Response to Reviewer 2

The authors have chosen to address a challenging topic and the results of this effort reflect the difficulty of simulating a heterogeneous, three dimensional, multi-process phenomenon in a 1-dimensional framework. While there is certainly much progress and future work to come in this area, the authors have presented a valuable analysis and initial framework that can be adapted and built upon in the future. As such, I am recommending that this work be accepted for publication in The Cryosphere, following some minor revisions that I believe will improve the quality and usefulness of the work.
We thank the reviewer for his constructive comments on the manuscript. Please find our detailed response to the issues raised by the reviewer below.

Major comment

Much of the framework relies on the interpretation of the results of Katsushima et al. (2013) as interpreted by Hirashima et al. (2014). They found that when water had reached the bottom of the profiles in their laboratory experiments, that the fractional area at 4 cm depth was smaller with larger snow grain size. The wetted fraction was interpreted here as the preferential flow fraction. As grain size decreased the wetted fraction increased and this was interpreted as the preferential flow fraction being larger. However, as this extends to the finest grain size, it is stated that no preferential flow was observed, and indeed if slower matric flow was happening, the deepest wetted area would be larger and more uniform. If that is what happened, and matric flow had extended beyond the 4 cm depth, then we have no clear transition defined between the two flow regimes. I would like the authors to discuss the interpretation of the experiment on which Figure 2 is based.
We partly agree with the interpretation of the experiments by the reviewer. We think that at some point, it may become not possible anymore to clearly define matrix flow and preferential flow regimes. If one observes that 90% of the snowpack is wetted during infiltration, could you still call it preferential flow and flow fingering, or is it just matrix flow where 10% of the snowpack is somehow not involved in matrix flow? However, we should also consider the possibility that the observations in the studies by Katsushima et al. (2013) and Hirashima et al. (2014) are not sufficient to conclude that both regimes cannot be clearly defined. It may happen that the finger width exceeds the size of the container used in the experiments such that the preferential flow finger is not identifiable as such anymore DiCarlo (2013). The diameter of the rings used in the experiments by Katsushima et al. (2013) was 5 cm and larger rings may be required to study water flow for small grains. On the other hand, in soil science, experiments also have demonstrated that for certain flow regimes and small grain size, no preferential flow forms and the flow is considered stable DiCarlo (2013). This was also found for larger grain sizes at low infiltration rates. We think that at this point, it is uncertain whether for fine snow grains, a regime exists were no preferential flow forms, or that preferential flow in snow forms under any circumstance for infilitration in snow. We hope that future experimental studies will address these issues. We did not explicitly mention in the manuscript the absence of preferential flow for fine grains in the observations by Katsushima et al. (2013) when discussing Fig. 2. We will revise the manuscript at this point. However, more importantly, the smallest grain size class used in the experiments by Katsushima et al. (2013) is not often found in the SNOWPACK simulations. This grain size class is mostly associated with new snow fall and subsequent metamorphism generally quickly increases the grain size. We plan now to amend the manuscript with an extra figure (see Fig. 1 in this document). In subfigure (e), the grain radius from the simulations is shown. Here, we coloured the grain radius black when it is below 0.16 mm, which corresponds to the lowest grain size class used by Katsushima et al. (2013). As can be seen, the important parts of the snowpack involved in preferential flow have a grain radius larger than the specific grain size class for which no preferential flow was observed.

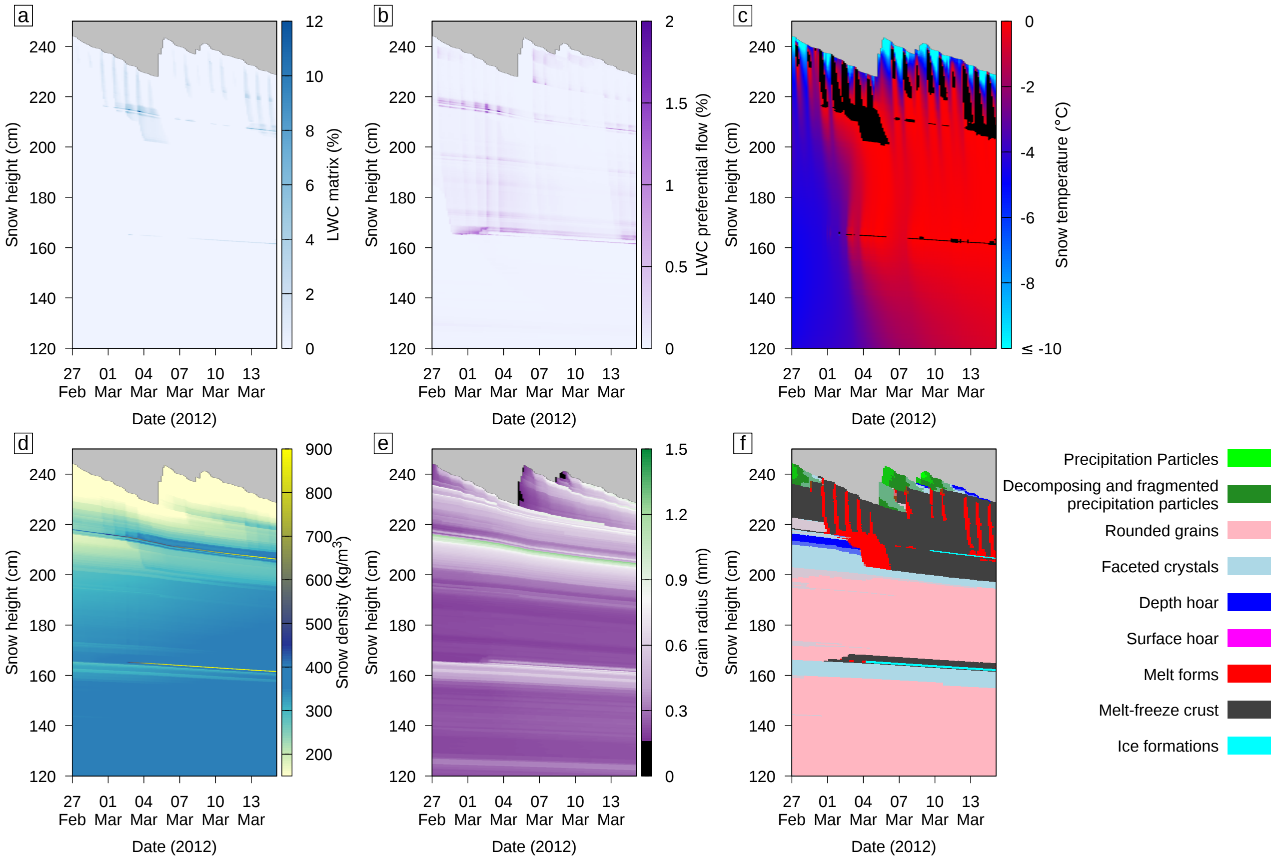

Figure 1: LWC in matrix flow (a), LWC in preferential flow (b), snow temperature (c), snow density (d), grain radius (e) and grain shape (f), depicting a detail of Fig. 5 in the original manuscript. Only the upper part of the snowpack is shown for the period 27 February to 15 March. In (c), snow at melting temperature is coloured black to highlight wet parts, in (e), grain radii smaller than 0.16 mm are coloured black, denoting the smallest grain size class used in *Katsushima et al.* (2013), and in (f), ice formations are defined as modelled dry snow density exceeding 700 kg/m^3.

Specific Comments:

- P2 line 15: Although I would not expect a through presentation of the work of Colbeck (1979) or Marsh and Woo (1985), the authors have stated that these works were not widely adopted. A couple of sentences summarizing the main concepts presented in these studies, and any weaknesses that may have resulted in their lack of adoption, would help to inform the readers about the need for progress in this area of research.
  We will revise the text as follows:
  "In early attempts by Colbeck (1979) and Marsh and Woo (1985) to describe preferential flow in snowpack models, the water flow in snow is described as a flow in multiple flow paths. In Colbeck (1979), flow paths are defined that differ in size and snowpack properties, which results in different percolation speeds in the individual snow paths when applying Darcy's law. In Marsh and Woo (1985), the snowpack is devided in flow paths of equal size and snowpack properties, but based on comparted lysimeter measurements, it is determined how much of the total flux is transported in each of the individual flow paths. Both approaches never found widespread adoption, probably because they require a-priori specification of the flow path variability (Marsh, 1999)."
  Why exactly the adoption of those approaches was limited is difficult to assess. It may be related to the uncertainty of the additional parameters needed for those descriptions. It may also be that most model developments focussed on processes that required more urgent attention. When taking the SNOWPACK model as an example: the model was orginally developed for avalanche warning purposes and was later developed to study other areas, like snow hydrology, catchment hydrology, Antarctic snow covers, the Greenland ice sheet, etc. Furthermore, there seems to be a strong increase in interest in liquid water flow in snow in the recent years. The availability of high quality laboratory experiments, new modelling techniques, etc., turned the attention again to preferential flow. We therefore do not want to interpret too much why the early modelling attempts did not find widespread adoption.

- P2 lines 19-21: I suspect that the specific application suggested in the example would require at least a 2-d model. The dual-domain approach is a good starting place. Given that the authors point to three flow regimes (matric flow, flow fingering and macropore flow), it would seem appropriate to include multiple flow domains in the future; one step at a time.
  Thank you for the comments. We actually think that the application for wet snow avalanche prediction may already benefit from the dual domain approach, not necessarily requiring a full 2-dimensional model. Here, acquiring accurate snow depth and meteorological forcing in avalanche slopes is probably a more important source of uncertainty than the liquid water flow modelling. We agree that potentially, the dual domain could be extended with additional domains for other water transport processes. However, we think that validation and calibration of the model will become increasingly difficult given the typical inaccuracies in forcing data (meteorological measurements) and validation data (snow lysimeter data) for snow models as well as deficiencies in model process descriptions.

- P7: I suspect that there would be an interaction between the layer thicknesses (resolution) and the threshold of defining ice lenses / layers. A thicker layer would require more water equivalent to reach a given threshold density, and this would be harder to achieve in terms of water and energy transfer. A higher resolution or finer layers should enable higher threshold density values for defining ice layers.
  Indeed, this is illustrated in Fig. 7 in the manuscript. For a specific threshold, the probability of detection is higher using thinner layers, which also means that for the same probability of detection, the threshold can be set higher for thin layers than for thicker layers. However, our interpretation of this figure is

still that the difference between high and normal resolution simulations is much smaller than the difference between observations and simulations. Many discrepancies between observed and simulated snow profiles in terms of layering are probably not caused by the initial layer thickness, but also by errors in meteorological forcing as well as model simplications and model process representation errors.

- P8 lines 7-17: This is an honest discussion of the performance issues but also shows that this methodology may be useful for future development.
  Thank you.

- P9 lines 1-5: Was the precipitation type generally known in the observations or was this 2003 event a specific example of a known case in which the model misclassified the precipitation type? In any event, if it was known that the precipitation was rain, the model's diagnosis could be overwritten for this event, and if precipitation type was generally known, the model could be fed rainfall and snowfall separately. It would remove a source of uncertainty from the results. If the precipitation type was not generally known, but merely appeared to follow expected patterns save for the 2003 example, then it is not necessary to make any changes, given the size of the dataset.
  Precipitation type is generally not known for the Weissfluhjoch measurement series, and the separation in rain and snow is done on the basis of air temperature. However, this particular event is a known case where rain on cold snow formed an ice layer which sustained throughout the snow season. We did experiments to force the model with rainfall for this period, but still the ice layer did not form. Probably the following reasons play a role: the SNOWPACK model is mostly run in 15 minute time steps and processes are solved sequentially. This means that at the beginning of the time step, new snow is added to the domain. Then the heat equation is solved to describe the temperature change over the 15 minute time step. Excess energy present after 15 min. is converted into snow melt, or, vice versa, in refreeze. Then, water flow is calculated over these 15 minutes. The fact that the equations are currently not solved in a coupled way but in a sequential way leads to the situation where water from rainfall can percolate out of the surface layer before there is time to refreeze the water. Furthermore, the water from rainfall is put in a layer of about 1.5 cm. In reality, one can imagine that rain drops hitting a snow surface below freezing will freeze directly at the surface, thereby creating an ice layer. It may even be the case that the ice layer thickens because the pore space at the surface vanishes, creating ponding conditions. Radiative cooling or rain with near surface air temperatures slightly below 0 °C may cause this ponding water to refreeze. If one really wants to force the model to form an ice layer when rain falls on a cold snowpack, we can think of a mechanism where the upper most layer of the snowpack is split and the upper part of the splitted layer is forced to take ice density. We will amend the manuscript at this point, to make clear that even forcing the precipitation to be rainfall did not create an ice layer in the simulations.

- P9 lines 7-13: I would expect the spread in simulated snow density to be somewhat smaller than observed with such a model, based on the fact that snow pits sample small spatial areas, and the presence of discontinuous ice layers and fingering may increase the number of samples necessary to obtain a reliable estimate of the mean density and its variability at a given depth. The PFP simulations in Figure 6 appears to capture the density distribution slightly better than the Richards equation alone. Have other statistical measures been explored as alternatives or in addition to r2? A comparison of the mean and variance may show closer agreement with the PFP simulations. There is Willmot's index of agreement or one of the revised formulations. The objective is not to choose the statistic that makes the model appear better, but r2 has been criticized as being insensitive to important factors of model performance.
  As snow density in the field is generally measured in sections of about 20-30cm depth, the presence of ice

layers, with typical vertical extents of less than 2 cm, is hardly detectable in the measurements. We think that simulations with preferential flow do have a slightly lower performance in reproducing snow density in the simulations, for the reasons mentioned in P9L13-17 in the manuscript. We now also tested using Willmott's index and also with this index, simulations with preferential flow perform slightly worse on overall density distribution (the index decreases from 0.84 *for simulations with Richards equation only to 0.83 for simulations with preferential flow). However, it seems that when changing from the 20-30 cm thick snow segments with measured density to high resolution vertical spacing, the simulations with preferential flow may better capture the variability by reproducing the ice layers. Measuring ice layer density is complicated (Watts et al., 2016) and we currently lack a good dataset of high resolution snow density measurements to validate the density distribution at the vertical resolution of the simulations (ca. 2 cm). Currently ongoing field campaigns using snow micro penetrometry (SMP, Schneebeli and Johnson (1998)) may change this situation soon and offers future potential to more in-depth verify the snow microstructure and snow density in the simulations in high detail. We will mention the test of Willmott's index in the manuscript and provide additional discussion about this point.*

- In looking at Figures 8 and 9, there are differences in the performance of the REQ and REQ+PF models from year to year. It would be interesting to compare the conditions against a ranking of differences in r2 (Fig. 8) and arrival date (Fig. 9). Are there specific snowpack or meteorological conditions that are correlated with the differences in performance between years? Knowledge of this may be useful for future model development.
  We provide here the same response as to Reviewer 1, who raises a similar issue: it certainly is an interesting suggestion to try to explain year-to-year differences in performance of snowpack runoff simulations. However, we did not find a statistically significant correlation between, for example, the r^2 *value or arrival date and the number of ice layers or the number of jumps in grain size or hardness observed in the profiles. We tested both linear correlation using Pearson correlation as well as rank correlation coefficients (Spearman, Kendall). One issue is that, as Reviewer 2 points out, meteorological conditions also vary from year to year. In some melt seasons, percolation speeds through the snowpack are high, in others it takes a few weeks for the whole snowpack to become wet. These differences arise from weather patterns: in some years warm weather prevails for several weeks in the melt season, leading to a quick wetting of the snowpack, whereas in other years, snow melt periods are interrupted by periods with colder weather and new snowfall amounts. We therefore also analyzed the maximum difference in the relative part of the snowpack that consists of melt forms in the observed profiles, between two subsequent snow profiles. This is an indication of the progress of the melt water front inside the snowpack. However, also this did not reveal any statistically significant information. We think that ultimately, there are many factors contributing to high or low* r^2 *values or good or poor estimate of the arrival date: warming rate of the snowpack, presence of capillary barriers and ice layers that may trigger preferential flow, errors in meteorological measurements and errors in snow lysimeter measurements. The snow lysimeter at the Weissfluhjoch measurement site has a surface area of 5* m^2. *As an illustration, Fig. 2 in Kattelmann (2000) shows that with this size, the variation coefficient is still quite large, and that an area of at least 10* m^2 *may be required to more accurately capture snowpack runoff from the snowpack. Furthermore, inconsistencies in the bi-weekly profiles are present due to the subjective component of judging grain size and shape, as well as the fact that multiple observers are responsible for the snow profiles. All these factors make the analysis of factors contributing to high or low model performance difficult. We will discuss this in the revised manuscript.*

Technical Comments:

- P1 line 18: Change 'extend' to 'extent'
  Changed, thank you.

- P1 line 20-23: Change 'Water may flow... instead (Eiriksson et al., 2013)' to 'Water may flow laterally over ice layers or crusts, which reduces travel times and has a significant impact on catchment scale hydrology; alternatively, preferential flow in snow may promote vertical percolation instead (Eiriksson et al., 2013)'.
  Changed, thank you.

- P1 line 23: This reads as if the reader is familiar with the increased melt on the Greenland Ice Sheet, which is likely to be true but, I would add a word or two to indicate the time frame.
  The reference we used to support this claim investigated a period of about 2 decades to relate changes in firn structure to warming trends, although the process may be occurring over longer time scales. We will amend the manuscript at this point.

- P1 line 24: Change 'extends in' to 'extents, on'.
  Changed, thank you.

- P2 line 6 (possibly elsewhere): I find the use of the term 'snow covers' to be awkward. I suggest the use of the term 'snowpack' or 'snowpacks'.
  Changed throughout the manuscript, thank you.

- P2 line 15: I would change 'deployment' to 'adoption'.
  Changed, thank you.

- P7 line 5: Change '... given the suspicion of problems...' to '... due to suspected problems...'.
  Changed, thank you.

- P9 lines 8-9: Change '.... is well reproduced...' to '... is reproduced well...'
  Changed, thank you.

References

Colbeck, S. (1979), Water flow through heterogeneous snow, *Cold Reg. Sci. Technol.*, *1*(1), 37–45, doi:10.1016/0165-232X(79)90017-X.

DiCarlo, D. A. (2013), Stability of gravity-driven multiphase flow in porous media: 40 years of advancements, *Water Resour. Res.*, *49*(8), 4531–4544, doi:10.1002/wrcr.20359.

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.

Kattelmann, R. (2000), Snowmelt lysimeters in the evaluation of snowmelt models, *Ann. Glaciol.*, *31*(1), 406–410, doi:10.3189/172756400781820048.

Marsh, P. (1999), Snowcover formation and melt: recent advances and future prospects, *Hydrol. Proc.*, *13*(14-15), 2117–2134, doi:10.1002/(SICI)1099-1085(199910)13:14/15<2117::AID-HYP869>3.0.CO;2-9.

Marsh, P., and M.-K. Woo (1985), Meltwater movement in natural heterogeneous snow covers, *Water Resour. Res.*, *21*(11), 1710–1716, doi:10.1029/WR021i011p01710.

Schneebeli, M., and J. Johnson (1998), A constant-speed penetrometer for high-resolution snow stratigraphy, *Ann. Glaciol.*, *26*, 107–111.

Watts, T., N. Rutter, P. Toose, C. Derksen, M. Sandells, and J. Woodward (2016), Brief communication: Improved measurement of ice layer density in seasonal snowpacks, *Cryosphere*, *10*(5), 2069–2074, doi: 10.5194/tc-10-2069-2016.