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.

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. 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 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.
- 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.
- 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.

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. 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.

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.

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.

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). 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. Perhaps breaking the analyses further into a daily or multi-day 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.

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.

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.

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.

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.

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.

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?).

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.

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.

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).

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.

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.

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. 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.

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).

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.

<u>Further specific comments on text and tables in the revised manuscript</u>

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

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

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

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

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

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.

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

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}).

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

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

P15 ln17 – change *"low"* to "small"

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.

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

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.

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

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

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

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.

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. 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.

P18 ln10 – *"(Fig. 6a)"*

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

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.

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.

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.

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

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.

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

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

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.

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).

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

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. 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)"

Table 3.

- What do the bold values in the last column indicate?
- Again please use consistent acronyms – H (BA_{pro}) H (EC) etc., rather than introducing new acronyms H_{EC}, H_b etc.
- Same comments as Table 2 regarding TKE.
- Some measure of the spread of results is needed (RMSE or mean absolute error) to characterize uncertainty in the fit.
- The regression coefficients for situations with very low correlation are meaningless and should be removed.

Figure 3.

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

Figure 4.

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

Figure 6.

- Check colours in legend.

- Change *“Kaimal”* to “Kansas curves”

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)
- Please use the abbreviations BA_p and BA_{eff} in the caption, rather the longer descriptions.
- Please clarify what the *“error values”* (+/- 1 standard deviations, 2 standard deviations?) are in the caption.
- 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.

Figure 8.

- Please use consistent terms BA_{eff} etc
- 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).
- 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.

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.