Review of Eythorsson et al. (2022; doi:10.5194/egusphere-2022-590)

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., 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.

Specific comments

- L9: Change *"remote sensing observations"* to *"remotely sensed observations"* (replace all).
- L10: Use either *"General Circulation Model"* or *"Global Climate Model"* (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.
- 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.
- L13: Change *"Snow17"* to *"SNOW-17"* in line with Anderson's naming convention (replace all).

- L20: Suggest changing *"climate"* to *"micro-climate"* also change *"significantly"* to *"strongly"* since significant has a specific statistical meaning in your manuscript.

- L21: Change *"correlate to changes in"* to *"are highly correlated with changes in"*.

- L30: Change *"at least mid"* to *"at least the middle of the"*.

- L33: Suggest changing *"duration of snow cover"* to *"snow cover duration"* which is perhaps more widely used (e.g. Notarnicola, 2020).

- 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.

- L40: Change *"and is"* to *"which is"*.

- L41: Change *"earths"* to *"Earth's"*.

- Fix typo *"thatasc"*.

- 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).

- L47: Change *"Snow Covered Area"* to *"Snow-Covered Area"*.

- 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.

- L51: Change *"and it"* to *". It"*.

- L55: Change *"calculated"* to *"estimated"* since this is an approximation.

- 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 Snow-Covered 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$\leq$ 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, 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.

- 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.

- 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.

- 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 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.

- 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.

- 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 straightforwards to extend your 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.

- L117: Change *"of trend"* to *"of trends"*.

- 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).

- 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.

- 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.

- Suggest changing *"ensemble average"* to *"ensemble mean"* which is perhaps more commonly used.

- 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.

- L141 Change *"trended upward"* to *"a positive trend"*.

- L142 According to your earlier description the inequality describing full *and* patchy snow cover should be SNC$\geq$ 2 not SNC$>$ 2.

- 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 rainfall rates to a first order by applying simple air temperature-based thresholds to delineate precipitation phase.

- Table 2: Please use $\times$ for multiplication not $*$. Also fix *"of p-values Statiscially 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.

- Figure 3: Please center the colormap around 0.

- L176: Once more you are just repeating the caption. Please shorten considerably and avoid redundant text as much as possible.

- 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.

- L190: Remove *"fig"*.

- L192: Why aren't the observed and simulated trends compared for an overlapping period, such as 1950-2021?

- L204: What does *"increase in snow cover"* mean here? Are you referring to frequency, duration, depth, SWE? Please be more specific.

- 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 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.

- 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.

- L218: Once more *"increase in snow cover"* is vague, please explain what is meant by this.

- L224: Change *"predictions which forecast"* to *"projections of"*.

- L229: Change *"parameters"* to *"variables"*.

- 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"*.

- Code/Data availability: Please provide references (with DOIs) or link to all the datasets used.

I'm looking forwards to reading the revised manuscript,
Kind regards,
Kristoffer Aalstad

References

Aalstad, K., Westermann, S., and Bertino, L.: Evaluating satellite retrieved fractional snow-covered area at a high-Arctic site using terrestrial photography, Remote Sens. Environ., doi:10.1016/j.rse.2019.111618, 2020.

Alonso-González, E. et al.: Snowpack dynamics in the Lebanese mountains from quasi-dynamically downscaled ERA5 reanalysis updated by assimilating remotely sensed fractional snow-covered area, Hydrol. Earth Syst. Sci., doi:10.5194/hess-25-4455-2021, 2021.

Ambaum, M. H. P.: Significance Tests in Climate Science, Journal of Climate, doi:10.1175/2010JCLI3746.1, 2010.

Fan, Y. et al.: Hillslope hydrology in global change research and Earth system modeling, WRR, doi:10.1029/2018WR023903, 2019.

Fiddes, J., Aalstad, K., and Westermann, S.: Hyper-resolution ensemble-based snow reanalysis in mountain regions using clustering, Hydrol. Earth Syst. Sci., doi:10.5194/hess-23-4717-2019, 2019.

Fiddes, J., Aalstad, K., and Lehning, M.: TopoCLIM: rapid topography-based downscaling of regional climate model output in complex terrain v1.1, Geosci. Model Dev., doi:10.5194/gmd-15-1753-2022, 2022.

Gascoin, S. et al.: Estimating Fractional Snow Cover in Open Terrain from Sentinel-2 Using the Normalized Difference Snow Index, Remote Sens., doi:10.3390/rs12182904, 2020.

Gunnarsson, A., Garðarsson, S. M., and Sveinsson, O. G. B.: Icelandic snow cover characteristics derived from a gap-filled MODIS daily snow cover product, Hydrol. Earth Syst. Sci., doi:10.5194/hess-23-3021-2019, 2019.

Liu, Y., Fang, Y., and A, M. S.: Spatiotemporal distribution of seasonal snow water equivalent in High Mountain Asia from an 18-year Landsat–MODIS era snow reanalysis dataset, The Cryosphere, doi:10.5194/tc-15-5261-2021, 2021.

Notarnicola, C.: Hotspots of snow cover changes in global mountain regions over 2000–2018, Remote Sens. Environ., doi:10.1016/j.rse.2020.111781, 2020.

Riggs, G. and Hall, D.: Continuity of MODIS and VIIRS Snow Cover Extent Data Products for Development of an Earth Science Data Record, Remote Sens., doi:10.3390/rs12223781, 2020.

Salomonson, V. V. and Appel, I.: Development of the Aqua MODIS NDSI Fractional Snow Cover Algorithm and Validation Results, IEEE Trans. Geosci. Remote Sens., doi:10.1109/TGRS.2006.876029, 2006.

Yılmaz, Y. A., Aalstad, K., and Sen, O. L.: Multiple Remotely Sensed Lines of Evidence for a Depleting Seasonal Snowpack in the Near East, Remote Sens., doi:10.3390/rs11050483, 2019.