Response to reviewer 1:

We would like to thank Peter Moore for his careful review and helpful comments. We hope we have addressed them adequately and as a result created a more impactful version of our manuscript.

Our responses are given in blue below and highlighted in the revised manuscript using track changes.

Specific comments:

These comments are broadly about improving the structure of the article in order to make it more impactful. This falls under two main categories:

1) The need for further explanation of the relevance of the debris distribution examples shown to allow the reader to ore clearly see the significance of these
2) Make some more generalized and hypotheses regarding the controls on likely debris thickness frequency distributions
3) Provide more concrete hypotheses regarding how ablation calculated using a given frequency distribution of debris thickness might be expected to differ from that calculated using mean debris thickness.

We address these as follows:

(a) Introducing some more speculative discussion of expected debris thickness distributions in the introductory section

"The limited available data shows the probability density functions or frequency distribution of debris thickness at a glacier or local scale to show varying degrees of kurtosis and typically a positive skew (e.g. Reid et al., 2012; Nicholson and Benn 2012), but the degree to which the frequency distribution deviates from normal, and the controls on the degree of kurtosis and skewness have not been well investigated. Nevertheless, some postulations can be made based upon the systematic and non-systematic variability components described above. As thick debris cover tends to form where there is little to no ice flux it follows that glaciers close to steady state will tend to be dominated by thin debris, causing the debris thickness frequency distribution to have a positive skew, while this might be expected to be less pronounced in sluggish debris-covered glacier termini, or even have a negatively skewed distribution on stagnant glacier tongues or rock glaciers, where ice flux is minimal. Glaciers with patchy debris at the surface are also more likely to have a positively skewed debris thickness distribution than continuously covered glacier surfaces due to gradual topographic inversion and lateral dispersal of debris from localised surface deposits (Anderson, 2000; Kirkbride and Deline 2013). Gently sloping smooth surfaced debris covered glaciers might be expected to experience less gravitational sliding than steeper or more chaotic glacier surfaces, and less gravitational reworking may favour relatively higher kurtosis than at sites where sliding and slope failures are common, and the frequency distribution of debris thickness can be rapidly reworked and potentially even develop multimodal distributions with many areas of thin, recently destabilized debris and also many areas of thick debris where material from slope failures has accumulated."

(b) More clearly describing the link between the debris thickness distributions and the Østrem curve

"Sub-debris ice ablation calculations are commonly performed using the mean debris thickness over a portion of the glacier surface derived, for example, from satellite thermal imagery (e.g. Fyffe et al., 2014) yet given a skewed local debris distribution, in conjunction with the asymptotic decline in ablation rate with increasing debris thickness (Fig. 1), calculations of sub-debris ice ablation rate and meltwater production using spatially-averaged mean debris thickness may differ substantially from the actual meltwater generated from a debris layer of highly variable thickness within the same area. Reid and others (2012) offered a first consideration of this effect when they applied a distributed glacier ablation model by assigning debris thicknesses to debris covered glacier pixels by random sampling of a probability distribution based on a set of high resolution field measurements. However, as yet no modelling study has explored in detail the interplay between the local debris thickness variability and the local Østrem curve, in terms of its net effect on calculated sub-debris ablation. Given the paucity of data on local debris thickness variability there remains a critical need to quantify not only mean supraglacial debris thickness, but also local debris thickness variability, and assess its impact on ablation rate in order to understand how debris cover is likely to impact glacier behaviour, meltwater production and contribution to local hydrological resources and global sea level rise."

(c) We considered adding some more generalized ablation modelling in which we fit curves to the available measured debris thickness frequency distributions shown in Figure 5, and make a first attempt to characterize the impact of various 'classes' of debris thickness frequency distributions on sub-debris ablation for the climate case we use here (August Himalayan meteorological forcing, and rock/debris properties representative for the Ngozumpa glacier). However as our cases are limited and may also contain some sampling bias, we prefer to focus on what these specific cases say. We have re-written parts of the results/discussion to tray and draw out more concrete conclusions based on these specific cases however:

"Clearly, while debris thickness shows small-scale variability in all cases on the Ngozumpa glacier, the details of that variability differ from site to site. This pattern of change agrees with the tentative hypotheses proposed in the introduction, whereby the downglacier progression of greater debris cover maturity, increasingly stagnant ice and increasing activity of gravitational reworking on the hummocks terrain studded with ice cliffs and ponds all serve to gradually reduce the skew and kurtosis of the debris thickness distribution.

This pattern is supported by data from other glaciers (Table 2; Fig. 5). The medial moraine on Haut Glacier d'Arolla emerged during glacial recession in the second half of the 20th century (Reid et al., 2012), offering an example of a recently developed debris cover. The debris-covered part of Suldenferner developed its continuous debris cover since the beginning of the 19th century, when the glacier was mapped with debris cover below ~2500 m and only surficial medial moraine bands extending up to 2700 m (Finsterwalder and Lagally, 1913). The Nepalese glaciers are thought to have been debris-covered for longer (Rowan, 2016), although it remains unclear when their debris covers first developed.

The Lirung glacier measurements appear broadly more similar to sites further downglacier on the Ngozumpa glacier. Debris thickness at the Lirung glacier, central Nepal, which like the lower Ngozumpa glacier supports a thick debris cover overlying stagnant ice shows a bimodal distribution not replicated at the other sites, but partially seen in the Ngozumpa Margin site (Fig. 5a). At Lirung, this is suspected to be at least partly due to sampling bias, as the measurements were made to test the GPR method rather than to characterize typical debris thickness at this glacier. However, the hummocky terrain of Lirung glacier (cf. Fig. 2b), dissected with ponds and ice faces, is likely to facilitate widespread debris slope failure, which would more readily cause multimodal distributions of debris thickness. In contrast, debris thickness variability at the Alpine sites shown here is more comparable to that of the upper Ngozumpa, The less mature debris cover on Suldenferner, in the Italian Alps, is generally thinner and the terrain is less hummocky, with relief primarily associated with incision by supraglacial streams Debris thickness measured across the whole debris-covered area by excavation, and along cross- and down-glacier transects by GPR, shows a substantially thinner mean than the Himalayan cases, with greater kurtosis. The GPR lines sampled at Suldenferner crossed thick medial moraines and this sampling bias may explain the distribution being less skewed that that determined from the excavations covering the whole debris covered area. This highlights a further problem in sampling strategy for meaningful determinations of debris thickness variability at a local and glacier scale, as the locally less skewed distributions are presumably applicable only to sections of the glacier surface containing these medial moraines, while the debris covered ablation area as a whole shows a more skewed distribution of debris thickness. The debris cover on the medial moraine of Haut Glacier d'Arolla in the Swiss Alps is even thinner with yet more pronounced skewness and kurtosis. This is inkeeping with its younger age and what might be expected from primary dispersal from the meltout of a localised moraine deposit."

We also added in the results of the ablation modelling:

"Coupled with the previous interpretations of how the skewness of debris thickness distribution relates to the relative maturity of the debris cover, this implies that the difference between sub-debris ablation calculated with a mean debris thickness of the thickness distribution will be greatest for recently developed or emerging debris cover."

Technical corrections:

L77. Add space between "has" and "been".

Done.

L202-203. "Ablation rate and surface temperature [delete 'is'] calculated for…"

Done.

L287. It is not clear what the "recurrence rates" refer to. Is this the repeated appearance of supraglacial ponds in a particular area? If so, is it distinct pond bodies, draining/refilling of unchanging basins, duration of ponding, or something else?

Changed 'recurrence rates are generally high' to 'seasonal ponds commonly reform at the same sites'.

L341. There seems to be a word like "of" missing between "kurtosis" and "debris".

Added 'of'.

L364.Change "order or the effect" to "order of the effect".

Done.

L367-368. One could argue that since the use of mean debris thickness seems to consistently underestimate composite ablation rates, it is not worthless but can still have value as a lower bound. Going one step further, even the debris thickness distribution derived from higher-spatial resolution measurements could include some spatial averaging, so at what point are we looking at a small enough area that refining the resolution even more wouldn't further increase ablation?

We did not really intend to label the mean debris thickness as worthless, but highlight its potential limitations. Therefore we now write: "This suggests that while modelled ablation using local mean debris thickness can provide a lower bound this and other measures of central tendency (tested but not shown here), are likely to be poor inputs for ablation modelling for typical debris cover."

The issue of defining an 'appropriate' resolution of measurements is a good one to raise and remains a little problematic. We now say that: " … sufficient data points of debris thickness to capture the local variability are likely to give a more reliable ablation estimate from model simulations." Although this only partially addresses the problem.

L499 and L517. A nitpicky stylistic thing, but I dislike the word "shallow" used as the opposite of "steep". I suggest "gentle" or "gradual" instead.

Changed 'shallow/er' to 'gentle/more gentle' throughout.

Section 5.4. I think the gravitational stability modeling is a reasonable piece to include in the analysis, but it would be prudent to present some assessment of the sensitivity of the model results presented (i.e., areal extent of predicted instability) to unknown values introduced to or inferred from the model, like the ice-debris friction coefficient or debris hydraulic conductivity. These could significantly change the results.

Thanks for suggesting the sensitivity assessment it was remiss of us to not include one. We now include a sensitivity test routine in which each of the parameters is perturbed in turn, and we include the description and findings of this in our methods and discussion as follows:

Additional section in Methods:

"In order to assess the robustness of the slope stability model, sensitivity tests were carried out for each study area, in which key variables of the slope stability model (ratio of densities of water to debris; saturated hydraulic conductivity; debris-ice interface friction coefficient; debris thickness and calculated daily melt rate) were perturbed, one at a time, by ± 10 %. The percentage of the study area classified as unstable, as well as percentage change from that study area's areal percentage instability (using the best estimate values given above), was recorded for each perturbation."

Additional section in Results and Discussion:

"Perturbing slope stability model input variables by 10% generally resulted in small changes of up to 1% in areal percentage slope instability, indicating the model is relatively robust. However, adjusting the debris-ice friction coefficient by 10% caused relatively large changes of up to 9%. Increasing melt rate and the density of water to the density of wet debris ratio cause areal percentage slope instability to increase. Increasing hydraulic conductivity, the debris-ice friction coefficient, and debris thickness cause areal percentage slope instability to decrease. It is interesting to note that the upglacier study area is most sensitive to input variable perturbation, presumably because debris thickness and therefore melt rate are greatest in the upglacier study area."

Figure 2's caption and the text on line 109 indicate that there should be a panel (b) for Figure 2, but none appears in the copy of the manuscript I've seen.

Thanks, we now include the photograph for Figure 2b as originally intended.

Finally, we would like to point out three changes that have been made further to those requested in the reviews.

1) While doing the additional sensitivity tests on the slope stability model suggested by reviewer 1, we noticed a coding error causing areal percentage slope tability/instability excluding ponds and ice cliffs to be wrong. We have adjusted the values accordingly in the manuscript and figures. This does not affect the conclusions of the paper, but rather strengthens our argument that relatively large areas of the debris surface are unstable, on the basis that the values that exclude ponds and ice cliffs are now more similar to those that include ponds and ice cliffs.
   This led to a change in the text as follows: "Slope stability modelling suggests that, under mid-August ablation conditions, the percentage of the debris-covered area interpreted as potentially unstable for the three study areas of Ngozumpa Glacier is between 13 and 34% including ponds and ice cliffs, and between 12 and 22% 10 and 32% if ponds and ice cliffs are excluded (Fig. 9)."
2) We also noticed that we had used the incorrect colour map in Figure 9d and this has also been corrected in the revised manuscript.
3) The reference to Del Gobbo (2017) was previously missing from the reference list, but has been added now.