Response to Reviewer

Dear reviewer,

We sincerely appreciate your thoughtful comments and detailed suggestions. Your feedback has significantly improved the clarity, rigor, and overall quality of our work. We believe we have adequately addressed all the major and minor comments, and the manuscript has been greatly improved. Our point-by-point responses to your comments are listed below in black.

The previous reviewers have provided excellent and thorough feedback on the manuscript, with which I largely concur. Upon reviewing your work, I believe your primary contribution to the snow science community lies in proposing a novel snow phenology determining method. The revised manuscript is substantial and well-developed around your proposed algorithm. However, several major issues should be addressed to further enhance the manuscript's clarity, rigor, quality, and impact.

Response:

Thank you for recognizing our work and for your suggestions. We have taken your comments, and the manuscript has been further enhanced.

Main comments:

1. Regarding research gap (Also noted by Reviewer 2#): Generally, the research gap of one publication should be demonstrated by prior researchers or by author himself. The research gap of your manuscript is currently underdeveloped. Your statement—"The fixed threshold for snow phenology fails to account for the variations in snow cover across the NH" and "Snow conditions are variable, but thresholds are always fixed, which can lead to inaccuracies or overestimation of in snow phenology"—reads more like an assumption than a well-supported argument. There is no cited literature, empirical evidence, or demonstration to validate this claim. To strengthen your manuscript, you must either: 1) cite prior studies that highlight limitations of fixed thresholds, and/or 2) Provide original analysis (e.g., comparative assessments) to substantiate this assertion. A compelling case for your work's necessity should be made early in the manuscript to convince readers of its significance.

Response:

Thank you for your suggestion. We have now revised the manuscript to include additional references to highlight the limitation of fixed threshold.

For the threshold of snow depth data, Gascoin et al. (2015) suggested using 10 cm, while other researchers reported best performances for lower values, e.g. 2–5cm (Thirel et al., 2012). Notarnicola (2020) argued that global snow cover analyses should consider snow characteristics such as accumulation, duration, and melt. The selection of thresholds is optimized according to different snow cover characteristics (in terms of regularity and maximum snow depth). The study showed that in areas where maximum snow depth is below 100 cm, the most suitable threshold ranges from 0 to 10 cm (average 9 cm), while in areas with deeper snow, the optimal range is 5 to 15 cm (average 12.4 cm). In addition, Gascoin et al. (2015) identified a clear seasonal trend in the optimal threshold. These findings suggest that the best threshold varies across space and time, and using a single fixed value may lead to uncertainty in the results.

The following are some auxiliary proofs. Previous studies on vegetation phenology have clearly identified the limitations of fixed-threshold methods (White et al., 1997; Mo et al., 2012). Moreover, the daily snow cover product of the Moderate Resolution Imaging Spectroradiometer (MODIS) relies on the Normalized Snow Cover Index (NDSI) threshold to determine whether there is snow at the grid points. Riggs et al. (2017) suggested that applying a single threshold without accounting for local snow properties, atmospheric conditions, and land cover types often reduces snow detection accuracy. These findings underscore the need for flexible, context-specific threshold approaches in snow cover studies.

Based on the above content, we have revised the manuscript.

Line 58:

Numerous studies have demonstrated that improving the accuracy of snow cover products is the primary means of enhancing snow phenology metrics (Frei et al., 2012; Estilow et al., 2015; Xiao et al., 2024). However, whether the extraction methods of snow phenology are reasonable has received little attention. Notarnicola (2020) suggested that global snow cover analyses should consider snow characteristics such as accumulation, duration, and melt. The selection of thresholds is optimized according to different snow cover characteristics (in terms of regularity and maximum snow depth). However, most studies still employ a fixed threshold to extract snow phenology in different regions and years (Brown et al., 2007; Gao et al., 2011; Yue et al., 2022; Tang et al., 2022). In fact, applying a single threshold without accounting for local snow *properties, atmospheric conditions, and land cover types often reduces snow detection accuracy (Riggs et al., 2017; Gao et al., 2019). Snow cover increases with latitude, with thick and stable snow cover at high latitudes and shallow and short-lived snow cover at middle and low latitudes, especially in the TP (Orsolini et al., 2019). In addition, the snow changes from year to year due to many aspects of the climate and the regional snow cover trends exhibit a heterogeneous and non-linear response to its regional warming rate (Blau et al., 2024). Snow conditions are variable, but thresholds are always fixed, which can lead to uncertainties in snow phenology. At present, it has been proven in the methods of extracting vegetation phenology that fixed thresholds cannot accommodate spatio-temporal heterogeneity, ignore inter-annual variations, and are not applicable to diverse vegetation types, among a series of other problems (White et al., 1997; Mo et al., 2012). We aim to propose a novel method that incorporates both spatial heterogeneity in snow cover and temporal variability to extract snow phenology, reducing the uncertainty associated with the fixed threshold method from a physical meaning perspective.*

2. Regarding the claim: “fixed thresholds always lead to inaccuracies in snow phenology” (also noted by Dr. Xiao). I partially disagree with this statement (or perhaps I have not fully grasped your reasoning). Conventionally, inaccuracies in snow phenology metrics are attributed to errors in snow cover products (e.g., snow cover extent, fraction, depth, or SWE). For example, Dr. Xiao’s work(https://doi.org/10.1016/j.isprsjprs.2024.07.018) and other studies focus on improving snow cover product accuracy as a primary means to enhance snow phenology metrics—a point also reflected in your Section 3.1. To support your argument, you must: 1) provide evidence fixed thresholds (independent of snow product errors) contribute significantly to phenology inaccuracies. 2) Acknowledge that snow cover product quality remains a critical source of uncertainty. A broader discussion on all potential uncertainty sources (thresholds, data quality, etc.) would provide a more balanced perspective.

Response:

Following the suggestions of Dr. Xiao and your feedback, we have revised the relevant statements. We now place greater emphasis on the limitations of using a fixed threshold. In particular, we highlight that a fixed threshold cannot capture the actual evolution of snow cover across different regions and years.

(1) Regarding the first question, the following are several examples we provided in the manuscript, in which we also expressed our concern about the use of fixed thresholds. As shown in Fig. 4, fixed thresholds are not suitable, especially in mid- and low-latitude regions. For example, in the Tibetan Plateau and the 30°N-40°N zone, SCF rarely reaches 50% throughout the year. For snow depth, the 2 cm mark often corresponds to the middle or even the peak of the snow curve, long after the snow season has begun. We can also find large differences in snow peak values between high and low latitudes, which makes the period when the threshold is reached different. These observations highlight a key limitation: fixed thresholds are limited for extracting snow phenology, regardless of the data type used.

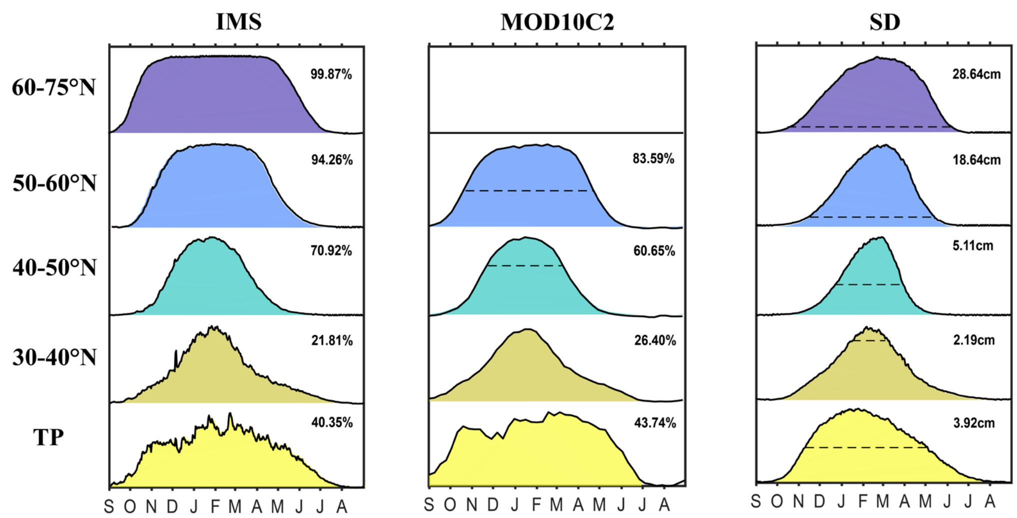

Figure 4. Intra-annual variations of IMS, MOD10C2, and SD in five latitudinal zones (including the Tibetan Plateau) from 2000 to 2018. The dashed lines in the MOD10C2 curve represent the MOD10C2 of 50%, and the dashed lines in the SD curves represent the SD of 2 cm. The MOD10C2 dataset north of 60°N is not analyzed due to

the effects of the polar night. The values in the graphs characterize the annual snow maximum over the respective areas.

Because the Tibetan Plateau is at low to mid-latitudes while also having areas of stable thick snow and unstable thin snow, we discuss the applicability of dynamic and fixed thresholds to the Tibetan Plateau separately in the manuscript (Fig. 10). Employing the dynamic snow threshold method in mountain areas led to a reduction from the original 2 cm threshold to 0.712 cm, which is more consistent with the position of the turn change in the curve slope. Conversely, snow in non-mountain areas remains shallow and unstable, with an annual average maximum SD below 2.5 cm. The 2 cm threshold nearly reaches the snow peak.

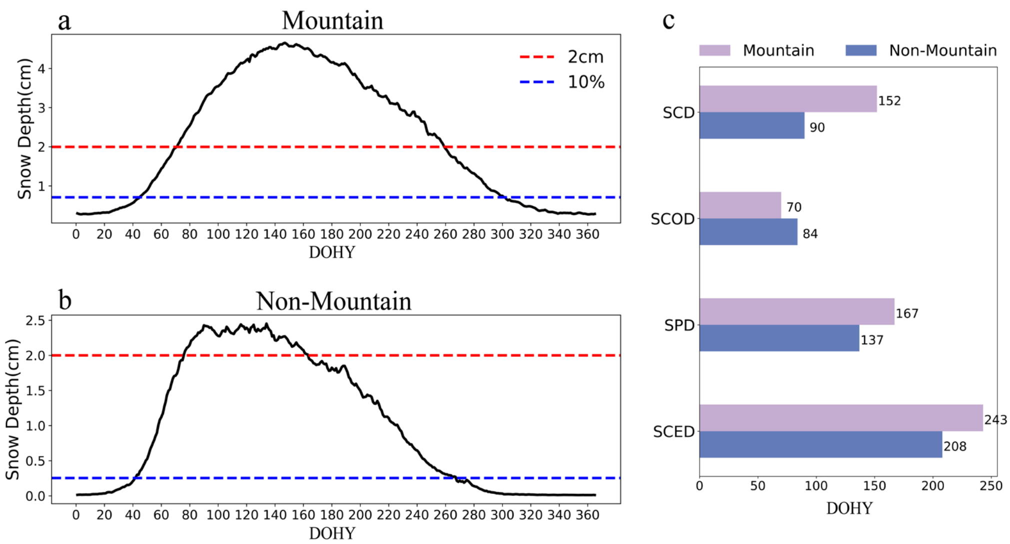

Figure 10. Intra-annual variation of snow depth in (a) mountain, and (b) non-mountain areas on the Tibetan Plateau (TP). The red dashed line represents the location of the fixed threshold of 2 cm, and the blue line represents the location of the threshold extracted using the snow dynamic threshold method. (c) Histogram of snow phenology in non-mountain and mountain regions of the TP using the dynamic threshold method. The unit DOHY is an abbreviation for day of the hydrological year, defined as September 1 through August 31 of the following year.

We validate the usability of the remote sensing data in the TP by comparing the passive microwave remote sensing data with in situ observations, and we select the nearest grid points based on the latitude and longitude of each weather station. Spatial maps of multi-year average snow depths show that passive microwave remote sensing snow depths are greater than observed data. Passive microwave remote sensing of snow depth captures the inter-annual variability of the observed data, but suffers from systematic bias. This shows that it is difficult to reach the 2cm threshold even in the case of large raw data, and it is even more difficult to reach the 2cm threshold if the data precision is higher.

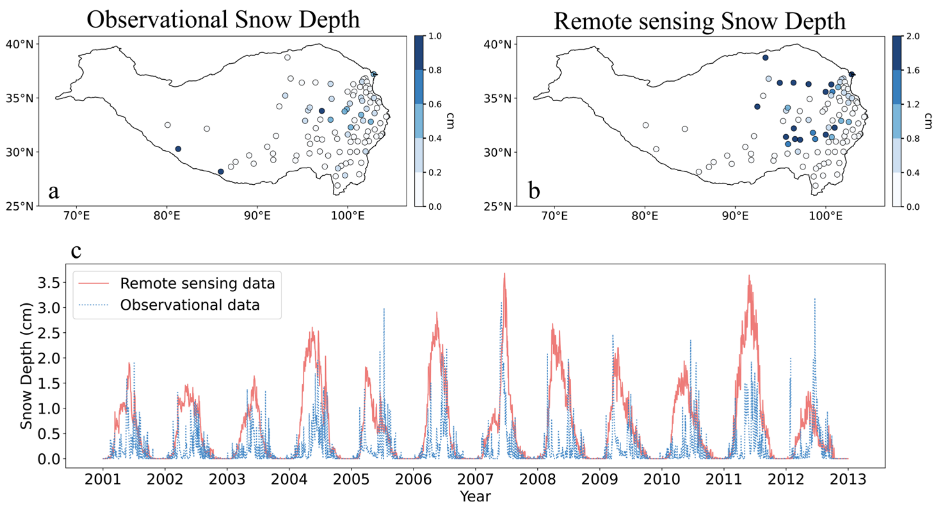

Figure R1. Spatial distribution of maximum snow depth at 30 meteorological stations of (a) shallow snow and (b) deep snow. (c) The probability distribution function for the maximum number of consecutive days of snow depth at shallow and deep snow weather stations. The blue line represents shallow snow, and the red line represents deep snow.

We plot both absolute and normalized snow depth over the mountainous regions of the Tibetan Plateau for the hydrological years 1988–2017. The timing and magnitude of peak snow depth vary greatly between years, yet the shape of the normalized curves remains consistent. This indicates that the time required to reach a 2 cm snow depth differs from year to year. In years with deeper snow, the 2 cm threshold tends to occur later. By contrast, the 10% dynamic threshold aligns more consistently across all 30 years, typically appearing at the inflection point where the snow line shifts from a gradual to a rapid rise.

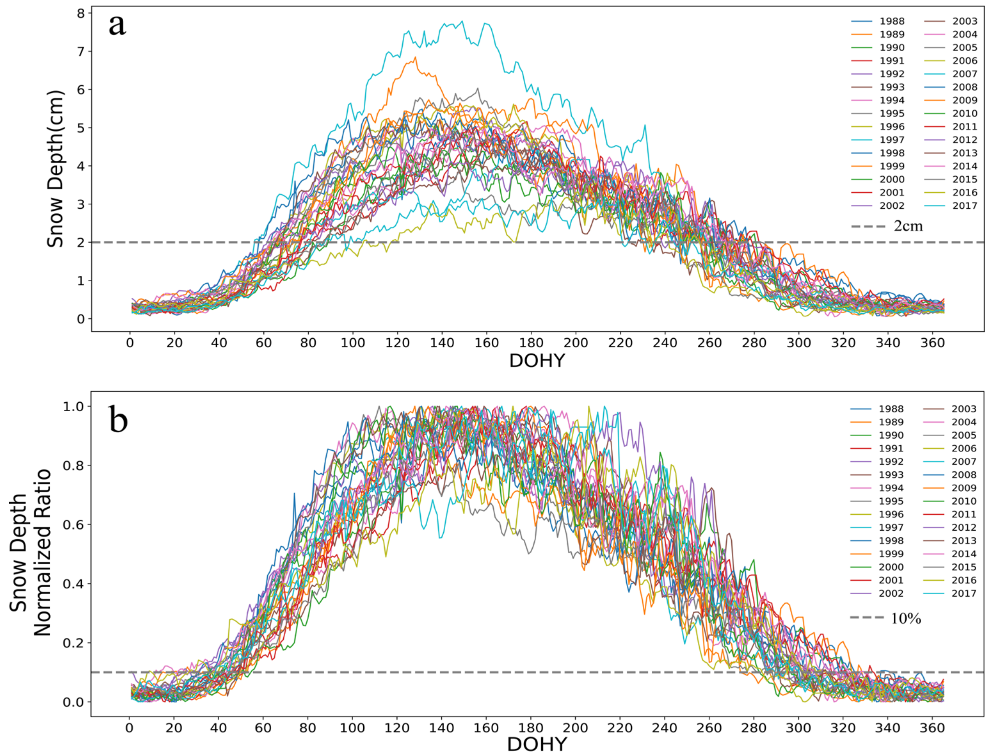

Figure R2. Intra-annual variations in (a) snow depth and (b) normalized snow depth on the Tibetan Plateau from 1988 to 2017. Gray dashed lines indicate 2cm and 10% respectively. The unit DOHY is an abbreviation for day of the hydrological year, defined as September 1 through August 31 of the following year.

These findings show that the uncertainty associated with fixed thresholds cannot be ignored. We have revised the manuscript accordingly.

Line 253:

Based on the above results, we believe that improving the accuracy of snow cover products is one of the primary means for enhancing snow phenology, and that the soundness of snow phenology extraction methods is crucial. In the next section, we aim to enhance the snow phenology extraction method using SD data to obtain more reasonable snow phenology in the NH.

(2) Indeed, we always acknowledge that uncertainties in snow cover products contribute significantly to snow phenology analyses, regardless of the thresholding method used. In Section 3.1 of our manuscript, we show clear differences in snow cover results derived from different products. Therefore, improving the accuracy of snow cover products is the first step to ensure the accuracy of snow phenology results. We also believe this point is well supported and have discussed it in detail in the first part of our discussion (Line 472- Line 487).

In order to provide the reader with a clearer understanding of our point of view, we first add a brief acknowledgement of the impact of snow cover products on the accuracy of snow phenology in the introduction section of the manuscript. Second, a more detailed discussion is provided in the discussion section.

Line 56:

Various types of snow data are used to extract snow phenology, including SDs, snow cover fractions, and snow water equivalents, leading to possible differences in identified snow phenology (Chen et al., 2015; Guo et al., 2022). Numerous studies have demonstrated that improving the accuracy of snow cover products is the primary means of enhancing snow phenology metrics (Frei et al., 2012; Estilow et al., 2015; Xiao et al., 2024). However, whether the extraction methods of snow phenology are reasonable has received little attention. For example, most studies have employed a fixed threshold to extract snow phenology in different regions and years (Brown et al., 2007; Gao et al., 2011; Yue et al., 2022; Tang et al., 2022).

Line 472:

Despite improvements in snow phenology extraction, variations in the data and definitions of snow phenology and hydrological years lead to differences in extracted snow phenological characteristics, which are further compounded by inherent data uncertainties (Xie et al., 2017; Ma et al., 2020; Guo et al., 2022). The fundamental principles underlying snow information acquisition vary across observation methods, impacting binary snow results (Hall and Riggs, 2007; Dietz et al., 2011; Zhang et al., 2024). Factors like the observational instrument's orbit and cloud cover can further *affect the accuracy of snow datasets (2005; Gao et al., 2010; Coll and Li, 2018). Second, the performance of snow data varies geographically. Since the accuracy of passive microwave detection increases with snow depth, the passive microwave remote sensing data is more effective for analyzing snow phenology in regions with consistent snow cover (Armstrong & Brodzik, 2001; Savoie et al.,2009). In areas with shallow snow with wet snow, the accuracy of passive microwave remote sensing data is reduced and the snow depth indicator may not accurately capture accumulation and melting processes. In addition, for the transient snow area, the snow depth curve is more volatile, which makes the assumed single-peak structure untenable. After comprehensive consideration, the snow cover fraction may be a more reliable indicator in such cases. Therefore, we perform another extraction of dynamic snow phenology using the snow cover fraction data, and the results are similar to SD, but with greater differences in TP (see Fig. S1- S5 in Supplement). Regardless of the threshold method, problems with the data itself increase the uncertainty of snow phenology analysis. Therefore, it is necessary to integrate ground observation data with different remote sensing data to form a more comprehensive and accurate snow phenology extraction system.*

3. Some suggestions on logical flow: The manuscript appears to follow a straightforward logic: propose a new algorithm and validate its performance through extensive analysis. However, as you noted in your responses, ground-truth observations for snow phenology metrics are lacking, making absolute accuracy assessments impossible. Thus, claims of your method’s "accuracy" should be tempered. Instead, focus on demonstrating its reasonableness (e.g., via consistency checks, comparative advantages, or physical plausibility). I recommend revising statements throughout the manuscript to reflect this nuance.

Response:

We appreciate the reviewer’s thoughtful comments. After receiving suggestions from previous reviewers and community members, we have also recognized that the absence of observational data on snow phenology prevents us from validating which method is absolutely more accurate. In addition, our aim is to improve on traditional snow phenology methods, not to discredit them. In response, we revise our terminology. While we propose a dynamic threshold method grounded in physical processes, we now refer to it as more reasonable, rather than more accurate. Compared with the traditional fixed threshold approach, the dynamic method offers clearer physical relevance and greater flexibility. Accordingly, we have replaced all instances of “accurate” with “reasonable” when describing the dynamic identification of snow phenology in the manuscript. We have also replaced “errors” and “bias” in the fixed threshold method with “limitations”.

Line 7:

Previous studies commonly employed fixed threshold methods to extract snow phenology, which cannot represent the differences in the beginning/end of the snow period under different snow conditions in the Northern Hemisphere, leading to potential uncertainties in snow phenology.

Line 70:

We aim to propose a novel method that incorporates both spatial heterogeneity in snow cover and temporal variability to extract snow phenology, reducing the uncertainty associated with the fixed threshold method from a physical meaning perspective.

Line 254:

In the next section, we aim to enhance the snow phenology extraction method using SD data to obtain more reasonable snow phenology in the NH.

Line 314:

This suggests that the method can dynamically adjust the threshold based on the annual and regional SD’s variations, thereby reducing the uncertainty in snow phenology extraction caused by large-scale climatic influences.

Line 377:

In addition, the snow phenology dynamic threshold method is more reasonable in areas with complex topographic and climatic features, such as the TP.

Minor comments:

1. Table1: “The four days” ---> “The fourth day”

The definitions in Table 1 seem tailored to your called “fixed threshold”. If so, clarify this in the table title. If not, explain how your proposed algorithm accommodates these definitions across different snow cover products in your updated manuscript.

Response:

Thank you for your correction, and the sentence and title has been modified.

Table 1:

the fourth day before which the pixel is first covered with snow.

Line 117:

Table 1. *Definitions of fixed threshold methods for snow phenology parameters with different datasets.*

2. Equation 2: The range of snow cover fraction is from 0 to 1 over the whole seasonal snow season. The equation includes “snow cover fraction”, but the variable “Snow_ratio” appears equivalent to “Snow (snow cover fraction)”. This raises questions about the equations applicability, especially given your note that it cannot be used with IMS data. Clarify how your algorithm derives phenology metrics in such cases.

Response:

Although both $Snow_{ratio}$ and snow cover fraction (SCF) range from 0 to 1, they represent different physical meanings. The SCF is spatial dimension and represents the distribution state of snow at a certain moment. In contrast, $Snow_{ratio}$ captures a key transition point over time. When the snow depth is input as snow element, if the maximum snow depth is 20cm (the minimum snow depth is 0), the 10% threshold corresponding to it is 2cm. When SCF is used as an element, if the maximum SCF is 50% (the minimum SCF is 0), its 10% threshold corresponds to 5%. $Snow_{ratio}$ represents the key turning point of the change in snow coverage in a grid/area, corresponding to a certain snow element value. Once this threshold is reached, we consider the grid to be snow-covered.

We do not apply this formula to IMS data because it has already been processed. The IMS product classifies each grid cell as snow-covered or snow-free based on a set threshold. As a result, we do not reassess snow presence or compute grid-level percentage changes at the grid level.

3. Fig. 7: “percentage of snow depth”--->“Normalized Snow depth”. The current term suggests a direct accumulation of values, which may mislead readers.

Response:

Thank you for your advice. To avoid misunderstanding, we have changed “percentage of snow depth” to “Snow depth Normalized Radio” in the figure.

4. One friendly suggestion: While borrowing concepts from vegetation phenology is innovative, fundamental differences between snow and vegetation dynamics (e.g., snow’s rapid appearance/disappearance vs. gradual vegetation changes) require careful handling. Explicitly address these discrepancies and justify any adapted methodologies.

Response:

Thank you for the suggestion. As you point out, snow and vegetation are similar in some ways, but fundamentally different. We use the analogy between vegetation and snow to highlight a broader point: diverse extraction methods are also needed for snow phenology. Vegetation phenology research began earlier and is now relatively mature. Many studies have proposed various methods to extract vegetation phenology (Zhang et al., 2003; Filippa et al., 2016; Kong et al., 2022). In contrast, snow phenology has typically relied on a single method with a fixed threshold. We believe this limits progress and innovation in the field. The dynamic threshold method builds on ideas from vegetation phenology but does not assume that snow and vegetation behave in the same way. For instance, vegetation studies often use a 15% threshold. In our work, we identify a 10% dynamic threshold for snow depth by considering both first-order derivatives and the underlying physical processes. Our research results show that the 10% threshold is applicable in most areas of the Northern Hemisphere. Moreover, because snow changes rapidly, we require more than just reaching a threshold to define snow phenology. We also apply a five-day continuity condition to ensure stability. This second step is crucial—it clearly distinguishes snow phenology from vegetation phenology. We will further clarify the key differences between snow and vegetation, while emphasizing the need for more flexible and physical methods in snow phenology research. Based on the above, we add the following sentences in the discussion section.

Line 488:

We propose a dynamic approach to defining snow phenology by adjusting the threshold for snow presence in this study. However, snow and vegetation differ in their fundamental dynamics. Vegetation grows gradually, while snow can change rapidly over short periods. Therefore, there is a second key step in the extraction of snow phenology, which requires that the threshold must be met for several consecutive days. This condition ensures that the detected event reflects a stable and meaningful snow presence and mitigates the influence of sporadic snowfall events.

References:

Filippa, G., Cremonese, E., Migliavacca, M., Galvagno, M., Forkel, M., Wingate, L., Tomelleri, E., Morra di Cella, U., & Richardson, A. D.: Phenopix: A R package for image-based vegetation phenology. Agricultural and Forest Meteorology, 220, 141–150. https://doi.org/10.1016/j.agrformet.2016.01.006, 2016.

Gascoin, S., Hagolle, O., Huc, M., Jarlan, L., Dejoux, J.-F., Szczypta, C., Marti, R., & Sánchez, R.: A snow cover climatology for the Pyrenees from MODIS snow products. Hydrology and Earth System Sciences, 19(5), 2337–2351. https://doi.org/10.5194/hess-19-2337-2015, 2015.

Kong, D., McVicar, T. R., Xiao, M., Zhang, Y., Peña-Arancibia, J. L., Filippa, G., Xie, Y., & Gu, X.: phenofit: An R package for extracting vegetation phenology from time series remote sensing. Methods in Ecology and Evolution, 13(7), 1508–1527. https://doi.org/10.1111/2041-210X.13870, 2022.

Mo, J., Zhu, W., Wang, L., Xu, Y., and Liu J.: Evaluation of remote sensing extraction methods for vegetation phenology based on flux tower net ecosystem carbon exchange data. Chinese Journal of Applied Ecology, 23(2), 319-327. https://doi.org/10.13287/j.1001-9332.2012.0072, 2012.

Notarnicola, C. : Hotspots of snow cover changes in global mountain regions over 2000–2018. Remote Sensing of Environment, 243, 111781. https://doi.org/10.1016/j.rse.2020.111781, 2020.

Riggs, G.A., Hall, D.K., Salomonson, V.V.: MODIS snow products user guide to collection 5. Digital Media 80, 1–80, 2006.

Riggs, G. A., Hall, D. K., & Román, M. O.: Overview of NASA's MODIS and Visible Infrared Imaging Radiometer Suite (VIIRS) snow-cover Earth System Data Records. Earth System Science Data, 9, 765-777. https://doi.org/10.5194/essd-9-765-2017, 2017.

Thirel, G., Notarnicola, C., Kalas, M., Zebisch, M., Schellenberger, T., Tetzlaff, A., Duguay, M., Mölg, N., Burek, P., & De Roo, A. : Assessing the quality of a real-time Snow Cover Area product for hydrological applications. Remote Sensing of Environment, 127, 271–287. https://doi.org/10.1016/j.rse.2012.09.006, 2012.

White, M. A., Thornton, P. E., and Running, S. W.: A continental phenology model for monitoring vegetation responses to interannual climatic variability. Global Biogeochemical Cycles, 11, 217–234. https://doi.org/10.1029/97GB00330, 1997.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F. F., Gao, F., Reed, B. C., & Huete, A.: Monitoring vegetation phenology using MODIS. Remote Sensing of Environment, 84(3), 471–475. https://doi.org/10.1016/S0034-4257(02)00135-9, 2003.