Reply to general comments of Reviewer 1:

The paper presents a new water transport scheme for the 1-D multi-layer physics based SNOWPACK model that accounts for preferential flow effects. The model bases on a dual-domain approach and solves Richards equation for matrix and preferential flow. The area of fingers is explicitly parametrized using results from previously available laboratory experiments. Exchange of water between the matrix and the preferential domains is ruled either by water entry pressure head or by water saturation. The approach is evaluated using an extensive dataset of rain-on-snow events (ROS) from two different locations within European Alps and some field experiments. The proposed scheme demonstrates an improved performance at the scale of single ROS events and at the scale of a snow season.

Including preferential flow in snow models represents an important goal for snow hydrology. This is because it can provide an efficient routing of liquid water through snow and can generate snowmelt runoff earlier than expected. A frequent limitation for snow modelers is that the process understanding is still limited. In this regard, the paper proposes a parsimonious approach that parametrizes the portion of area occupied by fingers and thus takes this process into account without using a full 3-D geometry. The evaluation strategy is extensive and thorough and the paper is generally well written. I have some suggestions for authors that may be included with little effort. I can therefore suggest publication of the paper pending some (minor) revision.

My main suggestion regards Section 3 (Results) and 4 (Discussion). While I generally found both sprinkling experiments and the focus on long-term datasets well motivated and discussed, I am unsure that the two natural ROS events will provide a specific insight into this evaluation. Discussing some "real-world" applications is clearly important, but authors already do that using around 100 ROS events from Davos and Col de Porte. Moreover, results are "partly contradictory" when compared to artificial ROS simulations and this may be understandable as the physics is complex and data may be noisy. This is why focusing on a larger number of events (Section 3.3) is clearly more meaningful. So I suggest that either authors elaborate on the implications of these two specific events, or they remove Section 3.2, move this focus in the Discussion and use it as a starting point for discussing future research.

In the Discussion, I would also try to comment a bit more extensively on the dual domain approach. For example, Eq. 1 relates the area of preferential flow to grain radius, which is for sure the most important variable ruling heterogeneity of water in snow. Because experimental observations of this process are still limited, may you suggest some directions for future research in order to improve this parametrization? May it also depend on supply rate or other conditions of the snow? More importantly, the model includes a parameter that needs to be calibrated. While calibration is helpful to compensate for a lack of physical understanding (and this is definitely the case with preferential flow), it may be interesting for other users to know how did you choose the value of this parameter, or which would be the best calibration protocol for it. This is especially important where lysimeter data are not available. Which is the sensitivity of your results on the value of this parameter?

We thank the reviewer for his constructive comments and ideas to improve the manuscript. Below, we give our response to the issues raised by the reviewer.

We agree that one could question the benefit of including the two natural events into this evaluation (Section 3.2). However, only for these event we have a multi-lysimeter setup, which raises the awareness that the observed processes can show considerably spatial heterogeneity as e.g. documented in (Figure 5). Further, these events show the limitations of the model. In the current version of the manuscript this point may not have been stated clearly enough and we will consider using these events for opening the discussions section as a starting point for discussing limitations of the preferential model and further steps needed in improving the model, as recommended by the reviewer. A more detailed discussion about further research needs and limitations of the model was also requested by Reviewer 2.

As stated by the reviewer, the area of preferential flow (Eq. 1) is likely to also depend on the water supply rate. Data using sandy soils from Glass et al. (1989), shown in DiCarlo (2013), suggests that with increasing system influx rates (100cm/min), the finger width of preferential flow is increasing, whereas it stays small for lower fluxes (20cm/min). However we are not aware of any experiments that have determined the area of preferential flow in snow for influx rates that are typical for natural ROS events. This might lead to false assumptions concerning the area involved in preferential flow and number of fingers, which is in turn important for the refreeze process. Even though we have used the lowest influx rates from Katsushima et al. (2013), these might still lead to bias concerning the preferential flow area. More experimental data under conditions with lower influx rates would be desirable. This is described in Wever et al. (2016) in detail and we will refer to this study more explicitly. Another source of inaccuracy for deriving the preferential flow area are the samples densities (all above 380 kg/m^3) used by Katsushima et al. (2013).

The calibrated parameters are related to a physical process (threshold for saturation and number of flowpaths). Ideally, they should not have to be calibrated for every model application but rather determined from laboratory experiments. We agree that the calibration of the parameters is an important part of the model, and therefore will refer to the study of Wever et al. (2016) more explicitly, where this topic is discussed in greater detail. Here we would like to focus on the discussion about the role of rain-on-snow for preferential flow.

Reply to specific comments of Reviewer 1:

Abstract: I found lines 11 – 17 a bit wordy. Could you try to summarize this? Furthermore, I would also specify the meaning of “balanced” (line 24) as it may be unclear for diagonal readers who are not going to screen the entire text.

Reply: We will rewrite this in the revised manuscript. The term „balanced“ means that for the extensive dataset of WFJ and CDP, the interquartile range is smaller with the PF model and time lag errors are smaller.

Line 29 page 2: may “capillary gradients” work better than “capillarity” alone?

Reply: Colbeck (1972) used the term capillarity. We amend the manuscript, changing „capillarity“ to „capillary forces“. We think that in this context having a general expression suits best.

Line 17 page 3: remains -> remain?;

Reply: It will be changed in the updated manuscript, thanks!

Line 20 page 4: may authors include a brief comment about the reason why snow depth is constrained to observed values in a hydrologic application?

Reply: It is true that for Weissfluhjoch, a dataset with undercatch-corrected precipitation data is available. Nonetheless, because the timing of snowpack runoff is essentially dependent on the snow height, we wanted to exclude this potential source of error to achieve the best comparability between the 3 water transport models. Because we focus on the event-scale we constrained the simulations to the observed snow height, such that we have an accurate snow depth at the onset of the events.

Section 2.2: I would probably be more explicit about the simulated effect of preferential flow on water velocity. In my understanding, the model accelerates liquid water flow in snow because it concentrates water mass in small fingers where unsaturated conductivity is larger than in the matrix domain (and where refreezing is not allowed). Is this a correct interpretation? If yes, I would write something similar in the text in order to clarify this point.

Reply: The model indeed accelerates liquid water flow in snow because it concentrates water mass in a smaller area where the saturation is hence higher and unsaturated conductivity is larger. This happens faster in the preferential domain, representing only a fraction of the snow cover. Additionally, refreezing is not taking place in the preferential domain in the current approach. We will add a clearer description in the updated manuscript.

Eq. 1: should the exponent be negative as in Wever et al. (2016) on TCD?

Reply: The exponent should be identical to the one presented in Wever et al. (2016), and this appears to be the case. If your pdf viewer shows different values it might be a technical error of the pdf document.

Line 19 – 20 page 6: may authors clarify which features of the sprinkler make it "especially developed for sprinkling on snowpack"?

Reply: The sprinkled area and sprinkling intensity depend on the water pressure at the nozzle and the distance of the nozzle to the snow surface. The sprinkling device was calibrated using different pressures at the nozzle and distances to the surface to achieve preferably low intensities within naturally occurring range and at the same time a uniform distribution of sprinkling intensity over the lysimeter area. The device was developed to easily be able to adapt the sprinkling height to the height of the snow cover, so that the distance stays in the calibrated area. It is also lightweight to be able to move, set up the device and conduct the experiments within one day. This is crucial for being able to conduct the sprinkling experiments, but might be of smaller relevance for this study. We therefore decided to delete this sentence, as the device is already described in Juras et al. (2013) and add another reference (Juras et al., 2016).

Line 18 and Table 1: did you choose different portions of snowpack for your experiments at the same sites?

Reply: The multi-lysimeter setup (3-4 at each site) allowed us to use every lysimeter just once. Because they were installed before the first snowfall, the snowpack on the lysimeters was undisturbed. In one case (Klosters, 8-Apr-2015) we used the same lysimeter twice, because the lysimeter became free of seasonal snow cover and the experiment was conducted on fresh snow which fell the day before.

Line 27 page 7: I think including cumulative plots in Fig. 2 may definitely help to understand this methodology;

Reply: We will replace Fig. 2 in the original manuscript by Fig. 1 in this response.

Line 28 page 8: is this Table 1 instead of 2?

Reply: Yes, sorry for causing confusion. This will be changed in the updated version of the manuscript. Thanks!

Section 3.3.2: may you include some additional information about the observed variance of these plots? This may be helpful to put these lines in context;

Reply: See Fig. 2 in this reply letter. The original Figure will be replaced by a similar Figure.

Lines 4 – 22 page 11: I found this paragraph a bit difficult to read. Could you please try to rephrase it and try to reorganize the information around the most important findings? This is a key step in the paper and therefore I think it should be very clear.

Reply: We will adapt the manuscript in this part. Thanks for the advice.

Lines 24 – 28 page 11: which is the temporal resolution of lysimeter data? May this temporal resolution play a role for this analysis?

Reply: The temporal resolution of the lysimeter data used for the extensive dataset in this study is 1 hour. The temporal resolution may clearly play a role for this analysis. Wever et al. (2014) have already shown that for comparing the BA and RE model, improvements by RE are found particularly for subdaily timescales, but are less important for daily sums. Especially R^2 values strongly depend on the correct representation of increasing or decreasing runoff at the given time step. In the revised manuscript, we will discuss the effect of the temporal resolution of the lysimeter by analysing the 30 minute lysimeter data from WFJ.

Line 28 page 12: may refreezing be another important process here? This may be also important at lines 6 – 15 page 13.

Reply: Indeed, refreezing should be discussed here. First, neglecting refreeze in the PF model leads to earlier runoff for the cold snow covers. However, the cold content should be consumed by the end of the event and therefore refreeze should not be accountable for the difference in total event runoff between the RE and PF model. This difference might be attributable to differences in water held in the capillaries. We still think that the main processes of overestimating total event runoff for the RE and PF model are underestimation of water held in the capillaries and high lateral flow, observed during the experiment for cold initial conditions. The effect of high lateral flow is also confirmed by SWE measurements before and after the experiments, which show little changes, being within normal spatial variability and measurement errors. Lateral flow likely lead to an effective loss of sprinkling water per surface area of the lysimeter, which of course cannot be reproduced by the models. The short time lag for the 1st natural ROS event at Davos, even having the coldest snowpack, is contributing to the assumption that refreeze is limited in preferential flow paths. We will add this to the discussion. Thanks for the advice!

Lines 1 – 6 page 14: Katsushima et al. 2013 used a limited range of snow density in their experiments, and this range mostly includes densities greater than 380 kg/m3. May this help to explain this correlation?

Reply: Indeed, this is a likely explanation for this correlation. We will add this to the discussion, as already stated in the reply to the major comments. This might also explain the bad representation of runoff for the natural events where densities ranged from 180-220 kg m^{-3} on Jan 3rd and 250-310 kg m^{-3} on Jan 9th. For the sprinkling events, densities were around 220-270 kg m^{-3} for the winter experiments and 300-400 kg m^{-3} for the spring experiments. Also here we see an improvement in runoff representation with density. See also Fig. 3 in this reply letter.

Figure 2: may the bar be above the x-axis?

Reply: Thanks! The caption will be changed accordingly.

Figures 3, 4, etc.: could authors use different colors for the PF or BA approaches? In these figures, they are very similar and this is not very clear;

Reply: The colours will be changed in all Figures with that problem.

Figure 6: is measured runoff black instead of red?

Reply: Thanks! The caption will be changed accordingly.

References:

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

Glass, R., Steenhuis, T., and Parlange, J.: Wetting Front Instability, 2, Experimental Determination of Relationships Between System Parameters and Two-Dimensional Unstable Flow Field Behavior in Initially Dry Porous Media, Water Resour. Res., 25, 1195-1207, doi: 10.1029/WR025i006p01195, 1989.

Juras, R., Pavlásek, J., Děd, P., Tomášek, V., and Máca, P.: A portable simulator for investigating rain-on-snow events, Z. Geomorphol., Supplementary Issues, 57, 73-89, 2013.

Juras, R., Pavlásek, J., Vitvar, T., Šanda, M., Holub, J., Jankovec, J., and Linda, M.: Isotopic tracing of the outflow during artificial rain-on-snow event, J. Hydrol., 541, 1145-1154, doi: http://dx.doi.org/10.1016/j.jhydrol.2016.08.018, 2016.

Wever, N., Würzer, S., Fierz, C., and Lehning, M.: Simulating ice layer formation under the presence of preferential flow in layered snow covers, Submitted to Cryosphere Discuss., doi: 10.5194/tc-2016-185, 2016.

Figures:

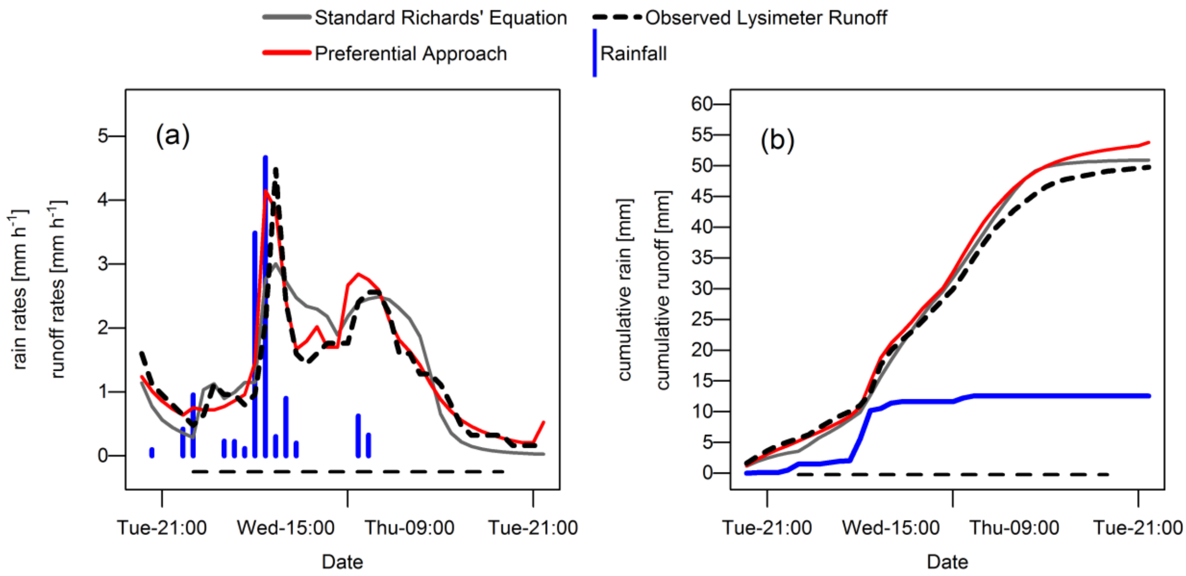

Figure 1: (a) Example of a ROS event occurring at WFJ. The entire extent of the x-axis refers to the evaluation period; the bar above the x-axis refers to the event length. (b) Cumulative version of the plot.

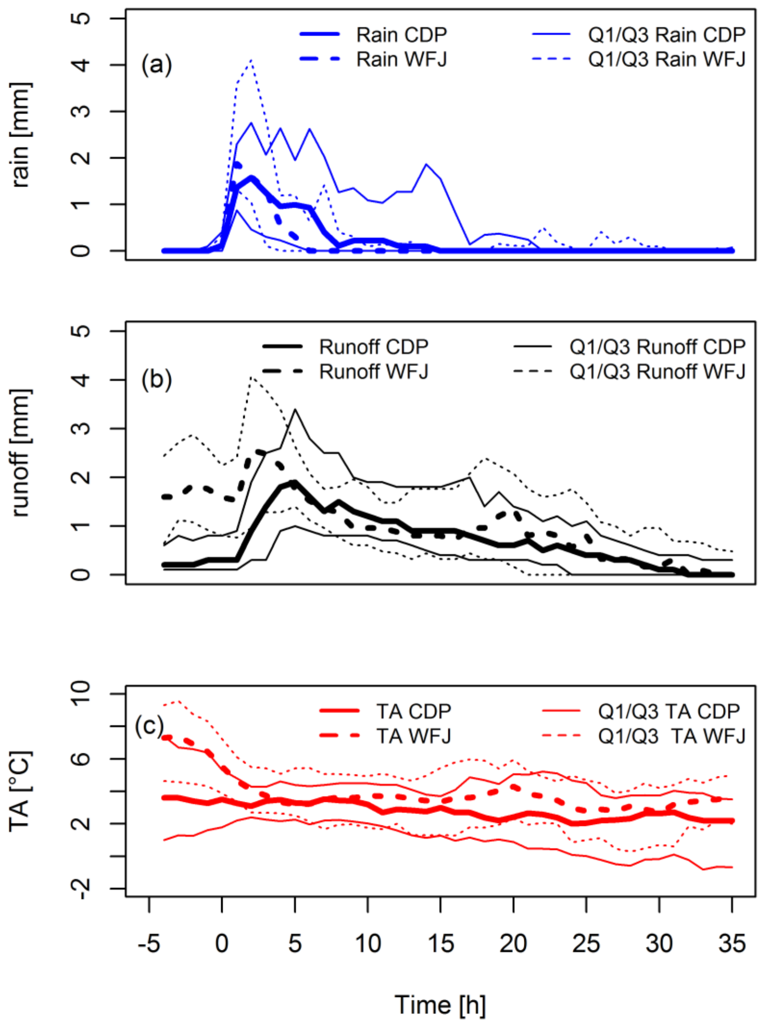

Figure 2: Course of median rain (a), measured snowpack runoff (b) and air temperature (c) for WFJ (dotted) and CDP (solid) for all 40 and 61 events respectively. The thinner lines represent the lower and upper quartiles, respectively. The displayed period is extended by 5 hours prior to event beginning according to the event definition (0 h).

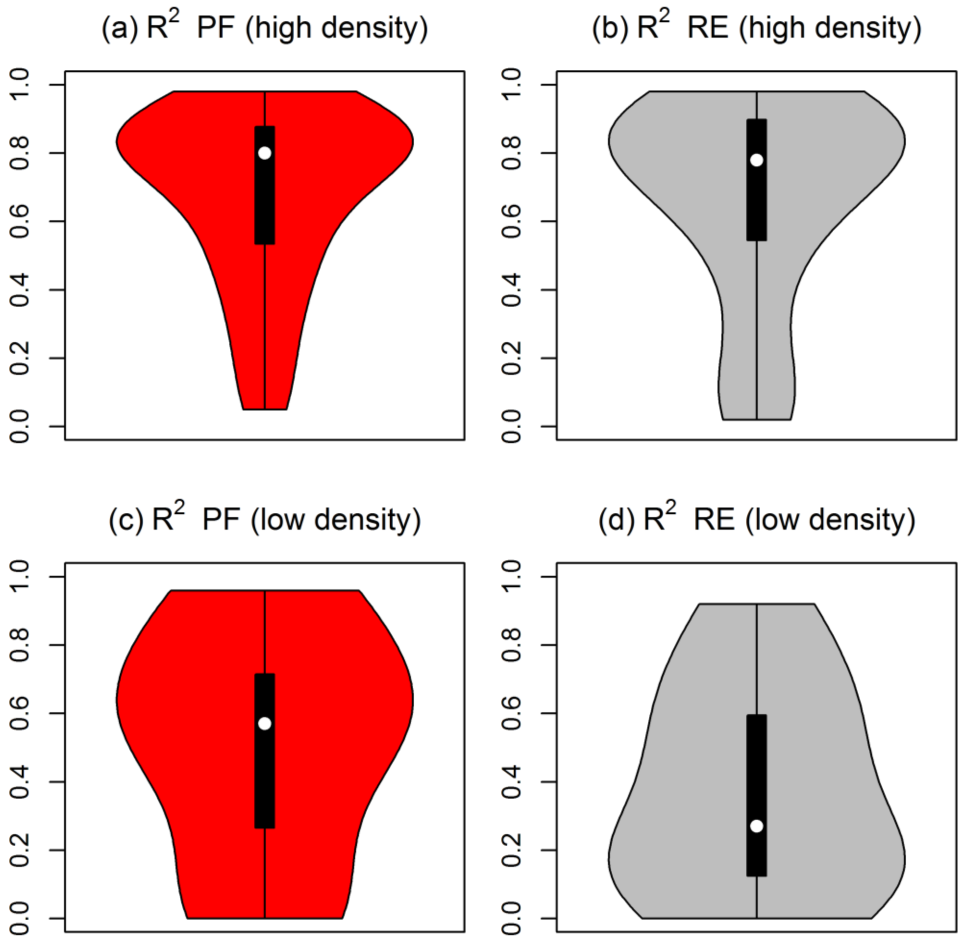

Figure 3: R^2 values for CDP events for the PF (a,c) and RE (b,d) model. The sample is split for bulk densities above 350 kg m^{-3} (a,b) and below 350 kg m^{-3} (c,d).