General comment:

The paper presents the use of reanalysis data to drive a distributed, physically-based modelling endeavor of the Saskatchewan Glacier. The authors cleverly combine a very limited on-site AWS record with longer term regional AWS records, and NARR product for multiple variables to model the long term (1979-2016, and 21st century climate perturbations) mass balance of the glacier. The analysis includes geodetic estimates of glacier change and point-scale estimates of mass balance, that are used to further constrain the model and investigate its performance. These results are used to assess the glacier's 21st century climate sensitivity across projected temperature and precipitation changes from the IPCC RCPs.

I found it challenging to follow the organization of the scientific information and I believe the manuscript should be significantly reworked. The text contains grammar and spelling mistakes that need to be addressed. The authors should clarify their use of terminology (i.e. 'altitude' vs 'elevation' are used interchangeably), homogenize variable names throughout the text (i.e. Ta for all air temperatures derived from NARR, from the regional AWS network, and the downscaled model input), and use consistent units (i.e. melt reported in mm w.e., cm w.e., and m w.e. at different stages). The information pertaining to the methods is disorganized causing difficulty in deciphering the approach (with omissions of details about downscaling key variables such as temperature and wind speed to the distributed glacier grid), and at some points systematic misuse of terminology ('downscaling' vs 'bias correction' vs 'extrapolation to model grid') obfuscates the underlying novelty of the approach to designing a spatially-distributed application of a physically-based melt model in a setting with limited input data.

Additionally, the use of reanalysis data from a single NARR grid node as model forcing is not adequately justified. The authors convincingly demonstrate that these data perform well in comparison to observational records of climatological variables. However, without additional information on the performance of either nearby nodes or an aggregate of these, it is impossible to determine whether this convincing performance is due to a modelling workflow which leverages a reanalysis dataset that is transferrable to other glaciers (especially those that are large enough to intersect multiple grid cells, or small enough and less fortunate in their geography to be equidistant from many grid nodes), or simply due to the coincident position of Saskatchewan Glacier relative to the nodes of the reanalysis product used. Some form of evaluation pertaining to the regional potential of the NARR product for direct incorporation into this melt-model workflow would help to better highlight this paper's relevance to cryospheric research.

For further context, I would highlight the interesting use of the melt modelling results to assess the sensitivity of Saskatchewan Glacier, specifically in extracting meaningful process-based explanations from the glacier's energy balance. These are obviously locally significant, and perhaps regionally so, but it is difficult to address the true scale of relevance of these results without significant improvements to the clarity and organization of writing, and rigor of the analysis. These improvements, combined with a more focused introduction presenting exactly the purpose of the study would greatly help in highlighting the potential use of the method for data-poor glacier settings. Additionally, improvements to clarity and organization would ensure the approach is reproducible, and as such, transferable to others.

Specific comments:

Section 1:

The introduction contains very detailed information on several topics related to the analysis that distract from a well-formulated motivation for this work. I would recommend using more direct wording to improve the readability of the text (i.e. 'the amount of precipitation remains unchanged' vs 'precipitation remains unchanged'; line 34, 'the so-called 'temperature index' or degree days' models vs 'temperature index models'; line 59). I would also recommend reducing the specificity of information on topics both tangentially and directly related to the paper topic.

With respect to tangential information: the details about GIC mass loss (lines 45-50) seem incongruous with the paper topic and the opening summary of glacier-freshwater relationships, details about the Canadian Rockies (lines 97-117) could be shortened by moving some information to the study area section.

With respect to direct information: detailed information on the temperature index to energy balance spectrum of melt models could be significantly condensed (lines 59-77), information on reanalysis products could be limited to NARR (lines 83-90), and information on their downscaling could be limited to the methods relevant to the study approach (lines 90-95).

The stated objectives highlight the 'comparatively unexplored' topics of ablation vs precipitation phase feedbacks to temperature sensitivity (lines 119-124) but offer no literature to situate this statement or the study. Please clarify what exactly is novel about this study's results and situate it in existing literature to really highlight the novelty of the study. The text at the very end of the paper (line 810-811) could be tied back to a statement here as it is more specific.

Section 3:

The data and methods section requires significant redesign. Section 3.2.3 contains most information pertaining to reanalysis data, and all pre- and post- processing associated with bias correction and downscaling across 5 climate variables with no internal divisions. Additionally, information on downscaling across the glacier model grid in a distributed manner is contained within the mass balance model section, which makes it hard to find and follow the logic of the approach. I would recommend a clear division of data (sources, types, any pre-processing strategies) and methods used in manipulating these data to create model inputs (downscaling, bias correction), followed by information specific to the mass balance model once all input variables (to the distributed model) have been presented. I like the authors strategy of presenting each element of the energy balance, but this could be shortened significantly if information relative to model input pre-processing were already made clear beforehand.

Additionally, the exact workflow used is very difficult to decipher. The authors first obtain NARR data from a node closest to the study area, then they bias correct these long-term timeseries of climatological variables (1979-2016) using a two year on-ice AWS record (excepting precipitation), then downscale the bias corrected NARR data to the model grid using lapse rates derived from regional AWS records. In the text, the terms bias correction and downscaling are used synonymously for work performed at the on-ice AWS location (lines 238-239), and the precipitation variable downscaling step is described as

'extrapolating to the model grid' in the mass balance model section (lines 262-263). See my technical comments for citations that may help in differentiating these terms. It is necessary to address the confusion in usage of terms such as 'downscaling', 'extrapolation', and 'bias correction', and distinguish between the steps for the transition of NARR climatological variables from the grid cell timeseries, to the AWS location timeseries, to the distributed model grid timeseries ultimately used to force the mass balance model for each variable, in order for this workflow to be reproducible.

Section 4:

The use of a single NARR grid node to drive the entire modelling study must be evaluated further. The authors demonstrate that temperature, solar radiation, and relative humidity perform well compared to a 2 year on-ice AWS record. My concerns and related recommendations here are two-fold:

(1) There needs to be evidence that demonstrates how the use of this grid node is justified compared to other regional grid nodes (for example, at better situated elevations), and/or a regional aggregate of NARR data, and/or using bilinearly interpolated NARR data at the AWS location (given that the nodes define a continuous field across the region of interest). Some recommendations include: commenting on differences between the coarse NARR topography (derived from surface geopotential variable in NARR for example) to credit/discredit the use of information at the AWS location, a sensitivity test of the model using a different grid node(s) and/or a regional aggregate of nearby grid nodes as forcing, and/or using AWS location bilinearly interpolated values to assess the geographic sensitivity of point-scale data as forcing (and assessing the relative cost/benefit of introducing interpolation to this approach).
(2) As it stands the novelty of this method seems limited by the possibility that it is only applicable to glaciers with nearby reanalysis grid nodes. Given that this study investigates a single glacier in the Canadian Rockies, it would greatly improve the significance of the work to demonstrate that this method can reliably inform physically-based models on other glaciers with limited AWS forcing data. To do this however, the above (1) needs to be addressed.

Throughout:

Methods details are missing:

- Temperature downscaling using lapse rates to the model grid is never explicitly detailed.
- The sensitivity of the model to poor wind forcing input is never evaluated (the downscaling of this variable to the model grid is not addressed and I am unsure if the AWS bias corrected NARR values are used to force the entire glacier surface or if these data are further downscaled to be distributed across the finer model grid). If only one wind speed timeseries is used to drive the entire distributed energy balance model, this strategy must absolutely be justified with supporting literature given a lack of distributed on-ice wind data to evaluate local variability. I do not have sufficient expertise to comment on the limitations of this in detail, but it seems important to demonstrate that this is a valid strategy for forcing wind speed, given how it fails to perform well due to the presence of katabatic and sloping wind signals in the AWS record (table 1), and the fact that WS factors heavily into the discussion of results relating to the energy balance of the glacier. See technical comment for more information.

- Smaller details of the workflow are missing (i.e. filling of precipitation record data gaps, see technical details) and are needed to ensure the study results are reproducible and the method transferable to other settings.

Presentation of results could be improved:

- Units change frequently throughout; this is distracting and must be made consistent.
- Please assign a unique term to each variable. For example, temperature that has undergone different steps of processing should have a different subscript, terms presented at one stage of the paper should be consistent throughout (i.e. Qh changes to Qs for sensible heat flux in methods/results), and variables in equations should have the same notation as those in the text that defines them. Whenever possible, variables relating to mass balance should be consistent with the mass balance glossary:
  Cogley, J. Graham, et al. "Glossary of glacier mass balance and related terms." *IHP-VII technical documents in hydrology* 86 (2011).
- Occasionally, statements are inaccurate of the literature cited for support, and the work of others is improperly cited. See technical comments.

Technical/detailed comments:

The abstract would benefit from being more focused on the paper's key results: removing information on the static vs dynamic extent performance (lines 20-21), details surrounding IPCC RCPs (lines 23-24), and details of the melt vs sensitivity (lines 26-27, 29-30), would both shorten the text and emphasize the key findings.

Lines 23-24: 'IPCC representative concentration pathways climate scenarios' is redundant, either climate scenarios or RCPs

Lines 44-49: Very dense information on SLR specific to GICs can be condensed to remain specific to small mountain glacier study

Line 51: Sentence lacks specific information, or citation if the importance of mass balance variable is necessary to argument.

Lines 52-53: 'Only a few' -> list how many. If this information does not exist, a citation pointing what is known would be appropriate. Time-consuming better described as resource intensive.

Lines 59-62: 'So-called 'temperature-index' or 'degree-day' models' is redundant, can just use temperature-index. More accurate would be to say these models use air temperature combined with an empirically derived melt factor, and require tuning to some observational target result(s).

Lines 62-63: Also require empirically determined melt factors and/or radiation factors that are modified by radiation inputs. This could be cut, but if kept more correct description of model formulations required.

Line 64: Please clarify what constitutes complexity, or reorder until after energy balance model is presented.

Line 67: 'is therefore questionable' is unclear, please state exactly what the issues of spatial and temporal transferability are.

Line 79: AWS is automated weather station, not ‘automated meteorological station’

Lines 79-95: consider condensing information to specific reanalysis product and methods relevant to this study

Lines 104-105 and 107: redundant

Lines 97-117: Consider reordering the information about Peyto Glacier to the end of paragraph to present information from the regional scale first, then the local scale. Or redistribute some regional info to study area section.

Lines 119-121: Great and clear

Line 121: Need citations after the ‘has been investigated in several previous studies’ statement

Lines 122-14: Is this truly unexplored? Citations needed to understand where exactly current studies decomposing energy balance in a climate sensitivity context lie. I recommend:

Rupper, Summer, and Gerard Roe. "Glacier changes and regional climate: A mass and energy balance approach." *Journal of Climate* 21.20 (2008): 5384-5401.

Anderson, Brian, et al. "Climate sensitivity of a high-precipitation glacier in New Zealand." *Journal of Glaciology* 56.195 (2010): 114-128. (already in citation list)

Figure 1: Shaded NARR grid is misleading given that the work uses a single node, and at a glance a reader might assume the shading delineates panel C. Perhaps shade the location of panel C and show locations of adjacent NARR nodes. Additionally, satellite imagery in panel C would help in assessing debris cover, ELA, for example. Could preserve the contours or provide hillshade off-ice to preserve information on shading.

Line 174: A division or set of sub-sections here between DEM sources/manipulation, and mass balance model simulations relative to different DEMs would help clarity of the approach.

Line 186: elevation not altitude

Line 188: GSC acronym is not previously defined, this will be confusing to international readers

Line 192: use ‘see section 3.2.2’, not ‘c.f.’ as no comparison is invited, rather referring to a subsequent step of the approach.

Line: 198: How were these anomalies introduced into the monthly record? More information on method used is needed to understand and allow reproducibility.

Line 203: How were the precipitation records combined? Here I am referring to the choices made (i.e. were the winter gaps in CI AWS filled with PR AWS, or how were overlaps handled). More information on method used is needed to understand and allow reproducibility. Also, note how much gap filling with NARR record was needed overall.

Line 203: Figure two shows precipitation is not recorded between 2005-2016 at either PR or CI AWS locations. The text does not indicate a lack of coverage during the study period. If no AWS precipitation is available between 2005-2016, is the NARR precipitation bias corrected using the pre-2005 historical data

multiplicative deltas? This information is crucial and needs to be reported, even if it is just the same factors as those derived during the historical time period, given that this means mass balance computed between 2005-2016 is exclusively derived from bias corrected NARR.

Line 205: use 'see section 3.2.3'

Lines 205-206 Need to specify the strategy used for summer months.

Lines 205-207: Need to show how the constant lapse rate elevation structure differs from the actual (not mean) lapse rate elevation structure. At least state their differences.

Line 210: its should be their for data, or its for timeseries or record of said data

Line 217: Bukovsky et al. explicitly state that NARR precipitation outperforms other reanalysis products for the continental US and warn against extending this conclusion to the rest of North America. They do suggest that improvements are possible with better precipitation coverage, therefore this statement and use of the citation requires rewording and/or an additional more recent citation to be accurate.

Lines 226-257: need subsections here when methods are introduced to improve clarity. Also need divisions by variable given methods used are different in each case starting line 229.

Line 238: Variable Ta has already been assigned to AWS air temperature (line 187). Need additional subscripts to differentiate air temperatures from different sources/stages of processing. This needs to be corrected across the manuscript, and other variables as well.

Lines 238-239: The data are downscaled to the point location of the AWS, by applying a bias correction. This precedes the downscaling of NARR node information to the finer glacier grid on which the energy balance model is run. These are not necessarily the same thing, although I believe translation of variables (without a reduction in scale) can qualify as downscaling. I recommend offering more information on the Matlab algorithm used to improve clarity.

This is an old, but helpful, reference to clarify terminology:

Hewitson, Bruce C., and Robert George Crane. "Climate downscaling: techniques and application." *Climate Research* 7.2 (1996): 85-95.

Line 257: information regarding downscaling of variables to the model grid should be presented here, before the mass balance model that uses it as forcing is presented.

Lines 261-262/Equation 1: I highly recommend using up to date terms for mass balance terms from the glossary of mass balance:

Cogley, J. Graham, et al. "Glossary of glacier mass balance and related terms." *IHP-VII technical documents in hydrology* 86 (2011).

Line 261: elevation not altitude

Line 261-262: extrapolation of precipitation to the model grid is in fact downscaling (from coarse NARR resolution to finer mass balance model grid). This needs to be explained before the mass balance model is presented. Additionally, the procedure for air temperature needs to be explained explicitly as well.

Lines 264-269: This is the method for obtaining accumulation from precipitation. This is a separate topic and should be in a separate subsection.

Lines 270-271: State grid resolution, present downscaled variable names (with different subscript).

Line 273: Melt is later reported in cm w.e. not mm w.e.

Line 278: Citation needed for statement that justifies ignoring Qg component of energy balance.

Line 283: downscaled to the AWS location (not at). G is downscaled to the AWS location, then extrapolated to model grid also seems like a downscaling step (from coarser to finer resolution).

Lines 294-295: Perhaps I have misunderstood, but does the assumption of spatially constant clear-sky ratio mean that the entire glacier responds to cloud cover at the AWS. If so, a citation or justification on the scales of cloud-cover change would be appropriate and required for reproducibility.

Line 298: It would be appropriate to report the maximum/mean lengths of time that the 'constant' clear-sky ratio is applied to cells that are illuminated after the AWS becomes shaded. If the AWS location is shaded and clear-sky ratio is inferred from that location for several hours at non-shaded locations, any systematic daily patterns in cloud cover could be missed (i.e. common summer afternoon clearing that is missed because all cells are assumed cloudy after the AWS is shaded). If the duration is small it should not matter, but if more than several hours I recommend a citation justifying this approach.

Line 324: describe why these values are 'optimum'

Line 326: LWin subscript

*Line 326: Explicitly describe what assumptions are used for the ice surface temperature (i.e. fixed to 0 when Ta > 0)

Line 331: Ta is multiply defined for AWS and NARR derived air temperatures, and here is neither. Downscaling of Ta to the model grid has not been presented yet, that would be a good place to define a new Ta variable used as input following processing.

Line 336: Downscaling of the WS variable to the model grid is absent, unless the bias corrected AWS location WS is used across the entire glacier. This needs to be explicitly detailed, and justified with literature supporting the use of local, point-scale WS data to drive a distributed energy balance model.

Line 338: WS variable in text, WSz in equation

Line 351: 'similar ice facies morphology between Peyto and Saskatchewan glacier' citation needed.

Line 355: I recommend adding WS/direction to analysis to analysis given the poor performance compared to AWS due to local wind patterns, and the important of sensible heat flux to the overall energy balance.

Line 368: Oerlemans 2001 citation unneeded

371-372 'different greenhouse gases emission scenarios', could just be previously defined RCPs, assuming these are where the T and P values are derived from? Strictly speaking the range is similar but does not encompass (0-8 vs 2.6-8.5)

Line 370: encompass not encompasse, would be more accurate to say approximate given previous comment

Line 372: this IPCC citation is to an annex of AR5. If this is where the values listed are obtained please disregard. Otherwise, I would recommend a reference for the primary AR5 report where RCPs are defined.

Figure two: Adding a shading to the data gaps would help visualize them.

Figure two: See comment for line 203 on precipitation record length. What is the origin of the precipitation data gap between 2005-2016, and what data are being used bias correct the NARR precipitation? It would be helpful to indicate, even if they are simply the values from the historical time period. It would be extremely helpful to color code the precipitation curve to show what data are from each of the PR and CI AWS stations, and what data is gap-filled from the NARR product.

Line 408: the bias-corrected NARR variables, before being downscaled to the mass balance model grid

Line 415: NARR Ta needs to be its own variable

Line 420: I see only a very modest improvement to RMSE and MAE and no difference in R in table 1 for WS (RMSE/MAE/R 2.21/1.72/0.37 vs 2.30/1.85/0.37) and precipitation (RMSE/MAE/R 4.96/2.82/0.30 vs 5.22/3.10/0.30), making the sentence 'Validation statistics were significantly improved using the scaling technique for precipitation and WS' seemingly incorrect.

Lines 419-425: having a call to an equation earlier in text that demonstrates what exactly the scaling method is would be helpful here.

Line 433: Here the data is stated to be the 'bias-corrected NARR', inconsistent with previous statements of downscaled.

Line 454: This seems plausible, but additional analysis could demonstrate it more clearly. I recommend demonstrating the difference in lapse rates between clear and bad weather days as a supporting argument to the statement.

Figure 4: This is a very interesting and helpful figure! I think it would be important to clearly identify the data derived from the regional AWS network (lapse rates and R) and from the Saskatchewan Glacier AWS (WS and WD) to avoid confusion. Perhaps could use different line widths and explicitly state this in caption.

Lines 480-481: I would encourage the use of subscripts, and appropriate terms from the mass balance glossary (see citation at line 261 comment) for all mass balance terms.

Lines 482-485: As previously mentioned (line 205), I recommend showing the difference between mean precipitation lapse rate from AWS network (used to downscale precipitation onto the mass balance model grid) and the elevation dependent lapse rate as a figure. Additionally, I would recommend showing temporal variability over the years with local mass balance validation measurements.

Line 485: explicitly define what is considered 'good model performance', or remind the linear distribution of model vs measured b values if that is the case.

Line 494: m^-1 missing in units

Line 495: use exact language when comparing study balance ratios to those from previous work: the balance ratio is triple (3x) that computed for the region by Rea 2009. This seems anomalous and should be explained further.

Line 510: List the results in text that constitute 'good general agreement'.

Lines 505-516: This section needs to include the geodetic estimate results to have a number to compare the model balance to.

Line 518: Unclear if Ta and P are AWS or NARR variables because of same variable term

Lines 520-522: The difference between mean and elevation dependent precipitation lapse rates needs to be addressed here.

Figure 6: An excellent and helpful figure. Consider renaming the Y labels to cumulative MB. Additionally, choice of colors in legend does not correspond to colors in plot, likely because of overlain curves with transparency. This makes it hard to decipher Z0snow and ice albedo. Change to better color scheme.

Line 539: Sensible heat flux is now defined as QS, it was initially defined as Qh at line 277. These terms absolutely must be consistent throughout to avoid confusion.

8Line 547: Is LW* net longwave? The text refers to it as radiative cooling, which I assume means outgoing longwave exceeds incoming longwave?

Lines 576-578: This should be just melt, not ice melt if I understand correctly that we are above the snowline and therefore no ice surface is exposed in the upper accumulation regions (and all values are in cm w.e.)?

Figure 9: Scale bar needs rescaling on panel A

Line 601: Conversaly - > Conversely

Line 606: were overlapped -> were overlain

Line 607: 'given by the latest projections from climate models' redundant

Line 610: m w.e. -> cm w.e.

Line 613: Static and dynamic mass balance values should be clearly reported and defined with separate subscripts at some point in the manuscript

Figure 10: Previously stated Temperature range was 0-8 C in methods, now 0-7 C. This does not technically encompass RCPs of 2.6-8.5C

Line 645: typo

Line 631-646: A table would be extremely helpful to navigating these results

Line 658-659: reword attractive to a word that is specific to the gain (i.e. efficient, affordable, etc)

Line 660: this has been stated and cited already

Line 662: simple is misleading considering the breadth of statistical downscaling methods. I recommend specifying the exact methods referred to here. Additionally, Clarke et al. 2015 used dynamic downscaling of their surface precipitation to drive their mass balance model and is improperly cited here.

Line 666: Same as above: statistical downscaling can include non-linear regression with multiple predictors, as well as the use of scaling factors, etc, with some methods not having calibration steps. Specify exact methods.

Line 668: replace 'reasonably good' with exact numbers

Lines 674-679: With respect to wind speed, the authors convincingly argue that glacier-scale katabatics that are not captured on the coarse NARR grid account for poor performance (line 410-412), and identify the impact of poor wind speed variable performance on temperature lapse rates (section 4.3). However, the wind speed variable performance is only further discussed as suffering from katabatics (lines 674-679), and the impact of this poor performance on turbulent heat fluxes computed in the energy balance model seem absent. Furthermore the results demonstrate that the sensible heat flux plays an important role in the overall energy budget of the glacier surface. I recommend further comments on the possible consequences of poor wind speed data forcing the model, and would suggest adding this variable to the earlier forcing variable sensitivity tests. Additionally, the impact of using point scale WS data to drive an entire glacier-scale energy balance model must be explicitly justified.

Line 694: biased -> raw

Line 697: list all assumptions explicitly for clarity

Line 700-705: present errors as fraction of modelled results to give context. Discussing an error of 78 cm w.e. a-1 is different from an error representing +/- 108% (77.6/-0.72 m w.e.).

Lines 704-705: Significant figures for this result are different from previous ones in the results section (different units too), make consistent.

Line 707: list numbers corresponding to these results for clarity and context

Lines 711-713: Again, I highly recommend assessing the sensitivity of the model to Wind Speed given the poor performance of this variable and the importance of fluxes using Wind Speed to the interpretation of the results.

Line 717: glacier models -> glacier mass balance models or energy balance models or glacier surface process models given that 'glacier models' includes dynamic models from the bed to the surface.

Lines 748-757: a table would help navigating this discussion point

Line 782: m w.e -> cm w.e.

Line 782-783: 'A value of 1 would occur if all precipitation were snowfall and there were no albedo feedback.' -> specify that these values are ratios and unitless (the previous sentence does so partially). The Oerlemans 2001 reference here seems generic and unneeded.

Line 786: units are now in mm

Line 787: and now in m w.e.

Line 798: MacDougall and Flowers warn that transferring model parameters between sites in a specific setting will result in increasing error depending on the number of parameters and inputs carried over. This statement inexact resulting in an improper citation.

Line 800: I believe this has been done before? Please include citations to identify where this work improves the existing work

Line 804: please include geodetic estimate here for clarity