Response to Reviewer 1:

We thank Dr. Montpetit for taking the time to read our manuscript and provide a constructive and thorough review. We did our best to address each comment and believe the revised manuscript will be greatly improved. Our responses to each comment were highlighted in pink.

1. General Comments

This paper calibrates a Random Forest (RF) model using dynamic and static variables to identify snow phenology. Among the dynamic variables, the authors use the MEaSUREs passive microwave (PMW) brightness temperatures (TB) and modeled snow cover (IMS/SnowModel). They then apply the RF classifications over a 35-year period to identify phenology trends.

The methodology is sound, and the manuscript is generally well written. However, some clarifications are needed to improve the manuscript, and a few details still require discussion and analysis. That said, this study provides valuable insights and will benefit the cryosphere research community.

Thank you for the overall positive review.

Although RF is an effective tool for multi-variate predictions, significant work has also been done using Neural Networks (NN), which are now widely accessible and have proven to be more robust in linking snow properties to PMW measurements (Forman and Xue, 2017). While it is not necessary to reprocess this study, NN should be considered for future implementations. NN are particularly effective at simultaneously identifying spatial and temporal relationships within datasets, which could benefit this study.

We agree with your comments. We will discuss the use of alternative machine-learning methods in particular NN in the revised manuscript.

That said, Xiao et al. (2021) did show that RF outperformed Artificial Neural Networks (ANN) in retrieving fractional snow cover (FSC) using the MEaSUREs TB dataset. This latter reference is relevant and should be cited in the manuscript.

We will cite Xiao et al. 2021 to our text on RF application in the cryosphere.

Finally, some questions remain regarding the accuracy of the calibrated MEaSUREs TB dataset. Since this study does not use radiative transfer modeling, the TB dataset is still appropriate; however, the calibration uncertainty should be addressed within the RF model and compensated for by including additional input variables.

2 Major Comments

The first comment I have is in the processing of TBs. Most studies using TBs at those frequencies will apply an atmospheric correction to remove the impact of the atmospheric water vapour. Dolant et al. (2016) mentions that at 37 GHz, the atmospheric contribution to the measured TB can be up to 10 K. This is 20 ~ 10 % of the signal which is not to be neglected. It will also have a significant impact on the RF model predictions since no atmospheric information is included in the predictor variables. I strongly suggest correcting TBs with the methodology of Liebe (1989), or at the very least include atmospheric water vapour from a reanalysis dataset as an input variable to the model to account for atmospheric conditions.

In the updated manuscript, we will incorporate total column water vapor from the MERRA-2 reanalysis product as an additional dynamic predictor variable in the RF model. This variable will be extracted at 6 PM Alaska Time to align temporally with the evening overpass of the satellite sensors used to derive TBs. Including this predictor allows the RF model to account for atmospheric variability without explicitly correcting the TBs, preserving consistency with previous studies while improving model robustness.

We will also add a discussion in the methods and limitations sections of the revised manuscript to address this point directly and clarify how atmospheric effects are now considered in our modeling framework.

Another concern I have about this study is that, it was shown by Meloche et al. (2022), using an RF model, that snow depth is strongly linked to topographic features. In this study, it seems it has little importance. I would suggest including similar topographic features from high-resolution DEMs to ensure that topographic information is well captured by the RF model. There are also known issues with TB measurements in high-topography areas (Xiong et al., 2022), and this could help improve the predictions in this study.

This leads to another concern, using TB data in high-topography areas. TB measurements in those areas show high errors due to many factors, one being local incidence angle variation. As mentioned above, adding more topographic information could reduce the uncertainty and could improve the RF model reliability in these areas.

We thank the reviewer for raising these important and insightful points. We fully agree that topography plays a critical role in snow distribution and microwave signal behavior, particularly in complex mountainous terrain.

In our original implementation, we used the GTOPO30 digital elevation model, which has a spatial resolution of approximately 1 km. We acknowledge that this coarse resolution may have failed to capture key variations in topographic features across the region, thereby limiting the RF model's ability to leverage terrain-related predictors effectively.

In response to your recommendation, we will incorporate higher-resolution topographic data derived from the ALOS World 3D-30m DEM. From this data, we will calculate additional terrain metrics such as the standard deviation of elevation and/or a topographic complexity index aggregated to the 3.125 km EASE-grid pixel scale used in the study. These metrics will provide an indication of sub-grid topographic heterogeneity, which is particularly relevant in mountainous areas where elevation changes rapidly within a single coarse-resolution grid cell.

By enhancing the topographic inputs to the model, we aim to better capture the terrain-driven variability in snow accumulation and TB behavior. This refinement is also expected to mitigate known TB retrieval uncertainties in high-relief regions, such as those caused by variable local incidence angles (Xiong et al. 2022). These additions should improve the reliability and spatial representativeness of the RF model across complex landscapes.

3. Minor Comments

Line 50:
It is not accurate to state that passive microwave sensors (PMWs) are insensitive to cloud cover. Please revise this sentence in accordance with the major comment above. PMWs can penetrate cloud cover; however, they are still affected by clouds and the total atmospheric water column. Atmospheric corrections should be applied.

We will modify this sentence to: 'The moderate frequency (~≤37 GHz) retrievals from operational satellite microwave radiometers are sensitive to snow cover conditions, providing nearly continuous, year-round data. Importantly, the propagation of microwave energy through the snowpack is responsive to changes in snow structure, including liquid water content (LWC), grain size and density, which are key indicators of snowmelt onset (Tedesco et al., 2015).' Removing the incorrect description of PMWs being insensitive to cloud cover.

We will also address the atmospheric impacts in a newly added section 3.2.4 as follows:.

3.2.4 MERRA-2 Reanalysis

Precipitable water vapor and precipitating clouds can affect PMW observations (Du et al., 2014) and assert adverse effects on snow retrievals (Dolant et al. 2016). To mitigate atmospheric influence on SSM/I and SSMIS evening observations, the water vapor estimates for 6 PM Alaska Time were extracted from the MERRA-2 Reanalysis dataset and integrated in the RF framework. The MERRA-2 data were collected from the NASA EarthData Explorer and resampled to 3.125 km using a nearest neighbor interpolation.

Added reference:

Du, J., Kimball, J.S. and Jones, L.A., 2014. Satellite microwave retrieval of total precipitable water vapor and surface air temperature over land from AMSR2. IEEE Transactions on Geoscience and Remote Sensing, 53(5), pp.2520-2531.

Lines 66–68:
This sentence is somewhat confusing. The paragraph discusses active microwave systems, but this sentence feels like it belongs in the previous paragraph.

We agree. We will remove this sentence and place it at the end of the previous paragraph.

Line 71:
Please justify the choice of only using SSM/I and SSMIS in the study. Why were AMSR, AMSR2, and SMMR not included to extend the time series and add more data?

Thank you for this important comment. We chose to use SSM/I and SSMIS because they are both part of the Defense Meteorological Satellite Program (DMSP) with similar sensor configurations, and the NSIDC MEaSUREs dataset provides a consistent, well-calibrated dataset with stable acquisition times and long temporal coverage. This consistency is particularly valuable for long-term trend analysis and ensures compatibility in terms of orbital characteristics, overpass times, and calibration standards.

We excluded SMMR due to significant data gaps and dropouts that we observed during preliminary processing. Incorporating SMMR would have required substantial temporal interpolation and gap-filling, which could introduce additional uncertainty and reduce the robustness of the training dataset. As for

AMSR and AMSR2, although these sensors provide valuable data, their records do not fully overlap with the long-term SSM/I–SSMIS time series, and their inclusion would require additional cross-sensor calibration steps to ensure consistency across the record (e.g., the need to account for different spatial resolutions and observation geometry). To maintain continuity and data integrity over the full 35-year analysis period, we focused on the SSM/I–SSMIS series.

We will further clarify this decision under section 3.2.1.

Lines 79–80:
This line appears to be misplaced. It discusses active microwave (backscatter), whereas the rest of the paragraph focuses on PMW (brightness temperature/emissivity). For passive microwaves, snow layer emissivity increases, especially at higher frequencies (see Dolant et al., 2016). At lower frequencies, TB may decrease due to loss of soil emission, but with a sufficiently deep snowpack at K-Ka bands, TB increases. In active microwave observations, volume scattering is lost with increasing snow depth, leaving only surface scattering. Much of the signal is scattered away from the radar sensor, resulting in lower backscatter (see Tedesco et al., 2015).

We appreciate the reviewer's observation and agree that the referenced sentence introduces active microwave terminology and principles, which is inconsistent with our study. In the revised manuscript, we will clarify this by rephrasing the sentence as follows:

'In contrast, as the liquid water content (LWC) within the snowpack increases, snow emissivity at lower frequencies likely decreases due to attenuated soil emission, while increasing at higher frequencies such as K and Ka bands due to enhanced emission from the wet snow layers (Dolant et al., 2016).'

Lines 89–91:
Consider including the review by Meloche et al. (2022) in this section. They also used a Random Forest model to retrieve snow depth, and their findings are relevant to this study.

We will add Meloche et al. 2022

Lines 311–312:
Please elaborate on how the 80/20 data split was performed. Was the split randomized? If not, correlated data may have affected model performance.

The 80/20 split was random and performed with replacement. The revised sentence will read as follows:

'In each iteration, a random 80/20 (training/testing) split with replacement ensured that the testing data represented a unique subset of high-quality observations from different years and locations, allowing the model's generalizability to be evaluated across a variety of conditions.'

Lines 323–325:
Discuss potential errors arising from differences in spatial scale among datasets. The RF model outputs are at kilometer scale, while GHCNd station data is point-based and may not capture broader spatial variability.

This spatial discrepancy is discussed at line 491-494. Here we discuss the limitations in point-based observations not fully capturing the broader spatial variability.

Lines 383–384:
Consider adding a short paragraph confirming variability in model outputs. Since the models are applied to different time frames, snow cover may be more stable during the IMS period than the SnowModel period. Are the histograms of both models similar, or does the SnowModel histogram show more spread? This may influence interpretation and RF model precision.

This is a great point. We will add a few sentences describing the variability in model outputs and include histograms in the Appendix. Reviewer 2 had similar comments.

Line 389:
Why are 7–10 days not categorized? Is this due to a lack of data in that range?

We apologize, rather than greater than 10 days, it should day 'greater than 7 days.' Thank you making this observation.

Lines 398–400:
Why do the temporal categories (number of days) differ between the two predictions?

The categories are based off a natural breaks classification which differed between snowmelt onset and snowoff. We would like to keep the two predictions in their respective categories and will add the following text to add clarity:

'Given that snow cover at higher elevations can persist for extended periods of time and through the summer months, it is not a surprise that these pixels are ranked as 'low.' The temporal error categories for snowoff differ slightly from those used for snowmelt onset due to differences in the timing dynamics of the two events—snowoff tends to occur more gradually and can span longer periods, especially in complex terrain, necessitating broader classification thresholds to reflect the distribution of prediction errors.'

Lines 403–405:
See earlier comment regarding TB measurement uncertainties in high-topography areas.

We will add to the discussion on the potential impacts in TB measurements in topographically complex areas.

Section 5.4:
Please clarify why the time series was split at 2006. The rationale is currently unclear.

Thank you for this comment. We split the time series at 2006 to divide the 36-year period (1988–2023) into two equal halves in order to examine whether any meaningful trends were present within sub-periods of the record. The rationale for this segmentation was that, although the full-period trends were generally nominal or non-significant, we hypothesized that shifts in snow phenology may have occurred within shorter timeframes that would be obscured when analyzing the entire dataset at once. By splitting the record evenly, we aimed to identify any potential intra-period changes or reversals in trend direction, which indeed became evident in the contrasting patterns observed before and after 2006.

We will add the following paragraph under section 4.2.1:

'Because trends over the full period of record were generally weak or non-significant, we further examined the potential for meaningful sub-period patterns by dividing the 36-year time series (1988–

2023) into two equal halves: 1988–2005 and 2006–2023. This split allowed us to evaluate whether changes in snow phenology were occurring within shorter timeframes that may have been masked by variability across the full record. Indeed, this approach revealed distinct trends within each period that were not apparent in the overall time series.'

Line 505–506:
This observation may also result from higher TB measurement errors in high-altitude regions.

Here at 505-506 we will discuss implications of higher TB measurements at higher altitudes.

Equations:
Equations throughout the manuscript are not numbered. Please revise to include proper numbering.

We will number the equations.

References:

Dolant, C., Langlois, A., Montpetit, B., Brucker, L., Roy, A., & Royer, A. (2016). Development of a rain-on-snow detection algorithm using passive microwave radiometry. *Hydrological Processes, 30*(18), 3184–3196. https://doi.org/10.1002/hyp.10828

Forman, B. A., & Xue, Y. (2017). Machine learning predictions of passive microwave brightness temperature over snow-covered land using the Special Sensor Microwave Imager (SSM/I). *Physical Geography, 38*(2), 176–196. https://doi.org/10.1080/02723646.2016.1236606

Liebe, H. (1989). MPM—An atmospheric millimeter-wave propagation model. *International Journal of Infrared and Millimeter Waves, 10*(6), 631–650. https://doi.org/10.1007/BF01009565

Meloche, J., Langlois, A., Rutter, N., Royer, A., King, J., Walker, B., Marsh, P., & Wilcox, E. (2022). Characterizing tundra snow sub-pixel variability to improve brightness temperature estimation in satellite SWE retrievals. *The Cryosphere, 16*(1), 87–101. https://doi.org/10.5194/tc-16-87-2022

Tedesco, M., Sartori, M., & Jeyaratnam, J. (2015). An overview of the current NASA operational AMSR-E/AMSR2 snow science team activities. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 4037–4040). https://doi.org/10.1109/IGARSS.2015.7326711

Xiao, X., Liang, S., He, T., Wu, D., Pei, C., & Gong, J. (2021). Estimating fractional snow cover from passive microwave brightness temperature data using MODIS snow cover product over North America. *The Cryosphere, 15*(2), 835–861. https://doi.org/10.5194/tc-15-835-2021

Xiong, C., Yang, J., Pan, J., Lei, Y., & Shi, J. (2022). Mountain snow depth retrieval from optical and passive microwave remote sensing using machine learning. *IEEE Geoscience and Remote Sensing Letters, 19*, 1–5. https://doi.org/10.1109/LGRS.2022.3226204