In the following we give the **point-by-point response to the Referee comments** structured as follows:

Our **point-by-point response to the Referee comments** is structured in the following sequence: (1) comments from referees, (2) our response, (3) our changes in the manuscript.

Response to reviewer #1

The paper, titled "openAMUNDSEN v 0.8.3, an open-source snow-hydrological model for mountain regions", describes a fully distributed snow-hydrological model optimized for mountain regions. The model provides a detailed simulation of the snow cover's seasonal evolution, including mass and energy balance across the snowpack. It's designed for varied spatial scales (from point scale to thousands of square kilometers) and temporal scales (from single events to climate change scenarios), incorporating features such as spatial interpolation of meteorological data, multi-layer snow simulation, and glacier ice response to climate change. The model's flexibility allows customization for specific applications, backed by a Python codebase and available as an open-source project for public use. The paper is concisely and clearly written, but I believe it requires some revisions. I am convinced that making these revisions will improve the paper. The comments on major revisions are as follows.

The gridding of meteorological elements through interpolation of weather station data in this paper is conceptually similar to the Micromet model (Liston and Elder, 2006). Therefore, I would like the paper to discuss the advantages of its methodology by comparing it with Micromet. Additionally, please explain the temperature lapse rate with elevation and the elevation dependence of precipitation using equations. I would also like you to describe how these values differ when compared to Micromet.

We will insert a new paragraph in which we will briefly describe MicroMet. There we will discuss the differences between the two approaches, in particular with respect to the computation of (i) the lapse rate of temperature and of (ii) the elevation dependence of precipitation.

We have inserted a paragraph describing MicroMet with some hints to the similarities and differences between the Barnes objective analysis scheme as realized in MicroMet and the IDW procedure as realized in openAMUNDSEN. This also includes the similarity/differences in determining dynamic lapse rates of temperature. For the latter, we have produced a new figure showing the dynamic lapse rates from openAMUNDSEN *(derived from hourly station observations in each time step of the simulation) compared to monthly averages from several regional contexts (Rofental, Upper Danube river catchment and Northern Hemisphere), to show the necessity to include local observations in the lapse rate determination. Precipitation dependence on elevation cannot be compared between the two systems due to the stepwise correction scheme applied in openAMUNDSEN; this is explained in detail as well. We also added important references for these aspects. Finally, we mention the scientific attractiveness of a comparison of the spatial interpolation procedures in openAMUNDSEN, MicroMet and MeteoIO which could be subject of a subsequent paper, mostly if it comes to determining the effect of different regionalization schemes on simulated snow.*

The method of determining the Snow Redistribution Factor (SRF) should be explained in figures such as Figure 1 or Figure 3. Furthermore, it seems that calculating SRF requires a fairly detailed DEM, so there should be a discussion on the maximum grid size for which SRF can be calculated. Additionally, expressing how SRF is used in a formula would allow for a better understanding, so I would like you to show the utilization of SRF in an equation.

We will include a paragraph for the discussion of the grid size effect of the determination of the SRF (with respect to scale limits), including the original references where its computation is described in detail (Hanzer et al. 2016, Helfricht 2014, Yokoyama 2002, Grünewald et al. 2014, Freudiger et al. 2017). We will try to find a proper balance between what is described in detail in our revised manuscript, and what remains referred to in the respective original literature.

We have re-organized the respective chapter and included an explanation of the computation of the SRF, including the original references where the details of the procedure are described (Hanzer et al., 2016, Helfricht, 2014, Yokoyama, 2002, Grünewald et al., 2014). The way the lateral snow redistribution modelling in openAMUNDSEN is presented now is a balance between what is described in detail in our revised manuscript, and what remains referred to in the respective original literature.

I believe the merit of this model lies in the estimation of the spatiotemporal distribution of snow water equivalent (SWE). On the other hand, the validation data consists of snow cover fraction and snow depth, and I think a comparison with SWE is essential to demonstrate the model's validity. I would like you to show the reproducibility of point SWE measurements. By doing so, it would be possible to verify to what extent the model can reproduce the spatiotemporal distribution of SRF and precipitation, so I would like to request additional validation.

We will use the conducted simulations for the Rofental using different model configurations to provide such a proof of validity. In the Rofental, mostly the Proviantdepot station provides SWE measurement that can be used for this purpose. At Latschbloder only snow depth is observed, and at Bella Vista a local jet with small-scale erosion and deposition effects is observed. We will consider that any comparison of distributed modelling of snow (in a particular grid cell) to local station recordings always suffers from the scale gap between the model resolution and the very particular (local) situation at the measurement device. In mountain regions this effect can become significant. See, e.g., the very recent paper by Haddjeri et al.: (https://egusphere.copernicus.org/preprints/2023/egusphere-2023-2604/).

What we additionally can do is a short description of the "multi-level spatiotemporal validation" developed by Hanzer et al. (2016), a systematic, independent, complete and redundant validation procedure based on the observation scale of temporal and spatial support, spacing, and extent. This approach provides quantitative measures for the validity of the openAMUNDSEN results for all dimensions in space and time.

In the revised manuscript we present simulations for the Rofental using different model configurations to provide the requested proof of validity. In the Rofental, mostly the Proviantdepot station SWE measurement can be used for this purpose. It should, however, be considered that any comparison of distributed modelling of snow (in a particular grid cell) to local station recordings always suffers from the scale mismatch between the model resolution and the very particular (local) situation at the measurement instrument. In mountain regions this effect can become significant. See, e.g., the very recent paper by Haddjeri et al. (2023): (https://egusphere.copernicus.org/preprints/2023/egusphere-2023-2604/).

Nevertheless, we show a comparison of simualted snow depth and SWE for two winter seasons and different model versions for the Proviantdepot site. It shows the general validity of the model simulations, and the most complex energy balance multilayer approach including the lateral snow distribution process providing the best match with the observations. For all model versions we also provide quantitative measures for the comparison to the observations (Pearson correlation/Nash-Sutcliffe efficiency/Kling-Gupta efficiency/RMSE).

In addition we refer to the "multi-level spatiotemporal validation" developed by Hanzer et al. (2016), which is a systematic, independent, complete and redundant validation procedure for openAMUNDSEN results, based on the observation scale of temporal and spatial support, spacing, and extent. This approach provides quantitative measures for the validity of the model results for all dimensions in space and time.

I do not fully understand the meaning of the sentences in lines 355-357, so I would like you to rewrite them more clearly.

We will provide a new formulation of these sentences to better explain the issue of the snow tower artefacts in simulations of decadal horizons or longer.

We have reformulated the respective paragraph (in section 3.6). Now the snow accumulation artefacts in high altitudes are explained better, and an important reference where this effect is explained in detail is included (Freudiger et al. 2017).

Thank you very much for supporting the improvement of our paper! All your comments are very valuable.

Response to reviewer #2

AMUNDSEN is a well-known snow model with multiple capabilities and applications, and this description of the open source version is a useful reference. There is little demonstration of model performance, but that is OK in this model description paper, and the model has been extensively evaluated elsewhere. There is no demonstration at all of the method for generating climate scenarios, and I wonder if so much description is warranted when it is not subsequently used in the paper. Otherwise, there are a few places where I would like to see some more detail, and I have noted some minor corrections.

Line 17

“manyfold applications” – many

This, of course, will be corrected.

Corrected.

64

How long are the ”longer time horizons”, and why is lateral snow distribution then especially important?

We consider „longer time horizons“ to be of decadal length or longer in the simulations, since then snow towers might build up in the upper regions where seasonal melt is smaller than the amount of previously accumulated winter snow. Such artefacts can be avoided by computing lateral snow redistribution processes, removing the snow from prominent summits and ridges, and accumulating it in the slopes and depressions below. This process is also required to correctly compute glacier mass balances in the long term; actually, a way to evaluate the process is to compare the accumulated snow masses with observed glacier mass balances as shown in Hanzer et al. (2016).

This will be re-formulated in the revised version of the manuscript.

This has been remarked by reviewer #1 as well. We consider „longer time horizons“ to be of decadal length or longer in the simulations, since then snow towers might build up in the upper regions where seasonal melt is smaller than the amount of previously accumulated winter snow. Such artefacts can be avoided by computing lateral snow *redistribution processes, removing the snow from prominent summits and ridges, and accumulating it in the slopes and depressions below. This process is also required to correctly compute glacier mass balances in the long term; actually, a way to evaluate the process is to compare the accumulated snow masses with observed glacier mass balances as shown in Hanzer et al. (2016).*

We have re-formulated the respective paragraph (in section 3.6) in the revised version of the manuscript, and we included an important reference with a detailed description of the phenomenon (Freudiger et al., 2017).

86

“v0.9” – the title and text otherwise refer to v0.8.3. (what is it going to take to commit to v.1.0?)

We wanted to wait for all potential improvements of the model during this review process, eventually also in the model code itself, and then name the version that is described in the final revision of the manuscript with number 1.0.

We have changed the version of the model code that is described in the revised version of the manuscript now to 1.0. Therefore also the title of the paper has slightly changed.

Figure 4

It would be useful to see the station locations (the same as Figure 2a?). What is the resolution of the interpolated grid?

The station locations will be inserted in the revised version of the manuscript.; The resolution of the interpolated grid is 20 m; we will add this to the figure caption.

The station locations are inserted in all panels of figure 4 in the revised version of the manuscript. The resolution of the interpolated grid is 20 m; this has been added to the figure caption.

271

I guess that Liston and Elder (2006) is used for (the vast majority of) catchments that are less well gauged than Rofental. How does this compare with the dynamic lapse rates?

We will compute temperature grids for an exemplary winter season, and compare the results achieved with the two approaches. Dynamic lapse rates can only be computed if the distribution of the weather stations is well covering both the area as well as its elevational extent, and if the time series of the forcing data are mostly complete. We agree: this might often not be the case, but for the Rofental it is.

This has been remarked as well by reviewer #1. We have inserted a new paragraph as Appendix A in which we describe the similarities of the two regionalization schemes implemented in openAMUNDSEN and MicroMet, respectively: the inverse distance weighting scheme considering elevation in the one, and the Barnes objective scheme in the other. For illustration purposes, we have produced a new figure showing the dynamic lapse rates from openAMUNDSEN (derived from hourly station observations in each time step of the simulation) compared to monthly averages from several regional contexts (Rofental, Upper Danube river catchment and Northern Hemisphere), to show the necessity to include local observations in the lapse rate determination. We also added te important references. Dynamic lapse rates can only be computed if the distribution of the weather stations is well covering both the area as well as its elevational extent, and if the time series of the forcing data are mostly complete. We agree: this might often not be the case, but for the Rofental it fortunately is.

272

How are the precipitation thresholds chosen?

The threshold was chosen empirically: a value of 0.5 °C wet bulb temperature with a transition extent from 0 °C to 1 °C produced reliable results in many numerical experiments with the model, in particular for the well-gauged site Rofental. We will include this in the text.

The threshold was chosen empirically: a value of 0.5 °C wet bulb temperature with a transition extent from 0 °C to 1 °C produced reliable results in many numerical experiments with the model, in particular for the well-gauged site Rofental. This has been included in the text.

282

The method for calculating multiple reflections from clouds and slopes is not described. These reflections contribute to measured radiation, so does the model not end up double counting?

Yes, this is correct. Multiple reflections from clouds and slopes are computed in the simulations, as they also contribute to measured radiation. In the further processing, cloudiness will be derived from the comparison of the simulation results to the observations as shown in Strasser et al. 2004 (doi:10.1029/2003JD003973).

Yes, correct. Multiple reflections from clouds and slopes are computed in the simulations, as they also contribute to measured radiation. It is necessary to include these reflections in the simulation since in the further processing the cloudiness will be derived in each time step from the comparison of the simulation results to the observations as shown in Strasser et al., 2004 (doi:10.1029/2003JD003973).

353

It is not clear what it means that “different length scales” are used in Figure 5; none are specified.

Figures 5 and 7 (especially 7d)

The two used length scales are 50 m (this is the scale of ridges) and 5000 m (this is the scale of entire mountains). We will provide the length scales used in the revised version of the manuscript and include a better explanation of the concept of the procedure. See also the respective comment to the remarks of Referee #1.

The two used length scales are 50 m (this is the scale of ridges) and 5000 m (this is the scale of entire mountains and valleys). We have provided a more detailed explanation of the length scales used in the revised version of the manuscript, and we included important references where the method is developed and described in detail (Helfricht, 2014; Hanzer et al., 2016). See also the respective comment to the remarks of Referee #1 who also commented on this aspect.

It is counterintuitive for the areas with more snow to be darker.

We have discussed the color scheme issue for figures of this type already for quite some time, and considered different variants for this. There are both advantages and *disadvantages of every color scheme. Finally, we ended up with the used scheme as being a compromise with least weakness.*

We have discussed the color scheme issue for figures of this type for quite some time during the processing of the figure, and tried out different variants of it. We found both advantages and disadvantages of every color scheme. Finally, we ended up with the used scheme as the compromise with least weakness (mainly in figure 7 where it is important). We know this is a compromise, but any other choice would be one, too.

The Figure 7 caption does not mention that the pink blobs are clouds (it is not a big problem, but there were better Sentinel-2 views on several other days in June 2019).

We have searched the archive and found a Sentinel-2 scene with less cloud coverage, stemming from 28/06/2019. However, this scene is less attractive as example for the comparison to the model results, because does not show important details of the snow cover distribution originating in the respective processes of precipitation distribution, lateral snow redistribution and energy supply for melt. Finally, we suggest to stay with the scene showing the 18/06/2019.

422

“orby satellites” – orbital? Or just “satellites” (ones that are not orbital, such as CryoSat-1, are of limited value).

It should be two words: „or by satellites“. This will be corrected.

It should be two words: „or by satellites“. This has been corrected.

630

I can’t tell what the missing early snowfall event is in Figure 9.

This will be added. There is an early snowfall event in the observation which all model versions underestimated. We will re-formulate.

We have replaced this figure with an illustration of snow depth and SWE for the two winter seasons 2020/2021 and 2021/2022 which better serves the purpose of comparing the model results to the observations. Hence, the problem of the early snowfall event and its recognition does not appear anymore.

There are no metrics given, but I might judge from Figures 7 and 9 that the most sophisticated EB + Multi + SRF configuration has the worst performance in comparison with observations.

We will include a short explanation of this. In Figure 9 the most complex model formulation has a low performance, but in Figure 7 one can see that it well captures the processes. At the Proviantdepot station frequent small-scale redistribution of snow by wind occurs, leading to effects which worsen the model performance (buildup of a corniche which modifies the SPA observations). We will include a better explanation in the caption of Figure 7.

We have replaced the illustration of the winter season 2019/2020 in figure 8, in which indeed the most complex model formulation produced a low performance, with the two winter seasons 2020/2021 and 2021/2022 (now figure 9). For these two seasons, the most sophisticated EB + Multi + SRF configuration produces the most reliable results. We now also provide quantitative measures for the comparison of the different model versions to the observations. In (the new) figure 8 the most complex model version also best captures the complex pattern of snow heterogeneity caused by the precipitation distribution, lateral snow redistribution, and energy supply for melt.

Thank you very much for supporting the improvement of our paper! All your comments are very valuable.