Review of Mattea et al.: "*Firn changes at Colle Gnifetti revealed with a high-resolution process-based physical model approach*", by Vincent Verjans

This study uses a model-based approach to simulate changes in firn temperature at Colle Gnifetti, a region in the Swiss/Italian Alps. The authors process a meteorological dataset from a nearby weather station, corrected and complemented by using datasets from other weather stations of the region. They use an existing coupled model of Surface Energy Balance (SEB) and firn processes, which they partly re-calibrate for their site of interest. Their model simulations span the period 2003-2018 and firn temperature results are compared with 25 published measurements of firn temperature profiles. They also investigate temporal and climatic patterns in the melt rates derived from their SEB model. Based on their results, they quantify the increases in firn temperatures and surface melt amounts at Colle Gnifetti over the period of their study.

I believe that this study investigates a valuable research question and demonstrates an appropriate and relevant use of SEB and firn models in conjunction with meteorological data. Clearly, the authors have worked thoroughly on the processing of the meteorological data and on the firn simulations. They analyse in depth the important features of their results, and they make a great effort to put these into the context of previous scientific studies at their site of interest. Their conclusions are supported by their results and the manuscript is well-structured. I believe that the study will be of interest to both the firn modelling and the ice core communities. Finally, I appreciate the amount of work that went into this study. For these reasons, my review is largely positive. Nevertheless, I believe that the presentation of the calibration process is somewhat weak. Discussing the calibration more in details, in combination with a sensitivity analysis of the model parameters would bring this study to the next level. My review includes one Major and some Minor comments that I expect the authors to address in their response, and Technical comments, which are only related to the presentation of the manuscript.

Major Comment: the model calibration

I understand that the authors use the model of van Pelt et al. (2012), mostly in its original form. In Section 3 and Table 4, it is explained that many of the model parameters are calibrated using the data from Capanna Margherita (CM), the Seserjoch station and the Colle del Lys station. If I am correct, they recalibrate 6 parameters of the EBFM model ($\alpha_{fresh}, t^*_{wet}, K, b, e_{cl}, z_{lim}$).

Firstly, the authors do not explain their decision to re-calibrate these specific parameters, while leaving many others to their default values (see Table 4). Some parameter values are taken directly from the existing literature, without discussing the sensitivity of model results to this choice nor the particularities of their site of interest. I list here a few examples:

- α_{firn} is taken directly from van Pelt and Kohler (2015), which focuses on Svalbard glaciers.
- z_0 is taken directly from Suter et al. (2004), where this parameter value is not discussed. Given the importance of the turbulent fluxes in the SEB (see Fig. 8), I suppose that model results would be quite sensitive to this parameter value.
- According to van Pelt et al. (2012), the formulation of t_a depends on the geographic location, and its parameterisation has a strong impact on results from the EBFM. Yet, the authors do not address the formulation of t_a, and I saw in the model code that the authors keep the same formulation and parameterisation for t_a as in the Svalbard study of van Pelt et al. (2012).
- Similarly, the temperature threshold between solid and liquid precipitation is not addressed and, from the model code, I noticed that it is taken from van Pelt et al. (2019), which is another Svalbard-specific study. The air temperature is regularly above 0°C in summer at CM (Figure 2), and I thus expect this parameter to be influential.

Secondly, the calibration method for the recalibrated parameters should be better explained than the simple statement "*Tuned from*" (Table 4). Calibrating 6 model parameters implies that the number of degrees of freedom in the calibration is high. How do the authors account for potential interactions between some of the parameters? And how do they reach their final recalibrated values?

Thirdly, the parameter z_{lim} is calibrated to 20 m firn temperature, and its final value is 4 m. Thus, the influence of the z_{lim} parameter at 20 m depth will be outweighed by the influence of the thermal conductivity parameterisation. For this reason, I find the relevance of tuning it to 20 m temperature questionable. Why not tune it to temperatures at shallower depths (e.g. 5 or 10 m)?

Finally, Table 4 mentions that some parameters were "*verified with CM AWS data*". What does that mean? Is there any quantifiable evaluation of the verification?

As written in the introduction of this review, I believe that a more thorough sensitivity analysis to parameter values would bring the study to a next level. The application of the EBFM in many Svalbard-specific studies has been particularly successful because of the robust calibration work that was carried out in each of these studies. I understand that a full sensitivity analysis would require a considerable amount of work. Thus, at least, I recommend that the authors provide:

(1) More details about the calibration method used for the parameters that they do calibrate. This could be included in an additional section in the Appendix for example.
(2) A discussion about the limitations related to the absence of calibration for other important parameters (see above).

Minor Comments

1) The spatial contextualisation

The manuscript is well-written and mostly easy to follow. However, I had difficulties with the numerous mentions to specific locations, which are not clearly identified and not all are shown on a map. Clearly, the authors are familiar with the region of Colle Gnifetti, but they should keep in mind that most of the readers are not. For this reason, I suggest a few possibilities to improve the spatial information.

First, the authors should put all the locations referred to on the Fig. 1a map. Some of the locations that can be added are the Grenzgletscher slope, the KCC core location and the Colle de Lys station (if applicable). I also recommend expanding the legend of Fig. 1a instead of splitting the information between the legend and the caption. Finally, in the discussion of their results, the authors use terms such as CG, the saddle point, and the CG saddle to designate different things (the location of the CG grid cell or the larger area of the saddle). I believe that using CG only for the CG grid cell and using a term such as "the saddle area" would make the text less confusing.

2) Quantification

Statements in the main text often lack a quantitative support. This is an important point, especially in the Discussion section. I list some examples here:

- line 63: "*significant interannual variability*", the author could quantify the variability
- Table 3: provide also annual mean values and standard deviations, as bases of comparison for the RMSEs
- line 165: quantify the threshold required for the "*last significant snowfall*"
- Table 5: can the authors give between brackets the number of profiles considered for the "*CG only*" and "*All*" evaluations? Also, I do not understand how they compute the RMSE and Bias statistics. Do they consider all depth levels of the measured profiles? If so, how many temperature measurements are considered per core? I have the exact same questions concerning the residuals shown in Figure 5.
- line 298: "*the depth of zero annual temperature oscillation, at about 20 m*". This cannot be evaluated by the reader due to the large temperature range shown in Fig. 6, thus a quantitative metric should be given. I recommend, for example, giving the shallowest depth from which the mean interannual temperature oscillations at CG, SK and ZS are below 1°C.
- line 307: "*appears to correlate*", the author could quantify the correlation
- line 321: the authors should give the uncertainty interval on their trend. And the units should be °C yr^{-1}.
- line 382: "*approximately 19 and 25 cm*", why not provide the ranges of annual values through the simulation? I believe this would be more relevant because the values are compared to observed ranges.
- line 390: "*values in excess of 1000 W m^{-2} are a common summer occurrence*", provide mean number of hours (or days) per year of such occurrences.
- line 406: "*significant melt happening at negative temperatures*", provide mean annual melt occurring at negative air temperature (in mm w.e. yr^{-1}) and/or mean fraction of the total melt occurring at negative air temperature.

3) The meteorological data processing.

The link to Mattea (2020) in the references leads to a website that cannot be accessed. As such, one cannot have the full details about the data processing method. I believe that the authors made the processing properly, but for the sake of scientific openness, all the details of the method should be available. I suggest two possible options: (1) the authors add an appendix explaining the method in details, or (2) the authors add a statement in the Data availability section guaranteeing that further details about the weather data processing is available upon request. These options are suggestions, and I believe that the editor has the final say on such issues.

4) The temperature and pressure lapse rates.

If I understand correctly, these lapse rates are calculated over a large elevation range (~2000 m according to the elevations given in Table 2). The same lapse rates are then used over the model domain, where the elevation range is much narrower. Is it realistic to assume same lapse rate values over two ranges of elevation that are so different?

5) The extrapolation of climatic variables
The gridding of climatic forcing is not fully explained. The precipitation model is clearly detailed, and the temperature and pressure fields are adjusted via the lapse rates. However, the model must take several other climatic fields as inputs (e.g. wind speed, relative humidity). How are these calculated over the entire model domain?

6) Equation (18)
Some information about the location of the KCC core is needed (see Minor Comment 1). Is it reasonable to relate accumulation anomaly from the KCC core to wind speed at CM? Can the authors provide an intuitive, physical interpretation on why higher median wind speeds at CM should be linked to lower accumulation (maybe a link with wind scouring)?
Also, I could not find the accumulation anomaly data from Bohleber et al. (2018). How did the authors get this information? If it is through personal communication from Bohleber et al., it should be specified in the manuscript.

7) Discussion of the cold bias (lines 307-317).
The authors conclude that the cold bias of the model is due to dense, thick refrozen firn layers generated at the surface, due to the parameterised percolation, that block further infiltration and latent heat release. While this may have an impact, I think that there is a more important factor at play. If summer accumulation is underestimated, this leads to an underestimation of heat advection in the modelled firn column. Firn layers are deposited at the surface temperature of the time step. Subsequently, they are buried into the firn column, carrying this temperature signature towards greater depth. If summer snowfall events are underestimated, the amount of heat transported towards depth in this way is greatly underestimated. In my view, this could be the primary cause of the cold bias, and I would welcome the opinion of the authors about this thought.

8) Calculation of refrozen ice fraction in the *unifr-2019* core.
The authors compare their value of 31 cm of ice layers to melt amounts. Do they account for the fact that 31 cm of ice layers is not equivalent to 31 cm ice equivalent of refreezing? Meltwater refreezes in firn that has a density >0 kg m^{-3}. As such, multiplying the ice layer thickness by the ice density does not give the amount of refrozen water. Also, I note that there is no data between 3.7 and 3.9 metres depth, which seems to be an ice rich section of the core. I believe that the authors should mention this in their discussion in section 5.3.

Technical Comments
line 2
Add comma after "*Thus*".
line 26
I believe that "climatic archives" would be more appropriate than "*atmospheric archives*".
line 27
Change "*Beside*" to "Besides".
line 27
I find the wording "*mass losses brought upon glaciers by*" strange. I suggest "glacier mass losses caused by".
line 32
Why "*naturally*"?
line 35
Change "*Then*" to "Thus,".
line 42
Provide date range of the Pleistocene.
line 42
Change "*more*" to further back".
line 50
Because the study of Haeberli and Funk (1991) is quite old, specify the date range over which the steady state conditions were observed.
line 53
Change "*exposing*" to "highlighting".
line 61

Add comma after "*Thus*".
line 68
Change "*independent on*" to "independent of".
Table 1
Change "*CM*" to "Capanna Margherita".
Table 2
Is it possible to add a column ΔT_{CM} (difference in annual mean temperature with respect to CM)?
Table 3 and Table 5
Change "*RMS*" to "RMSE" (root mean square in itself is something different).
Figure 2
I suggest showing the daily mean values rather than the hourly values. I think it would give a better picture of the short-term variability, but I leave this choice to the authors.
line 121
Change "*Beside*" to "Besides".
line 148
I believe it is more relevant to refer the reader to van Pelt et al. (2012), which includes the model equations.
line 170
I believe that "$max(T, T_{max,t^*})$" should be $|max(T, T_{max,t^*})|$ because temperature is expressed in °C.
line 172
Make sure to use T only for a single variable (it is used for air temperature in the rest of the manuscript, and not snow temperature).
Table 4
The notation b is used for two different parameters. And what is Q_{ground} used for in the model?
line 180
Is vapor pressure the same as relative humidity? If so, I recommend sticking to the same wording as in the rest of the manuscript (e.g. Table 2 and Figure 2).
line 180
The variable n was already defined.
line 180
The constant σ should be defined as the Stefan–Boltzmann constant.
line 184
Change "*independent on*" to "independent of".
lines 217-218
"*Accumulation measured at each stake over a single year was re-scaled to a mean annual estimate by using the overlaps with firn cores and GPR points.*" This is not clear to me.
line 246
I am not sure that the use of the word "*conspicuous*" is appropriate here.
line 255
Change "*shows*" to "show".
line 264
"*Across the CG saddle*": does that mean over the entire domain (see Minor Comment 1)?
line 272
Specify that "*10%*" refers to a percentage of the sum of all the absolute energy fluxes.
lines 273-274
But the NE domain region is also where the melt amounts are lowest (see Fig. 7). Are the authors certain that it is this region where melt corresponds to the highest fraction of net accumulation?
lines 282-283
"*micro-melt events under 4 mm w.e.*": specify "under 4 mm w.e. in a single day".
lines 281-283
In my opinion, it would be interesting to investigate whether the importance of micro-melt events in the total melt amount tends to increase/decrease over time. For example, the authors could provide the trend in the ratio Total melt from melt events below 4 mm w.e. per day divided by Total melt. This is only a suggestion.
line 284
Change "*relationship*" to "relationships".
lines 285-286

"*unlike cloud cover which appears to have almost no effect*". As far as I understand, this contradicts the next sentence and Fig. 11b. It seems clear to me that low cloud cover values are associated with higher melt amounts. If the authors refer only to the melt rates, then the sentence should be clarified.
line 289
Why "*long-term*"?
Figure 8
The term "*GHF*" is not defined. And I believe that the surface fluxes at CG could also be shown.
line 310
Specify "the accumulation model (Eq. (17))".
line 313
Specify "based on weather station measurements from lower elevations".
lines 313-314
I suggest rephrasing: "Thus, we expect an underestimation of the strong seasonal gradient that favours summer accumulation."
lines 321-322
The trend estimate should have units °C yr^{-1} (see Minor Comment 2).
line 325
I think that "*assumed*" should be changed to "measured".
line 334
Change "*refreezing heat release*" to "latent heat release".
line 335
Change "*correspond*" to "corresponds".
lines 339-340
I do not agree that the increase in percolation depth through the melting season is necessarily "*obvious*". For example, ice lenses could form and hinder future percolation of surface meltwater.
line 369
I recommend changing "*With daily melt amounts often close to the size of single crystals*" to "Because ice-equivalent thicknesses of daily melt amounts are often of the same order as the size of single crystals".
line 381
Is it necessary to provide yet other location names (see Minor Comment 1)?
line 386
Change "*sub-freezing temperatures*" to "sub-freezing 2m air temperatures".
line 396
"*This work marked*": maybe use present tense here.
lines 399-400
If the authors discuss the applicability of the EBFM to different scenarios, they must mention the limitations of the meltwater percolation scheme and of the highly site-specific calibration procedure.
line 405
Change "*also*" to "and".
line 406
Change "*negative temperatures*" to "negative 2m air temperatures".
line 409
I suggest changing "*on the edge of statistical significance*" to "close to statistical significance despite high inter-annual variability and the brevity of the time series".
line 410
Change "*site suitability*" to "suitability of the site".
line 423
Change "*first*" to "shallowest".