Firn changes at Colle Gnifetti revealed with a high-resolution process-based physical model approach
Enrico Mattea *et al.*

Response to final remarks from the editor

I went through the last version of the manuscript, and found the story to be very clear and well presented. The figures nicely support your findings – thank you also for incorporating the initial suggestions that I made, e.g. on Figure 2 – and I appreciate the 'honest' way in which your results are presented. The model capabilities are highlighted, but you are not 'hiding' what does not work and give an elaborate explanation concerning the possible discrepancies. I am convinced that some of your main findings (e.g. the role of micro-melt events) will be of large relevance to scientists working on firn modelling and those interested in deriving palaeoclimatic information from ice cores on alpine glaciers.

We wish to sincerely thank the editor and the referee for the positive comments on our manuscript. We have incorporated most of the suggested changes; exceptions are discussed below. Editor comments are reported in *black italic*, while our responses are in blue. Moreover, in the manuscript we have now updated the DOI link to the GitHub repository with the model code: we have cleaned up the code in terms of commented lines and indentation to make it easier to read (but without any functional change).

At this stage, I have formulated a list of final remarks and suggestions that I would like you to address when uploading your (most likely final) version of your manuscript. These are mainly easy-to-incorporate changes, although some may require a small amount of work:

- *l.2: "...need to improve understanding and further develop" → "...need to improve our understanding and to further develop"*
  Done.

- *l.37: order the references in chronological order*
  Done.

- *l.102: "...the coupled model in…" → "...the coupled model is described in…"*
  Done.

- *l.114-116: here you describe how other station data is used to perform quality checks, to fill gaps in time series and to determine parameters that were not measured at CM. This seems like quite a lot of work, which must not have been trivial. A few questions here:*
  - *Do I understand it correctly that you have performed this work yourself? If so, would be good to state this more explicitly*
    Indeed, we have performed this (quite time-consuming) work ourselves. We have updated the text to state this more clearly.

  - *Can you say something about the "quality checks"? Potentially give some numbers*
    In the manuscript, we provided a simple overview of our workflow for quality checks: "automated pre-filtering routine […] based on objective criteria (absolute values, rates of change, comparison with reconstructed series and reanalysis). […] potential outliers […] were then manually checked. The CM AWS was always processed last […]". We have now added a new table (Table 3) with the numbers and fractions of missing values/gaps

for each measured meteorological parameter, before and after the quality check of the CM series. For a more detailed (technical) description, the interested reader is directed (l.138-139) to the full description in Mattea (2020).

  - *The new dataset you have created through your approach seems to be quite unique–it is definitely than what is directly available from the data providers. As this dataset would be quite valuable for other researchers, could you make it directly available? So far you mention "the meteorological time series...should be requested from the respective providers"in the "code and data availability section": but what you provide here goes "beyond"this.*
    We have contacted all data providers, asking about this possibility. Unfortunately, MeteoSwiss policy explicitly prohibits users from providing a public download link to anything which is (partly) based on their paid data/services (in our case, the hourly series of Gornergrat and Monte Rosa-Plattje, which we used to validate and complete the CM AWS data). Thus, we are allowed to provide our processed series only to a third-party who has already obtained access to the original datasets by the respective providers.

- *Table 1: what is CAE? Please define*
  CAE is the name of the company which has supplied the CM AWS sensors. Apparently it is not an acronym (https://www.cae.it/eng/), so at present we do not have additional information which could be added to the table. An internet search for e.g. "CAE" "VV20" immediately provides the sensor data-sheet, but we welcome any suggestion as to possible rewording within the table.

- *l.129: "...within the model input", maybe rephrase to "...as model input"?*
  Done.

- *l.149: "...18 boreholes, some locations having been measured…" → "18 boreholes, where some locations have been measured…"*
  Done.

- *l.184-186: "...corrected for...calibration parameters.". I got a bit lost in this sentence and could not understand 'what is what'. Could you rephrase this, potentially by splitting the explanation in two different sentences?*
  Done.

- *Table 4: very nice and useful for the reader! In the caption, you mention "Additional parameters not listed here were kept at the default value". It would probably be useful for the reader to also have this information directly at hand, without having to dive into the three studies that you mention (and having to look for 'what is where'). Could you add this information in a separate table (e.g. in suppl. Mat.)?*
  We have added a table with the additional parameters, their value, unit and explanation in the supplementary material.

- *l.313: you initialize by running the model over the time period 2004-2011 for 8 times. Two questions here:*
  - *Why did you decide to go for 8 times, and for instance not 4 times, or 10 times? Would be good if you could give a hint.*
    We opted for 8 times because the resulting 64-year spin-up series is just long enough (with a small margin) for the whole grid to reach temperature and density equilibrium with the surface forcing, up to a depth of 20 m over the entire domain. Specifically, after the 64-year spin-up the entire grid down to a depth of at least 20 m consists of snow and

firn layers which have been added by snowfall during the spin-up period (moving layers scheme: Sect. 3.4). We also experimented with shorter spin-up runs, where we observed a temperature and density transient at depth, most notable on the grid cells with lowest accumulation (Fig. 3a). An even longer spin-up series does not provide any advantage (within the first 20 m of the domain) and would be very time consuming, since model spin-up is needed after each parameter change (thus also for sensitivity experiments), to avoid transient adjustments. We have added a brief mention of the role of spin-up duration to the "Model initialization" section.

- *How sensitive are you results to running this time period 8 times? Could you include a short analysis on this, for instance in Appendix B?*
  We have tested sensitivity with two model runs, initialized using 4x8 and 12x8 spin-up loops instead of the 8x8 used for the main run. With the longer spin-up, the model output is basically the same (due to periodic forcing at the surface, which produces periodic conditions within the firn: the temperature change is within 0.02 °C everywhere and typically one order of magnitude smaller). The shorter spin-up run results in a firn temperature and density transient at the beginning of the actual model run, within the depth region which has not yet been reached by the newly accumulated snow layers. Such a transient is strongly affected by the choice of artificial starting conditions imposed at the beginning of the spin-up run. By contrast, these artificial conditions have no effect on the actual initial conditions of the main simulation (thanks to the complete 8x8 spin-up run, which forces a complete replacement of firn layers within the first 20 m): thus, we think it would not be very informative to quantify (one arbitrary example of) this transient adjustment. Section "3.5 Model initialization" now explains that the used spin-up allows complete adjustment of the grid to the surface forcing.

- *l.315: you mention '20 m / 1h'and '100 m / 3h'as spatio-temporal resolution. I found this quite confusing at first, as this made me think that you change two things at once: the spatial resolution and the temporal resolution: how can you then discern the effect of both on your results? Subsequently, I kind of understood that you probably change the time resolution to ensure numerical stability (correct?). My question here: would the results differ if you would run with '100 m / 1h'vs. '100 m / 3h'(where the former is obviously computationally more expensive). I would appreciate it if you could provide some information on this, and potentially reconsider reformulating this to 20 m vs. 100 m, and only mention the time resolution separately, indicating that this is changed to ensure numerical stability (if this would the case of course, I'd gladly be corrected here, but I am trying to take away any possible source of confusion ☺).*
  During exploratory work we performed some tests at 3 h / 20 m, and we did not find a significant difference compared to the higher time resolution (1 h / 20 m). For our analyses we used the 20 m / 1 h results as often as possible (since some of the shorter and less intense micro-melt events could in principle be lost within a 3 h run; also, surface slope is more accurate on a 20 m grid). For the examination of the full sub-surface grids, the data volume of the model output was simply too large to process and analyze (our 55x70x250 grid on 140256 hourly time-steps produces about 135 billion values for each sub-surface variable), so for example we show the 100 m / 3 h result in Fig. 6.

- *l.355-356: "in every month sublimation is a more effective energy sink than melt": ok, nice. Probably not that surprising of a finding for specialists in the field, but was for me at when reading the sentence at first. Maybe also consider mentioning this explicitly in your conclusion?*
  Done.

- *l.360: "with only minor amounts...": could you quantify this statement? Would be useful, as is difficult to visually derive from figure 9*
  Done. The "minor amounts" of April and October together reach up to 1 % of the annual melt totals at ZS (less at lower-melt locations, where melt is even less frequent in these two months).

- *l.364: "... it becomes statistically significant over the rest of the period": maybe good to be more specific here directly–to have info without having to refer to figure explicitly. i.e. "... it becomes statistically significant over the rest of the period ($p < 0.05$ for 2004-2018)"*
  Done.

- *l.370: "The majority of melt happens under clear sky conditions...". This contrasts with the previous sentence, from which I had derived that the cloudiness / sky conditions do not play a big role ("...unlike cloud cover which appears to have almost no effect"). Maybe consider slightly rewording? Could potentially remove the clear sky info and just focus on the "slightly positive temperatures"?*
  Cloud cover appears to have almost no effect on melt rates (i.e. the melt intensity, in mm w.e. h^{-1}), but the total amounts also depend on the frequency of occurrence of each set of meteorological conditions, and especially on the correlation between variables, such as warm temperatures and clear skies. We believe that it is important to convey both findings: (1) that cloud cover does not strongly affect the intensity of a melt event (thus for example melt duration might be investigated as a proxy for total melt amount within a specific melt event, even with limited or absent cloud cover information); and (2) that it is under clear skies and slightly positive air temperatures that the majority of meltwater and refrozen ice are produced at our site (and not, for example, only during extreme heat waves with air temperatures above 5 °C). We have reworded the sentence, adding a mention of the warm air/clear skies correlation, to clarify the different behavior of melt rates and total melt amounts.

- *l. 375: "...mean melt rates, slightly decreasing the likelihood of melt under high winds": could you add a sentence on why the likelihood is decreasing in this case?*
  Done. The effect appears to happen only at air temperatures between -5 and 0 °C, thus it is likely related to turbulent losses.

- *l.392-400: you explain how the density does not increase as long as melt-refreezing occurs at same location, and how your model is not accounting for this. Could you provide a hint somewhere about how this could be solved? Not suggesting that this needs to be changed, but would be good to provide a possible solution (like you do for some of the other limitations that you nicely put forward!)*
  The issue could be addressed with a dependence of percolation depths on melt rates/amounts, possibly in the form of a threshold for the occurrence of deep preferential percolation. With this refinement, the very small meltwater amounts which in reality undergo repeated melt-freeze cycles would no longer be allowed to percolate to 4 m in the model: they would remain close to the surface where they could melt again within a later melt event, avoiding unrealistic density increases. We are adding a summary of these considerations to the section.

- *l.415: maybe reword to: "...be affected by the lack of SW radiation reflected from the surrounding terrain in the modeled SEB"*
  Done.

- *l.442-443: “temperatures were initialized with repeated model runs over 2004-2011”: see related question earlier. If you would have repeated this more (or less) than 8 times, how would this have affected your firn warming?*
  See answer to earlier question.

- *“Conclusions and outlook”: nice overview of your study! It would maybe be useful, for someone who has a quick look at the paper and directly looks at your conclusions, to state here which data is used for calibration/tuning, and which one for evaluation: e.g. for external reader it is not clear whether the fact that the firn temperatures are reproduced reasonably well is a result from the fact that you tune to this (i.e. calibration) or that this is just an outcome of your model without specific tuning to this (i.e. evaluation). Would be useful to shortly say something about this.*
  We have now mentioned in the conclusions that we tune the model to a single deep value at the saddle point, and evaluate the model over 25 profiles distributed over the domain.

- *l.555: ‘code and data availability’: would really be nice if you could also provide the processed series for GC directly, where the data gaps are filled and non-measured variables are derived from other station data (see comment on l.114-116)*
  See answer to comment on l.114-116.

- *Please acknowledge the three reviewers (Adrien Gilbert, Vincent Verjans and one anonymous) in the ‘Acknowledgments’ section*
  Done.