Referee review for manuscript tc-2016-93-manuscript-version4, under review for The Cryosphere.

"Surface-layer turbulence, energy-balance and links to atmospheric circulations over a mountain glacier in the French Alps" by Maxime Litt, Jean-Emmanuel Sicart, Delphine Six, Patrick Wagnon, and Warren D. Helgason.

General Comments

The revised manuscript is improved in many respects from the initial submission. The removal of some analyses and inclusion of others has improved the flow of the results. Many points have been clarified and issues addressed.

The authors have done well to refine the results and discussion within a more cohesive conceptual framework. However, more care needs to be taken in the discussion and conclusions to accurately reflect on whether the statements being made are well supported by data, or more speculative in nature. In their present form, these sections contain too much speculation, which undermine other more robust arguments that are made.

We wish to sincerely thank the reviewer to have provided such a detailed review. We propose herein a revised version which takes into account all the new comments and provides more explanations when necessary. We changed considerably some parts of the selection procedures and some calculations, and we set-up a new analysis (including a Monte-Carlo approach for the turbulent fluxes, as suggested). In this final version we had to modify profoundly some parts of text, including the abstract and conclusion to correctly reflect the results of our modified and/or new analysis.

The use of both weather patterns and TKE to categorize turbulence data is novel, but at times this analysis becomes ambiguous. At the moment the two categories are used somewhat interchangeably, which is not strictly correct and leads to ambiguity. Further work is needed to clarify the use of TKE categories and weather pattern categories.

We agree that this point was either not clear or led to ambiguity. The main underlying reason for the use of two different categories is that the weather pattern classification works fine at defining wind regimes (katabatic/anabatic), but only partly, at sorting high/low TKE runs: We formerly introduced the high/low TKE classes since we observed, for weak weather patterns, more frequent low TKE cases. For strong weather patterns high TKE cases were the more frequent. We thought that characterizing the flow directly in terms of TKE was more reliable than only on weather patterns.

According to the reviewer's comment, we modified the classification scheme. We now first classify hourly data with regard to weak/strong weather pattern and study the general characteristics of these two subsets in terms of local wind regimes: strong or weak winds, presence of katabatic winds, etc… (Mostly using 2009 data) Then, in order to study turbulence characteristics (only for the 2006 campaign during which we had EC data), we refine the selection procedure: we analyze only the weak forcing cases for which TKE<1 m^2s^{-2}, H>5 Wm^{-2}, wind-speed>2 ms^{-1} , and for the strong forcing subset, we sub-select only the cases for which TKE>2 , H>5 Wm^{-2}, wind-speed>2 m s-1. According to one of the reviewer's comments, we identified discrepancies in the RH values from the different instruments, and our selection procedure includes now an RH criteria. This is new selection process is described in section 4.2.1.

These data are used for the updated Figures 7 and 8 (in the manuscript, Figure 6 and 7 herein).

Alongside this, there seems to be some discrepancies between the periods used to select data in some results (e.g. Table 3 vs Figure 8).

The periods used to select data were the same in Table 3 and Figure 8. The reason why the displayed values are slightly different was that in figure 8, the full boxes presented the total melt as calculated with the surface energy balance (SEB, equ. 8). The shaded part of the box was the mean sum of the radiation components. The remaining part was considered to be the contribution of the turbulent fluxes, which is not strictly correct since the overall melt calculated from the SEB is considered equal to 0 when SEB is negative. Thus, the white boxes heights actually represented ($SEB|_{seb>0\&Ts=0}$) - mean(SWin-SWout+LWin-LWout), which is not equal to the mean sum of H and LE presented in Table 3.

Figure 8 has been replaced by two figures (8 and 9), the first one presenting the mean hourly H+LE for the new classes (our new classification that considers strong forcing with high TKE and weak forcing with low TKE cases) and calculated with the different methods, the second presenting the melt cumulated over all the periods when continuous measurements of EC and bulk data were available. See the answer to the specific comments below.

The addition of further analyses comparing turbulent heat fluxes measured by eddy-covariance and modelled using the bulk aerodynamic method (Figure 8) is useful. However, these new analyses have also highlighted the limitations in the datasets available, namely:

- Large uncertainties in measured melt that make it hard to show a significant improvement of one method or another when calculating melt using the SEB method. Other avenues need to be explored to illustrate where better flux calculations matter.

  We present now in Figure 9 melt calculated with the SEB method on periods for which there is no data interruption, and we compare that to the integrated melt observed over the related period, with errors derived from the Monte-Carlo approach. The presented data suggests that discrepancies appear in the calculations whether profile-derived roughnesses or effective roughnesses are used but an unambiguous statement cannot be made. We agree that in terms of total melt, the use of either method for turbulent fluxes or the other doesn't provide significant changes, but we do show that the total turbulent fluxes significantly change (Figure 8). This shows that in the case of high altitude or very dry sites, where the contribution of the turbulent fluxes to the overall energy balance can be more significant (include references here), the choice of the method might be important. We changed the abstract, introduction and conclusion to better reflect these results.

  Regarding uncertainties on measured melt, we provided extreme error estimates, but our surface energy balance calculations actually do well at representing the melt as derived from the sonic height ranger, as shown in the Figure 9.

- A large scatter in the BA fluxes compared to the EC fluxes, particularly for latent heat: individual points that are divergent in sign (-100 vs +50 W m^{-2}) point to deficiencies in measured gradients of air temperature or humidity. This deserves further scrutiny as it severely limits the confidence in the latent and net turbulent heat flux data presented.

  We agree that the scatter is large between the measured sonic fluxes and the evaluated profile fluxes. First of all, in the former figure 7 there was an issue in our data treatment - we were not comparing datasets at the same time – this was corrected. See the updated figure 7 attached, where we selected the data with as described above. Scatter remains, and we identify 2 main reasons for that:

  1) Thanks to the comment by the reviewer we checked the humidity measurements. This do highlights issues: The comparison of the humidity measured by the AWS sensor and that obtained from the LICOR shows large dispersion, growing with humidity. When both instruments provided difference in RH larger than 10%, we removed the concerned runs.
  2) Overestimation of the temperature by the AWS-G due to solar radiation contamination (Huwald et al., 2009), even though the shield was mechanically aspirated (also seen in Sicart et al., 2014).

  Filtering for RH<60 % filters for outliers. Some scatter remains though, probably due to the second point above, but statistically, the tendency is shown.

  The lack of concurrent flux and profile measurements limit the conclusions that can be drawn around the flux-profile relationship. Because the profile derived roughness lengths were calculated during a different season to the EC flux measurements, no unambiguous statements can be made about the performance of the BA method using these profile data.

  We agree with the reviewer that in the case of different seasons with different surface states one cannot transfer roughness values from one period to another. We highlighted this in the text and have mitigated our conclusions according to this comment. Though, in our case, they are various arguments that justify our approach. We provide these in the answer to the specific comment below, which relates to the same issue.

  The authors do not always seem aware of these limitations and make many inferences about the flux – profile relationship that are not supported. For instance – they interpret the underestimation of sensible heat fluxes by the bulk method in 2006 (using parameters determined in 2009) as confirmation that bulk method does not resolve a low-frequency contribution to the sensible heat flux, yet have no profile data to show this.

  We agree, this might be a bit speculative. Though, the hypothesis is solid and since it is supported by other studies in stable layers. It has been shown , not only on glaciers (Andreas, 1987, Smeets et al., 1998, 1999) that when outer layer large scale turbulent eddies interact with the surface turbulent flow, the turbulence no longer scales with surface parameters (roughness, temperature) since turbulent transport is non-negligible. The presence of outer-layer interactions is supported by the shape of our spectra which compares, at low-frequency to results obtained in similar conditions in other studies (Hogstrom 2002, Hojstrup 1987, Smeets et al, 1998…). We introduced more clearly these concept in the script and method, to better illustrate our discourse, and we mitigate the discussion and our conclusions to take the reviewer comment in consideration.

It is well established that the roughness length for momentum can change by several orders of magnitude over the course of a season on a single glacier, so it may be equally likely that a profile determined roughness length was simply different in 2006 versus 2009, especially in the context of a large (4 orders of magnitude) scatter in the measured roughness length.

We agree with the reviewer on this point, but observations during the campaigns show that the surface was quite similar and that its evolution during each campaign cannot lead to such drastic changes in the roughness parameters. We provide herein the results of our roughness lengths (Figure 9) derivation from the profiles, and pictures (Figures 1-8) of the surface during each campaign. We clarify our discourse by defining clearly the effective roughness (z_{eff}) concept (new Equation 8).

We set z_{eff} such as $\ln(z/z_0)\ln(z/z_{0t}) = \ln(z/z_{eff})\ln(z/z_{eff})$, z = measurement height.

So we can calculate the first part of this equation using the profile derived z_0 and z_{0t}, and find an effective roughness and compare it to the z_{eff} set in order for the bulk fluxes to match the eddy fluxes. Such **zeff** in our case is 0.001 m. Our measurements with profiles in 2009 show a predominant value of and z_0=0.001 and z_{0t}=0.00001 (even for the most chaotic surfaces), this leads to a z_{eff} of about 0.0001 considering 2m measurements. The surface was quite similar (see Figure 1-4 herein) during both campaigns and it evolves through the campaign, but z_{0t} derived from the profile never goes above 10^{-4} m – we think it doesn't make sense that the actual z_{0t} in 2006 was 2 orders of magnitude higher than in 2009, whereas the surfaces remained similar.

Figure 1: The surface state on installation, 10 July 2006.

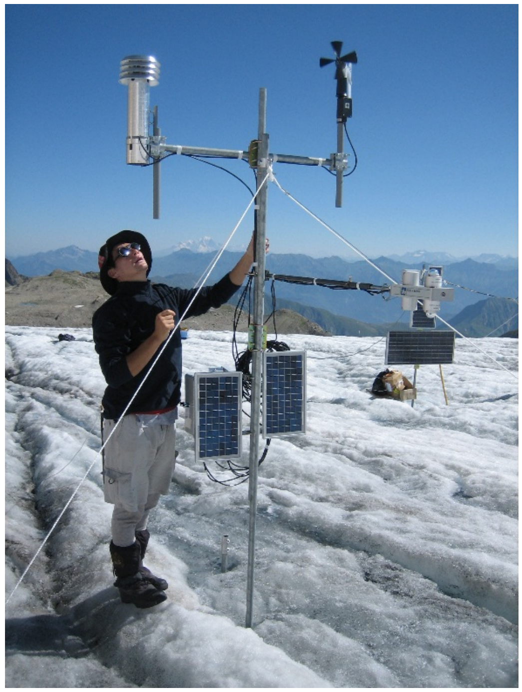

Figure 2: The surface on 19 July 2006.

Figure 3: The surface on 2006, 01 august

Comment from field report, 2006 12 August:

“Glacier clean and covered with 10 to 15 cm of snow”

Figure 4: The surface on 22 aout 2006

Figure 5: 2009 installation on 12 June z_0 profiles = 0.001 z_{0t} profiles =$10^{-4.5}$

Figure 5: 2009 surface on 21 July z_0 profiles =0.001 z_{0t} profiles =$10^{-4.5}$

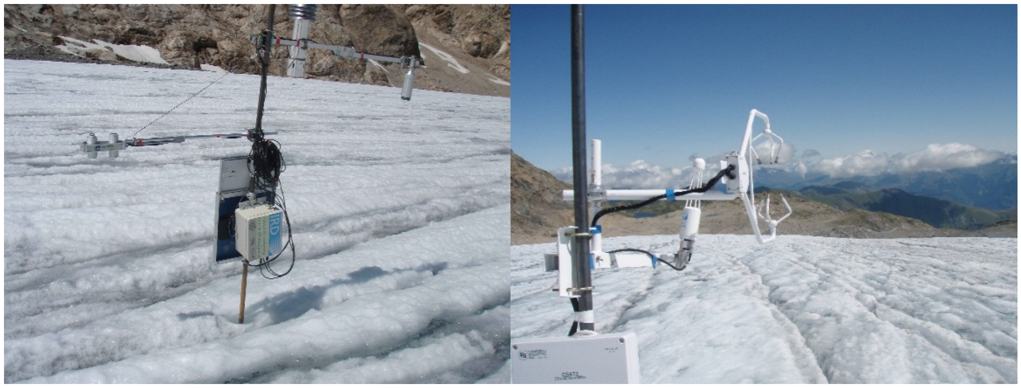

Figure 7: 2009 - 25 july z_0 profiles =10^{-3} z_{0t} profiles = $10^{-4.5}$

Figure 8: 2009 – 4 August $_{z0}$ profiles =10^{-3} z_{0t} profiles = 10^{-4}

We include herein a barplot of the z_0 and z_{0t} derived with the profiles in 2009 (Figure 9). The mentioned dispersion of 4 orders of magnitude was the actual range of remaining z_0 values after selection but the standard deviation $d\ln(z_0)$ is equal to about 2.5. This is the value we use for random error calculations. So the way dispersion on the roughness might impact the fluxes is taken into account in the random error. We updated the discussion according to these findings.

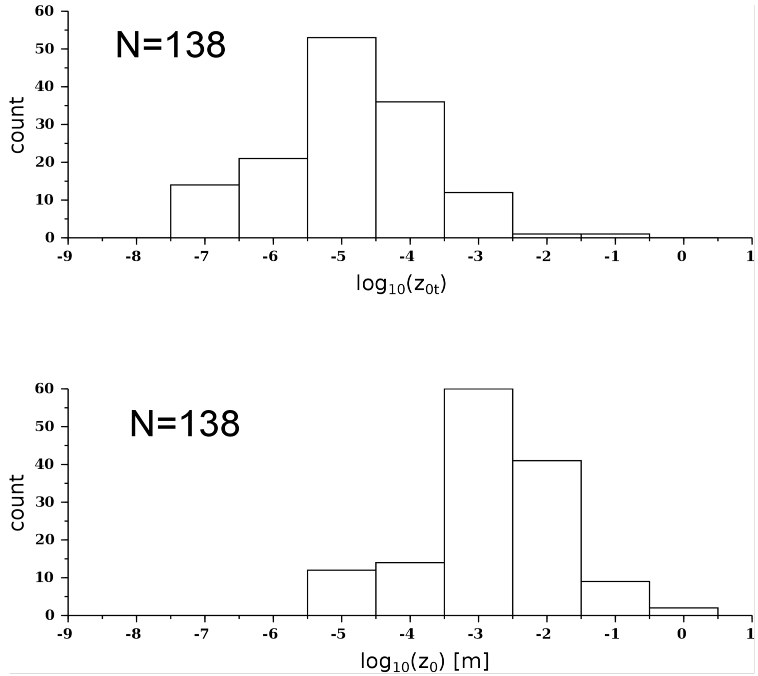

Figure 9: dispersion of the roughness measurements from the profiles in 2009

The introduction of an error analysis is encouraging, but this only takes into account random errors, while systematic errors are also likely and will affect the comparison of methods to a large degree. It is also well known that the BA method is extremely sensitive to the choice of roughness lengths, stability parameterization and surface temperature scheme used, but this is not assessed. A thorough comparison between EC and BA fluxes, taking into account of the full range of systematic and parametric uncertainty (roughness lengths, stability functions, treatment of surface temperature) needs to be made.

The former error study included an error on the emissivity of the surface = 0.01 that led to an uncertainty of 0.35K on the surface temperature (though this was not described). But anyway we now do a Monte-Carlo approach on top of which we add a random error calculation on the final fluxes. We calculate the whole possible combinations of:

1) Two Ts parameterization: blocking at 0 degrees – or one including a correction to include the effects of shortwave radiation on the longwave sensor: we remove 6% of the reflected longwave radiation.
2) (emissivity of the surface) set once at e=0.99 and once at e=0.95
3) z_0 fixed (0.001) but z_{0t}=0.00001 or 0.001 (z_{eff})
4) different stability function parameterizations:
   log-linear (a=5 and a=7), Brutsaert, Holtslag and de Bruin, Beljaars and Holtslag.
5) Inclusion of a SW error calculation scheme (based on the observed difference between the calculated potential radiation for non-tilted and tilted sensor)

On top, the resulting different turbulent fluxes are dispersed with random error calculation like in Litt et al. AMTD 2015. This now provide a broader analysis of the systematic errors, results are shown in figure 8 (new figure) which compares the mean H+LE bulk fluxes with the resulting dispersion with those derived from the Eddy covariance measurements. This is derived for the two subsets described above: weak forcing with low TKE and strong forcing with high TKE, for which we remove the cases for which RH values from the EC and the Vaisala values differ by more than 10%. The actual dispersion is large, larger than the random error on the classical log-linear method. In weak conditions the bulk method as well as the EC method provide similar net fluxes and the dispersion is too large to show any clear difference. In strong forcing, the EC method shows slightly larger fluxes than the pro method, but this could partly be due to some errors. The effective roughness method can explain properly the EC fluxes.

In weak forcing, the minimum values are obtained from log-linear with a=5 stability functions, epsilon = 0.95, and blocking method for T_s. The maximum values are obtained with Beljaars and Holtslag, emissivity=0.99, blocking method for T_s. In strong forcing, the maximum values are found for the Beljaars and Holtslag functions, emissivity=0.99, blocking T_s at zero method. Minimum values are found for emissivity = 0.95, blocking scheme for T_s, and log-linear a=7 when z_{eff} is used and Holtslag and de Bruin when z_{0t} profile is used.

In short, the authors need to present a more careful analysis and discussion for the manuscript to be acceptable. This would address the issues surrounding the limitations of datasets, a more thorough treatment of errors and a consistent and clear treatment of weather pattern or TKE classification.

We agree with the comment, we now present the limitations more clearly into the discussion and conclusions. Also, as mentioned above: We now classify data as weak/strong cases, and for the weak subset, we analyze only the weak forcing cases for which TKE<1, and for the strong subset, we analyse only the strong forcing cases for which TKE>2, so that our analysis is now unambiguous.

Critically, the authors also need to show some example of where the method used to calculate the turbulent fluxes impacts significantly on melt, as this is the fundamental premise of the paper (as outlined in the first two sentences of the abstract).

We do question the fact that choice of methods impact significantly the melt. In our dataset such a period of time where turbulent fluxes calculation method really impacts the ablation is lacking, still, if we take the period of strong forcing cases with high TKE we can see that the z_0 likely fails to capture the total turbulent fluxes (Figure 8). The error analysis on the melt (Figure 9 in the manuscript) shows that this result is not unequivocal and can be discussed, so the abstract-introduction-conclusions have been rewritten accordingly. The results show that situations where turbulent fluxes calculations should be taken care of are those for which very high winds and large sublimation can be observed. Such situations can be found on high altitude low latitude glaciers.

As it stands, Figure 8 does little to confirm to readers that they should consider the effect of deviations of turbulent heat fluxes from theoretical predictions when calculating melt using the SEB method.

True, we updated figure 8 by providing two figures (8 and 9). Figure 8 presents only the net turbulent fluxes with the results of the Monte Carlo approach, while figure 9 presents the melt obtained using only sections of more than 6 hour for which data were continuous, as suggested in the next comment.

Perhaps breaking the analyses further into a daily or multiday periods with the same weather pattern may highlight these effects? Using multi-day periods would reduce uncertainty in the back-calculated melt energy by increasing the absolute magnitude of melting for the same uncertainty in the sonic ranger measurements.

See response to the comment above. We now calculate the melt-fluxes over period for which the data set is continuous. But still this doesn't provide better evidence.

Alternatively, the authors may wish to present an analysis of the utility of the weather patterns for diagnosing temporal variations in the SEB and include the turbulent flux calculations as part of the uncertainty. The paper does present a good summary of how large scale weather patterns are related to surface layer turbulence and corresponding characteristic deviations from MOST. This provides a good base from which to present and discuss the temporal variation of SEB components in more detail. This may prove to be a more useful analysis given the limitations to the datasets noted above.

Apart from the wind conditions, it is hard to find a consistent tendency in the other parameter with the strong/weak forcing classification (see figures 10, 11 and 12 of this document). On the other side, we can show that the participation of the net turbulent fluxes is more important in the case of strong forcing: We compute the RATIO=mean(H+LE)/mean(SWinc-SWout+LWinc-LWout) for weak/strong subset and different methods for H+LE. Results suggest that strong forcing favors the participation of H+LE to the total flux exchange. So if errors affect them this might have some importance in these case. The ratio is much smaller in 2009 weak cases than in 2006 weak cases probably because the LW incoming is much larger than in 2006 for these conditions.

RATIO	z0-zt	ze	ec
2006 strong	0.20	0.27	0.25
2009 strong	0.24	0.34	NA
2006 weak	0.18	0.21	0.20
2009 weak	0.05	0.05	NA

A further general comment is that there is a general inconsistency in the use of abbreviations for various terms that at times makes the arguments hard to follow.

We apologize for the inconvenience, and updated the draft with respect to that comment.

I would recommend the authors carefully assess what can and cannot stated confidently with the datasets available, revise the results accordingly, position the discussion within this context and conduct a more thorough internal review for clarity and consistency (of text and figures) before resubmission.

According to this final comments, we reviewed more carefully the actual conclusions and discuss more clearly the issues to highlight the actual results of our study. The text has been clarified as demanded.

I have structured the detailed comments by first addressing the response of the authors to my earlier revision before making line comments on the revised text and tables in the new manuscript.

We provide below an answer to these comment to comments, in red.

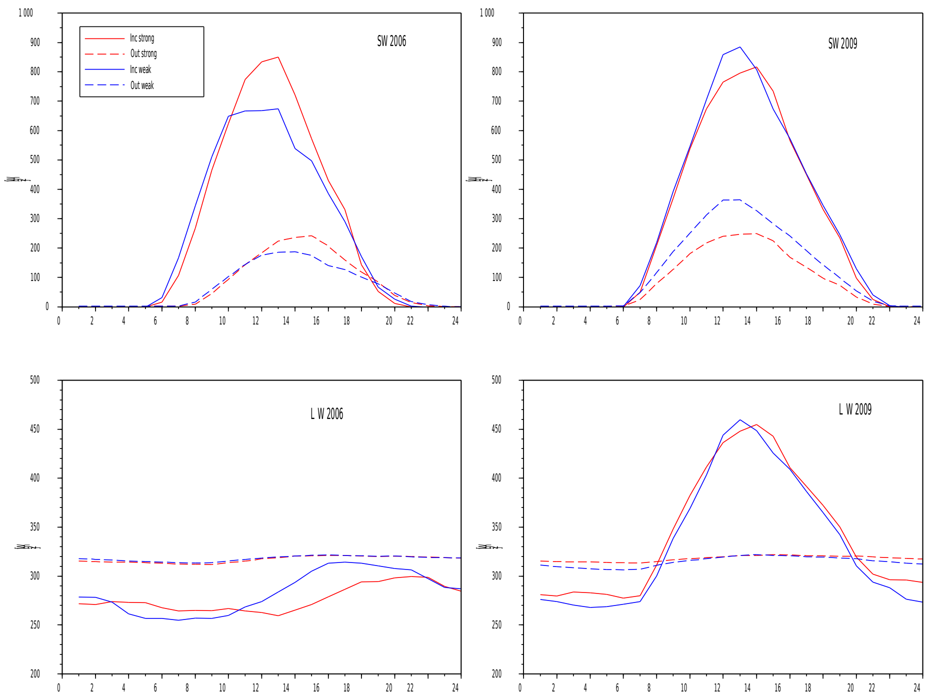

Figure 10: Mean daily evolution of the radiation components for weak and strong forcing conditions

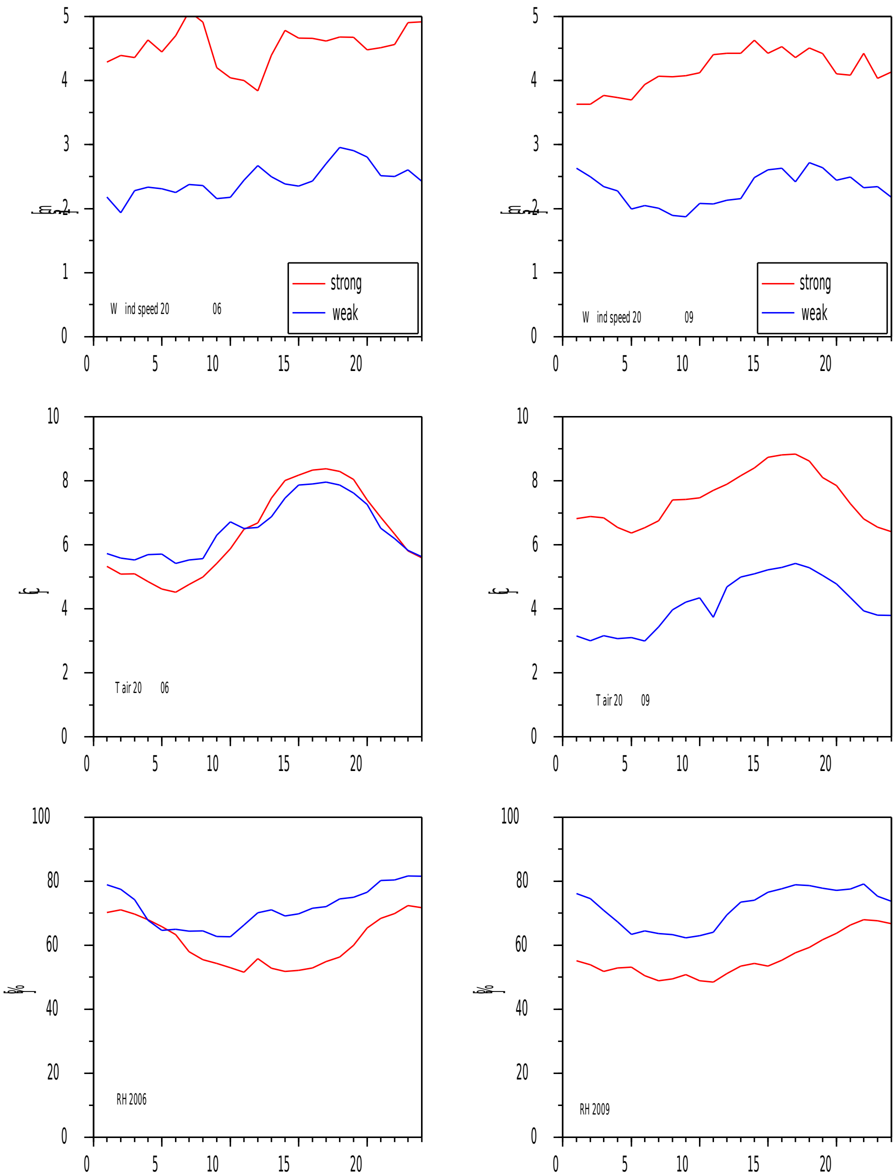

Figure 11: Mean daily variation of windspeed, temperature, and humidity, in 2006 and 2009.

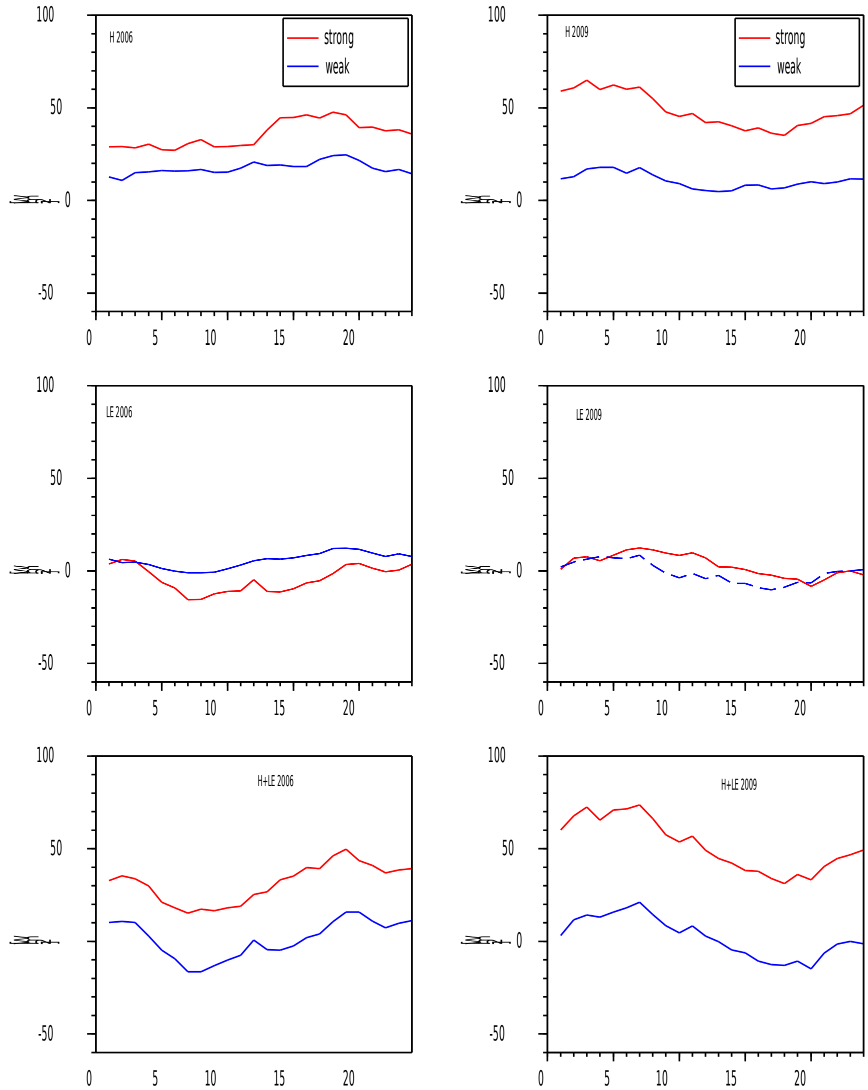

Figure 12: Mean daily variation of turbulent fluxes as measured with the bulk method and z0 derived from the profile method.

Comments on the response of the authors to comments on the first manuscript.

Note that, for brevity, comments are only shown where the initial comments were not fully addressed. Most comments were well addressed. The original comments are shown in bold type, author responses in blue type and the new comments as normal indented type.

The attempt to link these fine scales of turbulence to surface melt, through the calculation of fluxes in using the bulk aerodynamic (BA) method, is ambitious but makes sense in the conceptual framework of the paper. However, the formal links between each scale of analysis are not always well made and this reduces the confidence in the interpretations made.

As discussed above, this remains to be a central issue with the manuscript. Several weaknesses in the comparison of the BA and EC methods and their effect on melt (outlined above and discussed in more detail in comments below) need to be resolved.

Our improved analysis shows that we cannot really conclude that the choice of the sonic or profile method really impacts on the melt calculated – only suggests that errors might exist and should be taken care of on the longer term, and that strong forcing cases might be the cases for which that imports more. We updated our analysis, and the discussion and conclusions.

The large scope also makes the paper somewhat disjointed.

The paper is now less disjointed and provides a more cohesive conceptual framework for the reader.

There are clearly some interesting interactions occurring between the magnitude of net turbulent fluxes and the quantity of melt, which deserve further elucidating. The EC method gives the lowest melt of all methods, yet the has the same or higher net turbulent heat fluxes as the other methods. Why do we see a breakdown of theoretical predictions, yet no impact on modelled melt?

Actually, our results show the net turbulent fluxes are higher in Strong Forcing than in Weak Forcing conditions. Focusing on H (LE is small or erratic due to large random errors) we show the BA method fluxes are lower than the EC method fluxes in Strong Forcing conditions. Considering the EC method is more reliable than the BA method, we consider the BA method underestimates fluxes in this cases. This effect is even more pronounced if you consider a subset of only large TKE: turbulent fluxes H are quite different with the BA and the EC (Table 3, new version). The results suggest that an effective roughness length (ze, introduced in former studies) can be used as a tuning parameter to correct the BA method for this underestimation.

But, since the total melt results from turbulent exchanges and radiative exchanges, we are not sure that we clearly understand this comment. Relating turbulent fluxes to melt cannot be done without considering the whole surface energy balance, and especially the radiative balance which plays a key role on melt. The total melt can be low if the radiative balance is low, even with strong net turbulent fluxes, and vice versa. That explains why we do not find a clear relationship between turbulent fluxes magnitudes and melt. Actually, the radiative balance is controlled by the albedo. In section 5.2, the first sentence states: “During both campaigns, the SEB was mainly controlled by large radiative fluxes, regardless of large-scale forcing, but the contribution of turbulent fluxes to the SEB was significant”.

In section 5.2 which describes Fig. 9 (former version, now Fig. 8), we added: “Yet, changes in net turbulent fluxes resulting from the choice of calculation method remained too small in comparison with the radiative balance to yield significant differences in the SEB estimates.” We hope the issue is clearer now.

Perhaps I could have made myself clearer here, as the authors have not addressed my comment. My reference here was to Figure 9 in the original manuscript which detailed the melt predicted for each SEB method: 2006- strong forcing showed the EC method gave the lowest melt of all methods (which all have the same radiative fluxes). This seemed at odds with EC method having larger fluxes than the BA method. Is this because of large *LE* that cancels out a larger value of *H*, or is this because of how EC fluxes were used to calculate melt?

On closer inspection I now see the magnitude of *H+LE* is different in the figure (20-30W m^{-2}) compared to that predicted for the same period in Table 3 (30-40 W m^{-2}). Please explain clearly how the estimates of SEB in Figure 8 are made. Are the same time periods used? How have poor quality EC data been treated (gap filled? Excluded?).

The same period were used but results from Table 3 cannot be compared directly to the results from the SEB. In Table 3 we only showed mean turbulent fluxes for different periods. The sum of their mean is not equal to the mean of their sum. Furthermore in figure 8 (now 9) the SEB is sometimes found to be slightly negative whereas melt was observed, these values were forced to zero – this explains why the results of Table 3 and Figure 8 are slightly different. Poor quality EC data had been excluded. In the new melt comparison figure (i.e. figure 9) we used the new selection of both weather patterns together with TKE (see comments above).

I also note that in the updated Figure 8, the values of *H+LE* for the EC method and the average melt rate have changed between revisions. Please also explain why these net turbulent fluxes are different in the figure and table and why these have changed between revisions.

Values in the table are identical in between both version. In figure 8 values changed slightly because we wanted to include only full days of data. In first version we had included half days. We didn't change for Table 3.

Regarding the authors response to my comment here that – *"changes in net turbulent fluxes resulting from the choice of calculation method remained too small in comparison with the radiative balance to yield significant differences in the SEB estimates."* As presented this really does limit the usefulness of the analyses presented here to the glacier modelling community. Further analyses describing situations where it matters to get the fluxes right is needed.

Such results smight interest not only the glacier modelling community, but are also important for the atmospheric science community. Also, our results for Saint-Sorlin glacier in 2009 show that if the SF conditions are really frequent, differences between the flux calculation methods can grow, as well as the errors. Note that large sensible heat fluxes can observed on Scandinavian glacier, or very high sublimation at dry, high altitudes sites. In this latter cases better understanding of such situations can also have an impact on the mass balance calculations.

Further description of the processes occurring at the hourly and daily scale are needed to formally link uncertainties in the calculation method for turbulent heat fluxes to melt. Along with this, one of the key points introduced is the errors in the calculation of the turbulent heat flux, yet this is poorly addressed in later analysis. This would put the

paper in a much stronger position to comment on the conditions in which the violation of current turbulent flux modelling methods will have an impact on melt.

Regarding error calculation, we apply here methods to estimate the random errors resulting mainly from instrumental uncertainties that has been developed on Zongo glacier in Litt et al., (2015). We only briefly describe these methods herein, to avoid having a too long paper, but we could add more details if required by the reviewers. We encourage the reader to check the above mentioned paper where all the error calculation processes are described in detail. As presented in section 4.2.3, and discussed in section 5.1, page 18 lines 3-7 (new version), the main result here is that the random errors cannot explain the differences between the EC and the BA methods when TKE is high. We suggest the difference is due to the inability of the BA method to capture part of the flux in non-stationary conditions. We inserted the paragraph about errors into a new subsection (3.3), where we improved the method explanation. We also improved the interpretation of the results, especially for the turbulent fluxes, in relation with the turbulence characteristics, in section 5.1.

> While the manuscript is improved, the it still does not present a thorough comparison between EC and BA fluxes, taking into account of the full range of systematic and parametric uncertainty (roughness lengths, stability functions, treatment of surface temperature), which are known to have a large impact (see for example Giesen et al, 2008).
>
> We thank the reviewer for this useful reference, we now include a Monte Carlo approach for determining the impact of the choice of this different parameters on the overall melt. This is described in the new section 3.3.2.

P3 ln18. Were any corrections for tilting of the radiometer measurements necessary given the high melt rates? If so, please detail the procedure used to correct for tilting of the mast, or if corrections were not performed please comment on the effect of tilt on the radiation values.

Corrections were not performed, since tilting was not automatically measured. We assumed the potential correction on the radiometer remained low, since the sensor was rarely found to be out of alignment during each field visit (every 10 days), and it was levelled every time small deviations were noted.

> Please note in the text that corrections were not made. Also, the effect of even a few degrees' tilt on incoming shortwave radiation can be significant, so the assumption may not be justified. Please estimate the additional uncertainty introduced by the maximum estimated tilt angle and detail in the text.
>
> We now include into our analysis an error on the shortwave radiation. This is calculated by estimating the potential solar radiation (Pellicciotti et al., 2011) for untilted cases and for a maximum tilt of 10 degrees. The error is estimated as the relative difference between the two calculations.

P4 ln5. Please explain why 1-hour runs were chosen over standard 30-minute runs?

The choice was constrained by the parallel use of the off-glacier (on the moraine) meteorological station data for which sampling was set on 1-hour. Calculations of the fluxes have been done with 30 min based runs and did not changed the relative contribution of the fluxes to the SEB.

Please note this in the text.

Ok, this has been added.

P7 ln20. What runs were chosen for the analysis of roughness lengths, and how was stability taken into account? This is especially important given the low wind speed maximum observed. Also, why was the EC data not used to analyse roughness lengths?

The complete procedure, developed and presented in Sicart et al., 2014, is based on an iterative fitting of profiles between wind and temperature as described in Andreas et al., 2002. The selection of runs was based on a set of criteria, including neutrality (based on Richardson-bulk parameter analysis), quality of the fits (R_2>0.975), absence of a katabatic wind-speed maximum, and some others. We agree this is not clear in the text, so we included a statement referring to the Sicart et al., 2014 paper more explicitly: *"The method for roughnesses determination was inspired from Andreas (2002) and developed for the tropical Zongo glacier. It is detailed in Sicart et al. (2014)"*

We had not included the roughness values calculated with the EC system since we thought using only one point above the ground for this calculation was not reliable enough. Anyway, the dynamic roughness length had been evaluated with the EC system, by inverting the log-linear wind speed profile relationship. The values were dispersed but we found a mean value of z_0 around 0.02 m and z_t around

6.6 × 10_{-6} m when derived from the EC system. We now mention that in the new manuscript (page 8, line 17):

"We also evaluated the roughness lengths using the EC system and inverting equations 4 and 5, and selecting neutral runs. The median z_0 was 0.022 m, and the median z_t 6.6 × 10_{-6} m. We did not use these values to calculate fluxes through the BA method."

Interestingly, the factor $(\ln(z/z_0)\ln(z/z_t))^{-1}$ which is the denominator in the bulk formulation of the fluxes (assuming neutrality), calculated with the EC derived roughness lengths or the effective roughness is roughly identical (0.0165 and 0.0167). This shows that effective roughness lengths values can be used to compensate for the BA method underestimation of the fluxes. We now mention this results in the discussion section p18 line 6:

"This is supported by the values of the EC derived roughness lengths, considering they account for the additional mixing: the denominator $(\ln(z/z_0)\ln(z/z_t))_{-1}$ in the bulk formulation of the fluxes, calculated with the EC derived roughness lengths or the effective roughness length is roughly identical (0.0165 and 0.0167)."

The additional data presented here serve to illustrate the shortcomings of the profile technique, rather than the bulk aerodynamic method per se. The deviations of the spectra from classic theoretical predictions indicate that flux-profile relationships established in idealised circumstances are not appropriate in this environment.

That is exactly what we meant. The log-linear profile is valid when turbulence in the surface layer only result from the friction at the surface, and thus the aerodynamic roughness parameter has a physical sense and can be linked to the geometrical roughness, when turbulence does not depend only on interaction with the surface, the roughness parameter and the profile method fail. We added the comment in the text into the method section about turbulence characterization.

Thus, to correctly characterise fluxes, the roughness length should be diagnosed from EC measurements, rather than profile measurements. The authors need to frame their arguments in this light.

We indeed say that the profile fluxes correspond better to the EC fluxes if we take the EC-derived roughnesses to calculate the turbulent fluxes. Calculating the roughnesses from the EC is equivalent to calibrating the bulk aerodynamic method so that it fits to the observed fluxes.

P9 ln10. Please justify the exclusion of heat from precipitation. This is only reasonable if the contribution of this flux can be shown to be negligible.

This is a common assumption over alpine glaciers, with significant mass turnover. Also, heat advected by precipitation is generally negligible because the temperature difference between the rain and the ice is low, the rain intensity is small, and anyway rain is rarely observed. See Paterson (1994), and Oerlemans (2001). We added references in the text.

The exclusion of the heat advected from precipitation may be justified when considering total glacier-wide annual mass balance, but is not correct for shorter time periods, especially summer periods where convective rainfalls can add significantly to melt energy (Neale and Fitzharris, 1997). It is an oft used assumption, but is used for its convenience more than its robustness and should be stated as such. None of the references provide an analysis of the contribution (or lack thereof) of rain heat to the SEB, therefore can only be treated as reflecting the current practice, rather than established reality. Two of the papers (Six et al., 2009 and Oerlemans et al., 2009) simply cite the earlier text books, therefore should be removed. A more appropriate statement would read something like,

"The energy gains from precipitation were excluded from the analysis as they were assumed to contribute negligibly to the surface energy balance (Paterson and Cuffey, 1994; Oerlemans, 2001).

We thank the reviewer for this clarification, we updated the statement as suggested. Note that we anyway remove cases of precipitation from the analysis since they correspond low quality measurements. This has been added in the text.

Figure 9. These error bars seem unrealistically small – especially given the large spread in H from an uncertainty in surface temperature of +/- 1 K. Further justification of these errors is needed. Note that these are only errors on the mean turbulent fluxes. Since they are random errors, on average over all the available runs, the error is quite reduced, whereas it can be quite large for an individual measurement (see new figure 7). The error is calculated on the basis of an error of +/- 0.35 K on Ts (see Litt et al. 2015, and updated error methods section).

The effect of potential systematic biases on the average turbulent heat fluxes needs to be considered here, as they can be substantial and have a large impact on calculated fluxes – these will not cancel out when considering the uncertainty on the mean values. As noted earlier a more thorough treatment of systematic variations in roughness lengths, stability corrections, the treatment of surface temperature is needed to establish a more realistic uncertainty on both the average and hourly turbulent heat fluxes. The use of a Monte Carlo method may be useful way to include this uncertainty.

Our newly introduced Monte-Carlo approach addresses this issue. We now take into account different stability functions, surface temperature parameterizations and different emissivities, and potential errors due to tilting of the radiometer. Note that random errors on roughness lengths takes into account their large dispersion.

Further specific comments on text and tables in the revised manuscript

P10 ln14 - change *"wet"* to "moist"

Done.

P11 ln1 - *"10 June"* do you mean 20 June? The period 10 June to 20 June includes other and strong forcing.

Yes we meant 20 June, but there we are talking about weak forcing, which was observed between 20-25. This has been updated.

P11 ln8 - *"moderate air temperatures (5 C)"* – add "on-glacier"

Ok, this has been added.

P11 ln9 - *"generally cloudy conditions"* – this is not apparent in the figure. Please either clearly show this or modify the statement.

Ok, this has been removed.

P12 ln2 - change *"inexistent"* to "non-existent"

Ok, this has been changed.

P12 ln6 - *"However, some high and low TKE cases remained in the classification (Table 2)".* It is unclear what is being referred to here. Do you mean that after sub setting the 2006 data by TKE, some low (28%) and high (11%) cases remained in the SF and WF classes, respectively? If so, please state this.

Yes. We state that within the weak forcing runs 11% of the runs show TKE>2 and within strong forcing runs 28% of runs show TKE<1, this was changed to be clearer.

P13 ln3 - *"-5 W m^{-2}"* – do you mean +5 W m^{-2}

Yes, this was corrected.

P13 ln10, P14ln1, P15 ln3 – consistency in terms – *"horizontal", "longitudinal, "u".* Please choose one term and stick with it. Similar when introducing spectra and co-spectra (S_u, Co_{uw}).

Ok, we changed it to horizontal. Actually horizontal components refer to u (longitudinal) and v (lateral) components. We thus had to remove some "*lateral*" terms. These were simply noted as horizontal. The components symbols "u" and "v" are now used to differentiate the longitudinal and lateral cases.

P13 ln12 - Need to either explain the dependency on *z/L*, or remove the reference to it here.

Surface layer stability must control the position of the maximum spectra and cospectra by shifting it towards higher frequencies when increasing. We removed the statement to lighten the text.

P15 ln3 – change to *"…. affected the co-spectra of the momentum flux (Co_{uw}) (Figure 6d)"*

We changed to: "*affected the momentum flux as shown by the Co_{uw} cospectra (Figure 6d)"*, we hope the reviewer is ok.

P15 ln17 – change *"low"* to "small"

Done. Note that we now refer to weak forcing cases with low TKE.

P15 ln19 – *"…did not provide significantly different results"* – do you mean just for the average values of *LE* here? The BA_{eff} method gives an average value for *H* that is significantly larger than the EC flux, taking into account the uncertainty. Also the correlation (r) between EC and BA fluxes is different between methods (range 0.31 to 0.67) and this would indicate that there may be statistically significant differences in the fit, not just the mean value, which is relevant here. Please make sure your statements are consistent with the data presented.

We agree with the reviewer. This point as changed considerably, since we modified the selection procedure before studying the fluxes. We invite the reviewer to look at the new figure 8. With the complete Monte-Carlo approach, one cannot differentiate between the mean net fluxes from each method in the weal forcing, low TKE subsets. Regarding the issue with the statistical significance, we agree with the reviewer, but the classification has been changed and the results updated.

P15 ln21 – *"random errors were too low to explain such a discrepancy".* Further elaboration is needed on this point.

We are not sure to understand what the reviewer actually asks through this comment. We actually meant that the random error calculation on the turbulent fluxes could not explain the difference observed between the profile and the sonic method's net turbulent fluxes. Actually this is not still the case with the new classification and with the dispersion resulting from the Monte-Carlo approach (Figure 8 in the manuscript).

P15 ln23 – *"… as result of random errors on both methods…"*. Figure 7 indicates that many points lie well outside the range expected by random errors alone for LE. Please modify the statement and discuss further.

Figure 7 has been removed and this has been changed.

P16 ln1 – *"…better correlation … were found with BA_{eff}."* – The r values in Table 3 are identical for these methods. Please revise.

There was a mistake: referring to SF conditions, we wanted to refer to high TKE conditions. We admit that presentation of data was confusing. This sentence is not anymore relevant in term of the new classification. It was changed according to the new results.

P17 ln9 – please use consistent terms for each co-spectra and Kaimal…

We are not sure to understand the comment. Here we meant cospectra of u **with** w, that was changed. Otherwise all "*Kansas curve*" occurrences were replaced by "*Kaimal curve*".

P17 ln13 – please clarify that the peak in heat flux occurred at n= 10^{-2} in high TKE cases only.

The beginning of this paragraph was slightly changed to answer the query.

P17 ln16 – *"this explains why the BA_{pro} method systematically underestimated…"*. Unfortunately, this result is speculative, as concurrent profile and EC measurements are not available. Please modify the statement.

We updated the discussion in order to reflect this. We know present the statement as one of the possible explanations for the underestimations. We also add as an hypothese that the roughness length were simply different.

P18 ln3 – *"This suggests that the use in SEB models of an effective roughness length, larger than the profile-derived dynamic or thermal roughness lengths, respectively, in order to increase the turbulent fluxes so that the SEB matches the melt, is actually a way to compensate potential biases in the BA turbulent fluxes due to failure of the MOST when TKE is high.* The discussion needs to reflect on the use of an effective roughness length as a common practice.

We introduced a better description of the effective roughness in the methods.

Also I note that the effective roughness length used here (0.001 m) is the same as the dynamic roughness length derived from the profile measurement. Please revise.

Yes, but this effective roughness is not the same as the thermal roughness length derived from the profiles, it is two orders of magnitude larger. In the methods we describe now more clearly the roughness length issues.

P18 ln10 – *"(Fig. 6a)"*

Ok, this has been added.

P18 ln16 – please pose as a hypothesis *"…wind speed maximum would be expected to oscillate too… This would explain…"*

The text has been changed to reflect the fact that these remain hypotheses.

P18 ln20 – *"As a result, fluxes from … were de-correlated"* Further analyses would be needed to show that poor correlation of EC and BA fluxes in low TKE conditions are a result of the hypothesized oscillations in wind speed maximum. The poorer correlation is to be expected given that the smaller absolute values of the fluxes are more effected by measurement uncertainty. Please provide additional analyses, or modify the statement.

We agree with the reviewer comment. The text has been updated.

P18 ln27 – *"…too small to …. yield significant differences in the calculated SEB."* Figure 7 seems to show that the SEB estimates are significantly different from each other, just not significantly different from the melt. Please modify this statement.

New figures (8 and 9), with new classification provides different results. The text here has been significantly modified.

P19 ln6 – *"In these cases low-frequency oscillations…led to an underestimation of the turbulent heat fluxes by the BApro method"* – This statement is extremely speculative. No data are available to show what the actual turbulent heat fluxes were in 2009, while in 2006 no profile data are available to show the mean gradients. This may be occurring, but needs to be phrased as a working hypothesis.

We agree with the comment, the text was significantly modified, and all these statements are now presented as hypothesis.

P20 ln7 – *"we show that the BA method compared differently"* – Please point to the data that indicate

this.

Done

P20 ln 9 – *"for which temperature and cloud cover can be significantly different."* This is an important points and deserves further discussion.

P20 ln10 – *"cloud covered conditions"* – this was not shown in the data (see earlier comment). Please show more clearly the increase in cloud in SF conditions, or revise the statement.

Regarding these two comments, we agree this is important but actually this was not really evident (as shown by Figures 11 and 12 of this document). Our classification provides a distinction between high and low wind speed but not really for the other meteorological variables (different behavior between years). Another WP classification would be needed to more accurately represent the changes in the SEB. We updated the text in this section accordingly.

P21 ln2 – *"The non-equilibrium of the surface layer let to…"* Again, this is speculative and needs to be phrased as hypothesis.

See response to comments above: the text was significantly modified, and all these statements are now presented as hypothesis.

P21 ln3 – *"both methods"* please clarify which methods in the text.

Done

P21 ln4 – *"A systematic underestimation… in magnitude than the EC fluxes"* Again this is speculation as the data cannot show that the differences were not simply due to a systematic changes or errors in the dynamic or temperature roughness lengths, stability correction functions, or surface temperature. Please present these as hypothesis, or provide further analysis to support the statements.

This statement is actually reinforced with the data presented in the new figure 8. We kept it, but the discussion and conclusion are now more prudent on the explanations of this behavior.

P21 ln8 – *"During the 2006 campaign, using turbulent fluxes from the BA or from the EC method did not provide significantly different results…. Hence, the turbulent fluxes from both methods were small and not significantly different on average.""* – This statement is confusing as the fluxes from BA_{eff} and BA_{pro} look to be significantly different to each other (Table 3 and Figure 8). Also later in the paragraph it is stated that *"The turbulent fluxes calculated with BA_{pro} underestimated fluxes calculated with BA_{eff}"* – do you mean here that you interpret Figure 8 as showing significant differences, but only 2009? Please clarify which data indicate statistically significant relationships, and which data are used to illustrate or support the potential importance of processes during certain weather conditions (i.e. 2009 SF).

We have updated the text since the Figure 8 has changed (now Figure 8 and 9). In this paragraph we refer to Figure 9 (before to figure 8). We invite the reviewer to check the updated text in order to understand the new results. We included reference to Figure 9.

P21 ln13 *"the BA_{pro} method could not account… surface layer…"* Again, this is speculation and needs revised.

See response to comments above: the text was significantly modified, and all these statements are now presented as hypothesis.

Table 2.

- Please make sure the use of acronyms is consistent. Good quality runs -> GQR, SF, WF etc.
- It would be preferable to use the abbreviation TKE throughout rather than introduce the term e.

That has been changed throughout the text.

Also low-TKE and high-TKE would be clearer terms to use to rather than $e < 1$ m^2 s^{-2} etc.
– "2006-TKE" is confusing – perhaps use "TKE classes (2006 only)"

We hope the editing enables to clarify these issues

Table 3.

- What do the bold values in the last column indicate?

  They indicated the higher values of the regression coefficients. This has been removed.
- Again please use consistent acronyms – *H* (BA_{pro}) *H* (EC) etc., rather than introducing new acronyms H_{EC}, H_b etc.

  Done
- Same comments as Table 2 regarding TKE.

  Done
- **Some measure of the spread of results is needed (RMSE or mean absolute error) to characterize uncertainty in the fit.**

  Table 3 now includes the RMSE. The values support the results of Fig. 8.
- **The regression coefficients for situations with very low correlation are meaningless and should be removed.**

  We believe this still provides some information

Figure 3.

Change *"(a-f)"* to "(a, f)" etc.
- Change *"light shaded"* to "blue shaded"?
- Change *"gray lines"* to "dark lines"

Done

Figure 4.

- Please use AWS-G, AWS-M, SF and WF for consistency

Done

Figure 6.

- Check colours in legend.

  We are not sure to understand the comment.
- Change *"Kaimal"* to "Kansas curves"

We decided to keep Kaimal, as we homogenized all the manuscript by replacing "Kansas curve" with "Kaimal" curve.

Figure 7.

- Figure 7 is a good addition to the manuscript, but it would be better to provide separate plots for H and LE and similar figures for both BA_{pro} and BA_{eff} (i.e 4 sub plots)

  Done we have now 8 subplots since we separate for H and LE too.
- Please use the abbreviations BA_{p} and BA_{eff} in the caption, rather the longer descriptions.

  Ok
- Please clarify what the *"error values"* (+/- 1 standard deviations, 2 standard deviations?) are in the caption.

  This is not relevant anymore with the new figure.
- Please check the dates used for the SF and WF conditions: 10 - 26 July, 2006 seems to include a period of SF? Also it is ambiguous what dates/conditions are actually used – are the blue and red points selected using a combination of TKE and WP? I think the analysis needs to choose one or the other to avoid ambiguously and overly selecting data to compare. For a robust comparison, it would be preferable to present all data greater than/ less than certain TKE limits for the full 2006 period.

  We now consistently present data out of our new selection procedure, for all figures.

Figure 8.

- Please use consistent terms BA_{eff} etc

  Ok, done.
- Please clarify that *"mean measured melt"* here refers to the magnitude of melt energy back calculated from surface height changes at the AWS, not the actual melt (which is measured is mm w.e. or similar).

  Since the figure has been changed, the issue is no more relevant. We know present the actual melt.
- Why have the magnitudes of turbulent fluxes changed from the original manuscript? Have the analyses been updated? If so the changes that have been made should be carefully documented.

  We don't understand why the reviewer states the magnitude of the turbulent fluxes has changed in this figure. They are the same in both manuscripts.

References

Giesen, R. H., van den Broeke, M. R., Oerlemans, J., and Andreassen, L. M.: Surface energy balance in the ablation zone of Midtdalsbreen, a glacier in southern Norway: Interannual variability and the effect of clouds, Journal of Geophysical Research, 113, 2008.

Neale, S. M. and Fitzharris, B. B.: Energy balance and synoptic climatology of a melting snowpack in the Southern Alps, New Zealand, International Journal of Climatology, 17, 1595-1609, 1997.

We added reference to Giesen et al., since it inspired our Monte-Carlo approach. The second reference was not included since it was not directly used, even though interesting.