The RHOSSA campaign: Multi-resolution monitoring of the seasonal evolution of the structure and mechanical stability of an alpine snowpack

Neige Calonne[1,2*], Bettina Richter[1*], Henning Löwe[1], Cecilia Cetti[1], Judith ter Schure[1], Alec Van Herwijnen[1], Charles Fierz[1], Matthias Jaggi[1], and Martin Schneebeli[1]

[1]WSL Institute for Snow and Avalanche Research SLF, Davos Dorf, Switzerland
[2]now at Météo-France – CNRS, CNRM UMR 3589, Centre d'Etudes de la Neige, Grenoble, France
[*]These authors contributed equally to this work

Correspondence: Neige Calonne (neige.calonne@meteo.fr)

Abstract. The necessity of characterizing snow through objective, physically-motivated parameters has led to new model formulations and new measurement techniques. Consequently, essential structural parameters such as density and specific surface area (for basic characterization) or mechanical parameters such as the critical crack length (for avalanche stability characterization) gradually replace the semi-empirical indices acquired from traditional stratigraphy. These advances come along with new demands and potentials for validation. To this end, we conducted the RHOSSA field campaign, in resemblance of density (ρ) and specific surface area (SSA), at the Weissfluhjoch research site in the Swiss Alps to provide a multi-instrument, multi-resolution dataset of density, SSA, and critical crack length over the complete winter season 2015-2016. In this paper, we present the design of the campaign and a basic analysis of the measurements alongside with predictions from the model SNOWPACK. To bridge between traditional and new methods, the campaign comprises traditional profiles, density cutter, Ice-Cube, SnowMicroPen (SMP), micro-computed-tomography, propagation saw tests, and compression tests. To bridge between different temporal resolutions, the traditional weekly to bi-weekly snow pits were complemented by daily SMP measurements. From the latter, we derived a re-calibration of the statistical retrieval of density and SSA for SMP version 4 that yields an unprecedented, spatio-temporal picture of the seasonal evolution of density and SSA in a snowpack. Finally, we provide an inter-comparison of measured and modeled estimates of density and SSA for 4 characteristic layers over the entire season to demonstrate the potential of high temporal resolution monitoring for snowpack model validation.

1 Introduction

Regular snow monitoring programs are one of the cornerstones of snow science providing valuable time-series of snow properties (e.g. Reba et al., 2011; Morin et al., 2012; Landry et al., 2014; Wayand et al., 2015; Leppänen et al., 2016; Lejeune et al., 2019). Such time-series are indispensable for the development and evaluation of snow models (e.g. Fierz, 1998; Etchevers et al., 2004; Morin et al., 2013; Essery et al., 2016; Krinner et al., 2018) as well as for various applications such as snowpack stability assessment for avalanche risk forecasting (e.g. Schweizer and Wiesinger, 2001; van Herwijnen and Jamieson, 2007),

snowpack processes studies (e.g. Dumont et al., 2017), snow property retrievals from remote sensing (e.g. Leinss et al., 2016; King et al., 2018), water resources estimations (e.g. Jonas et al., 2009), climate studies (e.g. Takala et al., 2011), or instruments developments (e.g. Schneebeli et al., 1998). Worldwide many study sites have been established for snow monitoring (Ménard et al., 2019). Col de Porte in France (Lejeune et al., 2019), Sodankylä in Finland (Leppänen et al., 2016), and Weissfluhjoch

5   (WFJ) in Switzerland (Meister, 2009) offer some of the longest time-series of snowpack observations, ~~e.g.~~ dating back to 1936 ~~for the WFJ site~~in the case of WFJ. Regular snowpack monitoring programs rely on weekly to bi-weekly manual observations and measurements, ~~digging temporally and spatially consecutive snow pits~~ by digging snow pits along a profile line in the (nearly) homogeneous observation area. Observations comprise mainly traditional profiling with a characterization of layer properties (~~hand hardness....~~grain size, grain shape, hand hardness, and wetness) and measurements of ram resistance and snow

10 temperatures, all following standard procedures ~~.~~(Fierz et al., 2009). Those measurements are typically complemented by so-called snow stability test, such as the compression test (Jamieson, 1999; van Herwijnen and Jamieson, 2007), to monitor weak layers and snow mechanical properties in view of avalanche forecasting. Although these traditional characterization methods are well-established, they suffer from well-known problems of quantitative objectivity, limiting their use for physical snow modeling.

15   To address this issue, efforts have shown a clear tendency of replacing traditional measurements by newly-developed field methods to obtain more ~~objective, non-empirical~~ objectively-defined snow properties. Concerning the characterization of snow microstructure, the observer-biased estimate of traditional grain size ~~tends to~~ can be replaced by measurements of specific surface area (SSA) (Morin et al., 2013; Leppänen et al., 2015). It is defined by the ice/air interface surface area divided by the snow mass, which is inversely proportional to the optical grain size~~, and~~. SSA drives many snow processes as metamorphism,

20 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). Different field instruments were developed to measure SSA based on similar methods such as DUFISSS (Gallet et al., 2009), POSSSUM (Arnaud et al., 2011), ~~Iris~~IRIS (Montpetit et al., 2012), or IceCube (Zuanon, 2013). Concerning snowpack stability assessment, classical stability tests are now often complemented by the propagation saw test (PST), developed about a decade ago to objectively characterize the crack propagation propensity based on the critical

25 crack length parameter (Gauthier and Jamieson, 2006; Sigrist and Schweizer, 2007; van Herwijnen and Jamieson, 2005). The critical crack length corresponds to the length of a saw cut manually introduced in a buried weak layer leading to rapid crack propagation (e.g. Gauthier and Jamieson, 2008). Additional mechanical parameters can be obtained when combining PSTs with particle tracking velocimetry (van Herwijnen et al., 2016).

These latest advances in field measurements coincide with similar improvements in detailed snowpack models such as

30 Crocus (Brun et al., 1992; Vionnet et al., 2012) and SNOWPACK (Lehning et al., 2002b; Wever et al., 2015). The modeling of SSA as a prognostic variable was included in Crocus to replace the empirical grain size parameter (Carmagnola et al., 2014), and indirectly estimated in SNOWPACK from the grain size, dendricity, and sphericity (Vionnet et al., 2012). Modeling the SSA allows for an unambiguous comparison with SSA measurements. In addition, many snow properties can now be formulated using physical principles that naturally involve the SSA as a parameter. Likewise, a new model of the critical cut

length based on objective stratigraphic information was implemented in SNOWPACK (Gaume et al., 2017) and recently refined to support avalanche risk forecasting (Richter et al., 2019).

These advances, coherently developed in field techniques and modeling, come along with new demands for validation campaigns. If snow models are only validated against surface or bulk measurements instead of the full stratigraphy, the compensation of effects may prevent the detection of model errors (e.g. Essery et al., 2013; Lafaysse et al., 2017). However, only a few quantitative evaluations of density and SSA profiles exist (Morin et al., 2013; Leppänen et al., 2015; Wever et al., 2015; Essery et al., 2016). Presently, the evaluation of density and SSA is partly limited by the temporal and spatial resolution of measured profiles, which are typically conducted on a weekly to bi-weekly basis with a vertical resolution of 3 cm or set by the layers. In contrast, modeled profiles can be provided hourly and at sub-centimeter vertical resolutions. The gap in ~~resolution~~ resolutions between measurements and models precludes the evaluation of snow processes occurring on short time scales and/or locally in the snowpack, such as surface hoar formation (e.g. Stössel et al., 2010), faceting (e.g. Pinzer et al., 2012), or crust formation. Concerning the critical cut length, Richter et al. (2019) reported a good agreement between the temporal evolution of the critical crack length measured in the field and modeled from the refined parameterization. They also highlighted the capability of the parameterization to detect weak layers in simulated snow profiles.

Increasing the spatio-temporal resolution of measurements is still cumbersome due to inherent time-constraints for snow pits and manual measurements. Towards a remedy, recent studies utilized the micro-penetrometer SnowMicroPen (SMP) (Schneebeli et al., 1999) for both, microstructure characterization and stability assessment. Proksch et al. (2015) presented a statistical method to retrieve density and SSA from SMP data and Reuter et al. (2015) suggested an approach to estimate point snow instability from SMP data. These examples exploit key advantages of the SMP, namely fast profiling for frequent measurements and high vertical resolution~~such as~~, so that profiles are obtained at a considerably finer scale (mm) than possible with traditional means. Though principally promising, the use of the SMP within snow monitoring programs has never been assessed and would require a comprehensive comparison to other methods to evaluate uncertainties.

In the context raised above, the value of emergent, objective snow properties, their potential to replace traditional means in operational snow monitoring programs, and their requirements on temporal and vertical resolutions for model evaluations can ~~only~~ be investigated within a multi-resolution and multi-instrument dataset to facilitate comprehensive ~~cross-validation~~ cross-comparison analyses. We strive to provide such a resource in the form of the outcome of an extensive snow measurement campaign which is referred to as RHOSSA in resemblance of ~~density (~~the greek letter ρ ~~)~~ for density and SSA. The campaign was carried out at the WFJ site from December 2015 to March 2016 and comprises:

- daily ~~(full-depth)~~ profiles of density and SSA of ~~1~~ 0.5 mm vertical resolution derived from SMP measurements

- weekly ~~(full-depth)~~ profiles of density and SSA of 3 cm vertical resolution from manual snow pit measurements

- bi-weekly ~~(full-depth)~~ traditional profiles with layer-dependent vertical resolution, completed with PST and classical stability tests

- occasional ~~(selected locations)~~ profiles of the 3D microstructure at 18 μm vertical resolution from X-ray tomography (not full-depth, only on selected heights in the snowpack, mostly focusing on defined layers of interest).

Our main results comprise (1) ~~a newre-calibration of~~ new, re-calibrated parameterizations to derive density and SSA ~~retrievals~~ from SMP 4 measurements, (2) the evolution of density and SSA profiles at unprecedented spatial and temporal resolution, (3) the evolution of snow instability from various stability tests, (4) a comparison of the density and SSA estimates over time for distinct layers of the snowpack, and (5) a comparison between measured values of density and SSA and modeled ones from standard SNOWPACK runs that documents the state-of-the-art and highlights the potential of high resolution stratigraphy data for snow model evaluation and future developments.

The paper is organized as follows. Section 2 provides an overview of the design of the RHOSSA campaign. Section 3 and Section 4 describe the measurement methods and the simulations with SNOWPACK, respectively. Section ~~6~~ 5 presents specific data analysis methods applied to exploit the RHOSSA dataset, namely a re-defined statistical model for density and SSA retrievals from SMP 4 measurements and a layer tracking method to monitor the evolution of specific layers of the snowpack over the season. Section 6 provides a first analysis of the RHOSSA dataset in terms of stratigraphy, stability, density and SSA, including cross-comparisons between measurements and the evaluation of SNOWPACK simulations. Specific points are finally discussed in Section 7.

2   Campaign design

During the winter of 2015-2016, the snow observation program at the WFJ site, located in the Eastern Swiss Alps above Davos (elevation of 2536 m, latitude 46.82963°N, longitude 9.80925°E), was supplemented with additional measurements, forming all together the RHOSSA field campaign. We focused on the period ~~of dry snow~~ from beginning of December 2015 to end of March 2016 ~~,~~ to ensure measurements in dry snow condition as required by some of the used instruments. In addition, measurements were done in the morning typically starting at 8am. The RHOSSA campaign included traditional profiling, stability tests, density cutter measurements, IceCube measurements, SMP measurements, and tomography. Using such a wide range of measurement methods resulted in different temporal resolutions (frequency) and spatial resolutions (vertical along the snow profile), as synthesized in Table 1. SMP measurements were performed daily, density cutter measurements and IceCube measurements were performed once a week, traditional snow profiles were recorded on a weekly to bi-weekly basis and completed with stability tests. 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. Spatial resolutions range from 0.1 mm for the tomography-based properties to the size of the snow layer for the traditional profiling (typically from 1 to 30 cm).

The measurement field at the WFJ site is a flat area of about 20 m $\times$ 8 m ^{2} (Fig. 1). To ensure an efficient use of the snow field, measurements were performed within defined areas. The snow field was divided into three corridors, each 20 m long and 1.5 m wide, as illustrated in Figure 1. Throughout the season, sets of measurements were performed moving continuously along the corridor in daily steps, starting at one end of corridor 1 and ending at the end of corridor 3, two consecutive sets of measurements being at least 30 cm apart to avoid disturbances. A schematic of the location of three consecutive sets of measurements ("day 1", "day 2", and "day 3") performed in corridor 2 at mid season is shown in Figure 1. Each corridor was

divided lengthwise in 2 parts of 75 cm wide. One side was reserved for stability tests (red area in Fig. 1); the other side was used for all the other measurements. First, the five daily SMP measurements with a 15 cm spacing were performed perpendicular to the corridor direction (black dots in Fig. 1). Then, during a snow pit day as illustrated by "day 2" in Figure 1, the pit was dug such that the pit wall was parallel and a few centimeters behind the line that was formed by the SMP measurements. Density

5   cutter and IceCube measurements were done next to each other (blue and orange areas in Fig. 1), and complemented by a traditional snow profile when needed (green area in Fig. 1). Finally, for the occasional X-ray tomography, undisturbed snow blocks were extracted from the pit wall near the location of the other measurements.

Table 1. Overview of the RHOSSA campaign measurements.

Method	Frequency	Vertical resolution	Measured or derived properties
SnowMicroPen	daily (100 profiles in total)	~~1~~0.5 mm	penetration force ~~, density, SSA~~ (N), density (kg m^{-3}), SSA (m^2 kg^{-1})
Density cutter	weekly (15 profiles in total)	30 mm	density (kg m^{-3})
IceCube	weekly (13 profiles in total)	30 mm	SSA (m^2 kg^{-1})
Traditional profile	every 1 to 2 weeks (11 profiles in total)	variable	~~traditional layer parameters, temperature~~grain shape, grain size (mm), hand hardness, ~~ram resistance~~ temperature (°C), ram resistance (N)
Stability tests	8 times over the season	-	critical crack length (m), #taps until failure
Tomography	6 times over the season	0.1 mm	density ~~, SSA~~ (kg m^{-3}), SSA (m^2 kg^{-1})

3   Measurements

3.1   Traditional profile and stability tests

10   Traditional snow profiles were observed to characterize snow stratigraphy by hand hardness, grain size and grain type. In addition, ram resistance, snow temperatures, and water equivalent of the snow cover were measured (Fierz et al., 2009). Snow stability tests were performed to identify potential weak layers and evaluate the load required for failure. Specifically, we performed the compression test (CT; van Herwijnen and Jamieson, 2007), the extended ~~compression~~ column test (ECT; Simenhois and Birkeland, 2009) and the propagation saw test (PST; Gauthier and Jamieson, 2008). In a CT or an ECT, the

15   snowpack is progressively loaded by tapping on a snow shovel placed on the snow surface with increasing force (10 taps from the wrist, 10 taps from the elbow and 10 taps from the shoulder). If a failure occurs within the snow cover, the loading step,
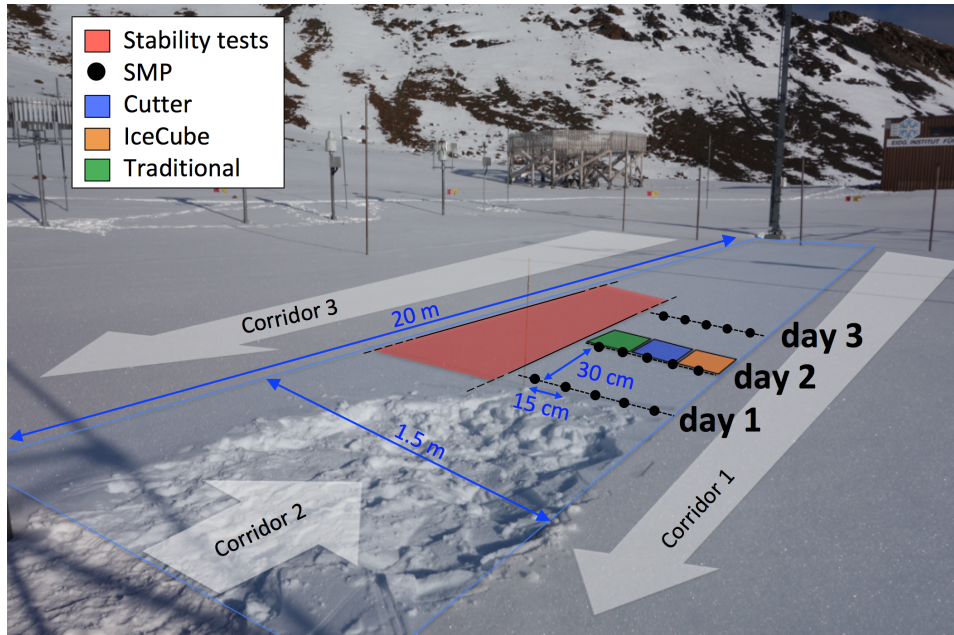
Figure 1. Picture of the snow field where measurements of the RHOSSA campaign were performed. The location of each measurement is illustrated for three consecutive days.

i.e. the number of taps at which the failure occurred, is recorded. In a CT, which consists of an isolated column of 30 by 30 cm, information describing the type of failure is also recorded (for more details see van Herwijnen and Jamieson, 2007). In an ECT, which consists of an isolated column of 30 by 90 cm, the propagation distance across the column is recorded as either no propagation, partial propagation or full propagation (for more details see Simenhois and Birkeland, 2009). CT and ECT are

5   thus used to identify potential weak layers and qualify the loading required for failure. The PST, on the other hand, is used to measure the critical crack length required for crack propagation in an a priori known weak layer. It consists of an isolated 30 cm wide column with a length of at least 120 cm, which has been excavated to below the weak layer of interest. An artificial crack is then created by drawing a snow saw through the weak layer until the critical crack length is reached and rapid crack propagation occurs. The critical crack length is recorded as well as the propagation distance, where END refers to cracks which

10   propagated to the end of the column (for more details see Gauthier and Jamieson, 2008).

3.2 Density cutter

A density cutter was used to manually record the density profile of the snowpack by performing successive measurements from the surface to the bottom of the snowpack with a vertical resolution of 3 cm. A box-type density cutter of 100 cm^3 ($3 \times 5.5 \times 6$ cm) (Carroll, 1977; Conger and McClung, 2009; Proksch et al., 2016), was used to measure density by weighing a snow

15   sample extracted from the cutter. A measurement error of about 10% can be expected (Carroll, 1977; Conger and McClung, 2009; Proksch et al., 2016), typical source of errors being the measurement of compacted snow volumes (overestimation) when

extracting light snow, and of incomplete snow volumes (underestimation) when extracting fragile snow (e.g. faceted crystals or depth hoar).

3.3 IceCube

The IceCube was used to measure an SSA profile of the snowpack by performing successive IceCube measurements from the surface to the bottom with a vertical resolution of 3 cm. The IceCube is an optical system commercialized by A2 Photonic Sensors (Zuanon, 2013) to retrieve SSA from measurements of the infrared hemispherical reflectance of snow (Gallet et al., 2009). Briefly, a snow sample is illuminated with a 1310 nm light diode and the light reflected by the snow surface is recorded. The signal is recorded as voltage values then converted in reflectance values based on a voltage-to-reflectance calibration curve obtained using certified optic standards. SSA values are finally estimated from the reflectance values using the parametrization of Gallet et al. (2009). The complete description of the measurement principle can be found in Gallet et al. (2009). Measurements were performed on cylindrical snow samples with a 6 cm diameter and 2.5 cm height, extracted from the snow pit following the method given by Gallet et al. (2009) and Zuanon (2013). Snow samples were very slightly compressed when inserted into the sample holder and attention was paid to have a flat snow sample surface. Measurement uncertainty was estimated to about 10% for SSA values below 60 m^2 kg^{-1}, as for the DUFISSS device. Additional measurement artifacts occur for snow with higher SSA that can lead to over-estimated SSA values (Gallet et al., 2009).

3.4 SnowMicroPen

The SnowMicroPen (SMP), a digital cone penetrometer, was used to measure the vertical penetration resistance profile of the snowpack. From that, density and SSA profiles were derived based on a statistical model and after a specific signal processing, as described in Section 5.1. The SMP consists of a motorized probe that is driven vertically into the snowpack at a constant speed of 20 mm s^{-1} to measure the penetration resistance exerted on a cone (diameter of 5 cm and cone half angle of 30°) located at the tip of the probe (Schneebeli et al., 1999). We used a version 4 SMP with a 2-meter rod and recorded penetration resistance with a vertical resolution of 1/242 mm. Two preliminary measurements were systematically performed to cool the SMP towards snow temperature before the five daily measurements were taken. The quality of each SMP profile was manually checked by evaluating the penetration resistance profiles. Signals showing strong drifts were discarded (e.g. frozen water in the SMP motor, defect of the force sensor, etc) . Signals that correspond to measurements in the air control of SMP force profiles was done manually by rejecting signals with 1) visible trends either in the air portion of the signal or over the entire depth, 2) high noise levels and unrealistic spikes, and 3) frozen tip problems revealed by a force response that appears to be activated only deeper in the snowpack. Most of these problems are caused by wet conditions. The air-snow and in the ground were truncated. No offset correction was necessary for this dataset. snow-ground interface were detected manually to remove air and ground regions from the signal.

3.5 Micro-computed tomography

X-ray micro-computed tomography was used to image the 3D microstructure of snow samples extracted from the snowpack at selected locations. Snow blocks of about $30 \times 30 \times 30$ cm^3 were cut out from the profile wall on 14 December, 13 January, 27 January, 10 February, 16 February, and 2 March. The location of the extracted blocks within the snowpack were chosen subjectively, either to ensure temporal continuity with a previously sampled block, or to re-focus on a particular layer of interest, mainly persistent weak layers. Extracted blocks were sealed in Styrofoam boxes and filled with dry ice (about -80$^{\mathrm{o}}$C) for transportation from the field site to the cold lab (duration approximately 1 h). In the lab, the blocks were stored at -25$^{\mathrm{o}}$C, and successively sub-sampled into sample holders of 7 cm height and 3.6 cm diameter. These samples were then scanned in a cooled micro-computer tomograph (μCT 80, Scanco Medical) with a resolution of 18 μm voxel size. ~~Reconstruction followed standard procedure.~~ The reconstruction utilized standard procedures with noise reduction by Gaussian filtering (support=2 voxels, width=1.2 voxels) and binary segmentation following the method of Hagenmuller et al. (2013). From the binary 3D images, density and SSA were computed over a moving window of 120 pixels height obtaining profiles at a vertical resolution of about 2 mm.

4 Simulations with SNOWPACK

To put the measurement campaign in context, we conducted standard simulations with the detailed snow cover model SNOW-PACK ~~(Lehning et al., 2002b)~~ (Wever et al., 2015; Lehning et al., 2002b) using version 3.4.1, revision 1473 (https://models.slf.ch/p/snowpack/). ~~SNOWPACK~~ The snowpack itself is considered to be a linear viscoelastic material, the settlement of which was calculated as described in Section 2.2.2 in Lehning et al. (2002b) and taking into account the impact of load rate. This new scheme also implies an altered viscosity parameterization (both unpublished). Liquid water flow in snow was solved using Richards equation recently implemented by Wever et al. (2014). Neumann boundary conditions were used at the snow-atmosphere boundary whereas a constant geothermal heat flux of 0.06 W m^{-2} was applied at the bottom of the 3 m deep soil column. 32 layers with thickness increasing from 1 to 40 cm with depth make up this column. A late summer iso-thermal temperature profile of 5°C was assumed. The simulation was initiated on 1 September 2015 with no snow on the ground until 14 October 2015 except for 1.5 days in September (snow height less than 11 cm). This results in a spin-up time of 43 days before the WFJ site was snowed in.

The model was driven with an optimized half-hourly dataset of meteorological and snowpack measurements from the automatic weather station at the WFJ site (WSL Institute for Snow and Avalanche Research SLF, 2015). The dataset is well described in Wever et al. (2015) and contains standard meteorological measurements including air temperature (ventilated), relative humidity (ventilated), wind speed, ~~shortwave and down-welling~~ short and long wave radiation ~~. The snow cover mass balance was driven with the~~ (both fluxes each), and undercatch corrected precipitation. The set also includes automatically measured snow height that was used to drive the snow cover accumulation, that is, by the increments of measured snow ~~depth. To estimate the occurrence of rainfall events if~~ height. The added mass is then obtained from the density of new snow computed using an empirical relation between air temperature and wind speed (Schmucki et al., 2014). To account

for rainfall we used the precipitation data whenever the air temperature exceeded 1.2 °C ~~, data on precipitation was used additionally~~(see Schmucki et al., 2014). Snow albedo was forced from the in-situ measurements of incoming and reflected shortwave radiation fluxes~~and snow height~~. The calculated values underwent a plausibility check and in case of a negative outcome were replaced by the model parametrization (less than 0.8% of the values). The surface sensible and latent heat

5   flux parameterizations are derived from Monin–Obukhov similarity (Lehning et al., 2002a). ~~Neumann boundary conditions were used at the snow-atmosphere boundary whereas a constant geothermal heat flux (0.06 W m^{-2}) was assumed at the bottom of the 3 m deep soil column. Liquid water flow in snow was solved using Richards equation recently implemented by Wever et al. (2014).~~

    The time step for the simulation was set to 15 min and output was written every 60 min. For this campaign, we were

10   particularly interested in evaluating the model in terms of density and SSA. The ~~density of new snow was obtained from an empirical relation between air temperature and wind speed (Schmucki et al., 2014). The snowpack itself is considered to be a linear viscoelastic material, the settlement of which was calculated as described in section 2.2.2 in Lehning et al. (2002b), using an altered viscosity parametrization. In addition, the effect of load rate was taken into account but any elastic effects were neglected. SSA was~~ latter was simply retrieved from the optical diameter of snow that is empirically derived from dendricity,

15   sphericity, and grain size according to Vionnet et al. (2012).

5   Data analysis methods

5.1   Deriving density and SSA from SMP

As a prerequisite to derive density and SSA from SMP measurements, it was necessary to modify the current statistical models of Proksch et al. (2015). When applying the parametrizations of Proksch et al. (2015), SMP-derived density and SSA compared

20   rather poorly to values from cutter and IceCube measurements respectively (Fig. 2). This is in part due to the fact that the parametrizations of Proksch et al. (2015) were derived from measurements with an SMP device version 2 whereas we used a newer SMP version 4 that contains different electronic components leading to different force correlations at small scale. We thus derived a re-calibration of the statistical models of Proksch et al. (2015) to better match our snow pit measurements. The obtained density and SSA parametrizations are called new parameterizations hereafter.

25   The idea of Proksch et al. (2015) was to relate a dataset of some relevant SMP micro-parameters to ~~reference values~~ a reference dataset of density (or SSA) ~~, both~~ from tomographic images using a statistical multi-linear regression model, all datasets being obtained from independent, co-located and co-temporal measurements~~, using a statistical regression model~~. The SMP micro-parameters consist of the median of the penetration resistance force $\tilde{F}$ and a characteristic length of the microstructure L (akin to the distance between two ruptures), as defined in the stochastic model of Löwe and van Herwijnen (2012).

30   Here we followed the same procedure but ~~we~~ took our cutter measurements as reference ~~values~~ data of density (ρ_{cutter}) and our IceCube measurements as reference ~~values~~ data of SSA (SSA$_{\text{ic}}$), ~~whereas Proksch et al. (2015) used values from tomography measurements~~instead of tomographic data. The statistical modeling was ~~thus~~ applied based on a sub-dataset of ~~15 days where~~ data from the days for which both SMP and snow pit measurements were available ~~. The SMP micro-parameters consist of~~

(15 days for density, 13 days for SSA). From each raw force signals, parameters $\tilde{F}$ and L were computed from the raw penetration force profiles over a sliding window of 1 mm with 50% overlap, yielding profiles of $\tilde{F}$ and L with a vertical resolution of 0.5 mm. Note that Proksch et al. (2015) used a sliding window of 2.5 mm, but tests with different window heights (1, 2.5 and 5 mm) did not show a significant impact. Next, for each day, the five daily profiles of $\tilde{F}$ and L of the same day were aligned by simply using snow surface as common reference and a median operation was applied to get one representative profile of $\tilde{F}$ and L per day, called the median profiles in the following. Next, each median profile was averaged vertically using a 3 cm window to match the vertical resolution of the snow pit measurements. Finally, the median 3cm-averaged profiles of $\tilde{F}$ and L and the profiles of ρ_{cutter} and $\mathrm{SSA}_{\mathrm{ic}}$ of the same day were aligned by using snow surface again as common reference and cropped to the length of the shortest profile. This way, all profiles of a given day are described on the same vertical scale and values of $\tilde{F}$, L, ρ_{cutter} and $\mathrm{SSA}_{\mathrm{ic}}$ can be paired for the statistical modeling, relying on a total of 590 paired-values for density and 497 for SSA.

Based on this sub-dataset, we applied a regression of the form

$$\rho_{\mathrm{smp}} = a_1 + a_2 \ln(\tilde{F}) + a_3 \ln(\tilde{F})\,L + a_4\,L \tag{1}$$

to estimate density from $\tilde{F}$ and L by least-squares optimization (ρ_{cutter} being the target). The following parameters were obtained: $a_1 = 295.8 \pm 0.3$, $a_2 = 65.1 \pm 0.1$, $a_3 = -43.2 \pm 0.4$, and $a_4 = 47.1 \pm 0.7$, where ρ_{smp} is in kg m^{-3}, L in mm and $\tilde{F}$ in N, and where the errors denote the standard errors of the regression. This regression has a R^2 coefficient of 0.79, a residual standard error of 40.8 kg m^{-3}, and p-values less than 10^{-3}. Differing slightly from the one suggested by Proksch et al. (2015), a regression of the form

$$\mathrm{SSA}_{\mathrm{smp}} = b_1 + b_2 \ln(L) + b_3 \ln(\tilde{F}) \tag{2}$$

was applied to estimate SSA by least squares optimization ($\mathrm{SSA}_{\mathrm{ic}}$ being the target). The following regression parameters were obtained: $b_1 = 0.57 \pm 0.05$, $b_2 = -18.56 \pm 0.04$, and $b_3 = -3.66 \pm 0.01$, where $\mathrm{SSA}_{\mathrm{smp}}$ is in m^2 kg^{-1}. This regression has a R^2 coefficient of 0.67, a residual standard error of 8.4 m^2 kg^{-1}, and p-values less than 10^{-3}.

The performance of the new parametrizations compared to the original parametrizations of (Proksch et al., 2015) 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 (Proksch et al., 2015) for the 15 days for which both data are available (same dataset as used for the statistical modeling). Similarly, the observed SSA from IceCube measurements are presented against the SMP-derived SSA from Eq. (2) and from (Proksch et al., 2015) for the 13 days for which both data were available (same dataset as used for the statistical modeling). 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 using the height of a thin persistent well-defined layer (described in Sec. 6) instead of the snow surface , leading to a better
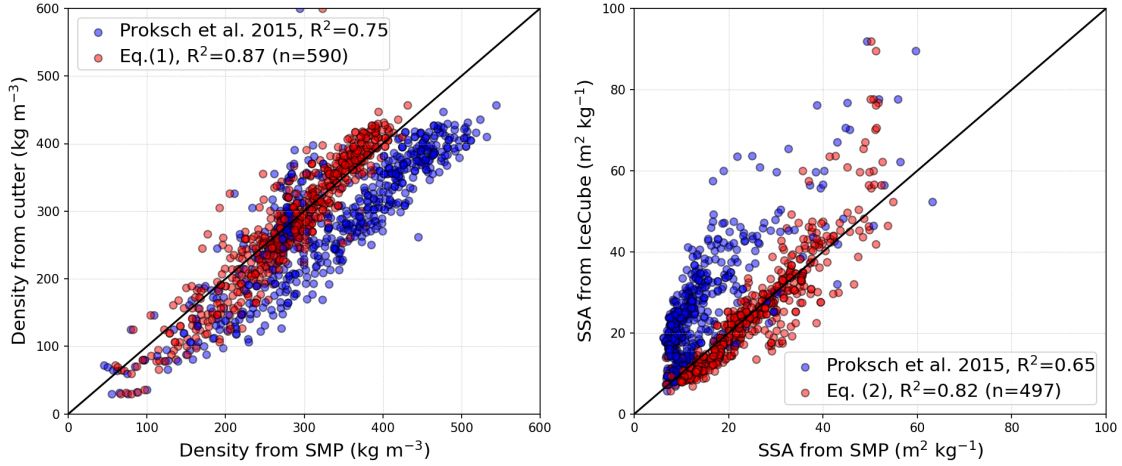
10

Figure 2. Left: Density from cutter measurements against density derived from SMP data using the parametrization of Proksch et al. (2015) (blue circles) and the re-calibrated present parametrization (red circles). Right: SSA from IceCube measurements against SSA derived from SMP data using both parametrizations.

vertical match of the profiles and thus a better correlation between estimates from SMP and snow pit measurements. with the snow surface and cropped to the length of the shortest profile. As expected, SMP-derived properties are closer to the snow pit measurements when using the present parametrizations. Between Applying a simple linear correlation between ρ_{cutter} and ρ_{smp}, a R^2 coefficient of 0.84 is 0.87 and a root-mean square deviation (RMSD) of 34 kg m^{-3} are found when using Eq. (1) against 0.73 a R^2 of 0.75 and a RMSD of 69 kg m^{-3} when using the parametrization of Proksch et al. (2015). Between SSA$_{\text{ic}}$ and SSA$_{\text{smp}}$, a R^2 coefficient of 0.81 is 0.82 and a RMSD of 7 m^2 kg^{-1} are found when using Eq. (2) against 0.64 a R^2 of 0.65 and a RMSD of 14 m^2 kg^{-1} when using the parametrization of Proksch et al. (2015). HenceIn the following, the present parametrizations Eq. ((1)) and ((2)) were applied to retrieve density the entire SMP data, so that a daily density profile and a daily SSA profile at 0.5 mm vertical resolution were retrieved from the daily median signal of $\tilde{F}$ and SSA from the entire SMP data$L$.

5.2 Layer tracking

We present a method to track particular layers of the snowpack throughout the season and retrieve their properties. This method allows evaluating will allow to evaluate later the measurement methods and /or simulation results by comparing the properties of the tracked layers, as presented in Section 6.2 and 6.3. To do so, them in selected layers.

The first step is to define which are the layers of interestwere defined by a upper and lower boundary, a boundary being detected by a significant, often , knowing that this method is only possible with layers that contrast well enough with their surrounding, so their boundaries can be easily identified by a rather sharp transition in the vertical profile of snow proper-

ties~~(either density, SSA, or penetration force). ~~. In this study, we chose to track four directly adjacent layers located in the bottom part of the snowpack called the DH-layer (depth hoar), the MF-layer (melt forms), the FC-layer (faceted crystals), and the RG-layer (rounded grains), from bottom to top, referring to the predominant grain shape observed in the layer. They are described in details in the next section. We chose these layers because they are among the main stratigraphic features of the snowpack observed during the winter, showed a wide range of snow types and properties, could be tracked over the entire winter, and were relatively easy to identify (rather sharp property transitions).

In the measurements data, the four layers of interest were defined by the height of their upper and lower boundaries. Boundaries were manually identified ~~for all measurement methods~~ by simply looking at the property profiles~~(density, SSA, or penetration resistance) and reporting heights. For the SMP data, this~~, looking for sharp and relevant transitions, and recording heights. This step was performed on ~~the median profiles of penetration resistance force computed~~ all the weekly density profile from the cutter and SSA profile from IceCube, as well as on all the daily representative profile of penetration force resistance obtained from the five daily SMP measurements. The identification of layer boundaries was sometimes challenging for weak stratigraphic transitions, e.g. the transition between a layer of fresh snow that fell onto a soft snow layer. To ~~this end, boundaries were~~ help in such cases, boundaries could be backtracked in time, starting from a profile where the layer ~~is older (typically 1 month after its deposition)~~ of interest is older and its boundaries more clearly detectable. ~~When approaching the date of the layer deposition~~Also, additional information, such as observed height of new snow, was ~~sometime~~ sometimes used to help delineate boundaries. ~~Once boundaries~~Based on the referenced boundaries, bulk properties of the layers of interest were ~~defined on all measurements of our dataset, layer properties were computed~~ computed for each date by averaging data within ~~heights given by the referenced boundaries~~the recorded heights.

~~We used a different method to identify layers~~ To identify the layers of interest in the SNOWPACK simulations~~based on the layer deposition date that~~, we used their date of deposition, which is one of the ~~layer properties~~standard layer parameter simulated by SNOWPACK. To do so, we attributed a time stamp (YYMMDD) to each of the defined layer boundary that corresponds to the date of deposition of the adjacent layer above ~~the given boundary (date of burial)~~it. Time stamps were determined using automatic weather station data as well as ~~the~~ daily manual observations of the snow surface. A layer of interest was then simply defined as all the simulated layers with a deposition date older than the time stamp of its lower boundary ~~but~~ and younger than the time stamp of its upper boundary.

~~Four distinct layers were~~ This way, the four layers tracked in this study ~~and consist in four directly adjacent layers located in the bottom part of the snowpack. We choose these layers because they are among the main stratigraphic features of the snowpack observed during the winter, showed a wide range of snow types and properties, could be tracked over the entire winter, and were relatively easy to identify (rather sharp property transitions). These tracked layers are called the DH-layer (depth hoar), the MF-layer (melt forms), the FC-layer (faceted crystals), and the RG-layer (rounded grains), from bottom to top layers, referring to the predominant grain shape observed in the layer. They are described in details in the next section. These four layers were~~ were identified based on four boundaries called 151201-boundary, 151202-boundary, 160102-boundary, and 160117-boundary, from the lower to the upper boundary, and the ground. ~~This way, the~~The DH-layer was comprised between

the ground and the 151201-boundary, MF-layer between the 151201-boundary and the 151202-boundary, FC-layer between the 151202-boundary and the 160102-boundary, and RG-layer between the 160102-boundary and the 160117-boundary.

6   Dataset analysis

This section presents a basic analysis of the RHOSSA campaign alongside with measurement inter-comparisons and a prelim-
inary evaluation of the SNOWPACK simulations. To ~~compensate for the inevitable height mismatches of the vertical property profiles, inherent to the snow spatial variability and measurements variability (e.g. Hagenmuller and Pilloix, 2016), all the~~ present the evolution of profile properties with time, profiles presented in the following were re-aligned such as $z = 0$ cm cor-
responds to the height of the upper boundary of the MF-layer (i.e. the ~~151202-boundary).~~ 20151202-boundary). Choosing this layer as a height reference leads to a better visual match than by simply taking the ground as reference (the field site ground at WFJ being uneven).

6.1   Evolution of weather, snow stratigraphy and stability

To provide background information for the origin of stratigraphic features of the season, Figure 3 shows the seasonal evolution of air and snow surface temperature as well as total snow height and height of new snow over 24 hours. The bi-weekly traditional profiles observed between 14 December 2015 and 15 March 2016 are presented in the upper ~~caption~~ panel of Figure 4. We can first note that winter 2015-2016 showed a below-average snow height, especially at the beginning of the season (Fig. 3). End of November, the winter started with a precipitation event after which the snow height reached approximately 40 cm. Thereafter, a dry period followed during which snow surface temperature remained between -20°C and -10°C, allowing large temperature gradients to build up across the shallow snowpack. Traditional profiles show that this basal layer recrystallized predominantly into depth hoar (dark blue colored layers below 0 cm in Fig. 4, upper panel), although faceted crystals and melt forms were sometimes also reported (light blue and red colored layers), and persisted throughout the season. This basal layer corresponds to the tracked layer referred as the DH-layer (Sec. 5.2). On the late afternoon of 1 December 2015, observers from the nearby ski resort reported rainfall up to 2600 m, and measured snow surface temperature reaching 0°C while the air temperature remained colder (see inset in Fig. 3) indicating freezing rain. This rainfall event led to the formation of a melt-freeze crust / rain crust at the snow surface, as reported in the traditional profile that followed on 14 December (Fig. 4, red and turquoise colored layer at 0 cm). This crust was persistent throughout the season and tracked as the MF-layer. Mid-December, about 10 cm of new snow accumulated on this crust and recrystallized into faceted crystals by the end of December, favored by a period of rather clear weather leading to low snow surface temperatures (Fig. 3). Again, this layer of faceted crystals was observed throughout the season (light blue colored layers between about 0 cm and 10 cm in Fig. 4, upper) and corresponds to the tracked FC-layer. January was generally characterized by a more cloudy weather with consistent precipitation events (Fig. 3). With the first snow falls early January, snow accumulated on top of the FC-layer and was quickly buried by the subsequent heavy precipitation events, being buried under around 75 cm of snow by mid-January. This layer was protected from significant temperature gradients and evolved into small faceted crystals and rounded grains (light blue and light red

13

colored layers between about 10 to 25 cm in Fig. 4). As this layer showed systematically a higher hand hardness (4 fingers against 1 finger) and a smaller grain size (not shown) than the FC-layer and DH-layer, this layer was named RG-layer for a sake of differentiation. Finally, after further precipitation events mostly occurring early February and early March, the snowpack height reached about 200 cm by mid March and consisted mostly of layers of rounded grains on a weaker base of facets and depth hoar.

The snowpack stratigraphy simulated by SNOWPACK is shown in the lower panel of Figure 4. Qualitatively, modeled stratigraphy compared well with observed stratigraphy. Indeed, although many subtle differences in grain shape and hand hardness exist throughout the season, the major stratigraphic features are well reproduced, notably the weak base layers (DH-layer and FC-layer) as well as the overlying slab which mostly consisted of small rounded or faceted grains for which the hardness increases from top to bottom. Note also the lower density of the base layer compared to the overlying slab. One major discrepancy is that the melt-freeze / rain crust which formed on 1 December (MF-layer) was not simulated by SNOWPACK (see dedicated comment in Sec. 7.3). Instead, SNOWPACK simulated around 3 cm of new snow, which later re-crystallized into faceted crystals.

Snow stability tests showed that the weak base, namely the DH-layer and FC-layer, were the most critical weak layers during most of the season. As shown in Figure 5, both layers consistently failed in CT and ETC until the beginning of February. Thereafter, these layers were not reactive anymore as tapping on the snow surface was not affecting the weak base buried below the hard and thick slab (black symbols in Fig. 5). From the PST, it was possible to follow the evolution of the critical crack length throughout the season (crosses in Fig. 5). Overall, the critical crack length increased steadily from about 20 cm in mid-January to around 60 cm beginning of March for both FC-layer and DH-layer, indicating weak layers less and less prone to crack propagation with time. Note that the critical crack length was consistently lower for the DH-layer than for the FC-layer.

6.2 Evolution of density

Figure 6 presents the evolution of the density profile during ~~the course of the~~ winter, as recorded from density cutter measurements, derived from SMP measurements, and simulated by SNOWPACK. Boundaries of the tracked layers are identified with solid black lines. The snowpack evolution is characterized by the punctual presence of new snow at the surface, showing the lowest density values down to about 50 kg m^{-3}. Overall, snow gets gradually denser upon deeper burial in the snowpack and as the season progresses, reaching density values as high as 450 kg m^{-3} in the middle of the snowpack by mid-winter. Despite located in the bottom of the snowpack, the persistent weak layers (DH-layer and FC-layer) remain significantly lighter than the adjacent layers. Finally, density of the MF-layer remains roughly constant throughout the winter at around 350 kg m^{-3}.

Although these features are consistently reported by both measurement methods, many stratigraphic details are only revealed by the SMP measurements and are not captured by the cutter measurements. The high temporal and spatial resolution of the SMP measurements allows indeed following almost continuously the evolution of density with time. For instance, we can clearly follow the density evolution of the 2 cm thick snow layer from its formation on February 22 showing density values around 350 kg m^{-3} (layer located at 145 cm height on February 22 in Fig. 6b) to mid-March when buried under about 40 cm
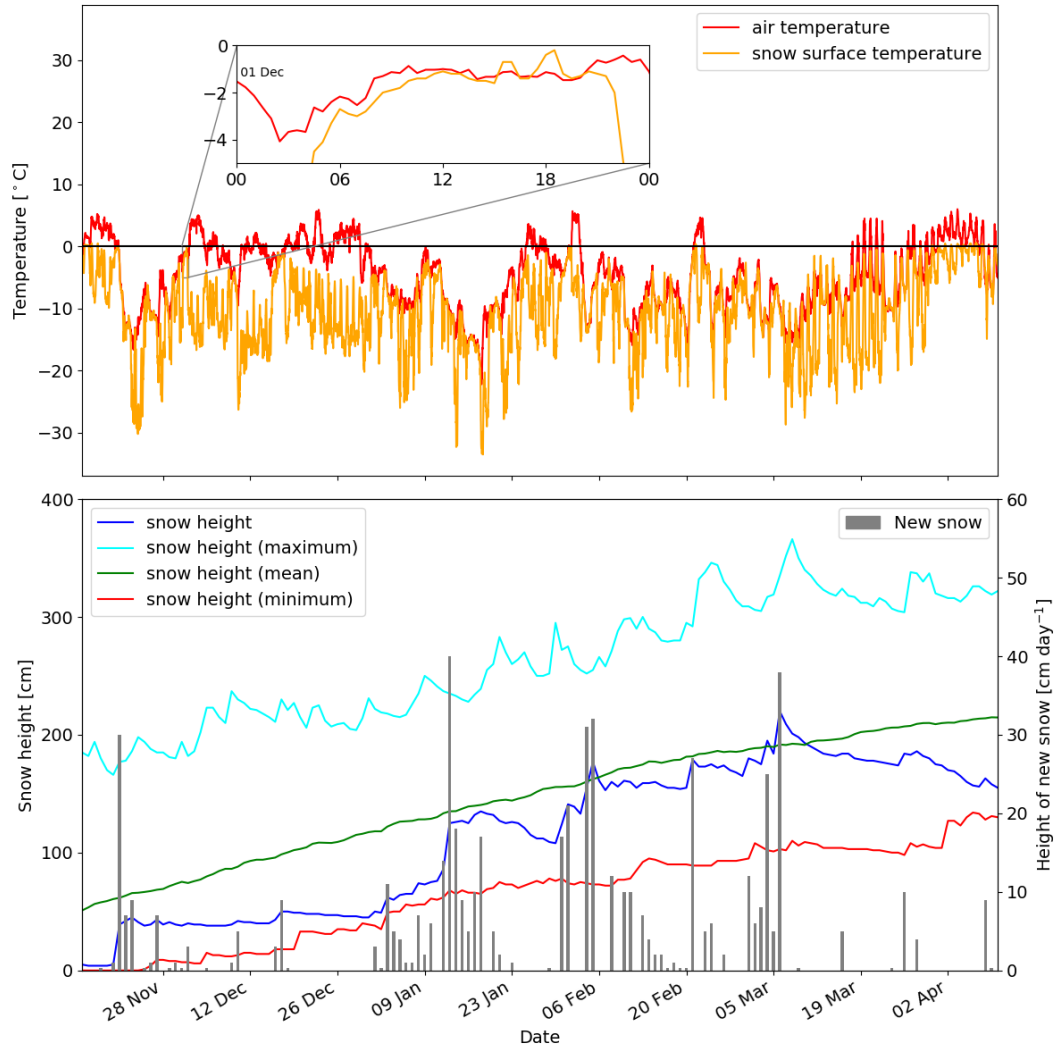
14

Figure 3. Top: Evolution of air temperature (red) and snow surface temperature (orange) at the WFJ site during winter 2015-2016. The inset shows data recorded on 1 December 2015 when the MF-layer formed. Bottom: Seasonal evolution of snow height (blue) and height of new snow (gray bars). For context, the 80 year daily maximum (cyan), minimum (red) and mean (green) snow height are also shown.
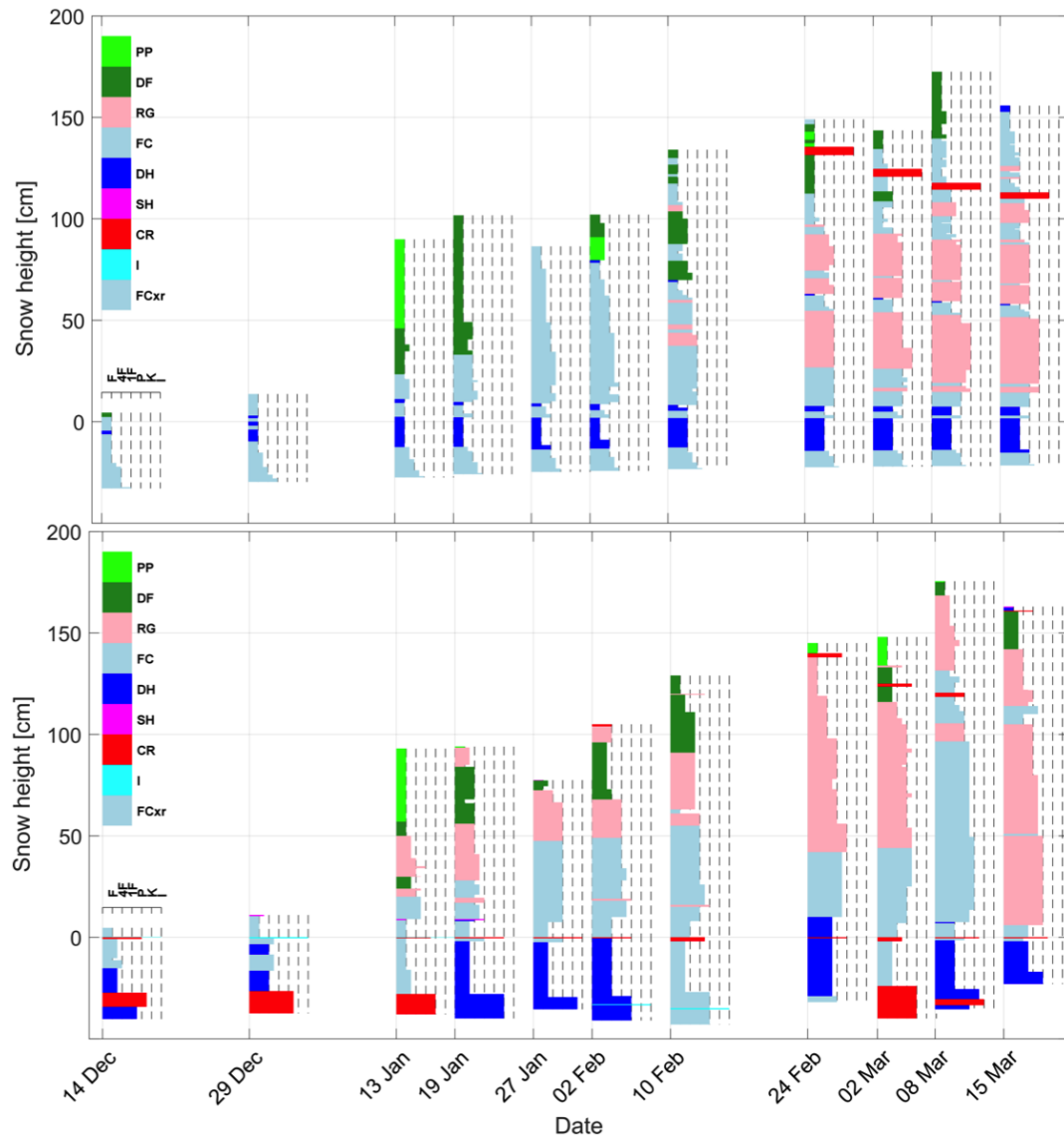
Figure 4. (top) Manual snow profiles observed during the 2015-2016 winter season. The colors indicate the major grain shape (red: melt forms, light blue: faceted crystals, blue: depth hoar, pink: rounded grains, green: decomposing and fragmented particles, light green: precipitation particles) and the width indicates the hand hardness. Snow height is relative to the top of the MF-layer. (bottom) Simulated snow profiles for the same dates.
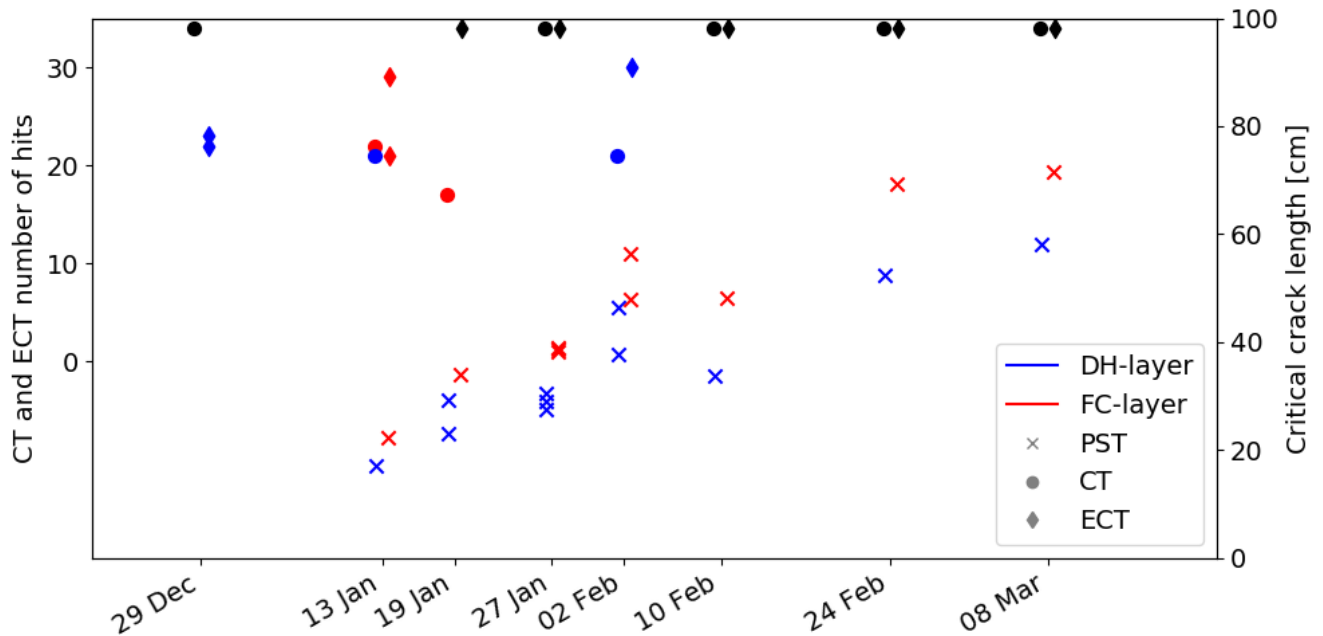
Figure 5. Stability tests results for the DH-layer (blue) and FC-layer (red). The number of hits for CT (circles) and ECT (diamonds) and the critical crack length obtained from the PST (crosses) are shown. Black symbols indicate that the CT or the ECT did not result in a failure in the layers.

of snow but still showing similar density values (layer located at 115 cm height on March 15 in Fig. 6b). The evolution of this layer is not or only diffusely captured by the cutter measurements. Note that this layer was reported in the traditional profiles from the 24th of February on as a layer of melt forms with a hand harness of one fist (Fig. 4).

~~Allowing further comparisons, Figure 7 provides a comparison of~~ The next figures allow comparing tomography, cutter,
5   and SMP measurement, as well as simulations from SNOWPACK, in greater details. Figure 7 shows the vertical profiles of density ~~on January 13, 2016 and March 2, 2016.~~ for six days of the season. Figure 8 shows the evolution of density for the 4 tracked layers DH-layer, MF-layer, FC-layer, and RG-layer throughout the winter. Both figures highlight an overall consistency between measurements. A slightly larger scatter is observed in the density evolution of the MF-layer (Fig. 8b), which might be partly due to uncertainties in the definition of the layer boundary (see Sec. 7.1). One can also note the decrease in density
10   recorded by the last two cutter measurements for the DH-layer and FC-layer (Fig. 8a and 8c). This might reflect a measurement bias that can occur when sampling fragile snow layers (under-sampling).

Simulations of the density profiles over the season agree overall well with the observations (Fig. 6c). The mis-modeling of the MF-layer, as mentioned earlier, leads however to large local deviations. Moreover, SNOWPACK seems to overestimate the densification rate of the DH-layer and FC-layer, leading to significantly higher modeled values by mid-March (Fig. 8a and 8c).
15   This overestimation can also be observed in the vertical ~~profile of March~~ profiles for both weak layers for example (Fig. 7~~b~~b-f).

17

Inversely, densification rate seems to be underestimated for layers evolving from fresh snow to rounded grains in the upper part of the snowpack, leading to simulated densities lower than the measured ones by mid-march, as shown in Figure 6 and 7~~b~~ f (layers from about 20 to 100 cm height). Finally, other inconsistencies can be observed locally in the simulated stratigraphy, such as the two relatively denser layers observed near the surface in March 2 at around 125 cm and 135 cm (Fig. 7b).

6.3 Evolution of SSA

Figure 9 shows the evolution of the SSA profiles over the course of the winter from IceCube measurements, from SMP measurements, and from SNOWPACK simulations. Note that IceCube measurements could not be performed on 19 January 2016 and 10 February 2016. SSA values range from about 70 m^2 kg^{-1}, for fresh snow layers at the surface, to about 5 m^2 kg^{-1}, in the bottom part of the snowpack. The MF-layer, well identifiable in terms of density (Fig. 6a and b), is here difficult to distinguish from the DH-layer and the FC-layer due to their similar SSA values. The general trend of the SSA evolution is an overall decrease with time and depth. The impact of the spatial and temporal resolution is again highlighted. For instance, the evolution of the layer deposited on February 22, easily identified by lower SSA values (greenish colors) than the ones of the adjacent layers, is clearly captured by the SMP measurements but only diffusely reported in the IceCube data.

To compare further, the vertical profiles of SSA ~~of 13 January 2016 and 2 March 2016~~ for 6 days of the winter are shown in Figure 10 ~~,~~ and the temporal evolution of the SSA of the 4 distinct layers (DH-layer, MF-layer, FC-layer and RG-layer) is presented in Figure 11. In particular, the latter figure allows analyzing the SSA decrease with time. The RG-layer shows the largest decrease, especially shortly after deposition when SSA evolves from about 45 to 20 m^2 kg^{-1} within one week. The SSA decay in the MF-layer and the DH-layer is slower, decreasing from about 15 to 10 m^2 kg^{-1} within the whole course of the season.

Both figures highlight significant disagreements between measurement methods. ~~To further investigate this issue, Figure ??~~ ~~presents vertical profiles of SSA for all the five dates when tomography, IceCube and SMP measurements were performed. In~~ ~~all profiles~~Looking at the vertical profiles (Fig. 10), SSA values from IceCube measurements are systematically higher than values from tomography measurements, by a factor of about 1.3. Besides this systematic bias, large deviations are found on 13 January 2016 in the upper half of the snowpack, for which SSA values from IceCube measurements range from 60 and 100 m^2 kg^{-1}, whereas values from SMP measurements do not exceed 50 m^2 kg^{-1} (Fig. ~~??~~10b, upper 60 cm). Possible causes for these deviations are discussed in Section 7.3.

Finally, SNOWPACK overall underestimates SSA compared to measurements (Fig. 9, 10, and 11). Deviations are higher with the IceCube data than the tomographic data, for which some good agreements can locally be found, for instance when looking at the SSA evolution of the tracked layers from mid-January on (excluding the MF-layer).

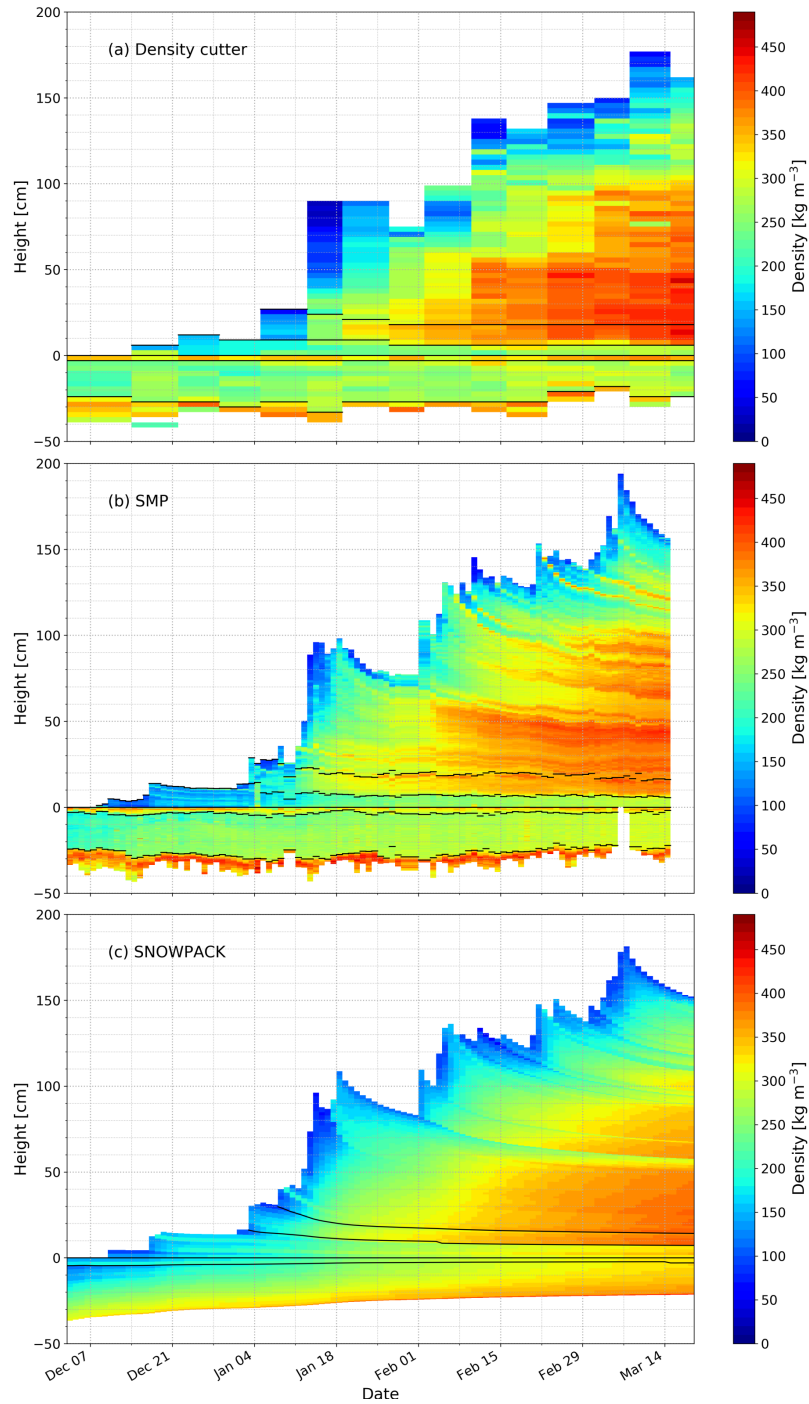~~Comparisons of SSA profiles from tomography, IceCube, and SMP measurements.~~

Figure 6. Evolution of the density profile during winter 2015-2016 (a) from cutter measurements, (b) derived from SMP measurements, and (c) simulated by SNOWPACK. Boundaries shown with black lines allow identifying the 4 tracked layers (DF-layer, MF-layer, FC-layer and RG-layer, from bottom to top). Measurements below the lowest boundary shown in SMP and cutter data were not considered part of DF-layer.
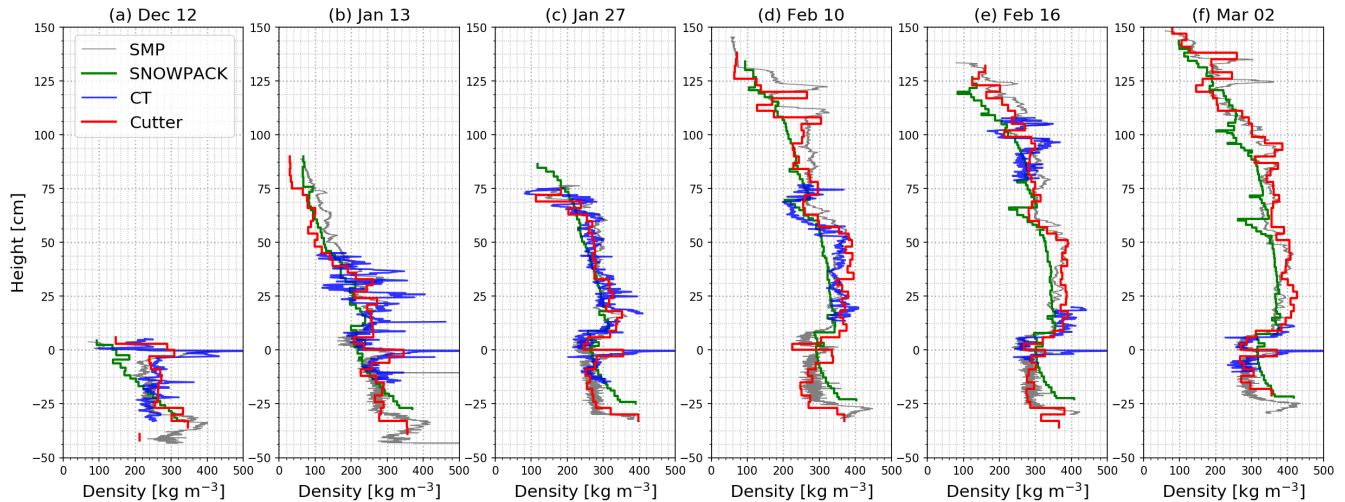
Figure 7. Vertical ~~profile~~ profiles of density ~~on 13 January 2016 and 02 March 2016~~ from ~~SMP~~tomography, ~~cutter~~IceCube, and ~~tomography~~ SMP measurements as well as modeled by SNOWPACK.

7 Discussion

7.1 The RHOSSA dataset for snow model evaluation

The presented dataset can be utilized as validation data for the evaluation of snow model outputs for the case of a dry alpine snowpack and over one winter season. Output parameters that can be evaluated are density, SSA, critical cut length, traditional snow pit measurements (grain size, grain type, hardness, temperature) and results from compression and extended compression tests. Snow models can be driven using the optimized forcing dataset, which includes meteorological and snow data from automatic and manual observations, provided in this study (Sec. 4). The RHOSSA dataset alone does not allow for robust and complete model evaluations, as model performances can vary depending on years and sites (Essery et al., 2013; Krinner et al., 2018). Yet, the snowpack monitored over winter 2015-2016 offered a wide range of alpine snow type and property variations throughout the season. It included typical persistent weak layers at the bottom of the snowpack (DH-layer and FC-layer) relevant for stability assessment for avalanche risk forecasting. Although the study focused on dry snowpack, some rain/melt events are also represented by the presence of several melt-refreeze crusts.

The specificity of the RHOSSA dataset is to provide time-series of density and SSA at a daily frequency and with a vertical resolution of ~~1~~0.5 mm, in contrast with previous validation datasets (weekly to bi-weekly, vertical resolution of 3 cm or higher) (e.g. Morin et al., 2013; Leppänen et al., 2015). Both temporal and spatial resolution are critical to account for in snow models because thin layers as well as processes occurring within short-time scales can have a significant impact on the snowpack behavior, e.g. on its mechanical stability (e.g Jamieson and Johnston, 1992). We highlight the need of high resolution datasets, as provided here, to evaluate the simulation of such features and processes.
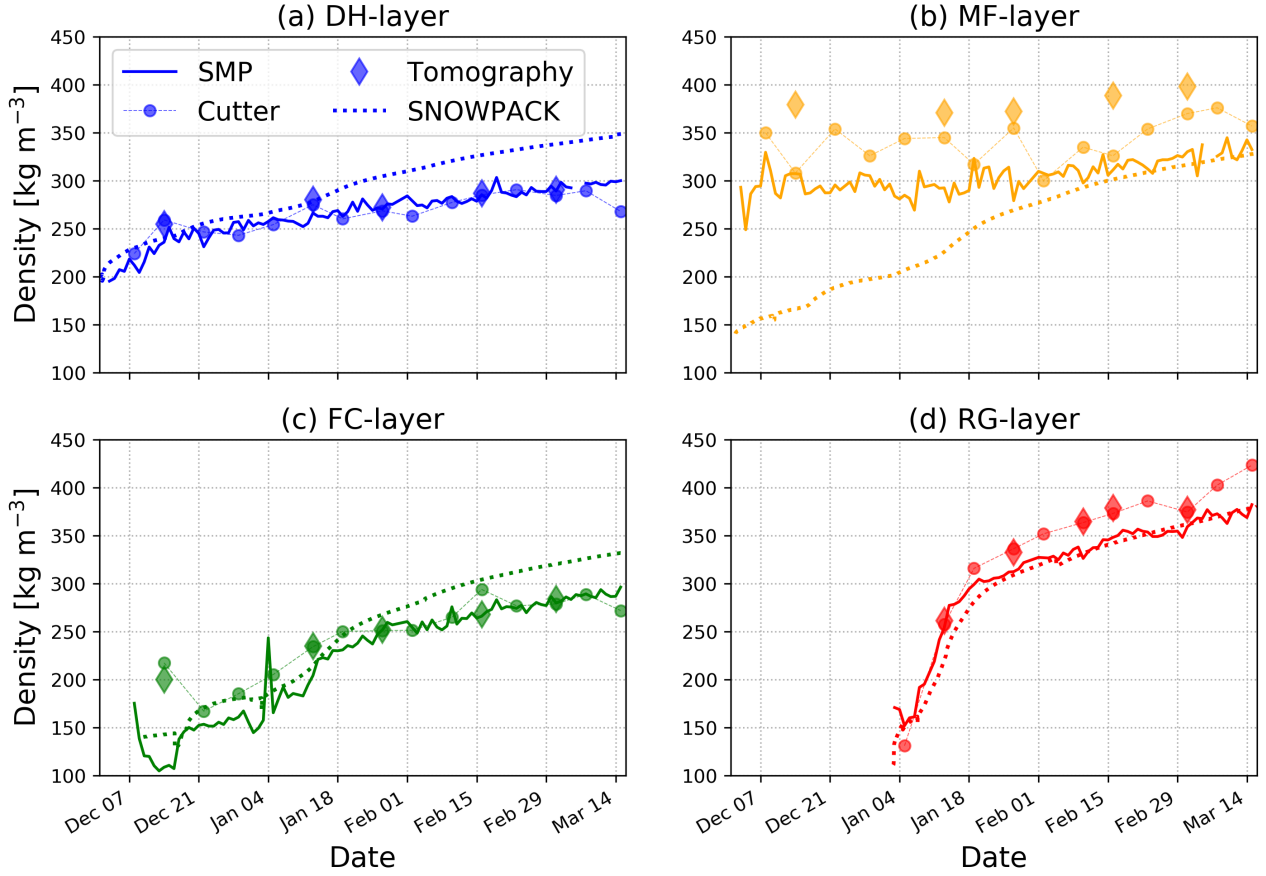
20

Figure 8. Density evolution of the 4 tracked layers from SMP, density cutter, and tomography measurements as well as modeled by SNOW-PACK.

In addition to validation datasets, comparison methods are also crucial when assessing models. Different methods were presented in the past to compare measurements and simulations: i) the comparison of averaged (bulk) values over the entire snowpack height (e.g. Landry et al., 2014; Leppänen et al., 2015; Essery et al., 2016), which is easy to implement but provides rather limited information, ii) the comparison of paired-values at the same height of the snowpack, which allows assessing the snowpack stratigraphy (e.g. Lehning et al., 2001; Morin et al., 2013) (as in Fig. 7 and 10), and iii) the comparison of values averaged within boundaries of specific layers of the snowpack, as used in Wever et al. (2015) and in this study (Fig. 8 and 11). This latter method seems particularly suitable to assess the skill of parameterizations of internal snow processes, e.g. temporal evolution of density and SSA of a fresh snow layer or of a buried layer of surface hoar. Layer properties evolution are indeed very close to the formulation of equations in a Lagrangian model. The method ii) and iii) bear with uncertainties from vertical mismatches that might contribute to the scatter between measurements and simulations and should thus be first corrected. When comparing paired-values at the same height, the prior alignment of the profiles is necessary. In the present case, we could simply
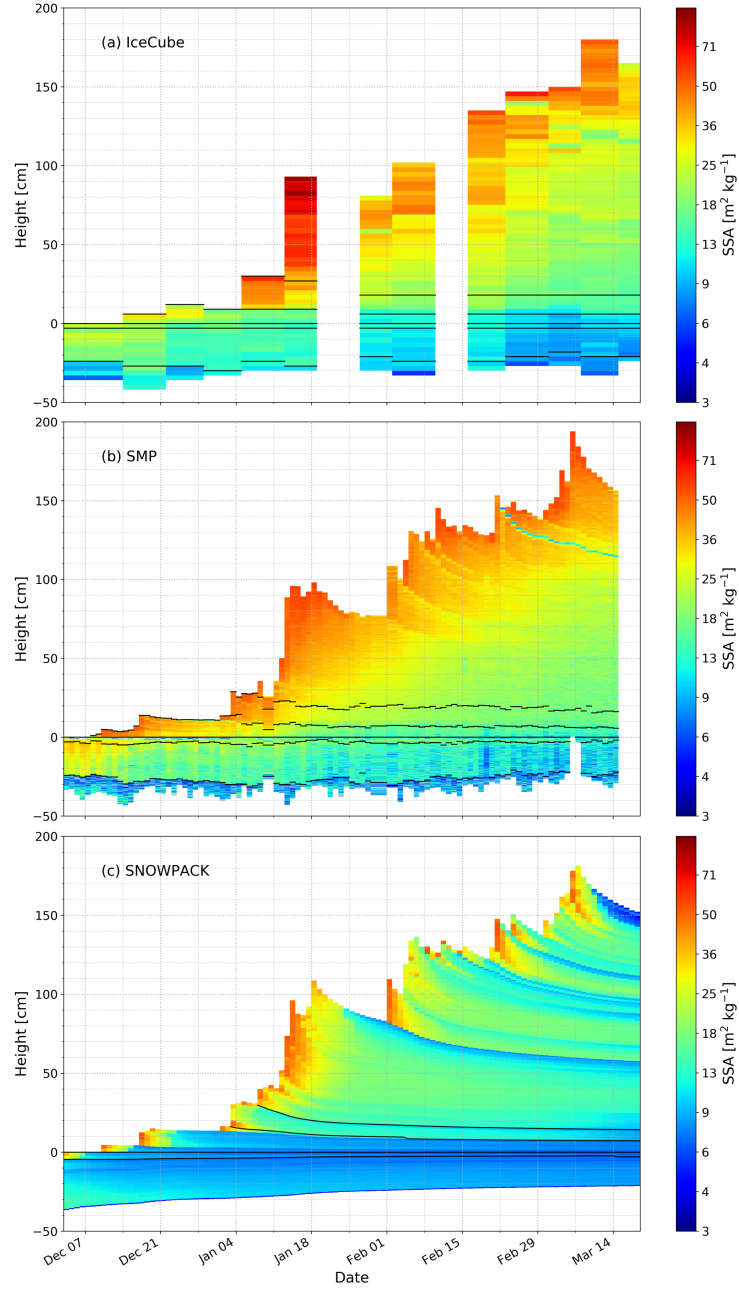
Figure 9. Evolution of the SSA profile during winter 2015-2016 (a) from IceCube measurements, (b) derived from SMP measurements, and (c) simulated by SNOWPACK. Boundaries shown with black lines allow identifying the 4 tracked layers as described for Figure 6.
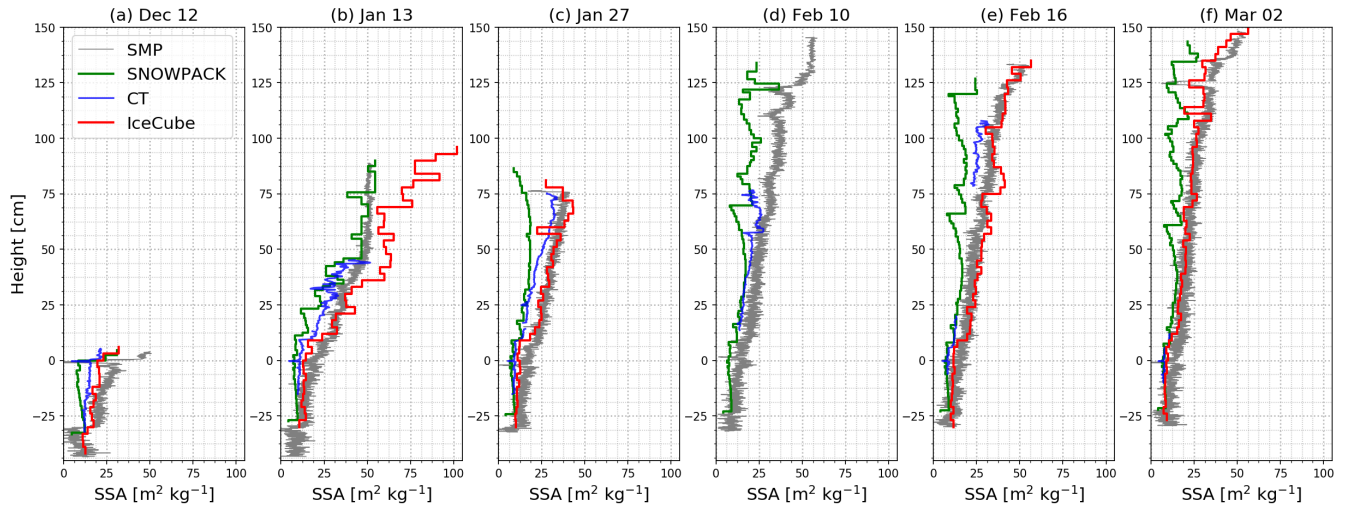
Figure 10. Vertical ~~profile~~ profiles of SSA ~~on 13 January 2016 and 02 March 2016~~ from ~~SMP~~tomography, IceCube, and ~~tomography~~ SMP measurements as well as modeled by SNOWPACK.

re-align the profiles thanks to the presence of the dominant MF-layer in all measurement methods and throughout the season. Slight vertical mismatches can however be found. For example, the density profile of March 2, 2016 (Fig. 7) shows two distinct denser layers at around 125 cm and 135 cm height which are well identified in both SMP and density cutter measurements but with a height mismatch of about 5 cm. This re-alignment method based on the identification of a persistent and well-defined

5 snowpack feature might however not be always applicable. A more systematic approach could be the algorithm presented by Hagenmuller and Pilloix (2016) to automatically match snow profiles by adjusting their layer thicknesses. This methods has a strong potential for quantitative comparison studies (Hagenmuller et al., 2018). When comparing properties of specific layers, the definition of the layers boundaries is critical. The ~~second-order~~ fluctuations observed in the evolution of density and SSA of the MF-layer (Fig. 8 and 11), especially visible in the SMP data, might possibly result from the boundaries definition of

10 this layer, in addition to the natural spatial variability of snow. Besides, the manual definition of boundaries is rather time-consuming if numerous layers are tracked. A more automatic method could be developed. In this respect, the RHOSSA data constitutes a valuable resource due to the continuity of the spatio-temporal picture of the seasonal evolution of stratigraphy.

7.2 The potential of daily SMP measurements

With daily SMP measurements, the RHOSSA campaign allows following the evolution of the internal structure of a snowpack

15 at a sub-centimeter vertical resolution almost continuously over 4 months - up to now inaccessible. An unparalleled, smooth picture of the spatio-temporal evolution of density and SSA is revealed, contrasting with data from the classical snow pit measurements (Fig. 6 and 9). Many thin stratigraphic features are indeed clearly visible in the SMP data but only diffusely shown by the manual measurements. This highly detailed picture of the snowpack evolution opens new opportunities for field
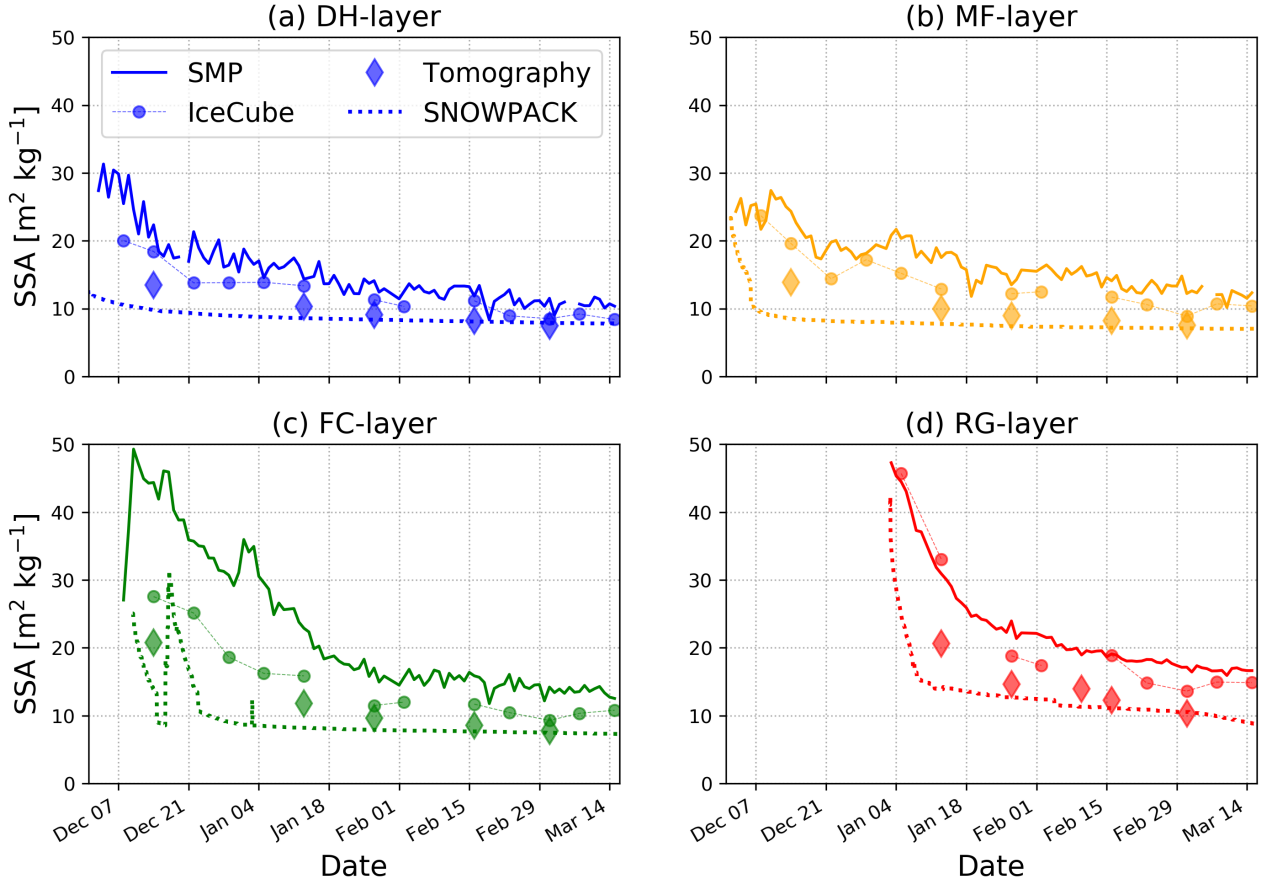
Figure 11. SSA evolution of the 4 tracked layers from SMP, IceCube, and tomography measurements as well as modeled by SNOWPACK.

studies on snowpack processes occurring over short-time scales (e.g. densification of fresh snow) or very localized (e.g. rain crust or surface hoar formation), as well as refined evaluation of snow models as already mentioned.

One advantage of SMP measurements compared to snow pit measurements is they are relatively faster (of the order of 30 minutes for five measurements) and thus more suitable for daily snowpack monitoring. It is however important to keep in mind

5 that density and SSA are not directly *measured* by the SMP but *derived* from the force signal based on parameterizations (Fig. 2), bearing additional uncertainties comparing to other more direct measurements. Several parametrizations were previously put forward to derive density and/or SSA from SMP signals (e.g. Pielmeier and Schneebeli, 2003; Dadic et al., 2008; Proksch et al., 2015; Kaur and Satyawali, 2017). Differences between the parameterizations of density and SSA of Proksch et al. (2015) and the ones presented in this study are ~~likely~~ due to the version of the SMP device which has undergone an update of the

10 electronics in version 4 that affected the inversion of the model Löwe and van Herwijnen (2012) through the force correlation function. We would hope that the parameterization Eq. (1) and (2) are generally applicable to an SMP version 4. However, without an independent validation by measurements under different snowpack conditions, it is not possible to state the range

of validity of the parametrizations presented here. In the long term, it would be desirable to improve the underlying stochastic-mechanical approach (Löwe and van Herwijnen, 2012) by an invertible model that contains density and SSA to retrieve these parameters from a more physical picture of the penetration process.

7.3 Comparing density and SSA estimates

As possible starting points to future dedicated studies, we sum up here the main deviations reported in this paper when comparing density and SSA estimates. 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.

We report a significant and systematic inter-measurement deviation in the SSA estimates. Values from IceCube and SMP are systematically higher than values computed on tomographic images, approximately by a factor 1.3 (Fig. ??10). A comprehensive comparison between optical methods, such as IceCube, and tomography seems very much needed to understand this systematic deviation. Besides, large disagreements were reported on the specific day of January 13, 2016, for which IceCube data range from 60 to 100 m^2 kg $^{-1}$ whereas SMP data show values around 50 m^2 kg $^{-1}$ (Fig. ??10b). That day, measurements were performed during a snowfall in light freshly-deposited snow. When measuring SSA of light snow, typically for values above 60 m^2 kg^{-1}, the emitted radiations can interact with the bottom of the sample holder during the measurement causing an overestimation of the SSA (Gallet et al., 2009; Zuanon, 2013). Another possible cause is that the present statistical model used to derive SSA from SMP measurements fails to reproduce the high SSA values of newly-deposited snow because of their under-representations (one day) in the IceCube dataset used for calibration; (similarly but to a lower extent, disagreements are found in the upper 20 cm of the density profiles of the same day (Fig. 7a): SMP measurements fail to capture the very low density measured by the cutter method (60 kg m^{-3} vs. 30 kg m^{-3})). Note that one major discrepancy between IceCube and SMP-derived SSA can be directly linked to the SSA calibration Eq. (2) that leads largely to overestimate 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) as a large part of the snowpack shows SSA values below 20 m^2 kg^{-1}.

Comparing SNOWPACK outputs against observations, one significant deviation is the absence of the MF-layer in the simulations. This is due to the fact that the precipitation forcing scheme used in the present simulations does not allow representing rain fall events occurring at negative air temperatures. This inappropriate forcing could be improved by using diagnostic atmospheric variables to detect such events Quéno et al. (2018). Also, SNOWPACK underestimates SSA in overall (Fig. 9, 10, and 11). A similar bias was reported at an arctic site (Leppänen et al., 2015). On the contrary, a systematic overestimation of the SSA simulated by Crocus was recently pointed out (Tuzet et al., 2017). Evaluations can however be challenged by the significant inter-measurement deviations observed, as discussed above. The agreement between simulations and estimates from tomography is better than between simulations and estimates from SMP or IceCube. ~~Finally, SNOWPACK seems to slightly overestimate the densification rate of persistent weak layers, as observed in our study for the~~ Finally, recent publications point to contradicting performance of SNOWPACK to simulate the properties of depth hoar layers. While some studies report rather poor performance in matching observed density in Arctic environments (Domine et al., 2019; Gouttevin et al., 2018), others showed that SNOWPACK captures fairly well the density of basal layers in Alpine snowpacks (Wever et al., 2015). This study

shows the first comprehensive comparison of the evolution of modeled and observed layer densities. Although SNOWPACK reproduces overall reasonably the low density values of the persistent weak layers DH-layer and FC-layer (Fig. 7), it seems to overestimate the densification rates leading to overestimated values of density by mid-March (Fig. 8a and Fig. 8c). Barrere et al. (2017) reported similar findings with the model Crocus. The discrepancies pointed out here suggest further investigations and might guide possible model improvements.

8 Conclusions

During winter 2015-2016, the standard snow observation program of the WFJ site (Eastern Swiss Alps, elevation 2536 m) was complemented by additional measurements and stability tests, bridging between traditional and newly-developed measurement methods. This campaign results in a multi-resolution and multi-instrument dataset of structural and mechanical properties of the snowpack, referred to as the RHOSSA dataset. The dataset includes time-series of density, SSA, and critical cut length, traditional snow pit parameters, and results from compression tests. Profiles of density and SSA were monitored daily and with a vertical resolution of ~~1~~ 0.5 mm based on SMP measurements. These high-resolution data offers an unprecedented smooth and continuous picture of the snowpack evolution throughout the season.

~~Our specific results~~ The first results of the campaign presented in this work comprise (i) re-calibrated parameterizations to estimate density and SSA from SMP measurements for version 4, (ii) the comparison of density and SSA estimates from state-of-the-art measurement methods (Cutter/IceCube, tomography, SMP-derived), and (iii) the assessment of the SNOWPACK model against measurements. ~~Results from the two latter point contribute describing current states and suggesting further investigations~~ Our results indicate that further investigations are required in the future to draw firm conclusions about the two latter aspects. Our study demonstrates the potential of high temporal and spatial resolution dataset for the evaluation of the detailed snowpack models as Crocus or SNOWPACK. In this view, the RHOSSA measurements campaign could be extended to other snow observation sites to cover different environments and conditions.

9 Code and data availability

The dataset presented in the paper will be available on the EnviDat database (doi will be provided upon acceptance).

Author contributions. All authors contributed to the field measurements. B.R., H.L., and N.C. wrote the paper with input from C.F. and A.V.H. A.V.H., B.R., C.F., H.L., J.S., and N.C. performed the analysis of the data and the simulations. N.C. and M.S. directed the project.

Competing interests. The authors declare that they have no conflict of interest.

5

References

Arnaud, L., Picard, G., Champollion, N., Domine, F., Gallet, J.-C., Lefebvre, E., Fily, M., and Barnola, J.-M.: Measurement of vertical profiles of snow specific surface area with a 1 cm resolution using infrared reflectance: instrument description and validation, J. Glaciol., 57, 17–29, https://doi.org/10.3189/002214311795306664, 2011.

5 Barrere, M., Domine, F., Decharme, B., Morin, S., Vionnet, V., and Lafaysse, M.: Evaluating the performance of coupled snow–soil models in SURFEXv8 to simulate the permafrost thermal regime at a high Arctic site., Geoscientific Model Development, 10, 2017.

Brun, E., David, P., Sudul, M., and Brunot, G.: A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting, J. Glaciol., 38, 13 – 22, 1992.

Carmagnola, C. M., Morin, S., Lafaysse, M., Domine, F., Lesaffre, B., Lejeune, Y., Picard, G., and Arnaud, L.: Implementation and evaluation
10 of prognostic representations of the optical diameter of snow in the SURFEX/ISBA-Crocus detailed snowpack model, The Cryosphere, 8, 417–437, https://doi.org/10.5194/tc-8-417-2014, 2014.

Carroll, T.: A comparison of the CRREL 500 cm 3 tube and the ILTS 200 and 100 cm 3 box cutters used for determining snow densities, Journal of Glaciology, 18, 334–337, 1977.

Conger, S. M. and McClung, D. M.: Comparison of density cutters for snow profile observations, J. Glaciol, 55 (189), 163 – 169, 2009.

15 Dadic, R., Schneebeli, M., Lehning, M., Hutterli, M. A., and Ohmura, A.: Impact of the microstructure of snow on its temperature: A model validation with measurements from Summit, Greenland, J. Geophys. Res., 113, D14 303, https://doi.org/10.1029/2007JD009562, 2008.

Domine, F., Taillandier, A.-S., Houdier, S., Parrenin, F., Simpson, W. R., and Douglas, T. A.: Interactions between snow metamorphism and climate physical and chemical aspects, in: P. C. I., edited by Kuhs, W. F., pp. 27 – 46, Royal Society of Chemistry, Cambridge, UK, 2007.

Domine, F., Picard, G., Morin, S., Barrere, M., Madore, J.-B., and Langlois, A.: Major issues in simulating some Arctic snowpack properties
20 using current detailed snow physics models: Consequences for the thermal regime and water budget of permafrost, Journal of Advances in Modeling Earth Systems, 11, 34–44, 2019.

Dumont, M., Laurent, A., Picard, G., Libois, Q., Lejeune, Y., Nabat, P., Voisin, D., and Morin, S.: In situ continuous visible and near-infrared spectroscopy of an alpine snowpack, The Cryosphere, 11, 1091, 2017.

Essery, R., Morin, S., Lejeune, Y., and Bauduin-Ménard, C.: A comparison of 1701 snow models using observations from an alpine site,
25 Adv. Water Res., 55, 131–148, https://doi.org/10.1016/j.advwatres.2012.07.013, 2013.

Essery, R., Kontu, A., Lemmetyinen, J., Dumont, M., and Ménard, C. B.: A 7-year dataset for driving and evaluating snow models at an Arctic site (Sodankylä, Finland), Geoscientific Instrumentation, Methods and Data Systems, 5, 219–227, 2016.

Etchevers, P., Martin, E., Brown, R., Fierz, C., Lejeune, Y., Bazile, E., Boone, A., Dai, Y.-J., Essery, R., Fernandez, A., Gusev, Y., Jordan, R., Koren, V., Kowalczyk, E., Nasonova, N. O., Pyles, R. D., Schlosser, A., Shmakin, A. B., Smirnova, T. G., Strasser, U., Verseghy, D.,
30 Yamazaki, T., and Yang, Z.-L.: Intercomparison of the surface energy budget simulated by several snow models (SNOWMIP project), Ann. Glaciol., 38, 150 – 158, https://doi.org/10.3189/172756404781814825, 2004.

Fierz, C.: Field observation and modelling of weak-layer evolution, Annals of Glaciology, 26, 7–13, https://doi.org/10.3189/1998AoG26-1-7-13, 1998.

Fierz, C., Armstrong, R. L., Durand, Y., Etchevers, P., Greene, E., McClung, D. M., Nishimura, K., Satyawali, P. K., and Sokratov, S. A.:
35 The international classification for seasonal snow on the ground, IHP-VII Technical Documents in Hydrology n 83, IACS Contribution n 1, 2009.

Gallet, J.-C., Domine, F., Zender, C. S., and Picard, G.: Measurement of the specific surface area of snow using infrared reflectance in an integrating sphere at 1310 and 1550 nm, The Cryosphere, 3, 167 – 182, https://doi.org/10.5194/tc-3-167-2009, 2009.

Gaume, J., Herwijnen, A. v., Chambon, G., Wever, N., and Schweizer, J.: Snow fracture in relation to slab avalanche release: critical state for the onset of crack propagation, The Cryosphere, 11, 217–228, 2017.

5 Gauthier, D. and Jamieson, B.: Understanding the propagation of fractures and failures leading to large and destructive snow avalanches: recent developments, in: Proceedings of the 2006 Annual Conference of the Canadian Society for Civil Engineering, First Specialty Conference on Disaster Mitigation, Calgary, Alberta, pp. 23–26, 2006.

Gauthier, D. and Jamieson, B.: Evaluation of a prototype field test for fracture and failure propagation propensity in weak snowpack layers, Cold Regions Science and Technology, 51, 87–97, 2008.

10 Gouttevin, I., Langer, M., Löwe, H., Boike, J., Proksch, M., and Schneebeli, M.: Observation and modelling of snow at a polygonal tundra permafrost site: spatial variability and thermal implications., Cryosphere, 12, 2018.

Hagenmuller, P. and Pilloix, T.: A new method for comparing and matching snow profiles, application for profiles measured by penetrometers, Frontiers in Earth Science, 4, 52, 2016.

Hagenmuller, P., Chambon, G., Lesaffre, B., Flin, F., and Naaim, M.: Energy-based binary segmentation of snow microtomographic images, 15 Journal of glaciology, 59, 859–873, https://doi.org/10.3189/2013JoG13J035, 2013.

Hagenmuller, P., Viallon, L., Bouchayer, C., Teich, M., Lafaysse, M., and Vionnet, V.: Quantitative comparison of snow profiles, in: Proceedings of the International Snow Science Workshop Innsbruck - 2018, 7-12 October, Innsbruck, Austria, pp. 876–879, 2018.

Jamieson, J.: The compression test-after 25 years, The Avalanche Review, 18, 10–12, 1999.

Jamieson, J. and Johnston, C.: Snowpack characteristics associated with avalanche accidents, Canadian Geotechnical Journal, 29, 862–866, 20 1992.

Jonas, T., Marty, C., and Magnusson, J.: Estimating the snow water equivalent from snow depth measurements in the Swiss Alps, J. Hydrol., 378, 161–167, https://doi.org/10.1016/j.jhydrol.2009.09.021, 2009.

Kaur, S. and Satyawali, P.: Estimation of snow density from SnowMicroPen measurements, Cold Regions Science and Technology, 134, 1–10, 2017.

25 King, J., Derksen, C., Toose, P., Langlois, A., Larsen, C., Lemmetyinen, J., Marsh, P., Montpetit, B., Roy, A., Rutter, N., and Sturm, M.: The influence of snow microstructure on dual-frequency radar measurements in a tundra environment, Remote Sensing of Environment, 215, 242 – 254, https://doi.org/https://doi.org/10.1016/j.rse.2018.05.028, http://www.sciencedirect.com/science/article/pii/S003442571830258X, 2018.

Krinner, G., Derksen, C., Essery, R., and Flanner, M. G.: ESM-SnowMIP: Assessing models and quantifying snow-related climate feedbacks, 30 Geosci. Model Dev. Discuss., 2018.

Lafaysse, M., Cluzet, B., Dumont, M., Lejeune, Y., Vionnet, V., and Morin, S.: A multiphysical ensemble system of numerical snow modelling, The Cryosphere Discussions, 2017, 1–42, https://doi.org/10.5194/tc-2016-287, http://www.the-cryosphere-discuss.net/tc-2016-287/, 2017.

Landry, C. C., Buck, K. A., Raleigh, M. S., and Clark, M. P.: Mountain system monitoring at Senator Beck Basin, San Juan Mountains, 35 Colorado: A new integrative data source to develop and evaluate models of snow and hydrologic processes, Water Resources Res., 50, 1773–1788, https://doi.org/10.1002/2013WR013711, 2014.

Lehning, M., Fierz, C., and Lundy, C.: An objective snow profile comparison method and its application to SNOWPACK, Cold Reg. Sci. Technol., 33 (2–3), 253 – 261, https://doi.org/10.1016/S0165-232X(01)00044-1, 2001.

Lehning, M., Bartelt, P., Brown, B., and Fierz, C.: A physical SNOWPACK model for the Swiss avalanche warning: Part III: meteorological forcing, thin layer formation and evaluation, Cold Regions Science and Technology, 35, 169 – 184, https://doi.org/https://doi.org/10.1016/S0165-232X(02)00072-1, http://www.sciencedirect.com/science/article/pii/S0165232X02000721, 2002a.

5 Lehning, M., Bartelt, P., Brown, B., Fierz, C., and Satyawali, P.: A physical SNOWPACK model for the Swiss avalanche warning. Part II: snow microstructure., Cold Reg. Sci. Technol., 35, 147 – 167, https://doi.org/10.1016/S0165-232X(02)00073-3, 2002b.

Leinss, S., Löwe, H., Proksch, M., Lemmetyinen, J., Wiesmann, A., and Hajnsek, I.: Anisotropy of seasonal snow measured by polarimetric phase differences in radar time series, The Cryosphere, 10, 1771–1797, https://doi.org/10.5194/tc-10-1771-2016, https://www.the-cryosphere.net/10/1771/2016/, 2016.

10 Lejeune, Y., Dumont, M., Panel, J.-M., Lafaysse, M., Lapalus, P., Le Gac, E., Lesaffre, B., and Morin, S.: 57 years (1960–2017) of snow and meteorological observations from a mid-altitude mountain site (Col de Porte, France, 1325 m of altitude), Earth System Science Data, 11, 71–88, https://doi.org/10.5194/essd-11-71-2019, https://www.earth-syst-sci-data.net/11/71/2019/, 2019.

Leppänen, L., Kontu, A., Vehviläinen, J., Lemmetyinen, J., and Pulliainen, J.: Comparison of traditional and optical grain-size field measurements with SNOWPACK simulations in a taiga snowpack, Journal of Glaciology, 61, 151–162, 2015.

15 Leppänen, L., Kontu, A., Hannula, H.-R., Sjöblom, H., and Pulliainen, J.: Sodankylä manual snow survey program, Geoscientific Instrumentation, Methods and Data Systems, 5, 163–179, 2016.

Löwe, H. and van Herwijnen, A.: A Poisson shot noise model for micro-penetration of snow, Cold Reg. Sci. Technol., 70, 62–70, https://doi.org/10.1016/j.coldregions.2011.09.00, 2012.

Meister, R.: Snow profiling at Weissfluhjoch, in: Schweizer J,(SLF)(ed) International snow science workshop, Davos, 2009.

20 Ménard, C. B., Essery, R., Barr, A., Bartlett, P., Derry, J., Dumont, M., Fierz, C., Kim, H., Kontu, A., Lejeune, Y., Marks, D., Niwano, M., Raleigh, M., Wang, L., and Wever, N.: Meteorological and evaluation datasets for snow modelling at ten reference sites: description of in situ and bias-corrected reanalysis data, Earth System Science Data Discussions, 2019, 1–34, https://doi.org/10.5194/essd-2019-12, https://www.earth-syst-sci-data-discuss.net/essd-2019-12/, 2019.

Montpetit, B., Royer, A., Langlois, A., Cliche, P., Roy, A., Champollion, N., Picard, G., Domine, F., and Obbard, R.: New shortwave infrared albedo measurements for snow specific surface area retrieval, Journal of Glaciology, 58, 941–952, 2012.

25 Morin, S., Lejeune, Y., Lesaffre, B., Panel, J.-M., Poncet, D., David, P., and Sudul, M.: An 18-years long (1993 - 2011) snow and meteorological dataset from a mid-altitude mountain site (Col de Porte, France, 1325 m alt.) for driving and evaluating snowpack models, Earth Syst. Sci. Data, 4, 13–21, https://doi.org/10.5194/essd-4-13-2012, 2012.

Morin, S., Domine, F., Dufour, A., Lejeune, Y., Lesaffre, B., Willemet, J.-M., Carmagnola, C. M., and Jacobi, H.-W.: Measurements and modeling of the vertical profile of specific surface area of an alpine snowpack, Adv. Water Res., 55, 111–120, https://doi.org/10.1016/j.advwatres.2012.01.010, 2013.

Pielmeier, C. and Schneebeli, M.: Stratigraphy and changes in hardness of snow measured by hand, ramsonde and snow micro penetrometer: a comparison with planar sections, Cold Reg. Sci. Technol., 37, 393–405, http://linkinghub.elsevier.com/retrieve/pii/S0165232X0300079X, 2003.

35 Pinzer, B., Schneebeli, M., and Kaempfer, T.: Vapor flux and recrystallization during dry snow metamorphism under a steady temperature gradient as observed by time-lapse micro-tomography, The Cryosphere, 6, 1141–1155, 2012.

Proksch, M., Löwe, H., and Schneebeli, M.: Density, specific surface area, and correlation length of snow measured by high-resolution penetrometry, Journal of Geophysical Research: Earth Surface, 120, 346–362, 2015.

Proksch, M., Rutter, N., Fierz, C., and Schneebeli, M.: Intercomparison of snow density measurements: bias, precision, and vertical resolution, The Cryosphere, 10, 371–384, 2016.

Quéno, L., Vionnet, V., Cabot, F., Vrécourt, D., and Dombrowski-Etchevers, I.: Forecasting and modelling ice layer formation on the snowpack due to freezing precipitation in the Pyrenees, Cold Regions Science and Technology, 146, 19 – 31, https://doi.org/https://doi.org/10.1016/j.coldregions.2017.11.007, http://www.sciencedirect.com/science/article/pii/S0165232X1730054X, 2018.

Reba, M. L., Marks, D., Seyfried, M., Winstral, A., Kumar, M., and Flerchinger, G.: A long-term data set for hydrologic modeling in a snow-dominated mountain catchment, Water Resour. Res., 47, W07 702, https://doi.org/10.1029/2010WR010030, 2011.

Reuter, B., Schweizer, J., and van Herwijnen, A.: A process-based approach to estimate point snow instability, The Cryosphere, 9, 837–847, 2015.

Richter, B., Schweizer, J., Rotach, M. W., and van Herwijnen, A.: Validating and improving the critical crack length in the snow cover model SNOWPACK., The Cryosphere, 2019.

Schmucki, E., Marty, C., Fierz, C., and Lehning, M.: Evaluation of modelled snow depth and snow water equivalent at three contrasting sites in Switzerland using SNOWPACK simulations driven by different meteorological data input, Cold Regions Science and Technology, 99, 27 – 37, https://doi.org/https://doi.org/10.1016/j.coldregions.2013.12.004, http://www.sciencedirect.com/science/article/pii/S0165232X13002036, 2014.

Schneebeli, M., Coléou, C., Touvier, F., and Lesaffre, B.: Measurement of density and wetness in snow using time-domain reflectometry, Annals of Glaciology, 26, 69–72, 1998.

Schneebeli, M., Pielmeier, C., and Johnson, J. B.: Measuring snow microstructure and hardness using a high resolution penetrometer, Cold Reg. Sci. Technol., 30 (1–3), 101–114, https://doi.org/10.1016/S0165-232X(99)00030-0, 1999.

Schweizer, J. and Wiesinger, T.: Snow profile interpretation for stability evaluation, Cold Regions Science and Technology, 33, 179–188, 2001.

Sigrist, C. and Schweizer, J.: Critical energy release rates of weak snowpack layers determined in field experiments, Geophysical Research Letters, 34, 2007.

Simenhois, R. and Birkeland, K. W.: The extended column test: test effectiveness, spatial variability, and comparison with the propagation saw test, Cold Regions Science and Technology, 59, 210–216, 2009.

Stössel, F., Guala, M., Fierz, C., Manes, C., and Lehning, M.: Micrometeorological and morphological observations of surface hoar dynamics on a mountain snow cover, Water Resources Research, 46, 2010.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J., Koskinen, J., and Bojkov, B.: Estimating Northern Hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements, Remote Sens. Environ., 115, 3517–3529, https://doi.org/10.1016/j.rse.2011.08.014, 2011.

Tuzet, F., Dumont, M., Lafaysse, M., Picard, G., Arnaud, L., Voisin, D., Lejeune, Y., Charrois, L., Nabat, P., and Morin, S.: A multilayer physically based snowpack model simulating direct and indirect radiative impacts of light-absorbing impurities in snow, The Cryosphere, 11, 2633–2653, 2017.

van Herwijnen, A. and Jamieson, B.: High-speed photography of fractures in weak snowpack layers, 43, 71–82, 2005.

van Herwijnen, A. and Jamieson, B.: Snowpack properties associated with fracture initiation and propagation resulting in skier-triggered dry snow slab avalanches, 50, 13–22, https://doi.org/10.1016/j.coldregions.2007.02.004, 2007.

van Herwijnen, A. and Jamieson, B.: Fracture character in compression tests, Cold Reg. Sci. Technol., 47, 60–68, https://doi.org/10.1016/j.coldregions.2006.08.016, 2007.

van Herwijnen, A., Bair, E., Birkeland, K., Reuter, B., Simenhois, R., Jamieson, B., and Schweizer, J.: Measuring the mechanical properties of snow relevant for dry-snow slab avalanche release using particle tracking velocimetry, in: Proceedings ISSW 2016. International Snow
5   Science Workshop, Breckenridge CO, USA, 3-7 October 2016, 2016.

Vionnet, V., Brun, E., Morin, S., Boone, A., Martin, E., Faroux, S., Moigne, P. L., and Willemet, J.-M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, Geosci. Model. Dev., 5, 773–791, https://doi.org/10.5194/gmd-5-773-2012, 2012.

Wayand, N. E., Massmann, A., Butler, C., Keenan, E., Stimberis, J., and Lundquist, J. D.: A meteorological and snow observational data set from Snoqualmie Pass (921 m), Washington Cascades, USA, Water Resources Research, 51, 10 092–10 103, 2015.

10   Wever, N., Fierz, C., Mitterer, C., Hirashima, H., and Lehning, M.: Solving Richards Equation for snow improves snowpack meltwater runoff estimations in detailed multi-layer snowpack model, The Cryosphere, 8, 257–274, https://doi.org/10.5194/tc-8-257-2014, 2014.

Wever, N., Schmid, L., Heilig, A., Eisen, O., Fierz, C., and Lehning, M.: Verification of the multi-layer SNOWPACK model with different water transport schemes, The Cryosphere, 9, 2271–2293, 2015.

WSL Institute for Snow and Avalanche Research SLF: WFJ_MOD: Meteorological and snowpack measurements from Weissfluhjoch, WSL
15   Institute for Snow and Avalanche Research SLF, https://doi.org/10.16904/1, 2015.

Zuanon, N.: IceCube, a portable and reliable instruments for snow specific surface area measurement in the field, in: International Snow Science Workshop Grenoble-Chamonix Mont-Blance-2013 proceedings, pp. 1020–1023, 2013.

Response to Review 1

The authors present a local-scale study aimed at characterizing seasonal snowpack evolution with traditional sampling (snow pits), advanced techniques (SnowMicroPen, IceCube, and Tomography) and model application (SNOWPACK). Applying a multi- scale approach, methods are intermixed to construct a daily time series of vertical variation in snow density and specific surface area. The methods are cross-compared to contribute a recalibration of the Proksch et al. (2015) SMP empirical model and to evaluate SNOWPACK simulations. Analysis of the dataset demonstrates clearly how recent advances in field methodology can support model evaluation at very high vertical resolutions. In particular, the details found in Figures 6 and 9, where SMP derived snow properties are introduce at daily time steps, show ability to track snow events and metamorphosis captured in SNOWPACK simulations. Overall, the paper provides a great summary of the campaign results and demonstrates how future model evaluations can benefit from applying similar seasonal framework.

Prior to publication, the paper would benefit from some restructuring to clarify proper-ties of generated the dataset and promote repeatability. These would be meaningful additions to allow application of this work to other environments:

- Recalibration of the Proksch et al. (2015) model uses collocated SMP profiles and density cutter measurements. No distinction is made between the training and testing data when evaluating Eqns 1 or 2. If the authors felt cross-validation was unnecessary, please include this information so that the reader can determine if the skill estimates may be biased (i.e. Test-Train are identical datasets).

→ The entire cutter and IceCube data have been used to "train" the SMP data and to obtain Eq. (1) and (2). The scatter plots shown in Figure 2 show the quality of these parameterizations for the same dataset, i.e. SMP derived data from Eq(1) and (2) versus cutter and IceCube data. The "Train" and "Test" dataset are thus the same. We aimed here at getting as close as possible to this particular cutter and IceCube dataset from our SMP data, and we did not evaluate the obtained parameterizations with other independent dataset. This is why we wrote page 23 line 10: "We would hope that the parameterization Eq. (1) and (2) are generally applicable to an SMP version 4. However, without an independent validation by measurements under different snowpack conditions, it is not possible to state the range of validity of the parametrizations presented here."

We improved Section 5.1 so that it appears more clearly that the test and train data are the same, p10, L6: "This plot [Figue2] 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 dataset as used for the statistical modeling). 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 (same dataset as used for the statistical modeling). 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."

- I'd like to better understand why realignment resulted in improved correlation between the cutter/IceCube measurements and SMP derived properties in Figure 2 as indicated in text (P9 L23). If alignment with the persistent layer defined in Section 6 resulted in a better vertical matching, why were the better alignments not used for the initial recalibration? Throughout the paper, descriptions of alignment could be improved and are noted in the extended comments below.

→ We thank the reviewer for pointing out this issue in the paper. We agree on the confusion about the alignment. For explanation, we would like to point out the difference between 1/ matching of profiles *of the same day* for statistical analysis, and 2/ matching for visualisation of the data such as the evolution of *profile with time*.

1/ Alignment of co-located, co-temporal profiles can be done by using the snow surface. This is convenient and always applicable (unlike using a specific layer) so it is a suitable method to use when doing a local re-calibration of the SMP parameterizations as in our study.

2/ Alignment of profiles when plotting their evolution with time requires another method of matching since profiles are then not co-temporal and do not share a common height/snow surface. One way is to re-align profiles with the ground. For sites showing a ground that is uneven or bumpy this method can however lead to a mediocre alignment. This was the case of the WFJ (ground is uneven) and we found out that a re-alignment based the crust MF-layer offers a qualitatively better match, when looking at plots of Fig 6 and 9 for example. Hence, we chose this alignment method for to present data in Figure 6, 7, 9 and 10.

→ As pointed out by the Reviewer, the first version of the paper showed an inconsistency related to the choice of the alignment method in Section 5.1. Indeed, method 1 (snow surface alignment) was used to develop the statistical model but method 2 (MF-layer alignment) was used to test the performance of the model (Figure 2).

We fully agree with the reviewer that it is confusing. Thus, we modified so that method 1 (snow surface alignment) is now used for both the statistical model and the analysis of the model performance. Method 2 (layer alignment) is only used later in the paper, in the Result part, for time-series plotting purposes.

Modifications throughout the paper have been done accordingly, especially:
  - Figure 2 has been redone, based on data re-aligned using the snow surface
  - R^2 coefficients associated to Figure 2 have been modified. They are slightly better than the previous version (from layer alignment to snow surface alignment: R^2 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. 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. 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 included now: "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)."

- While the layer tracking analysis is meaningful (Fig 8 and 11), description of the SMP tracking method is difficult (if not impossible) to reproduce. An enhanced description of how transitions in SMP signal were used to define layers would be a helpful addition.
→ Section 5.2 "Layer tracking" has been restructured and some reformulation has been made to improve the description of the method. Layers in SMP data were tracked in the same way as in the cutter and IceCube data, i.e. by a manual identification of boundaries in the snow property profiles. The paragraph now reads: "In the measurements data, the layers of interest were defined by the height of their upper and lower boundaries. Boundaries were manually identified by simply looking at the property profiles, looking for sharp and relevant transitions, and recording heights. This step was performed on all the weekly density profile from the cutter and SSA profile from IceCube, as well as on all the daily representative profile of penetration force resistance obtained from the five daily SMP measurements. The identification of layer boundaries was sometimes challenging for weak stratigraphic transitions, e.g. the transition between a layer of fresh snow that fell onto a soft snow layer. To help in such cases, boundaries could be backtracked in time, starting from a profile where the layer of interest is older and its boundaries more clearly detectable. Also, additional information, such as observed height of new snow, was sometimes used to help delineate boundaries."
Besides, we would like to point out that this method only works when tracking well-pronounced layers, so might be hard to use in a systematic way over entire snowpack profiles. To stress this point, we added p10, L23: "The first step is to define which are the layers of interest, knowing that this method is only possible with layers that contrast well enough with their surrounding, so their boundaries can be identified by a significant and rather sharp transition in the vertical profile of snow properties."

- I can confirm that the revised coefficients presented for SMP density are improved over those Proksch et al. 2015 for Arctic snow and snow on sea ice. However, local calibration with our SMP4 unit resulted in quite different coefficients and better RMSE over the use of global parameters (P23 L11). This may make it important to make clear the calibration methods so that they can be easily repeated for different environments or units(?).
We improved the description of the calibration method in Section 5.1, making sure that each step is clearly described.

General comments

P2 L5 – Suggest removing the 'e.g' and revising as 'data back to 1936 in the case of WFJ'.
→ Modified accordingly

P2 L8 – Please be explicit about which properties are characterized rather than using 'hard hardness . . ..'.

→ We modified accordingly; it reads now "grain size, grain shape, hand hardness, and wetness" (P2, L8).

P2 L9 – Remove the period between the citation and sentence.
→ Modified

P2 L14 – Can you clarify what 'non-empirical snow properties' means? This statement is unclear.
With "non-empirical properties" we refer to properties that are physically/mathematically-defined, such as density and SSA, in contrast to grain shape for instance which has no mathematical definition. We modified the term and use "objectively-defined snow properties" (P2, L15).

P2 L15 –Ideally traditional measurements would be supported with metrics such as SSA but the use of the word 'tends' seems to imply this IS a frequent practice. Could it rephrased with the word 'can' or similar?
→ Modified accordingly. The sentence reads now "Concerning the characterization of snow microstructure, the observer-biased estimate of traditional grain size can be replaced by measurements of specific surface area" (P2, L15).

P2 L19 – Capitalize 'IRIS'. Stands for 'InfraRed Integrating Sphere'.
→ Modified accordingly.

P3 L16 – Should the word 'such' be in this sentence?
→ We modified the sentence as "These examples exploit key advantages of the SMP, namely fast profiling for frequent measurements and high vertical resolution, so that profiles are obtained at a considerably finer scale (mm) than possible with traditional means." (P3, L17).

P3 L21 - It feels a bit discouraging to say that the stated goals are dependent on availability of a large dataset with many tools. As a suggestion, removing the word 'only' might lessen the tone. The wording 'cross-validation' could also be problematic as it refers to a specific statistics method. Later the wording 'cross-comparison' (P4 L8) is used which seems to be a better fit.
→ We agree with the reviewer and modified the sentence accordingly as "In the context raised above, the value of emergent, objective snow properties, their potential to replace traditional means in operational snow monitoring programs, and their requirements on temporal and vertical resolutions for model evaluations can be investigated within a multi-resolution and multi-instrument dataset to facilitate comprehensive cross-comparison analyses."

P4 L12 – Degree symbols should accompany the coordinate units.
→ Modified accordingly

P4 L14 – Consider revising the sentence to mention dry snow conditions only once.
→ Modified as follows "We focused on the period from beginning of December 2015 to end of March 2016 to ensure measurements in dry snow condition as required by some of the used instruments. (P4, L17)

P4 L23 – The second element of the measurement area description is squared. Was this intended?
→ Modified as follows "20 m x 8 m".

P7L7 – If the Zuanon (2013) methods were adopted, were any samples compressed to avoid over penetration of the laser? A sentence on how samples were extracted and prepared would be useful for future comparisons where this has become common practice.
→ The extraction of the sample was performed following the protocol described in Zuanon et al. 2013. In addition, we indeed systematically slightly compressed the extracted sample. We included this information in the paper: p7, L10: "Snow samples were very slightly compressed when inserted into the sample holder and attention was paid to have a flat snow sample surface."

P7 L8 – What about uncertainty with low SSA (i.e. DH or FC)? Standard deviation of the measurements in Figure 10a appears to increase with depth and is quite large relative to tomography.
→ As pointed out in the paper Section 7.3, we report a significant and systematic inter-measurement deviation in the SSA estimates. Although we did not study in details uncertainty of SSA measurements in weak layers, our results do not show that biases are more pronounced for DH or FC layers. We did not observe an evolution of the bias with depth. The paper however stresses that these inter-measurement deviations should be further investigated.

P7 L17 – Would like to see an enhanced description of what goes into the profile quality check. Previous studies have described linear trends while measuring in air while others have provided quantitative methods to apply a noise threshold. Which approach was used to determine drift or accept/reject a profile?
→ We improved the description of the SMP data processing. The paragraph now reads P7, L21: "The quality control of SMP force profiles was done manually by rejecting signals with 1) visible trends either in the air portion of the signal or over the entire depth, 2) high noise levels and unrealistic spikes, and 3) frozen tip problems revealed by a force response that appears to be activated only deeper in the snowpack. Most of these problems are caused by wet conditions. The air-snow and snow-ground interface were detected manually to remove air and ground regions from the signal."

P7 L20 – What were the qualities of the data, snow, or study site that determined the profiles could be matched without an offset correction? In section 6 the opposite seems to be stated that spatial variability required compensation to avoid height mismatches (P11 L13).
→ This seems to be a misunderstanding. We improved the description of the SMP data processing in Section 3.4. By offset correction we mean that the value of the force signal itself was not shifted by a given value as it can be sometimes observed (see previous comment). The force signal in the air was very close to zero (manual check) so we did not correct the force signal. This has no link with the height alignment performed in Section 6 for data visualisation.

P7 L29 – Suggest removing 'Reconstruction followed standard procedure' as it's described in the next sentence.

→ Modified accordingly

P8 L10 – May be helpful to indicate the rate of replacement.
→ During the period shown in this study (no melt out), only missing values of either incoming or outgoing SW or albedo values above 0.95 require a replacement. There are no missing values and the latter amount to at most 0.8%, predominantly at sunrise and sunset.

P9 L7 to 11 – Found this a bit of confusing. Is the single 'median' profile being used to train (1)? Perhaps the alignment sentence could be moved upwards in the paragraph to clarify. As it reads now I was not able to determine if 1 profile per pit is being used or if multiple A-S aligned and cropped profiles are being used.
→ We agree with the reviewer and modified the paragraph to describe more clearly each step of the process. It reads now, P9, L16: "The statistical modeling was applied based on a sub-dataset of data from the days for which both SMP and snow pit measurements were available (15 days for density, 13 days for SSA). From each raw force signals, parameters F and L were computed from the raw penetration force profiles over a sliding window of 1 mm with 50% overlap, yielding profiles of F and L with a vertical resolution of 0.5 mm. Note that Proksch et al. (2015) used a sliding window of 2.5 mm, but tests with different window heights (1, 2.5 and 5 mm) did not show a significant impact. Next, for each day, the five daily profiles of F and L of the same day were aligned by simply using snow surface as common reference and a median operation was applied to get one representative profile of F and L per day, called the median profiles in the following. Next, each median profile was averaged vertically using a 3 cm window to match the vertical resolution of the snow pit measurements. Finally, the median 3cm-averaged profiles F and L and the profiles of rho_cutter and SSA_ic of the same day were aligned by using snow surface again as common reference and cropped to the length of the shortest profile. This way, all profiles of a given day are described on the same vertical scale and values of F, L, rho_cutter and SSA_ic can be paired for the statistical modeling, relying on a total of 590 paired-values for density and 497 for SSA."

P9 L15 – Please provide the number of compared measurements to support of the significance test.
→ The number of compared measurements was 590 for density and 497 for SSA. We included that in the manuscript (see comment above).

P9 L16 – This differs substantially from Proksch et al (2015) where coefficients for SSA were not provided. This new equation requires no estimate of density from the SMP, which arguably is better if SSA is the target (minimizes bias from density coefficients and conversion from d0?). No action to take unless the authors wish to highlight the benefit of avoiding the conversion of L_ex to SSA.
→ We would agree with the reviewer that directly estimating SSA and not correlation length via the density as in Proksch et al. 2015, should lead to a better estimates (less errors). In the paper we simply pointed out this difference in the method by writing "Differing slightly from the one suggested by Proksch et al 2015, a regression of the from [Eq 2] was applied to estimate SSA ...".

P9 L23 – An enhanced explanation of why the values in Figure 2 do not reflect the error/skill assessment in this section is needed. Related questions: Why does correlation improve when Eqn. (1) was trained on a different set of comparisons? Why was Eqn (1) was not just trained on this better alignment to begin with?
→ As written in an above comment on the same issue, we agree with the Reviewer and modified Section 5.1. In the revised version, values in Figure 2 (and the associated correlation analysis) are based on the same set of data and same re-alignment with the snow surface than the values taken for the statistical modelling Eq 1 and 2. Besides, our statement that using the MF-layer alignment leads to better correlation of values in Fig 2 was a wrong statement. Slightly better R2 coefficients are indeed 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. This has been corrected in the revision and Section 5.1 is now consistent.

P8L29 – Remove one set of brackets around the Eqn.
→ done

P10 L2 – What was the statistical test that showed the boundary transition to be significant? If untested, consider removing the word 'significant'. See comments in the initiate statement about repeatability as well.
→ Boundaries were detected manually just from looking at the data, so there was no statistical test to identify them as well as to confirm that they are "significant". We deleted the work "significant" and it reads now, P10, L23: "The first step is to define which are the layers of interest, knowing that this method is only possible with layers that contrast well enough with their surrounding, so their boundaries can be easily identified by a rather sharp transition in the vertical profile of snow properties." We modified substantially Section 5.2 "Layer tracking", as described in a related comment above, so the method is better described now and can be repeated.

P10 L3 – Given that the boundaries were identified subjectively, will their heights be provided in the published dataset?
→ Heights of the tracked layers will be provided in the database of this study.

P15 L3 – I agree that the information is really useful to show the formation and evolution of these fine features. However, given that Figure 6b has no minor or major ticks for the initial date (Feb 22) it's fairly difficult to identity the feature. Could a label be provided for easy reference?
→ We prefer to leave the figures as is to avoid an emphasis on a single, annotated feature. Since the location is given exactly in the text and the x-axes of the subfigures are exactly the same, the birth of this layer could be easily taken from the SMP image above.

P23 L11 – I can confirm that the recalibrated density coefficients don't produce a best-possible estimates of snow density with our SMP for Arctic snow. Would be very interesting to combine datasets from multiple units to evaluate this uncertainty.
→ We agree that it would be very interesting to compare different sites to test the re-calibrations presented here.

P25L20 –Citation style should be a paraphrase.

Table 1 - List the number of measurements as a separate column. The large number of measurements is really smoothing to highlight! This will also be helpful in the future to frame comparison.
→ The number of measurements has been included in Table 1 (SMP: 100, Cutter: 15 profiles, IceCube: 13 profiles, Traditional: 11 profiles, Stability tests: 8 tests).

Figure 2 - Add N, R^2 and RMSE be added to these diagrams. Having a quantitative evaluation in the diagram provides a quick reference for the reader.
→ Done

Figure 4 – Please provide a colour legend for the grain type classifications even though they are standardized. Additionally, is it possible to provide sub-hatching for the hand hardness levels? It's challenging to determine the level past the first data.
→ Figure 4 has been modified accordingly.

Figure 6/9 – Has the SMP data been smoothed or aggregated? This does not appear to be mentioned in text but Figure 10 shows variability in SSA absent in Figure 9 at the 1 mm scale.
→ We used the same data with a resolution of 0.5 mm for the seasonal evolution plots as well as for the vertical profile plots (7 and 10).

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.

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% ucertainty 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.