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

Response to reviewer 1 (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.

We would like to thank the reviewer for the constructive and thorough review of our manuscript. Below, we provide point-by-point answers to the major and minor comments, as well as to selected technical comments. The review text is reported in *black italic*, while our responses are in blue. Figures in this response are labeled with Roman numerals to distinguish them from figures in the manuscript. Updated manuscript figures are shown at the end of the document.

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 (α_{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.

We made the choice of which parameters to re-calibrate based on the availability of local high-altitude measurements, relevance of the parameter for our site, and simplicity of verification within the model result.

Of the 6 re-calibrated parameters mentioned by the reviewer, the first 5 belong to the albedo and LW radiation routines. We made the choice to re-calibrate them because the high-resolution radiation measurements from Colle del Lys and Seserjoch (both above 4000 m and within 2 km from our site) were available for a direct comparison.

From both the Seserjoch radiation measurements and our field experience, we can affirm that exposed firn is never observed at our high-alpine site, so that local measured values of firn albedo are both not available and less relevant for the calibration. For the time-scales of albedo decay, a dry snow surface at 0 °C (corresponding to parameter t^{*}_{dry}, which we kept at the default value) is an extremely rare occurrence at our site. By contrast, both melting conditions and very negative surface temperatures are far more common, allowing a robust comparison between the measurements and the albedo model (see below for the method): therefore we re-calibrated them.

- 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.

The surface roughness length z_0 is poorly constrained at our site, due to the frequent scouring by extreme winds which alter the snow surface. As stated in Suter *et al.* (2004), “the surface roughness length for wind [...] was determined as 0.001 m from the [wind] profile measurements”; such measurements were performed at Seserjoch, under similar conditions to those found at Colle Gnifetti. The value of 1 mm is also within the ranges found by Essery and Etchevers (2004), who introduced the formulation of turbulent fluxes used in our study. For comparison, Gilbert *et al*. (2014) reached a calibrated value of 4 mm on alpine cold firn at 4250 m a.s.l., in a similar setting in the Mont Blanc range; the review of Brock *et al*. (2006) reports values from 0.1 to 50 mm over snow and ice of mid- and low-latitude glaciers.

We have performed two new model runs to assess the sensitivity of our simulation to surface roughness length, by changing the parameter value from 1 mm to respectively 10 mm and 0.1 mm. The resulting changes in deep firn temperature (Fig. I) are within [-1.6, +1.2] °C, with an overall firn cooling when the roughness length is increased (Fig. Ia) and vice-versa (Fig. Ib). At shallower depths, temperature deviations reach slightly larger values and follow a clear annual cycle, with a strong peak in late summer: this indicates that a change in simulated melt is likely driving the observed firn temperature changes. Indeed, the corresponding distributions of energy fluxes (Fig. II-III) clearly show an inverse dependence of simulated melt amounts on the chosen value of roughness length. In the case of low z_0 (= 0.1 mm), some melting is simulated at ZS over as many as 8 months per year (March to October), a result which is not supported by any evidence. Given the already high melt values simulated by the model (Sect. 5.3), we believe that this analysis points to a lower bound on the roughness length value in our simulation.

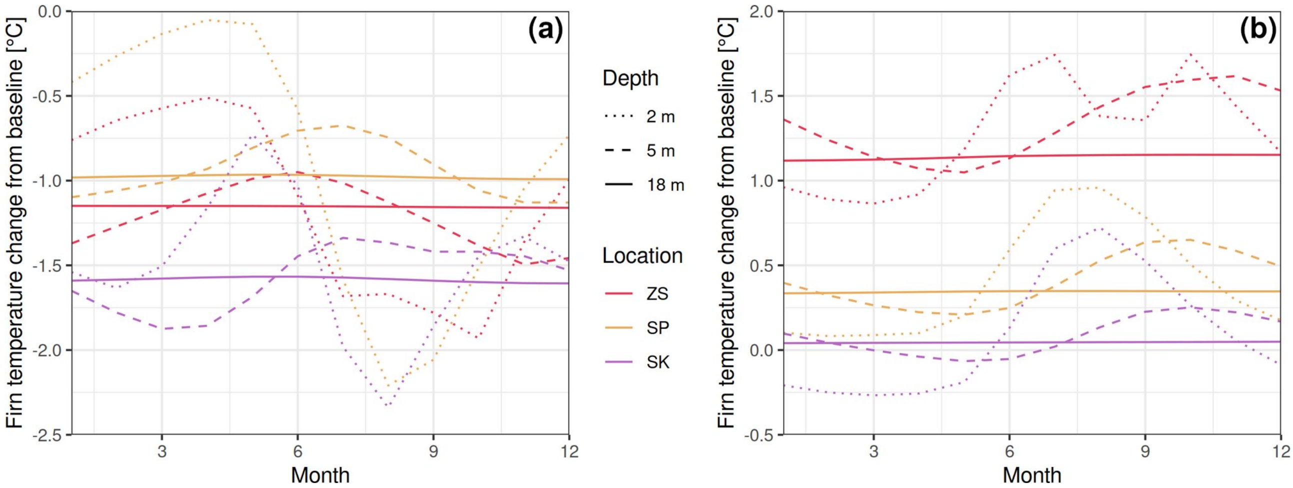


Figure I: *modelled firn temperature change by month, depth and location, after a change of roughness length from 1 mm to* ***(a)*** *10 mm and* ***(b)*** *0.1 mm.*

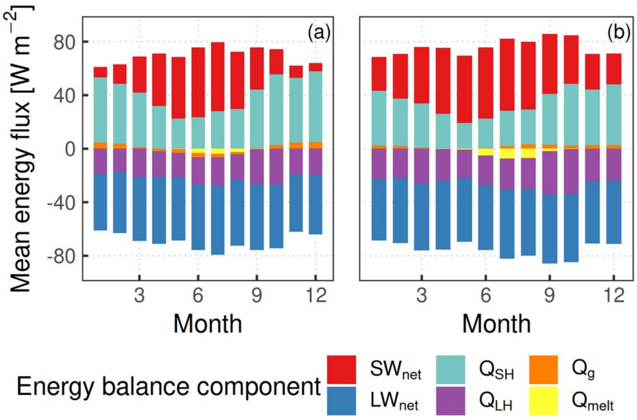


Figure II: *mean 2003-2018 monthly distribution of modelled energy fluxes at* ***(a)*** *SK,* ***(b)*** *ZS, with a roughness length value of 10 mm.*

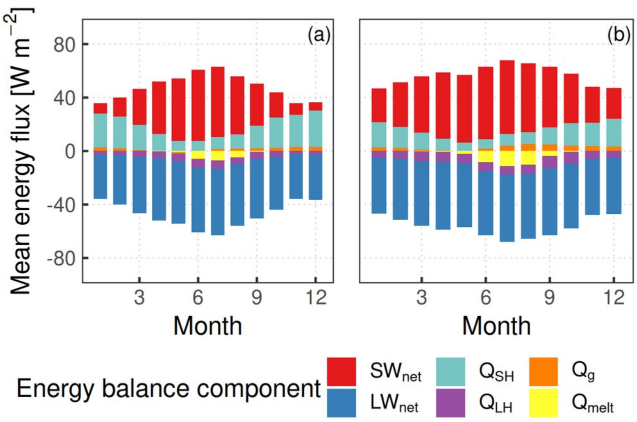


Figure III: *mean 2003-2018 monthly distribution of modelled energy fluxes at* ***(a)*** *SK,* ***(b)*** *ZS, with a roughness length value of 0.1 mm.*

Better constraining the snow roughness length at the wind-scoured CG site will be an important advancement. Model development on this front could include a time dependence of this variable, which in snow can span more than one order of magnitude over a single season (Brock *et al.*, 2006). In the revised manuscript we include a discussion of the limitations of having a constant and poorly constrained roughness length 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).

Calibration of t_a (as described by van Pelt *et al.*, 2012) may not be applicable to our case because we used the entire SW radiation series measured at CM to reconstruct the cloudiness series, by inverting the EBFM transmissivity routine (lines 110-111 of the original manuscript). Since the aerosol transmissivity t_a was already part of the equation in this reconstruction, its formulation could not be simultaneously tuned from our series. Moreover, such a tuning would not improve the SW simulation, which in our study is rather a re-computation of the measured SW values (with the model additions of gridding and topographic shading). The same applies to the gaseous and water vapour transmissivities (t_{rg} and t_w). By contrast, we did examine the cloud transmissivity t_{cl} (and finally set the parameters to Greuell *et al.*, 1997, instead of van Pelt *et al.*, 2012: see the t_{cl} discussion below), because the computed cloud cover also affects the long-wave balance, which was not simply recomputed by the model but actually simulated (Eqs. 7-9).

- 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.

Air temperatures in our 16-year hourly series are above 0 (1) °C during 2.6 (1.6) % of all time-steps. Likely due to a correlation between warm air temperatures and sunny conditions, the corresponding precipitation amounts are just 1.2 (0.7) % of the 16-year totals. Thus we expected that the rain/snow temperature threshold would have a minor impact on the overall energy balance. To quantify model sensitivity to this parameter, we have performed two new model runs, with changes of the threshold from the default 0.6 °C to respectively 1.2 and 0.0 °C. The corresponding firn temperature changes (Fig. IV) are very small: below the depth of annual variation, we observe at most ±0.13 °C at the high-accumulation ZS site, and one order of magnitude less at the two drier SK and SP (saddle point – see answer to minor comment 1). At shallower depth, deviations reach up to 0.18 °C, with a small seasonal cycle. This confirms the minor role of the rain/snow threshold at our high-altitude site. The parameter could become more relevant in the future as positive air temperatures become more common within a warming atmosphere (lines 321-322).

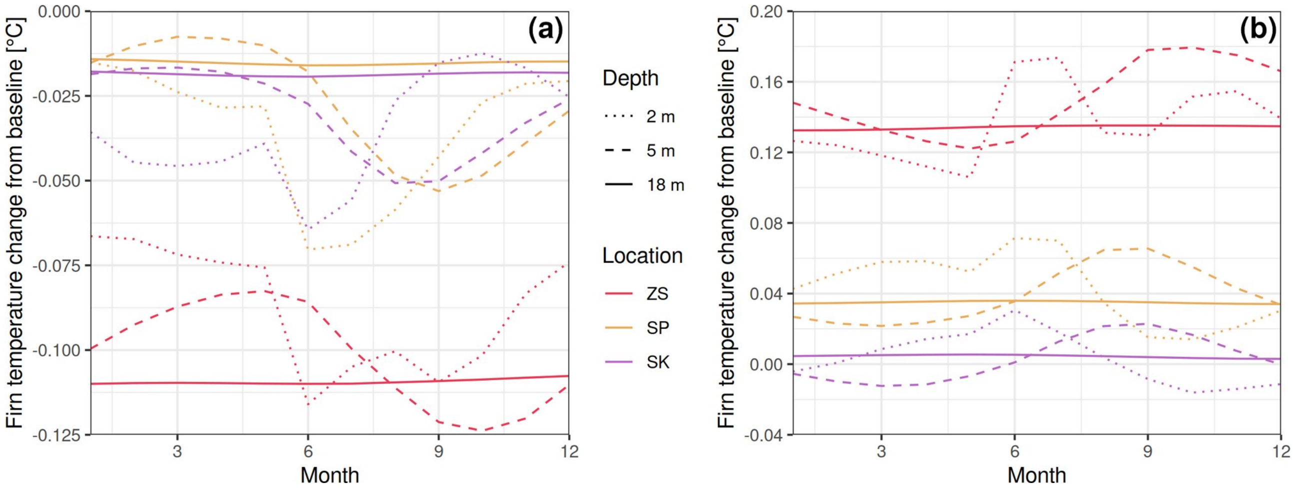


Figure IV: *modelled firn temperature change by month, depth and location, following a change in the rain/snow threshold temperature from 0.6 °C to **(a)** 1.2 °C, **(b)** 0.0 °C.*

In the revised manuscript, we explain our choice of re-calibrated parameters, including a summary of the new sensitivity results presented above.

Secondly, the calibration method for the re-calibrated 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?

For the 5 radiation parameters, we calibrated the radiation routines independently of the full model runs, using the measurements of Colle del Lys and Seserjoch. For the albedo routine, considering 10-minute measurements of albedo and snow surface temperature, we reproduced their series using the EBFM equations, tuning the α_{fresh}, t^{*}_{wet} and K values until we reached a satisfactory match (cancelling the bias and minimizing the RMSE). For the LW routine (parameters b and e_{cl}) we did the same over the corresponding measurements. The last parameter (z_{lim}) belongs to the sub-surface routine and has no effect on radiation; we tuned it until the 20 m temperatures simulated at the saddle point matched the measurements reported by Haeberli and Funk (1991). Thus, we believe that our calibration method – based on single routines rather than full model runs – is not affected by the interaction between parameters.
In the revised manuscript we provide this explanation of the calibration procedure.

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)?

Above 20 m depth, firn temperatures are affected by the annual cycle, thus the simulation of thermal conductivity could have an even more significant impact. For example, if the penetration rate of the temperature signal were inaccurate, calibration would be distorted due to a different phase of the temperature wave. Moreover, we would expect any long-term warming trend in the firn

to proceed from the (near-)surface towards depth; tuning to a relatively large depth corresponds to using measured values most unaffected by any such trend.
The revised manuscript presents these motivations explicitly.

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?

The two parameters in question belong to the cloud transmissivity parameterisation (Greuell *et al.*, 1997; Eq. 3 of the original manuscript). These parameters determine the fraction of clear-sky SW radiation which is transmitted by clouds, and were estimated (by both Greuell *et al.*, 1997, and van Pelt *et al.*, 2012) using a quadratic fit of cloud transmissivity versus cloud amount. In the two studies, the resulting minimum cloud transmissivity (corresponding to a cloud fraction n = 1) is respectively 0.352 and 0.526.
Direct measurements of cloud cover are not available at our site to perform a similar fit. Still, SW cloud transmissivities at CM can be estimated by computing the ratio of measured SW radiation to clear-sky radiation (itself computed with the EBFM parameterisations: van Pelt *et al.*, 2012, Sect. 4.1). The resulting distribution histogram (Fig. V, after filtering out time-steps with very low radiation) clearly shows that cloud transmissivities below 0.526 (the vertical line on the right in Fig. V) are a common occurrence at CM, thus the parameter values of van Pelt *et al.* (2012) would not explain the observed variability (over 33 % of the estimated transmissivities are below the allowed range of 0.526–1.0). We acknowledge that this is also true (but to a lesser extent) for the Greuell *et al.* (1997) minimum value of 0.352 (left line: about 18 % of the estimated transmissivities are lower). Still, it is reasonable to assume that cloud transmissivity does not have a sharp, well-defined minimum (due to the continuous variability of cloud optical thickness depending on cloud type: e.g. Greuell *et al.*, 1997), thus the simple quadratic parametrization will necessarily leave some unexplained variance in the measurements. We also observe (Fig. V) that the density of estimated transmissivities starts to decrease quite strongly below the 0.352 value. Considering the significant uncertainties involved in the estimation of cloud fraction (e.g. Silva and Souza-Echer, 2016), we believe that the SW measurements at CM support our choice of the Greuell *et al.* (1997) parameters.
The revised manuscript includes a summary of these considerations to justify the parameter choice.

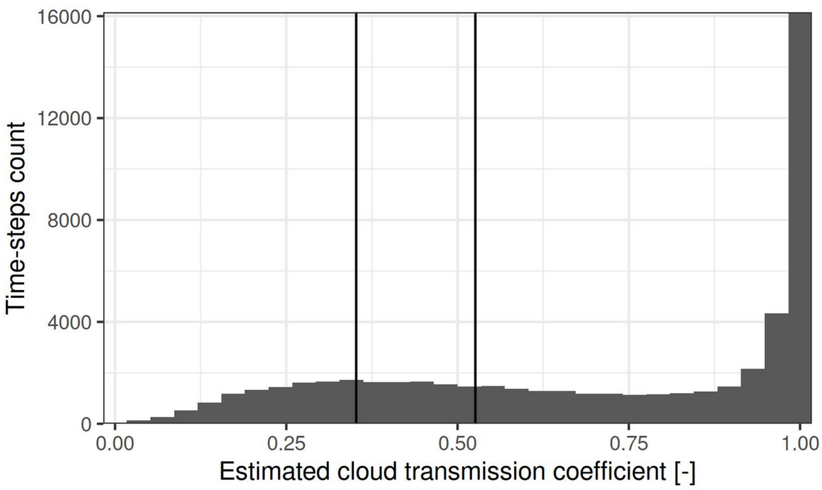


Figure V: *distribution of the estimated cloud transmissivities t_{cl} at the CM AWS (hourly series, 2003-2018). The vertical lines correspond to the lowest transmissivities allowed by the parameters of Greuell* et al. *(1997, left) and van Pelt* et al. *(2012, right).*

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).*

In the revised manuscript we add to Sect. 3 the description of our calibration choices and methods. We also include in the Appendix several sensitivity results to surface and sub-surface model parameters, with a discussion of their relevance for calibration. We hope that the changes address the reviewer's major comment.

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).

In the revised manuscript we add the Grenzgletscher slope and KCC core locations to the Fig. 1a map. The Colle del Lys station is located outside the map boundaries, and including it would make the simulation area (Colle Gnifetti) too small. The location of Colle del Lys relative to the Fig. 1a extent can now be seen in Fig. 1b.

I also recommend expanding the legend of Fig. 1a instead of splitting the information between the legend and the caption.

Done.

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.

CG is the standard abbreviation used in the literature to refer to the larger area of the saddle. In order to keep consistency, we have opted to rename the saddle point grid cell to “SP” throughout the revised manuscript.

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

In the revised manuscript we add a quantitative example of the variability (wind scouring which removes all snow in certain years, even eroding the previous year’s layer). A more detailed, numerical quantification (beyond the simple scope of the introduction) is also presented in Fig. 3a.

- Table 3: provide also annual mean values and standard deviations, as bases of comparison for the RMSEs

Done.

- line 165: quantify the threshold required for the "last significant snowfall"

Done.

- 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.

In the revised manuscript we add between brackets the numbers of profiles (respectively 19 and 25). We also clarify in the Table 5 caption the method which we used to compare measured and modelled profiles (linear interpolation of both to 1 cm resolution, then depth-averaging).

- 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.

In the revised manuscript we add quantitative metrics to the caption of Fig. 6, mentioning the depth at which oscillations vanish (that is, they are not observable against model quantization noise due to the discrete layers: 20 m), as well as the depth of a threshold amplitude of 0.1 °C (which corresponds to the typical accuracy of our firn temperature measurements; this depth is 15 m).

- line 307: "appears to correlate", the author could quantify the correlation

Done. Correlation coefficient is 0.42, which we would describe as “a moderate correlation”.

- line 321: the authors should give the uncertainty interval on their trend. And the units should be °C yr^{-1}.

Done. The new value for the air temperature trend at CM is (0.05 ± 0.03) °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.

The “observed ranges” which are reported for comparison are in fact the intervals of refrozen amounts reported by Lier (2018), as shown in Fig. VI. Rather than observed ranges, they represent reconstructed confidence intervals, obtained by comparing the core densities to the Herron-Langway dry densification model under three scenarios of the surface density parameter. Thus, they are not comparable to the mentioned *“range of annual values”* because they do not represent observed inter-annual variability, but rather confidence intervals of long-term mean values. In the revised manuscript, we mention explicitly this character of the given values.

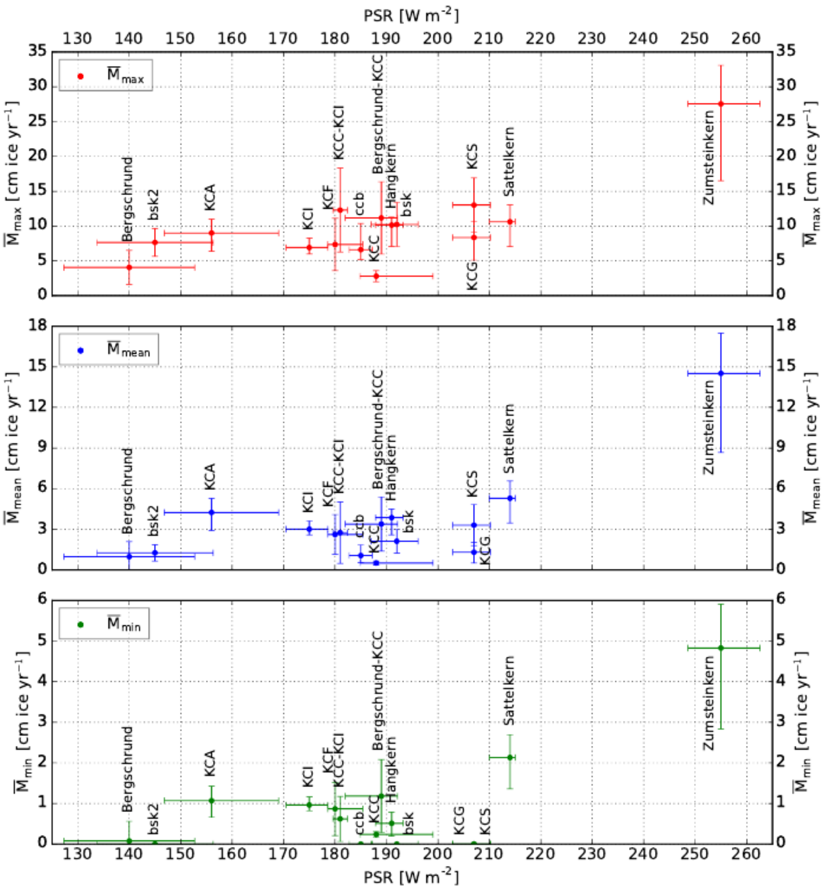


Figure 6.11: Results of the yearly amount of refrozen melt water $\overline{M}$ for all investigated firn cores as function of PSR (see section 6.1). **Top:** Maximum scenario: $\overline{M}_{max}$ with $\rho_{0,min}$ used in the H&L model. **Middle:** Mean scenario: $\overline{M}_{mean}$ with $\rho_{0,mean}$ used in the H&L model. **Bottom:** Minimum scenario: $\overline{M}_{min}$ with $\rho_{0,max}$ used in the H&L model. Note that various values of $\overline{M}_{min}$ are zero. The uncertainties of those data points are zero as well, since $\Delta\overline{M}$ is proportional to $\overline{M}$. Also note the exceptional low values of $\overline{M}$ of the KCC core in all three scenarios.

Figure VI: *refrozen amounts estimated at Colle Gnifetti from density anomalies with respect to ideal profiles of dry densification. Image and inner caption from Lier (2018).*

- 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.

Done. The mean value is 195 hours per year.

- 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.

The fractional value at the three representative grid cells of Fig. 1a amounts to 17, 22 and 34 % of the total melt amounts (respectively S-facing, flat and N-facing); we are adding these values to the results (Sect. 4.2).

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.

We agree that all the details of our methods should be made available as simply as possible. As of March 2021, the link to Mattea (2020) appears to be working normally. We will also add a statement to the *Code and data availability* section guaranteeing the availability of all details upon simple request.

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?

The elevation range for the calculation of lapse rates was usually about 1000 m (when possible, we computed lapse rates from the difference of CM to the highest stations: Stockhorn and Plateau Rosa, at 3400 m asl). We agree that such an elevation difference is still significantly larger than the ~300 m vertical extent of the simulation domain. Still, the availability of several other high-altitude stations allowed a more robust calculation of the lapse rate (computed as mean rate with respect to more than one station), as well as a quantitative estimation of the spread of such rates at each time-step. This spread – even when including stations below 3000 m a.s.l. – was usually one order of magnitude smaller than the computed rate, confirming the robustness of the calculation. Moreover, the close vicinity (in elevation) of the CM AWS to the domain, and the narrow elevation range of the domain itself, both help to minimize the impact of inaccurate lapse rate values. The revised text will include these considerations.

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?

We calculated those variables over the grid by considering their value (Fig. 2) constant over the whole domain, due to a complete lack of measurements to estimate their spatial variability. We actually experimented with simple models of atmospheric moisture variability along mountain slopes (following Feld *et al*., 2013), but eventually discarded them due to an excessive amount of out-of-range values (relative humidity far outside the [0,100] % interval on more than half of the time-steps). We recognize that the assumption of a uniform value in space is a substantial simplification; in mountain terrain, both the wind field and cloud cover are obviously affected by topography. The inclusion into the EBFM of spatialization algorithms for these variables will constitute an important development. In the revised text we describe the gridding of those variables explicitly.

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)?

We have now shown the KCC core location on the map of Fig. 1a. Wind scouring is indeed the process behind the relationship of accumulation anomaly to wind speed. This is suggested in the *Introduction* (line 60), the revised text will further clarify this point in Sect 3.3. A similar wind speed - accumulation relationship was already assumed by Suter and Hoelzle (2002, page 13) to derive a qualitative spatial "accumulation index".

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.

We computed the accumulation anomaly from the KCC core profile data (discussed in Bohleber *et al*., 2018) which were kindly provided by Dr. Josef Lier, as mentioned in the *Acknowledgements* section.

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.

Precipitation seasonality can certainly be a source of systematic under-estimation of the energy input in our model setup. In order to quantify its magnitude, we have performed a new model run with a different seasonal distribution of precipitation. A direct verification of the mechanism would have consisted in a further reduction (up to complete removal) of the already low winter precipitation amounts (last panel of Fig. 2). Unfortunately, in our model this strongly interferes with the simulation of albedo decay, because snow albedo is only reset after a certain precipitation threshold: thus, an even drier winter would lead to unrealistically low albedo values. Therefore we have opted to attempt a change in the opposite direction: for each year we have reduced precipitation by 50 % in the 6 warmest months (May-October) and redistributed those amounts over the other 6 months. The computed precipitation series has a more uniform distribution over the year (Fig. VII).

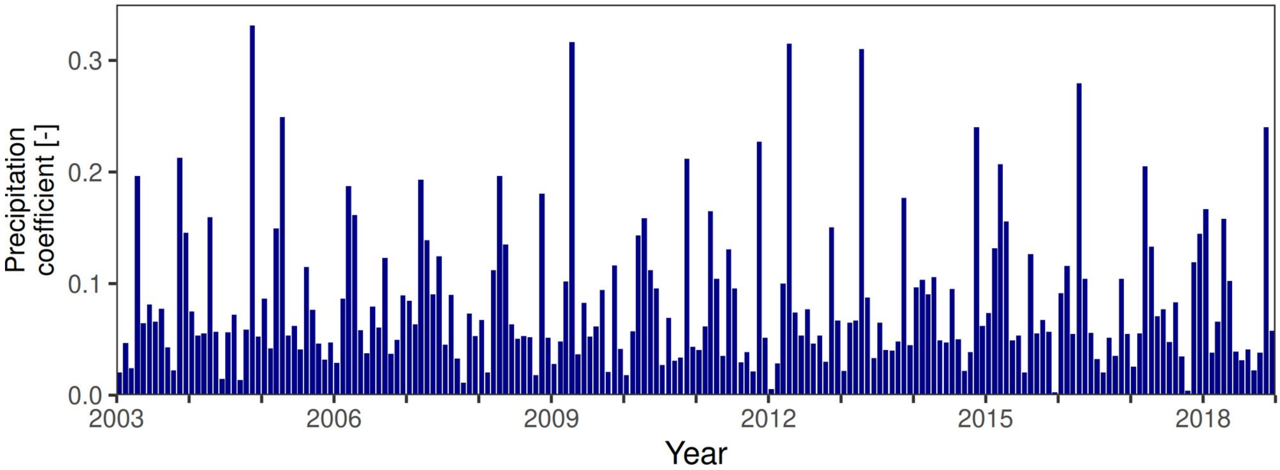


Figure VII: *precipitation coefficients (monthly means) after a 50 % reduction in May-October precipitation values, re-distributed over the rest of each year. As in the main model run, the 12 monthly values of each year add up to 1.*

The resulting changes in firn temperatures show a marked dependence on the location (Fig. VIII), and are strongly anti-correlated with the mean annual accumulation at the three sites (correlation coefficient -0.99). This suggests that a more realistic accumulation seasonality in the model – with a further reduction in winter and an increase in summer – would increase firn temperatures proportionally to the mean accumulation rates (except for the mentioned albedo issues). Since model bias is already positively correlated with the accumulation amounts, this correction could amplify the spatial pattern of the firn temperature biases. In the revised manuscript we mention precipitation seasonality as a systematic source of heat under-estimation in the model.

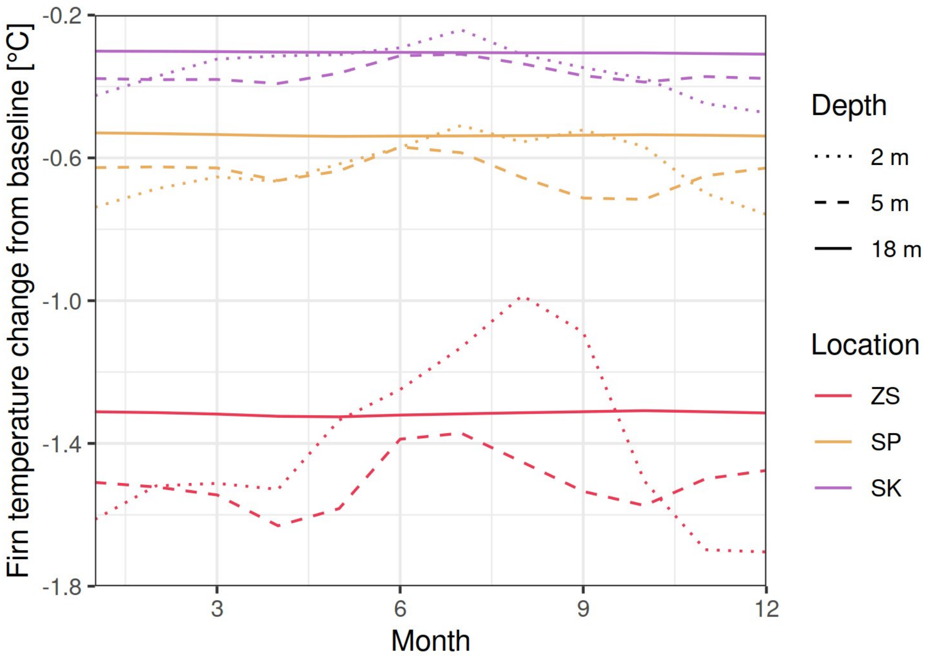


Figure VIII: *mean monthly deviation of the simulated firn temperatures (compared to the baseline) after changing the precipitation seasonality.*

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.

We fully agree, we are now making the comparison more informative by computing an adjusted estimate (14 cm w.e.) for the amount of refrozen ice in the core. We compute the correction from the mean density of the ice-free core sections. We acknowledge that such a computation involves a fair deal of uncertainty, due to the significant density variability in the ice-free sections of the profile, and also to the presence of “icy firn” which did not form well-defined ice layers. In the revised manuscript we update the discussion in both Sect. 5.3 and the Appendix. We also mention the possibility of some ice lost in the missing core section at 3.8 m depth.

Technical Comments

We wish to thank the reviewer for the careful examination of the technical aspects of our manuscript. We are implementing all recommendations from this section, except as discussed below.

line 32
Why "naturally"?

Our reasoning here is that cold firn is colder at higher elevations – hence more resilient to a warming climate.

line 42
Provide date range of the Pleistocene.

We think that the full range (starting 2,580,000 years B.P.) may not be very relevant in this context. We would instead provide the estimated age of the oldest ice at CG (19 kyr B.P.).

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.

We are not sure that we fully understand this comment. As we have mentioned in the text, steady state conditions were observed by Haeberli and Funk (1991) in a 1983 borehole profile (a single profile: we have now corrected the text from "borehole profiles"). Later, the first indications of non-steady state conditions were found by Lüthi and Funk (2001) in a 1995 profile. We believe that these dates correspond to the date range mentioned by the reviewer.

Table 1
Change "CM" to "Capanna Margherita".

The abbreviation is already introduced at the end of the *Introduction* and used in the text before Table 1: thus, we think that here the manuscript composition guidelines would prescribe the use of the abbreviation.

Table 2
Is it possible to add a column ΔT CM (difference in annual mean temperature with respect to CM)?

We have carefully considered this option, but (1) the stations cover different date ranges, thus a difference in annual mean temperature would have to be computed over inconsistent periods (sometimes very short ones) for the various stations; (2) when needed, the suggested difference can be estimated from the elevation difference which is reported in the table.

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.

At present the figure shows 140256 hourly values for each variable; daily aggregation results in 5844 values, a number which is still well beyond the resolution of the figure: we have verified that the visual change is almost undetectable.

line 148
I believe it is more relevant to refer the reader to van Pelt et al. (2012), which includes the model equations.

The study of van Pelt et al. (2012) is cited 8 lines before, at the very beginning of the section describing the model. Here, we make the point that the model version which we use is the more recent one, introduced by van Pelt *et al*. (2019).

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).

Good catch. We are renaming snow surface temperature to T_s.

Table 4
The notation b is used for two different parameters. And what is Q_{ground} used for in the model?

We are changing notation *b* for LW radiation into *c*. For Q_{ground}, as mentioned at line 144: "surface temperature and melt amounts [...], together with the lower boundary condition of geothermal heat flux, drive the sub-surface evolution."

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).

We are not sure that we understand this comment. In Eq. 9, *VP* is vapor pressure measured in Pa, as in the original formulation of Konzelmann *et al*. (1994). While a related concept, relative humidity refers to a fraction of the saturation vapor pressure at a given temperature.

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.

To extend coverage of the long-term mean accumulation measurements, we included some measurements taken at accumulation stakes (stake network of Suter and Hoelzle, 2002). These stake measurements span a single year of accumulation, therefore (due to inter-annual variability) they are in principle not representative of the long-term means. Thus, out of all the single-year stakes we took the ones which were more or less at the same location as a firn core or GPR point, and compared the accumulation between the single year (from the stake) and the long-term mean (from the core/GPR point). Then we used the mean ratio of these accumulation pairs to re-scale all the single year stake readings to the corresponding long-term mean. This provided some point estimates of long-term mean accumulation at locations where ice cores and GPR profiles were not available

(Fig. 3a). This method relies on the approximation (already used in the accumulation model) that the relative spatial patterns of accumulation are constant in time. The revised text includes a better explanation of the method.

line 246
I am not sure that the use of the word "conspicuous" is appropriate here.

We agree, we are replacing it with "extensive".

line 264
"Across the CG saddle": does that mean over the entire domain (see Minor Comment 1)?

This refers to the CG saddle, that is, excluding Seserjoch and the Grenzgletscher slope where the thermal regime is different. We are adding the mention "across the saddle" to the caption of Fig. 1 in order to clarify the naming.

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?

Yes, we have verified this by direct computation. This does not concern the very edge on the N side of the domain (where melt basically vanishes on the steep NE-facing slope), but rather the region to the east of the SK-saddle point line.

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.

This would certainly be an interesting analysis but we believe that it could be premature at this time: as we write in the conclusion, for the moment "more field observations are needed to verify the occurrence and improve the understanding of such events". We are currently working on a setup for continuous melt monitoring at Colle Gnifetti, whose results could prove helpful for such a trend analysis of micro-melt events.

line 289
Why "long-term"?

We agree that it is an unnecessary specification, we are removing it.

Figure 8
The term "GHF" is not defined. And I believe that the surface fluxes at CG could also be shown.

In the revised manuscript we are renaming all energy fluxes in the figure to be consistent with Eq. 1. We feel that adding a third panel to show fluxes at the saddle point would make this one-column figure too busy (too small panels, or too tall a figure), with little benefit as stated in the figure caption: the flux distribution at the saddle point is intermediate between the two (already very similar) distributions at the two other points.

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.

We agree, we are changing “obvious” into “often observed”.

line 381
Is it necessary to provide yet other location names (see Minor Comment 1)?

We understand that the many location names can prove confusing. However, *Sattelkern* and *Zumsteinkern* are the names of the ice cores from which the given refreezing estimates were derived (by Lier, 2018): we feel that mentioning them is necessary for the reader to understand the origin of these estimates, and this could simplify future comparisons of melt/refreeze amounts in the area. We are making the sentence clearer by replacing “saddle point” and “south facing slope” with the location names already used throughout the manuscript (SP and ZS).

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.

In the revised manuscript, we are adding a brief discussion of the need to overcome these limitations. We would place this later in the conclusion, where we discuss about “attempt[ing] model deployment at other cold firn/ice sites”.

References

Bohleber P., Erhardt T., Spaulding N., Hoffmann H., Fischer H., and Mayewski P. (2018). Temperature and mineral dust variability recorded in two low-accumulation Alpine ice cores over the last millennium, *Climate of the Past* 14: 21–37. https://doi.org/10.5194/cp-14-21-2018.

Brock B., Willis I., and Sharp M. (2006). Measurement and parameterization of aerodynamic roughness length variations at Haut Glacier d'Arolla, Switzerland. *Journal of Glaciology, 52*(177), 281-297. doi:10.3189/172756506781828746

Essery R., and Etchevers P. (2004). Parameter sensitivity in simulations of snowmelt. *Journal of Geophysical Research* 109(D20).

Feld S. I., Cristea N. C., and Lundquist J. D. (2013). Representing atmospheric moisture content along mountain slopes: Examination using distributed sensors in the Sierra Nevada, California: Representing Atmospheric Moisture Content in Mountains. *Water Resources Research* 49(7): 4424–4441.

Gilbert A., Vincent C., Six D., Wagnon P., Piard L., and Ginot P. (2014). Modeling near-surface firn temperature in a cold accumulation zone (Col du Dôme, French Alps): from a physical to a semi-parameterized approach, *The Cryosphere* 8: 689–703. https://doi.org/10.5194/tc-8-689-2014.

Greuell W., Knap W. H., and Smeets P. C. (1997). Elevational changes in meteorological variables along a midlatitude glacier during summer. *Journal of Geophysical Research: Atmospheres* 102(D22): 25941–25954.

Haeberli W. and Funk M. (1991). Borehole temperatures at the Colle Gnifetti core-drilling site (Monte Rosa, Swiss Alps), *Journal of Glaciology* 37: 37–46. https://doi.org/10.3189/S0022143000042775.

Konzelmann T., Vandewal R., Greuell W., Bintanja R., Henneken E., and Abeouchi A. (1994). Parameterization of global and longwave incoming radiation for the Greenland Ice Sheet. *Global and Planetary Change* 9: 143–164. https://doi.org/10.1016/0921-8181(94)90013-2.

Lier J. (2018). *Estimating the amount of latent heat released by refreezing surface melt water for the high-Alpine glacier saddle Colle Gnifetti, Swiss/Italian Alps*. Master's thesis, University of Heidelberg.

Lüthi M. P. and Funk M. (2001). Modelling heat flow in a cold, high-altitude glacier: interpretation of measurements from Colle Gnifetti, Swiss Alps. *Journal of Glaciology* 47: 314–324. https://doi.org/10.3189/172756501781832223.

Mattea E. (2020). *Measuring and modelling changes in the firn at Colle Gnifetti, 4400 m a.s.l., Swiss Alps.* Master's thesis, University of Fribourg. Available online at https://bigweb.unifr.ch/Science/Geosciences/GeographyTechnical/Secretary/Pub/Publications/

Geography/SelectedBachelorMasterThesis/2020/
Mattea_E._(2020)_M_Measuring_modelling_changes_Colle_Gnifetti.pdf.

Silva A. A., and Souza-Echer M. P. (2016). Ground-based observations of clouds through both an automatic imager and human observation. *Meteorological Applications* 23(1): 150–157.

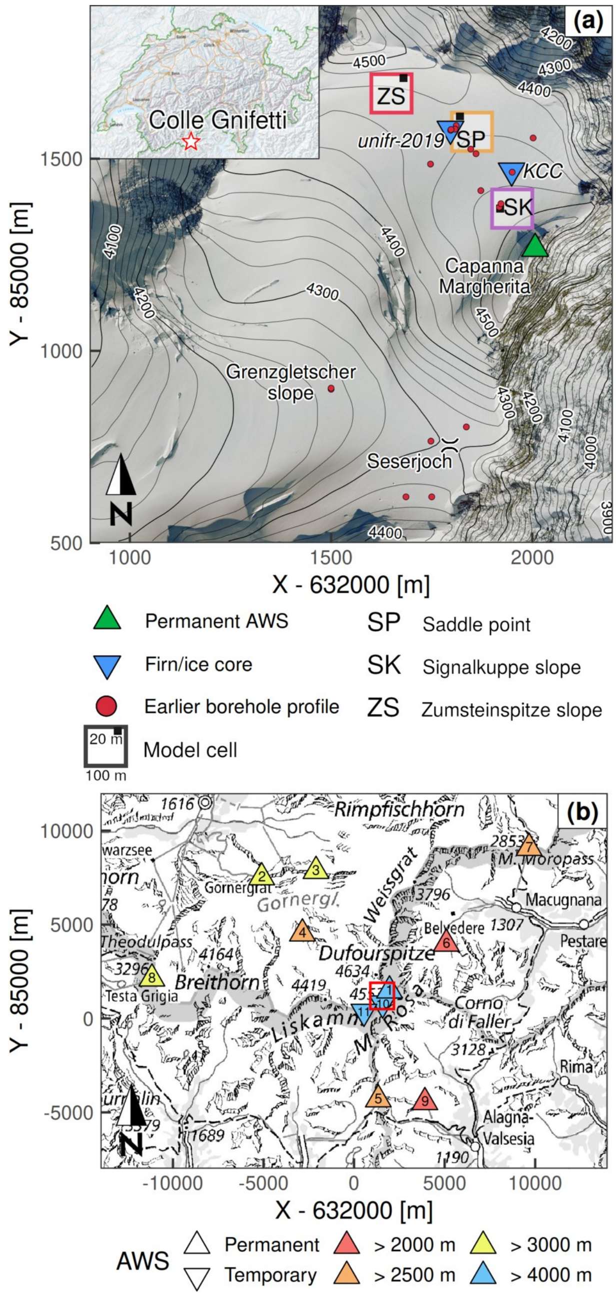
Suter S., and Hoelzle M. (2002). Cold firn in the Mont Blanc and Monte Rosa areas, European Alps: spatial distribution and statistical models. *Annals of Glaciology* 35: 9–18.

Suter S., Hoelzle M., and Ohmura A. (2004). Energy balance at a cold Alpine firn saddle, Seserjoch, Monte Rosa. *International Journal of Climatology* 24(11): 1423–1442.

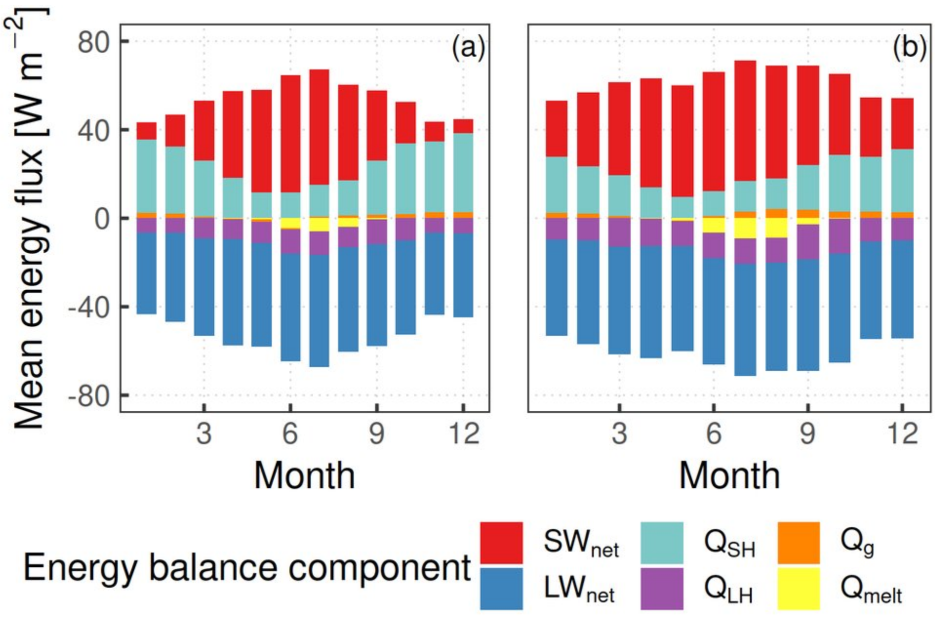
van Pelt W. J. J., Oerlemans J., Reijmer C. H., Pohjola V. A., Pettersson R., and van Angelen J. H. (2012). Simulating melt, runoff and refreezing on Nordenskiöldbreen, Svalbard, using a coupled snow and energy balance model. *The Cryosphere* 6(3): 641–659.

van Pelt W. J., and Kohler J. (2015). Modelling the long-term mass balance and firn evolution of glaciers around Kongsfjorden, Svalbard. *Journal of Glaciology* 61(228): 731–744.

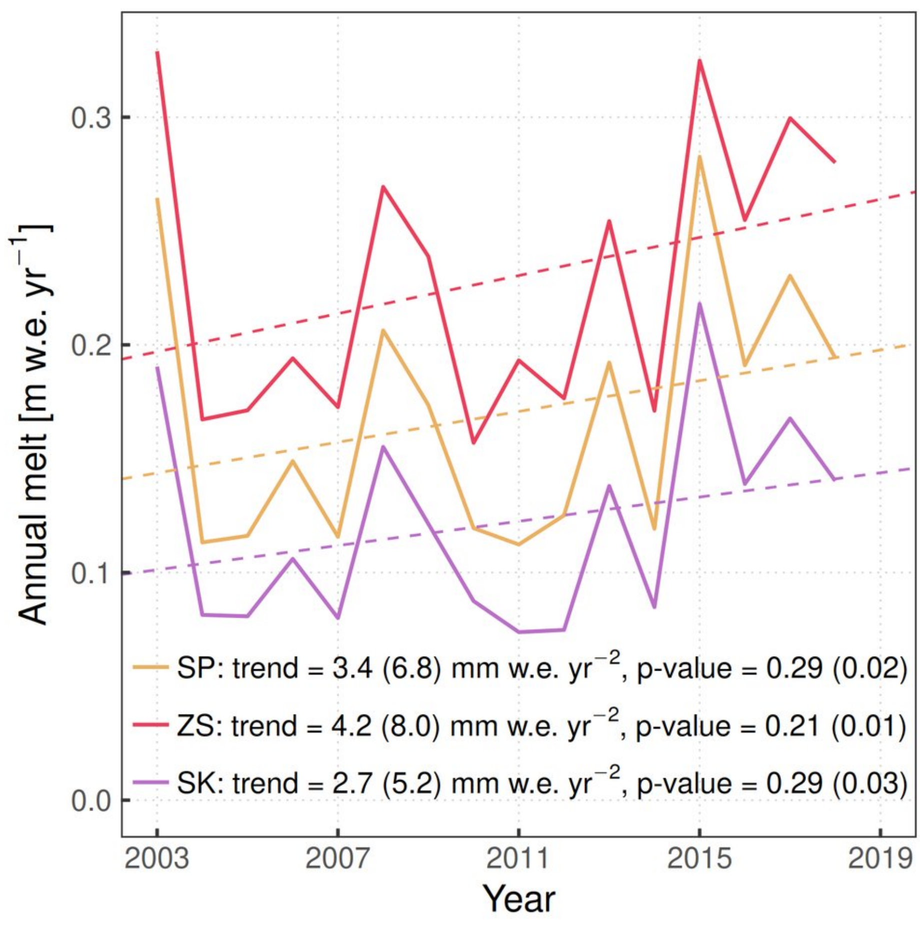
van Pelt W., Pohjola V., Pettersson R., Marchenko S., Kohler J., Luks B., Hagen J. O., Schuler T. V., Dunse T., Noël B., and Reijmer C. (2019). A long-term dataset of climatic mass balance, snow conditions, and runoff in Svalbard (1957–2018), *The Cryosphere* 13: 2259–2280,
https://doi.org/10.5194/tc-13-2259-2019.



Updated Figure 1



Updated Figure 8



Updated Figure 9