Response to Reviewer 1

We greatly appreciate the reviewer's insightful and constructive comments, which have significantly helped to improve the quality and clarity of this manuscript. The manuscript has been carefully revised accordingly. The reviewer's comments are shown in **black**, and our responses are provided in **blue**.

This manuscript presents the development and evaluation of the ChinaAI-FSC dataset, a comprehensive AI-ready MODIS fractional snow cover sample collection for China spanning 2000-2022. The objective of this work is to provide a standardized, large-scale, and high-quality benchmark for AI-driven snow cover mapping. The authors have undertaken a substantial effort in data integration, quality control, and validation, and the introduction of a novel "4L-4D-15A" evaluation framework is a notable strength. However, the manuscript in its current form has several issues that need to be justified. The most critical concerns revolve around the potential imbalance of samples across varying snow conditions and geographic regions, as well as insufficient discussion regarding the sources and mitigation of uncertainty. These aspects affect the perceived robustness and broad applicability of the dataset and must be thoroughly addressed before publication.

Response:

We sincerely thank the reviewer for the careful reading of our manuscript and for the insightful comments. We agree with the concerns regarding sample imbalance and uncertainty, which are indeed important for assessing the applicability of AI-ready FSC datasets. In response to these concerns, we have carefully revised the manuscript following these suggestions.

Regarding sample imbalance, we have revised the manuscript to explicitly acknowledge and discuss this issue in Section 5.2 (Data availability constraints and sample imbalance). We clarify that the construction of ChinaAI-FSC is fundamentally constrained by the availability of near-cloud-free, high-quality Landsat and Sentinel-2 observations, which are inherently uneven in space and time, especially over mountainous and persistently cloudy regions. Consequently, the dataset was assembled by collecting all available reference-quality observations that met strict quality criteria, rather than by imposing an explicit sampling strategy aimed at balancing snow conditions. As a result, some imbalance remains, particularly for extremely sparse and extremely dense snow cover fractions.

We also clarify that such imbalance reflects the intrinsic spatiotemporal characteristics of snow cover and the observational limitations of optical remote sensing, rather than deficiencies in the dataset construction methodology. To assess whether this imbalance critically limits AI-based FSC modelling, we conducted baseline experiments using simple benchmark models without specialized architectures or extensive parameter tuning. These tests indicate that, even under the existing sample distribution, AI-based FSC estimation can achieve satisfactory performance, suggesting that the dataset already provides meaningful learning signals. At the same time, we note that incorporating more advanced sample representativeness analysis techniques (e.g., hierarchical clustering, density-based selection, or reweighting strategies) may further improve model performance, and we explicitly frame ChinaAI-FSC as a reference dataset that enables such methodological investigations.

Regarding uncertainty, we have substantially expanded the discussion in Section 5.3 to systematically identify major uncertainty sources, including cloud contamination, terrain-induced illumination effects, forest canopy obscuration, mixed-pixel conditions, and cross-sensor inconsistencies. We explain how these uncertainties propagate into both FSC reference estimates and predictor variables, particularly in complex terrain and forested environments. While quality control and physically consistent feature selection help mitigate these effects, we acknowledge that residual uncertainties are unavoidable in optical snow remote sensing. We further discuss potential mitigation pathways, including uncertainty-aware modelling, ensemble learning, and probabilistic or Bayesian frameworks, to provide users with a clearer understanding of dataset limitations and appropriate usage.

Overall, the revised manuscript places greater emphasis on transparently characterizing sample imbalance and uncertainty, rather than minimizing or obscuring them. We hope that these revisions clarify the realistic scope, limitations, and methodological value of the ChinaAI-FSC dataset, and address the reviewer's concerns regarding its applicability and scientific grounding.

Major Comments:

What is the innovative aspect of this article? Is it a technological innovation or a methodological innovation?

Response:

Thank you for raising this important point. We clarify that the primary innovation of this study is **methodological rather than algorithmic**.

At present, there is no AI-ready FSC sample repository at either regional or hemispheric scale, nor a standardized and reproducible framework for constructing and validating FSC datasets explicitly designed for machine learning and deep learning. Existing FSC studies typically rely on locally collected samples or product-level validations, which are highly heterogeneous in reference generation, quality control, and data organization, thereby limiting reproducibility, cross-regional model generalization, and fair algorithm comparison.

The core innovation of this work is the establishment of a continental-scale AI-ready FSC dataset paradigm, together with a transparent construction and evaluation framework. Specifically, the methodological contributions include:

(1) A continental-scale AI-ready FSC sample repository

We construct ChinaAI-FSC over mainland China, a region encompassing most major Northern Hemisphere snow regimes, providing a representative and standardized benchmark that bridges the gap between local experimental datasets and global product-level archives.

(2) A complete and transparent AI-ready dataset construction workflow

We develop a transparent workflow covering high-resolution reference FSC generation, multi-source feature extraction, feature-target matchup construction, pixel- and tile-level consistency-based quality control, and standardized spatial tiling (128×128 MODIS pixels), producing learning-ready samples without ad-hoc preprocessing.

(3) Consistency-based multi-layer quality control

A pixel- and tile-level screening strategy is implemented to reduce unreliable or internally inconsistent samples, improving the robustness and generalizability of AI model training.

(4) A formal AI-readiness evaluation framework (4L-4D-15A)

We introduce an FSC-oriented extension of NOAA's AI maturity model to systematically assess dataset-level AI-readiness across multiple dimensions.

To avoid any ambiguity, we have revised the Introduction and Conclusion to explicitly state that the novelty of this work lies in dataset methodology, standardization, and AI-readiness engineering, rather than in proposing a new FSC retrieval algorithm.

The statements regarding the dataset's utility appear overstated or misaligned with its actual characteristics as presented. The authors should modify these claims to accurately reflect the dataset's demonstrated strengths and limitations.

Response:

We appreciate this important comment. We agree that some statements in the original manuscript regarding the utility of the ChinaAI-FSC dataset could be interpreted as broader than what is directly demonstrated. In response, we have revised the manuscript to ensure that all claims are grounded in experimental evidence and clearly acknowledge the dataset's limitations.

In particular, the final paragraph of the Introduction has been revised to better align with the analyses presented in the paper. The revised text now emphasizes that ChinaAI-FSC is a standardized and AI-ready MODIS FSC dataset for China, spanning 22 snow seasons, and highlights the methodological focus on constructing, validating, and evaluating FSC datasets suitable for large-scale machine learning and deep learning applications, rather than promoting speculative applications. This revision clarifies the dataset's purpose and positions it within a reproducible workflow, including feature-FSC matching, consistency-based quality control, sample quality assessment, and the AI-readiness evaluation framework (4L-4D-15A).

Additionally, Section 5.1 "Potential Applications of the ChinaAI-FSC Dataset" has been replaced with "Methodological Implications of AI-Ready FSC Dataset Construction", which emphasizes learning validity, physical consistency, and structural coherence in dataset construction rather than any untested downstream applications. The revised manuscript also clarifies key limitations related to observational constraints, sample imbalance, and residual uncertainties arising from subpixel heterogeneity, terrain effects, and canopy occlusion (as shown in new section 5.2 and 5.3). By making these aspects transparent, the dataset supports uncertainty-aware modelling and provides a reproducible, methodologically rigorous foundation for AI-driven FSC estimation.

Finally, while the current implementation focuses on snow cover, the introduced AI-readiness evaluation framework provides a transferable methodological reference for constructing and assessing AI-ready geophysical datasets more broadly.

Among these samples, what is the proportion of completely snow-free cases? If such samples account for an excessively high percentage, their contribution to research on fractional snow cover estimation may be limited.

Response:

Thank you for this important question. We acknowledge that the original manuscript did not explicitly report the proportion of completely snow-free samples, which is an important aspect for interpreting the value of the dataset for FSC modelling.

In fact, completely snow-free samples are not included in ChinaAI-FSC by design. During dataset construction, we applied explicit physical and statistical filtering to exclude tiles that carry little information for FSC learning. Specifically, we defined snow-covered pixels using an FSC threshold of FSC≥15%, and then removed samples whose mean snow-covered pixel fraction was <5% or >95%. These two classes correspond to nearly "snow-free" and "fully snow-covered" homogeneous samples, respectively. Such samples provide limited learning value for FSC estimation, because they contain little internal variability and do not represent mixed snow-land conditions that are essential for FSC modelling. Their inclusion would also bias the sample distribution toward trivial cases, weakening the ability of AI models to learn sub-pixel snow variability.

As a result, ChinaAI-FSC contains no completely snow-free tiles, and the dataset is explicitly designed to focus on spatially heterogeneous and physically informative snow conditions that are most relevant for FSC estimation.

To clarify this, we have added the following description in the manuscript between the sentences in the section 3.2.3.

In addition, pixels with FSC ⩾ 15% (Painter et al., 2009; Zhang et al., 2019) were defined as snow-covered, and samples with mean snow-covered fractions <5% or >95% were excluded, corresponding to nearly snow-free and fully snow-covered homogeneous conditions. Such samples contain little internal variability and provide limited value for learning fractional snow-land relationships. Consequently, ChinaAI-FSC contains no completely snow-free or fully snow-covered samples, but focuses on spatially heterogeneous mixed snow conditions.

The newly added text explains the FSC thresholding and tile-level screening (removal of <5% and >95% snow-covered samples) and explicitly states that homogeneous snow-free and fully snow-covered samples are excluded.

In addition, we found that the corresponding labels in original Figure 5 (now Figure 6 in the revised manuscript) were not sufficiently clear regarding this screening, and we have revised the figure accordingly to ensure consistency with the dataset definition.

The references on fractional snow cover retrieval are incomplete, lacking citations for mixed-pixel unmixing algorithms such as the MESMA-AGE algorithm.

Response:

We sincerely thank the reviewer for this valuable suggestion. We fully agree that including recent advances such as the MESMA-AGE algorithm is important to provide a more complete discussion of spectral mixture methods for FSC retrieval.

In response, we have added a description of the MESMA-AGE algorithm at the end of the second paragraph of the Introduction. The added text reads: "Recent advances, such as the Multiple Endmember Spectral Mixture Analysis with Automated Global Endmember selection (MESMA-AGE), have partially alleviated these limitations by dynamically selecting optimal endmember combinations from large spectral libraries and accounting for variability in snow, vegetation, soil, and illumination conditions. This strategy has enabled improved sub-pixel FSC estimation over complex mountain environments, and has been successfully applied to generate daily MODIS fractional snow cover products for the Asian Water Tower region, which is characterized by extreme terrain, heterogeneous land cover, and highly variable snow conditions (Pan et al., 2024)."

We have also added the corresponding reference in the reference list:

Pan, F., Jiang, L., Wang, G., Pan, J., Huang, J., Zhang, C., Cui, H., Yang, J., Zheng, Z., Wu, S., and Shi, J.: MODIS daily cloud-gap-filled fractional snow cover dataset of the Asian Water Tower region (2000-2022), Earth Syst. Sci. Data, 16, 2501-2523, https://doi.org/10.5194/essd-16-2501-2024, 2024.

This AI-ready MODIS FSC data needs to be evaluated with independent data. In the current version, I didn't get this.

Response:

Thank you for this important suggestion. In response, we have added an independent validation of the reference FSC using in-situ snow depth observations from 507 meteorological stations. Due to data availability, these observations could only be obtained for seven snow seasons (2013–2020), resulting in 5,016 SD-FSC pairs (new Section 3.3.3)

The validation shows strong overall agreement (OA = 0.944). Stratified analysis further demonstrates high consistency in general regions (OA = 0.940), complex mountainous regions (OA = 0.970), and forested regions (OA = 0.906). The slightly lower accuracy in forested areas is consistent with known canopy occlusion effects, and the limited number of validation samples in mountainous and forested regions is also acknowledged.

These results confirm the physical reliability of the FSC reference used in ChinaAI-FSC under heterogeneous surface and terrain conditions.

Section 3.1.1, Both Landsat and Sentinel-2 provide atmospherically corrected surface reflectance data, and the authors need to clarify what level of data from Landsat and Sentinel-2 is used. Did the authors use the available surface reflectance data or make atmospheric correction by themselves?

Response:

We thank the reviewer for raising this important point. We acknowledge that in the original manuscript, the description of the Landsat and Sentinel-2 surface reflectance data was not clearly stated. To address this issue, we have clarified these details at the beginning of Section 3.1.1: we used Landsat Collection 2 Level-2 Surface Reflectance (SR) products for Landsat-5 TM, Landsat-7 ETM+, Landsat-8 OLI, and Landsat-9 OLI-2, obtained from the USGS Earth Explorer platform (https://earthexplorer.usgs.gov/); and Sentinel-2A/2B MSI Level-2A SR products, obtained from the ESA Copernicus Open Access Hub (https://scihub.copernicus.eu/). These products are already atmospherically corrected and suitable for quantitative analysis (Masek et al., 2006; Vermote et al., 2016; Louis et al., 2016).

Line #99, during the snow accumulation period, cloud cover is often frequent. How are these samples obtained relying solely on the satellite sensors mentioned in the paper? I'm also confused on the seamless FSC data produced this work, since it mostly relies on MODIS band 1-7 reflectance data, which is contaminated with cloud cover.

Response:

We thank the reviewer for this important comment. We realize that the original manuscript did not clearly describe the data selection and processing strategy. We have addressed the reviewer's concerns regarding cloud interference as follows:

For the Landsat and Sentinel-2 imagery used to calculate FSC reference truth, we selected only scenes with overall cloud cover below 15% to approximate clear-sky conditions. Residual low-quality pixels in these scenes, including clouds, cirrus, and cloud shadows, were subsequently reconstructed using a spectrally constrained spatial gap-filling approach. Specifically, low-quality pixels in Landsat images were identified using the QA_PIXEL band generated by the CFMask algorithm (Zhu et al., 2015), while for Sentinel-2 images, cloud and shadow pixels were masked using the Scene Classification Layer (SCL) (Drusch et al., 2012). Masked or low-quality pixels in both datasets were subsequently reconstructed by estimating reflectance values from neighboring valid pixels with similar spectral characteristics, preserving local spectral consistency and spatial continuity (Chen et al., 2011). These details are now explicitly described at the beginning of Section 3.1.1.

For the MODIS surface reflectance, we used the Global 500 m seamless MODIS-derived dataset (SDC500) for 2000-2022 (Liang et al., 2024), as described in Section 3.1.2. Unlike the standard MOD09GA product, SDC500 reconstructs a continuous daily 500 m reflectance time series by correcting BRDF effects, detecting outliers, and filling missing values using phenology-guided spline interpolation. Snow and snow-free periods are treated separately to preserve seasonal reflectance dynamics. This preprocessing ensures that cloud contamination and data gaps are effectively mitigated, providing a reliable and seamless basis for FSC retrieval at the MODIS scale. The dataset demonstrates high accuracy, with a mean absolute error of only 0.043, further ensuring its suitability for quantitative FSC analysis.

In Figure 2, the first occurrence of any English abbreviation should be accompanied by its corresponding explanation.

Response:

We thank the reviewer for this valuable comment. All English abbreviations in Figure 2 now include their full explanations in the figure caption, ensuring clarity for the readers.

Section 3.2.4, The formula for (FSC, NDSI) should be provided. Additionally, what is the basis for determining the thresholds of FSC, NDSI, Refl, T_{base}, and others? This is also needs to be justified clearly.

Response:

Thank you for this detailed and constructive comment. We have revised Section 3.2.4 to explicitly address both aspects raised.

First, the mathematical formulation describing the relationship between FSC and NDSI has now been clearly provided. In the revised manuscript, Eq. (1) explicitly defines the Pearson correlation coefficient ρ(FSC, NDSI), which is computed across all pixels within each sample tile. This formulation serves as the core metric for identifying spectrally inconsistent samples during quality control.

$$\rho(\mathrm{NDSI},\ \mathrm{FSC}) = \frac{\sum_{i=1}^{128}\sum_{j=1}^{128}\left(\mathrm{NDSI}_{ij}-\overline{\mathrm{NDSI}}\right)\left(\mathrm{FSC}_{ij}-\overline{\mathrm{FSC}}\right)}{\sqrt{\sum_{i=1}^{128}\sum_{j=1}^{128}\left(\mathrm{NDSI}_{ij}-\overline{\mathrm{NDSI}}\right)^2}\sqrt{\sum_{i=1}^{128}\sum_{j=1}^{128}\left(\mathrm{FSC}_{ij}-\overline{\mathrm{FSC}}\right)^2}} \tag{1}$$

Second, the basis for determining all thresholds used in Section 3.2.4 has been explicitly clarified in the revised manuscript. Each threshold is introduced together with its physical or empirical justification in the corresponding subsection. Specifically, thresholds for FSC, NDSI, and surface reflectance are derived from well-established spectral characteristics of snow, and are consistent with commonly adopted values in MODIS-based snow cover studies (e.g., Dozier, 1989; Hall et al., 2002). These thresholds are used to exclude only physically implausible or spectrally contradictory conditions.

The temperature-related threshold (T_{base}) is anchored to the physical freezing point of water and is further adjusted using an elevation-dependent lapse-rate correction to account for topographic temperature gradients. This formulation reflects basic thermodynamic constraints on snow presence and is applied as a consistency check rather than a detailed energy balance model.

Importantly, all thresholds in Section 3.2.4 were selected conservatively and are applied solely for quality control purposes. Their role is to remove samples that clearly violate known physical relationships between snow cover, spectral reflectance, temperature, and topography, while retaining representative and physically plausible snow conditions. The justification for each threshold is now explicitly stated in the manuscript to ensure transparency, reproducibility, and clarity.

Section 3.4, the rationale behind the author's dataset partitioning requires further discussion. According to the author's partitioning logic, the validation set and test set could essentially be obtained by interpolating from the training set, which may lead to suboptimal results from such a partitioning approach.

Response:

We thank the reviewer for this constructive comment. In the revised manuscript, Section 3.4 has been substantially revised to clarify the rationale and implementation of the dataset partitioning strategy. Specifically, the dataset is partitioned by explicitly enforcing spatial disjointness across training, validation, and testing subsets within each major snow-climate subregion of China. A spatially disjoint 2:1:1 partitioning scheme is adopted to minimize spatial autocorrelation and prevent information leakage, ensuring that validation and test samples cannot be obtained through spatial interpolation from the training set.

In addition, each spatial subset spans 22 snow seasons, and samples from different snow seasons are treated as independent realizations of snow-environment interactions, reflecting strong interannual variability. This combination of spatial separation and long-term temporal coverage ensures that model evaluation assesses both spatial and interannual generalization rather than local interpolation performance. The revised partitioning strategy therefore provides a more rigorous and realistic benchmark for AI-based FSC models, avoiding overly optimistic performance estimates caused by spatial or temporal leakage.

In Figures 9 and 10, where does the MODIS result in the last column come from? Since the MODIS result is identified as cloudy, the author only provides feature variables without addressing cloud identification. Therefore, the cloud coverage in the other columns should be consistent with the last column. However, the author's results show retrieved fractional snow cover. Please provide justification for the validity of these results.

Response: We thank the reviewer for this important comment. The MODIS result shown in the last column of Figures 9 and 10 (new Figures 11 and 12 in the revised manuscript) represents FSC estimates obtained by directly applying the official fitting coefficients provided by Salomonson and Appel (2004) to the NDSI values in the MOD10A1 product. In contrast, the 2~7 columns are based on input data derived from the seamless MODIS surface reflectance product (SDC500) produced by Liang et al. (2024), which has been preprocessed to correct for clouds, outliers, and missing values, resulting in cloud-free and spatially continuous surface reflectance inputs. Therefore, although cloud contamination exists in the MODIS standard FSC product, the retrieved FSC shown in the other columns remains physically valid due to the use of cloud-corrected SDC500 inputs. We have added explicit clarification of this distinction and the corresponding justification in the revised manuscript.

Additionally, the AI results trained on the dataset provided by the authors show a significant visual discrepancy from the reference values. Why does this occur?

Response: We thank the reviewer for this observation. Visually, the AI-estimated FSC results do exhibit some differences from the reference truth values. However, they show clear improvements compared with the standard MODIS FSC product, and quantitative evaluation confirms that the models still achieve high estimation accuracy. The visual discrepancies mainly arise because all six benchmark AI models were implemented with standard, simple architectures without specific optimization, and the sample size is limited to a single snow season, which may reduce representativeness for some local conditions. Nevertheless, the results indicate that even simple models can achieve robust FSC estimation when trained on high-quality, physically consistent, and representative data, highlighting that the decisive factor for AI-based FSC performance is the dataset rather than the specific model architecture. Further improvements are expected by incorporating samples from multiple snow seasons and optimizing model structures.

Section 4.2, Line 489, How did the author incorporate the 'official retrieval algorithm'? Was it through direct adoption of the fitting coefficients or through refitting?

Response: We thank the reviewer for the comment. In the revised manuscript, we have clarified that FSC estimates were obtained by directly applying the official fitting coefficients provided by Salomonson and Appel (2004), without any refitting.

In Figure 12(b)(d)(f), why is there no cloud coverage shown in these images?

Response: We thank the reviewer for this comment. The reason no cloud coverage is visible in Figure 12(b,d,f) (new Figure 14 in the revised manuscript) is that our input data are derived from the seamless MODIS surface reflectance product (SDC500) produced by Liang et al. (2024), which has been preprocessed to correct for clouds, outliers, and missing values.

Minor Comments:

Line 324 is missing a period at the end. Please carefully review the entire text to avoid similar issues.

Response: Thank you for pointing this out. We have performed a thorough check of the entire text to ensure that all punctuation and formatting issues have been addressed.