Author's reply to Reviewer Comment #1

Black: Reviewer comments – **Blue:** Author replies as posted during interactive discussion – **Green:** Comments added or altered during revision

Overview:

Mazzotti et al. explore the interplay between topographic shading, forest cover, and three-years of varying meteorology on how snow evolves near Davos, Switzerland. The paper is well-written and contributes to understanding of how and why forest-snow process vary between years and different locations. I recommend publication following minor revisions. These fall in two categories. The first, listed under major comments, involves putting the results in a larger context. The second, listed under minor comments, involves improving clarity (by rewriting very long and complicated sentences). The authors are welcome to reach out to me if they have any questions about any of my comments: Jessica Lundquist, University of Washington, jdlund@uw.edu

We would like to thank Jessica Lundquist for the positive assessment of our work and for the constructive input. Please find our detailed replies to major and minor comments below. *Please note that line numbers indicated in our replies refer to the revised manuscript with tracked changes.*

Major comments:

1. This is a very thorough paper with a lot of years of analysis. Given the large focus on different components of the energy balance, I recommend that you comment on your relative confidence in your spatial calculations of SW, LW, turbulent fluxes. You keep saying sensible heat flux is a dominant term, but I think you are less confident about that than about your radiation calculations.

The capability of FSM2 to represent the dynamics of individual energy balance components at high spatio-temporal resolution was thoroughly assessed in Mazzotti et al. 2020 (doi:10.1029/2020WR027572). Therein, we would like to highlight Figure 4 (and similar figures in the Supporting Information) in particular, which present comparisons of modelled sub-canopy short-and longwave radiation, air and snow surface temperatures to data acquired with moving micro-meteorological sensors, at a temporal resolution of 15mins and spatial resolution of approx. 1.5m. Figure 7 additionally reports modelled and measured below-canopy wind speeds.

These micrometeorological measurements allowed direct validation of the measured fluxes and state variables; hence it is correct that we have highest confidence in the modelled spatial variability of sub-canopy incoming radiation fluxes. However, the spatial measurements of air and snow surface temperatures provided insights into the temperature gradients that determine sensible heat exchange (together with wind). Corresponding measurements allowed us to conclude that turbulent fluxes exhibit a less strong spatial variability than radiative fluxes, and the ability of FSM2 to capture snow distribution patterns in heterogeneous forest at different times during the snow season indicates that the order of magnitude of the modelled turbulent fluxes must be in the right order of magnitude (to arrive at realistic melt rates). In lack of a system to measure turbulent exchange with moving sensors, this was our best option to indirectly validate spatial turbulent exchange simulations, and we are not aware of any other such efforts in existing literature. These considerations reported in the discussion of Mazzotti et al. (2020). We now mention this aspect in the discussion (L590ff).

2. How transferable are your results to other regions of the globe? Further south, further north, or warmer or colder? How transferable to other modeling work? The following comments indicate some thoughts on how to more broadly interpret your work here.

Thank you for the below detailed remarks and inputs, which we reply to individually. Transferability of our results is addressed in the discussion (Section 4.2), where we state that the observed patterns may not transfer to other climates, but our process insights do. Limitations occur where the processes driving accumulation and melt are substantially different. These paragraphs have been extended to accommodate considerations based on your comments below.

The following are details related to overarching question number 2:

* Obviously, I'm biased by being me, but I'd be really curious to see you put your results in context of Figures 6, 7, and 8 from Lundquist et al. 2013. I was trying to get at what you've done here with a much simpler model, and I'm curious how much of your results confirm what the simpler model shows, and how much your results show that a more complicated model really is needed. (Motivation: we have a lot of forest and water managers in the U.S. who would like to use simpler models if possible — what are the trade-offs.)

The results from our study confirm the conceptual findings and process-level insights from Lundquist et al. (2013) but derive them based on 'real' heterogeneous forest structures and meteorological conditions. Consequently, our study is more concerned with small-scale variability at one site rather than differences between sites with varying climatic characteristics. We have incorporated these considerations in the discussion (L512ff and 566).

In view of forest management applications, the simpler model used by Lundquist et al. is a powerful tool to derive 'rule of thumbs' (e.g., does forest thinning accelerate or delay snowmelt?), while our approach would allow to draw conclusions on more specific management measures applied to an existing stand with its unique characteristics (e.g. if the impact of a specific forest thinning strategy is to be quantitatively assessed). The 'need' for a more complex model thus depends on the specific application; availability of data (both meteorological forcing and canopy structure information) is another factor affecting model choice and may motivate the choice of a simpler model over a more complex one. Forest management is now mentioned explicitly in L570.

* You say that you have to consider all three factors (canopy structure, topography, and meteorology), but could you consider topography as a local modification to meteorology? At your study site, does it do anything other than alter your solar exposure? For example, you state that Currier et al. 2022 could be further refined by accounting for topography, but the DHSVM model used in that paper does adjust all of the local meteorological inputs based on topography (including shading by surrounding slopes). (I'm less familiar with Broxton's paper, but I'm pretty sure the topographic effects are included in Currier et al. 2022.)

Yes, considering topography as a local modification to meteorology is a possible point of view. However, treating them as two separate factors allowed us to discuss two aspects:

1. Between-year variability caused by variability in meteorological forcing (while topography is constant), see Section 3.5.
2. The fact that the main topographic impact is on solar exposure leading to more complex dependencies between canopy structure and snow distribution patterns on the south- than on the north-exposed slope, see Section 3.4 (as an effect of the superposition of time-invariant factors only).

Regarding the adjustments of meteorological inputs to account for topographic effects, see our reply to your final comment in this section.

* You definitely don't need to add another citation to your paper, but you might be interested to look at Lundquist and Flint 2006 (https://journals.ametsoc.org/view/journals/hydr/7/6/jhm539_1.xml) which talks about topographic shading being more important when melt occurs earlier in the year. Look at Figures 12-15, which talk about the interplay of warming, latitude, and terrain on net radiation. Your multi-year

analysis takes this basic concept and expands to the impact of forest cover. I think that looking at these conceptual figures and discussion might help you put your results in a larger context.

Thank you for this input, indeed the conceptual figures and discussion are in line with our reasoning. This reference is particularly interesting in the context of Section 3.5 where we show how the response of the snowpack to variability in meteorological conditions may differ between points. We have added this reference to our discussion of snow regimes (L511), as it underpins the findings that different regimes occur on different slopes, and different years might result in different regimes at the same location.

* I'd also like to see a bit of discussion about the scale at which topography would become more important than canopy structure in controlling accumulation. For example, if you have a north-facing slope in a rain shadow, and an adjacent south-facing slope that experienced orographic enhancement (or vice versa). You have a unique situation in that your slopes experience the same precipitation, but differences have been observed even at very fine scales (e.g., see Minder et al. 2008: and Anders et al. 2007).

This is an interesting point. It is correct that differences in precipitation patterns that are strong enough to override heterogeneity caused by canopy structure can occur at the hillslope scale already. In such a situation, the comparison of points that 'feature the same canopy structure but are in different topographic settings' (Section 3.3 / Figure 6) would show accumulation differences, despite a similar impact of interception. Likewise, elevation controls on snow distribution likely override canopy structure effects if the domain considered spans a sufficiently large range of elevations. This has been reported in existing work (e.g. Tennant et al. 2017, doi:10.1002/2016WR019374).

However, the main purpose of our analysis is not to rank the importance of the factors topography, canopy structure, and meteorology in governing accumulation and ablation processes in absolute terms, but to assess specific combinations of these factors and the snow distribution patterns they create. This is particularly interesting in the context of sub-grid variability at sub-kilometric model resolutions, which are now more and more common in watershed- and regional-scale applications (e.g. the Swiss operational snow-hydrological model operates at a 250m resolution, Broxton et al. 2021, doi: 10.1029/2021WR029716 considered 100m pixels). In such situations, topography-driven variability in meteorological variables would be captured by the driving data fields in the model, while the canopy-structure-driven variability would need to be parametrized (see Clark et al. 2011 for an extensive discussion on these different process scales). Based on our results, different topographic settings featuring the same canopy characteristics may require different sub-grid variability parametrizations; but it is correct that this conclusion depends on the model resolution the parametrization is intended for. We have added a note on scales in Section 4.2, where the relevance of our finding is discussed in the context of sub-grid variability parametrizations (L533ff).

Minor comments:

There a large number of very long and complicated sentences. Some shortening for clarity would be helpful.

We have revised the language and shortened long sentences where appropriate. Specifically, we have adapted all the sentences marked in the annotated PDF (see below).

For example, this line around line 40 is super awkward to read "In view of ongoing changes in both, snow cover regimes due to increasing temperatures (Mote et al., 2018; Marty et al., 2017; Notarnicola, 2020; Bormann et al., 2018), and forest structure following manmade and natural disturbances (Bebi et al., 2017; Seidl et al., 2017; Goeking and Tarboton, 2020), it is also urgent and pertinent to adequate forest and water resources management strategies - particularly in regions where downstream water supply is dependent on snow resources from forested headwaters"

We have adapted and simplified this paragraph (L37ff.)

Not sure what you mean by this sentence: “However, it remains unclear whether landscape heterogeneity entails a variable response of snow cover dynamics to environmental change.” - line 42

We have reworded this sentence for better clarity (L44ff),

I really like Figure 3.

Thank you, we are glad you do!

Line 247: Regarding higher peak SWE not implying longer snow duration — note that these two are strongly linked across most of the western U.S., but these are also in snow packs with greater than 1 meter of SWE (and hence longer total melt out periods). With regards to greater context above, it might be nice to add some discussion to how different regions may or may not see similar results to what you show here.

This is a good point. In Section 4.1, we discuss the role of the relative strength of snow accumulation and ablation patterns in determining what snow cover dynamic regime exists at a specific location, and indeed the strength of these patterns can be related to local climate (especially precipitation amount and latitude/timing of radiation). Specific climate characteristics may favour or prevent the existence of any snow cover dynamic regime. We discuss transferability to other climates and provide examples in Section 4.2; snowfall is now mentioned as a factor impacting transferability / the occurrence of specific regimes (L563).

Additional sentences that would benefit from clarification are noted on the attached annotated PDF.

The corrections suggested in the annotated pdf have been implemented, and the highlighted sentences have been shorted and simplified. See the following sections in the manuscript: L38ff; L130ff; L216; L327ff; L347ff; L389.

Further notes to your comments in Section 2.4:

1. We have changed instances of ‘cumulative melt’ to ‘cumulative ablation’ throughout the manuscript to specify that both sublimation and melt are included, and we have revisited the use of ablation and melt.
2. Regarding the use of a 10mm threshold to define accumulation and ablation periods: yes, we use this threshold as if <10mm was ‘no snow’, hence it also affects the start of snow and snow disappearance days. The motivation behind this choice is now specified in the manuscript (L218ff). Omitting days that only featured snow throughout part of the day was relevant for 1) ablation rate calculations (to ensure that their calculation is only based on days that have enough snow left to ablate), and 2) calculations of energy fluxes (to ensure that these are only impacted by snow surface properties and not confounded by exposed soil).

Author's reply to Reviewer Comment #2

Black: Reviewer comments – **Blue:** Author replies as posted during interactive discussion – **Green:** Comments added or altered during revision.

Overview:

In the paper, the authors mainly presented an application of a new snow cover model FSM2 in a sub-alpine forest, and tried to understand the snow spatio-temporal distribution and its drivers mostly with the modeled results. Given still lacks wonderful snow models that can perform the complex snow dynamics for the mountainous terrain, especially with high resolution simulations, the topic in this paper is very interesting and important. The authors integrated the datasets of four observed snow depth data, Lidar-based canopy structure and elevation with the 2-m grid cell simulated results for 6 winters, and examined the snow energy balance affected by canopy cover and tograph. The paper fits with the scope of HESS and the results are overall well presented, however, the paper missed a lot of information on the study site, and the methods and modeling approaches are not clear. I have some following issues for the authors to consider.

Thank you for the overall positive assessment of our work. Please find our detailed replies to major and minor comments below. We are confident that these address your main concerns, and we thank you for pointing out any lack of clarity. We would like to note that this study aims at a scientific application of a model that has been introduced and discussed in previous work by the authors, especially Mazzotti et al. 2020a,b (doi:10.1029/2019WR026129, doi:10.1029/2020WR027572). While we have added missing methodological information where appropriate, we believe that the manuscript is already too long to accommodate detailed descriptions of the modelling approaches, and that directing reviewers to the previous two publications is sufficient. *Please note that line numbers indicated in our replies refer to the revised manuscript with tracked changes.*

Major comments:

1. Canopy structure:

Canopy structure refers to horizontal canopy cover and vertical canopy distribution (such as vertical layers, leaf area index LAI). Here in the study, the authors only focused on the canopy cover fraction, but did not pay a lot attention to other structure factors, such as canopy height and LAI. To avoid the confusion, I recommend the authors use 'canopy cover' to replace the word 'Canopy structure'. Did the authors find the impacts of canopy height and LAI on the snow interception? It is better to provide that information in the revision. I am not familiar with the sub-alpine forests. If you can provide more site information on the tree species (coniferous or deciduous species?) and forest structure (such as canopy layer and understory structure), it will help the readers better understand the background of the canopy cover and LAI. In addition, the simulation of snow dynamics in forests mainly depends on the radiation transfer model of canopy. In the paper it seems that the canopy radiation transfer model used is a single layer canopy model. Since the canopy height can be 5-35m mentioned in your supplement, the single-layer hypothesis will probably cause the model bias, especially for the understory snow dynamics. The authors should clarify those uncertainty and limitations.

The reasoning behind the choice to use canopy cover fraction when quantifying relationships between local canopy and snow descriptor metrics is explained in the methods (L221ff., Section 2.4). As we focus on the combined impact of canopy, topography, and meteorological conditions, one metric to quantify canopy was sufficient for our purpose. Detailed analysis of the relationships between individual processes, snow distribution, and sophisticated canopy descriptors, has been subject of earlier studies (e.g., Moeser et al. 2015, doi:10.1002/2014WR016724, for interception; Lawler and Link 2011, doi:hyp.8150, for sub-canopy incoming radiation; Hojatimalekshah et al. 2021, doi:10.1002/hyp.8150, for snow distribution) and is beyond the scope of ours. However, note that FSM2 does include more canopy descriptors than just canopy cover fraction (see Section 2.2). Therefore, the use of ‘canopy structure’ instead of ‘canopy cover’ is necessary due to the context of this study, where ‘canopy cover’ is a specific variable, while ‘canopy structure’ as a term to describe the general static state of the canopy/trees/forest that is represented by the different canopy descriptor variables (or ‘canopy structure metrics’). Thank you for pointing out this source of confusion. We have carefully reconsidered all instances of these expressions in the revision to ensure that the correct one is used. Site information (e.g. tree type and heights) is included in Section 2.1, and we have added information on the understory (L116). Finally, as you correctly note, the canopy in FSM2 is represented with one layer, but heterogeneities in canopy height are captured by including canopy structure metrics that integrate the hemispherical perspective at each modelled point, which accurately represent the three-dimensional features. This is now pointed out in Section 2.2 (L131). The calculation of diffuse and time-varying direct transmissivities for shortwave radiation both rely on this hemispherical perspective and are thus much more detailed than if a homogeneous ‘bulk’ canopy layer was used (e.g. as characterized by a site-averaged LAI). A detailed discussion of this radiation transfer modelling approach and its accuracy is provided in Jonas et al. 2020 (doi:10.1016/j.agrformet.2020.107903), and the LiDAR-based version is presented in Webster et al. 2020 (doi:10.1016/j.rse.2020.112017). The integration of hemispherical-image-based radiation transfer model and FSM2 and the associated model improvements achieved with this approach is subject of Mazzotti et al. (2020a). The reader is referred to this publication for further detail, where a thorough discussion of the accuracy of individual process representations has already been carried out and is therefore beyond the scope of this present study of the application of the model. Please note that the most relevant model limitations are addressed in the discussion (Section 4.3).

2. Wind:

In this paper, the authors mostly considered the meteorological factors of air temperature and solar radiation. However, for mountainous terrain, wind is widely recognized as one of the dominant controls of snow accumulation and distribution, including the alpine regions. Wind patterns interacting with topography determine the patterns of final snow distribution on the landscape, while the surface wind and turbulence also control the snow sublimation and melting processes via energy exchange. Unfortunately, I didn't find any relevant results and discussions in this paper. The authors should provide the above corresponding content in their revision.

Topographic influences on wind and precipitation fields are (at least partially) accounted for by the meteorological input fields, as described in Section 2.2, (L160ff). The effects of fine-scale topography (e.g. ridges) are not captured, and FSM2 does not (yet) include wind-driven snow redistribution processes. While snow redistribution is largest close to ridges (e.g. Mott et al. 2018, doi: 10.3389/feart.2018.00197), the tree line is often at lower elevations in alpine valleys. At our study sites, within-forest wind speeds are generally low and canopy structure controls on snow distribution outweigh signals of re-distribution of snow within the canopy by wind (see e.g. Mazzotti et al. 2020b, Koutantou et al. 2022, doi: 10.1016/j.coldregions.2022.103587). Impacts of wind-induced snow redistribution have been reported in some forest configurations in the Western U.S. (e.g. Currier and Lundquist 2018, doi:10.1029/2018WR022553, Dickerson-Lange et al. 2021, doi: 10.1029/2020WR027926). We now explicitly mention snow redistribution by wind as a subject of future work and mention ongoing efforts to couple FSM2 to 1. a high-resolution atmospheric model specifically suited for the simulation of small-scale wind and precipitation patterns in Alpine terrain (an upgrade of Gutman et al. 2016, doi:10.1175/JHM-D-15-0155.1) and 2. a snow redistribution scheme (Liston et al. 1998, doi: 10.3189/s0022143000002021), see L596ff. Corresponding publications documenting these developments are currently in preparation.

The impact of wind on turbulent exchange (including sublimation) within the canopy is included in FSM2. Details on the relevant parametrizations (which are based on commonly used flux-gradient relationships) are provided in the publications documenting FSM and FSM2 (Essery 2015 doi:10.5194/gmd-8-3867-2015, Mazzotti et al. 2020a,b). Mazzotti et al. 2020b showed that within-forest wind speeds are less dependent on local canopy structure than radiative transfer processes, therefore spatial patterns of sub-canopy snow sublimation are assumed to have a relatively small impact on snow distribution patterns. Note that sublimation is included in our calculations of cumulative ablation amounts and rates (Section 2.4).

3. Topography:

The authors used elevation, slope direction to represent the impacts of topography on snow dynamics. This is generally sound, however, in very steep terrain in particular, topography mostly can affect the evolution of snow water equivalent SWE, and thus patterns of meltwater generation. How did the author consider the gravitational snow redistribution during winter? I recommend the authors also provide the site information on slope angle to help the readers understand this.

This is a good point, thank you for addressing it. We have added maps of aspect and slope to the Supplement S2 and corresponding references in the main article. The modelling framework accounts for slope angle for the calculation of incoming solar radiation, but gravitational snow redistribution is not included. Note that very steep slopes are rare in our study domain, which limits the impact of omitting this process in our simulations. As in the case of wind-induced snow transport, established models do exist to represent these processes (e.g., Bernhardt and Schulz 2010, doi:10.1029/2010GL043086), efforts to couple such a model to FSM2 are ongoing at SLF and a corresponding publication is in preparation. We have commented on these limitations and developments in the discussion (L594ff).

Specific comments:

Methods

1. What's the time step or temporal resolution of your simulation? How did you intercept the data for your periodic results?

The FSM2 model runs at 1h time steps, which is now specified (L137). Temporal aggregation of the output can be specified by the user and was set to daily in this study.

2. How did you define the water year?

A water year (also referred to as 'hydrological year') is defined as the 12-month period between September 1st of any year and August 31st of the following year (e.g. water year 2022 started on Sept 1st 2021 and ended on Aug 31st 2022, see Wikipedia). The term is commonly used in hydrology, and we found it more convenient to refer to our simulations than a calendar year, as Northern-Hemisphere snow seasons (and therefore our simulations) usually begin at the end of a calendar year and continue through the first months of the next.

3. How about the uncertainties caused from the input datasets in your study?

As for every modelling study, uncertainty in the input data is unavoidable and doubtlessly affects model accuracy. Yet, we use the best available meteorological forcing and include downscaling methods that were specifically developed for complex terrain. The meteorological input fields can thus be regarded to be realistic, and spatially explicit simulation outputs are consistent with this meteorology. Model errors arising from uncertainty of the meteorological inputs are not relevant for this study because we do not aim at assessing model accuracy but rather at analyzing spatio-temporal patterns in the results, which generate from the modelled impact of canopy and topography. A discussion of input data uncertainties is thus beyond the scope of this study, but we now mention the issue as well as relevant literature (L193ff.)

4. Section 2.4, Regarding the SWE and Snow cover, were they the total of canopy interception and through canopy snowfall? Did you distinguish opening and canopy areas when identifying the SWE peaks and changing periods?

SWE and snow cover refer to snow on the ground, i.e. under the canopy, which is now specified in Section 2.4 (L197). In FSM2, as in any process-based mass- and energy-balance forest snow model, it is the result of snowfall, minus interception, plus unloading of snow previously intercepted in the canopy (c.f. Essery et al. 2009 doi: 10.1175/2009BAMS2629.1 for a general overview of forest snow processes included in such models, and Mazzotti et al. 2020b for the description of the parametrizations used in FSM2). Peak SWE was evaluated at each modelled point individually as maximum value achieved in a water year. The same algorithm to extract this value from the simulation results was applied for openings and canopy areas. Note however, that timing of peak SWE differed depending on a point's location and canopy cover (in fact, peak SWE timing is one of the variables we analyzed, see Section 2.4, L199ff.). Due to these within-site differences in peak SWE timing, also the timing of accumulation and ablation periods differed between points - but they were calculated in the same way regardless of whether a point was located below the canopy or in the open. We hope that these explanations answer your question, as we are not sure what you refer to by 'changing periods'.

Results and Discussions:

l159: What is the data collected frequency?

We now specify that snow depth data at stations is collected daily (L170). Planet imagery is available at approx. weekly intervals (see L172). Airplane and UAV borne LiDAR acquisitions occurred campaign-wise, i.e. at just a few points in time. We refer the reader to the respective publications for further detail on these campaigns (L178f.).

L159-165: Which place does the snow depth data represent for forests? On the top of the canopy or for the whole ecosystem?

If not specified otherwise, snow depth refers to the depth/height of snow on the ground (i.e. under the canopy or in canopy gaps). This is the general case for LiDAR derived snow depth datasets and thoroughly described in the referenced publications, which is why we did not explicitly mention it. It is now specified (L174).

L163, What does ASO stand for?

ASO is the Airborne Snow Observatory, the definition has been added (L146).

L263, Please provide the supporting materials for the statement on under canopy vs open vs gap areas.

This statement refers to the data shown in Figure 5b, expanding on some of the detailed peak SWE distribution patterns shown. We have added a reference to the figure to make this clearer (L282).

L268, Please provide the supporting materials for the statement on correlation.

The correlation coefficient was calculated based on peak SWE and canopy cover data over the entire modelled domain, which is now specified (L287). Figure 5 shows a subsection of these data, the full domain is shown in the Supplement (Figure S 3.2). We have added full-domain data of the snow cover descriptors for all modelled years to the Supplementary Material (S5) for completeness.

L308-309, Please clarify how did you conduct the canopy structure and topographic settings specifically?

As described in L317ff, these points are chosen as examples that serve to illustrate a systematic comparison of snow cover evolution and underlying energy exchange processes at contrasting canopy structure and topographic locations. We thus selected a set of points that covered the full range of canopy covers and included both north- and south-exposed locations, as well as vicinity to north-and south-facing forest edges (selected manually without quantitative criteria). The selection of these specific points rather than another set of points with the same set of characteristics is arbitrary. We now specify that this selection occurred manually in L316.

L406, How did you define 'semi-open conditions'?

This formulation is meant to be descriptive rather than exact and aims at contrasting the examples shown in Figure 6 to those discussed in this section and Figure 10. The former represent the end members of the canopy and topography ranges (see reply to the previous comment), hence their comparison would reveal rather consistent snow cover and energy partitioning pathways throughout the simulated years, while the latter can exhibit between-year differences (illustrated in Figure 10). Points chosen for this analysis were located in sparse canopy and relatively small gaps, as described in 433ff.

L410, What is the exact range of canopy cover?

The points shown in Figure 10 feature fractional canopy covers of fveg = 0.25-0.75, and sky-view fractions of 0.2-0.5.

L427, Here the citation of fig 10a may be not correct.

The reference to Figure 10a is correct. The sentence serves as introductory sentence to the following ones where the difference in melt out shown by the different lines in different years is explained.

L505-506, Snow water input is also affected by the slope angle in the mountain regions.

This is correct. As mentioned in reply to an earlier comment, we now specify the lack of consideration of snow redistribution in steep slopes angle as a limitation of this study and subject of future work (L594ff). We also mention that lateral exchange of mass between neighboring grid cells is neglected but would potentially improve simulations (L598ff). However, note that the recent study by Webb et al. (2019, DOI: 10.1002/hyp.13686) found only limited impact of lateral flow of liquid water through the snowpack below the tree line.

Figures and tables:

All figures: Please point out if the data is from observations or simulations.

Besides Figure 3, all results Figures only show simulations; See Section 2.4 (Analysis approach, L196). This is now also highlighted in the introductory paragraph of the results chapter (3, L228ff).

Fig 1, What does SLF stand for?

The SLF is the Snow and Avalanche Research Institute in Davos, Switzerland (Institut für Schnee und LawinenForschung), this is specified in L121. Note that the acronym is no longer used in Figure 1.

Fig 2, What do SSD,PSD,SDD stand for?

The acronyms are defined in Section 2.4, but we have added the definitions to the caption of Figure 2 to ensure it is stand-alone. Thank you for pointing this out.

Fig 3, It will be better to add an ANOVA test for fig3e. Why didn't you also show the south-exposed slope for Fig3e ? In fig3f, what are the differences between left and right panels?

As mentioned in the text (L157ff), the purpose of Figure 3 is to provide a model plausibility check that confirms what previous publications (Mazzotti et al. 2020a,b) have already shown: that FSM2 is an adequate tool to explore forest snow process variability and resulting snow distribution patterns. Figure 3 illustrates and summarizes this plausibility with examples including a variety of observational datasets, which allowed us to verify several aspects of model 'performance' (spatial patterns and temporal evolution). The figure is not meant to report the exhaustive evaluation that was done, and more subfigures would detract from its overall message, which is why we included only one of the slopes surveyed with the UAV LiDAR.

The omission of quantitative goodness-of-fit measures (such as an ANOVA) is conscious. Discrepancies between the model results and the data do exist and generate from a variety of factors: 1. uncertainty in meteorological forcing; 2. uncertainties in the data (e.g. influences of ground roughness on measured snow depth distributions, inaccuracies of the LiDAR datasets), and 3. uncertainties in the model itself. All these would impact the goodness-of-fit measures, but the first two are irrelevant for the design of this study and the use of FSM2 for its purpose. As noted in reply to one of your earlier comments, discussion of simulation error sources is beyond the scope of this study. To ensure that the manuscript storyline is streamlined and serves its scientific goal, we have therefore not included goodness-of-fit measures in the revision and explain our reasoning behind this approach in L193ff.

Fig 6a, The last two legends have the same titles.

The last two legend entries were the same on purpose, because both points are located under canopy and in a south slope, but we understand that this might have been confusing. We have changed this to two different entries, (canopy – S slope 1 vs. canopy – S slope 2) and outline the difference between them in the text (L321ff).

Fig 8, What do SWR, LWR, SHF, and LHF represent?

The acronyms are now defined in the figure caption, and the individual energy balance components are mentioned in the text (L392).

Fig 10b-d, I didn't find any legends.

Legends have been added.

Fig11a, Does it refer to the total melt depth or melt rate.

We have specified that this is the total amount of ablation in mm water equivalent in the figure caption and the text (L489).