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

Bresson et al.

Authors’ response (RC1)

We thank the reviewer for reviewing the manuscript and their constructive comments. We will revise the manuscript to clarify the scientific context and the methodology for bias adjustment, add an assessment of ERA5-Land SWE performance against observations, and present bias adjustment effect on the ensembles. We address each specific comment in detail below.

General comments

1. The introduction of the paper is rather short and does not present well the scientific context of the study. It should give more details about the bias-adjusting methodologies that are widely used in climate services. These methods are not necessarily well known by non-experts in climate projections. In addition, the introduction could give more detail about the challenges associated with the debiasing of snow variables that present a strong seasonality. The authors mention L 44 that the only debiasing method for SWE has been proposed by Michel et al. (2023). They could explain briefly what the characteristics of this method are and why does it work for snow variables. At the end of the introduction, the readers need to understand why bias-adjusted variables are crucial for climate services and the associated challenges for snow variables.

The points mentioned above will be taken into consideration and the introduction will be amended accordingly.

L20-33: “Impact models can be sensitive to small systematic biases that are present in the global climate models (GCMs) or inherited by regional climate models (RCMs) 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. Currently, this is only applied to a subset of variables for which statistical adjustment has been more extensively studied, namely near-surface temperature (2-m temperature) and precipitation and mainly on GCM simulations (e.g. Climate Portraits, climatedata.ca, Climate Atlas of Copernicus). Several climate indices are derived from these bias-adjusted variables. Extending the production of bias adjusted variables can be challenging. Some variables such as freezing rain have rare occurrences. Other variables such as snow on the ground or precipitation in regions with a monsoon regime exhibit strong seasonality. These features are difficult to handle with standard bias adjustment techniques, which often rely on quantile mapping. 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."

2. The authors use the SWE from ERA5-Land as the reference gridded product for debiasing. They only refer to performances reported in Kenda and Fletcher (2025) and Mudryk et al. (2024) to justify their choices. However, this justification remains vague. For example, they refer to the evaluation of ERA5-Land for different ranges of elevation detailed in Kenda and Fletcher (2025). However, this evaluation concerns Canada as a whole, including regions such as the Prairies or the Arctic where the snow conditions differ significantly from those found in Quebec. For this reason, I recommend the authors add an evaluation of the performances of ERA5-Land across Quebec using reference SWE measurements from the CanSWE dataset complemented by SWE data collected by the province of Quebec. Such rich dataset would allow a comprehensive evaluation of ERA5-Land in Quebec that would (i) strongly justify the choice of ERA5-Land as the reference product for debiasing and (ii) help identifying regions of lower performances for ERA5-Land where the result of the debiasing should be considered more carefully. In itself, a solid evaluation of ERA5-Land would be a very interesting outcome of this paper.

We performed an assessment of ERA5-Land against CanSWE (Vionnet et al., 2025) and SWE data collected by the province of Québec (GovQC hereafter; MELCCFP, 2020). Even if this work is not as thorough as, for example, Kouki et al. (2023) or Mudryk et al. (2015,2025), it helps justify the choice of ERA5-Land and identify regions with lower performance.
These results were added to the manuscript in the new Sect. 3.1 ERA5-Land assessment for Québec (**L222-244**).

3. A key component of this paper is the method used for the bias adjustment of SWE from regional climate models. In its present form, the paper does not explain well and

illustrate well why applying a bias adjustment is crucial before deriving relevant projections for different snow indices. Section 2.4 should be heavily revised to better explain the bias-adjustment method.

To provide a more complete and clearer description of the method for the bias adjustment and pre-processing of SWE, Section 2.4 Statistical Downscaling and Bias-adjustment Method was modified (**L167-220**). Parts of the text answering specifically the following points are mentioned after each point.

The following points should be considered:

- It is not clear how the method proposed by Michel et al (2023) addressed the issue associated with the seasonality of snow mentioned in the introduction as one of the main challenges for the bias adjustment of snow variables. (**L186-209**)
- It seems that the bias adjustment method allows for a downscaling from the native resolutions of the RCM (which are not mentioned in the paper) to the resolution of ERA5-Land.

The native resolutions of the RCM were added in Table 2 (**P7**).

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

- Can the author explain how such downscaling is possible? (**L168-170**)
- A pre-processing step is applied to the SWE time series during the melting season (P 6 L 135-142). It would be very valuable for the reader to have a figure that illustrates the impact of this pre-processing step. (**Fig. 3 P9**) In addition, is this pre-processing applied to each of the grid cells? Yes.

4. I understand that the authors want to focus their analysis on the bias-adjusted projections as part of group that provides climate services to stakeholders. However, from a scientific point of view, I recommend the authors to quantify if the projections of the different snow indices would have been different without the bias adjustment. Such analysis would highlight the importances of carrying out such bias adjustment when considering snow variables. For climate service, it may be obvious for certain variables such as near surface temperature, but it seems that bias adjustment is less usual for snow variables due to methodological challenges. This analysis would be a solid contribution from this paper that would extend beyond stakeholders interested in the future of the snow cover in Quebec.

A complementary analysis was performed regarding raw and bias-adjusted simulations. This information is in the new Sect. 4.1 Bias adjustment effect on annual cycle.

L302-313: "After the first selection of simulations discussed in Sect. 2.3.2, inter-model differences remain, especially in the SWE amplitude. Figure 15 presents the differences between the climatology (1991–2020) of the annual cycle of SWE for raw (dotted line) and bias adjusted (solid line) simulations for the five regions of interest. The raw simulations were downscaled to ERA5-Land grid using a bilinear interpolation.

The subregion with the larger discrepancies in raw simulations is Côte-Nord with maximum SWE from around 200 to 550 mm. The simulations produced with the RegCM4 model present a larger amount of SWE than the other simulations, except for Northern Québec. This is coherent with McCrary and Mearns (2019) conclusions on the bias in RCM's SWE being mostly induced by the RCM, less by the GCMs. As expected, bias adjustment mainly influences the SWE amplitude, and in a smaller part the SSS and SSE. The bias adjustment succeeds in reducing the small systematic bias from GCMs or inherited by RCMs from their drivers and helps reduce inter-model differences."

Specific Comments

P1 L 14: the terminology "the northern part of the northern hemisphere" is rather vague. Can the authors clarify? Maybe give a range of latitude.

The northern part of the northern hemisphere refers to the Arctic (66.6°N and up). Our sentence was unclear and was modified.

L13-16: "Changes in snow cover have significant feedback on climate (Déry and Brown, 2007; Flanner et al., 2011). The Northern Hemisphere witnesses a decrease in snow cover extent, a shortening of the snow season (Derksen et al., 2019; McCrary et al., 2022; Mudryk et al., 2020; Fox-Kemper et al., 2021) and a greater warming than in the rest of the world for the Arctic part (e.g. Hoegh-Guldberg et al., 2018)."

P 1 L21: what do they authors mean by "best understood"?

By "best understood", we meant the variables for which statistical adjustment has been more extensively studied, partially because temperature and precipitation have observations with the highest coverage and longest time span than the other variables. Some modifications were applied to the text to clarify this point (for example, **L23-26**).

P1 L22: "surface temperature": is it the actual surface temperature ("skin temperature") or the screen-level temperature (taken for example at 2 m above the ground)?

The surface temperature described in this paper is the 2-m surface temperature, above the ground. Clarification was added to the text (**L23-26**).

P1 L24: it would be highly relevant for the readers to add here references that illustrate how challenging it is to adjust the bias of variables with rare occurrences or strong seasonality.

Some modifications were made in the manuscript (**L22-33**).

P2 L 30-35: these sentences contain several statements that should be supported by appropriate references. For example, the statement "to better reproduce ... the processes such as sublimation or ablation" is really vague and must be supported by references.

The text was modified as follows.

L41-47: " LSMs can also be used in an offline-mode, which presents some advantages, like allowing a higher resolution to have more accurate orography. For example, ERA5-Land (Muñoz-Sabater et al., 2021) is run off-line, forced by ERA5 atmospheric fields (Hersbach et al., 2020), has a finer resolution (9 km) than ERA5 (0.25°), uses a daily environmental lapse rate to adjust air temperature for the altitude differences, and has an overall better performance than ERA5 for SWE. As bias adjustment of temperature and precipitation is better understood, it can also be performed on the drivers (GCMs or RCMs) of such offline LSMs (Luca et al., 2017; Morin et al., 2021). Using bias-adjusted inputs in offline LSMs adds a step in the simulation of snow at a high resolution. However, no feedback between the surface and the atmosphere are allowed when LSMs are used offline. "

P 2 L35: Offline simulations with snowpack schemes are often carried out at continental or global scales (such as the ERA5-Land product used in this study or the Crocus-ERA5 dataset

(Ramos Buarque et al., 2025).In this context, I recommend the authors to rephrase the sentence “Consequently, this method could be better adapted for specific purposes at a local scale”.

See previous answer.

P 2 L 43: the term “snow cover” used here is confusing since it is already widely used in the paper to refer to snow in general. Maybe use “snow cover fraction” since it is the variable of interest in the paper of Matiu and Hanzer (2022).

The term “snow cover” was modified for “snow cover fraction” (**L55**).

P 2L 48-49: can the authors explain briefly what are the problems that arise with snow simulations at high elevation?

At high elevation, the resolution used in the RCMs (10 to 45 km) can be too coarse to represent adequately the complex topography of mountains (orography is thus smoothed) and miss some of the orographic forcing on precipitation (**L61-63**).

P 2 L59: It would be interesting to know the mean elevation of the different subregions considered for the analysis. In particular, it would illustrate well the contrast between Southern Quebec and SLRV.

We added information about mean elevation for each subregion in Sect. 2.1. The values are: 468 m for Northern Québec, 391 m for Central Québec, 356 m for Southern Québec, 59 m for SLRV and 390 m for Côte-Nord (**L71-75**).

P3 L 71: what do the authors mean by “flexibility”? Would it be possible to reformulate to be clearer?

By flexibility we referred to a criterium that is not too strict. Indeed, with a 30-days threshold for example, some regions in southern Québec could have a no snow season at all due to the fragmentation of the snow cover by the end of the century. Considering more than two weeks for the SSS and SSE triggers can also limit the use of such indices for adaptation plan. The sentence was modified.

L84-89: “The 14 day threshold was chosen for two reasons: it was long enough to avoid an artificially short snow season when the snow cover is often interrupted by brief warm spells, and it was short enough to be of use for stakeholders and adaptation planners, in opposition to a 30-day threshold like the one used in Brown (2010). Indeed, with a 30-days threshold for example, some regions in southern Québec could have no snow season at all due to the fragmentation of the snow cover by the end of the century. Considering more than two weeks for the SSS and SSE triggers can also limit the use of such indices for adaptation plan."

P 4 L 89: what do the authors mean by "mismatch"? Between which datasets? What was the nature of this "mismatch"?

The term "mismatch" refers to differences between gridded datasets, such as Crocus, MERRA-2 or ERA5-Land. The sentence was modified.

L105: "However, spatial differences were observed between such databases (Mudryk et al., 2015, 2025)."

P4 L 91-92: this sentence should be reformulated since Figure 3 in Kenda and Fletcher (2025) presents an evaluation of the SWE from ERA5-Land across Canada (including region below 50N). Kenda and Fletcher (2025) did not only evaluate ERA5-Land in northern Canada above 50N.

This mistake was corrected and the sentence was changed.

L117-118: "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."

P 5 L 103: how many simulations were considered in this first ensemble?

In the first ensemble, 26 simulations were considered (9 with RCP4.5 and 17 with RCP8.5) (**L129-130**).

P 5 L 117: Why are the authors using the argument about the availability of SWE observations to justify focusing on the region below 50 N in their selection criteria? Indeed, SWE observations are not used in this study to evaluate ERA5-Land (see my second general comment) and are not assimilated in ERA5-Land.

ERA5-Land has an atmospheric forcing based on ERA5 and does not assimilate observations (Muñoz-Sabater et al., 2021). On the other hand, ERA5 assimilate various variables including temperature and humidity that compel the near surface temperature field (Hersbach et al., 2020). ERA5 presents a better performance in the meridional regions with more instrumentation and consequently more observations to assimilate. In Eastern Canada, most of the observations are in the South. Another motivation to focus on this region (< 50 N) was the population density, as one of the motivations to provide this dataset is to be used in an adaptation context (**L141-144**).

P 5 L 119: how many candidates were present in the initial ensemble? Such information is interesting to better understand how strict the selection criteria were.

There were 26 candidates before the selection. We tested different thresholds (Figure 2).

L145-146: "Different thresholds were tested. The selected thresholds allow to consider different driving GCMs, RCPs and RCMs combinations to cover uncertainties."

P 5 L 119-120: it would be good to know what the selection criteria were in McCrary et al. (2022) to justify why it makes sense to compare the two ensembles.

McCrary et al. (2022) rejected RegCM4 simulations from the NA-CORDEX ensemble because of unbounded snow accumulation over some mountainous regions and the southern Baffin Island (see Figure S1 in their Supplemental Information). The comparison between their ensemble and ours was to confirm that our selected simulations were not flagged for some specific issue in our region of interest (**L146-151**).

P 6 L 147: does the value of 375 mm refer to ERA5-Land?

Yes, the value refers to ERA5-Land (**L246**).

P 7 L 157-159: Has the bias correction already been applied when presenting the results for the two ensembles?

Yes, all results are presented for the bias-adjusted indices (**L89-90**). In regards of reviewers' comments, we added more information about the ensembles before the bias adjustment in the Discussion (see answer to General comment 4 and **L302-313**).

P 12 Figure 9: The dots on the different subplots seem to be rather noisy, especially for the northern and central domain. Is it because of very few events (even only one) over the 30-yr period are considered when computing the mean duration of noSCseq? Showing results aggregated longer time periods can potentially reduce the noise and make the figure easier to read.

As we grouped noSCseq by the day of year on which they start and took the 30 years mean of their length, we can indeed have some noise specifically when only one sequence is registered. This happens rarely during the snow season and could be due to interannual variability. The issue with an aggregated version of Figure 9 is the smoothing of the results and potentially removal of the period without noSCseq during summer in the Northern Quebec region during 1991-2020.

Technical Comments

Text

P2 L 32: “CROCUS” is not acronym and can be written “Crocus”.

This mistake was corrected (**L35**).

P 2 L 46: “water content in the snowpack” could be confusing. Maybe use “total water content of the snowpack” to make sure that it does not only refer to liquid water content in the snowpack.

The term “water content in the snowpack” was modified for “total water content in the snowpack” (**L59**).

P2 L 55: It could be worth mentioning the other Canadian provinces that are included in the simulation domain.

More details about the other provinces were added to the description of the domain of interest. The missing provinces were also added to Figure 1 (**P4**).

L68-69: “The study domain encompassed the Canadian province of Québec and included New Brunswick and parts of Labrador, Newfoundland, Ontario, Nova Scotia and northeastern United States (Fig. 1)”

P 5 L 104: the year is missing for the reference to Mearns et al.

This point was corrected (**L128**).

P 5 L105: Explain the meaning of the acronym “SM”. It should be changed throughout the document. I also recommend the authors to mention to which specific table or figure they are referring to in the Supplementary Material.

SM stands for Supplementary Material (**L130**).

P 5 L 108: the term “melt” or “melting” is often preferred to “thaw” when referring to snow.

Thaw was modified for melt (**L134**).

P 5 L 131: please double check to reference to (Themeßl et al., 2012). Is the family name written correctly?

After double checking, the name of the first author in this publication is correctly written.

P 6 L 144: this sentence can be included in the previous paragraph to avoid having a paragraph made of a single sentence.

This sentence was modified and moved (**L173**).

P 6 L 148: explain that the names of the regions such as Charlevoix, ... are shown on Fig. 2.

This information was added.

L246-249: "SWEmax is larger over the eastern part of Québec and Labrador (e.g 375 mm for Côte-Nord), and particularly over mountainous regions like the Gaspesian peninsula, Charlevoix, or the Torngat Mountains (presented in Fig. 1) than the rest of the domain (e.g. 216 mm for central Québec) (Fig. 6a)."

P 16 L 264: add the corresponding DOI.

The DOI was added (**L359**).

Figures

Figure 10: the different levels of transparency are not visible in this figure.

The figure was modified (**P19**).

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

Bresson et al.

Authors' response (RC2)

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 was added to the Supplementary Material, as well as the same figure for SSE.

L131-133: "A first inspection of the simulated SWE from RCMs showed large discrepancies between simulations (Fig. S1 and S2 in SM),

as already highlighted in the literature (e.g. McCrary et al., 2017), forcing a selection of simulations to build the ensemble."

As well, Section 2.3 was completed with Section 4.2 (**L114-123**).

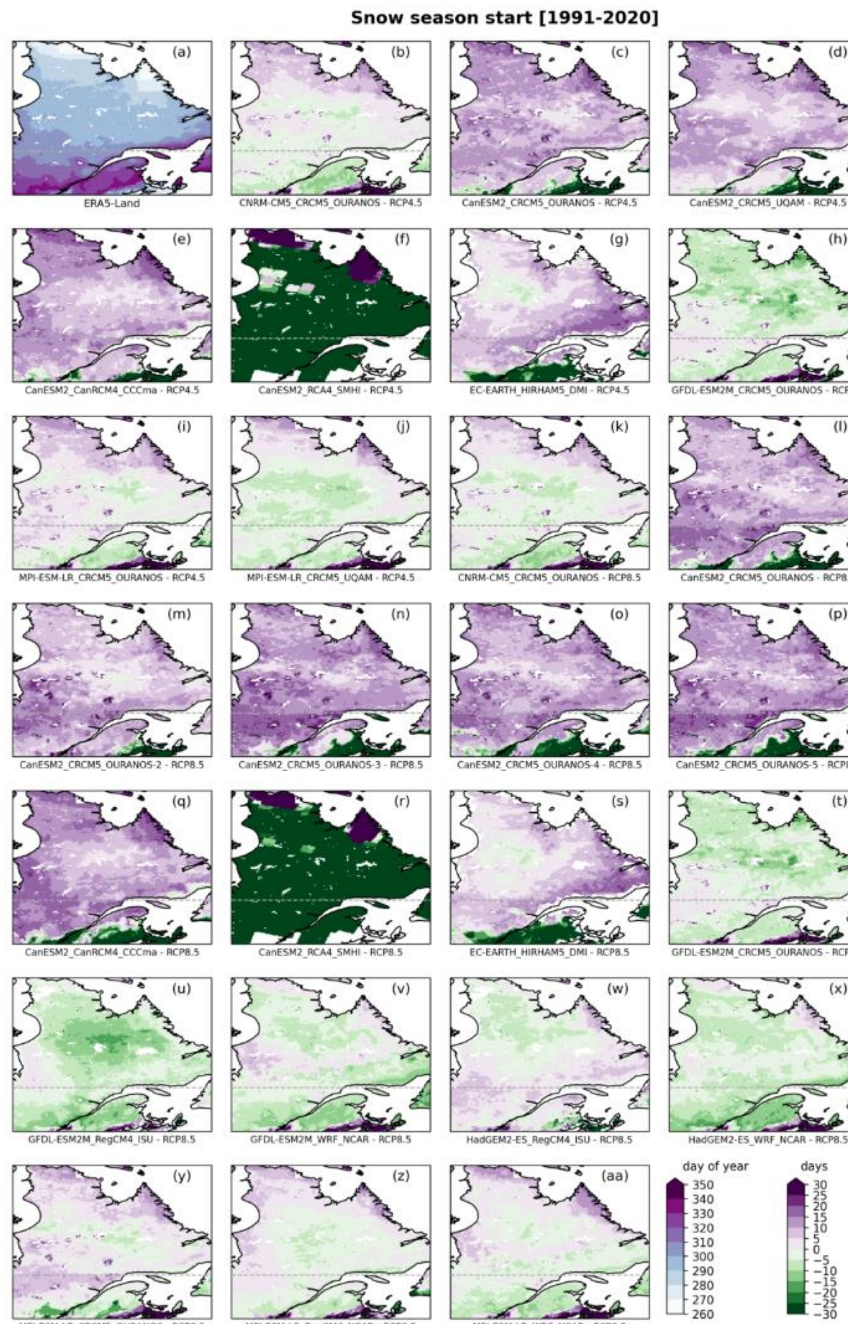


Figure 1. Annual mean of snow season start for ERA5-Land (a) for 1991-2020. Difference between ERA5-Land and simulations' annual mean of snow season start for 1991–2020 (b-aa). Results are presented for all available simulations, before applying the selection process presented in Section 2.3.2.

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 was added (**L262-268**).

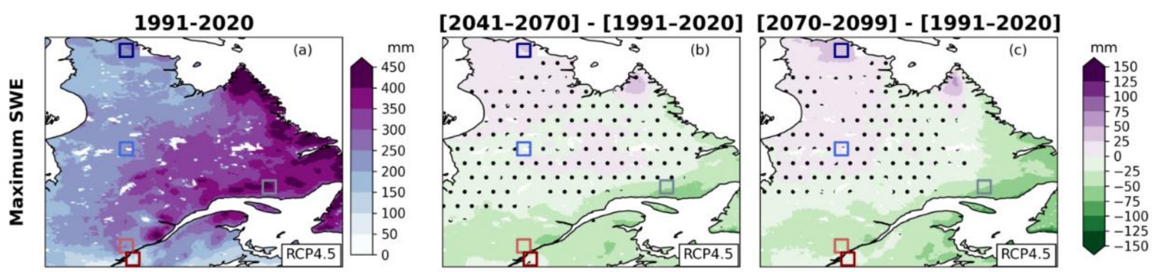


*Figure 2. Annual maximum SWE for 1991–2020 (a), and differences between the 1991–2020 period and 2041–2070 (b) or 2070–2099 (c). Dotted areas are those where less than **80 %** of the simulations in the ensemble agree on the sign of change. Results are presented for the RCP8.5 ensemble.*

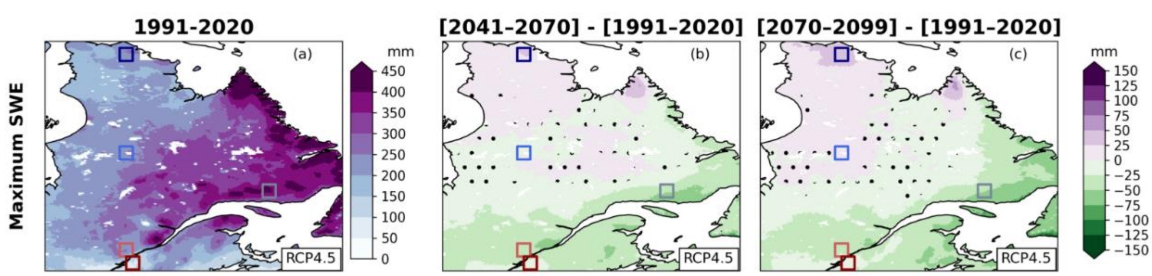


*Figure 3. Annual maximum SWE for 1991–2020 (a), and differences between the 1991–2020 period and 2041–2070 (b) or 2070–2099 (c). Dotted areas are those where less than **75 %** of the simulations in the ensemble agree on the sign of change. Results are presented for the RCP8.5 ensemble.*

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 added this information to the "Statistical Downscaling and Bias-adjustment Method" section.

L173-175: "Both the simulations and the ERA5-Land dataset were resampled to a daily frequency by taking the mean of sub-daily data. The quantile-based approach used for the bias-adjustment is described in what follows."

    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. (2023) 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 modified the text accordingly to include these details and better distinguish our method from Michel et al. (2023) (**L197-203**).

    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 (**L186-214**). An example is now shown in Figure 4 (**P9**).

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 enhanced the description of the bias correction process in Sect. 2.4 (**L167-220**).

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 (QDM). 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. EQM 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").

We revised the tense consistency and corrected the minor English errors.

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 Table 2 about noSCseq was added (**P4**).

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.

The native resolution of the RCM was added in Table 2 (**P7**).

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 were modified (**P13**). Lakes were masked in every map of the manuscript (**P13, 16, 17, and SM figures**).

Figure 5 is mentioned before Figure 4, rearrange order.

This was 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 was done (**L4**), and we kept the term “adjusted“ instead of “corrected” for consistency.

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

The modification was done (**L18**).

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

Some information was added.

L20-23: “Impact models can be sensitive to small systematic biases that are present in the global climate models (GCMs) or inherited by regional climate models (RCMs) 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 was modified.

L26-27: “Extending the production of bias adjusted variables can be challenging.”

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

The modification was done (**L26**).

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

This text was clarified (**L28-33**).

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

RCMs have resolutions lower than 50 km and GCMs higher than 200 km (**L39-40**).

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 (**L39-40**).

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.")

Thank you for the suggestion.

L56-58: “We will use the re-insertion of non-zero SWE values near the snow end season as proposed in Michel et al. (2023) before applying empirical quantile mapping to bias-adjust the SWE data.”

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 modified the Sect. 2.3.1 Reference Data by adding information previously in Sect. 4.2 (**L114-123**).

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.

The sentence was modified.

L106: “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.

L131-132: “A first inspection of the simulated SWE from RCMs showed large discrepancies between simulations (Fig. S1 and S2 in SM), as already highlighted in the literature (e.g. McCrary et al., 2017), forcing a selection of simulations to build the ensemble.”

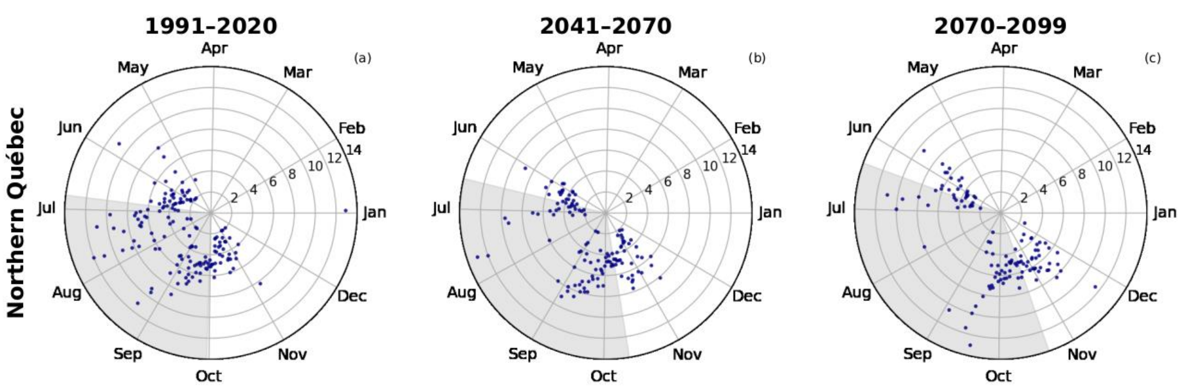
We considered some rephrasing in the text for "RCP ensemble" and explicitly defined it. **L151-152**: "Hereafter, when results are presented (Sect. 3) for RCP ensembles, it refers to the median of the simulations for a specific RCP."

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

The modification was done (**L189**).

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 (**L151-152**), whereas by "simulations" we look at each simulation individually (**L319**).