Dear Prof. Eng. Carlo De Michele,

We would like to thank you and both reviewers for their helpful comments which certainly helped to improve this study. Please find our discussion of the revisions we have made in response to the comments from the two reviewers on the following pages. The track-changes version of the manuscript with all changes marked in red is found at the end of this reply letter. Please note that the line numbers given in the specific replies refer to the original manuscript, whereas the references to changes made in the manuscript refer to the revised manuscript below.

During the revisions we decided to alter the representation of the soil column for WFJ as we discovered instances of ponding on the top soil layer which would constrain snowpack runoff. To allow better comparison with snow lysimeter data we now ensure that infiltration into the soil column is never blocked. Such instances of ponding were only observed for WFJ and with the RE model, consequently only performance metrics related to this model / site changed noteworthy (improved). This along with additional analysis as suggested by reviewer 2 provided better and more consistent insights into the performance of the different models as they relate to snow properties and its stratigraphy.

We are grateful for the interesting reviews and comments that in our opinion enabled to improve the manuscript considerably.

Thank you and with best regards,

Sebastian Würzer (on behalf of the authors)

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 events 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 6). Further, these events show some limitations of the model. In the original 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.

Changes: Because of the reasons stated above, we decided to keep the natural events in place. However, we extended the discussion about the value of the multi-lysimeter approach and the heterogeneity of runoff observations in section 4.

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. (2016b) 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).

Changes: We expanded the discussion about the role of system influx rates in section 4. See page 16 lines 3ff. in the revised manuscript.

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. (2016b) 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.

Changes: For the calibration of parameters, we refer to Wever et al. (2016b). However, we decided to use a threshold in saturation for preferential flow (return flow condition) of 0.06, which has been determined to reproduce runoff best in the sensitivity study presented in Wever et al. (2016b), which is now described in section 2.2 on page 6 lines 8ff.

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.

Changes: We revised the abstract in the revised manuscript, but mainly we decided to be more specific about the results and therefore replace the term "balanced".

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.

Changes: „capillarity" was replaced by „capillary forces" in the revised manuscript.

Line 17 page 3: remains -> remain?;

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

Changes: "remains" was replaced by "remain" in the revised manuscript.

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.

Changes: In the revised manuscript, we now justify the approach of constrained snow heights in section 2.1, page 4, lines 22ff.

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.

Changes: We added a short summary about the principle of the model as suggested by the reviewer in section 2.2, page 6, lines 14ff.

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.

Changes: No changes.

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).

Changes: We changed the sentence according to our reply.

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.

Changes: We revised the corresponding part in section 2.4 on page 7, lines 3ff. to clarify the use of lysimeters in the experimental procedure.

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.

Changes: The figure was extended by a cumulative plot in the revised manuscript.

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!

Changes: The numeration has been changed in the revised manuscript

---

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.

Changes: The figure was extended by the 1st and 3rd quartile of the corresponding values to provide some information about the observed variance.

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.

Changes: We revised the paragraph and we think that the results are now described in a more understandable way.

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.

Changes: In section 2.5, page 8, lines 4ff., we added information about resolution of lysimeter data. To assess the influence of the temporal resolution of the runoff data, we additionally compared simulations and observations at 30 minute time steps for WFJ, where meteorological and lysimeter data were available at higher temporal resolution. With similar values, this analysis was consistent with the results shown for the 1 hour resolution and we decided not to show the data.

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 led 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!

Changes: We added a comment on the role of refreeze in section 4 on page 14 lines 9ff.

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.

Changes: We extended the discussion in section 4 on page 15 lines 18ff. on the role of the experimental data on determining the parameters used in the respective models and their potential implications on the respective model performances. Please consider that the Fig. 3 in this reply shows data from the original manuscript, whereas Fig. 10 in the updated manuscript contains data from new simulations.

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

Reply: Thanks! The caption will be changed accordingly.

Changes: We **changed the caption of the figure 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.

Changes: The colours have been changed in the corresponding figures of the revised manuscript.

Figure 6: is measured runoff black instead of red?

Reply: Thanks! The caption will be changed accordingly.

Changes: We changed the caption of the figure 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:

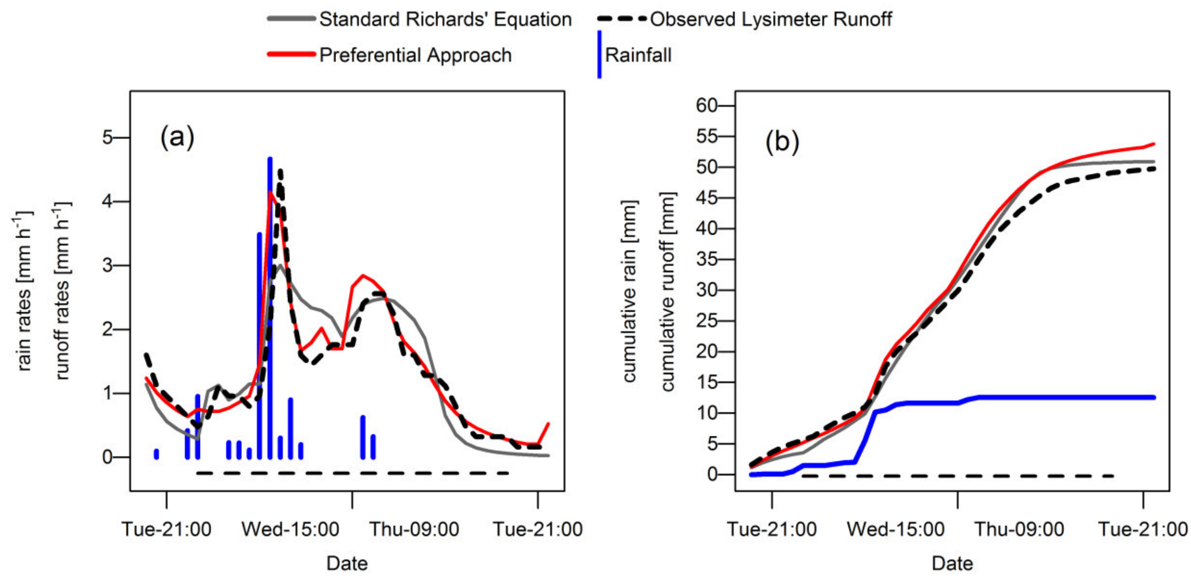


Figure 1: (a) Example of a ROS event occurring at WFJ. The entire extent of the x-axis refers to the evaluation period; the bar above the x-axis refers to the event length. (b) Cumulative version of the plot.

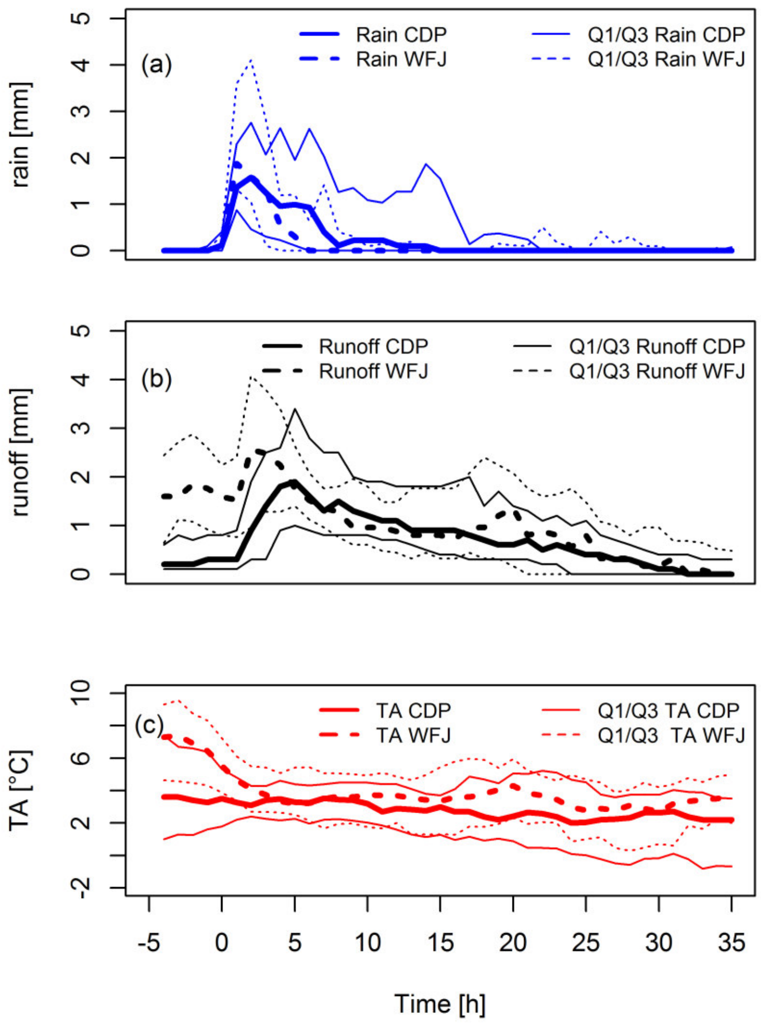


Figure 2: Course of median rain (a), measured snowpack runoff (b) and air temperature (c) for WFJ (dotted) and CDP (solid) for all 40 and 61 events respectively. The thinner lines represent the lower and upper quartiles, respectively. The displayed period is extended by 5 hours prior to event beginning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

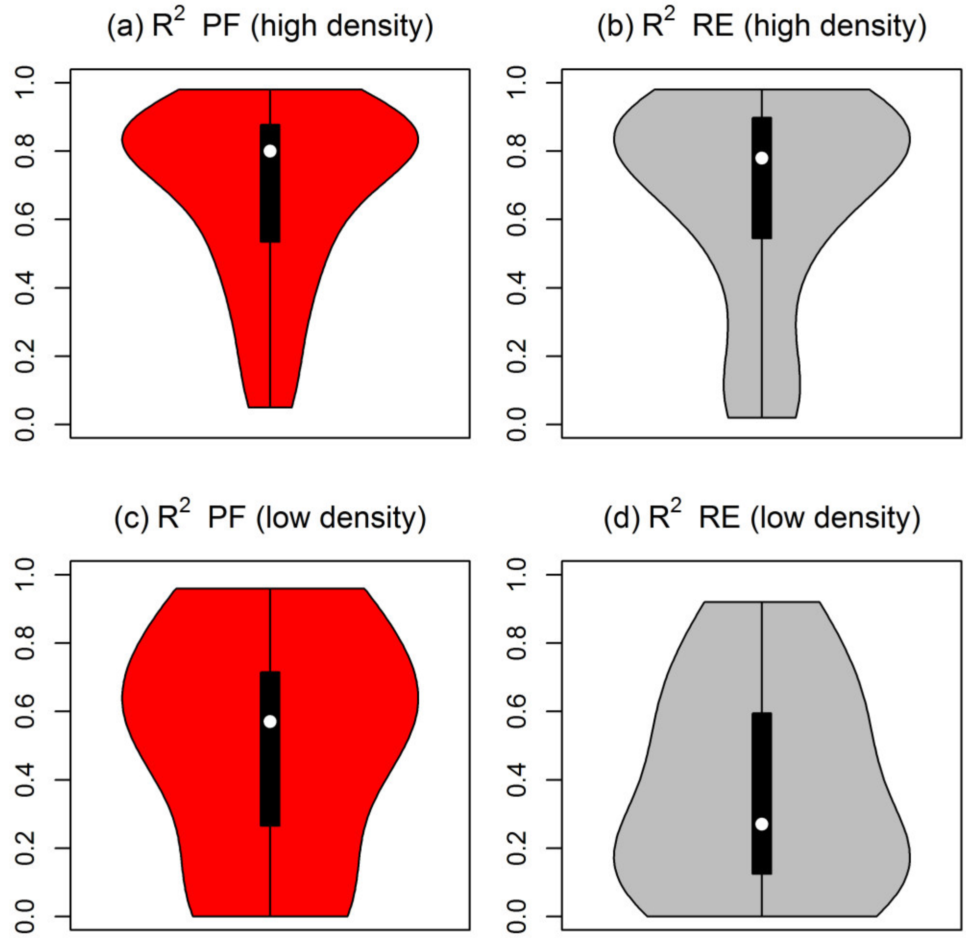


Figure 3: R^2 values for CDP events for the PF (a,c) and RE (b,d) model. The sample is split for bulk densities above 350 kg m^{-3} (a,b) and below 350 kg m^{-3} (c,d).

Reply to general comments of Reviewer 2:

Implementation of preferential flow process into one-dimensional model is challenging and important research. Accuracy of hydrological process will be enhanced by this improvement. Concepts and mechanisms of dual domain approach are described in more detail in the companion paper, Wever et al. (2016). Therefore, the main focus of this paper is the validation of the preferential flow model in terms of accuracy of runoff simulation. In this paper, authors performed the comparison with field data and showed the enhancement of accuracy in runoff by implementation of dual domain approach. The product of this study is appropriate to publish for HESS. On the other hand, although many contents of this paper described the success of this improvement, detailed analysis of the improved results are not sufficient. For example, information of snowpack was not shown and not considered in the discussion despite it affects significantly of the relationship between water supply and runoff. Information of snow stratigraphy helps to understand when and why PF model obtained better accuracy than RE model. In this study, authors used the SNOWPACK model. Therefore, it is not difficult to add the snowpack information. If there are observed data by snow pit observation, showing observed data is desirable. As well as showing snow stratigraphy, quantitative estimation of snowmelt amount is also necessary to discuss runoff as a response of the liquid water input. It can also be estimated from the output of the SNOWPACK model.

We thank the Reviewer 2 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 the snow stratigraphy is a very important factor influencing liquid water transport in the snowpack. It certainly can help understanding when and why the PF model obtained better accuracy than RE model. Eiriksson et al. (2013) for example stated the importance of ice layers and stratigraphic boundaries for runoff formation at the slope scale. Considering findings of Wever et al. (2016), we think that even though considering preferential flow enabled to simulate ice layers, the probability of detection is still not sufficient to analyse the effect of ice layers on runoff generation. So far, we focused on bulk snow cover properties derived by the simulations and their effect on the model performance, as shown in the discussion section in the original manuscript for snow height and bulk density. Additionally, Wever et al. (2016) show the good representation of density observations of RE and PF model at WFJ. We therefore plan to investigate the existence of marked grain size and density changes, representing possible capillary barriers, on the performance of the different models. Also for the experiments and natural events, snow stratigraphy information will be discussed, as far as available.

Changes: A paragraph about stratigraphic boundaries representing e.g. possible capillary barriers was added to the discussion on page 15 lines 22ff.

It is true that snowmelt is an important source of liquid water input for runoff generation. The total influx rate, consisting of both, rainfall and snowmelt can, besides snowpack stratigraphy, be an additional source of variability in the performance of the preferential model, since the area involved in preferential flow has been shown to be dependent on water influx rates for laboratory experiments in soil physics (DiCarlo, 2013). We will consider this in the updated manuscript.

Changes: The snowmelt amounts were very similar for all models for the extensive dataset and therefore cannot explain the different performance of the respective models. However, a paragraph about the influence of liquid water input rates on the model performance for the extensive dataset was added to the discussion on page 16 lines 3ff. The snowmelt amounts for the sprinkling experiments were added to the manuscript on page 9, lines 2ff and 6. For the Davos field site natural events, the differences between cumulative snowpack runoff of the respective models are mainly due to the different amounts of liquid water stored in the snowpack, which are considerably lower in case of the PF model.

Reply to specific comments of Reviewer 2:

P5 L18: This sentence describes Equation (1) is determined by Katsushima et al. (2013) and field observation data. Can you add the data used here (field observation data) in this paper? If it is already shown in previous paper, it should be referenced.

> Reply: The data used from Katsushima et al. (2013) is presented in a graph in Wever et al. (2016b). The field observation data was actually not used for the fit in the end and we will delete that part of this sentence.

Changes: The sentence has been adapted as stated in the reply.

Fig 1 Please indicate the position of the sections of Fig. 1 (b) and (c) in the Fig. 1 (a)

> Reply: Apparently we caused confusion here, as all the figures show different experiments and don't relate to each other. We will revise the caption. (b) relates to Serneus 26.02.15 and (c) relates to the experiment in Klosters on 26.03.15.

Changes: The caption for the figure was changed in the revised manuscript.

P8 L16 Snowmelt amount should be considered in the analysis. It can be estimated from output data of the SNOWPACK.

> Reply: In general, liquid water input and therefore snowmelt represents a substantial amount of liquid water input and is a key factor influencing the water transport. We will estimate the snowmelt from the simulation output. As stated in our reply to the major comments, we also plan to extend the discussion in this sense.

Changes: The snowmelt amounts derived by SNOWPACK simulations have been added to the results of the sprinkling experiments on page 9 line 6.

Fig 3-5 Information of snow stratigraphy had better be added in these figures because it affects the relationship between input water supply and runoff. Although snow depth, averaged snow temperature and water content are shown in Table 2, it is not sufficient because water infiltration process is affected by more complicated snow conditions such as existence of ice layer, grain size contrast and ratio of wet snow to dry snow.

Reply: As stated in the reply to major comments, we can extent the analysis by considering snowpack stratigraphy. In Fig. 1 of this reply letter, we provide the temperature and LWC profiles including identified crusts and ice layers for the sprinkling experiments. Unfortunately, no detailed information about grain size and shape is available for the snow profiles.

Changes: A Figure showing observed profiles of liquid water content, snow temperature and ice layers and crusts for the sprinkling experiments was added to the manuscript (Figure 3).

---

p9 L10-12 I guess that the reason of greater variability of snowpack runoff in highest located site is the existence of lateral flow due to ice layer or capillary barrier. In snowpack observation, are ice layer or capillary barrier existed?

Reply: Unfortunately, no detailed information about snow microstructure stratigraphy is available for the natural events. However, ice layers and the position of layer transitions were qualitatively assessed before the onset of rainfall. For the Event of Jan 03, no distinct ice layers were observed. The snow cover built up within the previous week with mostly very cold temperatures and just a light melt refreeze crust at the top. For the Jan 09 event, the previous event lead to distinct ice layers. But apparently, these differences in the snowpack layering for both events are not expressed by different behaviour of the lysimeter. They are in both cases very heterogeneous in runoff.

Changes: We added a comment about observed ice layers to the results and also discuss the spatial heterogeneity of measures runoff in the discussion section 4 on page 14 lines 21ff.

P10 L20 This sentence indicated that snowmelt affected runoff significantly. Therefore, snowmelt amount should be estimated. Analysis considering snowmelt amount will make better discussion.

Reply: As stated in our reply to the major comments, we plan to extend the discussion in this sense and provide information about snowmelt. We plan on analysing snowmelt as an important part of system input rates and analysing their implication on the performance of the water transport models.

Changes: We analysed the snowmelt contribution during the events. The snowmelt amounts of the respective simulations are very similar and therefore cannot explain the different performance of the models. However, we added a discussion about the effect of water input rates (sum of rain rates and snowmelt rates) on the model performance on page 16 lines 3ff.

P11 L4 RMSE(d,e) -> RMSE (b,e)

Reply: This will be changed in the updated manuscript. Thanks!

Changes: This has been changed in the revised manuscript

P11 L10 R2 value in WFJ was improved by PF model more significant than that in CDP. This result implies preferential flow predominated more in WFJ. Does the ratio of dry snow in WFJ was larger than that in CDP?

Reply: By applying a bulk threshold of 1 vol% LWC to separate between dry and wet initial snow conditions, 30% of events at CDP had a dry initial snow cover, whereas this was the case for just 1 event at WFJ. If looking at the ratio of wet layers to dry layers within the snowpack, this single event had 15% of the layers wet, whereas for all other events at least 99% of layers had a LWC of at least 1 vol%.

Changes: The decision to alter the representation of the soil column for WFJ led to improved performance metrics related to the RE model at WFJ. In the new simulations we show that the difference in PF and RE model performance is rather little for WFJ and in fact bigger for CDP, where we also observe a bigger variability in snowpack properties. The discussion has been extended concerning the effect of snowpack stratigraphy measures having possible implications on the singe event-R^2 on page 15 lines 22ff.

P12 In the discussion section, success of PF model was discussed mainly. However, discussions about shortcomings of the model and suggestion of required improvement are also important for future research.

Reply: Thanks for the advice. We will answer this with the last comment (P14, L30) in detail.

Changes: The discussion on page 16 lines 14ff. has been extended concerning this point.

P14L1 This sentence indicated that snow densities were spread from below 200kg m-3 up to 500 kg m-3 in CDP. But the accuracy of hydrological parameters (e.g. suction and hydraulic conductivity) of low-density snow in numerical snowpack model are insufficient because measurement of them is difficult. They are estimated using equations formulated based on the measured results using high-density snow. For this reason, inadequate accuracy is anticipated when low-density snow comprises a portion of snowpack. Can you add the discussion about the accuracy of runoff simulation in the case of snowpack with low-density snow? It may provide the information whether hydrological parameters of low-density snow should be measured in some way or is not important for runoff estimation.

Reply: Thanks for this comment and advice. This topic was also raised by Reviewer 1. The laboratory experiments from Yamaguchi et al. (2012) and Katsushima et al. (2013) were conducted on snow with densities of 380 kg m^{-3} and above (typically 400-600 kg m^{-3}), therefore much more in range of the densities at WFJ (around 450-500 kg m-3). This could explain the higher variance in R^2 values for runoff at CDP (including densities below 200 kg m-3). We will adapt the discussion accordingly. This suggests that the variable performance of RE and PF models at CDP may be associated with the existence of lower snowpack densities. Fig. 2 in this reply letter shows R^2 values for the CDP-events, split in two samples with initial densities below and above 350 kg m^{-3}, respectively. The PF model shows better performance for lower densities when compared to the RE model. In general, the low-density parameters need more experimental backing – as the reviewer suggested - and we will discuss this adequately.

Changes: A discussion of the variable model performance depending on snow density is now added on page 15 lines 15ff. Please consider that the Fig. 2 in this reply shows data from the original manuscript, whereas Fig. 10 in the updated manuscript contains data from new simulations.

P14 L30 Do you have any suggestion to improve the model? The companion paper, Wever et al., suggested some ideas to enhance the accuracy of ice layer formation. Suggestion to enhance the accuracy of runoff is welcome in this paper. Discussions considering snow stratigraphy help to provide idea for further improvement.

Reply: The suggestions provided in Wever et al. (2016) for improving the preferential flow model apply also for this study. This concerns especially the two parameters which have been calibrated: the threshold for saturation (Θ_{th}) and the number of preferential flow paths for refreeze (N). Laboratory experiments or detailed simulations using multi-dimensional snowpack models might be able to determine the number and size of preferential flow paths for lower input intensities and snow densities. Additionally, we think that very limited data of high temporal and spatial resolution snowpack runoff measurements are available, limiting validation possibilities. CDP and WFJ provide long-term measurements on an adequate temporal resolution; however, this data gives little information about spatial variability of snowpack runoff. Large area multi-compartment lysimeter setups might help improving estimating size, amount and spatial heterogeneity of flow fingers. The lysimeters should be at least 10 m^2 in total area to minimize the effect of preferential flow for the total lysimeter area (Kattelmann, 2000). Sprinkling experiments with low sprinkling intensities on such a device could fill a knowledge gap about water transport in snow under conditions naturally occurring, but under controlled conditions. Also the role of ice layers for vertical water transport is not yet resolved. During the sprinkling experiments in this study, the dry and cold snowpack showed highest lateral flows, but also clearly dominating vertical preferential flow. Our experiments for a ripe snowpack did not show pronounced lateral flow but distinct matrix flow, while Eiriksson et al. (2013) observed lateral flow in saturated layers during wet conditions on a slope. The implications of preferential flow or distinct lateral flow on ice layers or structural transitions can also be observed on the catchment scale. Rössler et al. (2014) had to adjust parameters, leading to very fast overland flow to be able to model the hydrological implications of a major ROS event in October 2011 in Switzerland. It is quite likely that this parameter had to be set that high to compensate for neglecting lateral flow within the snowpack or vertical preferential flow. To be able to better forecast such events, research has to be promoted from experimental lab and field studies over sophisticated multi-dimensional water transport models to simplified but operationally applicable 1D water transport models. We will amend the discussion in the manuscript regarding this point.

Changes: The discussion on page 16 lines 14ff. has been extended concerning current limitations and suggestions how to possibly enhance the accuracy of runoff are now provided.

References:

Eiriksson, D., Whitson, M., Luce, C. H., Marshall, H. P., Bradford, J., Benner, S. G., Black, T., Hetrick, H., and McNamara, J. P.: An evaluation of the hydrologic relevance of lateral flow in snow at hillslope and catchment scales, Hydrol. Process., 27, 640-654, doi: 10.1002/hyp.9666, 2013.

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

Kattelmann, R.: Snowmelt lysimeters in evaluation of snowmelt models, Ann Glaciol, 31, 406-410, Doi 10.3189/172756400781820048, 2000.

Rössler, O., Froidevaux, P., Börst, U., Rickli, R., Martius, O., and Weingartner, R.: Retrospective analysis of a nonforecasted rain-on-snow flood in the Alps-a matter of model limitations or unpredictable nature?, Hydrol. Earth Syst. Sci., 18, 2265-2285, doi:10.5194/hess-18-2265-2014, 2014.

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

Figures:

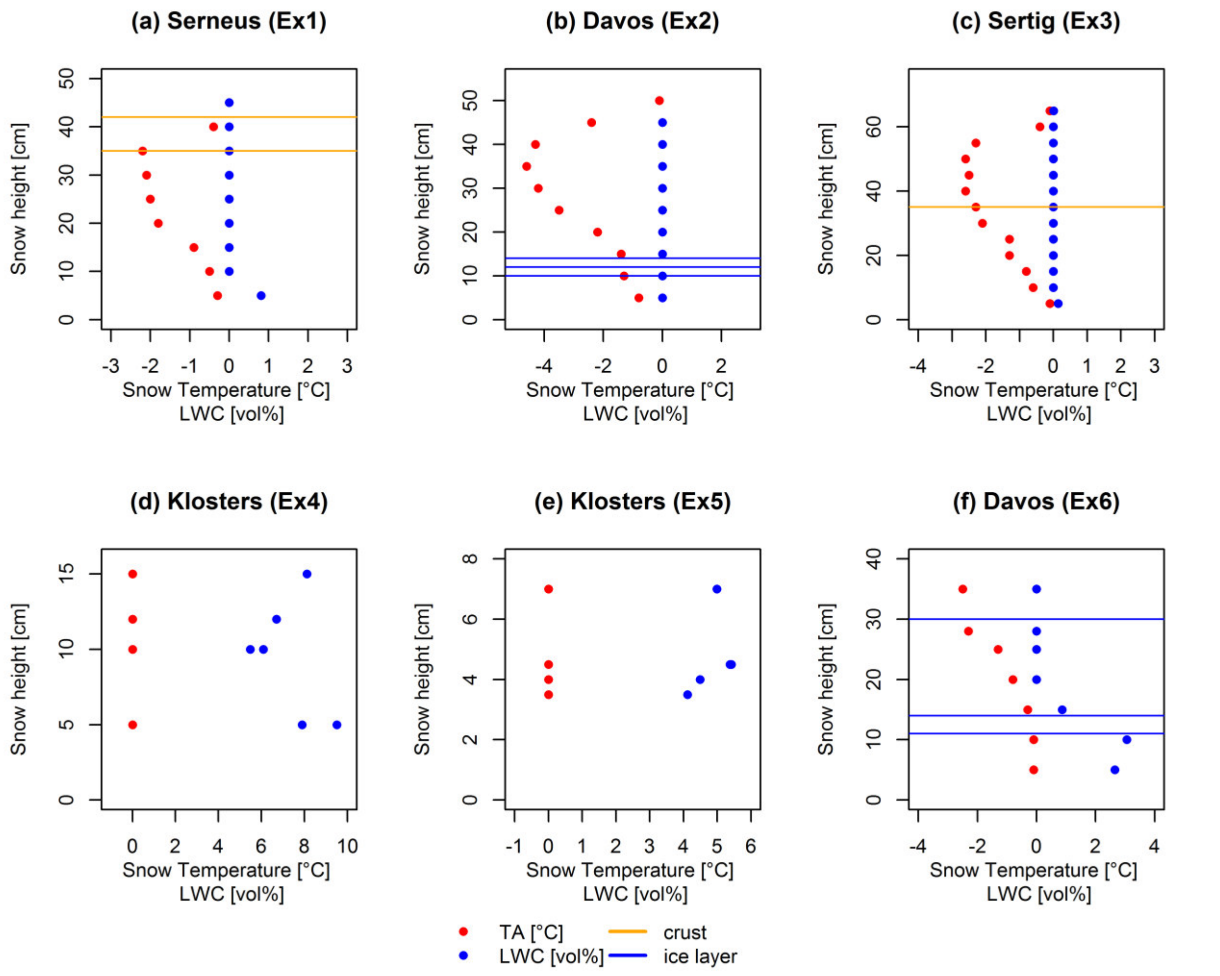


Figure 1: Snow temperature and LWC profiles representing the sprinkling experiment pre-conditions. The lines represent observed ice layers (blue) and crusts (orange).

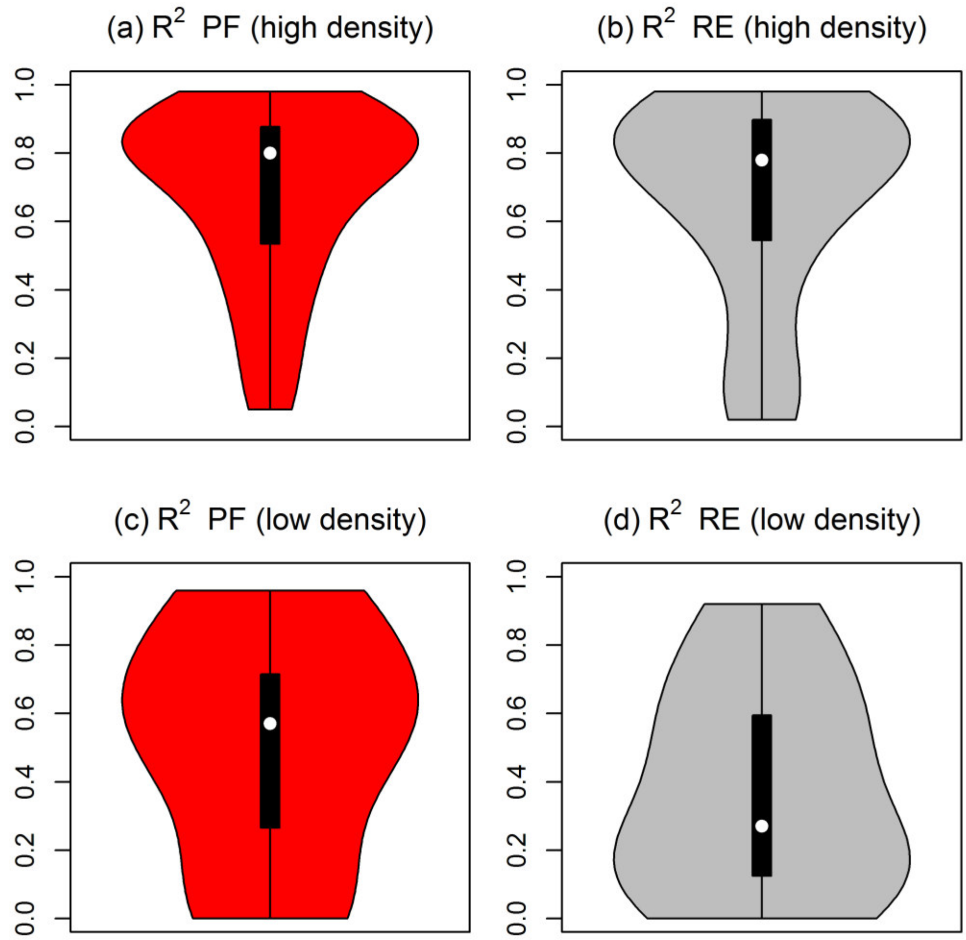


Figure 2: R^2 values for CDP events for the PF (a,c) and RE (b,d) model. The sample is split for bulk densities above 350 kg m^{-3} (a,b) and below 350 kg m^{-3} (c,d). The x-axis describes the density function of the data.

Modeling liquid water transport in snow under rain-on-snow conditions – considering preferential flow

Sebastian Würzer,[1,2*] Nander Wever,[~~1,~~2,1] Roman Juras,[1,3] Michael Lehning,[1,2] and Tobias Jonas,[1]

[1]WSL Institute for Snow and Avalanche Research SLF, Flüelastrasse 11, 7260 Davos Dorf, Switzerland
[2]École Polytechnique Fédérale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering, Lausanne, Switzerland
[3]Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Kamýcká 129, 165 21, Prague, Czech Republic

Correspondence to: Sebastian Würzer (sebastian.wuerzer@slf.ch)

Abstract. Rain-on-snow (ROS) has the potential to generate severe floods. Thus, precisely predicting the effect of an approaching ROS event on runoff formation is very important. Data analyses from past ROS events have shown that a snowpack experiencing ROS can either release runoff immediately or delay it considerably. This delay is a result of refreeze of liquid water and water transport, which in turn is ~~mechanisms in the snowpack. Water percolation is depending~~dependent on snow grain properties but also on the presence of structures such as ice layers or capillary barriers. During sprinkling experiments, preferential flow was found to be a process that critically impacted the timing of snowpack runoff. However, current one-dimensional snowpack models are not capable of addressing this phenomenon~~ correctly~~. For this study, the detailed physics-based snowpack model SNOWPACK is extended with a water transport scheme accounting for preferential flow. The implemented Richards´ Equation solver is modified using a dual-domain approach to simulate water transport under preferential flow conditions. To validate the presented approach, we used an extensive dataset of over 100 ROS events from several locations in the European Alps, comprising meteorological and snowpack measurements as well as snow lysimeter runoff data. The model was tested under a variety of initial snowpack conditions, including cold, ripe, stratified and homogeneous snow. Results show that the model accounting for preferential flow (PF) demonstrated an improved overall ~~and in particular more balanced~~ performance, where in particular the onset of snowpack runoff was captured better. While the improvements were ambiguous ~~small~~ for experiments on isothermal wet snow, they were pronounced for experiments on cold snowpacks, where field experiments found preferential flow to be especially prevalent.

Keywords: *snow cover, water transport, snowpack runoff, mountain hydrology, preferential flow, rain-on-snow, one-dimensional snow model*

1 Introduction

The flooding potential of Rain-on-snow (ROS) events has been reported for many severe floods in the US (Kattelmann, 1997; Kroczynski, 2004; Leathers et al., 1998; Marks et al., 2001; McCabe et al., 2007), but also in Europe (Badoux et al., 2013; Freudiger et al., 2014; Rössler et al., 2014; Sui and Koehler, 2001; Wever et al., 2014b) where for example up to 70% of peak flow events could be attributed to ROS events for Austria (Merz and Blöschl, 2003). With rising air temperature due to climate change, the frequency of ROS is likely to increase in high elevation areas (Surfleet and Tullos, 2013) as well as in high latitudes (Ye et al., 2008). Besides spatial heterogeneity of the snowpack and uncertainties in meteorological forcing, deficits in process understanding make the consequences of extreme ROS events very difficult to forecast (Badoux et al., 2013; Rössler et al., 2014). For hydro-meteorological forecasters, it is particularly important to know a priori how much and when snowpack runoff is to be expected. Particularly, a correct temporal representation of snowpack processes is crucial to identify whether the presence of a snowpack will attenuate or amplify the generation of catchment-wide snowpack runoff. Most studies investigating ROS only consider the generation of snowpack runoff on a daily or multi-day timescale, where an exact description of water transport processes is less important than for sub-daily time scales (Wever et al., 2014a). Water transport processes are further usually described for snowmelt conditions, but not for ROS conditions, where high rain intensities may fall onto a cold snowpack below the freezing point. In this study however, we particularly focus on snowpack runoff generation at sub-daily scales with special attention to the timing of snowpack runoff which is influenced by preferential flow.

Many studies have shown that flow fingering or preferential flow is an important water transport mechanism in both laboratory experiments (Hirashima et al., 2014; Katsushima et al., 2013; Waldner et al., 2004) as well as under natural conditions, using dye tracer (Gerdel, 1954; Marsh and Woo, 1984; Schneebeli, 1995), temperature investigations (Conway and Benedict, 1994) or by measuring the spatial variability of snowpack runoff (Kattelmann, 1989; Marsh and Pomeroy, 1993, 1999; Marsh and Woo, 1985). The variability of snowpack runoff is defined by the distribution and size of preferential flow paths (PFP), which are dependent on the structure of the snowpack and weather conditions (Schneebeli, 1995). Beyond its importance for hydrological implications, preferential flow may also be crucial for wet snow avalanche formation processes, where snow stability can be depending on the exact location of liquid water ponding (Wever et al., 2016a).

Most snow models describe the water flow in snow as a uniform wetting front, thereby implicitly only considering the matrix flow component. The history of quantitative modeling of water transport in snow starts with Colbeck (1972), who first described a gravity drainage water transport model for isothermal, homogeneous snow. This was done by applying the general theory of Darcian flow of two-fluid phases flowing through porous media, neglecting ~~capillarity~~capillary forces. Because water transport is not just occurring in isothermal conditions and snow can therefore not be treated as a classical porous medium, Illangasekare et al. (1990) were the first to introduce a 2D model being able to describe water transport in subfreezing and layered snow. A detailed multi-layer physics based snow model, where water transport was governed by the gravitational part of Richards` Equation described in Colbeck (1972) was introduced by Jordan (1991). With the implementation of the full Richards` Equation described by Wever et al. (2014a), the influence of capillary forces on the water flow was firstly represented in an operationally used snowpack model.

A model accounting for liquid water transport through multiple flow paths was developed by Marsh and Woo (1985), but not being able to explicitly account for structures like ice layers and capillary barriers. Recently, multi-dimensional water transport models were developed, which allow for the explicit simulation of PFP (Hirashima et al., 2014). These models are valuable for describing spatial heterogeneities and persistence of PFP, but have not yet been shown to be suitable for hydrological or operational purposes. In general, multi-dimensional models are limited by the fact that they are computationally intensive, thus not thoroughly validated for seasonal snowpacks and yet lack the description of crucial processes such as snow metamorphism and snow settling.

In snowpack models which are used operationally, PFP are not yet considered. The recently introduced Richards‘ Equation solver for SNOWPACK led to a significant improvement of modelled sub-daily snowpack runoff rates. For this paper, we further modified the transport scheme for liquid water by implementing a dual-domain approach to represent PFP. This new approach is validated against snow lysimeter measurements which were recorded during both natural and artificial ROS events.

The study aims to better describe snowpack runoff processes during ROS events within snowpack models that can be used for operational purposes such as avalanche warning and hydrological forecasting. This requires that the model results ~~remains~~ remain reliable, i.e. that improvements are not realized on the expense of a decreased model performance during periods without ROS, and that the model must not be too computationally expensive. This is the first study to test a water transport scheme accounting for preferential flow which has been implemented in a snowpack model that meets the above requirements.

Our analysis of simulations of over 100 ROS events targets the following research questions:

- Is snowpack runoff during ROS in a 1D model better reproduced with a dual domain approach to account for preferential flow than with traditional methods considering matrix flow only?
- Are there certain snowpack or meteorological conditions, for which the performance specifically benefits if preferential flow is represented in the model?

This paper is structured as follows: Section 2 describes the snowpack model setup, the water transport models, input data and the event definition. Results of the simulations are shown in Sect. 3. This includes data of sprinkling experiments of ROS (3.1), natural ROS events (3.2) and the validation of the model on a long-term dataset from two alpine snow measurement sites (3.3). The results will be discussed in Sect. 4, followed by the general conclusions found in Sect. 5.

2. Methods

All results in this study are derived from simulations with the one-dimensional physics based snowpack model SNOWPACK (Bartelt and Lehning, 2002; Lehning et al., 2002a; Lehning et al., 2002b; Wever et al., 2014a) using 3 different water transport schemes, described in Sect. 2.2. The model was applied to four experimental sites that were set up for this study in the vicinity of Davos (Sect. 2.3). These sites were maintained over two winter seasons between 2014 and 2016 where data was recorded for several natural ROS events. At the same sites, we conducted a set of 6 sprinkling experiments to simulate ROS events for given rain intensities (Sect. 2.4). Furthermore, we conducted simulations for two extensive datasets from the European Alps: Weissfluhjoch (Switzerland, 46.83° N, 9.81° E, ~~2540~~ 2536 m MSL, WSL Institute for Snow and Avalanche Research SLF (2015), abbreviated as WFJ in the following) and Col de Porte (France, 45.30° N, 5.77° E, 1325 m MSL, Morin et al. (2012), abbreviated as CDP in the following). These datasets provide meteorological input data for running SNOWPACK as well as validation data, including snowpack runoff. Both datasets have already been used for simulations with SNOWPACK (Wever et al., 2014a) and provide data over more than 10 years each.

Below, the SNOWPACK model and the different water transport models are described first, followed by the description of the field sites for ROS observation in the vicinity of Davos. Then, we detail the setup of the artificial sprinkling experiments. After summarizing the WFJ and CDP dataset, we finally present the definition of ROS events that is used in this study. Most analyses were performed in R 3.3.0 (R Development Core Team, 2016) and figures were created with base graphics or ggplot2 (Wickham, 2009).

2.1 Snowpack model setup

The setup of the SNOWPACK model is similar to the setup used for simulations in Würzer et al. (2016). For all simulations, snow depth was constrained to observed values, which means that the model interprets an increase in observed snow depth at the stations as snowfall (Lehning et al., 1999; Wever et al., 2015). Because the study focuses on the event-scale and snowpack runoff is essentially dependent on the available snow, this approach was chosen such that we have the most accurate initial snow depth at the onset of the events to achieve the best comparability between the 3 water transport models. The temperature used to determine whether precipitation should be considered rain (measurements from rain gauges) or snow (from the snow depth sensors) was set to achieve best results for reproducing measured snow height for precipitation driven simulations for the Davos field sites (between 0°C and 1.0°C). For WFJ and CDP, this threshold temperature was set to 1.2°C, where mixed precipitation occurred proportionally between 0.7°C and 1.7°C. Turbulent surface heat fluxes are simulated using a Monin–Obukhov bulk formulation with stability correction functions of Stearns and Weidner (1993), as described in Michlmayr et al. (2008). At the Davos field sites (Sect. 2.3) incoming longwave radiative flux is simulated using the parameterization from Unsworth and Monteith (1975), coupled with a clear sky emissivity following Dilley and O'brien (1998), as described in Schmucki et al. (2014). For the roughness length z_0~~z0~~, a value of 0.002

m was used for all simulations at the Davos field sites and WFJ, whereas a value of 0.015 was used for CDP. The model was initialized with a soil depth of 1.4, 2.2 and 2.14 m (for WFJ, CDP and Davos field sites, respectively) divided into layers of varying thickness. For soil, typical values for coarse material were chosen to avoid ponding inside the snowpack due to soil saturation. The ~~Soil~~ soil heat flux at the lower boundary is set to a constant value of 0.06 W m^{-2}, which is an approximation of the geothermal heat flux.

2.2 Water transport models

The two previously existing methods for simulating vertical liquid water movement within SNOWPACK are either a simple so-called bucket approach (BA) (Bartelt and Lehning, 2002) or solving the Richards' Equation (RE), a recently introduced method for SNOWPACK (Wever et al., 2014a; Wever et al., 2014b).

The bucket approach represents liquid water dynamics by an empirically determined irreducible water content Θ_r which defines if water stays in the corresponding layer or will be transferred to the layer below. This irreducible ~~residual~~ water content varies for each layer according to Coléou and Lesaffre (1998). The Richards' Equation represents the movement of water in unsaturated porous media. Its implementation in SNOWPACK and a detailed description can be found in Wever et al. (2014a).

The preferential flow model presented in this study is based on the RE model, but follows a dual-domain approach, dividing the pore space of the snowpack into a part representing matrix flow and a part representing preferential flow. For both domains Richards' Equation is solved subsequently. The preferential flow model is described by (i) a function for determining the size of the matrix and preferential flow domain, (ii) the initiation of preferential flow (i.e., water movement from matrix flow to preferential flow) and (iii) an return flow condition from preferential flow to matrix flow.

The area of the preferential domain (~~F~~F) is as a function of grain size (Eq. 1), which has been determined by results of laboratory experiments presented by Katsushima et al. (2013) ~~and field observations with dye tracer~~:

$$F = 0.0584 r_g^{-1.109} \tag{1}$$

where r_g ~~rg~~ is grain radius (mm). ~~F~~ F is limited between 1% and 90% for reasons of numerical stability. The matrix domain is then accordingly defined as (1-~~F~~F). Water is transferred from the matrix domain to the preferential domain if the water pressure head for a layer in the matrix domain is higher than the water entry pressure of the layer below, which can, according to Katsushima et al. (2013), also be expressed as a function of grain size. This condition is expected to be met if water is ponding on a microstructural transition (i.e. capillary barriers, ice lenses) inside the snowpack. Additionally, saturation was equalized between the matrix and the preferential domain, in case the saturation of the matrix domain exceeded the one in the preferential domain. To move water back into the matrix part, we apply a threshold in saturation of the preferential flow domain and water will flow back to the matrix domain once this threshold is exceeded. This threshold is used as a tuning parameter in the model.

Refreezing of liquid water in the snowpack is crucial for modeling water transport in subfreezing snow and may also be important for modeling preferential flow. The presented preferential flow model has also been used to simulate ice layer formation under the presence of preferential flow by Wever et al. (2016b). Thereby, a sensitivity study on the role of refreeze in the preferential flow domain and the return flow condition from preferential flow to matrix flow was conducted. It was found that neglecting refreeze led to the best results for reproducing ice layer formation, but did not significantly affect the performance in reproducing measured hourly snowpack runoff. Therefore, refreeze in the preferential domain is neglected in the presented study. The threshold in saturation for preferential flow (return flow condition) was also determined by the sensitivity study described in Wever et al. (2016b). While they determined a threshold in saturation ~~as~~ of 0.1 to reproduce ice-layer observations at WFJ best, a value of 0.06 was determined to reproduce observed seasonal runoff best. We therefore used the value of 0.06. In contrast to Wever et al. (2016b), we did not set the hydraulic conductivity in soil to 0, because this can lead to an inaccurate representation of observed lysimeter runoff due to modelled ponding on soil, which is not expected to happen on a snow lysimeter. Further details on the implementation of the PF model and its performance can be found in Wever et al. (2016b).

In summary, the PF model accelerates liquid water transport in the preferential domain by concentrating water mass in a smaller area, representing the area fraction of flow fingers in the snowpack. The saturation in the preferential domain is hence higher and unsaturated conductivity is larger. Further acceleration is achieved by disabling refreeze in the preferential domain.

2.3 Davos field sites

Four field sites have been installed within an elevational range of 950 to 1950 m MSL in the vicinity of Davos, Switzerland, with one meteorological station and 3-4 snow lysimeters each (15 in total, 0.45m diameter). The meteorological stations provided most data necessary for running the SNOWPACK model~~,~~ and missing parameters were estimated ~~according~~ as described in ~~to~~ Sect. 2.1. Lysimeters were installed at ground level with an approximate spacing of 10m horizontal distance. The lysimeters consisted of a funnel attached to a precipitation gauge buried in the ground, which monitored snowpack runoff with a tipping bucket. To block lateral inflow at the snow-soil interface, each lysimeter was equipped with a rim of 5 cm height around the inlet. The multiple snow lysimeter setups allowed analyzing the spatial heterogeneity of snowpack runoff. Snowpack properties (SWE, LWC, HS, TS) were manually measured directly before each natural ROS event so that the initial conditions of the snowpack are known in detail. LWC was measured with the “Denoth meter”, a device introduced by Denoth (1994). The onset of runoff was defined as the time when cumulative snowpack runoff (measured and simulated, respectively) has reached 1 mm.

2.4 Sprinkling Experiment Description

During winter 2014/15, a total of 6 artificial sprinkling experiments were performed on all four Davos field sites described above to be able to investigate snowpack runoff generation for different snowpack properties. For each experiment, a sprinkling device was placed above a snow lysimeter, covered by an undisturbed natural snowpackThe experiments were conducted , i.e. each lysimeter was only used for one experiment. with a sprinkling device especially developed for sprinkling on snowpack. The device used for sprinkling was a refined version of the portable sprinkling device described in Juras et al. (2013) and Juras et al. (2016). For each experiment, the device was placed above a snow lysimeter, covered by an undisturbed natural snowpack. The water used for sprinkling was mixed with the dye tracer Brilliant Blue FCF (concentration 0.4 g l^{-1}) to be able to observe PFPs within the snowpack. Sprinkling was performed in 4 bursts of 30 minutes each, interrupted by 30 minutes breaks. Sprinkling was conducted over a 2x2m plot centered above the lysimeters, and with an intensity of 24.7 mm h^{-1}, leading to a total of 49.4 mm artificial rain in each of the experiments. The intensities were determined by calibration experiments on lysimeters not covered by snow and are valid for a certain distance between the nozzle and the sprinkled surface and water pressure at the nozzle. Despite the fact that this value still represents a very intense ROS event, it is within range of natural ROS events and similar or much lower compared to previous studies (19 mm h^{-1}; Eiriksson et al. (2013); 48–100 mm h^{-1}; Singh et al. (1997)). For the sprinkling experiments, the exact timing of rain and intensities are known and the snowpack runoff measured at 1 minute intervals allowed precisely analyzing the performance of model simulations. Figure 1 shows a horizontal cut of a snowpack after the sprinkling experiment and a topview of the lysimeter after the snowpack was removed for cold and wet conditions, respectively. The blue color indicates where water transport took place and where sprinkled water was held by capillary forces or refrozen.

2.5 Extensive dataset for in-situ validation

Two long-term datasets from two study sites in the European Alps providing snow lysimeter data and high quality meteorological forcing data for running the energy balance model SNOWPACK were chosen to validate the different water transport models systematically. Datasets of both study sites used for the extensive in-situ validation are publicly available. The Col de Porte (CDP) site, located in the Chartreuse range in southeast France has been described in Morin et al. (2012) and the Weissfluhjoch site (WFJ) in the Swiss Alps has been described in Wever et al. (2015). WFJ (46.83° N, 9.81° E) is located at an elevation of 2540 2536 m MSL and CDP (45.30° N, 5.77° E) is located at 1325 m MSL. CDP experiences a warmer climate than WFJ and as a consequence the snowpack produces snowpack runoff more often throughout the entire snow season and ROS events are more frequent than at WFJ. A multi-week snowpack builds up every winter season at CDP, but is, in contrast to WFJ, interrupted by complete melt in some years. The WFJ site is equipped with a 5 m^2 snow lysimeter, which measures the liquid water runoff from the snowpack. It has a 60 cm high rim to reduce lateral flow effects near the soil-snow interface (Wever et al., 2014a). CDP is equipped with both a 5 m^2 and a 1 m^2 lysimeter. Here we use data from the 5 m^2 lysimeter, but include data from the 1 m^2 lysimeter to discuss the uncertainty associated with measurements of the snowpack runoff. The studied period for WFJ is from October 1st 1999 to September 30th 2013 (14 hydrological years). Because of possible errors in the lysimeter data in the winter seasons of 1999/00 and 2004/05 as described in (Wever et al., 2014a), these data were excluded from the study. For CDP the studied period is from October 1st 1994 to July 31st 2011 (17 winter seasons) according to the data availability from the 5 m^2 lysimeter. The temporal resolution of lysimeter data is 1 hour for CDP and 10 minutes for WFJ. Simulation results for CDP and WFJ as well as lysimeter data for WFJ were aggregated to an hourly time scale.

2.6 CDP+WFJ event definition

As the number and characteristics of ROS events are strongly dependent on the event definition, special care needs to be taken to determine begin and end of a ROS event. Being interested in the temporal characteristics of snowpack runoff during ROS, we need to include the entire period from the onset of rain to the end of ROS induced snowpack runoff. Here we use an event definition according to Würzer et al. (2016) with slightly decreased thresholds to identify ROS events. According to this definition, a ROS event requires a minimum amount of 10 mm rainfall to fall within 24h on a snowpack with a height of at least 25 cm at the onset of rainfall. While the event is defined to begin once the first 1 mm of rain has fallen, the event ends once there is less than 3mm of cumulative snowpack runoff recorded within the following 5h. This definition resulted in a selection of 61 events at CDP and 40 events at WFJ. The model simulations were subsequently evaluated over a time window that extends the event length by 5 and 10 hours at the beginning and end, respectively (Fig. 2). These extended evaluation periods allowed to also investigate a possible temporal mismatch between modelled and observed snowpack runoff.

3 Results

3.1 Experimental sprinkling experiments

During the winter period 2014/15, 6 sprinkling experiments (Ex1-Ex6) were conducted on 4 different sites to be able to investigate snowpack runoff generation for different snowpack properties. With distinct differences in snowpack properties but controlled rain intensities, these experiments were expected to reveal the influence of snow cover properties and differences between the water transport models best. For all experiments, initial snow height (HS), snowpack temperature (TS), and LWC profiles were measured (Table 1 and Figure 3). According to these measurements, the snowpack conditions on which the sprinkling experiments were conducted can be separated into two cases: The first 3 experiments were conducted on dry and cold (i.e. below the freezing point) snow and will be called winter experiments. The snowpack of Ex4 and Ex5 was isothermal and in a wet state. At the onset of Ex 6 however, the snowpack was not completely isothermal and had just little LWC. Nevertheless the snowpack already passed peak SWE and was in its ablation phase. Therefore the later 3 experiments (Ex4-Ex6) will be referred to as spring experiments in the following.

For all winter experiments (Fig. ~~3~~ 4 and Fig. ~~4~~5, (a,b,c)), both modeled and observed total event runoff remained below the amount of sprinkling water. Energy input estimated by the SNOWPACK simulations suggests that snowmelt was insignificant for the winter experiments, but refreeze led to significant retention of liquid water. Additionally some sprinkled rain was retained as LWC at the end of the experiments. During Ex3 no snowpack runoff was observed, ~~because~~ visual inspection afterwards revealed an impermeable ice layer covering both the lysimeter and the adjacent ground. During spring conditions, on the other hand, snowmelt (5.1, 8.4 and 27.4 mm respectively) ~~lead~~ led to snowpack runoff exceeding total sprinkling input, except for measured snowpack runoff in Ex6 (Fig. ~~3~~ 4 and Fig. ~~4~~5, (d,e,f)).

Additionally, Fig. ~~4~~ 5 shows, that ~~just~~ only the PF model was capable to reproduce all 4 peaks of observed snowpack runoff for winter conditions (Ex1+2), and even the magnitude of~~if~~ the first peak of Ex1 was ~~strongly underestimated~~captured well. For spring conditions however, all 3 models managed to represent 4 peaks corresponding to the four sprinkling bursts, but the PF model showed best correspondence with observed snowpack runoff (Fig. ~~3~~ 4 and Fig. ~~4~~ 5 (d,e,f); Table 1). Regarding the onset of snowpack runoff, the PF model especially led to faster snowpack runoff for the first 2 winter experiments, where the RE and BA models showed delayed snowpack runoff onset. For spring conditions the faster snowpack runoff response of the PF model led to a slightly early snowpack runoff. Maximal snowpack runoff rates for dry and cold conditions were generally overestimated by all models, ~~in case snowpack runoff was measured and snowpack runoff was simulated,~~ whereas wetter conditions led to a minor underestimation (except for Ex3, where no snowpack was measured).

Regarding the overall correlation between measured and simulated snowpack runoff, PF outperformed the other models (Table ~~2~~1), in particular during winter conditions. Summarizing, this initial assessment suggests that the PF approach has potential advantages in particular a) as to the timing of snowpack runoff and b) for cold snowpacks which are not yet entirely ripened.

3.2 Natural occurring ROS Events

In January 2015, two ROS events occurred in the vicinity of Davos. They were observed over an elevational range of 950 to 1560 m MSL on the same sites on which also the sprinkling experiments were conducted. Figure ~~5~~ 6 shows the course of cumulative rainfall and snowpack runoff for both dates and all sites. Pre-event conditions (HS, LWC, TS) were measured shortly before the onset of rain for both events and are shown together with coefficients of determination (R^2) for hourly snowpack runoff of the different models Table 2.

For the event of 03.01.2015 (Fig. ~~5~~6, upper row) the lower sites Serneus and Klosters (950 and 1200 m asl) showed a similar snowpack runoff dynamics regarding the delayed onset and the total amount (cumulative sum averaged over the 3 corresponding lysimeters: 20.3 mm and 21.1 mm, respectively). Also the heterogeneity between data from the individual lysimeters was relatively low (Range of 3.1 mm and 3.9 mm, respectively). For the highest located site (Davos), however, the snowpack runoff measured by all 4 lysimeters showed a greater variability (Fig. 6c) in the delayed onset of snowpack runoff (0 to 7 hours) and the total amount of snowpack runoff (mean 24.7 mm; range of 57.9 mm). The snow cover mostly built up within one week before the event. Cold temperatures led to a light melt refreeze crust at the top, but no distinct ice layers were observed. For the lower sites (Serneus and Klosters), the PF and RE models generated snowpack runoff too early (PF: ~~2 to~~approx. 3 hours; RE: 0.~~3~~ 2 to 1.4 hours). The BA model generated snowpack runoff rather too late (1.3 to 2 hours), but still within range of the variability of observed snowpack runoff for Serneus. However, the cumulative lysimeter snowpack runoff showed good accordance with modelled PF and RE snowpack runoff at Serneus, whereas ~~the~~PF led to an overestimation at Klosters and BA ~~always led~~ to an underestimation of cumulative snowpack runoff at all sites. At the higher elevation site Davos, the RE model led to a better representation of mean observed snowpack runoff amount, when compared with BA and PF. The mean observed snowpack runoff onset however was represented best by the PF model (0.3 hours early) if being compared to BA (3.~~7~~ 4 hours delay) and RE (1.~~2~~ 1 hours delay).

For the event of 09.01.2015 (Fig. ~~5~~6, bottom row) the lower sites showed again little temporal and spatial heterogeneity in lysimeter runoff (Range of 1 mm and 2.2 mm, respectively), whereas this was more the case for Davos again (Range of 13.3 mm) probably owing to ice layers that were formed after the event on Jan 3rd. Observed mean event snowpack runoff was more diverse for all elevations, where Klosters had the highest cumulative snowpack runoff (Serneus 13.3 mm; Klosters 17.7 mm; Davos 7.8 mm). If compared to observed total snowpack runoff, the PF model overestimated snowpack runoff for Serneus and ~~underestimated snowpack runoff for~~ Klosters, whereas the RE and especially the BA model underestimate event snowpack runoff for both sites. For Davos, all models were overestimating event snowpack runoff and led to early snowpack runoff. Except the RE model, which represented onset of snowpack runoff correctly for Serneus, none of the models were able to model snowpack runoff onset correctly for any of the sites.

3.3 Validation on an long-term dataset

3.3.1 Modeled and observed snowpack runoff for the whole dataset

Given the partly contradictory findings on the performance of the three model variants based on the above assessment for artificial ROS simulations under controlled conditions (Sect. 3.1), as well as natural ROS events (Sect. 3.2), further more systematic model tests were needed. Therefore we validate the different models based on extensive datasets from the two sites WFJ and CDP, as described in Sect. 2.4.

Before we focus on the specific performance of the PF model for a large number of individual ROS events, we first analyzed the overall model performance throughout the whole study period, i.e. over entire winter seasons. ~~Therefore~~ For this, we analyzed observed and modeled hourly snowpack runoff provided snow heights ~~were above~~exceeded 10 cm to ensure that lysimeter runoff was caused by snowpack runoff and not rainfall. For both sites, R^2 values for PF were slightly higher than for RE (Table 3), which both clearly outperformed the BA. Also the root mean squared errors (RMSE) of the PF model were lower compared to RE and BA. We can therefore conclude that the implementation of the PF approach slightly improves water transport over entire winter seasons.

3.3.2 ROS event characteristics of the extensive dataset

Average Median characteristics of the individual ROS events at CDP and WFJ are summarized in Fig. 67. The temporal course of median rain and snowpack runoff rates averaged overof all events at WFJ (40 individual events) and CDP (61 individual events) are shown in Fig. 6 7 (a, b). ROS events at WFJ showed, on average, generally higher maximum rain intensities than at CDP, leading to higher average median snowpack runoff intensities in the beginning of the events. Whereas at WFJ, ROS events tended to be short and intense, at CDP the event rainfall extended over a longer period of time. Interestingly, we observed relatively high initial snowpack runoff rates before the actual begin of the ROS event, especially for WFJ, which suggests that many ROS events at this site occurred during the snowmelt period. Averaged over all individual events,Median snowpack runoff reached a peak after 1 and 4 3 hours after the onset of rain for WFJ and CDP, respectively. At WFJ snowpack runoff and rain rates in the beginning of the events were generally higher than at CDP. The course of the mean median air temperature during ROS events at both sites is shown in Fig. 6 7 (bc). For both sites,Especially for WFJ, mean median air temperature (TA) dropped with the onset of rain. At WFJ, this drop was more distinct and mean median TA was higher than at CDP. The mean initial ROS event snow height (HS) for WFJ was 95 cm, which is approximately the average snow height during mid-June (for 70 years of measurements). The mean initial HS for CDP is 67 cm. With a SD of 42 cm, the variability of initial HS for WFJ was higher than for CDP (29 cm).

3.3.3 Modelled and observed snowpack runoff at the event scale

Below we investigate the performance of the three water transport schemes at the event scale. Modeled snowpack runoff was assessed against observations by the coefficient of determination (R^2) and the root mean squared errors (RMSE). To further analyze the representation of snowpack runoff timing, we defined an absolute time lag error (TLE) as the difference between the onsets of modelled and observed snowpack runoff in hours. The onset of snowpack runoff is defined as the time when cumulative snowpack runoff has reached 10% of total event-snowpack runoff.

Figure 7 8 shows boxplots of R^2 (a,d), RMSE (db,e) and absolute TLE (c,f) for all 40 ROS events at WFJ (a,b,c) and 61 events at CDP (d,e,f), respectively. For both sites, R^2 values show that the BA model performance was inferior to the RE model which was in turn slightly outperformed by the PF model. The interquartile range of R^2 values for CDP was generally higher than for WFJ and increased from BA to RE, whereas it was decreasing for PF. The PF also led to a reduction in RMSE by approximately 50% if compared to the BA, but less (9% for WFJ and 25% for CDP) if compared to the RE model. Whereas the median of TLEs for all models at WFJ was 0 and therefore all models reproduced the onset of snowpack runoff very well, the interquartile range decreased from BA to the RE and PF models. The same behavior in interquartile range decrease could be observed for CDP, where the magnitude of TLE was higher than for WFJ and mostly negative. The median TLE was again 0 for the PF and -1 h in the case of BA and RE, indicating that for these models, snowpack runoff was on average a bit delayed compared to the observations. For WFJ, TLE for BA was more often positive (early modelled snowpack runoff), which led to a very good median for BA, but also a larger interquartile range. Hence, the PF model showed the most consistent results, especially if regarding the interquartile range. For CDP we added the comparison between the 1 and 5 m^2 lysimeters installed at CDP (Sect. 2.5) as a reference to Fig. 8, referred to as RL. This comparison can be seen as a benchmark performance, as it represents the measurement uncertainty of the validation dataset. As expected, RL shows the highest overall performance measures, but while the results for both PF and RE were reasonably close to those of RL, the BA model performed considerably worse.

For WFJ, R^2 values show that the BA model performance was inferior to the RE model which was in turn outperformed by the PF model. The PF also led to a reduction in RMSE by approximately 50% and 20%, if compared to the BA and RE model, respectively. Whereas the median of TLEs for all models at WFJ was 0 and therefore all models reproduced the onset of snowpack runoff very well, the interquartile range decreased from BA via RE to the PF model. For CDP, a distinct increase in R^2 values could be observed between BA and both RE and PF which showed a similar median R^2. The interquartile range of R^2 values was generally higher than for WFJ and increased from BA to RE, whereas it was decreasing for PF. Also the RMSEs significantly decreased with RE and PF, compared to the BA model. Similarly to WFJ, the median TLE for CDP was zero, except in the case of RE, where negative median TLE indicates that modelled snowpack runoff was on average a bit delayed compared to the observations. Nevertheless, the PF model showed the most consistent results, whereas the BA model showed the largest spread in TLE for individual events. The magnitude of TLE was generally higher for CDP than for WFJ and mostly negative, which means that the modelled snowpack runoff was delayed compared to lysimeter snowpack runoff. For BA and PF, TLE was more often positive (early modelled snowpack runoff), which led to a very good median for BA, but also a larger interquartile range. The PF model led to the same median as the RE model, but showed the smallest interquartile range. As reference we added the comparison between the 1 and 5 m^2 lysimeters installed at CDP (Sect. 2.5) to Fig. 7, referred to as RL. This comparison can be seen as a benchmark performance, as it represents the measurement uncertainty of the validation dataset. As expected, RL shows the highest overall performance measures, but while the results for both PF and RE were reasonably close to those of RL, the BA model performed considerably worse.

The results shown in Fig. 7 8 may be influenced by both a time lag as well as the degree of reproduction of temporal dynamics. To separate both effects, we conducted a cross-correlation analysis, allowing a shift of up to 3 hours to find the best R^2 value. Figure 8 9 shows both the time lag, as well as the best R^2 value achieved. Interestingly, the BA model showed best correlations if the modeled snowpack runoff was shifted by 1 or 2 hours (consistently too early compared to observations). The RE model, on the other hand, showed best correlations for a shift in the other direction (consistently too late compared to observations). Neither was the case for PF with lags centered around 0.

The R^2 of the cross correlation analysis gives some indication of how well the temporal dynamics of the observed snowpack runoff can be reproduced, neglecting a possible time lag. The results in Fig. ~~8~~ 9 show an improvement in R^2 values for both sites and all models if a time lag is applied. Greatest improvements were observed for the BA model~~, which even outperformed the RE model at WFJ, albeit not for CDP~~ for both sites. The good timing with the PF model is confirmed by almost no lag for WFJ and only a small lag for CDP needed to maximize R^2. For CDP, ~~B~~both RE and PF had maximized R^2 values in range of the lysimeter comparison (RL).

4 Discussion

Even though preferential flow of liquid water through snow is a phenomenon that is known and investigated since a long time, it has not yet been accounted for in 1D snow models that are in use for operational applications. The results of this study show that including this process into the water transport scheme can improve the prediction of snowpack runoff dynamics for individual ROS events as well as for the snowpack runoff of entire snow seasons. Moreover, the representation of the onset of snowpack runoff is improved. This is particularly important at the catchment scale, where a delay of snowpack runoff relative to the start of rain may affect the catchment runoff generation, especially if the time lag varies across a given catchment.

During the sprinkling experiments, sprinkling intensities were higher than average rain intensities during ROS but still within range of peak rain intensities during naturally occurring ROS events in the Swiss Alps (Rössler et al., 2014; Würzer et al., 2016) and the Sierra Nevada, California (Osterhuber, 1999). The use of the PF model clearly led to a better representation of the runoff dynamics for all experiments, including shallow and ripe snowpacks during spring conditions as well as cold and dry snowpacks representing winter conditions. The improvements were strongest for winter conditions, suggesting that under these conditions accounting for preferential flow is most relevant. This is supported by observations of preferential flow paths during winter conditions (Fig. 1 (a)), which were not visible after the spring experiments. During winter conditions just a fraction of the lysimeter area was colored with tracer, indicating preferential flow of the sprinkled water (Fig. 1 (b)), whereas spring conditions left the whole cross section of the lysimeter colored (Fig. 1 (c)). While a fast runoff response can be expected for wet and shallow snowpack and may be easier to handle for all models tested, it is the cold snowpacks that both RE and BA models did not manage to represent well: runoff from these models was more than one hour delayed (Ex1 and Ex2), and missed approx. 10 mm of snowpack runoff within the first hour of observed runoff. This can partly be explained by the fact that BA and RE need to heat up the subfreezing snowpack before they can generate snowpack runoff, whereas refreezing is neglected in the preferential domain of the PF model and runoff can occur even in a not yet isothermal snowpack. Adjusting parameters like the irreducible water content Θ_r for the BA model could probably lead to earlier runoff under these conditions, but thereby lead to earlier runoff, for example for WFJ events, where TLE already is positive for several events.

Despite the improved representation of the temporal runoff dynamics of the PF model (Table 1), the total event runoff of both RE and PF models is very similar for most conditions. Notably, the total event runoff for dry snowpacks is mostly overestimated by all models, suggesting an underestimation of water held in the capillaries. In cold snowpacks, dendricity of snow grains may still be high, such that water retention curves developed ~~with~~ for rounded grains underestimate the suction. Additionaly, high lateral flow was observed during the experiment for those conditions (Fig. 1a). This leads to an effective loss of sprinkling water per surface area of the lysimeter, which of course cannot be reproduced by the models. Therefore, observed snowpack runoff likely underestimates the snowpack runoff that would have resulted from an equivalent natural ROS event and we assume that the performance of the PF and RE models to capture the event runoff is probably better than reported in Table 1. Note that neglecting refreeze in the PF model should not be accountable for differences in the total event runoff between the RE and PF model, if we assume that the cold content is depleted by the end of the event.

Interestingly, despite having the coldest snowpack, time lag for the 1st natural ROS event at Davos was shorter than for the other 2 sites. This relationship where a cold and non-ripe snowpack led to smaller lag times was also found during sprinkling experiments conducted by Juras and Würzer (unpublished data). We assume that this is an indication for the presence of pronounced preferential flow paths under those conditions, which is also supported by the high spatial variability of snowpack runoff. Glass et al. (1989) state that the fraction of preferential flow per area is decreasing with increasing permeability, which itself ~~has~~ was found to be increasing with porosity (Calonne et al., 2012). Therefore, with a decreasing preferential flow area due to lower densities, the cold content of a snowpack loses importance, but saturated hydraulic conductivity is reached faster within the preferential flow paths. The combination of those effects then is suspected to lead to earlier runoff. This behavior should be ideally reproduced by the PF model and indeed the onset of runoff is caught well for this event. Here, our multi-lysimeter setup raises the awareness that the observed processes can show considerably spatial heterogeneity as e.g. documented in Figure 6. Also the formation of ice layers underlies spatial heterogeneity. Moreover the creation of preferential flow paths is strongly dependent on structural features like grain size transitions leading to capillary barriers. Unfortunately, no detailed information about grain size is available in the observations to verify this.

The PF model led to improvements ~~for~~ in reproducing hourly runoff rates at CDP and WFJ for a dataset comprising several years of runoff measurements. This is an important finding, demonstrating that the new water transport scheme aimed at a better representation of preferential flow during ROS events, did not negatively impact on the overall robustness of the model. To the contrary, the overall performance over entire seasons could even be improved. ~~Whereas a~~All 3 models represent the overall seasonal runoff better for WFJ than for CDP (Table 3), ~~this~~ which was ~~not~~ also found on the event scale (Fig. ~~7~~8). ~~However~~Moreover, the CDP simulations exhibit a larger interquartile range in R^2 values and are therefore generally less reliable. The observed differences in model performance between both sites may either be caused by differences in snowpack or meteorological conditions or by issues with the observational data. Moreover, SNOWPACK developments have in the past often been tested with WFJ data, which could lead to an unintended calibration favoring model applications at this site. Despite an obvious contrast in the elevation of both sites, the average conditions during ROS events seem to vary. Figure ~~6~~ 7 suggests that at WFJ short and rather intense rain events dominate. The higher maximum rain intensities at WFJ, compared to CDP, are probably due to the later occurrence of ROS at this site (May-June), where air temperatures and therefore rain intensities are usually higher than earlier in the season (Molnar et al., 2015). Regarding mean intensities over the event scale, data shown in Fig. ~~6~~ 7 further imply that short and intense ROS events typically attenuate the rain input (ratio runoff to rain < 1), whereas long ROS event rather lead to additional runoff from snowmelt, which is in line with results presented in Würzer et al. (2016).

Snow height is generally higher at WFJ where the average initial snow height for the ROS events analyzed was 30 cm higher than at CDP. Ideally, the performance of the water transport scheme in the snowpack should not be affected by the snow depth. At both sites, the snowpack undergoing a ROS event is mostly isothermal with a mean initial LWC of 1.~~9~~ 8 vol% (CDP) and 3.~~3~~ 0 vol% (WFJ). The initial snowpack densities at both sites were quite different. At WFJ, densities for all ROS events are around 450-500 kg m^{-3}, whereas for CDP densities are spread from around ~~below~~ 200 kg m^{-3} up to 500 kg m^{-3}. This suggests that the variable performance of all models at CDP (Fig. ~~7d~~8d) may be associated with early season ROS events. At CDP, a linear regression fit suggests a positive, albeit weak correlation ~~of~~ between snowpack bulk densities and event-R^2 for the RE (R^2 of 0.2), but no correlation for both ~~and~~ the PF ~~model at CDP, but not for~~and the BA model. It seems that the RE model had some difficulties with low density snow, which was not the case for the PF model (Figure 10). This may explain why PF outperformed RE at CDP, but not for WFJ.

Remaining inaccuracies in the representation of runoff for low densities for both models applying the Richards`Equation may be explained by the fact that the water retention curve have been derived by laboratory measurements with high density snow samples (Yamaguchi et al., 2012). Also, the parameters defining the preferential flow area (*F*) have been developed from snow samples with a density mostly above 380 kg m^{-3} (Katsushima et al., 2013).

We further analysed snowpack stratigraphy derived from the SNOWPACK simulations, such as marked grain size changes (bigger 0.5 mm) and density changes (bigger 100 kg m^{-3}) in two adjacent simulated layers as well as the wet layer ratio (percentage of layers exceeding 1 vol% over layers below 1 vol%) and the percentage of melt forms (Table 4). These stratigraphy measures represent possible capillary barriers having implications on the singe event-R^2 and might help understanding the advantages and disadvantages of the different models. Any considerable correlation between the abundance of stratigraphy features and event-R^2 would be indicative of potential errors in the respective model. Negative albeit small correlations could be found between the number of grain size changes and the event-R^2 for WFJ. Similar correlations were noted with regards to the number of changes in density between layers for the RE and PF model. In both cases correlations were less negative for the PF model indicating a more balanced and ultimately less degraded performance with increasing number of potential capillary barriers. While at WFJ most events occurred with ripe snow this was not the case for CDP. There, positive correlations were found between the ratio of melt forms and the wet layer ratio with event-R2 for the RE model (0.33 and 0.44) and for the PF model (0.14 and 0.16). Also in this case the PF model showed more balanced results that were less influenced by the initial LWC, which is in line with our findings of the sprinkling experiments.

System input rates (sum of melt rates and rain rates) are known to significantly affect water transport processes. E.g. the area of preferential flow (Eq. 1) is likely to depend on the water supply rate. Data using sandy soils from Glass et al. (1989), shown in DiCarlo (2013), suggest that with increasing system input rates the finger width of preferential flow is increasing. Even though we have used the lowest influx rates from Katsushima et al. (2013), these rates still exceeded what seems representative of natural ROS events. We therefore analysed the effect of system input rates on the performance of our water transport models. Positive, albeit weak correlations (R^2 of 0.07 to 0.21) could be observed between event-R^2 and system input rates for all models, suggesting that they generally performed (slightly) better for higher influx rates. For the PF model this could probably be explained by the preferential flow parameters depending on laboratory measurements with high influx rates. For WFJ on the other hand, a clear correlation between R^2 and HS was found for the BA model (R^2 =0.44), but not for RE and PF model. This leads to the assumption that performance of RE and PF model is slightly better for higher densities, whereas the BA performance is primarily dependent on snow height.

In combination with the hydraulic properties for lower density snow samples, additional laboratory experiments might be able to determine the number and size of preferential flow paths for lower input intensities and snow densities. Especially the calibrated parameters threshold for saturation (Θ_{th}) and the number of preferential flow paths for refreeze (N) could benefit from such experimental studies. Even though CDP and WFJ provide long-term measurements on an adequate temporal resolution, this data gives little information about spatial variability of snowpack runoff limiting further validation opportunities. Large area multi-compartment lysimeter setups might help improving estimating size, amount and spatial heterogeneity of flow fingers. Sprinkling experiments with preferably low sprinkling intensities on such a device could fill a knowledge gap about water transport in snow under naturally occurring conditions.

5 Conclusions

A new water transport model is presented that accounts for preferential flow of liquid water within a snowpack. The model deploys a dual-domain approach based on solving the Richards‘ Equation for each domain separately (matrix and preferential flow). It has been implemented as part of the physics based snowpack model SNOWPACK which enables for the first time to account for preferential flow paths within a model framework that is used operationally for avalanche warning purposes and snow melt forecasting.

The new model was tested for sprinkling experiments over a natural snowpack, dedicated measurements during natural ROS events, and an extensive evaluation over 101 historic ROS events recorded at 2 different alpine long-term research sites. This assessment led to the following main conclusions:

Compared to alternative approaches, the model accounting for preferential flow (PF) demonstrated an improved overall ~~and in particular more balanced~~ performance, particularly for lower densities and initially dry snow conditions~~,~~. This led to ~~by showing~~ smallest interquartile ranges for R^2 values and considerably decreased RMSE for a set of more than 100 ROS events. When evaluated over entire winter seasons, the performance statistics were superior to those of a single domain approach (RE), even if the differences were small. Both PF and RE models, however, outperformed the model using a bucket approach (BA) by a large margin (increasing median R^2 by 0.49 and 0.48 ~~23 and 0.39~~ for WFJ and 0.~~47~~53 and 0.~~46~~48 for CDP). In sprinkling experiments with 30-min bursts of rain at high intensity, the PF model showed a substantially improved temporal correspondence to the observed snowpack runoff, in direct comparison to the RE and BA models. While the improvements were small for experiments on isothermal wet snow, they were pronounced for experiments on cold snowpacks.

Model assessments for over 100 ROS events recorded at two long-term research sites in the European Alps revealed rather variable performance measures on an event-by-event basis between the three models tested. The BA model tended to predict too early onset of snowpack runoff for wet snowpacks and a delayed onset of runoff for cold snowpacks, whereas RE was generally too late, especially for CDP. Combined with results from a separate cross correlation analysis, results suggested the PF model to provide the ~~most balanced~~best performance concerning the timing of the predicted runoff.

While there is certainly room for improvements of our approach to account for preferential flow of liquid water through a snowpack, this study provides a first implementation within a model framework that is used for operational applications. Adding complexity to the water transport module did not negatively impact on the overall performance and could be done without compromising the robustness of the model results.

Improving the capabilities of a snowmelt model to accurately predict the onset of snowpack runoff during a ROS event is particularly relevant in the context of flood forecasting. In mountainous watersheds with variable snowpack conditions, it may be decisive if snowpack runoff occurs synchronously across the entire catchment, or if the delay between onset of rain and snowpack runoff is spatially variable e.g. with elevation. In this regard, accounting for preferential flow is a necessary step to improve snowmelt models, as shown in this study.

Acknowledgements

We thank the Swiss Federal Office for the Environment FOEN and the scientific exchange program Sciex-NMSch (project code 14.105) for the funding of the project. Special thanks belong to Jiri Pavlasek for making it possible to conduct the sprinkling experiments, the extensive work and valuable exchange of ideas during the experiments. We also would like to thank Timea Mareková and Pascal Egli for their help during the experiments.

References

Badoux, A., Hofer, M., and Jonas, T.: Hydrometeorologische Analyse des Hochwasserereignisses vom 10. Oktober 2011 (in German), Tech. Rep., WSL/SLF/MeteoSwiss, 92 pp. [Available online at http://www.wsl.ch/fe/gebirgshydrologie/wildbaeche/projekte/unwetter2011/Ereignisanalyse_Hochwasser_Oktober_2011.pdf], 2013.

Bartelt, P., and Lehning, M.: A physical SNOWPACK model for the Swiss avalanche warning Part I: numerical model, Cold Reg. Sci. Technol., 35, 123-145, doi: 10.1016/S0165-232x(02)00074-5, 2002.

Calonne, N., Geindreau, C., Flin, F., Morin, S., Lesaffre, B., Rolland du Roscoat, S., and Charrier, P.: 3-D image-based numerical computations of snow permeability: links to specific surface area, density, and microstructural anisotropy, Cryosphere, 6, 939-951, doi: 10.5194/tc-6-939-2012, 2012.

Colbeck, S. C.: A theory of water percolation in snow, J. Glaciol., 11, 369-385, 1972.

Coléou, C., and Lesaffre, B.: Irreducible water saturation in snow: experimental results in a cold laboratory, Ann. Glaciol., 26, 64-68, 1998.

Conway, H., and Benedict, R.: Infiltration of water into snow, Water Resour. Res., 30, 641-649, doi: 10.1029/93WR03247, 1994.

Denoth, A.: An electronic device for long-term snow wetness recording, Ann. Glaciol., 19, 104-106, 1994.

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.

Dilley, A., and O'brien, D.: Estimating downward clear sky long-wave irradiance at the surface from screen temperature and precipitable water, Q. J. R. Meteorol. Soc., 124, 1391-1401, doi: 10.1002/qj.49712454903, 1998.

Eiriksson, D., Whitson, M., Luce, C. H., Marshall, H. P., Bradford, J., Benner, S. G., Black, T., Hetrick, H., and McNamara, J. P.: An evaluation of the hydrologic relevance of lateral flow in snow at hillslope and catchment scales, Hydrol. Process., 27, 640-654, doi: 10.1002/hyp.9666, 2013.

Freudiger, D., Kohn, I., Stahl, K., and Weiler, M.: Large-scale analysis of changing frequencies of rain-on-snow events with flood-generation potential, Hydrol. Earth Syst. Sci., 18, 2695-2709, doi:10.5194/hess-18-2695-2014, 2014.

Gerdel, R. W.: The transmission of water through snow, Eos Trans. AGU, 35, 475-485, 1954.

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.

Hirashima, H., Yamaguchi, S., and Katsushima, T.: 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, 2014.

Illangasekare, T. H., Walter, R. J., Meier, M. F., and Pfeffer, W. T.: Modeling of meltwater infiltration in subfreezing snow, Water Resour. Res., 26, 1001-1012, 1990.

Jordan, R.: A one-dimensional temperature model for a snow cover: Technical documentation for SNTHERM. 89, DTIC Document, 1991.

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.

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

Kattelmann, R.: Spatial Variability of Snow-Pack Outflow at a Site in Sierra Nevada, U.S.A., Ann. Glaciol., 13, 1989.

Kattelmann, R.: Flooding from rain-on-snow events in the Sierra Nevada, IAHS Publ., 239, 59-66, 1997.

Kroczynski, S.: A comparison of two rain-on-snow events and the subsequent hydrologic responses in three small river basins in Central Pennsylvania, Eastern Region Technical Attachment, 4, 1-21, 2004.

Leathers, D. J., Kluck, D. R., and Kroczynski, S.: The severe flooding event of January 1996 across north-central Pennsylvania, Bull. Am. Meteorol. Soc., 79, 785-797, doi:10.1175/1520-0477(1998)079<0785:TSFEOJ>2.0.CO;2, 1998.

Lehning, M., Bartelt, P., Brown, B., Russi, T., Stockli, U., and Zimmerli, M.: SNOWPACK model calculations for avalanche warning based upon a new network of weather and snow stations, Cold Reg. Sci. Tech., 30, 145-157, doi:10.1016/S0165-232X(99)00022-1, 1999.

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 Reg. Sci. Tech., 35, 169-184, doi: 10.1016/S0165-232x(02)00072-1, 2002a.

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. Tech., 35, 147-167, doi: 10.1016/S0165-232x(02)00073-3, 2002b.

Marks, D., Link, T., Winstral, A., and Garen, D.: Simulating snowmelt processes during rain-on-snow over a semi-arid mountain basin, Ann. Glaciol., 32, 195-202, doi:10.3189/172756401781819751, 2001.

Marsh, P., and Woo, M.-K.: Wetting Front Advance and Freezing of Meltwater Within a Snow Cover 1. Observations in the Canadian Arctic, Water Resour. Res., 20, 1853-1864, 1984.

Marsh, P., and Woo, M. K.: Meltwater movement in natural heterogeneous snow covers, Water Resour. Res., 21, 1710-1716, 1985.

Marsh, P., and Pomeroy, J.: The impact of heterogeneous flow paths on snowmelt runoff chemistry, Proc. East. Snow. Conf, 1993, 231-238, 1993.

Marsh, P., and Pomeroy, J.: Spatial and temporal variations in snowmelt runoff chemistry, Northwest Territories, Canada, Water Resour. Res., 35, 1559-1567, 1999.

McCabe, G. J., Clark, M. P., and Hay, L. E.: Rain-on-Snow Events in the Western United States, Bull. Am. Meteorol. Soc., 88, 319-328, doi:10.1175/bams-88-3-319, 2007.

Merz, R., and Blöschl, G.: A process typology of regional floods, Water Resour. Res., 39, doi:10.1029/2002wr001952, 2003.

Michlmayr, G., Lehning, M., Koboltschnig, G., Holzmann, H., Zappa, M., Mott, R., and Schöner, W.: Application of the Alpine 3D model for glacier mass balance and glacier runoff studies at Goldbergkees, Austria, Hydrol. Process., 22, 3941-3949, doi:10.1002/hyp.7102, 2008.

Molnar, P., Fatichi, S., Gaál, L., Szolgay, J., and Burlando, P.: Storm type effects on super Clausius–Clapeyron scaling of intense rainstorm properties with air temperature, Hydrol. Earth Syst. Sci., 19, 1753-1766, doi: 10.5194/hess-19-1753-2015, 2015.

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

Osterhuber, R.: Precipitation intensity during rain-on-snow, in: Proceedings of the 67th Annual Western Snow Conference, 67th Annual Western Snow Conference, South Lake Tahoe California, 1999, 153–155, 1999.

R Development Core Team: R: A Language and Environment for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing, Retrieved from http://www.R-project.org/, 2016.

Rössler, O., Froidevaux, P., Börst, U., Rickli, R., Martius, O., and Weingartner, R.: Retrospective analysis of a nonforecasted rain-on-snow flood in the Alps-a matter of model limitations or unpredictable nature?, Hydrol. Earth Syst. Sci., 18, 2265-2285, doi:10.5194/hess-18-2265-2014, 2014.

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 Reg. Sci. Technol., 99, 27-37, doi:10.1016/j.coldregions.2013.12.004, 2014.

Schneebeli, M.: Development and stability of preferential flow paths in a layered snowpack, in: Biogeochemistry of Seasonally Snow-Covered Catchments (Proceedings of a Boulder Symposium July 1995), edited by Tonnessen, K., Williams, M., and Tranter, M., 1995, 89-96, 1995.

Singh, P., Spitzbart, G., Hübl, H., and Weinmeister, H.: Hydrological response of snowpack under rain-on-snow events: a field study, J. Hydrol., 202, 1-20, doi:10.1016/S0022-1694(97)00004-8, 1997.

Stearns, C. R., and Weidner, G. A.: Sensible and Latent Heat Flux Estimates in Antarctica, in: Antarctic Meteorology and Climatology: Studies Based on Automatic Weather Stations, edited by: Bromwich, D. H., and Stearns, C. R., Antarctic Research Series, Vol. 61, American Geophysical Union, 109-138, 1993.

Sui, J., and Koehler, G.: Rain-on-snow induced flood events in Southern Germany, J. Hydrol., 252, 205-220, doi:10.1016/S0022-1694(01)00460-7, 2001.

Surfleet, C. G., and Tullos, D.: Variability in effect of climate change on rain-on-snow peak flow events in a temperate climate, J. Hydrol., 479, 24-34, doi:10.1016/j.jhydrol.2012.11.021, 2013.

Unsworth, M. H., and Monteith, J.: Long-wave radiation at the ground I. Angular distribution of incoming radiation, Q. J. R. Meteorol. Soc., 101, 13-24, doi:10.1002/qj.49710142703, 1975.

Waldner, P. A., Schneebeli, M., Schultze-Zimmermann, U., and Flühler, H.: Effect of snow structure on water flow and solute transport, Hydrol. Process., 18, 1271-1290, doi: 10.1002/hyp.1401, 2004.

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, Cryosphere, 8, 257-274, doi:10.5194/tc-8-257-2014, 2014a.

Wever, N., Jonas, T., Fierz, C., and Lehning, M.: Model simulations of the modulating effect of the snow cover in a rain-on-snow event, Hydrol. Earth. Sys. Sci., 18, 4657-4669, doi:10.5194/hess-18-4657-2014, 2014b.

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, Cryosphere, 9, 2271-2293, doi: 10.5194/tc-9-2271-2015, 2015.

Wever, N., Vera Valero, C., and Fierz, C.: Assessing wet snow avalanche activity using detailed physics based snowpack simulations, Geophys. Res. Lett., doi:10.1002/2016GL068428, 2016a.

Wever, N., Würzer, S., Fierz, C., and Lehning, M.: Simulating ice layer formation under the presence of preferential flow in layered snow covers, Cryosphere, doi:10.5194/tc-10-2731-2016, 2016b.

Wickham, H.: ggplot2: Elegant Graphics for Data Analysis, Use R!, Springer-Verlag New York, VIII, 213 pp., 2009.

WSL Institute for Snow and Avalanche Research SLF: Meteorological and snowpack measurements from Weissfluhjoch, Davos, Switzerland, doi:10.16904/1 2015. dataset

Würzer, S., Jonas, T., Wever, N., and Lehning, M.: Influence of initial snowpack properties on runoff formation during rain-on-snow events, J. Hydrometeorol., 17, 1801-1815, doi: 10.1175/JHM-D-15-0181.1, 2016.

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

Ye, H., Yang, D., and Robinson, D.: Winter rain on snow and its association with air temperature in northern Eurasia, Hydrol. Process., 22, 2728-2736, doi:10.1002/hyp.7094, 2008.

Table 1: Snowpack pre-conditions and execution dates for the sprinkling experiments as well as R^2 values for the different model simulations. Measured values are snow height (HS), bulk liquid water content (LWC), bulk snow temperature (TS).

Initial snowpack conditions					R^2 of hourly runoff of the simulations		
Experiment	HS [cm]	LWC [vol%]	TS [°C]	DATE	RE	PF	BA
Serneus (Ex1)	48.5	0.1	-1.3	26-Feb-15	0.~~22~~14	0.~~45~~59	0.09
Davos (Ex2)	54.5	0.4	-2.5	27-Feb-15	0.~~25~~24	0.62	0.08
Sertig (Ex3)	71.5	0	-1.6	28-Feb-15	NA	NA	NA
Klosters (Ex4)	15.7	6.9	0	26-Mar-15	0.~~78~~75	0.96	0.86
Klosters (Ex5)	7	4.9	0	8-Apr-15	0.~~71~~70	0.84	0.88
Davos (Ex6)	39.3	0.9	-0.6	10-Apr-15	0.~~52~~58	0.~~82~~83	0.~~37~~36

Table 2: Snowpack pre-conditions and R^2 for hourly snowpack runoff for natural events Jan 03 + Jan 09

	Site	Pre-event snowpack conditions			R^2 for hourly snowpack runoff		
		HS (cm)	LWC (vol%)	TS (°C)	RE	PF	BA
03.01.2015	**Serneus**	19	0	0	0.~~61~~63	0.35	0.83
	Klosters	24	0	-0.1	0.~~73~~72	0.~~66~~39	0.~~79~~78
	Davos	20	0	-0.4	0.~~28~~27	0.~~30~~33	0.~~13~~17
09.01.205	**Serneus**	14.5	0.1	-0.2	0.94	0.~~56~~57	0.79
	Klosters	18	0.1	-0.2	0.84	0.~~66~~73	0.~~70~~73
	Davos	19.5	0.1	-0.6	0.00	0.04	0.~~01~~00

Table 3: R^2 and mean absolute errors for hourly snowpack runoff for 17 and 14 years, for CDP and WFJ, respectively.

	R^2 hourly snowpack runoff			RMSE of snowpack runoff (mm h^{-1})		
	BA	RE	PF	BA	RE	PF
CDP	0.33	0.50	0.52	0.5756	0.4544	0.4140
WFJ	0.485	0.77	0.785	0.531	0.304	0.3028

Table 4: Correlations between event-R^2 and stratigraphic features at WFJ and CDP. Stratigraphic features are marked grain size changes (bigger 0.5 mm) and density changes (bigger 100 kg m^{-3}) in two adjacent simulated layers as well as the wet layer ratio (percentage of layers exceeding 1 vol% over layers below 1 vol%) and the percentage of melt forms. Bold numbers denote negative correlations, italic values denote positive correlations.

		R^2 between event-R^2 and			
		No. of grain size changes	No. of density changes	ratio of melt forms	wetting ratio
WFJ	PF	**0.19**	**0.20**	**0.03**	**0.04**
	RE	**0.29**	**0.22**	*0.03*	*0.02*
	BA	**0.31**	*0.03*	**0.01**	**0.01**
CDP	PF	**0.02**	*0.00*	*0.14*	*0.16*
	RE	**0.04**	*0.01*	*0.33*	*0.44*
	BA	**0.01**	**0.07**	*0.02*	*0.02*

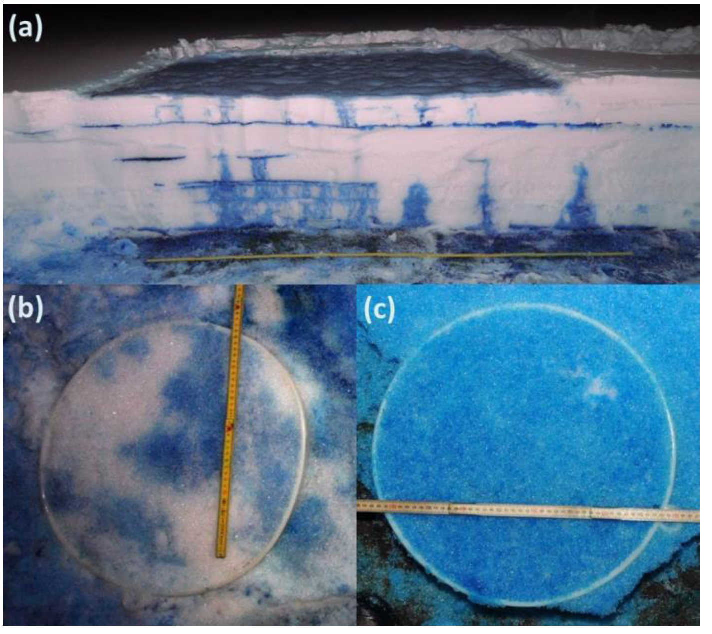
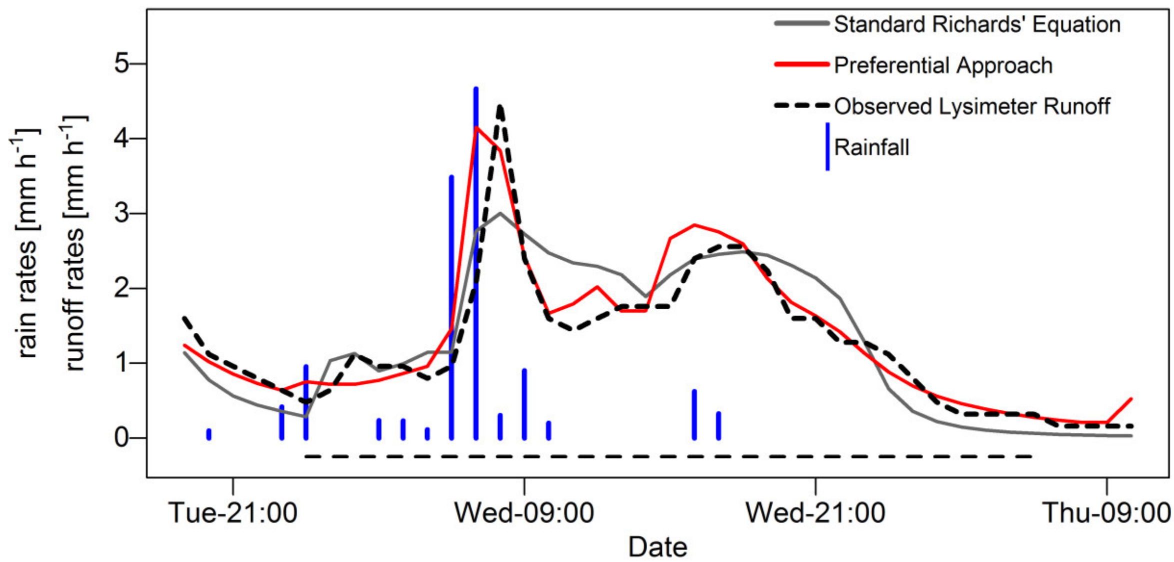


Figure 1: (a) Horizontal cut of a snowpack after the sprinkling experiment Sertig ~~Ex1~~Ex3 (26.02.2015). Lateral flow and the presence of PFP were observed. PFP were generated at regions with rain water ponding at ice layers and layer boundaries with a change in grain size (creating capillary barriers). (b) Lysimeter area after sprinkling during winter conditions (Serneus Ex1, 26.02.2015): Colored areas indicate the area where water percolated due to preferential flow. (c) Lysimeter area after sprinkling during spring conditions (Klosters Ex4, 26.03.2015): Colored area shows that water percolated uniformly, indicating dominating matrix flow.



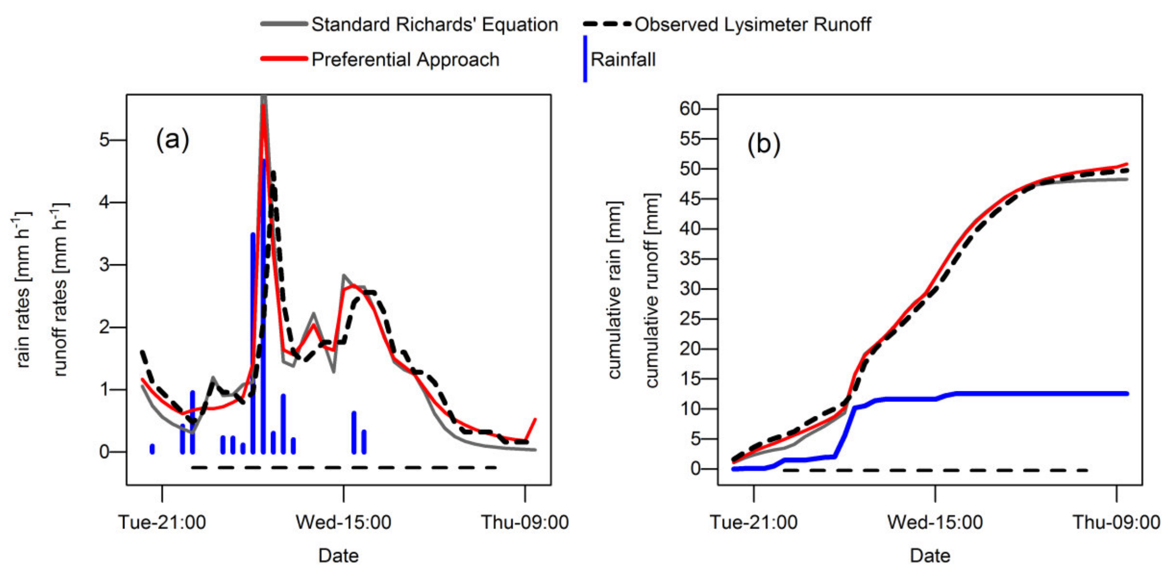


Figure 2: Example of a ROS event occurring at WFJ. The entire extent of the x-axis refers to the evaluation period; the bar ~~below~~ above the x-axis refers to the event length. (b) Cumulative version of the plot.

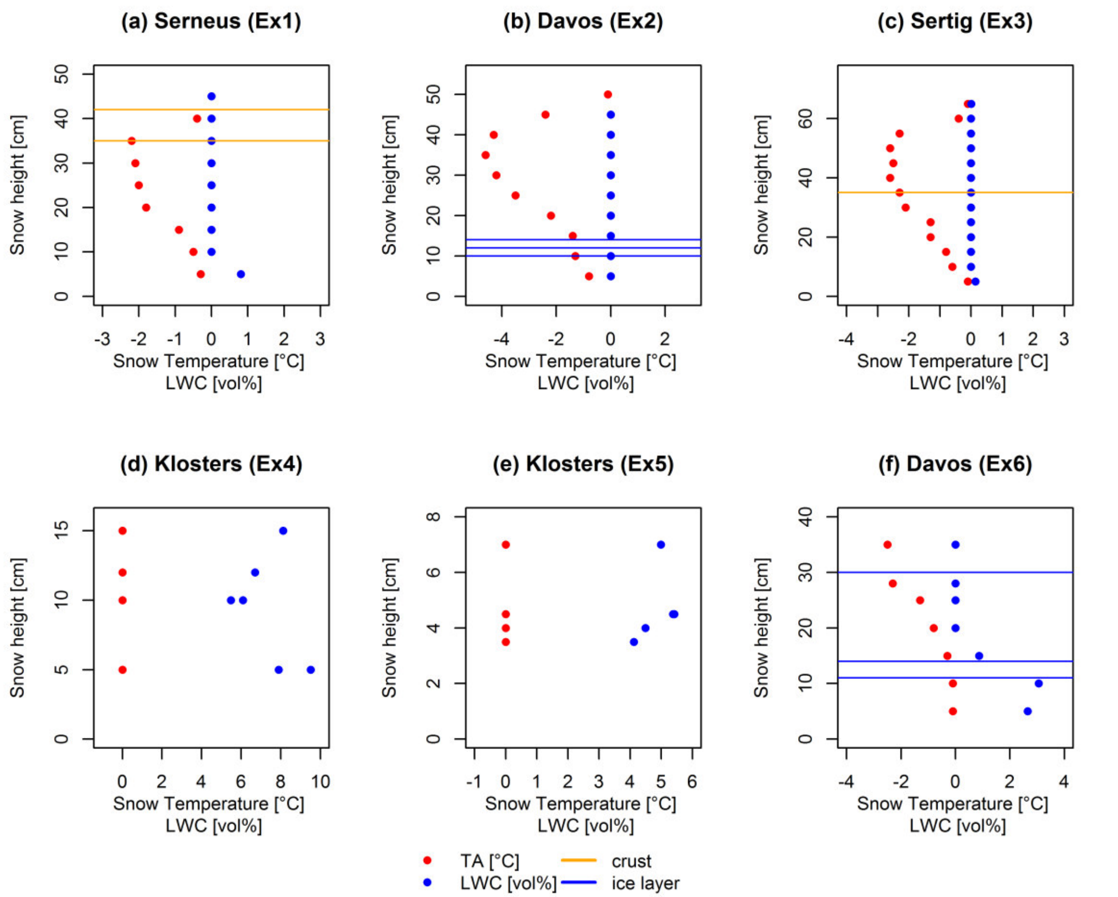
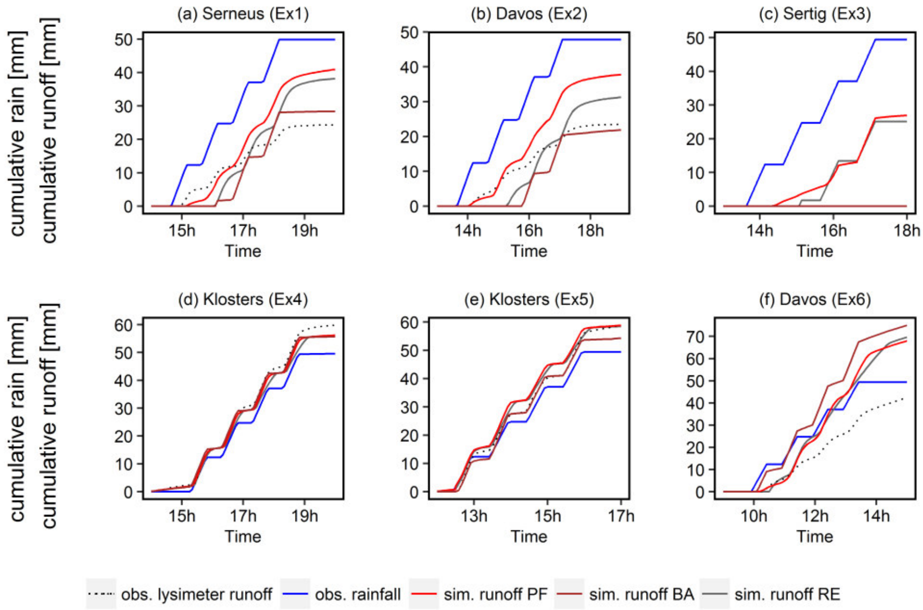


Figure 3: Snow temperature and LWC profiles measured directly before the sprinkling experiment started. The lines represent observed ice layers (blue) and crusts (orange).



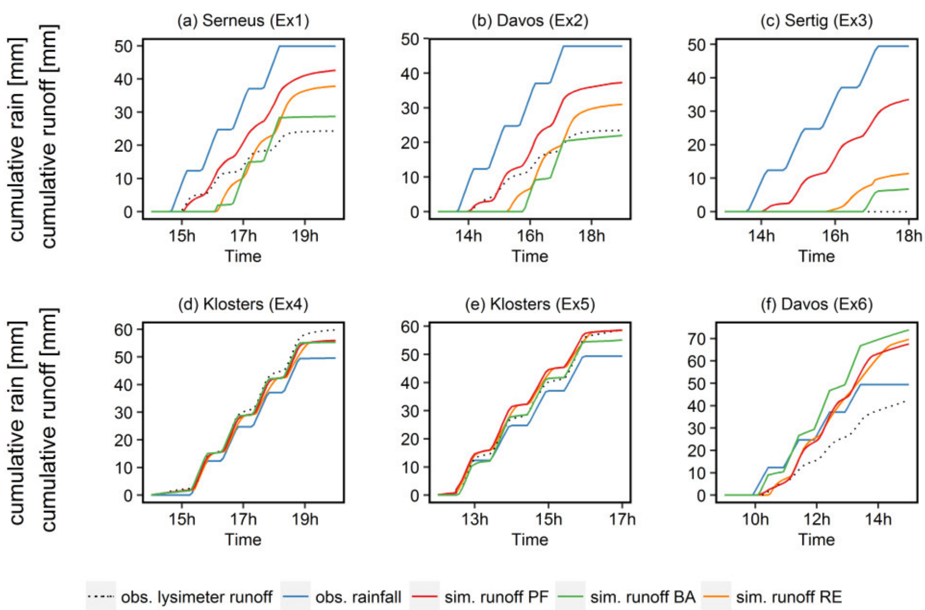
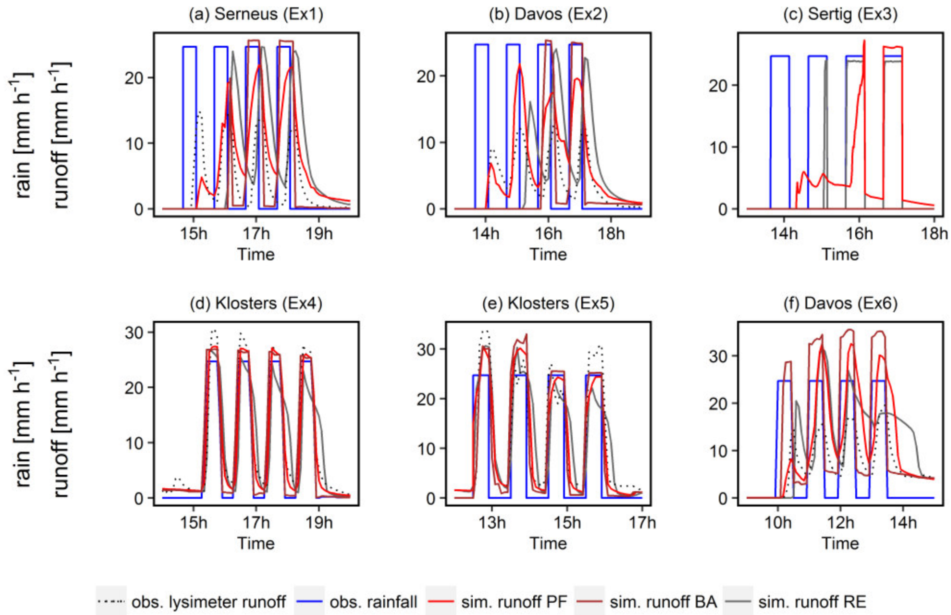


Figure ~~3~~4: Cumulative rain and snowpack runoff displayed for the six sprinkling events. Ex1 (a) - Ex3 (c) were conducted during winter conditions, Ex4 (d) – Ex6 (f) were conducted during spring conditions.



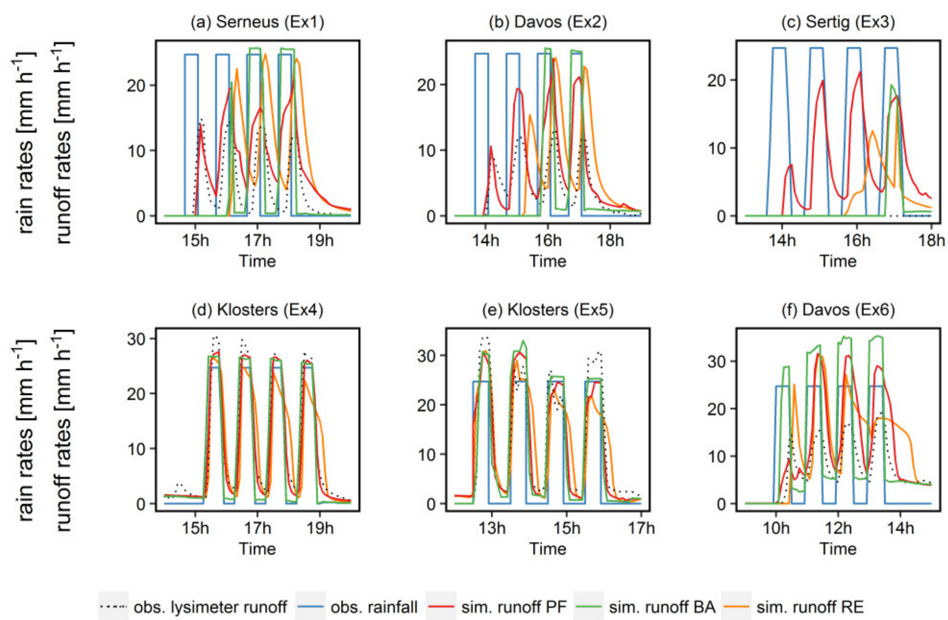
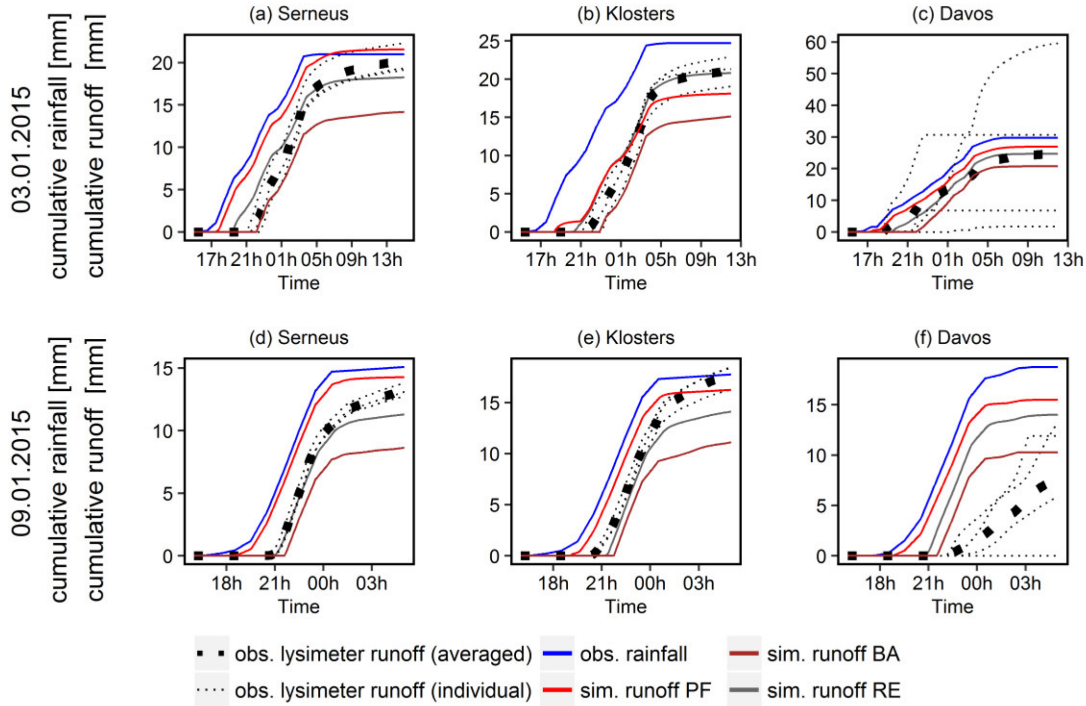


Figure 45: Rain and snowpack runoff displayed as hydrographs for the six sprinkling events. Ex1 (a) - Ex3 (c) were conducted during winter conditions, Ex4 (d) – Ex6 (f) were conducted during spring conditions.



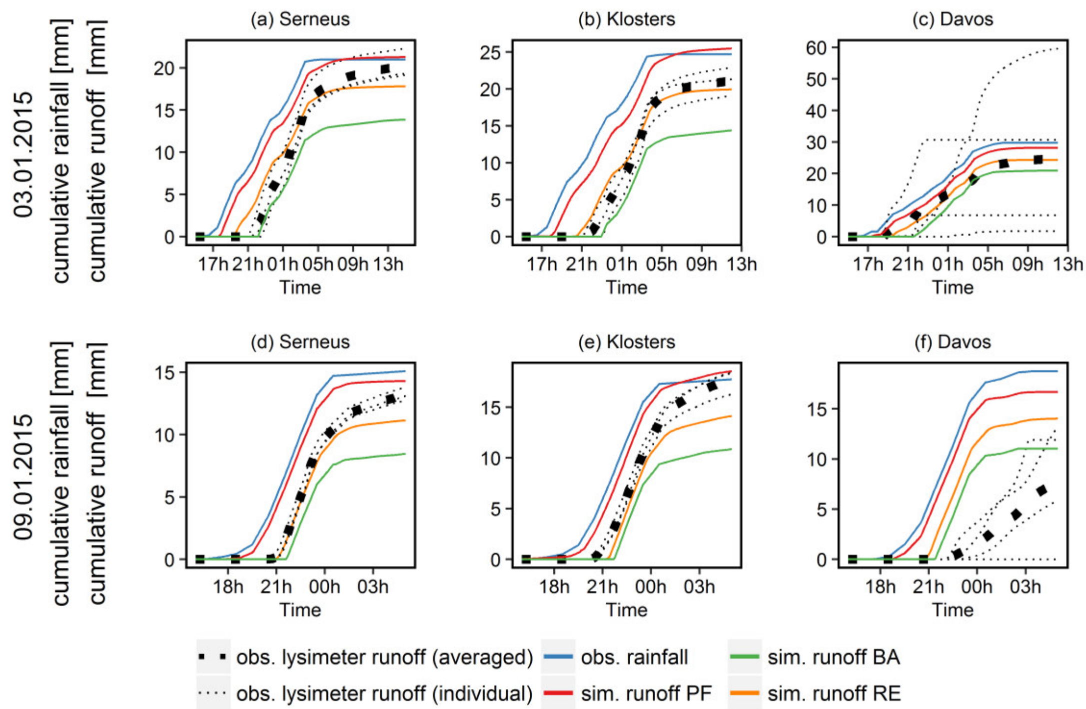
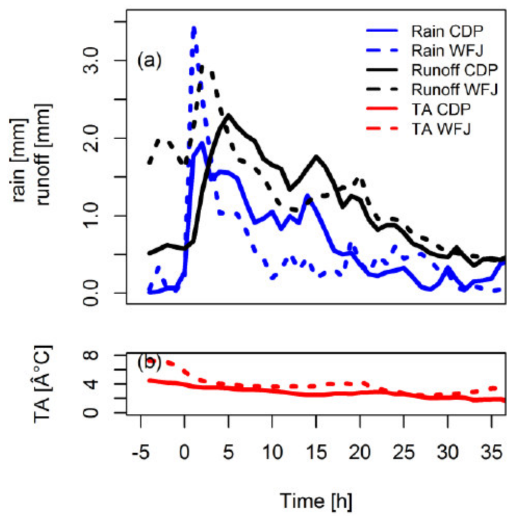


Figure 6~~5~~: Natural ROS events at 3rd and 9th of January 2015 in (a) Serneus, (b) Klosters and (c) Davos



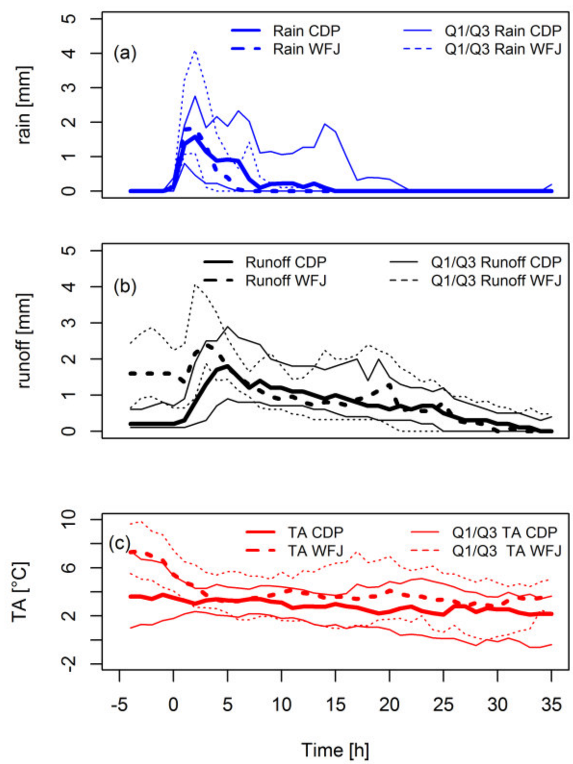


Figure ~~6~~7: (a) Temporal course of median rain (a), measured snowpack runoff (b) and air temperature (c) for WFJ (dotted) and CDP (solid) aggregated over all 40 and 61 events respectively. The thinner lines represent the lower and upper quartiles, respectively. The displayed period is extended by 5 hours prior to event beginning according to the event definition (0 h).~~Course of mean rain (blue) and measured snowpack runoff (red) for WFJ (dotted) and CDP (solid) for all 40 and 61 events respectively. (b) Mean air temperature for WFJ (dotted) and CDP (solid). The displayed period is extended by 5 hours prior to event beginning according to the event definition (0 h).~~

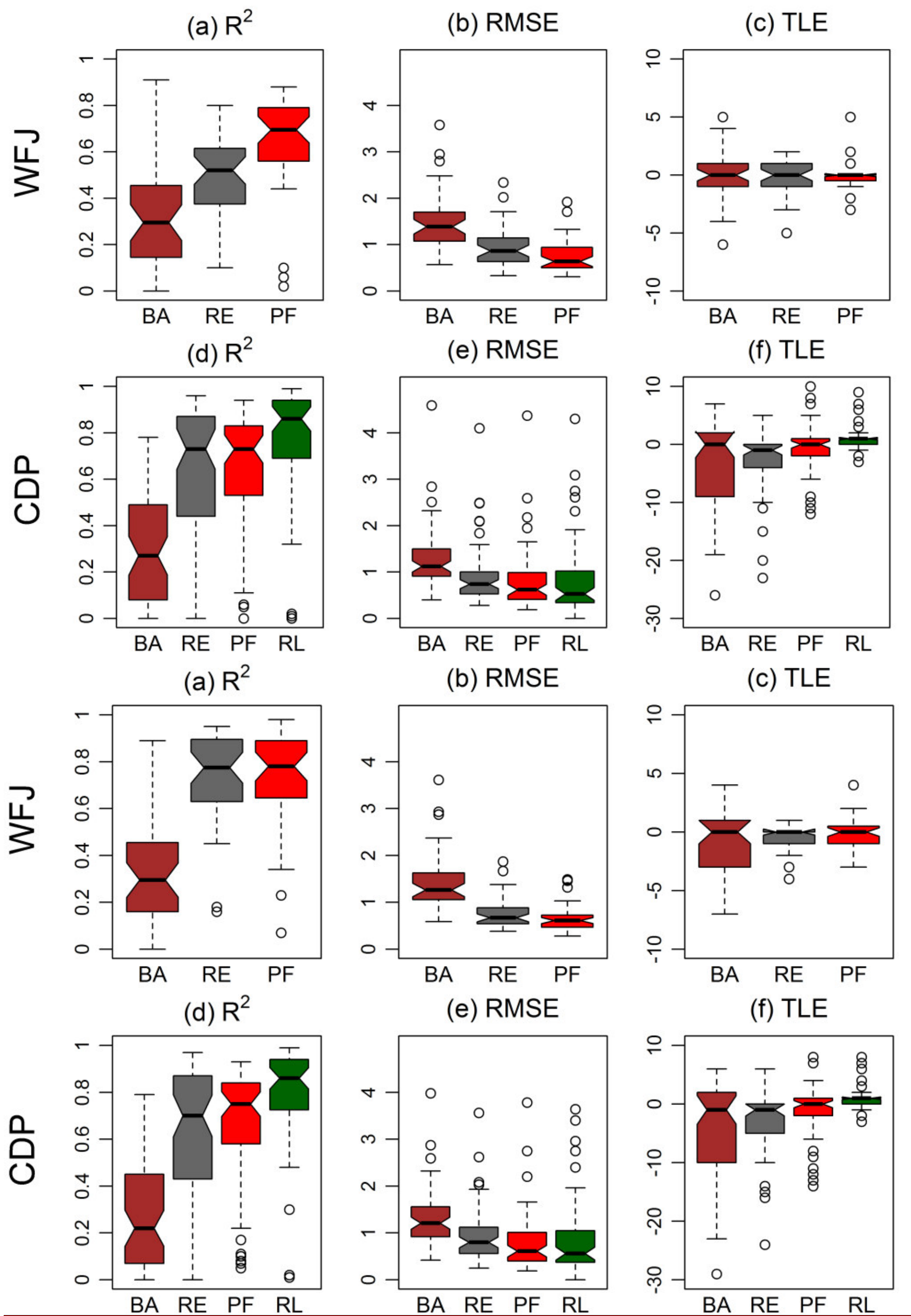


Figure ~~7~~8: RMSE, R^2 and TLE for simulations of 61 ROS events at the CDP site and of 40 ROS events at the WFJ site for all models (BA, RE, PF) and the reference lysimeter (RL) available only for CDP.

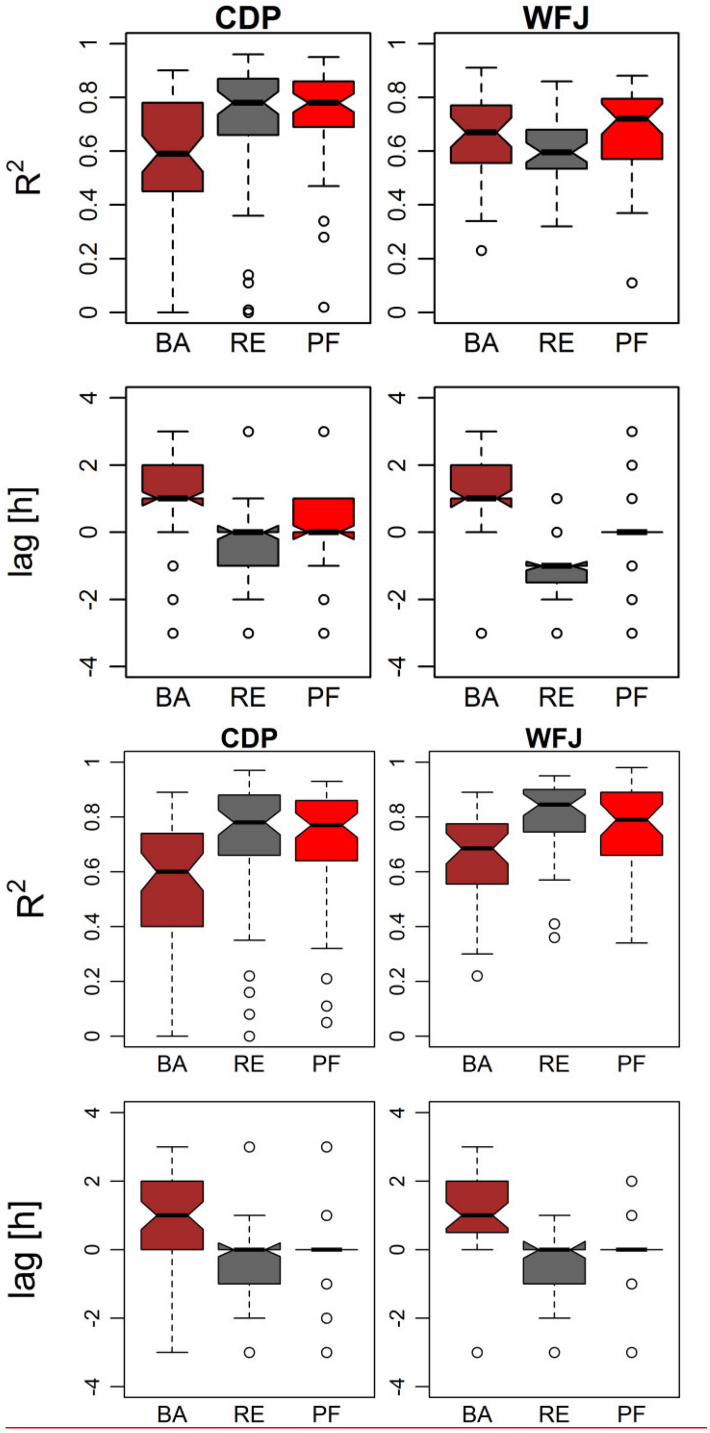


Figure 89: Best R^2 values and corresponding lags using a cross-correlation function allowing a time shift (lag) of max -/+ 3 hours.

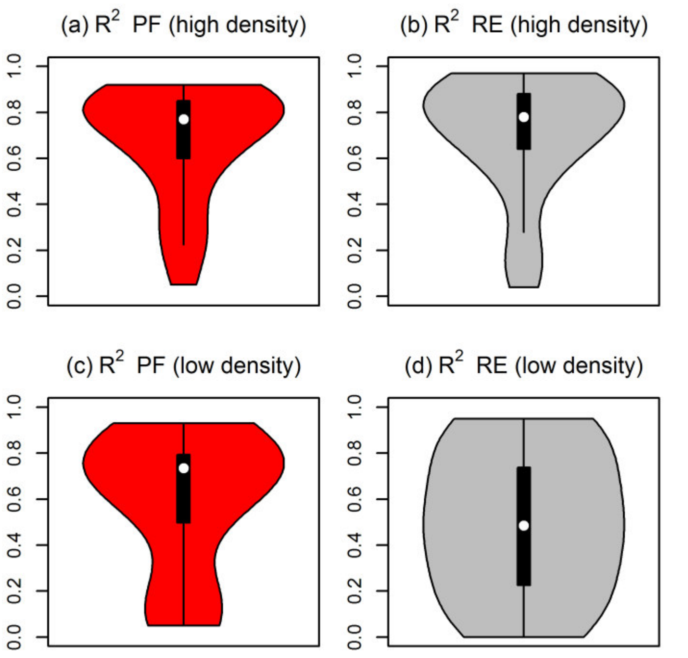


Figure 10: Distribution of event-R^2 for CDP events for the PF (a,c) and RE (b,d) model. The sample is split into initial bulk snow densities above 350 kg m^{-3} (a,b) and below 350 kg m^{-3} (c,d).