Response to Review 3

We thank very much Reviewer 3 for his comments that help improving the manuscript. Please find below our point-by-point replies in blue color.

General comments: The paper highlights results from a winter field campaign based out of the well-known WFJ site in Davos. The authors present a temporal analysis of snow microstructure and mechanical properties using state-of-the-art instruments that all have their advantages and limitations. Of particular relevance, repeated SSA and resistance measurements using an IceCube and the SMP are presented and com- pared against SNOWPACK simulations. The originality of the paper reside in a new calibration for the V4 of the SMP that will be indeed useful for international users such as my own group.

Overall, the paper is clearly written, with a very thorough analysis that certainly is worthy of publications. The expertise and reputation of the author's list is obviously excellent. I however, have several comments and questions that I would like to see addressed from my own perspective of being a SNOWPACK user in the Arctic with our own SMP and IRIS instrument since I think some elements need stronger analysis or at least physical explanations from the results presented in the paper given that very important science questions remain open.

Specific comments: In general terms, using SNOWPACK is not trivial. Yes the model can run virtually anywhere, especially in Switzerland where it was developed but cer- tainly harder elsewhere. A realization we came with as being users since 2002 is that the model remains very sensitive to 1) forcing dataset, 2) soil configuration and 3) ob- viously the internal physic calculations of microstructural elements that have changed from version to version over the years. For instance, a bias is observed in Canada on snow depth as a function of precipitation rate; or again bias in microstructure are not the same given the metamorphic process in place (kinetic vs equilibrium). Section 4 of the paper present the model in very general terms, I would suggest modifying this section to: SNOWPACK configuration where the authors would list: better description of the meteorological forcing dataset; soil configuration (type, roughness, how many soil layers?). There is also no mention of the spin up? Was the simulation initiated with a snow profile? It is obvious form the author list that the simulation is more than likely to be well parameterized, simply that I think there are more and more SNOWPACK users aware of potential problems, so more details on the simulation configuration I think would be very beneficial.
→We agree with the reviewer that the configuration of SNOWPACK needs a few more additions regarding initialisation, soil, etc. We adapted Section 4 accordingly. However, it is out of scope to present a detailed description of either the model or the data set in this study that has a quite different focus. Instead we will refer to Wever et al. (2015) that contains all information needed, except for the new settlement scheme. Indeed, and unfortunately, that part has only been presented in the frame of EGU 2011 but was not published yet. We are planning to do so soon. In addition, the dataset will be made available on Envidat upon acceptance.

Page 2, Line 8: 'spatially consecutive'...what is meant exactly? A clarification be appreciated. I assume the snowpit in such a confined space is useful for time-series, to avoid any variability due to spatial variability processes.
→ By "spatially consecutive" we meant those snow pits are dug consecutively during the season. We modified the sentence, P2, L6: "Regular snowpack monitoring programs rely on weekly to bi-weekly manual observations and measurements, by digging snow pits along a profile line in the (nearly) homogeneous observation area."

Page 2, Line 18: I would argue to add as a more general term the importance in surface energy balance, which in turn plays a critical role in freeze-thaw cycles for example. So the importance for large scale processes.
→ We agree and modified the sentence as follows: "It is defined by the ice/air interface surface area divided by the snow mass, which is inversely proportional to the optical grain size. SSA drives many snow processes as metamorphism, radiation interaction, air flow, chemical reactions and thus plays an important role in many large scale processes such as surface energy balance (e.g. Domine et al. 2007).

Page 3, Line 5: . . .change to gap in temporal resolution
→ Here we meant the gap in temporal and spatial resolution. We modified accordingly.

Page 3, Line 17-18: This was the whole idea behind the Snow Grain workshop held several years ago. Would the authors consider revisit some of the data?
→ We agree with the Reviewer that it will be a good idea to work with the data of the Snow Grain workshop. As far as we know, there are no plans in this direction for now.

Page 3, Line 29: how were selected the sites? It is mentioned that site were chosen on 'selected locations' but we all know site selection is critical. Some details on how the sites/samples were chosen be appreciated.
→ We agree and added more details on the selected locations, which were chosen to monitor the bottom part of the snowpack during the winter, i.e. the snow located around the persistent crust (MF-layer) and the weak layers (DH-layer and FC-layer). Once during the season we extend our sampling up to the slab on top of the FC-layer (including the RG-layer). These details are now provided in the paper so it reads now P3,L31: "occasional profiles of the 3D microstructure at 18μm vertical resolution from X-ray tomography (not full-depth, only on selected heights in the snowpack, mostly focusing on defined layers of interest)". Besides, we also include P4, L24: "X-ray tomography measurements of extracted, decimeter-sized samples were occasionally performed six times during the season at selected locations to image some defined layers of interest and allow further comparisons."

Table 1: add units to the measured/derived properties.
→ done

Page 5, Line8: ECT are extended column test, not extended compression test.
→ corrected

Section 3.2.: What was used to weigh the density cutter?
→ A digital scale was used to weigh the density cutter.

Page 7, Line 8: The 10% uncertainty is for IceCube or DUFISSS? IceCube was used, but the reference provided is for DUFISSS. What is the published accuracy of IceCube?
→ We provide here the uncertainty from Gallet et al. 2009, i.e. DUFISSS, assuming that uncertainty for IceCube is likely to be the same as the latter has been developed directly based on DUFISSS. As far as we know, we are not aware of a study specifying the accuracy for IceCube specifically.

Page 8, Line 8: I know the 1.2C threshold is used, likely well parameterized for WFJ. However I assume mixed precipitations are possible, what uncertainty can arise from such cases? A study by Ding et al. (2014) suggest that precipitation type are not only a factor of Tair, but also altitude and relative humidity. So how precise, at WFJ is precipitation phase parameterized?
→ The question of the impact on simulations from not considering the phase of precipitations cannot be answered straight away as we currently do not have observations permitting a proper attribution of precipitation phase at Weissfluhjoch. However, in preparation of the first SnowMIP around 2000, a dataset including the phase (liquid/solid, no mixed precipitations) and based on visual observations of the current weather could be constructed. The observations led us then to use a threshold of 1 °C. The threshold of 1.2 °C for Automatic Weather Station located above ~1000 m a.s.l. was introduced for operational use and proved to be well suited for Switzerland and Weissfluhjoch in particular (see Schmucki et al., 2014).
Along the period considered in this paper, there were no major precipitations associated with air temperatures above 0 °C though.
In summary, this threshold plays no role in the context of this study and it would be out of scope to discuss it further in the text. Nevertheless, we reformulated slightly that sentence in Section 4 of the paper.

Section 5.1: Our group is also doing just that with our own SMP this winter. Our concern is, that we are working on deriving a SMP(lc) method based on vertical 'z- axis' measurements from the SMP, with IceCube and density cutter that have a strong 'y-axis' component. We are asking ourselves if the SMP 'F' and 'L' parameter would be the same if we were to conduct a SMP profile in the 'y-axis' (i.e. in H instead of V)… From an anisotropy point of view, I think we can expect them to be different. Also we have an IceCube that includes a very thin layer being samples, with a density cutter that include a lot more snow… We are dealing with different scale, yet trying to correlate them together, I am fully aware that for now, this is the way to do it. Simply that I'd be happy to hear the authors ideas on this offline.

Section 6.1.: the problems linked to the vapor flux parameterization behind the growth of depth hoar is well known (Domine et al, 2019; Gouttevin et al., 2018). I'm also aware of the current work done in author's lab to correct that problem. Given the temperature and snow depth stated, yes I'm not surprise to see presence of depth hoar. Although, I'm pleased to see that SNOWPACK seems to react quite well to this, especially when I'm looking at Figure-7 where the depth hoar layer is indeed corresponding to a reduction in density as can be expected. This was a problem, that now looks much better. So my question is: did the authors used a different metamorphism parameterization to reach this? Or the standard version online was used without further modification?
→ Thank you for mentioning this. We added a sentence in the text to draw the reader's attention to it. However, neither changes nor adaptations to the metamorphism scheme

of SNOWPACK were implemented to reach this in our study. Indeed, whenever a deep depth hoar layer develops at the bottom of the snowpack at Weissfluhjoch, the resulting lower density of that basal layer is reasonably captured by SNOWPACK. For example, see winters 2015, 2005, and 2002 in Wever et al, (2015). A close inspection of the newly added Figure 7 however reveals that SNOWPACK still systematically underestimates the density of the slab while the density of the base is still overestimated the DH base. This effect is minor for this alpine snowpack but may still be emphasized for more extreme DH formation. We added some discussion on the performance of SNOWAPCK to simulate depth hoar layer point in Section 7.2

Figure-4 would be much easier to read with a legend.
→ A legend has been added to Figure 4.

Page 16, Line 2: Why does SNOWPACK overestimate the density of the DH layer? Is it because of the absence of vapor flux from the ground leading to the underestimation of the SSA?
→ The overestimation of density of the DH-layer from SNOWPACK is probably, at least partly, linked to the absence of the MF-layer in the simulations (not formed), Without this dense, stiff layer, we think that more load might have been transferred to the DH-layer leading to more densification. More work would be needed (we could force the simulation of a crust and compare simulations with and without it for example) to investigate the origin of this overestimation. Concerning SSA, SNOWPACK underestimates values in overall, so for all snow types and not more particularly for the DH-layer. The main cause is thus likely not only linked to an effect close to the ground, which would affect mostly basal layers, but maybe to the SSA parameterisation scheme implemented in SNOWPACK. However, as pointed out in Section 7.3, it is difficult to evaluate the SSA simulations in details because of the significant inter-measurement deviations observed. Dedicated works on that topic would be necessary.

Section 6.3: When using IceCube, it is very hard to sample properly depth hoar by the simple nature of the thickness of the hoar layer vs the sampler size. Any sampling difficulties were encountered using IceCube in these conditions?
→ We did not encounter major difficulties to sample the depth hoar layers that were indeed made of rather large crystals but also rather dense (around 300 kg/m3). The latter might have contributed to facilitate the sampling. For the layers that are difficult to sample, including depth hoar but also in our case fresh snow or crust, we might had to repeat the sampling and measurement so that consistent, reliable values were obtained.

Section 7.1, Lines 9-10: I would argue that yes there is a range, but it remains alpine where the processes governing stratigraphy, energy transfer is a different world from what we find in the Arctic, or even in other alpine regions of the world. I would argue to state that the snowpack offered a wide range of alpine snow conditions.
→ We agree and modified the sentence accordingly such as P21, L3: "Yet, the snowpack monitored over winter 2015-2016 offered a wide range of alpine snow type and property variations throughout the season."

Section 7.2.: With a snowpack having a temperature gradient important enough to lead to the formation of a depth hoar layer, can expect to have a decent variability in temperature vertically obviously. But, the effect of changing temperature as the SMP travels through snow is not discussed. I know the authors are aware of this problem, can they confirm this was not an issue in this environment?

→ We are indeed aware of this problem of the influence of temperature on the penetration resistance signal. We did not observed any anomaly or drift in values that could be related to this effect on the technical side. On the scientific side, for the present calibration we did not take into account any temperature dependence of the calibration parameters. In principle, such an extended analysis seems readily feasible from the present dataset by re-evaluating the structural data together with the (relatively robust predictions) of temperatures from SNOWPACK. It remains unclear though if temperature trends could be statistically discerned from microstructural (snow type) effects. But the present calibration must be considered as an average over all naturally occurring temperatures in the snow profile.

Page 24, Line 12: I think it is more a problem of the laser hitting the side of the sampler rather than the bottom, but this is a small detail.

→ We thank the Reviewer for the comment and would be happy to exchange more on this issue.

Again, this is a very nice contribution made by a very solid team at a site internationally known. I would suggest my comments to be minor, and would be happy to see this work published after the comments above are addressed.