Bias-adjusted projections of snow cover over eastern Canada using an ensemble of regional climate models

Authors' response (RC2)

Bresson et al.

We thank the reviewer for reviewing the manuscript and their constructive comments. We will revise the manuscript to clarify the selection of the simulations for the ensemble and the methodology for bias adjustment, and homogeneize the lake mask. We address each specific comment in detail below.

Major comments:

1. Simulation selection: In the Data section (2.3), the authors describe the process of excluding some model realizations from their ensembles. My understanding is that this was done before the main analysis of the study because of the bias-correction step. If this is the case, I think section (4.2) needs to be moved and combined with (2.3). I would also suggest a supplementary plot showing the selection variables compared to ERA5-Land for each model/realization considered (before exclusion). This would support the claim that a "first scan of the simulated SWE showed large discrepancies".

The simulation selection was indeed performed before the main analysis, as we performed the bias adjustment only on the set of 10 simulations.

Figure 1 (SSS) below will be added to the Supplementary Material, as well as the same figure for SSE: "A first scan of the simulated SWE from RCMs showed large discrepancies (Fig. S1 and S2), as already highlighted in the literature (e.g. McCrary et al., 2017)".

As well, Section 2.3 will be completed with Section 4.2.

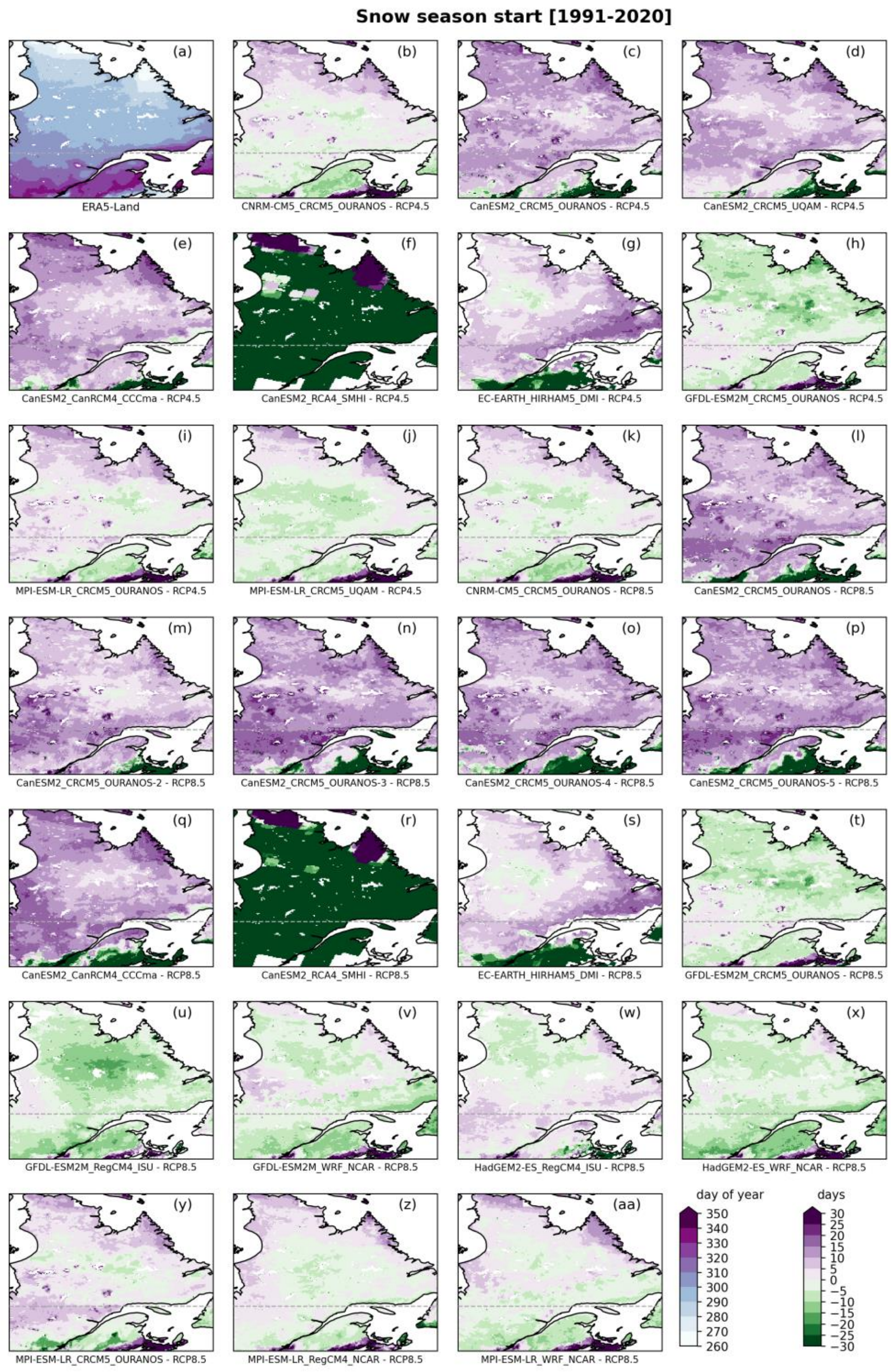


Figure 1. Annual mean of snow season start for ERA5-Land (a) for 1991-2020. Difference between ERA5-Land and simulations' annual mean of snow season start for 1991–2020 (b-aa). Results are presented for all available simulations, before applying the selection process presented in Section 2.3.2.

a. Additional comment related to this: the stippling in figures represents 80% agreement among models, but there appear to be 4 RCP4.5 simulations chosen and 7 RCP8.5 simulations chosen (Table S1). 80% therefore corresponds to consensus in the RCP4.5 simulations but agreement in either 6 or 7 out of 7 RCP8.5 simulations. I request that the authors clarify this point and discuss the implications.

Consensus between the simulations is reached when 80 % of the simulations agree on the sign of change. The 80 % threshold is based on IPCC work. In our case, the consensus is reached for RCP4.5 ensemble when 4 simulations out of 4 (100 %), and for RCP8.5 ensemble when 5 out of 6 (83 %) simulations agree on the sign of change. This imbalance needs to be kept in mind, as this consensus criterion is stricter for RCP4.5 than for RCP8.5 ensemble. There are no noticeable changes in the consensus when the threshold is set to 75 % instead of 80 %, except for SWEmax that presents more consensus in the region where the sign of the difference in future and past climatology changes with RCP4.5 ensemble (Fig. 2 and 3). This information will be included in the revised manuscript.

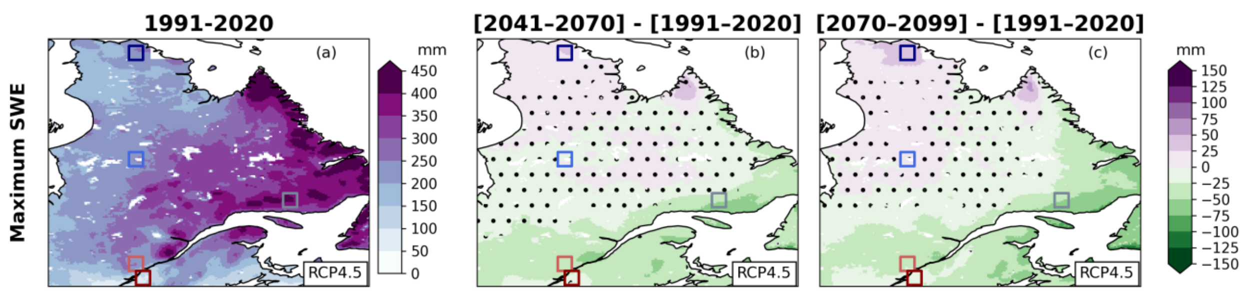


*Figure 2. Annual maximum SWE for 1991–2020 (a), and differences between the 1991–2020 period and 2041–2070 (b) or 2070–2099 (c). Dotted areas are those where less than **80 %** of the simulations in the ensemble agree on the sign of change. Results are presented for the RCP8.5 ensemble.*

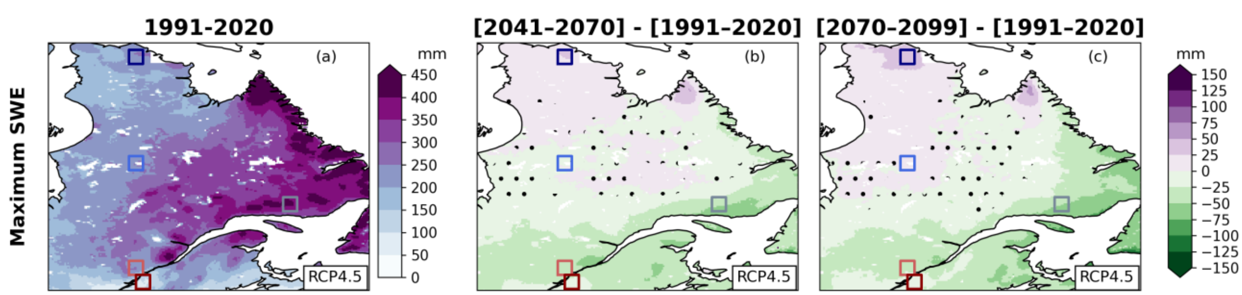


*Figure 3. Annual maximum SWE for 1991–2020 (a), and differences between the 1991–2020 period and 2041–2070 (b) or 2070–2099 (c). Dotted areas are those where less than **75 %** of the simulations in the ensemble agree on the sign of change. Results are presented for the*

RCP8.5 ensemble.

2. Total number of years: Given that ERA5-Land has a long back-extension (to 1950s), would there be value to using a longer training period for the statistical bias correction?

Extending the reference period over 30 years introduces a new issue where the climate change signal becomes apparent within the training dataset. It can be taken care of by detrending the reference data, but this comes with new complexities. We decided to keep the method simpler for this version of the dataset and only took the last 30 years. Doing so, we make the hypothesis that the variability in the reference data is assimilable to the natural variability. For sure, climate change does have an impact over the 1991-2020 period, but we assume here that this impact is not statistically significant in comparison to the natural variability.

3. Method description: There are several aspects of this study that need to be more clearly outlined within the paper.
   a. Was SWE data from the ERA5-Land reanalysis used for bias-correction? Is the model output hourly or daily? If different, how is this reconciled?

Yes, we used the SWE data from ERA5-Land. Sub-daily timeseries were resampled with a mean to a daily time step. We will add this information to the "Statistical Downscaling and Bias-adjustment Method" section.

   b. Was the full Michel et al. (2024) algorithm used? If it was, what data was used for the clustering variables (slope, aspect, altitude…)? If not, how was the EQM applied (how many quantiles were estimated, is there sufficient data to separately do this at each grid cell)?

No, we didn't use the full Michel et al. (2024) algorithm. We didn't explore spatial clustering. We used temporal clustering with a window of 15 days around each day-of-the-year. We now emphasize that we first and foremost adapted the decay procedure from this work. 48 quantiles were used, with less quantiles in the low tail to ensure a better stability in the bias-adjustment procedure. We will modify the text accordingly to include these details and better distinguish our method from Michel et al. (2024).

   c. Clarify the SWE replacement method (L138-143). This is the algorithm I imagine from reading: find the end of the snow season using a 1 mm threshold (by finding the last >1 mm value before 14 consecutive days below 1 mm); replace all values after that date by half the previous day's value.

Yes, exact. We also define a minimal value, and values of SWE below this minimal value are re-transformed to 0 at the end of the pre-processing step. We made adjustments to the text to clarify this point. An example is now shown in Figure 4.

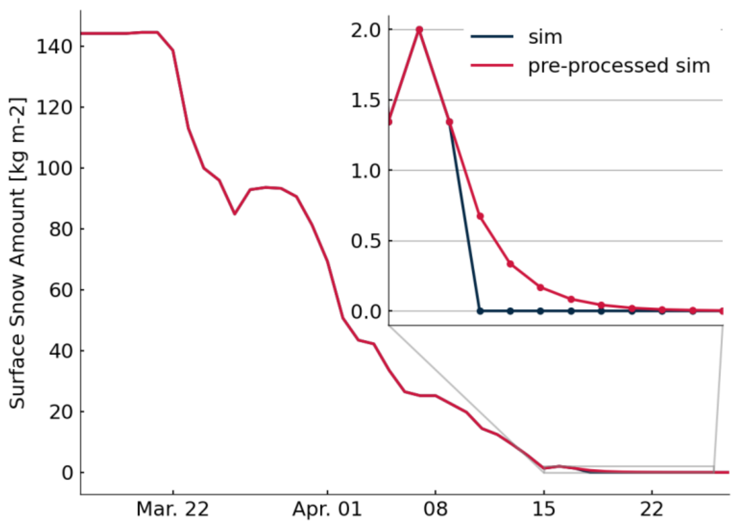


Figure 4. An example of the impact of the pre-processing on a simulation timeseries at the end of a season for a specific year. A sudden decay of SWE in the original simulation (sim) is smoothened with an exponential decay (pre-processed sim).

d. If half of 1 mm is replaced in the dataset up to a 0.001 mm cutoff, that yields roughly 6 more days of nonzero snow, but in the model selection stage, the authors look for models for which SSE can differ by up to 15 days. What impact could this have on bias correction?

By starting from 0.5 mm on the first day and decreasing by factors of ½, the tail will consist of 9 days (including this first value 0.5 mm). The snow-season-end computed in the pre-selection of models uses a 4 mm threshold, so it's not directly comparable. But, it is indeed the case that the method may not correct completely an early end of season, because not enough days are added. We opted for a more conservative approach. In the same way, we only added snow following the first sequence of 14 days below 1 mm. If there was a small increase in SWE above 1 mm after that, it was not considered.

e. The authors mention in L230 that varying the threshold for snow cover is outside the scope of the study. I wonder, though, what the difference would be if another threshold (e.g. 2mm instead of 1mm) were used for the snow insertion before bias-correcting. Were there any tests done for this?

What we mention near L230 concerns the threshold (4 mm) used for snow season end indicators. The 1 mm threshold mentioned above is related to the snow decay procedure. It is indeed related to the computation of a snow season end, but this is only to ensure that small peaks of SWE following the season end (defined with the 4 mm threshold) are ignored. We did not test with a different threshold, but the differences were small if we removed this criterion altogether. The situation avoided through this procedure is not frequent in any case, and it’s mainly a good measure for consistency.

f. Could snow be added to the snow start season (before bias correction)?

The hypothesis behind the snow addition in the snow season end is that some simulations might have abrupt end of seasons. In this case, we force a smoother season end with a exponential decay. The idea is that the melting mechanism in the surface model may have failed. In the case of the snow season start, the main driving mechanism is snowfall. In this case, it is expected that the snow season can start with a large amount of snow following a large snowfall. In this case, it is less evident how the timeseries could be modified in a coherent way. The statistical methods work well to extend the snow season end by inflating small non-zero values, and we help this process by adding more small non-zero values if the fall was judged too abrupt. But in the snow season start, extending the snow season could mean that a large amount of snowfall needs to occur earlier, and this is not evident to handle. In short, adding snow in the start of the season would require a different justification and most likely a different implementation.

g. I echo the first reviewer's call to explain the details of the bias correction process.

We will add more details to the manuscript (see the answer to the first general comment raised by Reviewer 1).

h. How was Quantile Mapping applied to the projections (2041-2070 and 2070-2099)? In these cases, there is no corresponding "true" distribution to match (e.g. ERA5-Land). Does the method adequately account for the climate change signal? If a quantile delta mapping was applied, please explicitly state this and how it works for a variable like SWE.

We opted not to use the Quantile Delta Mapping. The reasoning behind this choice will be added to the manuscript. The justification is related to important changes that can occur in snow cover. The distribution of SWE in the training vs. in the future where adjustment must be applied can be very different. This will occur specifically for day-of-years where we pass from a snow-abundant to a snow-scarce regime. Using QDM, we may apply the adjustment factor found for a substantial SWE amount in the training to a near-zero quantity in the future. QDM is indeed designed to take into account a shift in the distributions, but we argue that it should not be used to compare a distribution that is mostly zero-inflated (the future distribution) to one that isn't. QM was a practical way to avoid this issue. It would be interesting to experiment with new methods to address this issue more thoroughly.

i. How does the bias correction potentially affect an index like noSCseq? How are zero SWE values within the snow season season corrected? This factor might impact snow projections where zero values become more frequent during the snow season in some regions.

As pointed out in the previous point, we used empirical quantile mapping. This means that adjustment factors are attributed to specific values of SWE seen during training in the biased dataset. The way low values of SWE are treated during the historical (i.e. the training) period is the same in the future. If the frequency of low SWE values increases, then the corresponding adjustment factor is applied more times. See previous point for more details. We have thoroughly tested indicators such as the snow season start and snow season end that similarly rely on getting the zeros of the distribution right. noSCseq is expected to behave similarly.

4. Editorial comment: I would suggest to revise the paper for tense consistency and some other minor English errors (L25 "precipitations" instead of "precipitation", L117 "experimented" instead of "experienced").

These will be done in the reviewed manuscript.

Figures and tables:

Section (2.1) Are there any lakes or water bodies resolved at the 0.1 degree scale for any of the RCMs? I see that lakes are masked in most figures.

ERA5-Land doesn't have data on lakes and oceans. Consequently, when bias adjustment is performed on the simulations, data over bodies of water are set to NaN.

Table 1: should include noSCseq, defined as the average duration of snow-free conditions starting on each day of the year (excluding 1 day and 14+ day snow-free conditions).

A line in the table about noSCseq will be added.

Name	Definition	Unit
SWEmax	Annual maximum SWE	mm
SSS	Snow season start: first date on which SWE $\geq$ 4 mm for 14 consecutive days	day of year
SSE	Snow season end: first date (after SSS) on which SWE $<$ 4 mm for 14 consecutive days	day of year
SSD	Snow season duration: number of days between SSS and SSE	days
noSC	Number of days without snow cover (days with SWE $<$ 4 mm)	days
noSCseq	Average duration of snow-free conditions starting on each day of the year (excluding 1 day and 14+ day snow-free conditions)	days

Table 2: include resolution, otherwise unclear why MPI-ESM-LR-CRCM5 would be different when produced by a different institution.

A column with the resolution will be added. Indeed, the resolution is the main difference between both simulations with CRCM5 driven by MPI-ESM-LR.

Simulation name	RCP4.5	RCP8.5	Resolution
CNRM-CM5_CRCM5_OURANOS	X	X	0.22°
GFDL-ESM2M_CRCM5_OURANOS	X	X	0.22°
GFDL-ESM2M_WRF_NCAR		X	25 km
HadGEM2-ES_RegCM4_ISU		X	25 km
MPI-ESM-LR_CRCM5_OURANOS	X	X	0.22°
MPI-ESM-LR_CRCM5_UQAM	X		0.44°
MPI-ESM-LR_RegCM4_NCAR		X	25 km

Figure 2: the second and third columns have colours which are hard to see, might suggest cutoff of +/- 15% instead of 20%. Lakes are not masked in panel (g).

The intervals will be modified as follows. Lakes will be masked in every map of the manuscript.

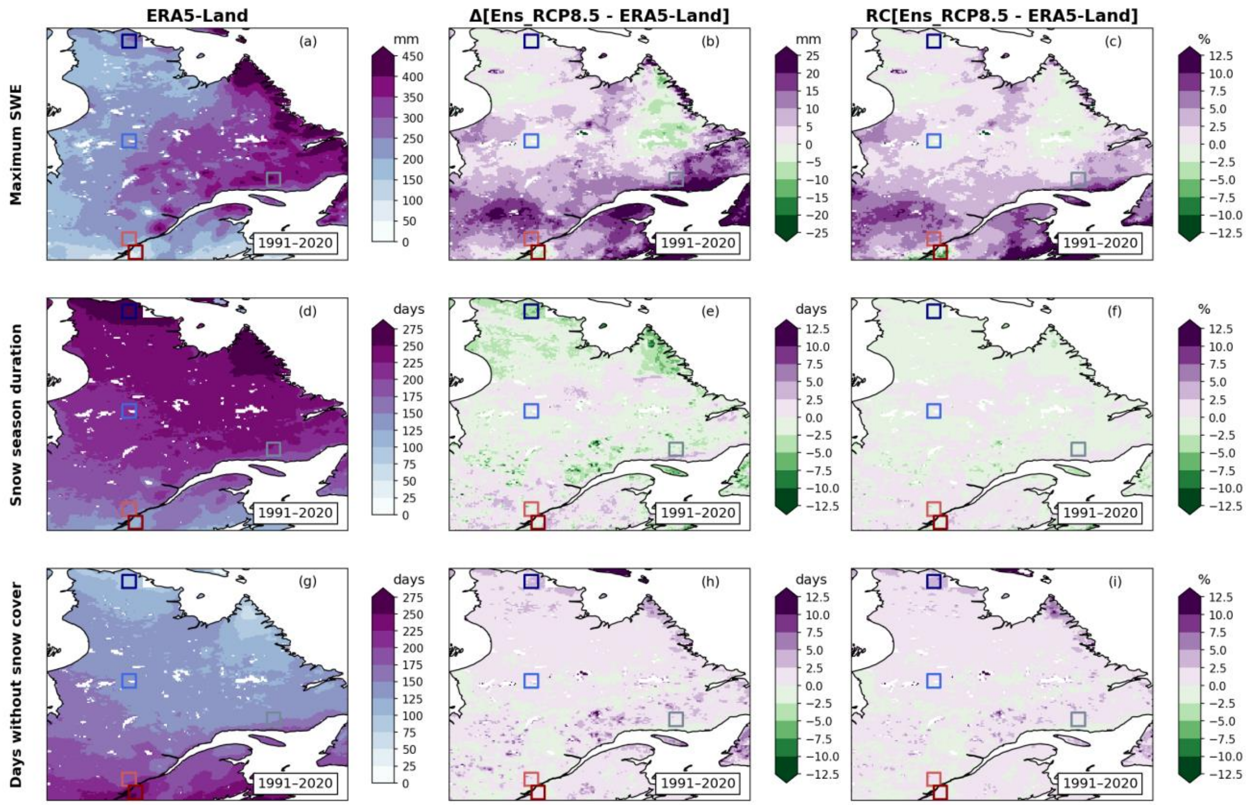


Figure 5 is mentioned before Figure 4, rearrange order.

This will be done.

Figure 8: It appears that coastal regions gain a lot more snow-free days than continental regions, especially so in the 2070s compared to historical. This is not mentioned in the discussion. Lakes are not masked in panel (a).

This pattern could be a result of a larger warming in the coastal regions regarding the continental, more mountainous regions. This will be added in the manuscript. The lakes will be masked in all the figures.

Figure 9: It's interesting that the period of the year with full snow cover (no fragmentation) can be seen from this figure. In Northern Quebec, full cover appears to last between early November and early May. It lasts a little longer in 2041-2070 (mid-Nov to Jun), and then shortens by 2070-2099 (Dec to mid-May). In the other regions, the full snow season shrinks uniformly over time or disappears completely (as in southern Quebec, SLRV). On the other hand, we do not see the duration of noSCseq get longer over time, indicating that there is still regular (short-lived) snowfall occurring throughout the winter months in all regions even at the end of century in these simulations.

This is an interesting point. We will add some information in the manuscript.

Minor points or suggestions for wording:

L4: Replace wording with "...an ensemble of statistically bias-corrected regional climate projections of snow water equivalent..."

This will be done.

L19: Remove "etc." and replace with "and more".

The modification will be done.

L20: Could add some comments on the value added by statistical bias correction of multi-model output.

The following sentence will be added: "Impact models can be sensitive to small systematic biases that are present in the global climate models or inherited by regional climate models from their driver (Muerth et al., 2013; Maraun et al., 2010). Bias adjustment helps reduce these biases, which also reduces the inter-model differences and eases their combination into multi-model ensembles."

L24: I think "datasets" here should be "methods".

The sentence will be modified as follows: "Extending the production of bias adjusted variables can be challenging."

L24: Replace ";" with "." or adjust capitalization.

The modification will be done.

L25: Clarify why these SWE is difficult to handle. Perhaps use L110-112.

This text will be added: "The standard bias adjustment techniques often rely on quantile mapping (QM). Declined in multiple variants (see Gutiérrez et al., 2019 for an exhaustive list), these methods generally divide the reference and the source into quantile-based bins and for each bin, find an adjustment factor that is applied either additively or multiplicatively, depending on the adjusted variable. Particularly, it is the zero-bounded aspect of surface snow that makes it more difficult as additive correction can create nonsensical negative values, and multiplicative correction is incapable of adjusting values that are zero in the source.”

L32: Could add how coarse the GCMs and RCMs tend to be.

RCMs have resolutions lower than 50 km and GCMs higher than 250 km.

L32: What "biases" need to be "addressed"? How is this typically done, perhaps for other variables?

There are known biases in the GCMs simulations, like in temperature and precipitation for example. These biases can propagate in the LSMs. The biases can be address with some bias adjustment method like quantile mapping or delta methods.

L44: Add some brief description of the method (e.g. "We propose using this method to insert nonzero snow near the snow end season before applying EQM to bias-adjust the SWE data.")

We will add a description, thanks for the suggestion.

L85-99: Highlight the dataset that was chosen (ERA5-Land), including its strengths and weaknesses, rather than first mentioning others that were not selected. Please also mention the data assimilation in the ERA5- Land model and perhaps the known biases (which are now in Section (4.2)).

Considering this comment, we will modify the Section 2.3.1 as follows : “Mudryk et al. (2025) showed that ERA5-Land had a good SWE performance for eastern North America among over products and it was the reference dataset used operationally in Ouranos’ climate services at the time of the project to consider this product as the reference data here. ERA5-Land was produced by a land surface model running offline and driven by ERA5 atmospheric fields. ERA5-Land has a 0.1 ° horizontal resolution and an hourly temporal resolution, and its SWE data is publicly available (named “snow depth water equivalent"). No SWE observations were assimilated in ERA5-Land. Leduc and Logan (2025) highlighted that the warm bias of daily minimum near-surface temperatures in ERA5-Land induces a bias in the climatology of the freeze-thaw events, which they define as a day when the minimum temperature is below 0 °C and maximum temperature is above 0 °C and such a bias could also have an effect on the SWE. Kanda and Fletcher (2025) analyzed the bias of ERA5-Land SWE across Canada against Canadian Historical Snow Water Equivalent observations (CanSWE; Vionnet et al., 2021), for three ranges of elevation. In Québec, the selected stations belonged in either low or medium elevation categories. In their results, they highlighted that ERA5-Land tends to underestimate large snow-packs (exceeding 300 mm, mainly high-altitude stations), well-estimate mid-elevation regions, and underestimate low-elevation regions. For a more local overview of ERA5-Land performance, an analysis was performed to assess ERA5-Land against CanSWE and MELCCFP datasets on our studied domain (see Section 3.1). Considering the literature and the additional analysis, the ERA5-Land product was selected as the reference data for this study."

L90: Mudryk et al. (2015) does not cover ERA5, ERA5-Land, or B-TIM datasets, so might not be directly relevant here. Please review this reference.

We will modify the sentence by: "However, spatial differences were observed between such databases (Mudryk et al., 2015, 2025)."

L106: The authors should clarify if they mean discrepancies between models, scenarios (RCP4.5 vs 8.5), or realizations. I suggest the use of the phrasing "forcing scenario ensemble" instead of saying "RCP ensemble" throughout the paper.

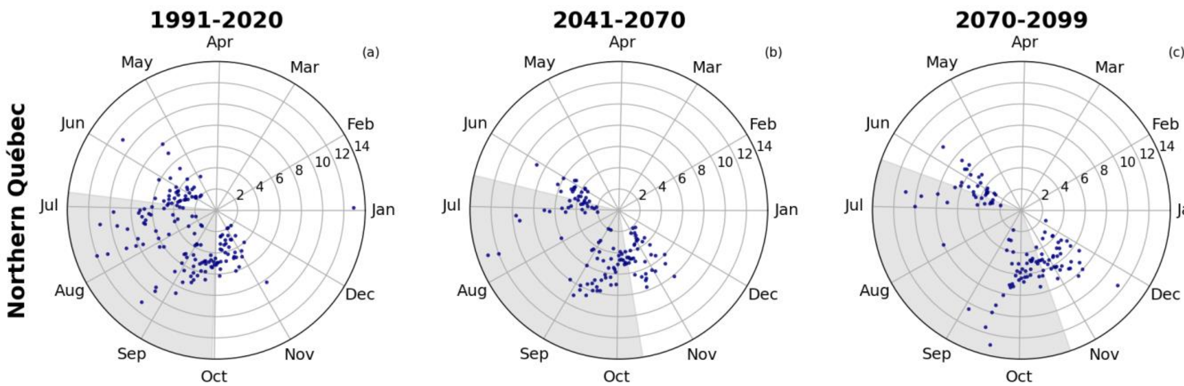
The discrepancies are between the simulations. This will be modified in the revised manuscript. We will also consider some adjustments in the phrasing regarding "RCP ensemble".

L127: Could say SSE instead of "end of its snow season"

The modification will be done.

L175: Is it possible that this asymmetry in the response for Northern Quebec could be related to inserting nonzero snow values only in the snow melt season?

Northern Quebec reacts differently to climate change than other regions in various aspects which are not related to the boundaries of the snow season and are not affected by the nonzero values insertion. The fact that the snow season start changes more rapidly is plausible in the North. Historically, there are years where there is a barely a pause in snow cover, and a partial thin snow cover below the 4mm threshold can still give an albedo that helps a snow build-up in the start of seasons. With climate change, it becomes more likely that Northern Quebec experiences more frequently a total absence of snow, which further delays snow season starts (see Figure below).



L206: unclear what it means to individually analyze "simulations" as opposed to "RCP ensembles".

By “RCP ensembles” we mean the median of the simulations with a specific RCP, whereas by “simulations” we look at each simulation individually. This will be clarified in the revised manuscript.