Response to Reviewer 2

We are grateful to the reviewer for the thoughtful and constructive comments. These suggestions have been very helpful for improving the clarity, rigor, and overall quality of the manuscript. We have carefully revised the paper accordingly, and our point-by-point responses are provided below. The reviewer's comments are shown in **black**, and our responses are given in **blue**.

The authors present the development and evaluation of the ChinaAI-FSC dataset, a comprehensive, AI-ready MODIS-based fractional snow cover (FSC) sample collection for China covering 2000–2022. The work aims to establish a standardized, large-scale, and high-quality benchmark for AI-driven snow cover mapping. Considerable effort is evident in data integration, quality control, and validation, and the introduction of the novel "4L-4D-15A" evaluation framework is a clear strength. Overall, the study represents a meaningful contribution to FSC retrieval from MODIS. However, several issues in the current manuscript need to be addressed to better align the presentation with the scientific contribution and to meet the expectations of a high-quality journal.

Response:

We sincerely thank the reviewer for the careful and thoughtful evaluation of our work, as well as for the constructive comments provided. We greatly appreciate the recognition of the efforts made in developing the ChinaAI-FSC dataset and the introduction of the "4L-4D-15A" evaluation framework. Your comments has been extremely valuable in helping us improve the clarity and presentation of our manuscript. In response, we have carefully revised the manuscript to better highlight its scientific contributions to AI-driven fractional snow cover retrieval and to ensure it meets the expectations of a high-quality journal. Detailed responses to each of your comments are provided below.

Major Comments

While the manuscript provides extensive detail on the "AI-ready" nature of the dataset, it repeatedly frames the work more as a project report than a scientific contribution to snow remote sensing. This emphasis, particularly in structure and narrative, risks misleading readers into viewing the paper as a technical documentation of an AI platform rather than a methodological advance in FSC retrieval. To better highlight your unique scientific contribution, I recommend significantly reducing discussion of the AI project framework and refocusing the manuscript on FSC data processing, algorithmic choices, and evaluation rigor. For example, Section 3.3 reads more like a description of an evaluation protocol than an explanation of how it advances FSC validation. Similarly, parts of the text give the impression that your team developed the evaluation methodology itself—please clarify what is novel (or are you just follow NOAA evaluation framework?) versus what is applied.

Response:

We sincerely thank the reviewer for this thoughtful and constructive comment. We agree that in the original version of the manuscript, the scientific contribution was not articulated as clearly as it should have been, and that the narrative at times placed excessive emphasis on the "AI-ready" framework, which could give the impression of a project or platform report rather than a methodological contribution to snow remote sensing.

We would like to clarify that the detailed description of the AI-ready dataset construction workflow was not intended to promote an AI project or platform, but to reflect the intrinsic methodological requirements of building a reliable and reusable FSC dataset for machine learning and deep learning. From an AI-ready Earth observation perspective, steps such as reference FSC calculation, feature extraction, feature-target matching, consistency-based quality control, sample quality assessment (including AI-readiness), and standardized spatial partitioning are essential for ensuring learning validity, reproducibility, and robust model evaluation. Nevertheless, we recognize that these elements were previously presented in an overly project-oriented manner.

In response to the reviewer's suggestions, we have taken several concrete actions. First, we have revised the Introduction and Conclusion to explicitly state that the novelty of this study lies in a methodological contribution; namely, the establishment of a continental-scale, standardized, and quality-controlled FSC dataset paradigm for AI-driven snow mapping, rather than in the development of an AI system or platform. Second, we have substantially reduced and refocused Section 3.3, rewriting it as a dataset-oriented quality and learning-validity assessment rather than an operational evaluation protocol. In particular, we now clearly distinguish between what is adopted from NOAA's AI maturity model and what is newly developed in this study, including FSC-specific adaptations, dataset-level attributes, and their implementation at pixel, tile, and dataset scales.

We believe these revisions significantly reduce project-style descriptions and more clearly highlight the scientific and methodological contributions of this work to the snow remote sensing community.

Section 5 is currently dominated by forward-looking statements about the dataset's future applications, which detracts from the core scientific message. Most readers, including myself, are primarily interested in the FSC dataset itself: how it was produced, its limitations, and how it improves upon existing products. The current discussion is confusing and lacks focus, particularly Section 5.1, which reads like a project roadmap rather than a scientific discussion. I suggest removing Section 5.1 entirely and redirecting the discussion toward substantive issues in FSC retrieval, such as: 1) Training sample selection and representativeness, 2) Impact of sample size and spatial/temporal distribution, 3) Challenges in complex terrain and forested regions, 4) How your approach handles subpixel snow in heterogeneous landscapes. These would strengthen the paper's relevance to the snow remote sensing community.

The arguments in Section 5.2 currently read as personal opinions rather than evidence-based discussion. Please support your claims with relevant literature. Without citations, the section lacks scientific credibility and appears speculative.

Response:

Thank you very much for this constructive and insightful comment. We fully agree that the Discussion section should focus on the scientific implications of the FSC dataset itself, including how the samples were constructed, their representativeness and limitations, and how the dataset advances FSC retrieval, rather than emphasizing forward-looking or project-style applications. In response to this comment, we have substantially revised and refocused Section 5 to address these concerns in the following ways:

The original application-oriented and roadmap-style content has been removed or rewritten. Section 5.1 has been refocused to discuss the methodological implications of AI-ready FSC dataset construction, emphasizing learning validity, feature-FSC consistency, and their relevance to FSC modelling, rather than future applications.

Following the reviewer's suggestions, we have substantially revised the Discussion to explicitly address key scientific issues relevant to FSC retrieval. Section 5.2 now focuses on data availability constraints associated with Landsat and Sentinel-2 observations and the resulting spatial and temporal sample imbalance, clarifying how uneven observation coverage leads to redundancy and imbalance across FSC intervals, and how these characteristics affect sample representativeness. Importantly, this section emphasizes that such imbalance reflects intrinsic snow spatiotemporal variability and optical observation limitations rather than deficiencies in dataset construction, and discusses the implications for AI-based FSC modelling and mitigation strategies.

Section 5.3 further discusses uncertainties and limitations in FSC modelling over complex surfaces, including the effects of subpixel snow heterogeneity, terrain-induced illumination variability, and forest canopy interactions. This section clarifies how these factors jointly influence both reference FSC estimation and predictor variables, and explains how ChinaAI-FSC is designed to make such uncertainties transparent and diagnosable, thereby supporting uncertainty-aware and physically consistent AI-based FSC modelling..

We believe these revisions have significantly improved the focus, scientific rigor, and relevance of Section 5 to the snow remote sensing community.

I would like to know the performance of your dataset in the forested area. If possible, I suggest you attach the relevant analysis results and discussion content.

Response:

We thank the reviewer for this valuable suggestion. In response, we have added an independent validation of the reference FSC in Section 3.3.3 using in situ snow depth (SD) observations from 507 meteorological stations, covering seven snow seasons from 2013 to 2020 and yielding 5016 independent SD-FSC validation pairs. The validation results were stratified by land-cover and terrain conditions to explicitly assess dataset performance in forested and complex environments.

Based on confusion-matrix analysis, the reference FSC shows strong agreement with in situ observations across the entire study area (overall accuracy, OA = 0.944). When stratified by surface conditions, high consistency is maintained across all categories. The highest agreement is observed in mountainous regions (OA = 0.970), indicating that the high-resolution reference construction effectively captures terrain-modulated and heterogeneous snow patterns. However, we explicitly note in the manuscript that the number of validation samples in mountainous regions is relatively limited, which may introduce additional uncertainty and partially inflate the estimated accuracy.

In forested regions, the reference FSC also achieves high agreement with in situ observations (OA = 0.906), although slightly lower than in non-forested areas. This reduction is consistent with canopy occlusion effects that weaken optical snow signals in forest environments. In addition, the smaller number of forested validation samples is acknowledged as a contributing factor to increased uncertainty in the estimated accuracy.

Overall, this independent validation confirms that the reference FSC used in ChinaAI-FSC is physically reliable across diverse land-cover and terrain conditions, while emphasizing that performance metrics in forested and mountainous regions should be interpreted in the context of limited in situ sample density. This limitation is now explicitly discussed in the revised manuscript.

Minor Comments

L12: Remove “mainland”.

Response: Done. The term “mainland” has been removed from Line 12 as suggested.

L54–65: Please add a brief review of prior FSC retrieval studies in challenging environments (e.g., mountainous or forested regions), such as Xiao et al. (2022, JAG).

Xiao et al. 2022. Estimating fractional snow cover in vegetated environments using MODIS surface reflectance data

Response: We thank the reviewer for this helpful suggestion. In response, we have added a concise review of prior FSC retrieval studies in challenging environments, particularly mountainous and forested regions, in the revised Introduction (Lines 63-65). This addition highlights algorithmic strategies that explicitly account for vegetation and terrain effects. Specifically, we now cite and discuss **Czyzowska-Wisniewski et al. (2015)** and **Xiao et al. (2022)**, including *Xiao, X., He, T., Liang, S., Liu, X., Ma, Y., Liang, S., and Chen, X. (2022), Estimating fractional snow cover in vegetated environments using MODIS surface reflectance data, International Journal of Applied Earth Observation and Geoinformation*, which directly addresses FSC retrieval under vegetated conditions using MODIS surface reflectance data.

L74: Consider removing “AI-ready” here to frame the research gap more broadly.

Response: We appreciate the reviewer's suggestion. In the revised text, "AI-ready" has been removed from point (1) to avoid premature introduction of the concept. The paragraph now emphasizes the lack of large-scale FSC datasets suitable for AI-based modelling

L74–80: The two stated objectives appear redundant. Clarify whether they represent distinct goals or rephrase to avoid repetition. Given that the primary output is an FSC dataset, focus the motivation on its scientific value—not its compatibility with AI workflows.

Response: We thank the reviewer for the comment. We have substantially revised the text to clarify that the two factors are related but distinct:

(1) the absence of large-scale FSC datasets suitable for AI-based modelling, highlighting the scientific value of creating a comprehensive benchmark dataset, and

(2) the lack of standardized protocols for dataset construction and evaluation, emphasizing methodological reproducibility and transparency.

These revisions clearly distinguish the scientific contribution of ChinaAI-FSC from the methodological framework for AI-ready construction and evaluation. The paragraph now also naturally leads into the formal introduction of AI-ready dataset principles in the following section.

L101: Suggest revising to: "Standardized AI-ready metadata and unified evaluation protocols."

Response: We thank the reviewer for the suggestion. The text has been revised accordingly to read: "Standardized AI-ready metadata and unified evaluation protocols" (new Line 108).

L106–123: Avoid restating the abstract. Provide a concise overview of the study's scope and structure instead.

Response:

Thank you for this helpful suggestion. We agree that the original ending of the Introduction resembled a condensed version of the abstract rather than a concise overview of the study's scope and structure.

In response, we have rewritten the final paragraph of the Introduction to avoid restating dataset details, model lists, or experimental specifics. The revised paragraph now provides a high-level overview of the study's scope, methodological focus, and paper organization, clearly emphasizing that the contribution lies in AI-ready FSC dataset methodology rather than algorithm development. It briefly outlines the workflow, large-scale validation over mainland China, and the methodological innovation introduced by the unified dataset paradigm and the 4L–4D–15A AI-readiness evaluation framework.

We hope this revision improves clarity, reduces redundancy with the abstract, and better guides readers into the structure and scientific focus of the paper.

Section 3.1.1: Clarify the acquisition and processing specifics of the two satellite datasets (Landsat and Sentinel-2).

Response: Thank you for this helpful comment. We have substantially revised Section 3.1.1 to clearly describe the data sources, product levels, and subsequent preprocessing steps for both Landsat and Sentinel-2 datasets. The revised text explicitly distinguishes between (i) the use of standard, atmospherically corrected surface reflectance products and (ii) additional quality control and reconstruction procedures applied by our team.

L156–159: Were surface reflectance data for Landsat and Sentinel-2 processed by your team, or were standard products used?

Response: We used standard, officially released surface reflectance products for both sensors. Specifically, Landsat Collection 2 Level-2 Surface Reflectance products (Landsat-5 TM, Landsat-7 ETM+, Landsat-8 OLI, and Landsat-9

OLI-2) were obtained from the USGS Earth Explorer, and Sentinel-2A/2B MSI Level-2A Surface Reflectance products were acquired from the ESA Copernicus Open Access Hub. These products are already atmospherically corrected using well-established algorithms (e.g., LEDAPS and LaSRC for Landsat; Sen2Cor for Sentinel-2), and no additional atmospheric correction was performed by our team. This clarification has been explicitly added in the revised manuscript.

L160–163: Clarify whether cloud masking was performed using your own implementation of CFMask (Landsat) and SCL (Sentinel-2), or if you relied solely on the native QA layers.

Response: Cloud and shadow masking relied solely on the native quality layers provided with the standard products, rather than a custom re-implementation of the algorithms. For Landsat imagery, clouds, cirrus, and cloud shadows were identified using the QA_PIXEL band generated by the CFMask algorithm included in the Collection 2 Level-2 products. For Sentinel-2 imagery, cloud and shadow pixels were masked using the Scene Classification Layer (SCL) provided with the Level-2A products. This has now been clarified in Section 3.1.1 to avoid any ambiguity regarding algorithm implementation.

L164–165: Was the interpolation of Landsat-7 ETM+ SLC-off gaps performed by your team, or did you use an existing gap-filled product? Please specify.

Response: The gap filling for Landsat-7 ETM+ images affected by the SLC-off failure was performed by our team, rather than using an existing pre-filled product. Specifically, we adopted a local neighborhood linear interpolation strategy following the method of Chen et al. (2011), in which missing pixels were estimated from adjacent valid observations along the scan-line direction and subsequently smoothed using a 3 × 3 spatial kernel to ensure spatial continuity and radiometric consistency. This reference and methodological clarification have now been added to Section 3.1.1.

Section 3.1.2:

1) Replace “MODIS data” with “MODIS series products” or similar for precision.

Response: Thanks for the comment. The text has been revised to replace “MODIS data” with “MODIS series products”.

2) Briefly describe the seamless surface reflectance processing algorithm to help readers understand that this product—rather than standard MOD09GA—is the foundation of your FSC retrieval.

Response: We thank the reviewer for this valuable comment. We have added a brief description of the SDC500 dataset. In the revised manuscript, Section 3.1.2 now includes the following description:

“The surface reflectance data were obtained from the Global 500 m seamless MODIS-derived dataset (SDC500) for 2000-2022 (Liang et al., 2024). Unlike the standard MOD09GA, SDC500 reconstructs a continuous daily 500 m reflectance time series by correcting BRDF effects, detecting outliers, and filling missing values with phenology-guided spline interpolation. Snow and snow-free periods are treated separately to preserve seasonal reflectance dynamics. The dataset demonstrates high accuracy, with a mean absolute error of only 0.043, providing a reliable basis for FSC retrieval.”

Section 3.2.2: Why were all input variables retained without feature selection? In many FSC applications, not all predictors contribute meaningfully, and including redundant variables can reduce model efficiency and interpretability (e.g., Xiao et al., 2022, JAG). Please justify your approach.

Response:

Thank you for this important comment. We fully agree that, for a specific FSC retrieval model, feature selection can improve efficiency and interpretability. However, in this study, our objective is not to optimize a single empirical model, but to construct an AI-ready, physically meaningful, and generally applicable FSC training dataset.

The 20 predictors (Ref1-Ref7, NDSI, NDVI, LC, LST, FTC, elevation, slope, aspect, terrain relief, surface roughness, longitude, latitude, and Julian day) were selected based on extensive evidence from previous snow-cover and cryospheric remote sensing studies, which consistently demonstrate their relevance to snow spectral behavior, vegetation masking effects, surface energy balance, and topographic controls on snow distribution. Collectively, these variables characterize the spectral and radiometric properties of snow, land-cover and canopy influences, thermal and energy-state constraints, topographic controls on snow accumulation and ablation, as well as the spatiotemporal context and climatic gradients of snow processes.

Rather than performing dataset-specific feature pruning, we intentionally retained all physically plausible predictors to avoid prematurely discarding information that may be critical under different climatic, ecological, or algorithmic settings. This design enables dataset users to flexibly explore feature selection strategies and model architectures, and to systematically assess the impacts of predictor combinations on FSC estimation performance and uncertainty.

Therefore, the retained feature set prioritizes generality, physical completeness, and reusability, which is consistent with the purpose of a benchmark-oriented FSC dataset.

Section 3.2.4: Support your threshold choices (e.g., 0.2 for Ref4, 0.4 for Ref2 and Ref6) with references or sensitivity analyses.

Response:

Thank you for this comment. In the revised Section 3.2.4, we have explicitly linked the reflectance thresholds (e.g., Ref4 $\geq$ 0.2, Ref2 $\geq$ 0.4, Ref6 $\leq$ 0.4) to the well-established spectral characteristics of snow reported in previous studies (Dozier, 1989; Hall et al., 2002). These values fall within the commonly reported reflectance ranges for snow in the corresponding MODIS bands and are consistent with prior MODIS snow-mapping literature.

These thresholds were chosen conservatively and are used exclusively as screening criteria in the quality control process. Our purpose is not to impose strict spectral constraints, but to remove samples exhibiting extreme spectral inconsistencies that contradict known snow physics. As clarified in the manuscript, the resulting dataset statistics and spatial patterns are robust to reasonable variations around these threshold values.

L283: Replace “violate” with “fail to meet.” Please check the entire manuscript for similar phrasing.

Response: Done. We have replaced “violate” with “fail to meet”. In addition, we carefully reviewed the entire manuscript and revised similar phrasing to ensure consistent and appropriate terminology throughout.

Equation 2: Provide a clearer physical or empirical rationale for the formulation.

Response:

We appreciate this helpful comment. In the revised manuscript, the physical rationale for the Equation (now Eq. (3)) has been clarified. Equation (3) is designed to capture the well-documented preferential accumulation and persistence of snow on north-facing (shaded) slopes compared to south-facing (sun-exposed) slopes due to reduced incoming solar radiation and lower melt rates.

The 50% ratio adopted in Eq. (3) represents a conservative lower-bound criterion rather than a strict quantitative relationship. Its purpose is to exclude only samples that clearly contradict the expected topographic modulation of snow distribution, while allowing for natural variability at local scales. This formulation is supported by established empirical findings on aspect-controlled snow distribution in mountainous terrain (e.g., Grünewald et al., 2013), as now explicitly stated in the manuscript.

Figures 7, 9, 10: Include spatial scales and clearly label the regions of analysis.

Response: We thank the reviewer for this comment. Figures 7, 9, and 10 (renumbered as Figures 9, 11, and 12 in the revised manuscript) each present a single 128×128 MODIS-pixel tile to illustrate spatial patterns of FSC and associated features. The corresponding geographic locations (i.e., the regions of analysis) of these tiles are explicitly indicated in the revised manuscript in Figures 3 and 10. By providing these reference maps, readers can identify the regions of analysis and understand the spatial context of the presented tiles.

Section 4 heading: Consider renaming to "Demonstration of Applications Using the AI-Ready FSC Dataset" or similar.

Response: We thank the reviewer for this suggestion. The Section 4 heading has been revised to "Demonstration of Applications Using the AI-Ready FSC Dataset" in the updated manuscript, as recommended.

L447–448: The current statement is too vague. Elaborate on the specific factors influencing FSC accuracy (e.g., illumination, forest structure, grain size).

Response: We thank the reviewer for this comment. The original statement has been revised to provide a more detailed explanation of the factors influencing FSC estimation errors. The updated text now specifies the contributions of spectral saturation, canopy occlusion, mixed-pixel effects, terrain-induced illumination variations, forest structure, and snow grain size, which collectively affect the accuracy of FSC retrieval.

L602–603: The claim that "expanded feature space enables AI models to better characterize complex snow–terrain–climate interactions" is speculative without evidence. Rephrase to clarify what you mean, e.g., which features improve representation of which physical processes?

Response: We thank the reviewer for this comment. We agree that the original statement claiming that the "expanded feature space enables AI models to better characterize complex snow-terrain-climate interactions" was speculative and lacked direct supporting evidence. In the revised manuscript, Section 5 has been substantially reorganized to remove such claims. The discussion now emphasizes methodological considerations, dataset limitations, and uncertainty characterization. Specifically, Section 5.1 highlights the methodological implications of AI-ready FSC dataset construction, focusing on learning validity, physical consistency, and structural coherence. Section 5.2 addresses data availability constraints and sample imbalance, while Section 5.3 discusses uncertainties arising from subpixel heterogeneity, terrain illumination, and canopy effects. Finally, Section 5.4 presents the AI-readiness evaluation framework, a transferable methodological approach for constructing and assessing AI-ready geophysical datasets beyond snow cover. This restructuring ensures that the discussion is grounded in the demonstrated evidence and methodological focus of the study, without making speculative claims about feature-process relationships.