A review of: "*The importance of model horizontal resolution for improved estimation of Snow Water Equivalent in a mountainous region of Western Canada*" by Samaneh Sabetghadam Christopher G. Fletcher and Andre Erler

Brief Summary: In this paper, the authors attempt to show the importance of high-resolution climate models for estimating SWE in mountainous regions. The study basin has 9 available CanSWE snow pillows with a continuous record. Comparisons of model SWE vs CanSWE records are made for 5 modeled datasets: WRF 9/3/1km, ERA5 and ERA5-Land. The authors chose to aggregate the 9 stations for the bulk of the analysis, while the individual stations are investigated in later sections. A brief investigation into the role of elevation and elevation bias in the models are made as well. The authors conclude that WRF at 1/3km resolution can provide a good estimate of SWE values in the South Saskatchewan River Basin and thus can be used for flood/runoff forecasting.

Overall Thoughts: The paper is well written and easy to understand. I have minor comments for some of the language used. I have two major concerns that I believe the authors should address prior to publication: I believe not enough snow pillow sites were used in the analysis, and the comparison of SWE point data to gridded model SWE is not justified adequately. I have included references of papers that expand upon these major concerns.

Major Comments:

1. This study uses only 9 snow pillow sites for an area of ~ 100km x 60km. I have concerns that the number of stations used is too low. The authors do not present any work to suggest that this number of stations is statistically significant and thus viable to compare against the gridded data set. With a sample size of only 9 stations, the aggregation of the stations to compare against the modeled SWE does not seem appropriate. Further work to show that these results are statistically significant and unbiased.
    a. In line with this comment, I suggest the authors include standard deviation and/or mean absolute error metrics for their results.
2. The SWE time series shown in figure 2 are based off the spatially aggregated station results. However, when you look at the SWE time series for the individual stations (figure 5), the results seem at odds with figure 2. There are several instances of the "best" data product (WRF 1k) differing from CanSWE by what appears to be 50% or more. Therefore, I suspect the aggregate results is likely averaging out the mischaracterized stations. This discrepancy is glossed over in the manuscript and takes away from any claims the authors attempt to make about the relative value of high-resolution WRF runs. Again, calculation standard deviation and mean absolute bias/error for the individual stations and expanding the discussion of those results would be helpful.
3. The comparison of point data SWE to gridded data SWE is difficult *(see Blöschl, 1999)*. By design, snow pillows are not representative of typical mountain terrain. Snow pillows are placed in (relatively) open and flat clearings where they can accurately report on the mass of falling snow while being protected from the wind from surrounding tree cover. However, the processes that impact the actual spatial distribution of snow in mountains

are not well represented by the local environment that is conducive to snow pillow placement. Many of the processes that govern snow depth in mountain environments are happening on sub 1km scales. Preferential deposition and vegetation presence are important factors that operate on the <100m scale, which is not resolved by any of the model resolutions in this study. See Clark et al., (2011) for more details on these processes.

The authors claim the higher resolution modeled runs are more accurate, but the results presented do not show this. They show that the spatial variability of the grid cell sizing is impacting the SWE estimate from the model. This can be clearly seen in the elevation bias component of the results section (specifically figure 10). Looking at Station 1, the elevation bias for the 3km and 1km model runs are ~+300m and ~+260m respectively. This results in a massive overestimation of snow accumulation by the model. Meanwhile, the 9km model run has an elevation bias of only -30m and shows the typical WRF underestimation of SWE (i.e. Wrzesien et al., 2018)

   a. How can the authors be sure their results are showing the differences in model resolution as opposed to the difference in spatial variability due to the elevation bias in the model?
   b. See *Scaling Issues in Snow Hydrology* (Blöschl, 1999)for more details.
   c. See Rice & Bales (2010)and Dressler et al., (2006) for more info on snow pillow site selection.

Minor Comments:

1. The authors do not refer to the model simulations with consistent language. For example, figure 3 refers to them as WRFD1/D2/D3. While the plot colors match figure 2 which refers to them explicitly as WRF(9km)(3km)(1km), the reader is forced to double check against the figure. I suggest the authors define a consistent naming format early in the paper and stick to it. It could be as simple as “the 9 km WRF runs will be referred to as WRF9K for the remainder of this paper” etc.
2. The authors bounce around between their use of increase/decrease, fine/coarse, high/low in a confusing manner. A more consistent verbiage would make it easier for the reader. Improving the precision of language would increase the readability of the paper. For example, phrases like "less than 4km resolution” are very confusing.
   a. Interpreted literally, one could assume they meant “a resolution with a numeric value lower than 4”. i.e., a finer resolution
   b. However, it is equally fair to interpret that as “a resolution with less spatial fidelity than 4km” i.e., a coarser resolution.
3. The introduction reads poorly. The layout does not flow well and feels disjointed in some of the paragraph transitions. Please consider restructuring.
   a. Line 97 has an extra “ERA5 and ERA5-land” in it.
4. The hypothesis of this entire manuscript feels unclear to me. I’m unsure if the authors believe WRF estimates of SWE are appropriate for use at basin scale, grid scale, or both.

    a. My inclination is that they believe it is appropriate for basin scale estimation, but discussion around the analysis of individual stations also leads me to believe they are implying some level of reliability in estimating SWE at grid/point scale.
    b. I highly suggest the authors refine the scope and goals of this manuscript and reframe the title/hypothesis/abstract/conclusion appropriately.
5. I feel that the inclusion of ERA5 and ERA5-L in the analysis is mostly left out. In the current state it feels like it has been added just to have an extra point of comparison. I feel that the authors should expand upon this analysis in a written discussion or remove it.
6. The investigation in elevation bias in section 3.3 is interesting but not fully explored. I suggest the authors expand on this section.
    a. Line 301 states "when elevation biases are largest (at 9km)" but both figure 10 and table 2 show that the largest bias (both positive and negative) occur in the 3km runs. Please clarify if this is an error or if some other metric (possibly mean absolute bias?) was used to justify this statement.

In conclusion, the work presented in this manuscript has significant issues in the underlying assumptions made by the authors. Should the authors reframe their hypothesis and conclusions based on their results, a further iteration of this work could be publishable. The defense of the statistical significance of this work with only 9 snow pillow sites is required. Changes in their justification of their comparison of snow-pillow SWE to gridded model results is necessary. A significantly expanded discussion of the individual station analysis is necessary as well. Further statistical error metrics are necessary for those individual station results. Finally, I recommend the authors expand upon the elevation bias section of the manuscript.

References:

Blöschl, G. (1999). Scaling issues in snow hydrology. *Hydrological Processes*, *13*(14–15), 2149–2175. https://doi.org/10.1002/(SICI)1099-1085(199910)13:14/15<2149::AID-HYP847>3.0.CO;2-8

Clark, M. P., Hendrikx, J., Slater, A. G., Kavetski, D., Anderson, B., Cullen, N. J., Kerr, T., Örn Hreinsson, E., & Woods, R. A. (2011). Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resources Research*, *47*(7), 2011WR010745. https://doi.org/10.1029/2011WR010745

Dressler, K. A., Fassnacht, S. R., & Bales, R. C. (2006). A Comparison of Snow Telemetry and Snow Course Measurements in the Colorado River Basin. *JOURNAL OF HYDROMETEOROLOGY*, *7*.

Rice, R., & Bales, R. C. (2010). Embedded-sensor network design for snow cover measurements around snow pillow and snow course sites in the Sierra Nevada of California. *Water Resources Research*, *46*(3). https://doi.org/10.1029/2008WR007318 open_in_new

Wrzesien, M. L., Durand, M. T., Pavelsky, T. M., Kapnick, S. B., Zhang, Y., Guo, J., & Shum, C. K. (2018). A New Estimate of North American Mountain Snow Accumulation From Regional Climate Model Simulations. *Geophysical Research Letters*, *45*(3), 1423–1432. https://doi.org/10.1002/2017GL076664