Response to the anonymous referee #2

Main comment:

The paper by Réveillet and others describe the results of the modeling of the surface mass balance of Saint-Sorlin glacier in the French Alps. The SMB is calculated with CROCUS, an energy balance model originally designed to calculate snow pack. This model is forced with SAFRAN reanalysis data that is corrected to the in-situ weather observation at the location of the glacier. The authors explore the sensitivity of the simulated SMB to different components of the climatic forcing and to various model parameters. Furthermore is the result of the energy balance model compared with the results of an empirical model. The study is well written and the results are interesting. I have the following comments, spit up in general comments, specific points and some technical points, the latter two per page and line number.

Authors' response:

Thank you for your positive comment on the interest of the paper. We carefully answered point by point to your general comments and specific/technical points. Note that replies from the authors are in green.

General comments

General comment #1:

The difference between energy balance model and empirical model is not just the model formulation, but also the way the models are calibrated. The parameters of the empirical model are calibrated by fitting calculated SMB to the observed SMB, while for the energy balance the meteorological input of the model is fitted to observed weather. Would the energy balance model perform better if its model parameters and input corrections are optimized to the observed SMB?

Authors' response:

You are right, the performance of the energy balance in modelling SMB could be improved by using SMB measurements for calibration (as for example in calibrating the z_0 to fit the SMB to observations as it has been done in several studies). However, this doesn't mean that the model is performing better for good reasons. SMB can be well simulated by compensating bias and/or uncertainties (as can be the case for the turbulent fluxes as mentioned above). In this study, we have many energy balance measurements, including turbulent fluxes. Our objective is to evaluate the performance of the model while ensuring that all energy components are well modeled. This is indicated in the introduction: "Then, the surface energy and mass balance model is calibrated using the measured energy fluxes to ensure that all the energy balance components are accurately represented."

The comparison between the energy balance model and the empirical model is done to evaluate the performance of these models considering the available data. The main idea is to identify and discuss the most important variables for each model and to provide clues as to the best approach to be used for future simulations, while limiting uncertainties.

General comment #2:

I do not fully see how the conclusion that empirical models would be better suited to model glacier SMB for periods, or glaciers, without AWS measurements.

The results of this paper show that the energy balance model performs better, mainly in the accumulation area, than the empirical model. This is also true for the period where the input of the *model is not directly corrected with the AWS data. So therefore I do not see why the empirical model should be more suited to model future glacier mass balance, when also no direct correction of the meteorological forcing with in-situ observation is possible.*

Authors' response:

We do not really agree with your comment; our results do not indicate that the energy balance model performs better than empirical approaches. Regarding the comparison between these approaches, results depend on the area (ablation *vs* accumulation) and the quality of the forcing. We have to note that this comparison have been done using an adjusting forcing (even over the period when no AWS are available to assess this correction).

Nevertheless, the study has been designed especially to assess the performance of energy balance and empirical models. For this purpose, we first evaluate the performance using the most accurate forcing we have (i.e. reanalysis adjusted with AWS measurements). Then, we performed a thorough study of sensitivity to meteorological inputs (section 4.2) and other parameters (surface roughness, …). Our modelling experiments and comparison with the *in-situ* measurements reveal a strong sensitivity of energy balance model to the wind speed, especially in the ablation area (i.e. over ice surfaces). Indeed, the use of uncorrected wind data (i.e. coming from SAFRAN reanalysis) leads to large differences in annual mass balance of the ablation area (1 to 1.7 m w.e. yr^{-1}), as shown in Figure 7 (blue curves). Note that the differences are smaller in the accumulation area (over snow surface). In addition, for the whole dataset of mass balance obtained from ablation and accumulation areas, the model performance (using the Nash and Sutcliffe coefficient) decreases from 0.67 to 0.27, with or without the corrected values of the wind speed data respectively. Similar conclusions are obtained for the long-wave radiation correction (please, refer to section 4.2.2.2). Although the sensitivity is low in the accumulation zone, these results clearly show that *in-situ* meteorological data are needed to ensure a good performance of energy balance model, as using reanalysis without adjustment considerably decrease the model performances. Consequently, the energy balance model cannot be transferred to another glacier without *in-situ* meteorological data. In addition, the current bias in reanalysis data highlights the complexity to model all the meteorological variables in the present; we thus expect to also have bias in future projections. In this case, and due to its strong sensitivity, an energy balance model could lead to significant bias for future simulations.

Otherwise, our study reveals a strong impact of roughness value on ice ablation and indicates the necessity of measurement to calibrate this parameter. Given that the surface roughness value is unknown in the future, this is a strong limitation of this kind of model.

Therefore, empirical approaches based on precipitation and temperature only, could be more appropriate for simulations of glaciers in the future, as performances presented in this paper are similar than for an energy balance model, in the ablation area. For the sake of clarity and to take into account this comment, the manuscript has been revised (see below).

However, despite the limitations of energy balance model for simulations in the future, this physical model remains crucial to study the processes to understand the physical relationships between the meteorological variables and ablation. In the conclusion, we added a sentence in this direction to mitigate the obstacles of energy balance model abilities to simulate the glacier mass balance in the future.

Changes made in the manuscript indicated below.

1: In the abstract section as follows:

"Given the uncertainties in the temporal evolution of the relevant meteorological variables and glacier surface properties in the future, empirical approaches based on temperature and precipitation could be more appropriate for simulations of glaciers in the future."

2: The conclusion was organized and completed as follows :

"This study has evaluated the performance of the Crocus snowpack model, which was fed with SAFRAN reanalysis data, thereby simulating seasonal and annual SMBs of Saint-Sorlin Glacier over the last 20 years. Using meteorological forcing adjusted with *in-situ* measurements, our results show very good performance of the model to simulate summer SMB in both accumulation and ablation areas. Performance of the model is lower for the 1996-2005 period due to the absence of *in-situ* meteorological measurements to adjust the forcing data.

Additionally, this study compared the performance of this energy balance model to an empirical approach using temperature and potential incoming solar radiation as inputs. Regarding simulations of summer SMB for the accumulation area, our results show better performance using the energy balance model, especially concerning simulations of snow and firn melting in the accumulation area. Regarding the ablation area of the glacier, the two approaches show similar performance when forced with meteorological data adjusted with nearby AWS measurements. When such measurements are not available in the vicinity of the glacier, performance of the empirical model in the ablation area is superior although the physical processes are not properly represented. However, the temporal stability of the calibration parameters of the empirical approach need to be assessed over a longer time period before using such an approach over several decades.

These conflicting conclusions about model performance in accumulation and ablation area emphasize greater importance of having meteorological data to correct the forcing in ablation area. Indeed, ~~According to~~ in our sensitivity study using forcing data, the results demonstrate that the Crocus model is highly sensitive to wind speed, especially for ice melt simulations. Indeed, using *in-situ* wind speed data instead of reanalysis data (where observed wind speed values larger than 10 m s^{-1} can be under-estimated by a factor 2 or 3) led to an annual mass balance decrease of more than 1.7 m w.e. yr^{-1}. Thus, without local wind speed measurements, the model's performance strongly decreases, even using wind speed data corrected via a quantile-mapping method. In addition this study confirms the findings by Dumont *et al.* (2012) concerning the importance of correcting the incoming longwave radiation from SAFRAN.

Model calibration represents an important step to improving model performance. According to the sensitivity study concerning model calibration, our results highlight the importance of calibrating the ice surface roughness using turbulent fluxes measurements. An increase in $z0_{ice}$ by a factor of 10 can have an impact of 1.5 m w.e. yr^{-1} on ice melting. Regarding the ice albedo, while having *in-situ* measurements to calibrate the model improved model performance; the sensitivity of summer SMB for this variable is lower than the sensitivity to wind speed over icy surfaces (the ice melt difference reaches 0.48 m w.e yr^{-1} when the ice albedo is divided by a factor 2). This could suggest a relatively low sensitivity to ice albedo change (due to dust or black carbon for example) for summer SMB variations in the future."

While both these approaches can provide good summer SMB simulations, winter SMB simulations need to be corrected using winter mass balance measurements. In any case, our results indicate a strong sensitivity of annual SMB to winter SMB. The understanding of the spatio-temporal variability of accumulation processes at the glacier surface needs to be more fully investigated in future work.

In conclusion, our study reveals the major role of wind speed, which controls the magnitude of turbulent fluxes, on melting. The results highlight a very serious obstacle for the modelling of future glacier mass balances, as this meteorological variable is highly unpredictable. Our results also suggest that the sensitivity of annual mass balance to accumulation and wind speed parameters is of primary significance, as compared to the sensitivity to snow and ice albedo changes. However, as such data are still difficult to represent in climatic models, the accuracy of their predictions are also questionable (e.g. Terzago *et al.*, 2017). We thus suggest a careful use of the physical approach for future long-term simulations, considering the uncertainties. Nevertheless, despite these limitations for future simulations, this physical model remains crucial to study and understand physical processes and interactions between atmospheric variables and ablation. Otherwise, although empirical approaches based on simple meteorological variables also have serious drawbacks, they could be more appropriate for simulations of glaciers in the future, especially to simulate summer SMB in ablation areas, bearing in mind the lack of availability of reliable information on future meteorological variables and surface roughness."

General comment #3:

It would be instructive include information and figures on the climate at Saint-Sorlin Glacier in the paper. And to include in these figures a comparison between the original climate given by the SAFRAN data and the climate as measured with the AWS.

Authors' response:

As also requested by reviewer 1, we propose to add a summary of meteorological conditions in a supplementary material. We choose to present the measured conditions on the moraine at 2720 m. Comparison with SAFRAN data are given all along the text, in section 2.3.4.

Based on 8 years of AWS_m records on the moraine (2006-2013), daily means of temperature (°C), relative humidity (%), incident shortwave radiation (W/m²), incoming longwave radiation (W/m²), wind speed (m/s) are presented on the figure (see below). Concerning wind direction, instantaneous half hourly data are classified and a percentage of number of data per direction is given.

We proposed to add in the manuscript, l.119 :

"A summary of the meteorological conditions at AWS_m is given in the supplementary material."

The related figure and the figure caption are:

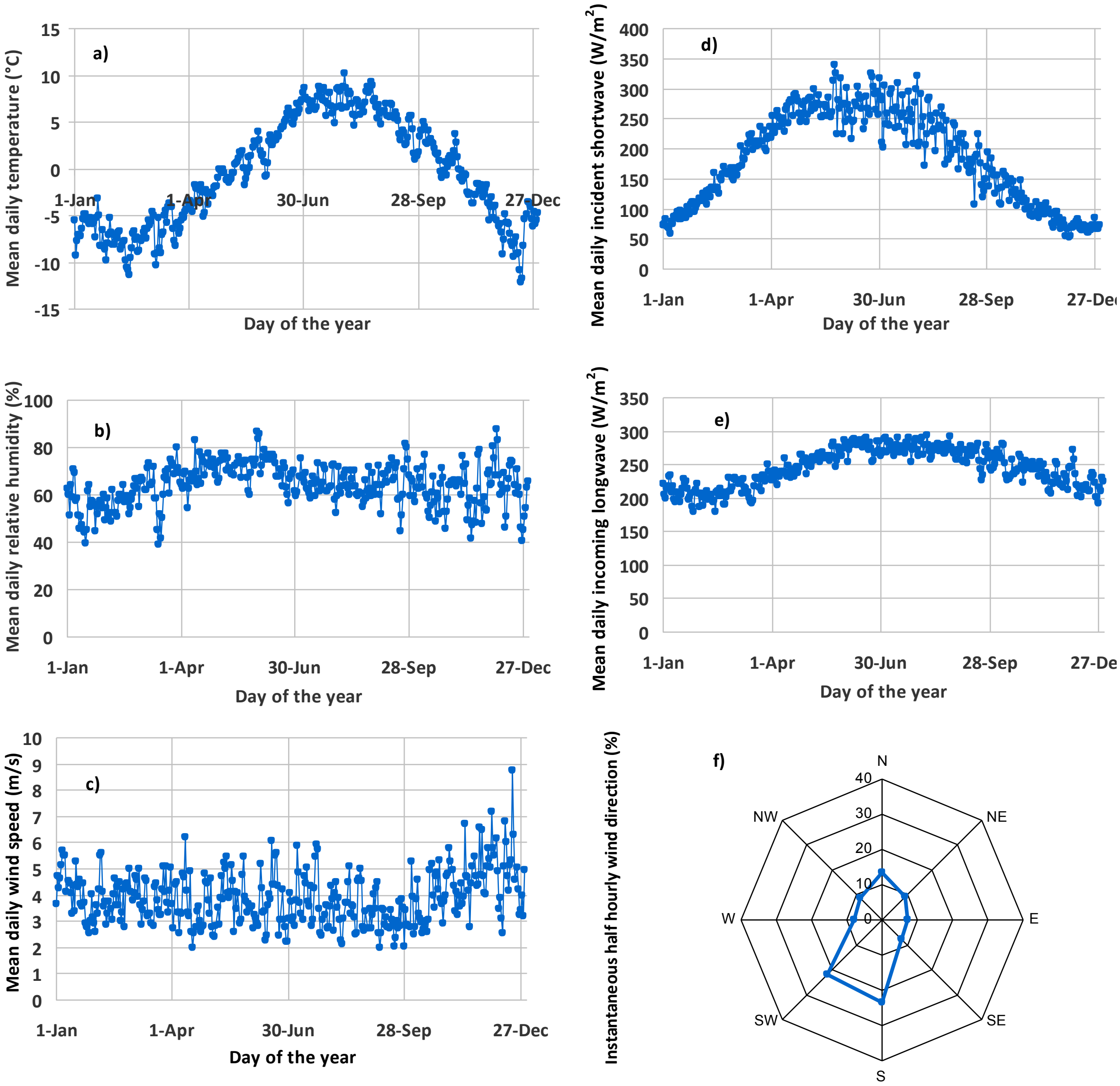


Figure 1: Daily means of a) temperature (°C), b) relative humidity (%), c) wind speed (m/s), d) incident shortwave radiation (W/m²), e) incoming longwave radiation (W/m²), f) percentage of instantaneous half hourly wind direction data in each direction (%). Data are calculated on the 2006-2013 period, at AWS$_m$ (2720 m)

Specific comments

Specific comment #1 (page 2 - l7):
In addition to the fact that it is questionable if the calibrated parameters are valid over long time periods, a disadvantage of the temperature-index models is that the parameters have no validity for other glaciers. Maybe you could add this here.
Authors' response:
A sentence and references have been added in the introduction in response to your comment:
In addition, transferring parameters determined for an instrumented glacier to another site decreases model performance (Carenzo *et al.*, 2009; Réveillet *et al.*, 2017).

Specific comment #2 (page 3 - l2-5):

Here you discuss variability, but you are, also, interested in the absolute value of the SMB, not only of the variations from year to year.

Authors' response:

Right. This is the reason why the beginning of the sentence is "the temporal variability of the annual SMB is mainly driven by summer SMB variability ». We also have precised that the simulation of summer SMB (absolute value) strongly depend on winter SMB.

According to your comment, we reworded as follow:

"In the Alps, the temporal variability of the annual SMB is mainly driven by summer SMB variability (e.g., Six and Vincent, 2014). For this reason, many studies have focused on ablation modelling. However, simulated summer SMB and associated uncertainties strongly depend on the winter SMB (Réveillet *et al.*, 2017), highlighting the need for a quantification of the sensitivity of annual SMB to both seasonal components."

Specific comment #3 (page 3 - l20):

I do not see that the aspects of Saint-Sorlin vary a lot in Figure 1, it is mainly N - NE - E slopes. Please be a bit more specific.

Authors' response:

In response to this comment as well as comment 4, a map has been added in Figure 1, that shows glacier topography with a higher resolution and the surrounding topography. In this map, the main glacier flowing lines are indicated with arrows to better represent the different glacier aspects. We chose to add a new map instead of adding information to the one on the article for sake of clarity.

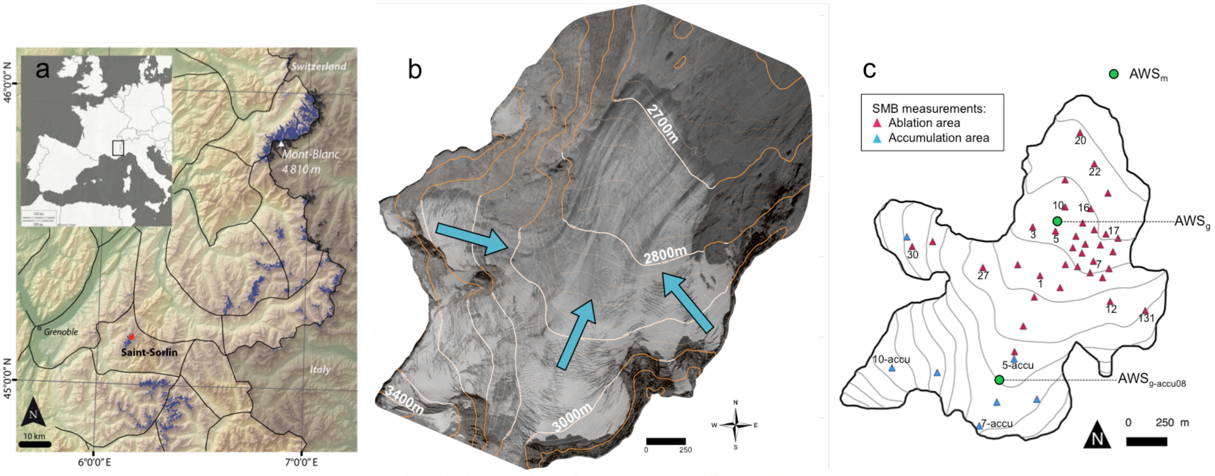


Figure 1. (a) Location of Saint-Sorlin Glacier in the French Alps. French glaciers are shown in blue except for Saint-Sorlin Glacier, used for the present study, which is in red. Black lines represent SAFRAN massif outlines (adapted from Rabatel *et al.*, 2017) (b) Aerial photo of Saint-Sorlin glacier. Blue arrows indicate the three main glacier flow lines. (c) Map of Saint-Sorlin Glacier with the network of in-situ SMB measurements (blue triangles in the accumulation area and red triangles in the ablation area). Locations of automatic weather stations used in this study are represented by green circles.

Specific comment #4 (page 4 - Figure 1):

Could you add to surrounding topography in Figure 1b? That could give a better understanding for *your discussion of wind speed and snow distribution.*

Authors' response:

As mentioned in the previous comment, a map representing the surrounding topography has been added.

Specific comment #5 (page 4 - l10):

- What is the accuracy of the DEMs?

- Do they give the same values for stable ground?

Authors' response:

- These DEMs were derived from aerial photogrammetry using a 10-m spatial resolution. In particular DEM for 2003 have been done based on the same method than in Thibert et al., 2008.

Thibert E., R. Blanc, C. Vincent, N. Eckert. Glaciological and volumetric mass balance measurements: error analysis over 51 years for Glacier de Sarennes, French Alps. (2008). Journal of Glaciology, 54 (186), 522-532).

In this study, performed with aerial photographs coming from the same aerial campaign, errors due to internal stereoscopic measurements and roughness have been assessed to 1.26 m and 0.35 m respectively (i.e. 1.31 once combined) and are assumed to be similar for Saint Sorlin glacier. The orientation error is 0.22 m.

- Recent studies (unpublished) compared DEM 2003 and 2014 of Saint Sorlin glacier and indicate a mean difference of 0.52 m outside the glacier. (R. Basantes, personal communication)

In our study, DEMs have been resampled to a 200m resolution DEMs. Due to the low uncertainties mentioned above, the uncertainties due to the DEM acquisitions are negligible, in particular in this case (i.e. considering the resample).

Specific comment #6 (page 5 - l23):

So SAFRAN is not gridded, but has one output per mountain region as given in Figure 1a? And this is then distributed following elevation and aspect?

Authors' response:

Yes, SAFRAN data are not gridded but are given for the different massifs, altitudinal ranges (every 300m) and aspects (7 orientations). This has been clarified by the following additions to the sentence:

'… that are assumed to be homogeneous within a given massif (in particular within the Grandes Rousses massif where the Saint-Sorlin Glacier is located, Figure 1a) and depend only on altitude (one data every 300 m) and aspect (7 orientations available: N, NE, NW, S, SW, SE and 'Flat').'

Specific comment #7 (page 7 - l5):

High correlation doe not say it all, as you can have a significant bias while having a high correlation. Could you indicate how well the values correspond, for example with a RMSE?

Authors' response:

Yes, we agree with your comment. This comment was also made by the second reviewer. Therefore, RMSE values have been computed and added in this section as follows:

- "SAFRAN and AWS$_m$ hourly air temperatures over the ablation and accumulation seasons are well correlated (R^2 = 0.98 (summer) and 0.99 (winter), both significant at the 99% confidence level (Student's t test), RMSE = 0.7°C (summer) and 0.76°C (winter)). Hourly SAFRAN relative humidity is also in good agreement with the AWS$_m$ data (R^2 = 0.74, significant at the 95% confidence level and RMSE = 13.6%.)"

- "Using this correction, the correlation between AWS$_m$ incoming LW radiation and corrected LW radiation from SAFRAN increased the correlation from R^2 = 0.71 to R^2 = 0.83 and decreased the RMSE from 44.3 W m^{-2} to 29.7 W m^{-2}."

Note that in the previous version, the correlation regarding the LW was indicated as R and not R^2. This mistake has been corrected in the current version.

- "Correlations between daily incoming shortwave radiation ($R^2 = 0.81$) are significant at the 99% confidence level (Student's t test) and RMSE = 77.2 W m^{-2}."
- "A poor correlation ($R^2 = 0.19$, RMSE = 3.8 m s^{-1}) between SAFRAN wind speed (considered at 2-m)…"
- "Since the correlation between the measured wind speed on the foreland and on the glacier is high (R^2=0.97, RMSE=1.7 m s^{-1}) "

Specific comment #8 (page 7 - l9-10):

How does a lack of low-altitude clouds in SAFRAN explain an overestimate of incoming long-wave radiation in SAFRAN? I would expect that low altitude clouds are warmer than high-altitude clouds and thus emit more long wave radiation, such that not including these low-altitude clouds would lead to an underestimate of the incoming long wave radiation.

Authors' response:

We completely agree. This was a mistake that slipped into the original manuscript. The sentence initially read:

"The comparison between SAFRAN and AWS_m incoming long wave radiation indicates an overestimation of SAFRAN data for **low** cloudiness conditions. This can be explained by local orographic features and/or **low**-altitude clouds that are not considered in SAFRAN reanalysis."

It should be, in reality: "The comparison between SAFRAN and AWS_m incoming long wave radiation indicates an overestimation of SAFRAN data for **low** cloudiness conditions. This can be explained by local orographic features and/or **high**-altitude clouds that are not considered in SAFRAN reanalysis." This has been changed in the new manuscript.

Given that this sentence was neither clear nor correct, we decided to rewrite it as follows: "The comparison between SAFRAN and AWS_m incoming long wave radiation indicates an overestimation of SAFRAN data for **low** cloudiness conditions. This can be caused by **high**-altitude clouds that are not considered in SAFRAN reanalysis and an incorrect vertical discretization of the atmosphere in SAFRAN."

Finally, we also add the RMSE values (in accordance with comment 7) on the bias between measured and calculated incoming LW as follows:

"Using this correction, the correlation between AWS_m incoming LW radiation and corrected LW radiation from SAFRAN increased the correlation from $R^2 = 0.71$ to $R^2 = 0.83$ and the RMSE decreased from 44.3 W m^{-2} to 29.7 W m^{-2}."

Specific comment #9 (page 7 -l17):

Also here, could you give an estimate of the bias/rmse in addition to the correlation?

Authors' response:

Done, please refer to the response to comment #7.

Specific comment #10 (page 7 - l22):

How likely is it that you measured the katabatic wind on the site of AWSm? It is located off the glacier and well above the nearest glacier surface (glacier extends to below 2700 m and the AWSm is at 2720 m), while the katabatic wind on a small glacier can be quite shallow. If you look at the wind direction, do you then see a consistent down-slope wind?

Authors' response:

This was based on other studies. This is now indicated in the paper (see below) and a sentence has been added concerning wind direction in response to your comment and those of the other reviewer.

"This underestimation is likely due to both non-consideration of katabatic wind and local effects due to orography (Dumont *et al.*, 2012). As mentioned in Litt *et al.* (2017), when large-scale atmospheric forcing was strong, intense downslope winds were observed, aligned with the main glacier flow (*i.e.* coming from the South, see Figure 1)."

Specific comment #11 (page 10 - l30):
It is not clear why you perform the calculations on a 200 m resolution. Here you refer to 2.2.2, where in turn is referred to 2.3.3 where it is stated 'we linearly interpolated on the 200-m horizontal resolution grid' without any further reason why this interpolated to this 200 m. Please explain.
Authors' response:
"See section 2.2.2" has been replaced by "see sections 2.2.2 and 2.3.3.
In addition, the following information regarding the reason for this grid size has been added:
"Note that a 200 m resolution was chosen as a compromise to be sufficiently precise to consider the spatial variation of Saint Sorlin glacier (in particular the variation of aspect) and capture variability between stakes, while maintaining relevance regarding the meteorological forcing (given that values are available every 300 m of elevation)."

Specific comment #12 (page 11 - l10-33):
I would add the content of these section 3.2.2.x to the respective subsections in the results section 4.
Authors' response:
As according us no results are presented in this section, we would like to keep this content in the method section.

Specific comment #13 (page 12 - l10-11):
- Here you write that the interpolation used to create the precipitation can explain the difference between measured and simulated winter SMB. But if all winter SMB measurements are included in the determination of the precipitation fraction maps, then how does the interpolation between these points affect the model results on the stake locations?
- And how do melting events in the accumulation period explain the difference? They should be in the simulations as well?
Authors' response:
- The interpolation method is the krigging: it allows interpolating the data with the best fit at each measurements point; but as the grid resolution is 200 m, it could exist differences between measured and interpolated points. In addition the exact position of each stake can be distinct from one year to an other (few meters), and can lead to differences.
- Regarding the melting events, you are right, it is considered in the simulations. Nevertheless, it doesn't mean that the model is able to simulate the exact rate of melting, and it therefore exist an uncertainty.

Specific comment #14 (page 12 - ln29):
This contradicts the conclusion that when no correction in the forcing is possible due to lack of AWS data, the ATI performs better than the energy balance simulations
Authors' response:
In fact, these results indicate that "Over the period 1996-2005, considering all the point data over the entire glacier, Crocus performs better than the ATI model". But as mention later, this is explained by the low performance of the ATI in simulating ablation in the accumulation part: "Here again, summer SMBs simulated with the ATI model in the accumulation area are under-estimated." And if we consider only the ablation part, the ATI performs better: "On the other hand, when considering the

ablation area only, results from the ATI model better fit the summer SMB measurements (NS is 0.36 for Crocus and 0.59 for the ATI model)."

Indeed, it is difficult to draw conclusions for the entire glacier, but note that these conclusions are drawn based on a correcting forcing (even if this correction hasn't been validated over this period). As mentioned in the main comment #2, the sensitivity of meteorological forcing and surface roughness is significant, and in particular for ice surface (which mean in the ablation area; see sections 4.3.2.2, 4.3.2.3 and 4.2.3.1). As these data are really difficult to be modeled in the present, it should lead to significant uncertainties in the future. Due to these uncertainties and the sensitivity of the model to these variables, we suggest that an empirical model, requiring only the temperature appears to be a good option, especially to model the summer SMB in the ablation area.

Nevertheless, in response to you comment, it has been specified in the conclusion that ATI model is probably more appropriate for futures summer SMB simulations, but in particular for the ablation area:

"Otherwise, although empirical approaches based on simple meteorological variables also have serious drawbacks, they could be more appropriate for simulations of glaciers in the future, especially to simulate summer SMB in ablation areas, bearing in mind the lack of availability of reliable information on future meteorological variables and surface roughness."

In addition some changes have been made to clarify our conclusions (see main comment #2).

Specific comment #15 (page 14 - Figure 4):

You do not plot the correlation here.

Authors' response:

"Correlation" has been replaced by "comparison" in response to your comment.

Specific comment #16 (section 4.1.3):

I think the results of this section could be clarified with a scatter plot where you plot annual SMB vs summer SMB with constant winter SMB and one plot with annual SMB vs winter SMB for constant summer SMB.

Authors' response:

In response to your comment, and to be consistent with suggestions made by the other reviewer, additional information have been added into the text. Note that we decided not to add a new figure, as there are already a lot.

It hasn't been mentioned in the previous version, but the study has been done at several stakes. Thus mean and standard deviation are now mentioned. The paragraph has been re-written as follow:

"The tests (described in section 3.2.1) of the annual mass balance sensitivity to seasonal mass balance using the Crocus model were performed at seven stakes in the ablation area, ranging between 2700 m a.s.l. and 2870 m a.s.l.. For the sake of clarity, only the results for stake #10 (located at 2760 m a.s.l.) are presented in Figure 6, but conclusions are similar for all the stakes.

Regarding the sensitivity of annual SMB to summer SMB (Figure 6a), the results show that the simulated annual SMB was the least negative with 1995 summer conditions (green curve) and the most negative with 2003 summer conditions (red line). The difference in annual SMBs between these two extreme summers for the stake #10 was 4.1 m w.e. yr^{-1} at the end of the hydrological year. Similar results are found for the other stakes: the mean difference is 4.4 m w.e. yr^{-1} with a standard deviation of 0.41 m w.e. yr^{-1}.

The sensitivity of annual SMB to winter SMB is illustrated by Figure 6b. Note that for the sake of clarity, only the two extreme years of the time series (2000-2001, highest winter SMB (pink line) and 2008-2009, lowest winter SMB (blue line)) are presented in Figure 6b. The difference between these two years on 15 April is 1.2 m w.e. at stake #10 (and on average 1.1 m w.e. with a standard deviation

of 0.13 m w.e. considering all the stakes). Using the same summer conditions, the difference at the end of the hydrological year is 2.4 m w.e. (*i.e.* twice the difference at the end of the winter season). Here again results are similar for the all the stakes considered: the mean difference is 2.2 m w.e. yr^{-1} with a standard deviation of 0.21 m w.e. yr^{-1}.
The same test was performed using the extreme 2003 summer conditions instead of the mean summer conditions. In this case, the difference at the end of the hydrological year for all the stakes was considerably larger (3.4 m w.e. in mean, standard deviation 0.45 m w.e. yr^{-1}; results not shown). These results confirm that the annual SMB variability is mainly driven by the summer SMB variability (*i.e.* differences are larger when we considered a mean winter and all the summer conditions than the contrary). Nevertheless, the annual SMB appears to be very sensitive to the winter SMB, in particular for extreme years."
In addition we think that this figure shows clearly the entire range of the sensitivity, and id easy to read. We worry about confusing thing with point representation.

Specific comment #15 (page 16 - l6-9):
This part is not so clear to me. How can differences up to 25% be explained by 'only slightly affect the simulated SMB for a limited number of stakes'?
Authors' response:
We agree with your comment, this point is unclear in the current version. Indeed, the maximum is 25%, which is quite high. Nevertheless, it is only affect few stakes (maximum 5) and the mean (also mentioned in this paragraph) is much lower (and generally lower than the measurement uncertainty). Our conclusions saying that it only "slightly affect the simulated SSMB" have been drawn considering this mean. This has been re-written to make it clearer:
"The highest differences between simulations and measurements are obtained for the stakes located in the lower part of the glacier tongue, using 1998 and 2007 DEMs (*i.e.* where geometric changes are the greatest). Simulations performed with 1998 and 2007 DEMs led to a mean difference in simulated summer SMBs of 0.19 m w.e. yr^{-1} (~5% of the SSMBs) and reached 0.64 m w.e. yr^{-1} for the lowest stakes (~15% of the summer SMBs and ~20% of the annual SMBs). Simulations performed with 2007 and 2014 DEMs, led to a mean difference of 0.15 w.e. yr^{-1} (<5% of the SSMBs) and a maximum of 0.47 5 w.e. yr^{-1} for the lowest stakes. Note that the differences in simulated summer SMBs *vs.* measurements in the accumulation area are larger when considering the DEMs from 2014 and 2007 than with 1998 and 2007 DEMs and can reach 0.38 m w.e. yr-1 (~20% of the summer SMBs and ~25% of the annual SMBs). Despite changes in glacier surface topography over the entire study period, such changes only affect the simulated summer SMB (*i.e.* considering changes larger that measurement uncertainty) for a limited number of individual stakes (maximum 5). Considering the entire glacier, these changes in the simulated summer SMB are negligible as the mean is lower than the measurement uncertainty."

Specific comment #16 (page 23 - l24):
I do not really understand the conclusion that empirical models would be better fitted to model future SMB. From your results it is clear that using observations to correct forcing the energy balance model performs better than the empirical model, also for the period where no AWS is available. It is unclear whether the forcing corrections remain valid in the future, but the same holds for the parameters in the empirical model (as you have pointed out). So why is then the empirical model more reliable for future projections?
Authors' response:
Please refer to our answer to your main comment #2 and the specific comment #14.

Technical points

Technical point #1 (page 1):

Why not include Saint-Sorlin in the title, replacing the French Alps?

Authors' response:

We agree and have changed the title:

"Relative performance of empirical and physical models in assessing the seasonal and annual glacier surface mass balance of Saint Sorlin glacier (French Alps)."

Technical point #2 (page 3 - l23):

Here and elsewhere in the paper: I am not a big fan of these acronyms
WSMB, ASMB, SSMB. I feel it provides easier reading by just writing out 'winter SMB',
'annual SMB', and 'summer SMB'.

Authors' response:

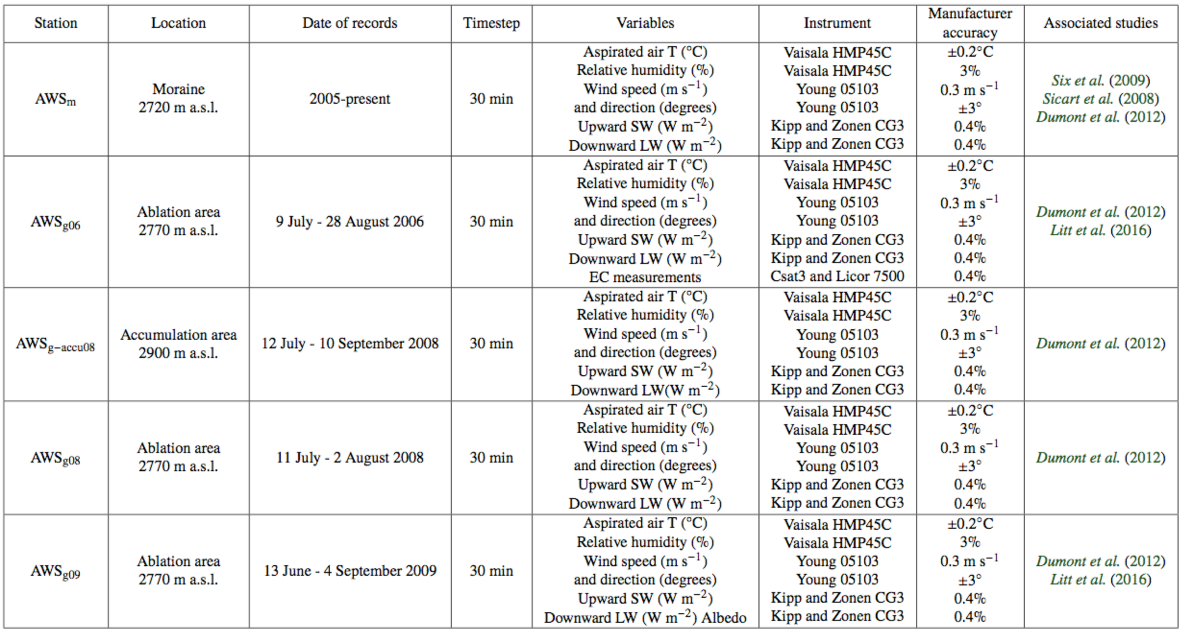
Done.

Technical point #3 (page 3 - Table1):

Add some separation between the different stations, like a line, or a blank text line.

Authors' response:

Done.



Station	Location	Date of records	Timestep	Variables	Instrument	Manufacturer accuracy	Associated studies
AWS$_m$	Moraine 2720 m a.s.l.	2005-present	30 min	Aspirated air T (°C) Relative humidity (%) Wind speed (m s^{-1}) and direction (degrees) Upward SW (W m^{-2}) Downward LW (W m^{-2})	Vaisala HMP45C Vaisala HMP45C Young 05103 Young 05103 Kipp and Zonen CG3 Kipp and Zonen CG3	±0.2°C 3% 0.3 m s^{-1} ±3° 0.4% 0.4%	*Six et al.* (2009) *Sicart et al.* (2008) *Dumont et al.* (2012)
AWS$_{g06}$	Ablation area 2770 m a.s.l.	9 July - 28 August 2006	30 min	Aspirated air T (°C) Relative humidity (%) Wind speed (m s^{-1}) and direction (degrees) Upward SW (W m^{-2}) Downward LW (W m^{-2}) EC measurements	Vaisala HMP45C Vaisala HMP45C Young 05103 Young 05103 Kipp and Zonen CG3 Kipp and Zonen CG3 Csat3 and Licor 7500	±0.2°C 3% 0.3 m s^{-1} ±3° 0.4% 0.4% 0.4%	*Dumont et al.* (2012) *Litt et al.* (2016)
AWS$_{g-accu08}$	Accumulation area 2900 m a.s.l.	12 July - 10 September 2008	30 min	Aspirated air T (°C) Relative humidity (%) Wind speed (m s^{-1}) and direction (degrees) Upward SW (W m^{-2}) Downward LW(W m^{-2})	Vaisala HMP45C Vaisala HMP45C Young 05103 Young 05103 Kipp and Zonen CG3 Kipp and Zonen CG3	±0.2°C 3% 0.3 m s^{-1} ±3° 0.4% 0.4%	*Dumont et al.* (2012)
AWS$_{g08}$	Ablation area 2770 m a.s.l.	11 July - 2 August 2008	30 min	Aspirated air T (°C) Relative humidity (%) Wind speed (m s^{-1}) and direction (degrees) Upward SW (W m^{-2}) Downward LW (W m^{-2})	Vaisala HMP45C Vaisala HMP45C Young 05103 Young 05103 Kipp and Zonen CG3 Kipp and Zonen CG3	±0.2°C 3% 0.3 m s^{-1} ±3° 0.4% 0.4%	*Dumont et al.* (2012)
AWS$_{g09}$	Ablation area 2770 m a.s.l.	13 June - 4 September 2009	30 min	Aspirated air T (°C) Relative humidity (%) Wind speed (m s^{-1}) and direction (degrees) Upward SW (W m^{-2}) Downward LW (W m^{-2}) Albedo	Vaisala HMP45C Vaisala HMP45C Young 05103 Young 05103 Kipp and Zonen CG3 Kipp and Zonen CG3	±0.2°C 3% 0.3 m s^{-1} ±3° 0.4% 0.4%	*Dumont et al.* (2012) *Litt et al.* (2016)

Technical point #4 (page 5 - l10):

Please rephrase this sentence on the instrumentation height adjustment. You probably mean that you lowered the instruments in order to keep the distance between instruments and the ice surface constant.

Authors' response:

This part has been re-written in response to your comment and the comments of the other reviewer:

"Due to ice melt, instrument heights are not constant over time. However, at each station (except for AWS$_{g08\text{-}accu}$ where melt is limited), a sonic ranger was set up and helped determine the melt over each recorded time step. The heights of the instrument were then adjusted in our simulation using the melt determined by the sonic ranger. Every 10 to 15 days, instruments were re-adjusted manually to a set height of 2 m."

Technical point #5 (page 5 - l27):
'emitted long wave radiation and reflected short wave radiation', the earth surface does not emit short wave radiation.
Authors' response:
The sentence has been re-written in response to your comment:
'…but the impact of emitted long wave radiation and reflected short wave radiation by surrounding slopes is not considered.'

Technical point #6 (page 7 - l9):
could you replace 'explained' with 'caused'? 'explained' would require a more in-depth analysis
Authors' response:
Done.

Technical point #7 (page 9 - l24):
If I understand correctly, you have changed the lower albedo limit from 0.7 to 0.5, keeping the time decay. Then I suggest to replace 'fixed at' with 'set to', as 'fixed' could indicate that you eliminated the time evolution of the albedo.
Authors' response:
You are right, 'fixed at' has been replaced by 'set to'.

Technical point #8 (page 14 - Figure 4 caption):
Replace the hyphen in 'blue (a-c)', 'orange (b-d)', etc. with a comma 'blue (a,c)'
Authors' response:
In response to your comment and the remark made by the second reviewer, caption as been changed: "**Figure 4.** Correlations between simulated (blue (a and c) for the ATI model and orange (b and d) for the Crocus model) and measured summer SMBs at each stake of Saint-Sorlin Glacier over the 2006-2015 period (a and b) and the 1996-2005 period (c and d). Circles represent measurements in the ablation area and solid dots represent measurements in the accumulation area."