Reviewer 1

We would like to thank the reviewer for his detailed and helpful comments and remarks, of which some made us rethink parts of the analysis. We point out our detailed response to the issues raised by the reviewer below.

Major comments

In addition to comparisons discussed in this manuscript, suggestions of studies to reduce discrepancies should be discussed for future reference. It will be informative to decide what observation and experiment is necessary to improve SNOWPACK model.

This is an interesting suggestion and we will amend the manuscript with an outlook section. We think that the recommendations for future research to improve the model, that seem most relevant in the context of this manuscript, can be separated into two parts: (i) the liquid water flow and (ii) melt behaviour in spring.
(i) Currently, simulating liquid water flow only considers a 1-dimensional component, assuming homogeneity in the horizontal dimensions. This is, however, a very strong simplification, as in reality, liquid water flow exhibits strong variation in 3 dimensions, due to preferential flow paths or flow fingering. Numerical experiments (Hirashima et al., 2014) and laboratory observations (Katsushima et al., 2013) have provided indications that these processes can be described using Richards equation in 3 dimensions. At the same time, several processes that do appear in 1-dimensional simulations, as for example the ponding of liquid water on capillary barriers, seem to be essential in forming preferential flow paths. This possibly allows for a parametrisation of preferential flow in the SNOWPACK model, that is closely linked to physical processes. Validation could be achieved by more detailed snow lysimeter studies, for example from measurement sites with multiple neighbouring lysimeters, improved laboratory experiments or further exploiting the upGPR data.
(ii) At the measurement site WFJ, we found a consistent overestimation of melt rates in spring, which is indicated by an underestimation of SWE compared to the manual snow profiles. The difficulty is that the SWE depletion in spring is dependent on many factors, such as snow density and wet snow settling, influencing snow heat capacity, internal heat fluxes and the penetration of short wave radiation, as well as the surface energy balance and liquid water flow. These processes are difficult to investigate separately. For the surface energy balance, ideally, repeated cold content measurements could be performed. This could be done using the calorimetric method: melting the snowpack and determining how much energy is required. However, these measurements are rather cumbersome to perform in the field. We are currently analysing measurements of turbulent fluxes, which so far have revealed that the constant flux layer assumption is often violated and it is hoped that an improved turbulent fluxes calculation scheme can be developed (Schlögl et al., 2015). Snow compaction (settling) in spring could be assessed with in-situ snow harps or snow profiles at a higher temporal resolution than only biweekly. However, recent advances in snow micro penetrometer (SMP) could also be helpful, allowing to achieve density measurements at high temporal and high spatial resolution with relatively little effort (Proksch et al., 2015). A drawback of that method is that SMP measurements are unfortunately not suitable for wet snow conditions. Finally, a more detailed analysis of upGPR data and snow lysimeter measurements can help to highlight discrepancies in the modelling of liquid water flow. Heilig et al. (2015) show an example of how bulk liquid water content measurements from upGPR in combination with snow lysimeter measurements could improve snowpack models.

However, in terms of wet snow area, residual water content seems to affect more rather than calculation scheme. According to author's previous paper (Wever et al., 2014), residual water content was not constant and not larger than 0.02. I think values or temporal variation of residual water content of bucket scheme and RE scheme should be shown in this paper. If the difference of θ_r between two schemes is large, discussions about suitable value of θ_r is also necessary.

The reviewer is using the term residual water content for both the bucket scheme and Richards equation. It is true that both water transport schemes have a parameter that may seem related to each other, but they are different. We therefore consistently used the terms water holding capacity or irreducible water content for the bucket scheme and residual water content for the Richards equation. The residual water content is a hypothetical value, principally not reached by water transport alone as it is associated with an infinitely small pressure head, but only due to phase changes. On the other hand, the water holding capacity in the bucket scheme refers to the typical liquid water content reached in snow, and refers to the size of the buckets. In RE, the actual liquid water content in the simulation is near or above the residual water content, whereas in the bucket scheme, the actual liquid water content is always at or below the water holding capacity. So the two values are not comparable. We will provide the following short note on this when revising the manuscript:

> Note that the residual water content in the water retention curve, which is the dry limit, is not comparable to the water holding capacity or irreducible water content in the bucket scheme, which refers to wet conditions.

Specific comments

P8 In section 2.2, improvement of a calculation scheme for soil was discussed. It is one of the updated contents of the model in this paper. However, the effect of this improvement seems to be not shown in this paper. Is there large differences at the snow-soil boundary between before and after improvement of soil scheme? Probably it will be verified by comparison between simulated and observed soil water content profiles. It may be future works.

We did not compare the new soil module to the old one in the manuscript, as we think that this comparison is not important for the outcome of the manuscript. Basically, we did not intend to claim that the new soil module is better regarding the simulation of the snow-soil interface temperature, but only wanted to show that the approach is also providing a correct lower boundary condition for the snow cover. Both the old and the new model produce very similar results. Furthermore, the original soil module has been applied often in permafrost and rocky terrain, whereas we are interested in soils, as the upper part of the ground at Weissfluhjoch is soil rather than rock. We are indeed planning a comparison of soil moisture measurements, but as those measurements have not been carried out at Weissfluhjoch for logistical reasons, we consider it to be out of the scope of the manuscript.

P16 L20 Table1 showed average values for more than 10 years. I think it varies from year to year. Therefore, information of fluctuation from year to year is also necessary in Table 1. For example, standard deviation of annual average is calculated and added in Table 1 as 'average ($\pm$SD)'

This is an interesting suggestion, so we modified the table accordingly (see Table 1 in this document).

P19 L13 True, snowpack runoff is strongly coupled to the LWC distribution, but good agreement of runoff does not mean good agreement of water content in the snowpack. Do you have any water content profile data obtained in snow pit observation? Direct comparison of water content is important even if it is discontinuous and destructive.

We agree with the importance of verifying vertical liquid water content distributions inside the snowpack. For Weissfluhjoch, the only dataset available for the full studied period is the wetness reported by the observers from the biweekly manual snow pits. However, a few issues can be identified with this data:

- *Wetness is reported in only a few classes, that span a wide range of liquid water content. For example, following the international classification (see Fierz et al. (2009)), wetness class 3 spans 3-8% LWC.*
- *Judging wetness of the snowpack is generally difficult and has a subjective component. From our own experience (not published), we did see discrepancies between snow wetness reported by observers, and measurements by Denoth, SnowFork or upGPR.*
- *Manual snow profiling is generally done in the morning hours, thus before the onset of snow melt. Generally, no repeat snow profiles are made during the day to follow the changes in LWC distribution during the day. In contrast, bulk LWC derived from upGPR is able to follow diurnal cycles and a comparison with SNOWPACK simulations have been made (see Heilig et al. (2015)), showing a relatively good correspondence between simulated and observed bulk LWC. However, it is not (yet) possible to derive the vertical spatial distribution inside the snowpack from the radar signal.*

In Figure 1, an example is shown of the LWC distribution as simulated, together with the LWC reported by the observers. Although observers may report 5 wetness classes, we decided to only show data in three classes (0% LWC (dry), 0-3% LWC (mois) and ≥3% LWC (wet)) because of the aforementioned reasons. We will add this data in all the figures in the manuscript and supplement when revising.

P20 L8 NSE coefficient of snow-height driven simulation is better. I agree accurate percolation time is one of the reasons of it. In addition to this, does the difference of date of snow disappearance affect NSE? In many years, snow disappeared faster in precipitation driven simulation than that in snow-height driven simulation. In NSE estimation, did you consider the period after snow disappearance in precipitation driven simulation? Also, Table 1 had better include difference of date of snow disappearance.

This is an interesting question. In order to test the sensitivity of the NSE coefficients to the period chosen, we calculated the coefficients in two ways: either taking the melt-out date from the measured snow height or from the simulated snow height (both defined as less than 5 cm of snow remaining). The latter approach was used to calculate the values in the table. On the average NSE coefficients for the studied period, this has not a large impact (typically influencing average NSE coefficients by less than 0.01). However, in individual years, differences may be larger (up to 0.16), in particular for precipitation driven simulations. For this simulation type, the melt out dates are not as well predicted as for the snow height driven type. Nevertheless, the differences between simulation setups within either snow height driven simulations or precipitation driven ones are smaller than the differences between both simulation types. This implies that the same conclusions can be drawn, regardless of the choice of calculation period.

The influence on r^2 values is larger, due to the fact that the last days of melt out are often associated with large snowpack runoff. We will introduce a section in the manuscript discussing the effects of the choice of calculation period. Following the reviewers' suggestion, we added the difference in melt out date in the table, which will help interpreting these results (see Table 1 in *this document). Note that this table contains slightly different values for NSE and r^2 compared to the original manuscript, as there were a few discrepancies how it was determined which period should be analysed. They were revealed when performing the analysis as recommended by the reviewer. Furthermore, we have to apologize for reporting correlation coefficients (r) instead of the coefficients of determination (r^2) for cold contents, isothermal part and avg. grain size, although the table suggested otherwise. This is also corrected now.*

P20 L16 In Fig6a, different Y-axes were used for different depth of sensor of soil temperature. Before I aware the difference of y-axis, soil temperature seemed to be stable around 3 and 6 degree Celsius at 30cm and 50 cm in depth, respectively. To avoid misunderstand, caption is necessary at the right side, and scales of right side should show 0 at the zero-point.

We changed the figure as suggested by the reviewer (see Figure 2 in this document). We are sorry to have caused confusion here.

P22L3 Is Figure 8a misdescription of Figure8? Also, It needs caption on right side and showing 0 at the zero-point as well as previous comment for Fig. 6.

Figure 8a should indeed be Figure 8. We changed this figure also as suggested by the reviewer, which entail similar changes as shown in Figure 2 in this document. We are sorry to have caused confusion here.

P24 L1 Fig. 11 showed comparison of measured and simulated snow density separating upper, middle and lower part. I think it is suitable using relative height (1 at the top and 0 at the bottom) like relative date used in Fig. 2, and compare at the specific relative heights (e.g. 0.1, 0.5 and 0.9).

We changed the figure as proposed by the reviewer, showing the snow density in the lower part (0-25% of snow height), the middle part (37.5%-62.5% of snow height) and the upper part (75%-100% of snow height) of the snowpack. See Figure 3 in this document. We did not scale the time axis between 0 and 1 for the beginning and end of the snow season respectively, as snow density is not a continuous measurement, but only a biweekly one. That means that there are typically 12-16 snow profiles per snow season, and scaling these between 0 and 1 is in our opinion not improving the figure as the time resolution is too low. As we do not think this figure is conveying a clearer message than the one in the manuscript, we do not plan to make a change here.

P25 L7 According to fig. 13a, simulated increase of average grain size during melt season seems to be smaller than that in observation. Also, according to fig. 13b, simulated SD was smaller than observed SD. Although display of average and SD express overall trend, it is not easy to find the reason of discrepancies. Can you add the example of direct comparison of grain size profile in supplement figures?

Similar to liquid water content, we will add the biweekly profile data for grain size now in the figures, as shown, by way of example, in Figure 4 in this document.

P21 L11 and P27 L26 Isothermal part of snow temperature relates the wet snow area. In terms of isothermal part of temperature, simulation result of RE scheme corresponded better with observations than bucket scheme. On the other hand, comparing with the upGPR measurement, bucket scheme corresponded better with observation. Can you explain the reason of this contradict results?

The contradictory result is in our opinion a consequence of the short period of upGPR data (only 4 snow seasons), compared to the full period of 15 years used for determining the statistics. Furthermore, the statistics are determined for all snow profiles, also those made in the beginning of the snow season (October and November), where regularly snow melt is occurring. See for example Figures S3a,b,g,h and S4i,j, and S5c,d,g,h in the Supplement. As this comment points us to the importance of having grain type information in the manuscript, we will include the grain types from the simulations and the observed profiles in the online Supplement for completeness. We originally did not plan to extensively discuss grain type evolution by the model, as the manuscript is already quite long. Furthermore, grain types from the SNOWPACK model can be regarded as a post-processing of the microstructural parameters in the model, whereas the other variables discussed in the manuscript are explicitly evaluated in the simulations.

P29 L19 In conclusion, you wrote "updated soil module can provide a correct lower boundary for snowpack in the model". Is it written in the main text? I could not find the discussion of difference of reproducibility at the boundary with calculation schemes in section 4.3 and Figure 6b.

It is true that it was not explicitly stated in section 4.3. We will amend the sentence on p. 2674, L27 as:

> Figure 6b shows that the simulations capture the variability in early season soil-snow interface temperature to a high degree in most years and that the soil module in SNOWPACK is providing an accurate lower boundary for the snow cover in simulations.

References

Fierz, C., R. Armstrong, Y. Durand, P. Etchevers, E. Greene, D. McClung, K. Nishimura, P. Satyawali, and S. Sokratov (2009), The International Classification for Seasonal Snow on the Ground (ICSSG), *Tech. rep.*, IHP-VII Technical Documents in Hydrology No. 83, IACS Contribution No. 1, UNESCO-IHP, Paris.

Heilig, A., C. Mitterer, L. Schmid, N. Wever, J. Schweizer, H.-P. Marshall, and O. Eisen (2015), Seasonal and diurnal cycles of liquid water in snow - Measurements and modeling, *J. Geophys. Res.*, doi:10.1002/2015JF003593, accepted.

Hirashima, H., S. Yamaguchi, and T. Katsushima (2014), A multi-dimensional water transport model to reproduce preferential flow in the snowpack, *Cold Reg. Sci. Technol.*, *108*, 80–90, doi:10.1016/j.coldregions.2014.09.004.

Katsushima, T., S. Yamaguchi, T. Kumakura, and A. Sato (2013), Experimental analysis of preferential flow in dry snowpack, *Cold Reg. Sci. Technol.*, *85*, 206–216, doi: 10.1016/j.coldregions.2012.09.012.

Proksch, M., N. Rutter, C. Fierz, and M. Schneebeli (2015), Intercomparison of snow density measurements: bias, precision and spatial resolution, *Cryosphere Discuss.*, *9*(4), 3581–3616, doi:10.5194/tcd-9-3581-2015.

Schlögl, S., R. Mott, and M. Lehning (2015), Assessment of turbulent flux parametrizations over snow using different stability corrections, in *IUGG General Assembly*.

Yamaguchi, S., T. Katsushima, A. Sato, and T. Kumakura (2010), Water retention curve of snow with different grain sizes, *Cold Reg. Sci. Technol.*, *64*(2), 87 – 93, doi: 10.1016/j.coldregions.2010.05.008.

Yamaguchi, S., K. Watanabe, T. Katsushima, A. Sato, and T. Kumakura (2012), Dependence of the water retention curve of snow on snow characteristics, *Ann. Glaciol.*, *53*(61), 6–12, doi: 10.3189/2012AoG61A001.

Table 1: Average and standard deviation (in brackets) of bulk snowpack statistics over all snow seasons for various simulation setups (bucket or Richards equation (RE) water transport scheme, snow height (HS) or precipitation (Precip) driven simulations, Y2010 (*Yamaguchi et al.*, 2010) or Y2012 (*Yamaguchi et al.*, 2012) water retention curves, and arithmetic or geometric mean for hydraulic conductivity) for all simulated snow seasons. Differences are calculated as modelled value minus measured value, ratios are calculated as modelled value divided by measured value. The isothermal part is only considered during the melt phase (from March to the end of the snow season).

Variable	Bucket	RE-Y2010AM	RE-Y2012AM	RE-Y2012GM	Bucket	RE-Y2012AM
	HS driven (2000-2014)				Precip driven (1997-2014)	
RMSE HS (cm)	4.16 (1.73)	4.00 (1.56)	4.11 (1.64)	4.12 (1.71)	20.86 (12.31)	23.12 (11.38)
Difference HS (cm)	1.33 (2.24)	0.87 (2.09)	0.88 (2.17)	0.89 (2.21)	-1.23 (12.31)	-5.24 (11.38)
Difference melt out (days)	-0.67 (1.45)	-0.73 (1.44)	-0.73 (1.44)	-0.73 (1.44)	-3.94 (6.08)	-7.00 (6.83)
RMSE SWE (mm w.e.)	39.28 (15.51)	39.62 (14.71)	39.78 (15.50)	39.39 (15.45)	84.96 (36.34)	99.03 (36.23)
Difference SWE (mm w.e.)	-5.67 (27.20)	-7.08 (27.04)	-9.29 (27.05)	-8.06 (27.14)	-16.14 (67.61)	-36.00 (66.91)
Ratio SWE (mm w.e.)	1.01 (0.09)	0.99 (0.08)	0.99 (0.08)	0.99 (0.08)	0.97 (0.19)	0.91 (0.17)
Ratio runoff sum (-)	1.08 (0.28)	1.14 (0.28)	1.13 (0.28)	1.13 (0.28)	0.98 (0.31)	0.98 (0.31)
NSE 24 hours (-)	0.72 (0.32)	0.73 (0.32)	0.73 (0.32)	0.73 (0.32)	0.66 (0.32)	0.67 (0.31)
NSE 1 hour (-)	0.13 (0.37)	0.57 (0.35)	0.59 (0.34)	0.58 (0.34)	0.02 (0.39)	0.39 (0.34)
r^2 24 hrs runoff sum (-)	0.85 (0.11)	0.87 (0.10)	0.87 (0.10)	0.87 (0.10)	0.84 (0.12)	0.85 (0.13)
r^2 1 hour runoff sum (-)	0.52 (0.06)	0.78 (0.08)	0.78 (0.08)	0.78 (0.08)	0.48 (0.07)	0.68 (0.11)
Lag correlation for runoff (h)	-1.47 (0.79)	-0.20 (0.37)	-0.17 (0.31)	-0.13 (0.30)	-1.72 (0.79)	-0.44 (0.48)
RMSE cold contents (kJ m^{-2})	627 (274)	529 (244)	554 (285)	551 (277)	786 (556)	742 (509)
Difference cold contents (kJ m^{-2})	-129.0 (312.9)	11.1 (326.2)	-30.5 (336.2)	-36.7 (322.9)	-46.0 (604.0)	62.4 (565.0)
r^2 cold contents (-)	0.76 (0.36)	0.78 (0.36)	0.79 (0.36)	0.78 (0.36)	0.77 (0.36)	0.78 (0.36)
r^2 isothermal part (-)	0.64 (0.33)	0.74 (0.36)	0.74 (0.36)	0.73 (0.35)	0.65 (0.32)	0.74 (0.36)
r^2 avg. grain size (-)	0.47 (0.31)	0.45 (0.30)	0.45 (0.30)	0.45 (0.30)	0.39 (0.29)	0.37 (0.28)
Mass balance error (mm w.e.)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.09 (0.25)	0.02 (0.03)
Energy balance error (W m^{-2})	0.03 (0.08)	0.06 (0.08)	0.06 (0.08)	0.05 (0.08)	-0.05 (0.07)	0.05 (0.08)
CPU time (min)	0.57 (0.07)	1.39 (0.26)	1.44 (0.36)	1.45 (0.37)	0.61 (0.11)	1.55 (0.45)

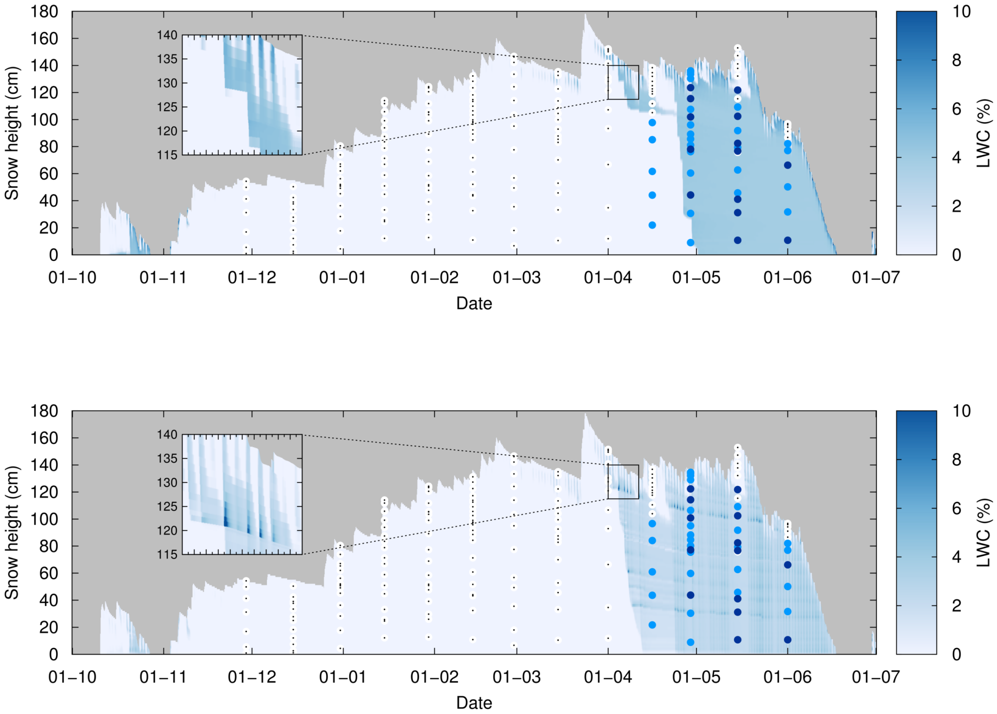

Figure 1: Snow LWC (%) for the snow height-driven simulation with the bucket scheme **(a)** and with Richards equation using the *Yamaguchi et al.* (2012) water retention curve and arithmetic mean for hydraulic conductivity (RE-Y2012AM, **(b)**), for the example snow season 2014. Dots denote layers that have been reported as dry (0% LWC, white with black center dot), moist (0-3% LWC, light blue) or wet, very wet or soaked (≥3% LWC, dark blue) from the biweekly snow profiles. When layers are reported as "1-2" (dry-moist), it is considered moist. In the zoom insert, major and minor x-axis ticks denote midnight and noon, respectively.

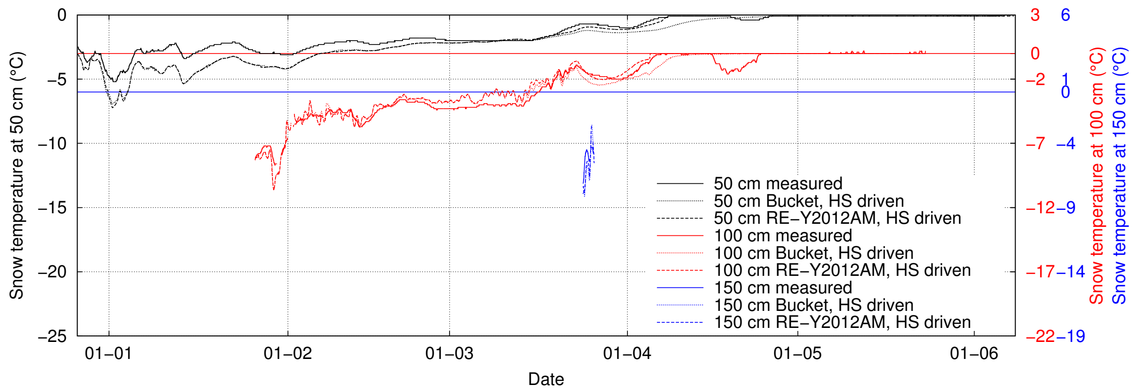

Figure 2: Measured and modelled snow temperatures at 50, 100 and 150 cm above the ground for snow height-driven (HS driven) simulations using the bucket scheme or Richards equation using the *Yamaguchi et al.* (2012) water retention curve and arithmetic mean for hydraulic conductivity (RE-Y2012AM) for the example snow season 2014. Values are only plotted when the snow height was at least 20 cm more than the height of the temperature sensor. Note that the x-axes for 100 and 150 cm depth are staggered by 3 °C to prevent overlap.

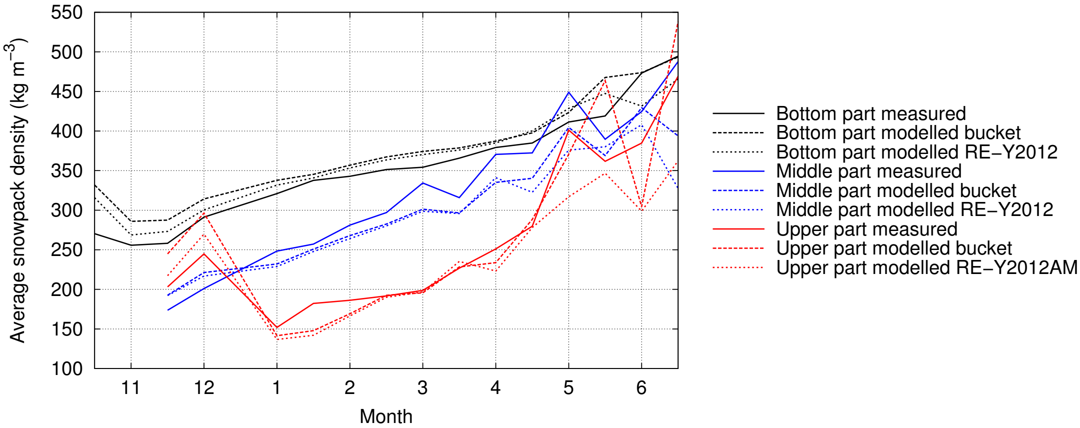

Figure 3: Average simulated and measured snow density (kg m^{-3}) for the relative lower (0 – 25 %), middle (37.5 – 52.5 %) and upper part (75 – 100 %) of the snowpack height.

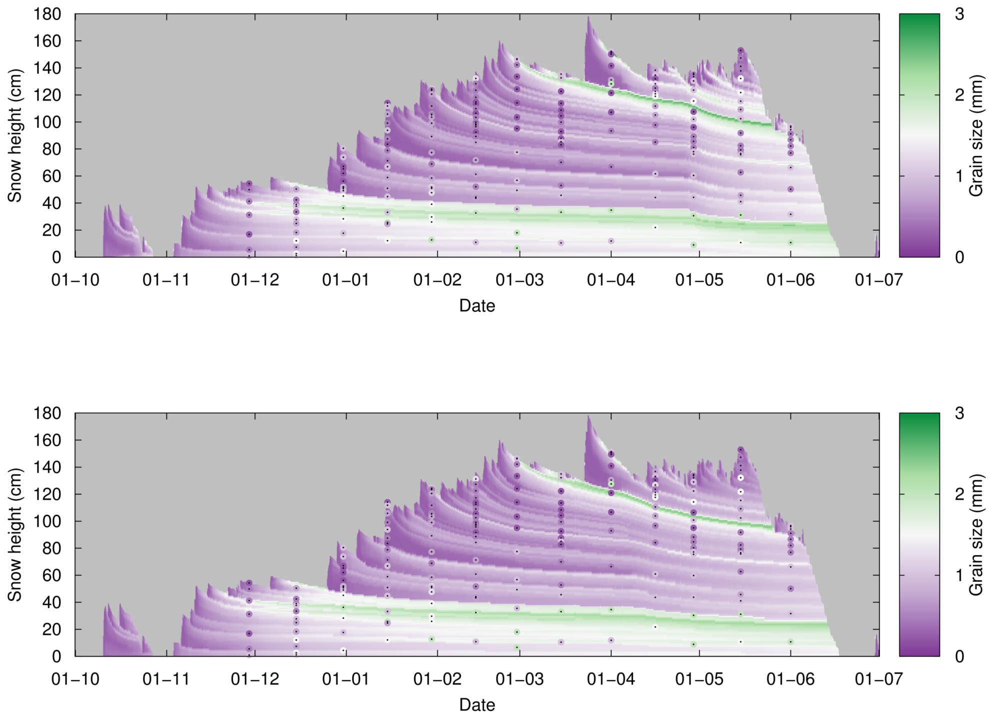

Figure 4: Grain size (mm) for the snow height-driven simulation with the bucket scheme **(a)** and with Richards equation using the *Yamaguchi et al.* (2012) water retention curve and arithmetic mean for hydraulic conductivity (RE-Y2012AM, **(b)**), for the example snow season 2014. Dots with a black center point indicate observed grain sizes reported from the biweekly snow profiles, where the black center point is located in the middle of the observed layer.

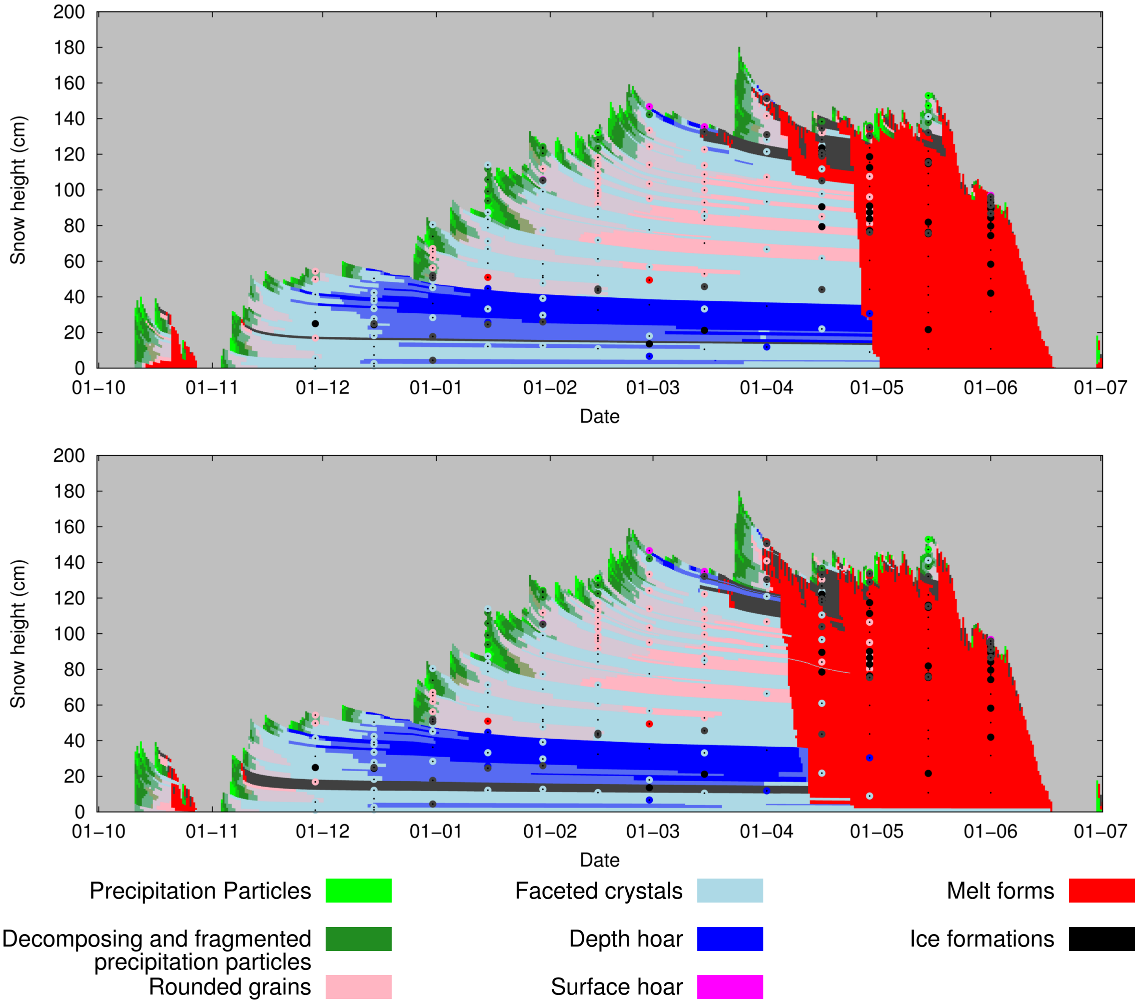

Figure 5: Grain type for the snow height-driven simulation with the bucket scheme **(a)** and with Richards equation using the *Yamaguchi et al.* (2012) water retention curve and arithmetic mean for hydraulic conductivity (RE-Y2012AM, **(b)**), for the example snow season 2014. Dots with a black center point indicate observed grain types reported from the biweekly snow profiles, where the black center point is located in the middle of the observed layer.