To Editor:

We greatly appreciate the reviewers taking the time to provide constructive feedback and valuable suggestions. The implementation of these suggestions has greatly improved the quality of the manuscript, enabling us to make significant improvements. Every revision suggestion and opinion proposed by the reviewer has been carefully considered. Of course, there are also some parts that we do not agree with, and we will annotate them in segments.

We have also prepared and supplemented the documents one by one according to your six requirements. Original comments and suggestions quoted from this decision letter highlighted in black italics, and our response is in blue. The revised content is marked in red text in the manuscript.

Response to the Reviewer comments 1

1 The last section of the introduction is excessively long and should be made brief.

Reply 1: We sincerely appreciate the reviewer's constructive comment on the length of the last section of the introduction. In strict accordance with your suggestion, we have thoroughly revised and streamlined this section: we removed all redundant descriptive content, including the tedious step-by-step elaboration of methodological details that do not fall within the scope of the introduction, while retaining only the core research value of our work and the necessary description of the paper's structural arrangement. The revised section is significantly shortened, with more concise and focused expression, which fully conforms to the standard writing norms of the introduction section. All relevant modifications have been clearly marked in the revised manuscript. **(Lines: 98-106)**

To address the key challenges of SD inversion in sparsely observed regions, especially those with complex mountainous terrain, this study proposes a more accurate and robust SD retrieval method. The core innovation lies in the integration of multisource data, including downscaled SD data and ground-based measurements, into the PyCaret AutoML framework. This approach not only improves SD retrieval accuracy but also overcomes the limitations of existing methods that struggle with data scarcity and terrain complexity. Moreover, the automatic model selection feature of PyCaret helps mitigate human bias in model choice, ensuring the robustness and generalizability of the results. The findings provide reference for improving SD estimation techniques, particularly in cold regions with complex topography. The remainder of the paper is organized as follows: Sections 2 and 3 introduces the study region, data and methods; Section 4 describes the results; Sections 5 and 6 present the discussion and conclusions.

2 Please Harmonize reported performance metrics in Abstract and Conclusion. For example, R and R2 are both used.

Reply 2: We sincerely thank the reviewer for this pertinent and rigorous comment, which accurately points out the non-standardized expression of performance metrics between the Abstract and Conclusion sections, and is critical for improving the academic rigor and readability of our manuscript.

Strictly in full accordance with your suggestion, we have thoroughly standardized the expression of correlation-related performance metrics throughout the manuscript. Specifically, we have unified all relevant correlation coefficients (the mixed use of R and R^2) in the Abstract, Results, and Conclusion sections to the coefficient of determination (R^2). This revision completely resolves the inconsistency of performance metric reporting across different sections, and fully meets your core

requirement of harmonizing the metrics in the Abstract and Conclusion.
Meanwhile, we have further optimized the presentation of accuracy metrics for different underlying surface types in the Abstract, and supplemented the snow depth (SD) estimation accuracy results of grassland and water areas. This revision enables a complete and systematic presentation of the SD estimation performance for all four underlying surface types, and further enhances the integrity and consistency of performance metric reporting throughout the manuscript.
All relevant revisions have been clearly marked with red text in the revised manuscript, and these modifications have significantly improved the rigor, standardization and readability of the metric reporting in our paper.

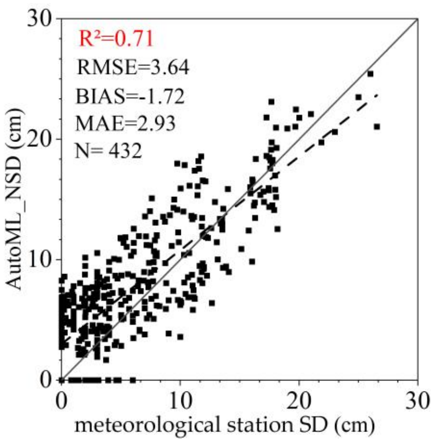
(Lines: 10-26)

Snow depth (SD) is a crucial parameter for describing the spatiotemporal variations of snow cover, and passive microwave (PMW) SD products (10-25 km) are widely used for monitoring SD changes. However, as one of the three major snow-covered regions in China, the Qinghai-Tibet Plateau (QTP) has complex terrain and rapid changes in snow cover with strong spatial heterogeneity, making it difficult for coarse-resolution SD products to accurately describe its spatiotemporal characteristics. This study proposes a high spatial resolution (500 m) SD estimation method based on AMSR-2 brightness temperature (BT) data and Automated Machine Learning (AutoML). Firstly, using Pearson correlation coefficients, 19 key factors influencing SD, including AMSR-2 BT, slope, and surface roughness, were selected as input data (independent variables) for AutoML. Meanwhile, PMW downscaled SD data and ground-based SD measurements were introduced as dependent variables for AutoML. The AutoML model was then trained separately for four different types of snow cover surfaces (forest, grassland, water, and bare land). Finally, through ten-fold cross-validation, the optimal machine learning model for SD estimation under each type of underlying surface coverage was selected, thus generating sequential SD datasets for the ten-year snow cover period of the QTP from 2012 to 2021. Results show that (1) the estimated SD values are consistent with ground-based observations (R^2=0.71), with high accuracy, reflected by an RMSE of 3.64 cm. (2) SD estimation demonstrates the highest accuracy in unused land (CatBoost, R^2=0.82), followed by grassland (CatBoost, R^2=0.77, RMSE=3.11 cm), water (ET, R^2=0.75, RMSE=2.20 cm), and forest (XGBoost, R^2=0.71, RMSE=3.30 cm). (3) A comparison with snow cover extent (SCE) derived from Landsat-8 optical imagery reveals that the estimated SD spatial distribution is consistent with the SCE, providing reliable data for monitoring snow cover changes in mountainous regions.

(Lines: 322-323)

As shown in Figure 7, the Auto_NSD **results exhibit** a **strong agreement** with meteorological station **measurements**, **yielding** an **R^2** value of 0.71.

(Figure 7)

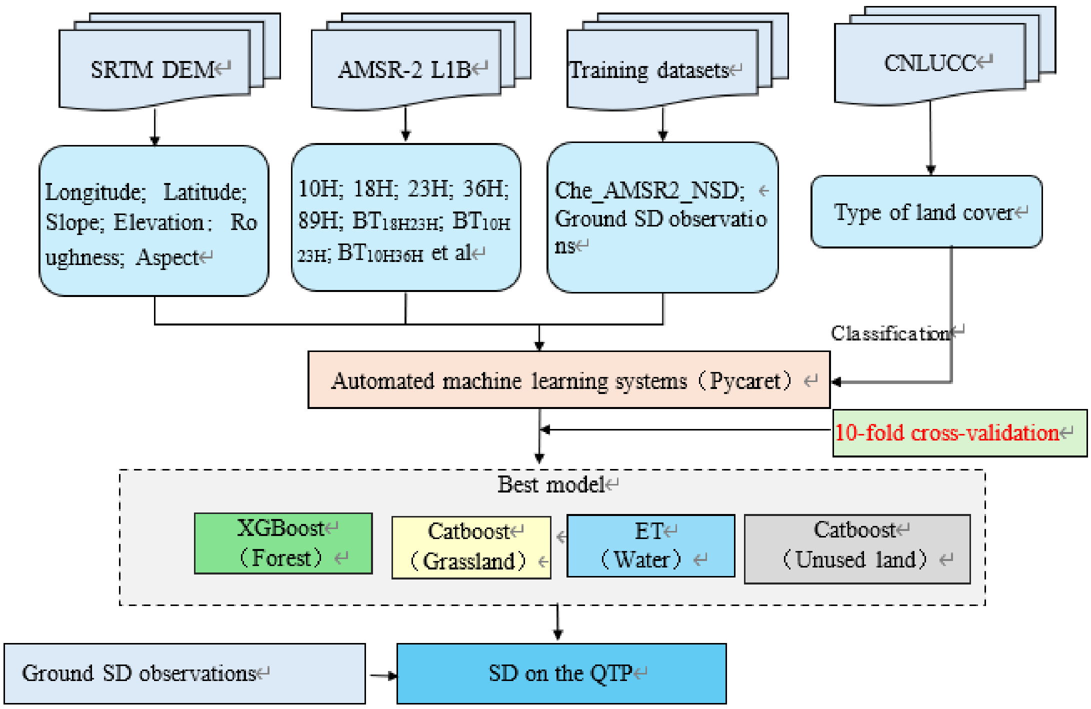


(Lines: 536-538)

The findings suggested that (1) the SD estimates derived from ML techniques exhibited superior accuracy in **characterizing** the snow distribution in the QTP region, closely **matching** ground observations, with R^2=0.71, RMSE=3.64 cm, BIAS=-1.72 cm, and MAE=2.93 cm;

3 "10% off cross-validation" in Figure 2 is "10-fold cross-validation"? Please standardize the terminology.

Reply 3: We appreciate the reviewer's correction. We confirm the typo in Figure 2 and have revised "10% off cross-validation" to the standard term "10-fold cross-validation". Relevant terminologies in the paper have also been standardized, with all changes marked clearly.



4 The feature selection is based on correlation coefficients, why not the importance score?

Reply 4: We sincerely appreciate the reviewer for raising this insightful and valuable question regarding feature selection methods. To address collinearity among variables, we adopted the traditional correlation coefficient method in this study, instead of using feature importance scoring or variance inflation factor approaches. The core consideration in our study design is that the

AutoML framework inherently performs a secondary round of feature selection based on the input data, which automatically identifies the most predictive features via its built-in statistical and model-based importance metrics. Thus, to ensure that the AutoML framework could mine and retain as many useful features as possible during its iterative process, we intentionally selected the classic correlation coefficient method for preliminary data preprocessing and feature screening in the early stage.

We fully acknowledge the rationality of the insightful perspective you put forward, and we will adopt more robust methods to eliminate collinearity in our future research, rather than relying entirely on the "black box" mechanism of AutoML for this step. Thank you again for this crucial comment—we have supplemented the detailed rationale for the selection of the correlation coefficient method in the revised manuscript, and we commit to exploring more rigorous collinearity elimination strategies in subsequent work.

(Lines: 524-529)

(3) Third, the feature selection process employed a traditional approach based on correlation coefficients to identify influential factors. Future research will explore more advanced feature selection techniques, such as feature importance and variance inflation factor, to improve model accuracy.

5 What the influence of land cover use on SD estimation? The manuscript uses CNLUCC 2020 for the entire 2012–2021 period, and the discussion should explain this.

Reply 5: We sincerely appreciate the reviewer's valuable comment on the influence of land cover on SD estimation and the use of the CNLUCC 2020 dataset. In response, we have added a dedicated section in the revised Discussion (Uncertainty Analysis) to explicitly clarify two key points, and also elaborate on the potential uncertainties associated with land cover data:

We expound the specific role of land cover in SD estimation: land cover types are incorporated into the SD inversion process through categorical stratification modeling, which effectively reduces the surface heterogeneity of the study area and significantly improves the robustness of the SD estimation model.

We clarify the rationale for using the CNLUCC 2020 dataset for the entire 2012–2021 study period: considering that the dominant land cover types over the Qinghai-Tibet Plateau exhibit a slow change trend in the short term, the adoption of a consistent single-year land cover dataset avoids additional uncertainties caused by interannual classification inconsistencies of multi-period land cover data.

Additionally, we have explicitly identified and discussed the potential uncertainties in SD estimation that may be induced by local small-scale land use changes in the study area. All the above supplementary contents have been clearly marked in the revised manuscript.

(Lines: 475-482)

This section evaluates the principal sources of uncertainty in SD estimation, including land cover representation, reference dataset selection, and snow cover identification, and examines their impacts on model stability and retrieval reliability.

The CNLUCC 2020 dataset is used to represent land cover conditions for 2012 – 2021. Land cover over the QTP changes slowly, particularly for dominant classes such as grassland, forest, and unused land, which occupy most of the region. Consequently, interannual variations during the study period are limited and are unlikely to substantially affect SD estimation at the spatial scale considered. Using a single and consistent land cover dataset also avoids uncertainties arising from inconsistencies among different products. Although local land use changes may introduce some uncertainty, their overall influence is limited because land cover is mainly used for categorical stratification rather than as a continuous predictor.

To Editor:

We greatly appreciate the reviewers taking the time to provide constructive feedback and valuable suggestions. The implementation of these suggestions has greatly improved the quality of the manuscript, enabling us to make significant improvements. Every revision suggestion and opinion proposed by the reviewer has been carefully considered. Of course, there are also some parts that we do not agree with, and we will annotate them in segments.

We have also prepared and supplemented the documents one by one according to your six requirements. Original comments and suggestions quoted from this decision letter highlighted in black italics, and our response is in blue. The revised content is marked in red text in the manuscript.

Response to the comments 2

Please ask author to polish their manuscript by a native english speaker.

Reply: Thank you for this valuable suggestion. We fully recognize the importance of clear and fluent English expression for effective scientific communication. In response to the reviewer's comment, the manuscript has been thoroughly revised and polished with the assistance of a native English speaker with expertise in academic writing. Throughout the revision, particular attention was paid to improving grammar, sentence structure, terminology consistency, and overall readability, while preserving the original scientific meaning. All language-related revisions throughout the manuscript have been highlighted in red for the reviewer's convenience. We believe that the language quality of the manuscript has been substantially improved.

1 Innovation and Contribution: The authors state that the novelty of the study lies in "(1) the adoption of ..., and (2) the first application of the PyCaret AutoML framework..." However, this description falls short of clearly articulating a significant and unique contribution to the field of snow remote sensing. Merely being the first to apply a particular AutoML tool does not, by itself, constitute scientific novelty—especially when numerous recent studies have achieved highly accurate snow depth retrievals, some even outperforming the results presented here. Given the competitive landscape of snow remote sensing research, the manuscript must better justify why this work advances the field beyond incremental technical implementation. I encourage the authors to explicitly restate the core scientific innovation and its unique value to the community.

Reply 1: We sincerely appreciate the reviewer's insightful and constructive comments on the innovation and contribution of our study. We fully agree with your critical point that merely the first application of a specific AutoML tool (PyCaret) does not independently constitute substantial scientific novelty in the field. In response to this valuable suggestion, we have thoroughly revised and refined the presentation of our study's core innovation and scientific contribution in the manuscript.

We now explicitly emphasize that the novelty of this work does not lie in the simple application of the PyCaret AutoML framework, but in our unique multisource data integration strategy that embeds downscaled SD data and ground-based observational data into the PyCaret AutoML framework. This tailored integration is specifically designed to address the long-standing challenge of SD estimation in complex mountainous regions with sparse observational data—a key issue that has not been fully resolved by previous relevant studies. Additionally, the automatic optimal model selection capability of the PyCaret framework is leveraged to eliminate the inherent human biases and result variability associated with traditional manual model selection methods. We believe this integrated approach effectively enhances the accuracy, spatial resolution, and generalizability of SD retrieval methods in complex, data-scarce alpine environments, which is the core scientific contribution of our study to snow remote sensing research.

We hope the above revisions fully address your concerns and clearly demonstrate the unique scientific value of this work in advancing the field of snow remote sensing.

(Lines: 98-106)

To address the key challenges of SD inversion in sparsely observed regions, especially those with complex mountainous terrain, this study proposes a more accurate and robust SD retrieval method. The core innovation lies in the integration of **multisource data**, including **downscaled SD data** and **ground-based measurements**, into the PyCaret AutoML framework. This approach not only improves SD retrieval accuracy but also overcomes the limitations of existing methods that struggle with data scarcity and terrain complexity. Moreover, the **automatic model selection** feature of PyCaret helps mitigate human bias in model choice, ensuring the robustness and generalizability of the results. The findings provide reference for improving SD estimation techniques, particularly in cold regions with complex topography. The remainder of the paper is organized as follows: Sections 2 and 3 introduces the study region, data and methods; Section 4 describes the results; Sections 5 and 6 present the discussion and conclusions.

2 Motivation and Research Gap: The revised text in Lines 77–85 does not convincingly establish the study's motivation. While machine learning methods are indeed diverse, the authors have not clearly defined the specific scientific or methodological gap their snow depth retrieval algorithm addresses. The justification should focus on "what problem remains unsolved" in current approaches and how this work provides a meaningful solution. Please reframe the motivation around a well-defined research gap.

Reply 2: Thank you for this insightful comment. We agree that the motivation and research gap of this study were not sufficiently articulated in the previous version, particularly in clarifying what specific limitations of existing snow depth retrieval approaches remain unresolved.

In the revised manuscript, we have reframed the motivation to emphasize a clear methodological gap in current machine learning–based snow depth estimation studies. Although numerous ML models have been applied to passive microwave snow depth retrieval, most existing approaches rely on manually selected models, heuristic feature selection, and trial-and-error hyperparameter tuning. Such practices introduce strong human-induced bias and often lead to overfitting and limited robustness, especially in complex mountainous regions like the Qinghai–Tibet Plateau, where observational data are sparse and unevenly distributed.

We further clarify that this study does not aim to propose another standalone ML model, but rather to address the lack of a systematic, objective, and reproducible framework for model selection and optimization in snow depth inversion. By introducing an AutoML-based approach, this work provides a practical solution to reduce subjective intervention, enhance model generalizability, and improve robustness under data-scarce mountainous conditions. These revisions are intended to clearly define the unresolved problem and the scientific rationale for adopting AutoML in this context.

(Lines: 75-87)

Although ML-based approaches have improved SD retrieval accuracy, several methodological challenges remain unresolved, particularly in complex mountainous regions. Most existing studies rely on a limited number of manually selected ML models, empirical feature selection strategies, and trial and error hyperparameter tuning. Such practices introduce substantial human-induced bias and often result in overfitting, especially under conditions of sparse and unevenly distributed observations, as is the case over the QTP (Du et al., 2020). Consequently, model performance is frequently optimized for specific regions or datasets, while robustness and generalizability across heterogeneous mountainous environments remain limited (Feurer et al., 2015; Hernandez et al., 2025). These issues indicate that the primary challenge lies not only in model performance, but also in the lack of a systematic and objective framework for model selection and optimization in SD inversion.

Automated Machine Learning (AutoML) has emerged as an effective solution to address these limitations by autonomously executing data preprocessing, feature evaluation, model selection, and hyperparameter optimization without extensive human intervention (Ribeiro et al., 2024; Hernandez et al., 2025). The fundamental design philosophy of AutoML is to reduce human bias and enhance the robustness, stability, and reproducibility of ML models (Benghzial et al., 2023).

3 Reference Dataset (Che_AMSR2_NSD) and Its Uncertainties: My previous concern regarding the use of Che_AMSR2_NSD as a reference ("true") snow depth dataset has not been addressed. If this product serves as the ground truth in model training and validation, the authors must discuss its known uncertainties (e.g., spatial representativeness, retrieval errors, temporal mismatch) and explain how these may affect the stability, reliability, and generalizability of their retrieval model.

Reply 3: Thank you for this important comment regarding the use of Che_AMSR2_NSD as a reference dataset. We agree that Che_AMSR2_NSD should not be regarded as a strict ground truth, and its uncertainties require explicit discussion. In the revised manuscript, we have added a dedicated uncertainty analysis to address this concern.

We first discuss the known limitations of Che_AMSR2_NSD, including retrieval errors over complex terrain, spatial representativeness associated with passive microwave observations, and potential temporal mismatches with in situ measurements. To assess how these uncertainties affect model stability and reliability, we conducted a comparative analysis at four representative sites across the Qinghai–Tibet Plateau, covering regions with varying station densities. Model performance was evaluated under three reference scenarios: using station observations only, using Che_AMSR2_NSD as a supplementary reference combined with station data, and using the original 25 km Che_AMSR2 SD product as the reference.

The results show that the combined reference strategy consistently achieves higher R^2 values and lower RMSE across all sites, indicating improved robustness compared to single-source reference datasets. While uncertainties in Che_AMSR2_NSD may propagate into SD estimation, the AutoML-based framework exhibits stable performance across different reference scenarios, suggesting a degree of resilience to reference data uncertainty. These analyses and discussions have been added to the revised Discussion section.

(Lines: 483-498)

The Che_AMSR2_NSD SD product contains inherent uncertainties, including retrieval errors over complex terrain and potential temporal mismatches with in situ observations. Therefore, it serves as a supplementary reference rather than strict ground truth, particularly in regions with sparse meteorological stations. To evaluate the impact of reference data uncertainty, four representative sites across the QTP with contrasting station densities were selected, and model performance was assessed under three reference strategies: station-only observations, coarse-resolution Che_AMSR2 SD data, and a combined reference integrating station observations with Che_AMSR2_NSD (Table 3). The combined strategy consistently performs best, with R^2 values of 0.84–0.87 and RMSE of 3.8–4.2 cm, compared to lower R^2 (0.64–0.70) and higher RMSE (4.6–5.0 cm) for the other two strategies. These results indicate that incorporating Che_AMSR2_NSD mitigates the limitations of sparse stations and coarse-resolution satellite data. Direct comparisons between Che_AMSR2_NSD and ground-based observations show a moderate correlation with systematic biases (R=0.72, Figure 13), suggesting potential uncertainty propagation into SD estimation. Nevertheless, the AutoML-based framework maintains stable performance across reference scenarios, demonstrating robustness to reference data uncertainty.

Table 3: Sensitivity of SD estimation performance to different reference data sources at representative sites across the QTP.

Sample Point	lon	lat	method	R^2	RMSE	information
1	101.58	33.74	Combined reference	0.8615	4.1862	Close to meteorological stations
1	101.58	33.74	Station-only reference	0.6809	4.6499	
1	101.58	33.74	Che_AMSR2 reference	0.6707	4.7069	
2	85.97	28.18	Combined reference	0.8481	3.8060	Sparse meteorological stations
2	85.97	28.18	Station-only reference	0.6535	4.9448	
2	85.97	28.18	Che_AMSR2 reference	0.6421	5.0047	
3	97.79	36.06	Combined reference	0.8686	3.8528	Dense meteorological stations
3	97.79	36.06	Station-only reference	0.7030	4.8212	
3	97.79	36.06	Che_AMSR2 reference	0.6933	4.8813	
4	84.35	36.42	Combined reference	0.8431	4.0984	Without meteorological stations
4	84.35	36.42	Station-only reference	0.6526	4.9521	
4	84.35	36.42	Che_AMSR2 reference	0.6427	5.0047	

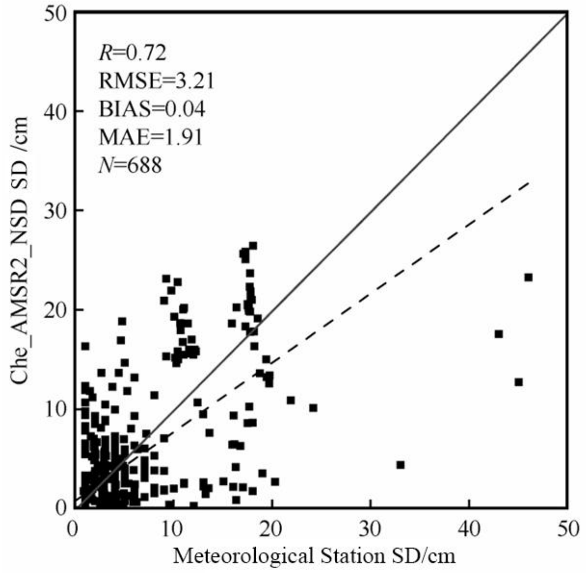


Figure 13: Scattered plot of SD observed by meteorological stations and SD by Che_AMSR2_NSD.

4 Uncertainty in Snow Cover Data: The manuscript mentions a daily cloud-free snow cover dataset that fuses MODIS and passive microwave-derived snow depth (based on Huang et al.'s data). However, it still lacks any quantitative assessment of the uncertainty associated with snow cover identification from Huang's dataset. Please provide a clear evaluation—ideally with error metrics or sensitivity analyses—of how uncertainties in this snow cover input propagate into your final results and affect their interpretation.

Reply: We thank the reviewer for this insightful comment. In the revised manuscript, we have clarified the role of the daily cloud-free snow cover dataset and explicitly addressed its associated uncertainty. This dataset is not used as an input variable in model training or validation; instead, it serves only as a binary mask to distinguish snow-covered and snow-free pixels. Therefore, uncertainties in snow cover identification do not directly propagate through the regression process, but may influence the results only through potential snow/no-snow misclassification.

To evaluate this effect, we added a sensitivity-based uncertainty discussion. We show that such misclassification primarily affects snow extent rather than snow depth retrieval itself, and its overall impact is constrained. Moreover, quantitative SHAP-based analyses indicate that snow cover–related variables play a secondary role in the uncertainty budget. These additions provide a clearer assessment of how snow cover uncertainty affects the final results and their interpretation.

(Lines: 499-509)

Uncertainty associated with snow cover identification is comparatively minor. The daily cloud-free snow cover product is used solely as a binary mask, with SD set to zero for snow-free pixels and snow-covered pixels processed through the proposed framework. It is not included in model training or validation; thus, its uncertainties do not directly propagate through the regression process, affecting results only through potential snow/no-snow misclassification. SHAP-based feature importance and sensitivity analyses (Figure 14) show that elevation and latitude dominate model predictions, while snow cover－related variables have near-zero mean Shapley values. SHAP dependence plots further confirm that FSC variations do not induce systematic changes in SD estimates relative to terrain and geographic predictors. Overall, snow cover - related uncertainties play a secondary role in the uncertainty budget, and the robustness of the proposed framework is primarily determined by dominant predictors rather than snow cover identification errors.

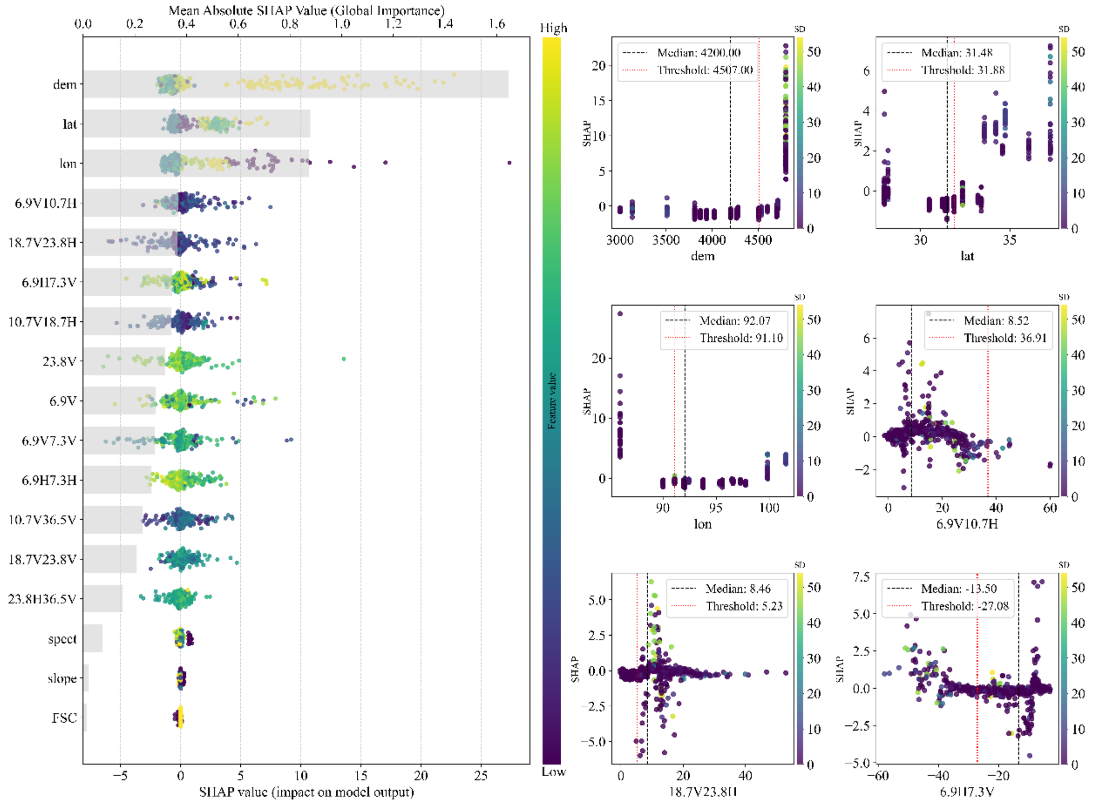


Figure 14: SHAP dependence plots.

5 Please add additional tese to quantify the contribution of each variable used in model training stage and assess the influence of snow cover properties.

Reply 5: We appreciate the reviewers' suggestions. To quantitatively assess the contribution of each variable during model training and to evaluate the snow cover characteristics and the impact of Huang's snow cover data on snow depth inversion, we conducted SHAP-based feature importance and dependency analyses, which have been added to the uncertainty analysis. Details can be found in ***Reply 4***.

6 Please correct "SMER2" to "AMSR2" throughout the manuscript.

Reply 6: Thank you for bringing this issue to our attention. It was caused by a typing error. Having checked the entire manuscript, we found four similar errors and replaced ' SMER2' with 'AMSR2' throughout.