Response to Reviewer 3

I thank the reviewer for his / her valuable comments on the manuscript. My response to the comments and the changes to be made in the revised manuscript are detailed below. For clarity, the comments are in blue font, while my response is in black. In some cases, I have included the text planned to appear in the revised manuscript in red font.

This is one of the most thoughtful and comprehensive reviews that I have ever received for any of my manuscripts. Due to practical constraints, I unfortunately cannot implement the two major requests of the reviewer (regarding spatial and temporal resolution and the more detailed interpretation of the snow-on-ground-fraction term) to the extent that the reviewer might have hoped. However, I have done my best to do what can reasonably be done, and I hope that my justifications for not doing more make good sense.

The manuscript "Changes in March mean snow water equivalent since the mid-twentieth century and the contributing factors in reanalyses and CMIP6 climate models" uses a decomposition-based approach to evaluate and diagnose changes in historic snow water equivalent (SWE) trends in the northern hemisphere for two historic periods in two reanalysis products, the GlobSnow product, and the CMIP6 ensemble. SWE is decomposed into contributions from total precipitation, snow fraction, and snow-on-ground fractions. The primary findings include high degrees of spatial correlations but individual variation due to model-to-model differences in snow-on-ground fraction. Total SWE has gone up in many far northern locations due to increases in precipitation, though CMIP6 appears to overestimate SWE increases. Biases in all data sources are discussed, and the paper highlights how difficult SWE is to correctly simulate in climate models. The topic is within the scope of The Cryosphere.

Overall, I like the methodology as an interesting approach to breaking down drivers of SWE. However, my first major concern centers around choices in the spatial aggregation step and the monthly time step that may lead to very different results. Huge efforts and computational resources have been expended on improving spatial resolution of snow (and many key climate system variables) in global models, and it is unfortunate to not see how finer resolution may improve the insight provided by the work (it also may not, but it needs to be shown). My second major concern is the lack of energy balance components in SWE trends is perhaps implicit in the G term, but this is not apparent from the methodology description.

My response to these comments is divided to three parts, labeled as (1)-(3) below.

(1) I start from the G term because this explanation may also shed light on why the time resolution is not important. In this study, G is a purely diagnostic quantity. From Eq. (1) one gets

$$G = \frac{SWE}{\int_{t_0}^{t} FPdt'} = \frac{SWE}{\int_{t_0}^{t} Snowfall\ dt'} = \frac{SWE}{accumulated\ snowfall}$$

For calculating *G*, one only needs the SWE and the accumulated (i.e., time-integrated) snowfall from the beginning of the winter season. Both the SWE and the snowfall are directly available from reanalysis and climate model output. To clarify, the paragraph preceding Eq. (1) will be rewritten as:

Our diagnostic framework follows Räisänen (2008, 2021a). The only three variables that are needed from a reanalysis or a model simulation are monthly means of SWE, snowfall, and total precipitation (*P*). The monthly snowfall is first rewritten as *FP*, where *F* is the fraction of precipitation that falls as snow. SWE in month *t* then becomes

$$SWE = G \int_{t_0}^{t} FPdt' \quad (1)$$

Clearly, there are many different processes and meteorological factors that affect G. A detailed analysis of these factors in the individual reanalyses and climate models is well beyond of what can be reasonably included in the current paper. However, I will point out the importance of this issue by adding the following paragraph to the end of Section 4 in the revised manuscript:

Table 4 also shows that the inter-model differences in *G* are in relative terms larger than those in *P** and *F**. This is perhaps unsurprising, since *G* may be affected by a multitude of factors. As defined by Eq. (1), *G* reflects the balance between the source (accumulated snowfall) and sinks (snowmelt plus sublimation) of snow. The accumulated snowfall depends on both the amount and phase of precipitation, whereas snowmelt and sublimation are ultimately determined by the amount of energy that the land surface model allocates to them. The latter, in turn, is constrained by the downward solar and thermal radiation reaching the surface, the exchange of sensible and latent heat between the land surface and the atmospheric models, the description of the surface albedo and emissivity, and the use or release of energy associated with temperature changes within the snow-ground-vegetation system. As many of these processes are described differently in different land surface models, it is perhaps unsurprising that the simulated SWE may vary substantially even between land surface models that share the same atmospheric forcing (Mudryk et al., 2015). A more detailed understanding of the causes of variation of *G* within the CMIP6 ensemble is an important target for future research.

(2) The low spatial resolution of the analysis leads to a loss of potentially valuable local information from the high-resolution analysis products. However, the 2.5° × 2.5° grid is sufficient when the focus is on large-scale trends, as is the case in this paper. Importantly, the trend results in this coarser grid are insensitive to the order between grid remapping and the decomposition (Eqs. (1)-(2) in the manuscript). In other words, if one starts with remapping all the required input data to the 2.5° grid before applying Eqs. (1)-(2), the resulting trends are very similar to those obtained by first doing the calculations in a finer grid and then remapping the trends to the 2.5° grid. The only (expected) exception are 2.5° grid boxes that are partly covered by sea in the original grid, because the required input data are only available over land. This similarity, as well as the loss of small-scale information, is illustrated in the **new Figure B1, included in the next page**.

In any case, redoing the analysis on a finer resolution would have a low benefit-to-effort ratio, because I have only archived a large fraction of my CMIP6 data in the 2.5° grid. Furthermore, as most of the CMIP6 models only have a resolution of 1-2°, using a grid finer than 2.5° would not help to draw much new information from the models.

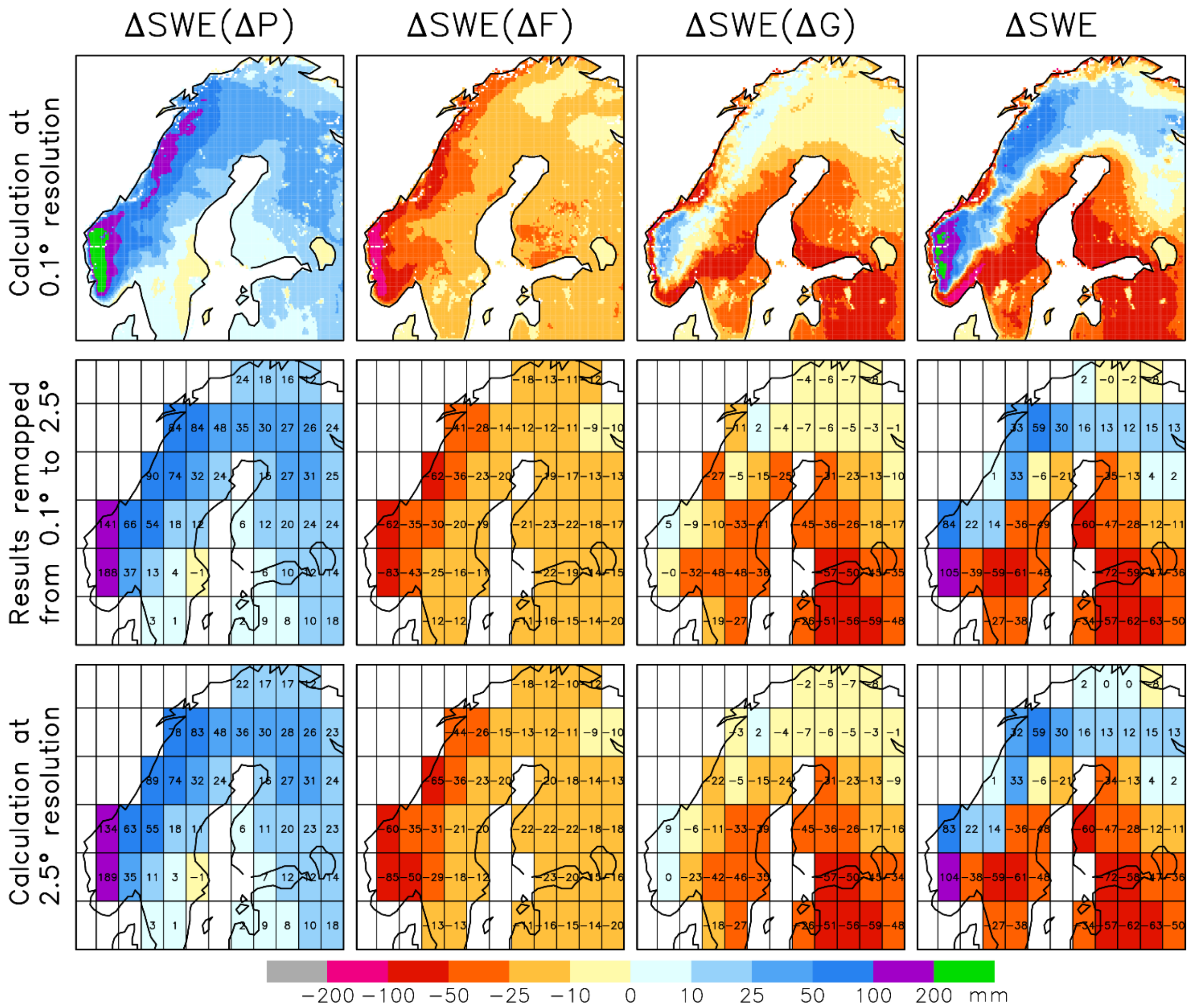

Figure B1. Trend in March mean SWE in Scandinavia from 1951 to 2022 (mm (71 yr)$^{-1}$) (column 4) and the contributions to it from changes in total precipitation (column 1), snowfall fraction (column 2) and snow-on-ground fraction (column 3). Top: calculation using ERA5L data at their native 0.1° resolution. Middle: values from the top row remapped to 2.5° resolution. Bottom: calculation using ERA5L data remapped to 2.5° resolution before the trend decomposition.

(3) The choice to use monthly rather than daily data in the analysis is also a pragmatic one. Using daily data would increase both the data volume and the computing time by a factor of 30. In principle, it would be better to do the calculation with daily than monthly mean data, because this would allow a more precise evaluation of the time-integrated snowfall in Equation (1). However, in practice, this difference is very unlikely to be important.

To get an idea of the potential impact of the time resolution of the input data, I recalculated the 1951-to-2022 trends for ERA5-Land using an even coarser time resolution: 2 months instead of 1 month. Constrained by the 2-month resolution, this analysis was done for the trends in the February-March mean SWE rather than March mean SWE. In Figure R1 below, the results are compared to the corresponding February-March mean trends calculated with 1-month time resolution.

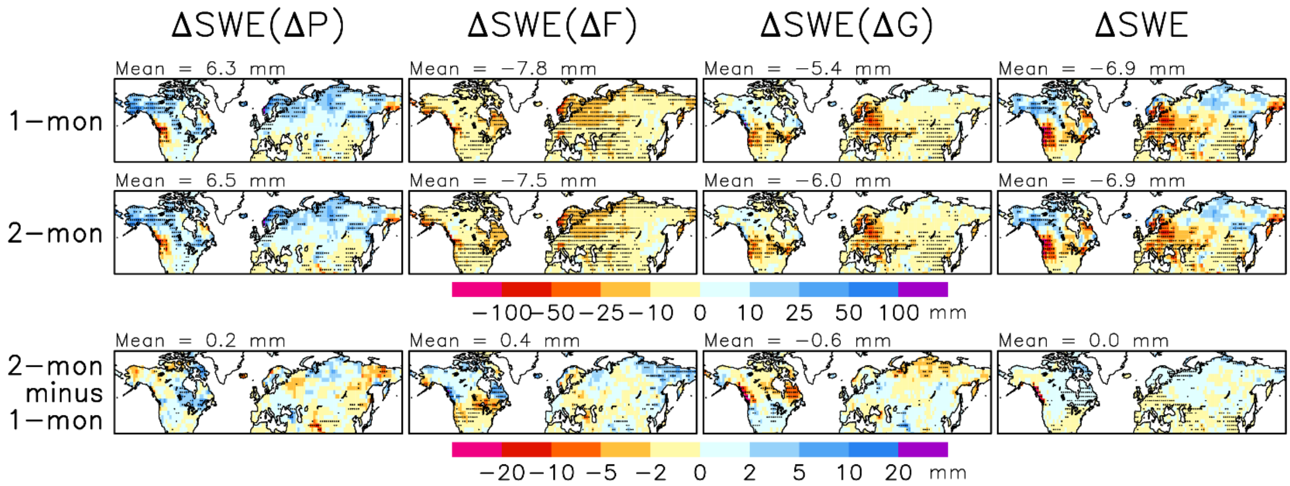

Figure R1. Decomposition of 1951-to-2022 trends in February-March mean SWE, as calculated by (top) monthly and (middle) 2-monthly ERA5-Land data. The difference is shown in the bottom row, using factor of 5 smaller intervals in the color scale.

The results with 1- and 2-month time resolution are very similar. The spatial correlations for the trends in $\Delta SWE(\Delta P)$, $\Delta SWE(\Delta F)$, $\Delta SWE(\Delta G)$ and SWE are 0.99, 0.96, 0.95 and 0.99, respectively. The reason why the SWE trends are not identical is the subtraction of the SWE in the first time unit (August for 1-month but August-September for 2-month time resolution) when calculating the "seasonal component" of the February-March mean SWE. Generally, the time resolution mainly affects the trend decomposition results in those areas where non-negligible snowfall already occurs in September. Elsewhere, the differences are negligible.

Because climatic conditions such as the amount of snowfall change more between two months than within a single month, this comparison likely gives an upper estimate of the errors that monthly time resolution may lead to.

The paper is written well. For the audience I expect to find this paper of the most benefit (e.g., modeling and observational groups focused on snow processes), the figures are satisfactory. There are numerous occasions where qualitative information is given (e.g., L232 "slightly larger") or implied (L235 "exceeds") when numerical values would be welcomed as they are more informative.

There is a delicate balance between the exactness and the readability of the text. I have added several more numerical values in the revised manuscript text, where I found this appropriate. In other cases, I have added more explicit references to the figures or tables from which the values can be seen (the L235 case mentioned in this review comment falls to this category). Still in other cases it would be potentially misleading to try to give a precise number or even a range of values, particularly when referring to features seen in maps (of this article or earlier ones). Since verbal definitions of areas (e.g., "easternmost Siberia") will be interpreted differently by different persons, the meaning of seemingly well-defined numerical values would also be interpreted differently.

The paper would also benefit from additional citations to shore up arguments and to increase its comprehensive coverage of the topic. With revisions to address the major, minor, and specific comments, I believe this manuscript will make a solid contribution to the literature.

Many additional citations have been added. I apologize if (as is likely) I have missed something important, but sometimes it is surprisingly difficult to find good references to results that appear to be established "silent knowledge” in the community. Also, as the paper is quite long already, making it even longer with a comprehensive discussion of earlier literature would make it less attractive to many readers.

Major Comments:

1. The aggregation via interpolation of observational data and CMIP6 simulations to 2.5° horizontal resolution likely will cause a lot of loss of potentially valuable information, especially in mountainous areas where SWE can vary over much smaller spatial scales. I am concerned that the results and interpretations may be very different if different resolutions are used. There may be good reason to do this aggregation as performed, but this needs to be shown. In other words, how much do the results vary if the ERA5, MERRA2, CMIP6, and GlobSnow are all re-gridded to 50 km instead of such a coarse 2.5° resolution? This exercise may also highlight where models are doing better/worse across locations in the Northern Hemisphere and be valuable to modeling groups who want to improve their results.

See above. I agree that a trend analysis on a finer grid would give valuable insights on regional SWE change dynamics in (e.g.) many mountain areas, but this is a topic for another study. The focus in this study is on large-scale features, the 2.5° resolution is sufficient for this, and the trend decomposition results on this scale would not be materially altered even if higher-resolution data were used before re-gridding (rows 2-3 in the new Figure B1, included in Page 3 above). Also note that the resolution for some of the CMIP6 models is close to 2.5°, so no higher-resolution information is available for them.

1. What physical components control the SWE on the ground (the G term in eq 1)? Does this term accumulate the effects of the surface energy budget? More explanation is needed as this is a critical component of SWE trends (the “demand side” whereas F and P are the “supply side”). While the focus on the supply side of SWE trends is valuable, adding an analysis of G, especially if broken down into its physical components, would greatly elevate the papers contribution to understanding what is driving SWE trends as well as be helpful for the modeling community to improve the representation of these key processes (L270 and L349 acknowledges the challenges of snowmelt modeling and role of how this step leads to differences in results). This would allow the paper to nicely ‘bookend’ the supply and demand sides in the CMIP6/historic reanalysis framework.

See above. This study only attempts to answer how trends in G have affected SWE. Diagnosing why G has changed in the way that it has would require very extensive additional analysis. This is far beyond the scope of the present manuscript. Nevertheless, a brief discussion of the factors that may affect G will be added to the end of Section 4 (the paragraph in red font on Page 2 of this file).

Minor Comments:

1. It was unclear what timestep of observational and model output was used, was it monthly March means or the date of peak snow water equivalent in March (daily). Similar to Major Comment #1, how sensitive are the results to the use of monthly versus daily values? After reading further, it appears March monthly mean was used, but was this necessary for some reason? Why not use the peak March value (which likely would vary in time across the month by latitude, elevation, a given year, etc)?

This was already discussed above. The choice of monthly time resolution was largely pragmatic. To be more explicit about the time resolution of the input data, the following sentence will be added before Eq. (1):

The only three variables that are needed from a reanalysis or a model simulation are monthly means of SWE, snowfall, and total precipitation (*P*).

1. L229: This difference in mean SWE could be readily examined by using finer (native) resolutions, making a stronger point than the speculation provided at present.

Yes, it might be possible to analyze the values of SWE together with model orography in major mountain regions to get more insight on this. However, at least from my point of view, this is out of the main focus of the study, and I have therefore not pursued this further.

1. It would help to have a bit more comprehensive literature review on SWE comparisons between reanalysis products such as MERRA2 and ERA5L across locations and scales; this may also be helpful when discussing the results found.

Although there are several studies in which SWE in different data sets is compared, the most salient information condenses on two points: (1) how large is the overall variability between data sets and (2) what can be said about the likely bias in ERA5-Land, MERRA2 and the CMIP6 models relative to the “real” real-world SWE. Thus, the revised manuscript will include two additions in this purpose. First, the following sentences will be added to the Introduction (in the paragraph starting “Remote sensing of SWE” on P. 3).

On the other hand, model biases and limitations in the observational input result in a large spread between the SWE estimates from different analysis products (Mudryk et al., 2015; Mortimer et al., 2020). Mudryk et al. (2015) found more than a factor of 1.5 range even in the Northern Hemisphere total winter peak snow mass among the five datasets that they evaluated (their Fig. 3a).

Second, the following paragraph will be added in the middle of Section 4, where the discussion is about whether the CMIP6 models over- or underestimate the average SWE:

Assuming that GlobSnow and the other observational estimates used by Kouki et al. (2022) are correct, the average SWE is too large in both the CMIP6 MMM, MERRA2 and (especially) ERA5L. A comparison of GlobSnow with three other bias-corrected estimates of the total snow mass in Northern Hemisphere non-alpine areas (Table 1 of Pulliainen et al., 2020) supports this assessment: all the four estimates are within 7.4 %, GlobSnow being the highest. However, SWE in mountainous areas is known less well and may be severely underestimated in many gridded analyses (Snauffer et al., 2016; Wrzesien et al., 2018). Despite the higher mean SWE in ERA5L than in MERRA2 and GlobSnow, Munoz-Sabater et al. (2021) found ERA5L to underestimate SWE by ~50 % at the five Earth System Model – Snow Model Intercomparison Project (Krinner et al., 2018) alpine reference sites.

1. I liked the two directions posed in the conclusion, however the way the conclusion was organized, the paper felt like it ended abruptly. I suggest adding a short closing section to finish the article or reorganizing this section.

Again, this is a good comment. Re-reading the original manuscript, I realized that it missed a discussion point on what stakeholders should do with the SWE projections, and the research needs related to this. This will therefore be added to the end of the Conclusions section:

Another important question is, how should water resource managers and other stakeholders needing SWE projections use the information available from climate model ensembles. On the face of it, the answer seems disappointing. As real-world SWE trends appear to have hitherto been further away from the model-simulated trends than the latter are from each other, this might also well apply to the trends in the future. It would therefore seem prudent to allow for all the uncertainty suggested by the variation within multi-model ensembles, possibly adding a safety margin for systematic model errors. However, this conclusion may need re-evaluation if future research is able to reduce the uncertainty in the observed trends. Moreover, the practice of giving all models the same weight in projections of future SWE change is probably sub-optimal, particularly for longer-term projections in which model differences grow increasingly dominant over internal variability. For example, Räisänen (2008) found a dependence between model-simulated baseline winter temperatures and projected future SWE changes, which makes increases of SWE more likely in models with a cold than a warm temperature bias. He also used inter-model cross validation to show that this physically expected dependence could be potentially used for improving probabilistic projections of SWE change. Furthermore, inter-model variations in the Northern Hemisphere snow albedo feedback are strongly correlated between seasonal and climate change time scales (Hall and Qu, 2006; Qu and Hall, 2014). This does not guarantee that a similar cross-time-scale constraint would be available for SWE as well, but it suggests that the possibility is worth exploring.

Specific Comments:
L23-32: Suggest to add citations to the first paragraph; all good info but offers opportunity to point readers to recent work on various sub-topics.

Several new citations have been added to the first paragraph in the Introduction (see below). I would have been pleased to find even more, but it turned out to be surprisingly difficult to find good references even to some seemingly well-known results. Part of the difficulty is that the literature on model-simulated changes in snow conditions (which is the theme of this paragraph) is more limited than that on observed changes.

Simulations of greenhouse gas induced climate change by global climate models feature both warming and, in the northern mid-to-high latitudes in winter, an increase in precipitation (Lee et al., 2021). These changes have opposing effects on snowpack. An increase in precipitation, if acting alone, would increase the amount of snowfall and snow on ground. However, an increase in mean temperature favours the occurrence of above-zero at the expense of below-zero temperatures, particularly where and when the mean temperature is relatively close to the freezing point (de Vries et al., 2013; Räisänen, 2016). Therefore, a smaller fraction of precipitation falls as snow (Krasting et al., 2013; Kapnick and Delworth, 2013; Räisänen, 2016) and the snowpack is reduced by more frequent and intense melt events during the winter (Musselman et al., 2021). The net effect of these changes in climate model simulations is a shortening of the snow season (Brown and Mote, 2009; Zhu et al., 2021) and a decrease in the snow water equivalent (SWE) in most areas (Mudryk et al., 2020). However, in the coldest regions such as eastern Siberia and northern Canada, the increase in total precipitation tends to dominate and thus leads to an increase in SWE at the height of the snow season (Räisänen, 2008; Brown and Mote, 2009).

L41: Add “beneficial” before “ecological impacts”?

Suggestion accepted.

L45-54: This paragraph would be stronger with additional numerical quantification of historic change and more local specifics, e.g., what are the decreasing trends?

I found two numerical values that are sufficiently easy to understand and well-defined enough for addition to this paragraph:

… Pulliainen et al. (2020) evaluated trends in snow mass in 1980-2018 using the Global Snow Monitoring for Climate Research (GlobSnow) v3.0 analysis. Focusing on non-mountainous areas north of 40ºN, they found a statistically significant decreasing trend in March mean snow mass in North America (best estimate: -4 % decade^{-1}) but a near-zero trend in Eurasia. … Using observations for the years 1981-2010, Mudryk et al. (2017) found the monthly mean snow cover extent in the extratropical Northern Hemisphere land areas to decrease throughout the snow season by (1.9 ± 0.9) × 10^6 km^2 for each 1°C increase in the mean temperature in the same month and area.

L69: Suggest removing “three-dimensional” as weather prediction and climate models are four-dimension (3-d in space plus a time dimension). Would also benefit to add some examples of these to give readers examples of their spatial and temporal resolution of output.

The word “three-dimensional” will be removed, and the following example of spatial and temporal resolution will be added: Reanalyses and land surface models produce spatially complete SWE simulations, in some cases with high spatial and temporal resolution (e.g., hourly data at 9 km resolution for ERA5-Land (Muñoz Sabater et al., 2021)).

L81: add the scale this paper addresses (Northern Hemisphere) to “recent trends in SWE” just to set the stage for going into the questions.

Suggestion accepted.

L107-112: Instead of this table-of-contents explanation of upcoming sections, I think this text could be markedly condensed and some initial insights into what the work provides the community could be given, thus helping motivate the work better.

Table of contents will be replaced with this sentence:

A key finding of this research is a reasonable agreement between the ERA5-Land reanalysis and the CMIP6 models on the March mean SWE trends and their contributing factors in the period 1951-2022 (Section 5) but a worse agreement between various observational data sets with both each other and the CMIP6 models on the trends from 1981 to 2022 (Section 6).

L138: Please quantify “thick”
L138: How much is this underestimation of SWE?

The essence of the problem is that the microwave signal saturates when SWE reaches ca. 150 mm, and the uncorrected GlobSnow SWE ceases to increase when this limit is exceeded. The corresponding sentence will be reformulated as:

These corrections are based on comparison with snow course measurements, and they improve the SWE estimates especially in areas with thick snow, where the non-corrected data systematically underestimate SWE due to the saturation of the microwave signal when SWE exceeds ca. 150 mm (Pulliainen et al., 2020).

L141, 167: Tables 1 and 2: Please add a column of the native spatial horizontal resolution of these products

A column of the native horizontal resolution will be added in Tables 1 and 2.

L154: Please define the SSP scenarios briefly and add a reference.

SSP = Shared Socioeconomical Pathways, and the chosen SSP2-4.5 scenario is often characterized as a “middle-of-the-road” future, although the difference from other scenarios only becomes important in the longer-term future. The sentence will be rewritten as follows:

Climate model simulations from the 6th phase of the Coupled Model Intercomparison Project (CMIP6) (Eyring et al., 2016) were also used, concatenating historical simulations for the years 1950-2014 with simulations for the Shared Socioeconomical Pathways “middle of the road“ 2-4.5 scenario (Fricko et al., 2017) for years 2015-2022.

L165: Please add a reference for conservative remapping (even if just the function used in the programming language) to ensure repeatability.

A reference will be added to

Jones, P. W.: First- and second-order conservative remapping schemes for grids in spherical coordinates, Mon. Weather Rev., 127, 2204-2210, https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2, 1999.

L172: Table 2: suggest making the delta-T subscripts hyphenated (51-22) to be more intuitive.

Suggestion accepted.

L180: What fraction of areas in the Northern Hemisphere have permanent snow cover? Has this number changed substantially? It would be OK to give the mean or median value for a given date range to give readers an idea even if this is calculated for each specific year.

I chose to answer a slightly modified question that is more important for the current work: in how large a fraction of the analysis area does the choice to subtract the August mean SWE affect the results?

To make Eq. (1) also applicable in areas where snow cover regularly or sporadically survives to the late summer, we subtract the August mean SWE from the left-hand-side, thus focusing on the seasonal component of SWE. For reference, in ERA5L about 7 % of the Total Snow Area (red and yellow shadings in the bottom-left panel of Fig. 1) has non-negligible (> 5 mm) time mean August SWE in the 2.5° × 2.5° grid, largely in mountainous and Artic areas. For MERRA2, this fraction is only 1 %.

L189: Please define “very small”.

Typically: two orders of magnitude smaller than the other terms:

Thus, the anomaly in SWE is decomposed to contributions from the total precipitation (ΔP), snowfall fraction (ΔF) and snow-on-ground fraction anomalies (ΔG), plus a non-linear term that is typically two orders of magnitude smaller than the others.

L239: Alternatively, could the ERA5L SWE being “too large” be a product of it doing a better representation of SWE at high elevations as well?

Yes, it could. The following paragraph will be added to the middle of Section 4, to acknowledge among other things that the SWE in mountainous areas is not well known:

Assuming that GlobSnow and the other observational estimates used by Kouki et al. (2022) are correct, the average SWE is too large in both the CMIP6 MMM, MERRA2 and (especially) ERA5L. A comparison of GlobSnow with three other bias-corrected estimates of the total snow mass in Northern Hemisphere non-alpine areas (Table 1 of Pulliainen et al., 2020) supports this assessment: all the four estimates are within 7.4 %, GlobSnow being the highest. However, SWE in mountainous areas is known less well and may be severely underestimated in many gridded analyses (Snauffer et al., 2016; Wrzesien et al., 2018). Despite the higher mean SWE in ERA5L than in MERRA2 and GlobSnow, Munoz-Sabater et al. (2021) found ERA5L to underestimate SWE by ~50 % at the five Earth System Model – Snow Model Intercomparison Project (Krinner et al., 2018) alpine reference sites.

L255: Missing period at end of sentence.

Corrected.

L281: Suggest to be more specific in terms of temperature (and radiation), also add citation(s) here unless this finding is directly shown, in which case please reference the figure or table here.

The sentence that this comment refers to ("This term is the most negative in mid-latitude North America and in a zone extending from eastern Europe to southern Scandinavia, where the main snowmelt season is ongoing in March and has been advanced by rising spring temperatures.") is based partly on analysis of the ERA5-Land data, partly on my knowledge of the climate in the mentioned areas. I could not find a good earlier reference to this result, but I find it physically intuitive. Also note that, although seasonal increases in both temperature and solar radiation contribute to the spring snowmelt, the greenhouse gas induced climate change is expected to affect temperature much more than the insolation.

L284-285: Add references; also would help to provide some additional physical reasoning about the logic involved in this sentence.

The reasoning is that, as far as the mean temperature is well below zero, the frequency of above-zero temperatures that lead to significant snowmelt remains small:

Although warming is generally expected to enhance snowmelt, this effect is modest where the mean temperature in March and in the preceding winter months is well below zero, so that above-zero temperatures remain uncommon despite the warming (e.g., Räisänen, 2008).

L317-320: Would help to give trend values in the text for easy comparison.

The trend values depend on the precise delineation of the areas, which is necessarily diffuse. Therefore, I just added a reference to the appropriate figure panels in this sentence.

L331-334: Does this suggest the models are getting temperature signals correct? I see later this is noted in a few paragraphs (L343); perhaps consider reorganizing the text in this section and the other to allow it to flow more naturally?

I considered the reorganization, but it would create another problem because the finding also applies to the CMIP6 simulations that are only discussed after this sentence. Therefore, I chose to add a brief explanation in this sentence (which, as argued below …) and keep the full discussion where it was.

Thus, ERA5L and MERRA2 agree reasonably well on the SWE trends caused by changes in the snowfall fraction (which, as argued below, is a relatively straightforward response to warming), but much less well on the trends associated with changes in total precipitation and the snow-on-ground fraction

L370: change "Both" to lowercase

Corrected.

L371: all or nearly all, can you rephrase to give a specific number?

Rewritten as: In MERRA2, changes in total precipitation make a smaller positive contribution to the SWE trend since winter 1981 than any of the models simulate, and the trend in ERA5L is also exceeded in 20 of the 22 models (Table 4).

L396: Change to: “With regards to the components…” or “Regarding the componen”

Changed to “Regarding”.

L428: To what extent might relative humidity (or direct moisture variables) be at play in the snowfall fraction? There has been a lot of recent work highlighting the importance of atmospheric moisture in determining precipitation phase (see e.g., Jennings et al. 2018 Nature Communications). Could be worth mentioning here.

It is true that both the relative humidity and the lower tropospheric lapse rate also affect the phase of precipitation. However, in the context of climate change, changes in these two are very likely to be less important than temperature change. I therefore decided to include these additional factors only in an implicit way, by adding two references in the following sentence:

This is most likely (i) because the phase of precipitation is primarily (Auer, 1974) although not completely (Sims and Liu, 2015; Jennings et al., 2018) determined by temperature and (ii) because the observational uncertainty is smaller (Gulev et al., 2021) and the signal-to-noise ratio is higher (Räisänen, 2001; Hawkins and Sutton 2009, 2011; Lehner et al., 2020) for temperature than precipitation changes.

L465: Is this comparison of spatial correlation performed at 2.5° resolution or at finer scales?

The revised text will explicitly mention that the correlations were calculated at the 2.5° resolution: For example, the spatial correlation between the CRU and ERA5L trends in the 2.5° grid is 0.69 for temperature but only 0.39 for precipitation (Table B2).

L474: This is interesting discussion; have other authors suggested, found similar results, or further investigating these aspects in CMIPx? Worth adding if so.

Yes. Lehner et al. (2020) confirm these results for the more recent CMIP5 and CMIP6 simulations, and I will therefore add this reference:

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., and Hawkins, E.: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, Earth Syst. Dynam., 11, 491–508, https://doi.org/10.5194/esd-11-491-2020, 2020.

L501: Instead of "Height" would ”maximum extent” or “maximum depth” be better terms?

Rewritten as: The focus was on SWE in March, when the Northern Hemisphere snow mass is the largest (Pulliainen et al., 2020).

L513: Here’s a great example of where numerical values would help readers and would be useful for citing this work! How much compensation of increased SWE from precip is there in the other direction (loss) from FP and G declines? OK to give an average and the range (e.g., 1 standard deviation).

Average values for ERA5-Land will be added. Further details would make the text too complicated:

Trends from winter 1951 to 2022. ERA5L and the CMIP6 models agree qualitatively well on the dynamics of SWE change. Although increasing total precipitation has acted to increase SWE in most of the extratropical Northern Hemisphere (average contribution in ERA5L in the Total Snow Area: 7.1 mm), this has been more than compensated by reduced snowfall fraction (-8.0 mm) and, in most areas, reduced snow-on-ground fraction (-7.4 mm).

L540: Per Major Comment 2, did this approach address how snowmelt occurs and what drives it becoming more rapid or frequent as the climate warms (I'm not sure "efficient" is the right word here, either, though I suppose if SWE is broken down into melt and sublimation terms, water input to the surface from snowmelt would be less efficient per unit SWE if more water was lost to sublimation).

More efficient will be replaced by enhanced snowmelt in a warmer climate. However, as discussed above, a more detailed discussion of the drivers of the decreasing snow-on-ground fraction is beyond the scope of the current analysis, which simply diagnoses this fraction from the ratio between SWE and accumulated snowfall.