

This is an interesting alternative to the NPR method. Essentially, look for the minimum TbH for the season, extract the soil temperature on that date then use that as the threshold temperature for freeze-thaw. The time series seem to show a clear response in most examples. Following the reasoning, it is only possible to retrospectively identify periods of melt and freeze and that thresholds differ by soil temperature depth as well as by year. It is worthwhile exploring further. There are a number of major points of clarification, reorganization, expansion, and presentation changes that should be addressed prior to publishing. The discussion nicely points out that there is a physical reason why the microwave signal responds differently as the soil thaws and that a physically based approach is better than an arbitrary threshold. One would imagine that this paper provides an entrance point to using more physics to effectively detect FT state from passive microwave remote sensing.

It was not clear from the results if a single threshold is recommended or varies; how much the threshold vary by site and year (add to Table A1). Most thresholds that were discernible in the figures appear to be within the error bounds of the instruments (~ 0.1 oC). Modeled soil temperatures during winter are well understood to be problematic, particularly with overlying snowpacks (see Moradi et al. 2023 for an overview). The Holmes et al. (2012) reference also notes that and only provides performance metrics for non-winter conditions (April forward or $T > 0^{\circ}\text{C}$). If the authors wish to suggest that a product can be used to set a threshold, then the GEOS-5 surface temperature errors should be established - at least for these sites.

It also isn't clear that the numerous comparisons to the "gold standard" makes sense. That is, simply setting the temp threshold to the gold standard would give perfect performance. A better comparison would be to acknowledge that both thresholds are likely flawed and that the authors do not have the dataset that might be called "truth". Section 3.3 focuses on improvement, but without a clear measure of soil frost, at best this is a comparison among methods and should be stated as such. As it stands the high accuracy with the revised method appears to suggest that the "gold standard" is fine. A better approach might be to highlight agreement and disagreement among the three methods. Could the soil moisture sensors be brought in to provide an alternate measurement for comparison?

The SurFC vs SurTCs shown in figure 3 and 5 are quite interesting and the hysteresis that is exhibited may be the most novel finding of the work. For the same soil temperature there are a wide range of TbH values. The TbH evolution follows different paths for thawing versus freezing. Section 3.1 is titled *SurFCs and SurTCs from in situ temperature sources* but the results and the figure captions appear to merge the two phenomena. It would be useful to explore the multiple pathways of thaw versus those of freeze. Recommend that section 4.1 be greatly expanded and more results about the curves be presented earlier. Be clear throughout when the general curves are being referenced versus SurFC or SurfTC. Dan Glaser's suggestions should be helpful.

There appears to be an acknowledgement that snow melt and soil thaw are contributing to the signal in section 4.1. There is a nice discussion about the snow wetting decreases during refreeze (crust and grain metamorphism) in some of Joan Ramage's work (and others). It would be useful to better connect the dots regarding what is occurring and why a single Tb is a better approach than NPR or DAV, for that matter. Recalling that following snowpack ripening, melt onset has two different phases, it is entirely possible that there is a DAV phenomenon that is occurring which results in a strong response prior to seasonal snow melt, water release from the snowpack, and soil moisture. One does wonder about how DAV would perform compared to the other methods but recognize that is a significant scope expansion.

It is challenging to determine whether the microwave signal is due to snowpack changes or soil state changes. Information on snowpack conditions appears to be available (Figure 13) and would be extremely useful in understanding both the snowpack and the soil evolution. Figure 13's SWE time series shows ablation periods that do not align with expected air temperature values.

The example in 3.3.1 appears to show a flaw in the method for extracting the minimum TbH. The minimum occurred long after the soil had thawed. In March, there was a snowmelt period that likely maintained temperatures near zero for the melt duration. The early May period appears to be a local minima value that would be used.

Other comments:

Abstract wanders quite a bit and needs to be rewritten.

L 52 remove duplicate text

2.1.3 It isn't clear which GEOS-5 FP Teff was used. What is the GEOS-5 FP surface temperature? What time? Were C and K values noted used? Is MERRA from Holmes et al. the same product as referenced here? Again Holmes et al.'s errors are not representative for this study area.

L195 to 199 lots of typos (see throughout the document).

L 198 needs a reference

Figure 1 Add SMAP pixel outlines

L 250 which temperatures? Air or soil

LL 259-260 Are you trying to capture the wet snow or the soil thaw? This could use a reference.

Figure 2 Need to seem more examples with the same pixel and multiple years or neighboring pixels from the same year (more comments above).

Figure 3 No need to invert the y-axis; match figure 2

Figure 4 Need to show how TbH min and associated temperatures were obtained – suggest clearly indicating on figure rather than burying in very small font

Figure 5 Are there multiple thaw branches? If yes, would like to see some of these labeled. Sect 3.3 redefine the “thermodynamic temperature”, recommend replacing the section as noted above. If not, Is this using GOESS? What are the thresholds for the various years and pixels?

LL 392 What does this mean?

Conclusions: Because the study location is fairly limited and do not clearly demonstrate that the 0.15oC classifier is not useful in many or most locations, the conclusions should better reflect the findings. Conclusions need to add caveats about modeling errors, locations (describe where, when types of conditions); the push towards SFC and comparisons to observations is not presented in results only discussion and seems like an afterthought – albeit one that should be pursued more rigorously. L515 point 2 was not shown – it happened to work fine and the performance difference was not quantified. “snow-influenced hydrologies where snowmelt dominated the L-band signature.” Was not shown – almost no information was provided about snow and snowmelt. Ll 532-534 Is winter snowpacks and end of the season rainfall common? Did NPR perform under these conditions? L540 airborne is extremely unlikely because of the number of overpasses required to capture the signal – microwave is only used to determine the threshold.

Moradi, M., Cho, E., Jacobs, J.M. and Vuyovich, C.M., 2023. Seasonal soil freeze/thaw variability across North America via ensemble land surface modeling. *Cold Regions Science and Technology*, 209, p.103806.