobble size, with occasional boulders of < 1m size."

  As requested by the reviewer, an optimization was performed using a constant value for $z_0$ of 0.018 m. Table R2-1 shows that the thermal conductivity is at the upper bound for 8 of the 10 sites, which is similar to the optimization allowing $z_0$ to vary for each site. The noticeable difference between optimizations is the increase in average $z_0$ from 0.014 m to 0.018 m causing the average albedo to decrease from 0.32 to 0.27. This makes intuitive sense as a higher surface roughness removes more energy from the debris, which is compensated by a lower albedo that will cause the surface to absorb more energy. The performance of the model appears to be comparable with the average $R^2$ of the surface sites being 0.73 for the constant $z_0$ compared to 0.74 when $z_0$ is allowed to vary for each site. Furthermore, the total sum of squares only varied by 5%. Without detailed knowledge of albedo, surface roughness, and thermal conductivity, it is difficult to determine which model performs better.

Table R2-1. $LE_{Rain}$ model optimization allowing $z_0$ to vary and setting it at a constant value of 0.018 m.

| | $LE_{Rain}$ | | | $LE_{Rain}$ | | |
|---|---|---|---|---|---|---|
| Site | α | k[1] | $z_0$[2] | α | k[1] | $z_0$[2] |
| 4 | 0.26 | 1.62 | 0.006 | 0.10 | 1.62 | 0.018 |
| 5 | 0.40 | 1.62 | 0.014 | 0.40 | 1.62 | 0.018 |
| 6 | 0.40 | 1.09 | 0.036 | 0.40 | 1.39 | 0.018 |
| 11 | 0.37 | 1.62 | 0.006 | 0.26 | 1.62 | 0.018 |
| 13 | 0.10 | 0.92 | 0.025 | 0.20 | 0.61 | 0.018 |
| 14 | 0.39 | 1.61 | 0.015 | 0.38 | 1.62 | 0.018 |
| 15 | 0.38 | 1.62 | 0.006 | 0.29 | 1.62 | 0.018 |
| 17 | 0.30 | 1.62 | 0.006 | 0.19 | 1.62 | 0.018 |
| 19 | 0.33 | 1.62 | 0.019 | 0.34 | 1.62 | 0.018 |
| 20 | 0.28 | 1.62 | 0.006 | 0.18 | 1.62 | 0.018 |
| Avg | 0.32 | 1.50 | 0.014 | 0.27 | 1.50 | 0.018 |
| Std | 0.09 | 0.26 | 0.010 | 0.11 | 0.32 | 0.000 |

[1]units of W m-1 K-1; [2]units of m

This analysis does stress the importance of accurately measuring the albedo, surface roughness, and thermal conductivity. For example, Sites 6 and 13 show how varying $z_0$ causes the thermal conductivity to change by 0.30 W $m^{-1}$ $K^{-1}$, which would have a significant impact on ablation rates. Future work should strive to measure all three of these parameters. Specifically, it is important to determine the scale at which surface roughness is equal to aerodynamic roughness, understand how albedo varies throughout the day and throughout the melt season, and improve measurements of thermal conductivity. The latter will likely be hampered by the limited ability of the temperature sensors to capture the small temperature changes with respect to time at deeper depths in the debris and our understanding of soil moisture within the debris profile(s). One promising method is to measure the albedo and surface roughness in conjunction with meteorological data and ablation stakes, such that the ablation stakes may be used to approximate the thermal conductivity.

Changes to the manuscript: A paragraph has been inserted into the manuscript in Section 4.2 as follows: "Future work should seek to compare these estimates of surface roughness with aerodynamic roughness to determine the scale at which these two values agree. Brock et al. (2006) found there to be no significant difference between the use of a 3 m and 15 m transect; however, they did state that a shorter pole would be unlikely to capture a sufficient sample of roughness elements if the vertical changes are greater than 1 m. The use of hundreds of transects over a ~4 $m^2$ grid has the benefit of expanding the number of surface roughness elements that can be captured compared to a traditional single transect. However, Brock et al. (2006) was comparing microtopographic and aerodynamic roughness over snow, slush, and ice, which is significantly different from the hummocky and heterogeneous terrain on debris-covered glaciers. Therefore, it will be important to determine the scale or fetch length at which the surface roughness agrees with the aerodynamic roughness. Nonetheless, the method developed in this paper provides an objective approach to select an obstacle height and yields consistent and reasonable estimates of $z_0$ for various grain sizes independent of the resolution of the DEM."

*Section 4.1: Calculation of thermal conductivity:*

Why were only the near-surface thermistors down to 20 cm used and not the deeper thermistors at sites 4 and 11? This implies that the k calculation is biased towards the openwork clast layers at the surface and not the more compact and humid lower debris layers. It could be reasoned that as void spaces of deeper layers are filled with water and fine rock material the thermal conductivity here would be higher than near the surface where void spaces are filled with low conductivity air. This implies an underestimation of the full depth k due to the use of temperature data only from the upper layers. This reasoning is supported by the fact that in the calibration of model parameters in Section 5.1 the optimal k value is at the maximum for most sites, implying the true value is greater than 1.62.

- The method from Conway and Rasmussen (2000) is based on approximations of T" ($\delta^2T/\delta Z^2$) and $\dot{T}$ ($\delta T/\delta t$) using standard centered finite-difference expressions. While Conway and Rasmussen (2000) show in the Appendix that the analysis is not affected by the thermistor position and calibration errors, this is referring to errors in measured position and errors in temperature readings between different sensors. The use of a standard centered finite-differences means the ideal set up is to have three sensors that are all more or less equidistant from one another. For example, three temperature sensors at 10 cm, 20 cm, and 30 cm allow $\delta T/\delta Z$ at 15 cm and 25 cm to be approximated, which is then used to approximate $\delta^2T/\delta Z^2$ at 20 cm. This coincides with the 20 cm temperature sensors, which records $\delta T/\delta t$. Nicholson and Benn (2012) note that deeper sensors are also problematic as temperature sensors may not be sensitive enough to capture the small temperature changes that occur, which was found to be the case with the temperature sensor at a depth of 0.83 m at Site 4. Therefore, the combination of the lack of sensors spaced at a reasonable distance to use the standard centered finite-difference approach and the thermistors not being sensitive enough to record temperature changes at deeper depths caused the deepest thermal conductivity to be reported at 20 cm. This could be improved in the future by spacing sensors similar to Nicholson and Benn (2012) and by using more sensitive temperature sensors.

The reviewer brings up an important point that the lack of thermal conductivity calculations at depth may cause the actual thermal conductivity at the site to be underestimated, as the deeper layers that may be more compact and humid within the debris are not considered. Nicholson and Benn (2012) estimated the effects of 10% and 20% of the void space being filled with water and found the thermal conductivities increased to 1.42 and 1.55 W $m^{-1}$ $K^{-1}$ compared to the summer dry debris value of 1.29 W $m^{-1}$ $K^{-1}$. This lends confidence to the higher average values of $k$ found in this study at Sites 4 and 11 of 1.44 and 1.62 W $m^{-1}$ $K^{-1}$. However, adjusting the thermal conductivities based on the percent of void space filled with water requires detailed knowledge of the soil moisture within the debris and how the soil moisture varies with depths, which unfortunately was not measured in this study.

To address the comment that the thermal conductivities may be underestimated, an optimization was performed for all three models with $k$ unbounded and only using the surface temperature sensors at each site (in response to a comment from the other reviewer). In this optimization, the thermal conductivities at Sites 4, 11, and 13 were also allowed to vary (as opposed to being held constant at their average measured value) to avoid the bias associated with only measuring thermal conductivities near the surface. The results of the unbounded thermal conductivity optimization are shown in Table R2-2.

Table R2-2. Model optimization with $k$ unbounded

| Site | $LE_{Rain}$ | | | $LE_{RH100}$ | | | $LE_{Dry}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | α | k[1] | $z_0$[2] | α | k[1] | $z_0$[2] | α | k[1] | $z_0$[2] |
| 4 | 0.13 | 2.51 | 0.012 | 0.21 | 2.34 | 0.008 | 0.10 | 3.30 | 0.024 |
| 5 | 0.35 | 3.99 | 0.006 | 0.31 | 4.50 | 0.006 | 0.35 | 4.20 | 0.006 |
| 6 | 0.40 | 1.09 | 0.036 | 0.40 | 1.34 | 0.036 | 0.40 | 1.20 | 0.036 |
| 11 | 0.20 | 3.40 | 0.006 | 0.22 | 3.18 | 0.006 | 0.15 | 3.90 | 0.012 |
| 13 | 0.10 | 0.92 | 0.025 | 0.10 | 1.77 | 0.021 | 0.10 | 2.10 | 0.024 |
| 14 | 0.31 | 2.95 | 0.009 | 0.38 | 2.42 | 0.009 | 0.35 | 2.70 | 0.012 |
| 15 | 0.18 | 3.64 | 0.006 | 0.17 | 3.76 | 0.006 | 0.20 | 3.60 | 0.006 |
| 17 | 0.26 | 2.03 | 0.006 | 0.27 | 1.96 | 0.006 | 0.25 | 2.10 | 0.006 |
| 19 | 0.20 | 2.66 | 0.026 | 0.29 | 2.24 | 0.017 | 0.15 | 3.00 | 0.036 |
| 20 | 0.26 | 1.80 | 0.006 | 0.27 | 1.75 | 0.006 | 0.25 | 2.10 | 0.006 |
| Avg | 0.24 | 2.50 | 0.01 | 0.26 | 2.53 | 0.01 | 0.23 | 2.82 | 0.02 |
| Std | 0.10 | 1.04 | 0.01 | 0.09 | 0.99 | 0.01 | 0.11 | 0.95 | 0.01 |

[1]units of W m-1 K-1; [2]units of m

Table R2-2 shows the optimized values of *k* are as high as 4.5 W $m^{-1}$ $K^{-1}$. The lithology of the debris cover located in the Everest region is predominantly granite, gneiss, and pelite (Hambrey et al., 2008). Robertson (1988) reports a thermal conductivity of solid granite gneiss to be 2.87 W $m^{-1}$ $K^{-1}$. Therefore, many of the values reported in Table R2.1 appear to be unreasonable, especially when considering the fact that the pores of the debris are filled with air and water, which would cause the effective thermal conductivity to be smaller than that of solid rock. Interestingly, Site 13 has the lowest value of thermal conductivity for all three models, which is consistent with the estimations of thermal conductivity compared to Sites 4 and 11. However, its value (~1.20 W $m^{-1}$ $K^{-1}$) is still more than twice as high as the average *k* measured (0.47 W $m^{-1}$ $K^{-1}$). Therefore, while the unbounded analysis yields some interesting results, it appears to be an unreasonable method to optimize the thermal conductivity at each site as many of the sites have values of *k* that are unreasonable.

The unbounded analysis shows that it is important to have an upper bound for the optimization. Figure R2-1 shows the variations in total sum of squares for all the surface sites for various values of thermal conductivity. The total sum of squares appears to plateau around 2.0 W $m^{-1}$ $K^{-1}$ and does not appear to significantly change for values above 1.6 W $m^{-1}$ $K^{-1}$. In fact, the difference between the minimum sum of squares, which occurs at 2.6 W $m^{-1}$ $K^{-1}$, and the minimum associated with 1.6 W $m^{-1}$ $K^{-1}$ is only 3%. This study found the thermal conductivity to vary from 0.42 to 2.28 W $m^{-1}$ $K^{-1}$ with the highest average thermal conductivity being Site 11 with a value of 1.62 W $m^{-1}$ $K^{-1}$. As previously discussed, these measurements may not be representative of the entire debris as they were measured closer to the surface; however, the value of 1.62 W $m^{-1}$ $K^{-1}$ is higher than those previously reported in the Everest region (Conway and Rasmussen, 2000; Nicholson and Benn, 2012). This value is also similar to the value reported by Nicholson and Benn (2012) of 1.55 W $m^{-1}$ $K^{-1}$, which assumed that 20% of the pore space was filled with water. Therefore, a value of 1.62 W $m^{-1}$ $K^{-1}$ appears to be a reasonable upper bound as higher values do not significantly reduce the total sum of squares and it is a reasonable value of effective thermal conductivity. The uncertainty associated with the thermal conductivity and the upper bound will be addressed via the sensitivity analysis, which will use an uncertainty of +/- 0.4 W $m^{-1}$ $K^{-1}$ as this covers the range of the value where the sum of squares levels off (2.0) and the effective thermal conductivity reported by other studies (1.22).

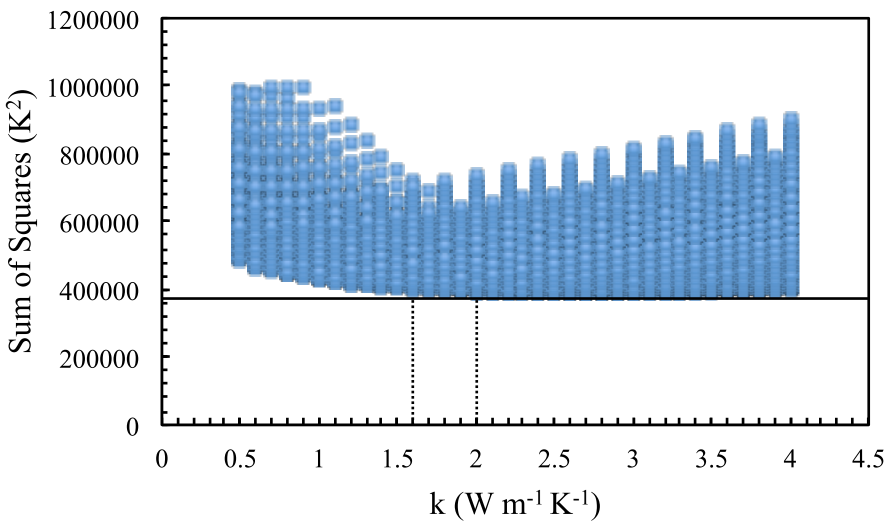


Figure R2-1. Total sum of squares for all surface sites as a function of *k*

Changes to the manuscript: A discussion regarding the upper bound and the calibration of k has been added to Section 5.1 as follows: "For the $LE_{Rain}$ and $LE_{RH100}$ model, 7 of the 10 sites had a thermal conductivity at the upper bound (1.62 W $m^{-1}$ $K^{-1}$), while for the $LE_{Dry}$ model 9 of the 10 sites were at the upper bound. An additional calibration was performed allowing the thermal conductivity to be unbounded and found 3 or more out of the 10 sites for each method had thermal conductivities greater than 3.0 W $m^{-1}$ $K^{-1}$ with one thermal conductivity as high as 4.5 W $m^{-1}$ $K^{-1}$. The lithology of the debris cover in the Everest region is predominantly granite, gneiss, and pelite (Hambrey et al., 2008). Robertson (1988) found the thermal conductivity of solid granite gneiss to be 2.87 W $m^{-1}$ $K^{-1}$, so the unbounded thermal conductivities do not appear to make physical sense when one considers that the thermal conductivity of debris should be much lower than solid rock due to the pore spaces being filled with air and water. Furthermore, an optimization performed using the total sum of squares of all the surface sites reveals that increasing the thermal conductivity from 1.6 W $m^{-1}$ $K^{-1}$ to its minimum of 2.6 W $m^{-1}$ $K^{-1}$ only reduces the total sum of squares by 3%. Therefore, the results reported in this study use an upper bound of 1.62 W $m^{-1}$ $K^{-1}$ and the importance of accurately measuring the thermal conductivity will be discussed in the sensitivity analysis."

You use values from Nicholson and Benn (2012) for debris characteristics in the k calculation. Are these values representative for the sites you measured, and how sensitive is k value to e.g. a 10% change in porosity?

- Nicholson and Benn (2012) was performed on Ngozumpa glacier, which is located ~25 km away from Imja-Lhotse Shar glacier. The characteristics of the debris cover on Ngozumpa glacier appear to be very similar to those on Imja-Lhotse Shar glacier, which lends confidence to the use of the same values as their study. A 10% change in porosity would result in a 5% change in thermal conductivity.

*Section 4.3*

It probably doesn't need pointing out, but it is a shame that most of the stakes melted out. Why didn't you drill the stakes in deeper, e.g., 2 m?

- A mechanical drill was used to drill the ablation stakes and after drilling ~1 m into the ice, the drill bit would get stuck. A thermal drill would be preferable for future work.

*Section 5.1*

A problem with the method of tuning model parameters separately for each of the 3 different methods of estimating turbulent fluxes is that the parameter values will compensate for any errors in the turbulent flux calculation. For example, in the $LE_{Dry}$ method, neglect of the energy used in evaporating water could lead to overestimation of the conductive heat flux, however this is offset by the relatively high albedo and $z_0$ optimized for this run (Table 3), which reduce net shortwave radiation and increase sensible heat transfer away from the surface during the daytime. This makes interpretation of the differences in performance of the model with three different LE flux formulations difficult. Maybe you should select one set of optimal parameters and look at the difference in performance between the three model formulations again. At least this issue needs some discussion.

- The optimization of each model certainly allows the $LE_{Dry}$ model to compensate for the lack of a latent heat flux term through higher thermal conductivities and surface roughnesses, which allow more energy to be removed from the surface. The difficulty in assessing model performance, as the reviewer commented, was discussed in the text (p3520, line 20). As albedo, thermal conductivity, and surface roughness were not measured at each site, this study lacks the ability to assess which model most accurately estimates the debris properties. However, the latent heat fluxes have been found to be a significant energy sink after rain events as discussed in the paper, so the instantaneous and daily average over the melt season latent heat fluxes may be used to show that the $LE_{Dry}$ model is not physically accurate.

*Section 5.4*

I think you should redo the sensitivity analysis, varying each parameter in its range of uncertainty rather than a flat value of +/- 10%. It is obvious a priori that varying $z_0$ by +/- 10% will have little effect on the magnitude of the turbulent fluxes when this parameter can vary across several orders of magnitude. In contrast, a variation in albedo of 10% is quite a significant change as its range of variation is much smaller. This would give a better assessment of model uncertainty which reflects the possible range of values of the input parameters.

- The range of uncertainty is difficult to assess as there are a limited number of measurements of the debris properties in this region. However, as the reviewer comments, it would be more beneficial to perform the sensitivity analysis by trying to account for the uncertainty within each parameter as opposed to using a set value of ± 10%. The uncertainty in thermal conductivity, as discussed above, will be ±0.40 W $m^{-1}$ $K^{-1}$. The uncertainty associated with the surface roughness, will be ±0.010 m, which is the approximate standard deviation associated with the $z_0$ values from the model calibration for each of the three models and similar to the standard deviation between the four sites where $z_0$ was measured (±0.013 m). Lastly, the uncertainty of the albedo will be estimated as ±0.10, which is the approximate standard deviation within the model calibration for each of the three models.

  Table R2-3 shows the percent changes in the total melt (m) as a function of the uncertainty associated with each parameter. The $LE_{Rain}$ model is used as the baseline case and the average value for each of the calibrated parameters ($\alpha$,k,$z_0$) from the model optimization is used for each site. The use of the average calibrated parameters at every site was required because a -0.010 m adjustment to $z_0$ for sites with low $z_0$ values (0.006 m) would cause $z_0$ to be negative, which is impossible. Nonetheless, Table R2-3 shows how the uncertainty within each parameter effects the total ablation rates. The total ablation is most sensitive to changes in thermal conductivity, where a ±0.40 W $m^{-1}$ $K^{-1}$ change in thermal conductivity causes on average a ±20.5% change in total melt. Total ablation is also moderately sensitive to changes in albedo, where a ±0.10 change in albedo caused a ±12.0% change in total melt. The effect of surface roughness is quite interesting as the model is quite sensitive to decreases in surface roughness, i.e., a -0.010 m change in $z_0$ caused a +15.0% change in melt; while, a +0.010 m increase in $z_0$ caused only a -7.3% change in total melt on average. This reveals that the model is much more sensitive to decreases in surface roughness compared to higher values of surface roughness. surface roughness compared to increases.

  Table R2-3. Sensitivity analysis showing percent changes relative to the total melt (m) as a function of the calibrated parameters ($\alpha$,k,$z_0$) for all sites over the study period using the $LE_{Rain}$ model in conjunction with the average calibrated parameters for all the sites.

| | Parameter | α | | k | | $z_0$ | |
|---|---|---|---|---|---|---|---|
| | *Adjustment* | *+ 0.10* | *- 0.10* | *+ 0.40* | *- 0.40* | *+ 0.010* | *- 0.010* |
| Site | Total Melt (m) | % Change | | | | | |
| 4 | 0.30 | -12.3 | +13.0 | +30.1 | -29.4 | -9.7 | +21.7 |
| 5 | 0.90 | -12.6 | +13.0 | +22.3 | -24.0 | -9.2 | +19.4 |
| 6 | 2.70 | -11.1 | +11.3 | +13.1 | -16.3 | -4.1 | +7.9 |
| 7 | 0.92 | -12.6 | +12.6 | +22.0 | -23.5 | -8.8 | +18.8 |
| 8 | 1.03 | -11.6 | +11.9 | +20.7 | -22.5 | -7.7 | +16.1 |
| 10 | 1.61 | -12.2 | +12.1 | +18.9 | -21.2 | -7.7 | +15.3 |
| 11 | 1.12 | -12.0 | +12.5 | +20.3 | -22.1 | -7.7 | +16.2 |
| 12 | 1.30 | -11.8 | +12.0 | +19.8 | -21.9 | -7.8 | +16.1 |
| 13 | 1.05 | -12.4 | +12.7 | +21.2 | -22.9 | -8.5 | +17.8 |
| 14 | 1.72 | -11.9 | +11.8 | +18.1 | -20.7 | -7.0 | +14.1 |
| 15 | 0.90 | -12.4 | +12.9 | +21.8 | -23.4 | -8.6 | +17.7 |
| 16 | 1.76 | -11.6 | +12.1 | +18.3 | -20.2 | -6.7 | +14.0 |
| 17 | 2.77 | -10.8 | +11.1 | +12.0 | -15.0 | -2.9 | +5.7 |
| 19 | 2.04 | -11.6 | +11.7 | +16.5 | -19.4 | -6.3 | +12.3 |
| 20 | 1.81 | -11.4 | +11.4 | +16.5 | -19.4 | -6.1 | +12.1 |
| | **Average** | **-11.9** | **+12.1** | **+19.4** | **-21.5** | **-7.3** | **+15.0** |

Changes to the manuscript:

Table 4 has been replaced by Table R2-3, which has the updated values incorporated into the sensitivity analysis. Additionally, the text accompanying the sensitivity analysis (Section 5.4) has been revised accordingly to read:

"A sensitivity analysis was performed to assess how albedo, thermal conductivity, and surface roughness affect the total ablation (Table 4) based on the uncertainty with respect to each parameter. The uncertainty in thermal conductivity was ±0.40 W $m^{-1}$ $K^{-1}$ as described above. The uncertainty associated with the surface roughness was ±0.010 m, which is the approximate standard deviation associated with the $z_0$ values for each of the three models (Table 3) and similar to the standard deviation between the four sites where $z_0$ was measured (±0.016 m). Lastly, the uncertainty of the albedo was estimated as ±0.10, which is the approximate standard deviation within the model calibration for each of the three models and is the difference between the mean and median albedo measured by Nicholson and Benn (2012) on Ngozumpa glacier. The $LE_{Rain}$ model was used as the baseline case and the average value for each of the calibrated parameters (α, k, $z_0$) from the model optimized is used for each site.

Table 5 shows the total ablation is most sensitive to changes in the thermal conductivity, where a ±0.40 W $m^{-1}$ $K^{-1}$ change causes a ±20.5% change in total ablation on average. The uncertainty associated with the thermal conductivity is also more sensitive to thicker debris, which is consistent with the findings of Nicholson and Benn (2012). Total ablation is also moderately sensitive to changes in the albedo, where a ±0.10 change causes a ±12.0% change in total ablation. Lastly, the total ablation is least sensitive to changes in increasing the surface roughness, as a +0.010 m increase in $z_0$ caused only a -7.3% change in total ablation. However, the model was quite sensitive to a reduction in the $z_0$ of -0.010 m, which caused an average change in total ablation of +15.0%. The sensitivity associated with $z_0$ also appears to increase with an increase in debris thickness. These results highlight the importance of properly estimating the thermal conductivity, but also show the surface roughness and the albedo are important as well."

**Minor Corrections**

Paper title. You need to insert glacier between 'Debris-covered' and 'energy'. As it stands, the title literally means that the model is debris-covered, not the glacier. Make this change everywhere the phrase 'Debris-covered energy balance model' appears in the paper, e.g., section 3.2 title.

- Good point. The changes have been made to the title and throughout the entire text.

P3506/17: I think you mean partial density of water vapour, not water vapour pressure. If vapour pressure was the same at 2 m as at the surface, there would be no vapour pressure gradient and hence no latent heat flux.

- Correct. It has been changed to "water vapour partial pressure".

P3507/19: 'cobble and gravel'.

- Correct. It has been changed to "cobble and gravel".

P3509: Section 2.2, make it clear that the Pyramid Station is an off-glacier station.

- The manuscript has been changed to "… located off-glacier, next to the Khumbu glacier, approximately 14 km northwest of Imja-Lhotse Shar glacier."

P3519/23: '...this particular temperature sensor…', which one? Be specific.

- Site 13 at 20 cm depth. The manuscript has been changed to include "(Site 13, 20 cm)."

P3521/14: I think you mean overestimate not underestimate.

- The other reviewer also commented on this. The language has been changed such that we refer to overestimating temperatures as a positive bias in temperature to avoid confusion with overestimating the nightly low, which could be thought of as lower temperatures.

P3522/12: You can't say that thin debris 'promotes ablation' as you have not measured bare ice melt rates you can't determine that the melt rate beneath thin debris is greater than that for bare ice. In all likelihood even the thinnest of your debris layers is reducing ablation through insulation.

- Good point. The sentence has been modified to read as "… as thin debris has higher rates of ablation compared to thicker debris, which insulates the ice to a greater extent thereby further retarding ablation."

P3519/23: These are not surface temperatures but temperatures at 0.01 m depth – an important distinction.

- Changed from "surface temperatures" to "temperatures close to the surface" and "near the surface".

P3523/20: You can't say the model agrees with the ablation stake data, when the stakes have melted out. All you can see is that the stake data do not contradict the model calculations.

- The sentence has been modified accordingly to read as "All the other model estimates of ablation were near to or greater than 1 m, which was also observed by their respective ablation stakes as they completely melted out of the ice."

Figure 1: What is the background image?

- "Landsat 8 panchromatic image from 14 Nov 2014" has been added to the figure's caption.

Figure 2: State the size of the target discs in the caption.

- The target discs are 19 cm in diameter. This has been added to the caption.

Figure 5: Why is the ablation for $LE_{Dry}$ so much higher than the other two methods for the thinnest debris layer, compared with the other sites? There is an inconsistency in terminology with the text here: $LE_{Dry}$, $LE_{Zero}$.

- The ablation for that particular thin debris layer (Site 6) was much higher than any of the other sites as a result of the individual calibration process. The three models each have an albedo of 0.40 and a $z_0$ of 0.043, so the only way for that site to compensate for the lack of a turbulent heat flux was to increase its thermal conductivity, i.e., from 1.29, 1.35, and 1.31 W $m^{-1}$ $K^{-1}$ for the $LE_{Rain}$, $LE_{RH100}$, and $LE_{Dry}$ models, respectively. The variations in thermal conductivity cause large differences in melting. Other sites adjusted the albedo and $z_0$ to compensate for the lack of a latent heat flux terms, which is why the differences in the melt for Site 6 were so much greater compared to those differences at other sites. It is important to note that the values associated with the maximum $z_0$ changed as a result of re-doing the $z_0$ analysis, so the melt rates at Site 6 also changed accordingly. The differences at Site 6 are no longer as drastic.

$LE_{Zero}$ was a typo and has been changed to $LE_{Dry}$ throughout the text.

**References**


Hambrey, M.J., Quincey, D.J., Glasser, N.F., Reynolds, J.M., Richardson, S.J., and Clemmens, S.: Sedimentological, geomorphological and dynamic context of debris-mantled glaciers, Mount Everest (Sagarmatha) region, Nepal, Quaternary Sci Rev, 27, 2361-2389, 2008.

Robertson, E.C.: Thermal Properties of Rocks. US Geological Survey (No. 88-441), 1988.