Dear Reviewer,

We thank you for your valuable comments and suggestions to improve the quality of the manuscript. We address your comments (dark grey) with our responses (blue) in the following. We think that the outlined additional analysis based on your recommendations will help to improve the study.

General comments

In this manuscript, variables related to snow cover are compared between station measurements and four models in Southern Germany. The manuscript is mostly well-written with clear figures and presents interesting results. However, some points could improve the paper and strengthen the results.

I think the aims of this study and the justification of the research setup should be more clearly stated.

Aims:

Thank you for this hint. We will emphasize the research aims more clearly. The main question of the study is: Can new-generation high-resolution regional climate models represent snow depth dynamics at high temporal (daily) and high spatial detail?

How is that motivated and why is this important?

Impact research needs information about impactful events at the local scale. Climate change affects the dynamics and conditions, which is why observation-based analyses are limited. Often, coarse-resolution RCMs or even GCMs have been used to drive snow models at local/regional scale. However, bias adjustment, statistical downscaling and the de-coupling of the interactions of snow dynamics and climate (snow simulations do not feed back into the climate simulation) induce additional uncertainties and limitations.

The “new generation” of high-resolution RCMs could potentially directly provide snow depth information from their internal land surface / snow modules, which leads us to the guiding question: How good are they at representing snow depth?

Setup:

To answer the question, we need to:

1) explore high-resolution RCM simulations,
2) which cover not only single years but climatological periods (~ 30 years) to represent the variability and extremes
3) and which are driven by reanalysis in order to be able to compare to observations
4) define a baseline (in our case ERA5L)
5) define a reference (in our case in-situ observations)

1) – 3) strongly limits the choice of available simulations. The CCLM and WRF simulations are the simulations, which we found publicly available.

4) ERA5L as global land reanalysis is the state-of-the-art reanalysis product at 9km resolution, which is also driven by the same climate (ERA5 atmosphere) and therefore comparable as baseline.

5) For snow depth, in-situ observations are the typical validation reference (see e.g. https://tc.copernicus.org/articles/15/1343/2021/). However, based on your suggestion (later in the text) and the other reviewer's suggestion, we will add remote sensing data for snow cover validation of the gridded simulations.

In addition to 1) – 5), we added the AMUNDSEN simulations driven by CCLM at the point scale. This setup was added in the course of the evaluation, where the CCLM showed strong systematic underestimation of almost any snow variable, while representing the climate better than the other models (lowest biases and errors). Hence, driven by the perspective of impact research, we wanted to explore how the separate snow model AMUNDSEN can make use of the well representative CCLM climate.

The authors have decided to use only one variable (snow depth) instead of using e.g. SWE, even though SWE might be a better variable for estimating snow accumulation and snow melt. They state (L211) that snow melt is assumed as the main driving process for snow depth reduction without adding any reference. I think this should be further discussed.

Based on the suggestion of the other reviewer, we remove this analysis of snow accumulation and melt due to the uncertainties of compaction. We agree that SWE would help to support our analysis. However, this variable is not available from the stored simulations (WRF & CCLM).

Also, I think the authors should state more clearly why this study area (Southern Germany) was chosen. It would be interesting to see the analysis cover also mountainous regions.

The study area is exposed to impactful snow depth dynamics, but not as strongly as alpine regions. The choice of the study area is motivated by the impact perspective. Impact and risk are related to exposure and vulnerability. While snow depths in alpine regions will be more extreme, the exposure in terms of affected people is higher in Southern Germany.
However, we fully agree that such an analysis would also be interesting for alpine areas. The other reviewer has given a reference to https://doi.org/10.5194/tc-17-3617-2023, where a 12.5km and a 2.5km simulation is evaluated over the Alps.

Why are only in situ observations used in the comparison and not remote sensing data? I think the authors should clearly state the justification for using only in situ data. As three of the used models show gridded snow cover estimates, using e.g. satellite-based gridded estimates as a comparison would make sense.

For snow depth, in-situ observations are the typical validation reference (see e.g. https://tc.copernicus.org/articles/15/1343/2021/). However, based on your suggestion, we will add remote sensing data for snow cover validation of the gridded simulations. We aim to compare the gridded simulations to MODIS TERRA snow cover (MOD10C1: https://modis-snow-ice.gsfc.nasa.gov/?c=MOD10C1) at daily resolution. This evaluation will cover the winter seasons 2000/2001 to 2017/2018.

Specific comments

Table 1 is a bit confusing and it is hard to see which model is in which line, as the names that are used throughout the manuscript are not clearly listed. I suggest adding a column on the left with the model name (WRF, CCLM, AMUNDSEN, and ERA5). Also, consider organizing the table in the same order as the text (first Regional climate models, second ERA5-Land, and last AMUNDSEN).

Thanks for this hint, we will adjust the table accordingly and add the model names.

Figure 3. I suggest adding the name of the model on top of each plot instead of just mentioning it in the caption. This applies to all figures throughout the manuscript.

We will add the model names for all figures, where subplots are categorized by the model.

Figure 3. Consider adding a diagonal line (from lower left to upper right corner) to each plot so it is easier to see whether there are biases in the models. This also applies to all similar figures throughout the manuscript.

Diagonal will be added to all scatterplots.

Figure 4. Consider adding a bit darker horizontal and vertical lines at x=0 and y=0.

We will enhance the 0-lines accordingly (darker/thicker).

Figure 5. Please add somewhere (in the text or in this figure) that you use the acronym DWD for the observations.

Will be replaced by “OBS” in all relevant figures and explained in the caption.

Figure 10. Is the difference relative or absolute difference?

The differences of white Christmas are absolute differences of the percentages. We will clarify this in the caption.

L381. Change 0.15 m to 15 cm. I think it is better to use the same units that are used in the figure.

True, we will adjust the units to cm in the text.

I suggest the authors check the grammar carefully. The manuscript is overall very well-written but some mistakes (e.g. incorrect prepositions) exist.

Thank you for the feedback – we will check the grammar in the course of the revision.