Ma et al. investigate the use of machine learning for track-to-area transformation of satellite observations of SWE in swathes that cover a small fraction of a watershed (it is worth noting that point-to-area transformation with horizontal and vertical distance weighting has long been used for assimilation of in situ snow observations in numerical weather prediction). This study is entirely synthetic, but interesting in anticipation of real remote sensing opportunities. Some clarifications and numerous minor corrections are required. Explainability of the ML results is promised, but there is very little physical explanation in the discussion of the input feature sensitivity tests.

LiDAR can produce continuous SWE maps; the Airborne Snow Observatory does for this very basin.

"TTA transformation could be achieved by leveraging snow pattern repeatability", but that is not what is done here.

Is a 50% perturbation in temperature applied to the Kelvin temperature (which would be enormous) or Celsius temperature (which would be meaningless)? A 50% error in air pressure is also unphysical. To be clear, is the perturbation a fixed bias (as described) and not a random error? If so, why does the DNN not just learn the bias?

What is intended by saying "a simple ANN ... is capable of learning non-linear relationships"?
... while DNN ... can learn more complicated relationships”? More complicated than non-linear?

Levenberg-Marquardt is not the cost function, it is the algorithm used to minimize the cost function.

Time series of SWE observations would not be available for wet snow after peak SWE.

There is no discussion of why removing a particular variable will sometimes increase and sometimes decrease MAE in Figure 8.

The biases in Figure 9 are fractions, not %. Why are there non-zero changes of MAE for zero bias in WY 2015 and 2017?

I do not follow the argument for why positive longwave radiation errors cause larger errors. Is it that decreasing longwave radiation cannot increase 1 April SWE above the cumulated snowfall, but increasing longwave radiation can decrease SWE all the way to zero?

The legend “Observation coverage” could be removed from Figure 10 so that it does not overlap the ground tracks (and four decimal places is excessive for the percentage).

Minor corrections

“it therefore, mitigates” – delete comma

“whether large amounts of snow accumulate”

“spatial, rather than point observations” – delete comma
“except during periods of precipitation”

“topographically complex areas etc. (Le et al. 2017),” – delete etc. and comma

“snow dominated”

“Each of the experiments used one of four algorithms”

“We investigated the performance of the SWE TTA estimation”

“from the closest previous observation day”

“for all four algorithms”

“The reason for larger values”

“for all three years”

It would make more sense to say that April 1st SWE is highly correlated with cumulative winter precipitation.

“The error range is larger in WY2008 than in the dry year”
“underestimation tends to occur”

The information in this first sentence is repeated in the next sentence.

“(dry, average and wet)”

Delete “are more obvious”

“are shown in Fig. 7”

“becoming more apparent”

“that provides the most useful information”

“The dominance of precipitation is most significant”

“larger biases have larger impacts”

“so SWE estimates from the network are slightly different”

“In addition, we propose”
“is larger than two”

581
“if more than two ground tracks pass”

603
“more satellite overpasses do not improve the estimates much”

621
“can be used”

628
“and wet years”

634
“The DNN method is the most accurate”

635
“And reduction in the training data size”