#### **Response to Referee #1:**

We thank Dr. Jonathan Conway for his detailed review that helped us to significantly improve the manuscript. Our responses to his comments are given in bold font, while any quotes from the revised manuscript are copied in italics. Before we list those changes, we summarize the major changes in the revised manuscript that incorporate comments from both referees:

- Restructured (streamlined) the methods and results sections. In particular, all evaluated bulk methods in the study are clustered into two main types: C-methods and methods based on katabatic models. C-methods consist of four subgroups: C\_log, C\_Rib, C\_M-O and C\_SR, whereas the second cluster consists of the C\_kat method and K\_Int method. All the bulk methods depend on mean meteorological variables only (i.e. there is no application of OPEC data in the bulk methods). All other variants of the bulk methods (i.e. those with OPEC-derived variables) are now presented as part of the sensitivity analysis;
- 2) Removed the section on newly derived stability functions (following the comments from Referee #1);
- 3) Improved the discussion on possible spurious self-correlations and removing the results where the self-correlation was present;
- 4) Introduced a proxy variable for the background temperature lapse rate by using the near-surface air temperature observations from the two nearby meteorological stations, at different altitudes and in the glacier vicinity, in addition to the temperature measurements from the glacier station. The usage of these data led us to identify a dependency of parameters in the K\_Int method on the proxy variable (i.e. difference in off-glacier and on-glacier near-surface air temperature). This empirical relationship allowed us to assess the parameters purely from the mean meteorological variables, instead of using the OPEC-derived stability parameter (z/L) as was the case in the initial manuscript;
- 5) Quantified the errors in the modelled sensible heat fluxes resulting from the radiative overheating of the temperature sensor;
- 6) All results (sensible and latent heat fluxes) are now evaluated for the cases with 30-min wind speed exceeding 1 m/s (instead of wind speed exceeding 3 m/s as was the case in the initial manuscript); and
- 7) Revised several figures and added a few that explicitly show dependencies among variables: z/L, Bulk Richardson number, wind speed and air-surface temperature difference.

#### **Responses to Referee #1's comments:**

• Many useful analyses are presented, and the paper should be of interest to researchers modelling the surface energy and mass balance of mountain glaciers. However, several areas need to be addressed if the most important results for this community are to be highlighted. In particular, more focus needs to be made on methods that can be implemented solely with mean wind, temperature and humidity, otherwise the results would be more suited to a meteorological journal addressing the underlying theoretical framework for the schemes. Several of results discussed contain ambiguities and potential spurious self-correlation (such as the relationship between u\* and the newly derived stability functions). These need a much fuller discussion elsewhere, and distract from the main thrust of the paper. That being said, I do think the analysis of turbulence data produces some key results (e.g. that u\* does not relate well to mean wind speed and that z/L is poorly predicted by bulk models), which could be presented more explicitly. This could take the form of a dedicated comparison of z/L with that predicted by M-O, correlation with bulk Richardson stability parameter etc. These results

would better set the context for the performance of each parameterisation. One other key result of the paper needs to be further highlighted - the good performance of the simple katabatic model compared to the bulk model.

We appreciate the constructive criticism aimed at tailoring our presentation of results to the glaciological community readership. We thus revised the manuscript by highlighting more explicitly the key findings related to the bulk methods that can be implemented solely with mean wind, temperature and humidity. Also, we now explicitly present the relation among measured u\*, U, z/L, H and the bulk Richardson stability parameter (see the figures below). As indicated by the referee, we hope that, by explicitly presenting these relations, we improve the interpretation of the bulk methods' performance.



Figure 5: Bias (modeled minus observed values) in  $u_*$  and  $Q_H$  versus the wind speed. Red points show values during the neutral conditions ( $|\frac{z_{v,t}}{L}| < 0.1$ ) only, and blue points indicate values during the prevailing katabatic flow.



Figure 7: Modeled z/L with the fixed-point iterative scheme in the  $\operatorname{L}_{M-O}$  method and the Bulk Richardson number ( $R_{ib}$ ) against the OPEC-derived stability parameter (z/L obs). Dashed black line shows 1:1 line. Also shown is a dependency of the 30-min OPEC-derived stability parameter (z/L obs) on the wind speed ( $U_z$ ) and the near-surface air temperature ( $T_z$ ).

• The large and somewhat overwhelming quantity of analysis makes the manuscript hard to follow at times. Many of the panels in the later figures (Fig. 9, 10 and 12) deserve to be their own figure as they address a distinct point from other panels. As discussed above, I would prefer to see a dedicated examination of the relationships between U, u\*, H, z/L, rather than the dispersed results at present. This would help the reader to evaluate different assumptions

within the theoretical framework, and point more convincing toward new theories. At present, there is a danger that readers are confused by the various ways in which the eddy covariance data are used.

We agree on the overwhelming appearance of the analysis and, following the referee's suggestions, we streamlined the results (e.g. put more focus on explicitly showing the relationships among U, u\*, H, and z/L) and reorganized the paper to focus on what is most important. We also clustered the bulk methods into two main categories (where only mean variables are used in the method), addressing the remaining bulk methods (with OPEC-derived variables) mainly as a part of a sensitivity analysis.

• Most of the analyses are clear, appropriate and well supported by theory. The analyses of the new stability function and Kint, however, introduce another layer of theoretical framework that deserves further analysis and discussion. Some of this is beyond the measurements available to the authors (i.e. profile measurements) and could be beyond the scope of the journal. Several aspects of these are quite uncertain and need more discussion. This includes the derivation of Kmax and Hk – does the variation of one, when the other is held constant, reveal the inadequacy of the method? The dependence shown in Fig. 12 (d) and (e) is between z/L and u\*/U – and as z/L contains u\*, thus there is potential for spurious self-correlation.

We agree that a more detailed analysis regarding the Kint method in assessing the turbulent fluxes is beyond the scope of this study, mainly because of the limited amount of data on this site. We make a more clear distinction in the revised manuscript between the findings fully supported by our data and the findings which are partly supported but insightful enough to trigger further research when/if the data become available. We now removed the analysis with the self-correlation. Instead, in the revised manuscript, we identified a better way to relate the parameters in K\_Int method with the use of near-surface air temperature measurements from the two stations in the glacier's vicinity.

• In regard to the calculation of new stability function – to properly address this requires profile measurements, otherwise it is simply a circular way to use the measured fluxes to correct the bulk method. It appears that as stability increases, momentum flux decreases while H does not – this points to the influence of a wind speed maximum, where u\* goes to zero at close the height of the maximum, which will be lower for lower wind speed (Denby and Greuell, 2000). The authors need to reflect on the sensor heights in regard to flux-profile relationships in presence of katabatic, and how these may be affecting the observed relationships between u\*, z/L and H.

We agree that, if one is to derive new stability functions to be used for glacier studies, more measurements (in particular, profile measurements) are needed. Our initial goal was not to derive the new stability functions, but to compare the performance of the optimized stability function (one that assures the best match between modelled and observed fluxes) with the performance of the stability functions commonly used in the glacier studies. However, considering the length of the manuscript and the fact that more observations are needed to adequately analyze the effect of stability, we now removed this section. Instead, we extended the discussion, as suggested by the referee, on the potential influence of a wind speed maximum on the bulk method performance. *Page 22/23 in the revised manuscript (copied here in the Latex form):* 

While the original C-methods, however, overestimate \$Q\_H\$ during the katabatic conditions, the C-methods with stability corrections and measured \$u\_\*\$ underestimate the fluxes (Table

 $\label{eq:solution} $$ and Fig. \eff fig: scatter plot with sensitivity tests $$). The overestimation of $$Q_H$ during katabatic flows has also been shown in \cite{Denby_Greuell2000} and explained by a failure of M-O theory in the presence of shallow katabatic wind speed maximum. At the wind speed maximum, measured $u_*$ approaches zero, while the C-method assumes constant momentum flux in the surface layer and therefore overestimates $u_*$. The overestimation is less pronounced for $Q_H$ than for $u_*$ because the reduced turbulence at the wind speed maximum leads to an increase in the air-surface temperature difference, and subsequently an increase in the measured $Q_H$. However, when measured $u_*$ is used in the $\mathrm{C_{M-O}}$ or $\mathrm{C_Rib}$ method, assuming that the eddy diffusivity is as effective as eddy viscosity ($Pr$=1), the C-method underestimates $Q_H$ since the air-surface temperature difference alone can not compensate for the effect of reduced momentum flux. To correct for this bias in $Q_H$, $Pr$ would need to decrease, i.e. the C-method would need to account for more effective eddy diffusivity than eddy viscosity at the given height. In the absence of wind profile measurements, we can only assume that these effects take place at our site, but we have no observational evidence for the presence of the wind speed maximum.$ 

• The authors need to be careful that the key results outlined in the abstract and conclusions are explicitly analysed in the paper. At present, there is some support for the alternate parameterisation schemes presented here, but they depend too much on in-situ turbulence measurements to be used widely. These results are still worth presenting, they just need to be more thoroughly analysed (perhaps elsewhere) before definitive statements can be made. It is good to see movements toward developing new turbulent heat flux parameterisations for mountain glacier environments, which is an essential step for the community.

### A valid point raised by the referee. We streamlined the conclusions to reflect more about our own data and what they show.

• Specific Comments (page-line):

1-8: "the bulk method" – please clarify what is meant by this term. In general, the terminology used needs clarification. The term K-approach is not likely to be familiar for most readers of The Cryosphere, and is easily confused with the KInt approach introduced later. Consider using a different term here to distinguish the bulk methods in which K scales with u\*, perhaps "M-O bulk schemes".

# We now clarify what is meant by the bulk methods in the abstract. The term 'K-approach' is removed, and instead we refer to a more commonly known terminology, i.e. bulk methods based on the gradient transport theory or K-theory (Stull, 1988). We also label these methods in the text as C-methods.

• 1-12: "The OPEC-derived 30-min momentum flux is linearly related to the measured wind speed, contrary to the proposed quadratic relation by the commonly used bulk methods." – This result is not shown but rather hinted at (26-4). Needs to be clearly analysed within the paper for this statement to be supported in the abstract.

Actually, the results have been shown (Figure 10c in the original manuscript) where u\* is proportional to  $\sqrt{(Uz)}$ , with a constant of proportionality being a function of the M-O stability parameter. This is equivalent to momentum flux being proportional to Uz, since the u\* is defined as the square root of the momentum flux. Nevertheless, since this finding is not the key finding in the study, we removed it from the abstract.

• 1-15: "In agreement with the katabatic flow model, we show that in a more stable atmosphere the bulk exchange coefficient for momentum is smaller." Again, the relationship is not analyses specifically, so it is hard to see this a key result. Please revise.

# We added more analysis on this issue in the text (dependency of the parameters in the K\_Int method on the proxy variable for the environmental lapse rate). In the abstract, however, this sentence is now removed since it is not one of the key findings.

• 1-16: "The sensible heat flux can be more successfully modeled if the bulk exchange coefficients for momentum and heat are allowed to follow different parametrization schemes, rather than assuming equal schemes as is the case in the common bulk methods." But the data don't seem to show a large improvement for the more complicated schemes when only mean wind speed and temperature are used. These schemes often rely on measurements of z/L so aren't easily transposed in space and time. Please revise.

The conclusions here follow our findings on the Kint method (section 3.3.4 in the original manuscript): "The Kint -approach, however, does not perform as well for the sensible and latent heat fluxes as it does for the friction velocity. This led us to conclude that while the linear-Gaussian parametrization of K(z) works well for the momentum flux, it works poorly for the heat flux, indicating a shortcoming in the commonly used assumption that KM and KH can share the same parametrization." We now improved the analysis on the similarity between eddy viscosity and eddy diffusivity, and provided more discussion on the varying eddy Prandtl number. *Page 25 in the revised manuscript (copied here in the Latex form):* 

As already illustrated in Fig.\ \ref{fig: scatter plots with basic bulk methods}, the \$\mathrm{K\_{Int}}\$ method gives the best estimate of the friction velocity and the poorest estimate of \$Q\_H\$ among all the bulk methods. This performance pattern indicates that \$K\_M\$ and \$K\_H\$ do not necessarily share the same parametrization. Setting Pr=0.75 for katabatic conditions, the \$\mathrm{K\_{Int}}\$ method performance is simulating \$Q\_H\$ improves (Table 3).

• 2-15: References needed here.

#### Done. We added Hock (2005) as an appropriate reference here.

• 2-21: Please consider adding Guo, X., Yang, K., Zhao, L., Yang, W., Li, S., Zhu, M., Yao, T., and Chen, Y.: Critical Evaluation of Scalar Roughness Length Parametrizations Over a Melting Valley Glacier, Boundary-Layer Meteorology, 139, 307-332, 2011.

#### Included. Thank you for the reference.

• 3-10: "valley glaciers" – do you mean mountain glaciers? Also, it could be worth consistently referring to mountain or alpine glaciers if the two are to be treated similarly (see 4-26 & 4-27).

## Thank you for pointing out this inconsistency. Yes, we mean mountain glaciers, and have corrected this term accordingly throughout the manuscript.

• 3-12: z0v is a mathematical variable that relates the flux and the gradient, and, as such, is not always related to the turbulence generated by roughness elements at the surface. Outer-layer turbulence can, for instance, increase the momentum flux in the surface layer, thereby changing z0v, while the surface roughness elements remain constant. Please revise.

We agree, but in our case we refer to a turbulence generated by a friction drag. Although the roughness length for momentum is defined as a mathematical concept, it relates to the physical quantity (roughness of surface elements; Stull, 1988) when the turbulence is generated by a friction drag. We revised the text to avoid any ambiguity.

#### Page 3 in the revised manuscript (copied here in the Latex form):

While this 'effective' roughness length (Braithwaite, 1995) works well as a tuning parameter when the modeled turbulent fluxes are optimized to match the observed ones, it differs from its 'actual' z0v which is dependent only on the geometry and distribution of the roughness elements, assuming turbulence being generated by a friction drag.

• 5-10: Please provide model numbers of the instruments.

#### Added.

• 6-13: Was the sonic temperature corrected for the effects of water vapour?

# Yes. The sonic temperature was corrected for humidity in the EddyPro software, using a correction based on Schotanus et al. (1983), revised in van Dijk et al. (2004) – this point appears in the revised manuscript.

• 7-4: WPL corrects for fluctuations in the water vapour density induced by high-frequency changes air temperature, so is associated with turbulent eddies. Please revise.

#### Revised. Thank you for this correction.

• 7-28: The assumption of a melting surface, even on temperate glaciers, does not always hold during the melt season. It would be better to use the SEB, or an air temperature threshold to screen periods in which the surface is likely to be melting to validate the use of this assumption. One period in September 2012 shows air temperature well below freezing, so would almost definitely have lower surface temperature. Also, periods in late August and early September 2012 have air temperature close to 0°C, so the surface temperature is likely to be less than 0°C during these periods. Please discuss further and consider filtering bulkmethod results based on periods in which high confidence can be had in the melting surface assumption.

We agree that it would be better to measure surface temperature directly rather than using this assumption. Unfortunately, we do not have these measurements, while using the SEB closure to derive surface temperature turned out to be unreliable (mainly because the measured radiative fluxes have large errors). In particular, using the NR-Lite net radiometer sensor to estimate the outgoing longwave radiation from the measured net radiation and measured shortwave incoming and reflected radiation, turned out to be erroneous. In the evaluation of bulk methods, we did exclude the 30-min segments for which air temperature was below 1°C (now we clarify this in the text). Note that we also provide an error analysis in the calculated fluxes from the bulk methods assuming random errors in surface temperature.

#### Page 7 in the revised manuscript (copied here in the Latex form):

In the absence of direct measurements, the surface temperature ( $T_{0}$ ) was assumed to be at melting point ( $0^{A}\circ$ ) and the surface vapor pressure at saturation (6.13 hPa). The assumption of consistent melting is corroborated with the sonic ranger measurements showing persistent surface

lowering throughout the observational period. To assure that the assumption holds we use only the data for which  $T_{z_t} > 1^{(circ)}$ . In general, assuming that  $T_{0}=0^{(circ)}C$  works well on temperate glaciers during a melting season, and is more accurate than estimating the surface temperature from the longwave radiation measurements  $citep{Fairall_etal1998}$  or from a SEB closure  $citep{Hock2005}$ . Nevertheless, when the surface is not consistently melting, SEB closure can give much better results than the assumption of the melting surface (e.g. Conway and Cullen, 2013). Estimating  $T_{0}\$  from our radiation data, proved to be unreliable because of the poor accuracy of NR-Lite net radiometer. As part of our uncertainty analysis, we will quantify errors in our results due to the assumed rather than measured surface conditions.

• 8-1: Assuming a melting surface in preference to outgoing longwave radiation measurements or SEB closure is predicated on there always being ample energy available for melt. Where this is not the case, SEB closure can give much better results than the assumption of the melting surface (e.g. Conway and Cullen, 2013). I agree that when the surface is most definitely melting, then making the assumption of 0°C is a good way to remove uncertainty in the calculation of surface temperature. Please revise this statement.

### The sentence is now revised to include the point that SEB closure can give much better results than the assumption of the melting surface (e.g. Conway and Cullen, 2013).

• 8-25: Please refer to the later choice to extend the z/L range to z/L = 2 (11-13).

#### **Revised accordingly.**

• 10-5: It would be useful to introduce the filters used to select the 30-min periods used to compare fluxes.

#### **Revised accordingly.**

• 10-7: The term K-approach is introduced with no background. Is there a less ambiguous term to use for this family of bulk approaches, given that none of the acronyms include K, and you introduce KInt later as a separate method?

## We now label the common bulk method (Clog, C\_M-O, etc) as C-methods. The term 'K-approach' has been removed.

• 12-7: Why were errors associated with the air temperature not included in the error analysis?

# Our statement was initially incorrect and is now revised to read that the errors in the air-surface temperature differences are used in the error analysis (not just the errors in surface temperature).

• 15-8: Why were low wind speed periods not included in the analysis? It would be more appropriate to only include the stationarity and wind direction filters, as the other filters are specifically designed for retrieving roughness lengths, rather than removing bad flux estimates. I would expect this to change the comparison significantly, especially the inclusion of low wind speed periods.

#### Initially, we included this filter for the consistency to evaluate the bulk methods over the same

filtered data used for the derivation of roughness length. In particular, we aimed to include only the 30-min segments with well-developed turbulence ( $u^* > 0.1$  filter) in the presence of relatively strong shearing (Uz > 3 m/s). In the revised manuscript, we now include the 30-min segments with Uz > 1 m/s (but keeping the  $u^* > 0.1$  filter). Note that the additional set of points did not alter our conclusions.

#### Page 9 in the revised manuscript (copied here in the Latex form):

To assure that the bulk method evaluation is performed on the high-quality measurements, all of the filters above are applied to the OPEC measured  $u^*$ , sensible heat ( $Q_{H}$ ) and latent heat ( $Q_{E}$ ), except the 'neutrality' and 'wind speed' filters. The latter two filters are modified so that that are modified: all runs with  $\frac{1}{L} < 2$  are included in the calculation of fluxes, as well as all runs with  $U_z$  is no stable stratification are commonly defined up to  $\frac{1}{L} = 2$  ( $L_{E}$ ), which represents strongly stratified stable regime.

• 23-16: The static stability referred to in the OG model is that of the background airmass. As we have no information on the background stability we don't know if this is necessarily reflected in the stability of the surface layer (z/L). Please revise.

A good point and statement now corrected. We further elaborate on the issue in the results section on the K\_Int method. Note that we altered our analysis of K\_Int dependence on static stability by incorporating the near-surface air temperature measurements from the two stations in the glacier vicinity (AWS\_up and AWS\_low). This new approach now made Kint method independent of the OPEC-derived z/L.

• 26-14: Is there a way to evaluate Kint/Kmax without measured z/L? For consistency, it would be useful to discuss if this is possible.

Please see the response above. We have modified the way we evaluate Kint/Kmax in the revised version. Instead of using measured z/L, we evaluate K\_max by its dependency on a proxy variable for background temperature lapse rate, which in our case is a difference between on-glacier and off-glacier near-surface air temperature. This proxy variable is empirically related to the strength of the katabatic flow, therefore, serving as a robust constraint for the K\_max parameter.

#### Page 14 in the revised manuscript (copied here in the Latex form):

To adequately determine  $K_{max}\$  and  $H_K$ , detailed observations of the wind speed and temperature profiles with height are required. In the absence of these observations, and similarly to  $C_kat\$  method, we derive the parameter values through the optimization method that incorporates dependency of the two parameters on the proxy variable for  $\sigma$  amma(i.e. the difference between on $glacier and off-glacier temperature). Since these two parameters (<math>K_{max}\$  and  $H_K$ ) are interrelated  $citep{Parmhed_etal2004}$ , it is sufficient to optimize one parameter while keeping the other constant. The assumed constant value for  $\operatorname{optimize} (K_{max})\$  on an Icelandic glacier  $citep{Parmhed_etal2004}$  as this was the only study we found that directly measured the two parameters. Keeping one parameter constant, the other is estimated from Eq.  $eqref{eq:u star}$ related to K Intusing 30-min OPEC-derived  $u_*\$  and log mean value of OPEC-derived  $z_{0v}\$ (from neutral conditions only). In this way we derive optimized  $K_{max}\$  (or  $H_K$ ) value for each 30-min segment. The final step is to regress the 30-min optimized parameter values against the proxy variable. The regression is performed on the 2012 data with identified katabatic flow conditions only, while the 2010 data is used to validate the regression model. • 27-9: Why are those with measured z/L but not those with measured u\* included? I think it would be better to only present schemes that do not use any time-varying information from the OPEC system, as these are the parameterisations that are of use to those wishing to use the bulk method.

# The methodology has now been modified, so that it addresses the bulk methods that rely on the parametrizations with the mean variables only (i.e. from AWS measurements, and not from OPEC sensor). The OPEC-derived z/L and u\* are only applied in the sensitivity analysis in order to identify potential causes of discrepancies in the bulk method performance. *Page 9 in the revised manuscript (copied here in the Latex form):*

In the gradient-flux relation, the eddy viscosity is parameterized as a function of  $z^{, u_*} and M-O$ stability parameter ( $\frac{1}{2}L^{, u_*}$ ). Because  $u_*$  and  $\frac{1}{2}L^{, u_*}$ , in the *C*-methods, are modeled rather than directly measured, any error in these modeled values can propagate into the flux estimates. Our goal in this section is to investigate the influence of the two variables,  $u_*$  and  $\frac{1}{2}v_*$ *L*, on the bulk method performance. To do so, we estimate the turbulent fluxes from each of the four bulk schemes using the OPEC-derived  $u_*$  and Obukhov length ( $L^{, u_*}$ ).

• 29-15: It is unclear which filtered periods were used here. Please clarify.

# We now clearly state which filters are applied on the fluxes used in the inter-comparison. This explanation is given, as suggested by the earlier referee's comment, already in the methods sections so there is no need to reiterate it again in the discussion.

Page 9 in the revised manuscript (copied here in the Latex form):

To assure that the bulk method evaluation is performed on the high-quality measurements, all of the filters above are applied to the OPEC measured  $u_*$ , sensible heat ( $Q_{H}$ ) and latent heat ( $Q_{E}$ ), except the 'neutrality' and 'wind speed' filters. The latter two filters are modified so that all runs with  $||frac{z_{v,t}}{L}| < 2$  are included in the calculation of fluxes, as well as all runs with  $U_z$  is chosen because the universal stability functions for stable stratification are commonly defined up to ||c|| < 2 citep{Foken2008}, which represents strongly stratified stable regime.

• 29-27: There is a need to discuss potential systematic biases in both surface and air temperature and how these could propagate into the calculation of H. This could include additional screen using the SEB to identify melting periods only to compare to OPEC fluxes, and/or a comparison of sonic-temperature with the unventilated air temperature measurements or application of corrections for low wind speed (Huwald et al. 2009).

As mentioned earlier, we exclude the 30-min segments that can potentially have surface temperature below 0°C (using the filter for near-surface air temperature to exceed 1°C). In addition to the assessment of errors in the modelled fluxes originating in the random errors in the surface temperature, we now also assess the errors due to the radiative overheating of the unventilated temperature sensor. To do so, we compare sonic-temperature measurements with the measurements from the naturally ventilated sensor, and derive a bias-correction function for temperature measurements.

#### Page 14 in the revised manuscript (Section 2.5 Uncertainty analysis):

We quantify uncertainties in the modeled 30-min fluxes due to (1) the choice of roughness lengths, (2) the assumption that the surface temperature is at a melting point, and (3) the systematic error in air temperature due to radiative heating of the temperature sensor.

• 31-7: "fails to successfully simulate QH." I would be careful making this statement, as you could argue that it does simulate H fairly well, not just as well as some other, more site-specific schemes. The main failing of the bulk schemes presented here, is the failure to model u\*. Please revise.

#### **Revised accordingly.**

• 31-8: "Note that the new stability correction acts in the opposite direction than those commonly used for glacier studies: in our case, the modeled QH needs to increase, rather than be suppressed, as the stability increases." Yes, but only because u\* is overestimated. H still decreases relative to stable conditions. Thus, the result is not so much about the effect of atmospheric stability, but the trouble with specifying turbulence in the presence of katabatic flow. Please revise.

#### **Revised accordingly.**

• 33-4: This paper is not available to the reader at the present time, so it is hard to assess this statement.

#### This paper is now published and we cite it accordingly.

• 33-13 to 16: This is a key result and should feature more highly in the manuscript.

#### We agree that this is one of the key findings and now highlight it more clearly.

• 33-22: "Applying the KInt approach to assess u\*, which is then used in the K-approach with the newM-O stability function to assess QH gives the best performance across all the bulk methods we tested". Yes, but the fit between u\*, Uz and Kint is informed by measurements of z/L (which contains H) as is the fit between u\*, z/L and H in the stability function, so it is not surprising that this function works the best. Please discuss the self-correlation and revise.

# This part of the methodology is now modified (see our responses about the K\_Int method and its relation to the proxy variables for the background temperature lapse rate). In the revised version, the parameters in K\_int method are calibrated independently of measured z/L (Page 14 in the revised manuscript).

• Figures 7 and 11: Consider including the numbers assigned to each parameterisation above the columns of each figure panel to aid the reader.

#### **Revised accordingly.**

• Figure 12: The order of parameterisations in panel (a) needs to be consistent with Table 3 – i.e. the first parameterisations introduced at the top. As with Fig 7, needs to have the numbers assigned to the schemes next to the y-axis labels.

#### **Revised accordingly.**

• Table 1: Please include the units for the roughness lengths in here.

#### Done as suggested.

• Table 3: Consider removing parameterisations 17, 19, 21 and 23 as they are essentially duplicates of 16, 18, 20 and 22. Also consider adding lines between the sub-sets of parameterisations.

## We revised the table according to the new clusters of bulk methods and have excluded the duplicates.

We sincerely thank Referee #1 for all the editorial comments. All of them have been incorporated in the revised manuscript.