Response to Review 2

We thank very much Reviewer 2 for his/her comments that help improving the manuscript. Please find below our point-by-point replies in blue color.

The paper presents the RHOSSA campaign focusing on snow density, SSA and sta- bility measurements over one winter in Weissfluhjoch, Switzerland. Modern methods such as SMP and IceCube are compared with traditional snow pit measurements and SNOWPACK modeling. Measurement results demonstrate how modern methods can increase temporal and vertical resolution in snow profiling compared with traditional measurements. This kind of data sets allow proper evaluation of modeling results, which is not possible using traditional measurements due to their poor temporal and vertical resolution. The main result is the recalibration of Proksch et al. 2015 model for deriving SSA and snow density from SMP data.

The snow stability part is a bit disconnected from the main text, which focuses on SSA and density. The authors could consider dropping the stability measurements.
→ We understand the concern raised here. However, although the mechanical properties are not analysed in as many details as for the structural properties, we think providing the complete dataset, including mechanical properties, can be very useful for other studies, as studies related to avalanches for example. It is important to keep in mind that traditional snow observations have a long tradition in avalanche research, which supports daily snow observations in alpine regions (such as Switzerland or France). And since nowadays stability predictions become feasible from high-resolution density profiles we definitely want to keep it. The full dataset is made available through a doi given in the paper.

p4r4 Section 6-> Section 5
→ Corrected

p4r12 Degrees missing from coordinates.
→ Corrected

p6r15 The snowpack was sampled with 3 cm resolution. What did you do with layers thinner than 3 cm? This explains why the 22 Feb layer is "only diffusely reported in the IceCube data" (p16 r17), if it is mixed with grains from other layers. How did you sample the MF layers? They are very difficult to get into sample holder without breaking them. Were the low density layers compacted to avoid measuring the sample holder?
→ Density and SSA profiles were recorded at regular height intervals of 3 cm, without considering the layering (we did the same for the SMP data with a vertical resolution of 1 mm and with the tomography data with a resolution of 18 μm). Using regular vertical grids and not following defined layers allows comparing data from different measurements and simulations, solely based on height (objective), without the need to identify layers (which can be subjective). We agree that a vertical resolution of 3 cm can lead to sampling in layer transitions leading to a more diffuse picture of the density or SSA profile. The MF layers were not too difficult to sample in our case (not overly dense) and procedure was the same as for other layers. Unfortunately, we are not aware of the method of compacting low-density layers to avoid measuring sample holder.

p8r8 Why exactly 1.2 ◦C?

→ 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.

p9r10 What is the justification for selecting different method for matching the profiles here than later in the paper (p9r24)? If re-aligning profiles using the MF layer resulted in "better correlation between estimates from SMP and snow pit measurements", why didn't you use the same method here to derive the parameters?

→ We thank the reviewer for pointing out this issue in the paper. We agree on the confusion about the alignment methods used in Section 5.1. The revised version of the paper was modified so that alignment done for the statistical modelling (Eq 1 and 2) and, later, to compare cutter/IceCube data and SMP data (Fig 2) is *the same and based on the snow surface*. Using the snow surface to re-align co-located and co-temporal profiles is the more convenient and systematically applicable method that can be done by others in the same way (unlike using a specific layer of the snowpack).

Besides, our statement that using the MF-layer alignment leads to better correlation of values in Fig 2 was erroneous. Slightly better R2 coefficients are found when using the snow surface than using the MF-layer for re-alignment (from layer alignment to snow surface alignment: R2 changes from 0.73 to 0.75 for density and from 0.81 to 0.82 for SSA, using Eq 1 and Eq 2 respectively). This makes sense as Eq 1 and 2 have been developed based on a snow surface re-alignment.

Finally, the re-alignment based on the MF-layer is now only used in the Result part, for time-series plotting purposes for which snow surface alignment is not relevant as profiles are not co-temporal anymore (evolution of snowpack height over the season).

Modifications concerning the alignment method were done throughout the paper, especially:

- Figure 2 has been redone, based on data re-aligned using the snow surface
- R2 coefficients associated to Figure 2 have been modified. They are slightly better than the previous version (from layer alignment to snow surface alignment: R2 changes from 0.73 to 0.75 for density and from 0.81 to 0.82 for SSA, using Eq 1 and Eq 2 respectively). This actually makes sense as Eq. 1 and 2 have been developed from data aligned with the snow surface.
- Section 5.1 reads now, p10, L6: The performance of the new parametrizations compared to the original parametrizations of Proksch2015 is presented in Figure 2. This plot shows the observed density from cutter measurements against the SMP-derived density obtained from Eq. (1) and from Proksch2015 for the 15 days for which both data are available (same data as used for the statistical

modelling). Similarly, the observed SSA from IceCube measurements are presented against the SMP-derived SSA from Eq. (2) and from Proksch2015 for the 13 days for which both data were available (again, same data as used for the statistical modelling). To do so, and as done for the statistical modeling, SMP-derived properties were averaged over 3 cm resolution and SMP and snow pit profiles of the same day were re-aligned with the snow surface and cropped to the length of the shortest profile. "

- In the introduction to the Result part, p12, L3, we explained further the choice of the MF-layer alignment to do temporal plots: "To present the evolution of profile properties with time, vertical profiles presented in the following were re-aligned such as z = 0 cm corresponds to the height of the upper boundary of the MF-layer (i.e. the 20151202-boundary). Choosing this layer as a height reference leads to a qualitatively better match than by simply taking the ground as reference (the field site ground at WFJ is uneven)."

p9 The model parameters are derived from IceCube measurements. Later (e.g. Fig 11) you show that there are big differences between IceCube and tomography measurements. Please comment on the accuracy of SMP-derived SSA values.
→ Comparisons between SSA measurements are described in Section 7.3. The SMP-derived SSA values inherits from 1/ the accuracy of the IceCube measurements (since the SMP-derived SSA values come from a fit (Eq 2) of the IceCube data) and 2/ the quality of the statistical model (how good is the fit).
Concerning 2/, the quality of the model is described in Section 5.1 and in Figure 2 (scatter plot). To describe further correlation of values in Fig 2, we included the RMSD values. P10, l14 now reads: "Applying a simple linear correlation between rho_cutter and rho_smp, a R2 coefficient of 0.87 and a root-mean square deviation (RMSD) of 34 kg m^{-3} are found when using Eq. (1) against a R2 of 0.75 and a RMSD of 69 kg m^{-3} when using the parametrization of Proksch et al. (2015). Between SSA_ic and SSA_smp, a R2 coefficient of 0.82 and a RMSD of 7 m^2 kg^{-1} are found when using Eq. (2) against a R2 of 0.65 and a RMSD of 14 m^2 kg^{-1} when using the parametrization of Proksch et al. (2015)."
Also, in Section 7.3 (line 14 page 24), we raise the point that the present statistical model used to derive SSA from SMP measurements fails to reproduce the high SSA values of newly-deposited snow, and that this could be because of their under-representations (only one day) in the IceCube dataset used for calibration.
Point 1/ is mentioned in Section 7.3 such as "First, we recall that density and SSA derived from SMP data were obtained to best match results from the cutter and IceCube measurements, so they necessarily inherit their performances" (p 23, L. 27). This implies that any discrepancies between IceCube and tomography data will necessary be also found between SMP-derived data and tomography data.

p11r2 choose->chose
→ modified accordingly

p11r20 caption->panel
→ modified accordingly

p22 Fig 11. The difference between SSA derived from SMP and tomography varies between different layers. Do you think the snow structure (grain type) has something to do with that? Should the SSA model be calibrated separately for different grain types?

And why are there big differences between IceCube measurements and SMP, if IceCube data was used in the fitting, shouldn't they agree better?

→ From Figure 11, we think that the variations in the differences between SSA derived from SMP and tomography depend more on the range of SSA values considered, rather than on the layers considered and so on the snow structure. Indeed, the quality of Eq 2 is better for some SSA ranges than other. In particular, looking at Figure 2b at SSA values below 20 $m^2 kg^{-1}$, we see that most of SMP values are slightly overestimated compared to IceCube values (cloud of values slightly below the 1:1 curve). Back to Figure 11, this bias clearly appears for most layers, for which SSA values are all mostly below 20 $m^2 kg^{-1}$. We add a sentence in the paper about this comment, which reads, P24, L7: "Note that one major discrepancy between IceCube and SMP-derived SSA comes from that the calibration used (parameterization) leads largely to an overestimation of the SSA values below about 20 $m^2 kg^{-1}$ by the SMP compared to IceCube (see Figure 2b, data cloud is mostly located below the 1:1 curve). This can be clearly seen in our results (Figure 9, 10, and 11) since a large part of the snowpack shows SSA values below 20 $m^2 kg^{-1}$."

→ Regarding the differences observed between SMP estimates of SSA and IceCube, they are directly link to the quality of the prediction Eq 2. To explain why a better regression could not be obtained, we think one point is that some snow type might not be well captured because of the under-representation in the IceCube measurements for some snow types, such as fresh snow in our case (this was the case of only 1 day on measurement for which fresh snow was measured in the first cm of the snowpack). To improve that, the calibration dataset should be extended so that all snow type is rigorously covered.

→ Regarding the grain type: a large part of the motivation of this work is making a step away from (subjective) indices. Thus re-introducing grain-type dependent calibration coefficients is, from our perspective, the wrong way to go. But it is true that the microstructure has an impact on the performance of the calibration model. This is the reason that only by introducing the SMP parameter L into the model, a significant improvement of the calibration (in particular in depth hoar) could be made over the old approaches of just using the median of the SMP force. Similar things are expected to happen for other snow types. The fact that the SSA point cloud in Fig 2 is not straight but slightly curved further supports that the present calibration model is still missing essential physics.

p23 Fig 12. Please add SNOWPACK profiles as well.

→ SNOWPACK simulations were added in Fig 12.