Response to anonymous reviewer's 2 comments on tc-2016-93.

We kindly thank the reviewer for her/his positive review that helped us to considerably improve the paper. We provide an edited version of the paper, including track changes (in red, added or modified text, and suppressed text), and a final version. We answered to all general and specific comments (reviewers comments are in black bold, our response in blue) and made changes in the text when necessary. Many similar issues were also raised by the other reviewer, we provide the same answers in these cases.

The paper uses eddy covariance and vertical wind-speed and temperature profiles to quantify the errors inherent in using the bulk aerodynamic method for energy balance modeling. Errors were estimated by comparing model outputs using the three approaches over two measurement periods. The authors used large scale weather patterns to constrain their comparisons. The authors should be commended on their use of interesting datasets. However, they need to refocus the Results and Discussion sections to better elucidate the conclusions of the paper.

We considerably changed section 4 (Results) and section 5 (discussion), in order to assess this overall comment and that of the first reviewer. In section 4, Table 2 has been split in two tables. The first one is describing the main characteristics of the weather pattern based and the TKE based classifications in terms of meteorology, and the second one is containing the turbulent fluxes information for each class. This second table includes previously missing information about the regression coefficients between measured and modeled turbulent fluxes. Meteorology and wind regimes (4.1) are now described differently: first for weak forcing, then for strong forcing. The subsections in 4.1 have been removed. The TKE description (4.2.1) has been simplified and the related Fig. 6 has been removed. We profoundly changed the description of the turbulent fluxes (4.2.3), we removed Fig. 8 and added a figure comparing the measured and the modelled fluxes (Fig. 7 in the new version). The discussion about the turbulence in the surface layer (5.1) is now better framed, and we provide a better estimate of the error on the melt energy as derived with the sonic ranger (5.2).

Table 1 has been updated with more information about the sensor heights.

Minor changes have also been made in the Introduction, 2nd and 3rd paragraph know provide a better overview of the katabatic flow issues. In methods section (3), we slightly modified the TKE description (section 3.2, paragraph 1), we provide a better description of the roughness lengths derivation method (section 3.2, paragraph 2) and include the EC-based method for roughness lengths derivation (related additional results are discussed in section 5.1, 2nd paragraph). Description of the error derivation method has been improved and has now its own subsection, 3.3.

We hope the new version will be more straightforward to read.

General comments

- **Impact of changing sensor heights**

In Section 2.2 a description of the changing sensor heights is given, as well as the manual lowering used to overcome these changes. The following remains unclear: was the 60-90 cm change accounted for in the flux calculations, or was a fixed height used (is unclear in Section 2.3)? Are the 'mean' heights given the height the sensors were changed to, the height after 5 or 7 days of change,

or other? How do changing heights impact the flux results? Were different height values tested during calculation of the fluxes to understand possible errors?

The value found in Table 1 is the mean over all the campaign. The changing sensor heights were taken into account in the flux calculation and the related uncertainty was included in the random error calculation. In order to obtain the day by day height changes, measurements from the SR50 (sonic height ranger) were used. The following statement can be found in section 2.3 : *"The changing instrument heights were derived from the sonic height ranger and regular field visits and controls"*. We now provide the range of variation and the standard deviation of the heights in Table 1.

The way the uncertainties were computed is described in detail in Litt et al. (2015b), where a complete analytical method had been developed to compute the random errors on the BA method and the EC method. The paragraph about errors in section 3.2 was transformed in a new section 3.3, and the description of the error assessment has been improved.

- **Surface properties**

Sections 2.2 and 3.2 describe the changes to the surface observed in each campaign. Can the fetch properties also be added to Section 2.2 so that it is clear that the instrument heights are appropriate for the corresponding homogeneous surface (especially for the eddy covariance measurements).

The glacier ablation area, over which the sensors were installed, showed homogeneous surface characteristics in all directions for distances of several hundred of meters. We assumed the fluxes and the turbulent characteristics derived at the measuring site were representative of the turbulent characteristics all over the ablation area. We added, p 5 line 15, *"The glacier surface characteristics remained homogeneous in every direction on hundreds of meters from the measuring site."*

- **Calculation of roughness lengths**

Section 3.2 describes the computation of roughness lengths using the profile data. Can you provide an explanation as to why the lengths were not calculated using the eddy covariance measurements, or, compare the lengths calculated using each method. Also, it would be useful to include an error range of the roughness length, as whilst it is stated that it 'did not change significantly', it would be useful for the reader to be able to evaluate the range (and associated error).

We had not included the roughness lengths 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 scattered but we found a median value of z_0 of 0.02 m and z_t of 6.6×10^{-6} m when derived from the EC system, values higher than the one derived from the profiles. We mention that in the new manuscript lines, p 8, line 17.

Interestingly, the factor $(\ln(z/z_0)\ln(z/z_t))^{-1}$ in the bulk formulation, calculated with the EC derived roughness or the effective roughness, is almost identical (0.0165 and 0.0167). This supports the fact that the effective roughness lengths are used to compensate for the BA method underestimation of the fluxes. We mention this result now in the discussion section, p 18 line 6.

The roughness length evaluated with the profiles are scattered between 0.1 m and 0.00001 m, which prevents the observation of significant change since the maximum observed height changes at the surface reached 0.1 m at the end of the campaigns. These large errors are likely due to large uncertainties on the temperature measurements (Sicart et al., 2014). See the following insertion, p 8 line 5:

"The results show a large scatter, about four orders of magnitude. The median values are z_0 = 0.001 m and z_t = 0.00001 m. The scatter can be attributed to poor accuracy of the temperature measurements leading to large random uncertainties (Sicart et al., 2014). The scatter was too large to observe significant changes in the measured roughness lengths during the 2009 campaign, in spite of snow falls or snow melt that uncovered the ice surface, or appearance of small gullies of about 0.1 − 0.3 m height variations on a few meters horizontal scale that could also have impacted the roughnesses (Smeets and Van den Broeke, 2008a, b)."

The fluxes error calculation included an error range on the roughness lengths, as mentioned in the original manuscript, in the error calculation description. This error on $z_{0,t}$ was evaluated as the result of the errors on temperature and wind speed measurements (Sicart et al., 2014) and is set to $d\ln z_0$ = 1.5. This value corresponds roughly to the standard deviation of the profile derived roughness lengths. We invite the reviewer to check the Sicart et al. (2014) publication for further details on the procedure used to derive the roughness parameters and the associated errors.

-Role of sublimation

Section 3.3 shows the formula used to calculate melt based on the SR50 measure-ments. Can you please add a comment to describe the role of sublimation, and why it is excluded from a mass loss calculation.

When the latent heat fluxes are at their maximum (for the high TKE subset), we find mean *LE*=-10 W m^{-2} (Table 3, former Table 2). This corresponds to a daily ablation of 0.3 mm w.e., if we consider this is sublimation and the rate is sustained during a whole day. That is clearly negligible. We added a sentence in the text to explain, p10 line 5: *"Ablation due to evaporation or sublimation at the surface was considered negligible: mean absolute latent heat fluxes remained below 10 Wm^{-2} for the most turbulent subsets (Table 3) which corresponds to a daily ablation of only 0.3 mm w.e."*

-Results

Currently the Results section combined comments related to both timing of events as well as classification by weather type. Please restructure to only focus on the weather patterns (as that is what is used in the Discussion section). Changing between time and pattern becomes confusing for the reader, as it becomes a lot of information to keep in mind.

Since the classification by weather patterns doesn't always reliably work for describing turbulent conditions in the surface layer, we decided to focus on a characterization of the turbulence and turbulent fluxes using a TKE classification. Since the strong WP tend to select high TKE conditions and low WP, low TKE conditions, the TKE classification helps understanding the processes occurring in the WP-related classification. We emphasized this point as much as possible throughout the text. In the new version, this point is evidenced in the results section 4.1, then most of the results are presented in terms of the TKE classification, and we finally discuss how that impact on the results for the WP classification.

-Results Section 4.1

Currently Section 4.1 is quite dense and difficult for the reader to pull apart. Please restructure in a way that is easier for a reader to follow. For example: "SF conditions are characterized by ___________. This was observed ______ (example). The impact on melt is ________." In this way, the reader can focus (and retain) the most relevant information associated with each weather pattern (rather than interchanging between time periods and weather patterns).

We significantly modified the text in this section 4.1 to take into account your comment together with the other reviewer comments. Meteorology and wind regimes (4.1) are now described differently: first for weak forcing, then for strong forcing. The subsections in 4.1 have been removed. The whole text is now much shorter and clearer, the whole section 4.1 included 42 lines, it now contains only 21.

Results Section 4.2.2

As the paper is directed at a general glaciology audience and not at a specialized micro-meteorology community, there needs to be a better introduction to spectral analysis within this section as a reminder. Perhaps the addition of one or two sentences describing why the analysis is necessary and what relevance it has to the theme (ie that it will help to describe the turbulent boundary layer) is necessary to ease the reader into the section. Even though it is briefly described in Section 3.2, it would be beneficial to restate the information here. Also, the term 'Kansas curve' and 'Kaimal curve' is used interchangeably throughout the document – please standardize.

We agree with the comment, and we added a sentence, which include reference to the methods sections, p13 line 9: *"The Fourier analysis of the 2006 EC data of the wind speed components and temperature were compared to the Kansas curves, to see if the surface layer was in equilibrium (Sect. 3.2)"* And modified the following sentence to account for this inclusion.

We standardized to "Kansas curve".

-Discussion Section 5.3

If there are concerns regarding the comparison with the weather patterns, these should be quantified.

Actually the concerns are limited, but the text wasn't very clear. The weather patterns do well at distinguishing typical wind conditions, and generally reflect differences between high and low TKE. Regarding the total SEB calculation, the WP classification is maybe less efficient. We exchanged the position of the sentences to reflect this. We added, in this section: *"can be useful to assess turbulent surface fluxes characteristics"* and *"It also provides a rough classification of the turbulence conditions in the glacier surface-layer, since SF and WF conditions are, in the glacier surface layer, likely associated with high and low TKE conditions, respectively. We show that the BA method compared differently with the EC method for different WP. More generally for SEB studies, we highlight some limitations."* We removed *"is only partly adapted for SEB studies"*.

-Results-Discussion-Conclusions disconnect

In the current form, there appears to be a disconnect between the Results, Discussion and Conclusions sections. Often, results required to support statements made in the conclusions (eg: erratic discrepancies between results during certain conditions) are not supported in the Results section, and additionally, certain sections of the Results (eg. Meteorology and wind regimes) are not fully utilized in the Discussion/Conclusions sections. I would recommend to the authors to visualize the conclusions and include only the results necessary to address those points.

The inclusion of table 3 with the correlations between the fluxes from the BA method and the EC method, the change of fig 8 to a scatter plot of the turbulent fluxes, and the overall revision of the text, now provide a much clearer and focused discussion. We removed the figure related to TKE and the related discussion, which was sufficiently framed and lacked supporting data. We didn't significantly modified the conclusion, but with the other modification they now better reflect the content of the results/discussion.

Specific comments

P1 L3 Please clarify the time period in which each dataset was collected – from the description it seems that eddy covariance data and the profile data was collected during both the 2006 and 2009 study periods.

Ok - done

P2 L2 Change 'englacial' to 'glacierized' (or similar)

Done

P5 L5 Why was a 1-hour period used instead of the standard 30-min run?

The choice was constrained by the parallel use of the outside meteorological station data for which sampling was set on 1-hour. Calculations made on a 30 min basis didn't change the relative contribution of the different turbulent fluxes to the SEB.

P5 L19 Please explain how the data was 'analyzed and compared'.

The weather pattern series has been kindly provided by "Electricite de France" (EDF), and we did not undergo this analysis ourselves. The paragraph provides a brief explanation of the whole method. For the specific step of analysis and comparison, each day, the shape of the observed geopotential height field over Europe is characterized by the geopotential height at 0h and 24h at the 700 hPa level and the 1000 hPa level for 110 grid points (a total of 440 points). This field is compared, in this 440 dimensions space, to the 8 geopotential height fields types which have been obtained by the statistical sorting of measured rain patterns in south-east France. The WP of the day is determined by finding the WP for which the geopotential height pattern is the nearest to the observed geopotential height field in the 440 dimension space, using the Teweles-Wobus score (Garavaglia et al., 2010).

Detailing this complex procedure is beyond the scope of this paper. We provide reference to the paper from Garavaglia et al. 2010. We modified the text to add a short description for the selection of the WP: *"Then, for any day (inside or outside the period used to characterize the decomposition, e.g. 1956-1996) the observed geopotential height field shape over Europe is characterized by the observed geopotential height at 0h and 24h, at the 700 hPa level and the 1000 hPa level for 110 grid points (a total of 440 points). This field is compared, in this 440 dimensions space, to the 8 geopotential height fields proposed by the rain-patterns analysis. The nearest of the 8 fields provides the WP."*

We added the following sentence at the end of the paragraph*: "Details of the procedure can be found in Garavaglia et al., 2010."*

P6 L11 Change Table 3 to Table 2

Done. Note that table 2 has been split in table 2 and table 3 and table 3 now contains more results: errors on mean turbulent fluxes, regression parameters between modelled and measured fluxes, as well as correlation coefficients.

P9 L4 Please minimize the use of brackets. It makes the sentence confusing for the reader.

Ok, removed.

P12 L9 Change 'again a after' to 'again after'

This detailed description of the melt behavior has been removed

P13 L9 Please minimize the use of brackets. It makes the sentence confusing for the

Done.