Referee comment	Author response
General comments	
This manuscript presents an impressive multi-faceted analysis of the historical and (possible) future snow cover on Iceland based on in-situ observations, remotely sensed data, and climate simulations. The fact that much of the analysis, including the snow modeling, seems to have been carried out in the cloud was particularly novel to me and should be of great interest to the wider cryospheric community. Among the main results of this study, historical observations show an increasing trend in snow cover frequency (SCF) over Iceland during the last decades whereas projections from the simulations indicate that SCF will decrease substantially in the future especially under the more aggressive emissions scenario. This could suggest a threshold effect related to changes in precipitation phase under a warmer and wetter future climate. While the manuscript was generally well written it is quite brief and parts of it felt a bit rushed. For example, the text could have benefited from more examples of recent related work on remote sensing and modeling for snow cover reanalysis (e.g. Alonso-González et al., 2021; Liu et al., 2021) and projections (Fiddes et al., 2022). The authors could also have taken advantage of their efficient conceptual snow model combined with downscaling (see Fiddes et al., 2022, and references therein) to conduct simulations at resolutions nearer to the hillslope scale (Fan et al., 2019) to try to better capture the spatial variability of the snowpack. Moreover, it did not seem as though the parameters in the SNOW-17 model were well calibrated using local information. This is a lost opportunity, as the MODIS data would have been a good candidate to calibrate many of the parameters (e.g. Fiddes et al., 2019; Alonso-González et al.,	The authors are grateful to the referee for thorough and insightful comments. We have addressed all of the referees' general and specific comments and believe that this has resulted in substantial improvements to the paper. These include • Expansion of the introduction section e.g. to highlight more recent studies in the field. • Revision and extension of the methods section e.g. clarify and elaborate of the modelling strategy and the description and processing of observational data. • Improved presentation and discussion of the results e.g. revision of all figures and the introduction of a new Figure 3 and discussion on the threshold effect mentioned by the referee. • Revisions to the discussion section for improved clarity. We agree with the referee that are many modelling strategies that can be applied to study snow cover. The objective of this study was to analyze trends in decadal average snow conditions over centennial time scales given of future climate scenarios derived by an ensemble of model projections. The modelling strategy used in this study was developed to optimally achieve that objective within the limits of the computational framework at our disposal. We agree that it would be a valuable effort to simulate Icelandic snow conditions at hillslope scale resolutions and that the MODIS data could be a good source to estimate some the SNOW-17 parameters at that resolution. This would most likely provide more accurate short-term simulations of the snowpack. However, as the objective of the study was to consider long term trends in average snow conditions given different emission scenarios from many models and several of the variables required for estimating the SNOW-17 parameters were not available at high spatial resolution (e.g. net SW) the model resolution was set at 0.2-degree. The parallel nature of the computing platform used in the study was also a limiting factor for model resolution when applied across centennial timescales for each member in a large ensemble of climate models.

2021; Liu et al., 2021) such as the gauge under-catch factor and the range on the melt factor. This could have helped to constrain uncertainty in the SNOW-17 model and the forcing. It was also odd that the GDDP forcing data was treated almost as a reanalysis dataset during the historical period rather than as (downscaled and bias-corrected) climate model output. In particular, this data is meant to represent the the correct climate on decadal timescales but not the day to day weather or even inter-annual variability. As such it is not fair to compare observations directly to SNOW-17 forced with this dataset. For such an exercise it would have made more sense to use actual reanalysis data, such as ERA5, to force the SNOW-17 model. Such simulations could be compared directly to the observations. Furthermore the climate (i.e. decadal moving averages) from the ERA5-driven simulations could be compared to the GDDP-driven simulations in the historical period (1950-2021) to gauge the performance of or bias correct the latter (Fiddes et al., 2022). I also had some concerns about the way the MODIS data was processed, particularly that no gap-filling was conducted before calculating SCF, despite previous work on such methods by some of the authors (Gunnarsson et al., 2019). This should at least be justified in the text. As such, I suggest that this manuscript should undergo major revisions. Nonetheless, I would like to commend the authors for this valuable work that fits well within the scope of The Cryosphere and I encourage them to address these general comments and the specific comments below.	Furthermore, as the objective of the study was to study trends in long term snow conditions under a changing climate, the model was evaluated on its ability to simulate variability and average snow conditions on a decadal time scale. Figure 4 (now Figure 5) and the text describing it has been updated to highlight this point. We agree that it would be an interesting research effort to apply the SNOW-17 model to reanalysis data such as ERA5 or GLDAS-2 and also at higher spatial resolution and that both should be addressed in future works. However, we feel that the extent of such a study would merit its own publication. We do note that reanalysis data from the GLDAS-2 dataset was used for parameter estimation, so weather data was used to adapt the model to the study area, this has been clarified in the text. We agree that the gap-filling product by Gunnarsson et al., 2019 provides an excellent analysis of plausible snow coverage. However, the approach taken in this study was to use the underlying research data as published without further manipulation such as e.g. interpolation of missing data. This has been clarified in the text. Please see responses to the specific comments below.
Specific comments	
L9: Change “remote sensing observations” to “remotely sensed observations” (replace all).	This correction has been made throughout the manuscript
L10: Use either “General Circulation Model” or “Global Climate Model”	„General Circulation Model“ is now used throughout the manuscript

(preferably the former) but not “Global Circulation Model”.	
L11: One immediately wonders why the CMIP6 version of this dataset was not used. This is not a criticism per se, but the existence of a newer version of the NEX-GDDP dataset should at least be mentioned somewhere in the manuscript.	Good point. A mention of this has been added in the manuscript.
L13: Representative Concentration Pathway 4.5 (RCP4.5) indicates a scenario where the radiative forcing will be 4.5 Wm−2 at the end of the century. So please use the “.”, it’s RCP4.5 not RCP45. The same holds for RCP8.5.	Correct. „.“ Have been added to all mentions of RCP scenarios in the text.
• L13: Change “Snow17” to “SNOW-17” in line with Anderson’s naming convention (replace all).	This correction has been made throughout the manuscript
L20: Suggest changing “climate” to “micro-climate” also change “significantly” to “strongly” since significant has a specific statistical meaning in your manuscript.	Good points, both corrections have been made.
L21: Change “correlate to changes in” to “are highly correlated with changes in”.	This has been changed.
L30: Change “at least mid” to “at least the middle of the”.	This has been changed
• L33: Suggest changing “duration of snow cover” to “snow cover duration” which is perhaps more widely used (e.g. Notarnicola, 2020).	We agree, this change has been made
L34: Best in what sense? A more qualified statement would be to say that a good balance of spatial and temporal resolution as well as temporal coverage is a valid reason for using MODIS. You could also consider citing a paper (e.g. Aalstad et al., 2020, and references therein) that evaluates satellite-based snow-covered area products from MODIS and give an idea of what uncertainty you would expect at the pixel scale.	We agree. This has been revised in the text (L40-41)
L40: Change “and is” to “which is”.	Changed
• L41: Change “earths” to “Earth’s”.	Changed
• Fix typo “thatasc”.	Changed
• L45: This could be a good place to mention other studies that have performed detailed future snow cover projections under different scenarios (e.g. Fiddes et al., 2022, and references therein).	Good point. The introduction has been restructured and extended to include a discussion of other studies of future snow cover projections.
L47: Change “Snow Covered Area” to “Snow-Covered Area”.	This has been changed

• L49: It would be natural to mention that SNOW-17 is a conceptual model built around the degree day (also known as the temperature index) approach. Currently this is not mentioned anywhere in the manuscript. Since it is a degree day model, it is not only more efficient but it also has less requirements in terms of forcing data than full energy balance models. On the other hand, given that the entire energy balance is lumped into a single term, one would think that calibrating the degree day factor is critical for the model.	Good points. The end of the paragraph in question has been revised to illustrate these points. We agree that calibration is beneficial for the optimal performance of any model in simulating the short-term response of a system to its environment. Here, the model parameters are a function of its environment which is expected to change over the long term, therefore, the initially calibrated model would be rendered biased over longer timescales, see e.g. Melsen and Guse, 2021. Therefore, the model parameters were estimated using best practices from globally available and comparable environmental datasets.
L51: Change “and it” to “. It”.	Changed
L55: Change “calculated” to “estimated” since this is an approximation.	This has been changed
Section 2.1.2: The formulation in this section should be improved. A reader that is unfamiliar with these MODIS snow cover products could think that ’NDSI_Snow_Cover’ is meant to measure Fractional SnowCovered Area (FSCA) directly. This is not the case, instead the valid pixels in the ’NDSI_Snow_Cover’ field merely contains the NDSI value (scaled by 100, see e.g. Riggs and Hall (2020)) for pixels that could possibly contain snow (positive NDSI) or a value of 0 for pixels that probably do not contain snow (NDSI≤ 0) and have passed various screens (not deemed to be cloudy etc. . .). This is readily verified by comparing it to the ’NDSI’ field which will have equal values once negative values are set to zero and missing (cloudy etc...) pixels are masked out. The NDSI is of course related to the FSCA of a pixel, but they are not identical and converting NDSI to FSCA usually involves some form of transformation. In particular, it is often the case that even pixels with an intermediate positive NDSI value (say equal to 0.6) can be fully snow-covered (FSCA=1). It would be worth making this clear to the reader and discussing the commonly used linear relationship between the two (Salomonson and Appel, 2006; Fiddes et al., 2019; Alonso-González et al., 2021), along with some uncertainty estimates form the literature (Aalstad et al., 2020,	Good points. Section 2.1.2 has been restructured and revised to include the clarification between NDSI and FSCA and a discussion of commonly used transformations there between.

and references therein). Indeed an advantage with the C6 (versus C5) MODIS snow cover products is that users can customize the NDSI-FSCA relationship for their own use case (see the MODIS Snow Products Collection 6 User Guide user guide). Work with Sentinel-2 (Gascoin et al., 2020) has shown that other types of NDSI-FSCA relationships (e.g. sigmoid) can also perform well	
L73: How much did glacier outlines change during the MODIS era (2001-2021)? One would assume that this effect is small, but make it explicit that you assume these outlines to be constant in time for this period.	The changes to the glacier outlines are observed in our results, shown in Figure 3a (Now Figure 4a). Glacier outlines were not used for modelling or analysis as the land surface was assumed to be constant. The reference to this dataset in the data section was inappropriate as it is only used for illustration in Figure 3 (Now Figure 4).
• L78: It would be nice to have at least a rough idea of what areal scales local snow cover and surrounding mountain snow cover represent? Moreover, the term "snow cover" is very general (see the NSIDC glossary) and a bit vague in this context in that it can implicitly refer to many different snow variables. You clarify your use of the term a bit later (L82) as being some kind of a snow cover classification (snow/patchy/no-snow) that lies between binary snow cover (snow/no-snow) and FSCA. I would still recommend calling it something more specific than "snow cover" such as "snow cover status" or similar	Good points. Section 2.1.1. has been revised to describe the scale the measurements of SNC and SNCM are designed to represent. The text was revised throughout to use the term „snow cover status" when referring to the SNC and SNCM snow cover observations.
L91: This sentence seems unnecessarily convoluted. Based on your classification system, an equivalent but more concise way to define "snow-covered ground" would be with the inequality SNC> 0 (or equivalently SNC≥ 2 since there is no class SNC= 1) and SNCM> 0 for local and surrounding mountain snow cover, respectively.	Agree and good point. This sentence has been revised and the classification system changed as per the referee suggestion.
L99: This is implicitly assuming that NDSI> 0 corresponds to FSCA> 0. While this is often true, in the sense that low but positive NDSI in a pixel can be due to a patchy snow cover, you may have some false positive 'snow-covered' pixels as a result of this, since some non-snow surfaces may also have a low but positive	We agree that assuming NDSI > 0 = FSCA > 0 holds the potential to include false negatives. We recognize the potential of an NDSI threshold value to exclude some valid observations. In this study we made the approach to analyze the wealth of data available without further data manipulation to the data as it is published by the

NDSI. Did you look at the sensitivity of your results to this threshold, for example by trying a threshold of NDSI> 0.1 instead which may be less sensitive to such false positives. Moreover, you could quantify the false positive rate by for example comparing your MODIS-based snow cover status classification to that from the station measurements. This could help to calibrate the NDSI threshold. Perhaps it turns out that 0 is good choice of threshold threshold, but testing this can help to strengthen your analysis and the resulting conclusions.	original producers of the dataset, which has already undergone their internal QA. Gunnarsson et al 2019 performed a validation of MODIS snow cover in Iceland using manned observations and higher spatial resolution remote sensing data (Landsat/Sentinel 2). Overall, a good agreement was found between the daily combined MODIS Terra／Aqua data set and the validation data sets from Landsat 7/8, Sentinel 2 and in situ observations in Iceland. To account for outliers in the MODIS snow cover data the Sens slope method was applied for trend analysis on the MODIS data as it is less sensitive to outliers and has shown good performance in snow cover trend analysis with the presence of outliers (Eythorsson et al., 2019). The text in section 2.2.4 has been extended to clarify this point.
L102: By excluding December and January from the analysis your SCF may end up incurring a negative bias (i.e. be underestimated). In particular, the months of December and January often have a higher SCF than many other months in the year. Since this is likely to be the case, your results may (at least on average spatially) end up being closer to the true annual SCF if you gap-filled your observations using either simple interpolation techniques or more sophisticated algorithms. More generally, applying gap-filling could also make your results more robust to a potentially uneven temporal distribution of cloudiness throughout the year. Given the large overlap in authorship, I found it surprising that you did not use the gap-filling method of Gunnarsson et al. (2019) which has already been applied successfully over your domain.	In this study we made the approach to analyze the wealth of data available in the area without performing any manipulation to the data from the original producers of the source datasets. We acknowledge that systematic phenomena such as the polar night may introduce a negative interannual bias but recognize that the interpolation across large periods would also be a source of uncertainty in the SCF estimate. As the study objective was to study and project long term trends in snow conditions under different climate scenarios and not investigate or develop methods for interpolating observational gaps to derive more plausible gridded snow cover datasets, we opted for using the unmanipulated data. This has been clarified in the text in Section 2.2.4
To properly test the performance of SNOW-17 given weather (rather than climate) forcing I would strongly encourage the authors to also run SNOW-17 for the historical period (1950-2021) with forcing from the ERA5 reanalysis. Note that this product is also available on Google Earth Engine, so it should be relatively straightforward to extend your	We agree that testing the performance of the SNOW-17 model given weather forcing data such as e.g. those in the ERA5 or GLDAS-2 reanalysis products, both of which are available in the GEE data catalog would be interesting. However, as the intent of this study was to investigate future snow conditions in Iceland given

analysis. This would be a fairer model simulation to compare with in-situ observations, which (disregarding scale-mismatches) experience (roughly) the same weather as in ERA5 rather than the weather simulated by GDDP which aims only to produce the correct climate (i.e. weather statistics). This ERA5 simulation would also a reference at roughly the same spatial scale as your existing GDDP-forced simulations that you could use to validate the climatic evolution (i.e. at the decadal timescale) of the snowpack in the historical period.	different climate scenarios, we opted to gauge the model performance in replicating the long-term average behavior Icelandic snow conditions. Additionally, we note that the ERA5 daily aggregates published in the GEE data catalog extend from 1979 to the present and not from 1950, thus not covering as much of the historical period. (see https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_DAILY#description)
L117: Change “of trend” to “of trends”.	This has been changed
L119: It could be more instructive to cite studies that have applied similar trend analysis methods for snow cover such as Yılmaz et al. (2019) and Notarnicola (2020).	Good point. The citations have been revised.
• L120: This statement is misleading. When doing null hypothesis significance testing you are not testing the alternative hypothesis. You are not even really testing the null hypothesis. For more about the caveats of significance testing and the error of the transposed conditional see Ambaum (2010). To be more specific, I would urge the authors to explain what their p-values actually quantify.	We agree that this sentence was unclear. Section 2.2.4 was revised, the statement removed, and an explanation of the trend test p-values was added.
Figure 2: Please improve the resolution and visibility of this figure. For example, text should not be visibly pixelated and it should be easy to differentiate markers without zooming excessively	The size, resolution and visibility of the Figure 2 has been improved.
Suggest changing “ensemble average” to “ensemble mean” which is perhaps more commonly used.	We agree, this has been changed throughout the manuscript.
• L126-130: This is essentialy a repetition of the figure caption. Please shorten considerably, this should only briefly describe the results that the figure shows.	All figure and table captions have been revised for brevity and clarity.
L141 Change “trended upward” to “a positive trend”.	Changed
L142 According to your earlier description the inequality describing full and patchy snow cover should be SNC≥ 2 not SNC> 2.	Correct, this has been corrected.
L146 Instead of speculating about what the increase in precipitation could lead to in terms of snow accumulation, you could analyze the changes in snowfall and	Good point. We have added a new analysis, presented in the new Figure 3 and text thereof to illustrate the effect of increased precipitation and

rainfall rates to a first order by applying simple air temperature-based thresholds to delineate precipitation phase.	temperature on the snow/rain temperature given a partition threshold.
Table 2: Please use × for multiplication not ∗. Also fix "of p-values Statiscally significant. . . " in the caption. Moreover, I would recommend setting all the text in the table to normal font apart from significant p-values which can be in bold.	Good points. All of these changes have been made.
Figure 3: Please center the colormap around 0.	Figure 3 (Now Figure 4) has been revised and the colormap centered around 0.
L176: Once more you are just repeating the caption. Please shorten considerably and avoid redundant text as much as possible.	All captions of tables and figures have been revised.
Figure 4: To make the long-term weather observations and the GDDP-driven climate simulations of SCF more comparable it would be natural to instead plot the decadal moving mean (serving as a low pass filter) of the observations. The corresponding moving standard deviation would then help to visualize the observed internal climate variability. In that way, you would be more fairly comparing observed climate to the simulated climate. In particular, as previously alluded to, GDDP is not a reanalysis so we can not expect it to reproduce the correct weather	Good point. Figure 4 (now Figure 5) has been updated to show the decadal moving average of the observations and the text describing it has been revised and expanded. The revisions highlight the fact that the objective of the study was to study long term climate impacts on Icelandic snow conditions.
• L190: Remove "fig".	This sentence has been revised
L192: Why aren't the observed and simulated trends compared for an overlapping period, such as 1950-2021?	The referee seems to have misunderstood. L192 refers to the MODIS observations which are only available since 2001. Figure 4 (Now Figure 5) shows the IMO observations over the period compared to simulations for the period 1950-2021 and the MODIS observations for the period 2001-2021 The entire text around Figure 4 (Now Figure 5) has been revised for clarity on this and other points.
L204: What does "increase in snow cover" mean here? Are you referring to frequency, duration, depth, SWE? Please be more specific.	Good point. This sentence has been revised to refer specifically to SCF and 1st April SWE.
L205: This is not shown clearly in Figure 2, although maybe it will be easier to see when the Figure is sharpened. I would also recommend to add a panel to Figure 2 containing a scatter plot that compares	Figure 2 has been redone for better visibility.

the MODIS SCF to in-situ SCF observations for both local and mountain snow cover.	
L210: Please make it clear here that although the simulated SCF magnitude is comparable to the observations, the trend is in the opposite direction.	This has been clarified at this point and in the revised discussion Figure 4 (Now Figure 5).
L213: It's of course not straightforwards to directly compare SWE to snow depth given variations in snow density. Moreover, Figure 2 shows annual mean observed snow depth whereas Figure 4 shows simulated April 1 st SWE. Please clarify what you are comparing here. While the unrepresentative (mainly) low-lying station locations can account for some of the differences, you could perhaps make the simulations more comparable to the observations by excluding higher elevation grid cells	We agree that this comparison was unclear. This paragraph has been revised as well as the discussion of these results in text around Figures 4 (now 5) and 2.
L218: Once more "increase in snow cover" is vague, please explain what is meant by this	Agree, this has been changed to „increase in SCF"
L224: Change "predictions which forecast" to "projections of ".	Changed
L229: Change "parameters" to "variables".	Changed
L230: Although the acronym GHG is well-known, this is the first time it is used in this manuscript. Consider writing it out in full or just change "GHG emissions" to "emission scenarios".	Good point. This has been changed to „emission scenarios"
Code/Data availability: Please provide references (with DOIs) or link to all the datasets used.	Section 7 has been revised and extended. A table has been added with links to all underlying data used.