The Cryosphere
Discussions

Open Access

EGU

Interactive comment on "Historical Northern Hemisphere snow cover trends and projected changes in the CMIP-6 multi-model ensemble" *by* Lawrence Mudryk et al.

Lawrence Mudryk et al.

lawrence.mudryk@canada.ca

Received and published: 24 April 2020

Thanks to both reviewers for their constructive comments on the manuscript. Changes prompted by these comments have improved the clarity and impact of the paper. Please note that we found and corrected an error in the selection of model simulations analyzed for future projections which affected Figures 8, 9, and 10. Previous figures were being generated using only a single model (which had metrics close to the ensemble median); the updated figures are based on all available models. None of the conclusions were affected, however we performed some additional analysis related to the seasonality of snow sensitivity (Figures 9 and 10) which was incorporated

as additional commentary as shown in the tracked-changes version of the manuscript. Our response to each comment is outlined below in **bold**. Revised text is in *red italics.*

General Comments:

Overall I thought this was a good paper and will make a good addition to the literature, especially since snow is a critical aspect of the natural environment (with obvious links to transportation, recreation, water supply, etc.) and is projected to decline with continued anthropogenic climate change. I have only minor comments, as enumerated below.

Line 77: Poor spatial resolution likely plays a key role over mountainous areas but is not explicitly mentioned here.

Good point. We added:

Despite increases in model resolution since CMIP-3, even the highest resolution models in CMIP-6 are still expected to have challenges simulating snow in mountain regions without appropriate downscaling (Verfaillie et al., 2018).

Line 108: Please note somewhere, for example in Table 1, the resolution of the gridded snow mass data sets before regridding to 0.5 degree resolution.

Resolution information was added to Table 1.

Line 112: Please provide information that would enable the reader to understand to what extent this weighting/mixing was performed. For example, what percent of active grid cells had a GlobSnow weight of 0.5 or less?

We added the following line to this section:

In midwinter, approximately 25% of snow-covered grid cells are blended with less than 50% GlobSnow.

Line 117: 4 mm seems somewhat arbitrary. Are your results sensitive to using a differ-

ent value? For example, does figure 3 only look good because 4 mm was used?

The snow extent climatologies of the component datasets do depend on the choice of SWE threshold, however only their variability is incorporated into the mean time series. The climatology is that of the NOAA/Rutgers dataset, which we assume to be more accurate than the reanalyses and GlobSnow. This is in contrast to the mean snow mass time series, for which no preferred climatology exists. The snow extent anomalies of the component data sets are also affected to a smaller extent by the choice of SWE threshold, however the difference between a 4mm and a 10mm threshold affects the mean trends to less than about 5%. Threshold choices less than 2mm will begin to affect things more substantially and are probably unrealistic (see Figure 10, for example, from the following reference): *Krinner et al., ESM-SnowMIP: assessing snow models and quantifying snow-related climate feedbacks, Geosci. Model Dev., 11, 5027–5049, 2018.*

Line 120: You lost me in this part about merging the data sets, not tactically (what you did) but strategically (why you did it). The initial motivation for using multiple datasets was that any one data set is uncertain, and by using multiple data sets you take the uncertainties in observational estimates into account. But now you are removing some of the differences between the data sets, which seems to obviate the motivation for using multiple data sets in the first place. I suggest you augment this section with an explanation of why you chose to take this approach given that one of your motivations was to look at the spread of results across observational data sets, and that this methodology seems to diminish that spread.

Thanks for bringing this up! We have added the following additional text to Section 2.1 which hopefuly provides a more complete rationale for the method:

The ideal dataset for comparison with coupled model ensembles would be a single, bias-free, observational record with uncertainty due only to interannual variability. Analysis of such a record would yield an accurate trend with a simply interpretable un-

certainty due to record length and sampling uncertainty and would provide the most straightforward way to compare the observed trend with those from coupled model ensembles (which contain further spread due to model structure and parametrization as well and internal variability). In reality, observational records are not ideal and different datasets may contain additional uncertainty in their variability due to technical issues. Our goal here is to combine anomalies from an ensemble of observation-based estimates in order to mitigate the most prominent trend biases of the individual datasets and estimate the irreducible interannual variability.

We also note here that because the focus of this paper was on evaluating CMIP6, we have refrained from a more thorough discussion of the full uncertainty/spread in trends among component gridded datasets. However we plan to present such an analysis in a subsequent contribution.

Line 148: Off the top of my head, I would have said that the *fractional* variability had likely not changed, but (along with the diminished snow cover in the recent period already seen) the actual variability might have. I.e., if you have more snow overall I'd expect first order that you'd have more variability, but perhaps the same fractional variability. Maybe it's just wording, but it would be good to be clear on this point. Presumably the actual amount of variability scales with the amount of snow, but the fractional variability would be more constant.

We have rephrased the original text, which was misleading. The standard deviation of the Brown index (as well as its climatology) is actually sampled over 1972-1991 and it is this value that is scaled to the ensemble average standard deviation sampled from the 1981-2014 period. Hence the two periods used are reasonably close and differences in the choice of period are covered by the additional uncertainty used to plot the blue shading (furthermore, most of this additional uncertainty is related to the choice of sample period used to match the NOAA climatology, rather than the sampling period for the standard deviation). The paragraph now reads:

C4

For certain months, the Northern Hemisphere snow extent time series can be extended back to 1922 through the interpolation of a fixed network of surface snow depth observations, as described in Brown (2000). These observations have sufficient spatial coverage to estimate relative changes in snow extent from year to year in the form of an index, but are not estimates of the actual snow extent. Limitations in the mutual availability and coverage of surface observations over North America and Eurasia mean that hemispheric estimates are only available for March, and April. March and April indices are converted into units of snow extent separately over each continent using the same process described above. Each index is standardized based on the 1972-1991 period. Then the index climatology is matched to that from NOAA sampled over the same period. We chose the 1972-1991 reference period because we expect that the NOAA climatology over 1972-1991 is more representative of earlier 20th century snow cover (whereas snow cover will have already responded to warming temperatures by the latter portion of the record). We rescale the variability of each index using the 6-member ensemble mean standard deviation for the appropriate month and continent that was used to compute adjusted time series for the other data sets. This process has an implicit assumption that the variability sampled in the Brown index over 1972-1991 is comparable to that from the ensemble mean SCE time series over the 1981-2014 period. The rescaled indices for each continent are then summed together to obtain a historical estimate of Northern Hemisphere snow cover extent from 1922-1991. Uncertainty in the Brown time series is calculated using the standard error of forecast, however we consider additional uncertainty due to the choice of reference periods used for matching the climatology and standard deviation. 95% uncertainty bounds are calculated for all possible selections of sequential 20 year climatological periods from the NOAA record in combination with all possible sequential 20 year estimates of variability. The maximum and minimum uncertainty bounds calculated from these combinations are used to define the range for each year.

Line 161: Although you're not wrong about the emissions following RCP 8.5 from 2006-2014 (mostly, but dropping below that towards the end), in fairness, the global mean

temperature between, say, RCP 4.5 and RCP 8.5 are statistically indistinguishable during this period.

We have rephrased the sentence to say observed emissions 'reasonably follow' those of RCP8.5.

Line 164: Typo, should say "SSP3-7.0", not "SSP3-3.7". **Changed.**

Line 166: Regarding "r1i1p1f1", note that the CMIP6 gateway has a warning right at the top of the page that states "WARNING: Not all models include a variant "r1i1p1f1" and across models, identical values of variant label do not imply identical variants! To learn which forcing datasets were used in each variant, please check modeling group publications and documentation provided through ES-DOC". If that is true, I'm not sure that it is necessarily correct to compare r1i1p1f1's across models. I suppose as the community gains more experience with CMIP6 this will sort itself out. I assume all the models you analyzed did, in fact, have an r1i1p1f1 member?

Sampling differences of climatological snow extent and snow mass across different realizations can be expected to be small. Specifying r1i1p1/r1i1p1f1 is a simple way to choose a single realization from each model keeping in mind that some groups only submit a single model. For trends the sampling differences will be much larger, which is why the median trend is selected from available model realizations as stated in the text. We have also added the following additional specification to the text:

"A handful of models had no r1i1p1f1 variant. In these cases the r1i1p1f2 variant was chosen (e.g. the CNRM models, MIROC-ESM, UKESM) and barring that the r1i1p1f3 variant (HadGEM). NorESM realization r1i1p1f1 had 10 years of missing data and was ignored for trends; realization r2i1p1f1 was used for its climatology."

Line 177: Would I be correct in presuming that models calculate their own snow cover fraction by applying a threshold to their SWE? If so, is there any reason to believe

C6

different models used the same threshold? If this is the case and different models used different thresholds, might it have affected your results by adding an element of inconsistency between the analyses of different models?

Yes, however this is a genuine parametrization choice made by different modeling center and bears scrutiny as examined in the Discussion Section.

Line 183: The wording does not make clear whether you used the CNRM-CM6-1 mask for all models, or just the subset of models that did not have the needed masks available. Please clarify. Additionally, if the CNRM mask was only used for a subset of models, please indicate somewhere (in text or in table 2 perhaps) which models used the CNRM mask instead of their own.

The CNRM masks were used throughout for consistency. We've rephrased the sentence to read:

We therefore used the masks from the CNRM-CM6-1 coupled model (Voldoire et al., 2019) for all other models by interpolating these masks to the corresponding grid of each model.

Line 191: "IPSL CICLAD computer center" is not sufficiently descriptive. Please augment the acronyms with more specificity, such as spelling it out, location, and institution. . **Acronyms specified.**

Line 261: "there is clear improvement in mean model agreement with observations at the hemispheric scale..." I presume this is based on Fig. 3, not figure 4 (which shows a degradation in model agreement with observations) or Fig. 5 (which does not show observations at all)? If so, please be a little more thorough on line 261 by noting that it is snow cover extent that shows clear improvement while snow mass shows less agreement with observations.

Fair point. This has been changed to:

In summary, while some regional differences in SCF are evident, the improved hemi-

C7

spheric statistics for snow extent seen in Figure 3 are generally consistent with the geographic patterns shown in Figure 5.

Please note, however that observations are shown in Fig 5 with pink contour.

Line 308: Does each dot in figure 9 represent one year averaged across all models, or what? Please add a bit more clarity to the text or caption. **We have changed the part of the caption to the following:**

Each data point represents the average snow extent from one CMIP-6 simulation (first ensemble member for each available model). The average is taken over all pentads from the given simulation which fall into a given temperature bin.

Line 371: Please add an X axis label to Fig. 11.

Changed.

Throughout: I suggest using "autumn" rather than "fall", since "snowfall" is a different thing and too similar to "fall snow".

Changed.

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2019-320, 2020.