Reviewer 1

We thank Bertrand Cluzet for taking the time to read this manuscript and the thoughtful comments and suggestions. We have updated the submission accordingly and added responses to each comment below.

In snowmelt-dominated water catchments, operational snowpack modelling is essential for water resource management and flood risk assessment. Most of the operational modelling services use highly calibrated temperature-index snow models, which may fail to capture changing snowpack dynamics amid climate change. In this paper, the authors make use of a physically-based snowpack modelling framework (AWSM, Havens et al., (2020)), whose core is the iSnobal model (Marks et al., 1999), and which includes a dedicated downscaling scheme, to perform 4 years of simulation at high resolution (50m) over a high alpine watershed (1373km2), tributary of the Colorado river. Compared to the Havens et al., (2020), the main innovations reside in the use of HRRR, a high resolution (3km2) weather forecast model to force iSnobal, and the use of a dedicated downscaling scheme for wind fields.

Model performance is evaluated against in-situ measurements of Snow Depth, snow depth high resolution aerial maps from the Airborne Snow Observatory (ASO) across the entire basin, and inflow measurements at the basin outlet. The output from the operational temperature-index model SNOW-17 is used as a benchmark.

Results show that iSnobal captures well the accumulation of snow, but seems to have a too low ablation rate, potentially due to a bad representation of net shortwave flux.

The authors conclude that this snowpack modelling chain seems suitable for further operational use, provided some improvements to the snow model are made.

Overall, I think that this paper makes a nice contribution to the snowpack modelling community, by offering a promising model workflow to perform high-resolution, distributed, physically-based simulations using downscaled numerical weather prediction models. Although this topic has already been assessed by other teams, an excellent introduction makes a very good statement for it, and the paper was, overall, very pleasant to read. I also appreciate the effort for reproducible, open-science by the authors, which is in line with GMD standards. I find, however, that the manuscript is not suitable for a journal like GMD in the present form, for the following main reasons. (1) The degree of scientific/modelling innovation is too limited for the scope of GMD (Havens et al., (2020) would probably have been a good fit). (2) I also am a bit disappointed by the fact that despite that the authors downscale the model to a very high resolution (50m), none of the processes that would justify to use such a high resolution (namely snow-vegetation interaction, wind drift, and potentially gravitational redistribution of snow) seem to be accounted for in the model. (3) The evaluation itself is not consistent with these modelling hypotheses, as comparison with ASO data should discard forested pixels, which seem to cover an important fraction of the domain.

I am very positive, though, that this work is suitable for publication, probably in another journal considering the major adjustments I suggest in the following. The authors will also find suggestions of adjustments further down, as well as inside the comments on the manuscript pdf.

We believe that this paper is well suited to the aims & scope of the GMD journal as it fits into the 'Full evaluations of previously published models' category. This work is the first to use the latest development stage of the iSnobal model forced with the HRRR NWP data and evaluates its performance with in situ point observations and spatial measurements. Notably, no evaluation or application using HRRR as the forcing data source is currently published. It also is the first to compare the HRRR-iSnobal outputs to those from the SNOW-17 temperature-index model, operationally used by the CBRFC in forecasting. It presents a path for introducing a spatially distributed physically based snow model within the CBRFC forecasting framework. Lastly, it highlights and suggests improvements to iSnobal. All these aspects combined make GMD an ideal journal with its broad reach into the snowpack modeling community.

This submitted revised paper version incorporated the suggested improvements to the paper. Namely, we would like to highlight:

- *Section 3 (Model description) was rewritten and expanded the iSnobal model description and how the vegetation and wind data is an essential input to the energy balance calculation*
- *Reasoning for the selected resolution of 50 m was expanded (see the response to bullet point 1. below)*
- *Revised and expanded discussion on why the introduction of the HRRR-iSnobal combination is an essential next step for the CBRFC daily operations (section 6.3)*

1. Consider reducing the model resolution to a scale that is consistent with the modelling hypotheses for further publications (or include relevant processes at 50m resolution)

We added the reasoning for the higher resolution:

This spatial resolution was guided by the recommendations of Winstral et al. (2014), which found that iSnobal output resolutions of around 100 m kept the errors for SWE and SWI below +/-5% during melt season compared to coarser resolutions. We further selected the 50 m resolution to keep the resampling of the vegetation data from LANDFIRE minimal and removed the need to interpolate the spatial snow depth product used in validation (section 4 Model Comparison). This resolution additionally better represents the terrain complexity, relevant for radiation calculations and interpolation of the HRRR supplied forcing data within SMRF. Lastly, we also found that the 50 m resolution does not inhibit daily run times and scales for the operational requirements.

2. Put more stress on the use of spatial data from ASO for evaluation in the body of the paper (see below).

We moved the spatial comparison figure for 2018, 2019 and Figure S6 to the main paper.

3. Increase the level of rigor and the amounts of details in the model description section (see below).

Please see the revisited model section 3 and the line-by-line responses below

Inline and Minor Comments

(General)

- please avoid using the term of "coupling" between HRRR and iSnobal, as the model relation between HRRR and iSnobal is only one-way (no feedback from iSnobal to HRRR, HRRR is "forcing" iSnobal). This is essential as atmosphere-Snow coupling is a big challenge that several teams are currently addressing (e.g. Sharma et al,. (in review) and references therein), and while it is perfectly fine to just use HRRR as a forcing, we should avoid ambiguities.

Changed the wording to HRRR-iSnobal combination throughout

Consistently, I would strongly suggest the authors to change all instances of "iSnobal-HRRR" to "HRRR-iSnobal" which would be more faithful to the causality between the meteorological forcing and the snow model.

The wording 'HRRR-iSnobal' was changed throughout the text.

- Overall, I think that the paper would strongly benefit from including more figures in the body of the paper (in particular S3-S5, and S6) rather than in supplements. S3-S5: ASO data are a wealth of information, and the paper is too shy about it!

(Abstract)

consider using the keyword of AWRS somewhere in the abstract, as this seems to be the key software that makes these simulations possible, and this study is a nice advertising 'scientific application' for it.

We understand that the reviewer is referring to the 'AWSM' module. It highlighted the need to update the description of AWSM and its role in the overall model architecture. We revisited the model description and highlighted that AWSM is only one component but not the core piece of iSnobal itself.

(3.1)

- Please revisit the model description section as it is a bit shaky and raises more questions than it actually clarifies the framework. It should clearly appear on the body of the paper whether the model accounts for wind redistribution of snow and models the canopy. This information may be available in the references (after 30mins of browsing through Marks et al, 92, 99 and Havens et al., 2017,2020, my understanding is that it is not), but such essential information should explicitly appear in the paper.

We revisited and reorganized the model section 3. to improve presentation of the necessary setup and expanded the high-level description. Please see the line-by-line comments below and the revised sections 3.1 to 3.5 in the paper.

- similarly, the description of the study area (Sec. 2) l 114-116 is too vague, in particular in terms of vegetation cover type: it seems from Hubbard et al., (2018), and Fig. 3., that the forest covers an important part of the watershed and may impact the snowpack dynamics (canopy interception, increased longwave radiation...). If that's the case, and based on my understanding that the model does not account for the presence of a canopy, then I would strongly recommend to:

* include a rough landcover type to maps on Fig. 1

* add the two unrepresented sites on Fig. 3

* discard forested areas when comparing to any observation, in order not to compare apples and oranges.

Vegetation is accounted for by the model and the role of the vegetation data was expanded in the paper. Please see the line-by-line comments below and the revised sections 3.3 and 3.5 in the paper for more details.

L 13 (suggestion) Seasonal snowpack dynamics?

Adapted

L 16 area?

Updated to:

To evaluate and improve the CBRFC modeling options, this work ran the physically based snow energy balance iSnobal model, forced with outputs from the High-Resolution Rapid Refresh (HRRR) numerical weather prediction model across four years in a Colorado River Basin forecast region.

L 17 These absolute values are not really speaking for those who don't know typical snow depth amounts for this region (e.g., this would be a huge error for an arctic snowpack). Using a percentage of improvement of this metric compared to the operational CBRFC, or, if so, just saying that performance is similar would be more informative.

We removed the absolute numbers and expanded the result section with percentage values per year and in-situ measurement stations. The updated presentation gives a better picture on how the agreement varied between the years and measurement locations.

The result section:

For the two sites within the ERW boundaries (Butte and SP), the simulated depths had a mean difference of 12 % (Butte) and 1 % (SP) in 2018, -2 % (Butte) and -15 % (SP) in 2019, - 16 % (Butte) and -21 % (SP) in 2020, and -4 % (B) and -3 % (SP) in 2021 until peak SWE (Supplement Figure S2), while the UT site had consistently higher than observed snow depth: 37% in 2018, 23 % in 2019, 28 % in 2020, and 20 % in 2021. (Supplement Figure S2).

L 30 typo

Corrected

L 32 I wonder whether it would be worth expanding. To my understanding, this is mostly true for calibrated/degree-day models, but as you say earlier in the abstract, physically-based models should much less be "overfit" to historical data. To date, some operational some operational snow-hydrology services actually use physically-based snow models

Updated to:

This presents a challenge from a modeling perspective, especially in operational settings, when historical data are the basis to create forecasts. Here, a consistent and accurate prediction of snowpack runoff is getting harder with the increasingly different snow accumulation, snow melt, and snow cover extents across the seasonally snow-covered mountain ranges supplying freshwater to the downstream regions.

L 48 (Detail) I would delete this, as a TI model with additional meteorological variables would likely not be a TI model anymore, right?

Adapted this suggestion

L 50 ... and are not explicitly calibrated on historical data (although some of their physical parameterizations rely on a limited set of measurements

We added this suggestion:

One option to improve forecasting results is the use of physically based models that incorporate more meteorological measurements and are not calibrated on historical observation data.

L 53 If I had to pick three of the most important weather observations for snowpack modeling, I would probably not include relative humidity

We updated this sentence to:

In general, physically based models use additional weather observations outside of temperature and precipitation such as relative humidity, wind speeds, and radiation, to resolve the mass and energy balances of the snowpack, which determines the snowmelt rate and meltwater runoff (Marks and Dozier, 1992).

L 54 (typo) ...runoff (

Corrected

L 55 Vionnet et al., 2012 would probably be more appropriated now. (doi: 10.5194/gmd-5-773-2012)

Updated

L 60 Even though we can guess from the name that this is the water supply modelling workflow you're untegrating iSnowbal into, please provide a bit more context on AWSM here to make it easier to understand. (something like: AWSM is the operational water supply model framework of the CBRFC,...)

L 61 I would move this one sentence up, before the introduction of AWSM

We updated the section to address the comments from L60 and L61. AWSM is part of iSnobal that would be integrated into the CBRF workflow. The new section:

Initially, iSnobal was developed as a two-layer snowpack model for a single point location (Marks and Dozier, 1992; Marks et al., 1992) and evolved later to a spatially distributed version (Marks et al., 1999), maintaining the design and input data requirements of the point-level version. A recent addition to the iSnobal modeling pipeline is the Spatial Modeling for Resource Framework (SMRF, Havens et al., 2017), which assists in distributing the forcing data across the model domain. To streamline the iSnobal data preparation and execution workflow and increase reproducibility, the Automated Water Supply Model (AWSM) combined iSnobal and SMRF into a single modelling framework (Havens et al., 2020). To this date, iSnobal has been successfully deployed to simulate snowpacks in watersheds sizes from less than 1 km^2 to over 1000 km^2 (Garen and Marks, 2005; Hedrick et al., 2018, 2020; Kormos et al., 2014).

L 69 This problem has been tackled long ago, I think that Durand et al., 1993 (doi: https://doi.org/10.3189/S0260305500011277) is probably a nice seminal reference

The paragraph context is presenting the use of meteorological data produced by numerical weather predication models (NWP). The noted paper by Durand et al. (1993) was the first novel evaluation of available meteorological data measured from observation networks as inputs for a snow model. This is a fundamentally different approach and requires different techniques to prepare the data before using them in a snow model. We think that the addition of this citation is not appropriate, especially since this paper is not using in-situ observation data outside of model validation.

L 72 (typo) \citet{...}

Corrected

L 78 Not very useful

We expanded to be more specific 'fourth and final' iteration. We think this is a useful piece of information, indicating the maturity of the model.

L 79 Potential users may be interested in HRRR domain (I guess it is not global)

We added this information: The HRRR product outputs are delivered at the 3 km spatial resolution for the continental United States and the Alaska domain.

L 95 I think that this sentence is not useful (this is a weak argument in contrast to the more solid scientific arguments you're using. Saying that the "Bureau of Reclamation" supports the use of physically based models is probably way enough

Changed to:

A recent report by the United States Bureau of Reclamation evaluated iSnobal as a "flight qualified" product and supports the increased use of physically based models.

L 98 references?

The examples are given later in the sentence

L 115 This paper does not appear in the references.

The missing reference was added

L 117 measuring HS

Updated to:

There are three Snow Telemetry (SNOTEL) stations with available snow depth measurements that are operated by the United States Department of Agriculture National Resource Conservation Service (USDA-NRCS) in the modeled domain.

L 121 indication on the altitude boundaries could be interesting for the reader

We added the elevation bands to the text

Each HRU is based on elevation (2896 m/9500 ft < middle < 3353 m/11000 ft) and are modeled independent from each other (Council, 2006).

L 126 Please point towards the corresponding reference to complement this very rough model description

A list of references is given on line 184-185:

A detailed description of the energy balance equations and data preparation can be found in Marks et al. (1992), Marks et al. (1998, 1999), Link and Marks (1999), and Havens et al. (2017; 2020).

L 128 are the soil layers part of iSnowbal, or of an exeernal model? Is it a "canonical soil, or is it parameterized with physiographic data from the " Basin setup tool"? please clarify.

We changed the description to:

The bottom layer acts as an interface for mass and energy exchanges between the top and a single soil layer.

L 133 (curiosity) wind direction, really? interesting, is there any lateral redistribution of snow by wind, or a topographic parameterization of turbulent fluxes? this would definitely be worth mentioning on the model description

We added more information on the use of wind data and options in SMRF:

From the downscaled wind data from the Katana module, SMRF interpolates the wind speeds to the configured output resolution. The SMRF options include the ability to account for precipitation redistribution using the speed and direction wind data according to Winstral et al. (2002). However, this option was not used in this study and reduced potential error sources when comparing the HRRR precipitation input data to Snow-17, which is explained in section 4 below.

L 133 already said earlier, I think you could delete.

Removed

L 133 use "shortwave radiation" instead of solar radiation throughout. Also, I thik that this is 'net solar radiation' in your case.

We adapted the use of 'shortwave radiation' where appropriate.

L 134 "downscaled" is much more appropriated (and rings a bell) since you go from 3km to 50m. Whether/how much driect/diffuse shortwave and longwave downscaling account for the topography is also worth mentioning there.

We expanded on the shortwave and added the longwave flux determination by SMRF:

The cloud cover itself is determined by the supplied incoming shortwave radiation from HRRR. An additional adjustment to the incident shortwave radiation reduces the value by vegetation type and height information from the vegetation metadata following Link and Marks (1999). The amount of reflected shortwave radiation is calculated using a simulated snow albedo derived with a time-decay function based on the time duration since last snowfall. The difference between the cloud and vegetation-adjusted incident shortwave and the reflected shortwave results in the net shortwave radiation at the snow surface.

L 135 please list them

We added:

The variables include information for precipitation phase, precipitation temperature, percentage of snow in precipitation, and the vapor pressure.

L 136 I don't get it, (1) isn't snow albedo a state variable from iSnowbal? If so, how can you then compute the net solar radiation externally from iSnowbal? (2)isn't iSnowbal forced by total incoming shortwave rather than net shortwave?

We re-visited and re-organized the entire section 3. This hopefully improves the overall description which energy inputs and energy fluxes are calculated at what stage in SMRF and iSnobal. To answer the questions here:

(1) The albedo is calculated with each time step and saved. The albedo calculation is done in the SMRF module and further used (see below answer to 2.).

(2) The iSnobal model input is net solar energy, which is calculated from incoming shortwave radiation and the snow albedo by the SMRF module. It is calculated as part of the data preparation step to run iSnobal.

L 145 a few words on what is a docker container would be appreciated

We added:

A Docker container is a virtualized environment that combines the required operating system, libraries, and the application itself into one executable entity.

L 151 This is not just tool, it's a core component of your modelling setup, (downscaling part). As such, I would include its description rather in the model description section. I also recommend higlighting that this is an addition compared to previous studies using using the same modelling framework.

We expanded on the use of the wind data with the updates to the text from comment L 133. Albeit unpublished to date, Katana has been part of the model execution chain and is not newly introduced with this work. We therefore do not want to put too much focus to this element. The interested reader is referred to the publication by Forthofer et. al (2014) for more details on WindNinja.

Figure 2: the two arrows out of HRRR are disturbing at first sight. Consider labelling the bottom one with "wind" or "(u,v)" in order to raise ambiguities

The figure was revisited and also adapted the suggested labeling of "wind"

L 162 I'm not a native speaker, but this formulation sounds awkward to me.

Changed to: Computing Environment

L 173 downscaling?

Corrected

L 174 Try to be more rigourous/specific here. I think that you mean the direct/diffuse partition of incoming shortwave radiation 174-177: please revisit the description of the handling of shortwave radiation in a more rigorous/precise manner.

The section 3 was completely revisited and addressed the comments for Line 174 - 175. Please see response to Line 133 and Line 134 for the updated text.

L 174 Why don't you go all the way to the model 50 resolution? this deserves an explanation.

The section 3 was completely revisited and addressed the comments for Line 174 - 175. Please see response to Line 133 and Line 134 for the updated text.

L 175 what is the cloud factor? do you mean, "nebulosity"? L 177 on? L 177 and what about the diffuse radiation?

The section 3 was completely revisited and addressed the comments for Line 174 - 175. Please see response to Line 133 and Line 134 for the updated text.

L 179-182: Whether your model includes wind drift/ preferential deposition of snow is really essential here as according to Winstral et al., (2014), 100-250m would have been enough if these processes are disregarded.

We added the reasoning for the selected output resolution. Note that this section has moved to 3.2 and is now earlier in the tex:

This spatial resolution was guided by the recommendations of Winstral et al. (2014), which found that iSnobal output resolutions of around 100 m kept the errors for SWE and SWI below +/-5% during melt season compared to coarser resolutions. We further selected the 50 m resolution to keep the resampling of the vegetation data from LANDFIRE minimal and removed the need to interpolate the spatial snow depth product used in validation (section 4 Model Comparison). This resolution additionally better represents the terrain complexity, relevant for radiation calculations and interpolation of the HRRR supplied forcing data within SMRF. Lastly, we also found that the 50 m resolution does not inhibit daily run times and scales for the operational requirements.

L 181 Which processes? Please list them down.

The sentence was revisited as a response to the comment on Line 179

L 183 How is this vegetation data used by the modelling chain? It does not appear in the model description section.

We expanded the description for shortwave and longwave radiation determination by SMRF that now also indicates where the vegetation data is used. Please see the response to Line 134 for the updated text.

L 186 I assume that the snowpack runoff SWI is the daily integral of melt, and not the end-of-day value

Correct, and we added this information:

The model was configured to store the end-of-day values (e.g. snow depth) or accumulated daily values (e.g. SWI), matching the temporal resolution of the available in-situ comparison observations and the SNOW-17 inputs and outputs, which are further explained in the model comparison section.

L 188 I don't get it, do you mean that you did a one-year spinup? But why do you need that? I don't see any mention of a ground model/perennial snow on the domain that would justify such a spinup.

The sentence was revisited to indicate that the two dates reflect the start and end date for one model year:

A single simulation year starts on the 1st of October and ends on the 30th of September. This date range is a 'water year' and a standard definition for hydrologic forecasting in the United States, defined by the USGS.

L 196 Because of the rather rough model description, it is not clear to the reader which physiographic features may influence the model output, and which not. My point is that at such fine scale (50m), if vegetation and wind redistribution are disregarded, topography alone (temperature and SW radiation) should not induce significant differences in smooth terrain.

We expanded the description as a response to comment Line 134 and 179. The vegetation data is used to correct shortwave and longwave radiation. Please see the updated text in response to Line 134.

Figure 3: which one of the three SNOTEL sites?

Added the name of the site: Schofield Pass

L 231 Reference to Fig. 4 is missing

This opening paragraph is a summary for all the compared years. The reference to Figure 4 is found later in the text, when describing the shown years 2018 and 2019.

L 233-239 still understanding that wind drift and vegetation are not accounted for, I could not wrap my head around the considerable variability in the modeled snow melt rates of the 4 neighboring pixels , at only two of the three field sites (Fig. 4 and S1).

We updated the text on how the vegetation data and turbulent fluxes are relevant processes in the model. Please see the updated text to the comment on Line 134.

L 237 "observation" is a bit misleading, please rephrase

We updated this to:

This site-specific result was consistent across all comparison years and not affected by annual differences in the accumulation magnitudes.

L 240 this paragraph is very difficult to follow.

We updated this paragraph to:

A cross-comparison of snow depths on ASO flight days between the observed in-situ SNOTEL sites, ASO remotely sensed, and HRRR-iSnobal simulated values showed a mixed result for agreement. The early season ASO flight in 2018 had HRRR-iSnobal with higher snow depths around the Butte (Mean difference

(MD) +0.35 m) and SP (MD +0.20 m) sites compared to ASO. The difference between HRRR-iSnobal and SNOTEL were small, with the Butte site ranging from +0.10 m to +0.13 m and the SP site from +0.04 m to +0.08m. The late-season flight, where snow was only present at the SP site, had the HRRR-iSnobal snow depths closer to ASO (MD -0.07 m) and above the SNOTEL values (+0.20 m to +0.54 m). The UT site was not included in both ASO surveys in 2018. For 2019, the early season ASO to HRRR-iSnobal snow depth differences varied with higher HRRR-iSnobal values at Butte (MD +0.25 m) and UT (MD +0.39 m) and lower values at SP (MD -0.22 m). The HRRR-iSnobal to SNOTEL differences were low at Butte (0.00 m to 0.04 m), negative at SP (-0.32 m to -0.43 m), and positive at UT (+0.16 m to +0.61 m). The late-season flight in 2019 had ASO higher than HRRR-iSnobal at Butte (MD +0.19 m) with no snow measured by SNOTEL, while SP had HRRR-iSnobal lower than ASO (MD -0.31 m) and an overlapping range with SNOTEL measured (+0.11 m to -0.66 m). The UT site had HRRR-iSnobal much higher snow depths than ASO (MD +1.25 m) and SNOTEL (+0.71 m to +1.55 m). Overall, using ASO as an additional snow depth reference data set at discrete point locations in the model domain showed no consistent over or under-simulation for HRRR-iSnobal.

L 254 ... but this delay varied...

We updated this sentence to:

Notably, the date for snow disappearance was simulated later relative to observations across all years and varying difference of days.

L 429 Acknowledgements

Updated

L 432 Fair enough. I have nothing against this statement, but from a non-Northern-American perspective, this statement may look somewhat political and I wonder whether this should take place in the acknowledgements section of a scientific paper. And what tells us that the USDA is actually fair?

This added sentence is a required element for publications with associations to the USDA and its employees.

L 505 Please check reference formatting (the journal name is missing)

Added the journal

L 546 repetition

These are the two publications describing the initial design of the Snobal model

(Figures)

- I found the interpretation of figures (S3 and Fig. 5) a bit difficult due to the unusual computation of model bias/difference (even though I appreciate the effort of the authors to indicate model under/overestimation, I think that it would ease the readability just to follow conventions). Usually, people take the observation as a reference, and compute “bias = model – obs”, resulting in positive bias for model overestimation and negative bias for underestimation. I strongly recommend using this convention throughout.

We updated the comparison to the suggested convention

- In S3, blue (resp. red) is usually the color used for model overestimation (resp. underestimation of snow amounts, since it looks wet (resp. dry).

We adapted the suggested color convention

Reviewer 2

We thank reviewer 2 for their time and comments. We have revised the submission accordingly and added responses to each comment below.

This manuscript presents a nice contribution to the snowpack modeling community, through a robust assessment of the operational capability of the HRRR-forced iSnobal model to reproduce observed snow depths (point-based measurements from snow telemetry stations and lidar snow depth maps) and stream gauge discharge in the East River Watershed (Colorado). I am, however, left a bit disappointed by the lack of in-depth evaluation of the iSnobal model versus its current operational alternative, namely the temperature-based index model SNOW-17. In the introduction, the reasons for evaluating HRRR-iSnobal are very well presented, but I find that after reading the paper, the benefits of using HRRR-iSnobal over SNOW-17 are not made clear enough. I also do not understand why the authors use such a fine resolution (50 m) for the iSnobal model even though most of the relevant processes at these scales (such as interactions between snow, wind and vegetation) are not currently represented in the model. Justification for downscaling the model to such a fine scale is currently not clear to me. Apart from these two general comments, I have only minor comments, highlighted below. I therefore recommend publication of this manuscript once both my general comments and these minor comments have been addressed.

The advantages to use HRRR-iSnobal have been expanded in the discussion section 6.3. The section highlights the aspects on why we believe the introduction of HRRR-iSnobal will result in operational benefits for the CBRFC:

Accurately simulating the seasonal snow accumulation and depletion across multiple years and different topographic regions through a single model is still an unsolved challenge. However, physically-based models are better suited to fulfill this need when compared to temperature index-based models. Existing studies using long-term records of SNOTEL measurements have identified a need for physically based modeling approaches as they can better show the impacts on water supply from current and projected climatological changes (Harpold et al., 2012; Musselman et al., 2021; Trujillo and Molotch, 2014). Physically based models are capable of accounting for scenarios such as shifting snowmelt rates (Musselman et al., 2017), changes to the length of snow accumulation and melt season (Trujillo and Molotch, 2014), accelerated melt due to darker snow (Skiles et al., 2018), and rain on snow events (Marks et al., 1998).

The HRRR-iSnobal combination in this work closely simulated the snow depth observations from one region across different seasonal characteristics (average vs. non-average year) without a priori region-specific knowledge. All required forcing data per water year only used the HRRR forecasted observations of that year, replacing the sparse in-situ measurements that are interpolated and calibrated with historical data. The interpolation and calibration process is an essential step in the current Snow-17 workflow of the CBRFC and requires forecaster experience per region. Voiding this need gives a major operational advantage, simplifying model application and enhancing the ability to adapt to environmental changes across different regions and water years.

In addition to the streamlined forcing data preparation, the advantages during the water year simulation include the replacement of region-specific parameters required by index models, such as melt-factor in Snow-17, by the energy balance calculations in physical based models. The index model parameters depend again on forecaster experience and require forecaster changes in response to changing environmental conditions (average vs. non-average). In this work, there was no need to change the model configuration between the 2018 (average) and the 2019 (above-average) years in HRRR-iSnobal. The seasonal differences were accounted for explicitly as the snow energy balance calculation includes influences such as shortwave (Marks and Dozier, 1992), longwave radiation (Link and Marks, 1999), and turbulent fluxes from terrain-dependent wind speeds (Winstral and Marks, 2002).

Lastly, iSnobal is a spatially distributed model run with a user-defined pixel size (50 m in this study) and set up with site-specific elevation and vegetation data information from the model domain. This configuration negated the need to divide the model domain into uniform and simplified HRUs to simulate the topographic differences of the model domain accurately. In summary, physical based models remove the dependence on historical data, reduce the need for model application knowledge, and better account for physiographic factors, which result in improved scalability across different seasonal and terrain-dependent snowpack dynamics. Gradually adding these models into operational settings, with architectures presented in this work, can provide snowpack response information with current environmental perturbations. Their addition will enhance the quality and expand the ability to adapt to current and future water supply forecast needs.

We have also expanded the text on why we chose 50 m resolution. The addition includes added details to the model description in section 3.3, where the relevant processes in the model include longwave radiation and wind. Below is the expanded paragraph from the manuscript on the reasons for the 50 m resolution:

This spatial resolution was guided by the recommendations of Winstral et al. (2014), which found that iSnobal output resolutions of around 100 m kept the errors for SWE and SWI below +/-5% during melt season compared to coarser resolutions. We further selected the 50 m resolution to keep the resampling of the vegetation data from LANDFIRE minimal and removed the need to interpolate the spatial snow depth product used in validation (section 4 Model Comparison). This resolution additionally better represents the terrain complexity, relevant for radiation calculations and interpolation of the HRRR supplied within SMRF. Lastly, we also found that the 50 m resolution does not inhibit daily run times and scales for the operational requirements.

Inline Comments

To me, the use of the word "coupling" is incorrect, as in this work iSnobal is simply forced by HRRR data.

Changed the wording to HRRR-iSnobal combination throughout

l. 30: the full stop is missing at the end of the sentence.

Corrected

l. 42: "require" should be plural.

Corrected

l. 72: please consider revising the citation to: "in Havens et al. (2019)".

Corrected

l. 143: "and is available".

We revisited section 3 and adapted this suggestion

l. 196: the term "ASO" has not been defined yet.

We moved the definition to the first occurrence

l. 247: there is a word missing here: "and positive at Upper Taylor".

Updated

Figure 4: please consider enlarging the figure, as currently, the labels and text inside the figure are too small. For example, the two sub-figures could be placed on top of each other instead of side by side.

We increased the figure height to improve the readability and kept the width. Keeping the width enabled the side-by-side comparison for an 'above-average' (2019) vs. an 'average' (2018) snow depth year.

l. 293: there is a word missing: "across all sites in 2020".

Corrected