Dear Prof. Jinkyu Hong,

Many thanks for handling the review process for our manuscript. The time and effort devoted to our manuscript by you are very much appreciated. We thank the reviewers for their insightful and constructive feedback. We have carefully addressed each point to improve the clarity, robustness, and accessibility of our manuscript. Below are a summary of our responses and the corresponding changes made to the manuscript.

EDITOR COMMENTS

1. Emphasize the G6multi–G6single contrast (snow-scheme-only attribution). Please focus your analysis and narrative on the differences between G6multi and G6single, as this pair most directly isolates the impact of the multi-layer snow scheme within the same forecast system. Your core claim concerns the effect of introducing a multi-layer snow scheme on seasonal forecast skill and land–atmosphere coupling; therefore, the primary attribution pathway should be consistently supported by evidence from G6multi − G6single, and—where appropriate—the corresponding offline LSM contrast JULESmulti − JULESsingle. At present, it is difficult to locate, within the figures and accompanying discussion, the key results that directly support the argument as stated.

➔ *We agree that the G6multi–G6single contrast is the most direct way to attribute changes solely to the snow scheme. In the revised manuscript, we have restructured Sections 4.1 to prioritize this comparison. While G5single remains as a historical baseline to show overall system evolution, the narrative now centers on the GloSea6 experiments to support our core claims regarding snow insulation and its subsequent impacts on soil moisture and coupling. We edited Fig. 2 to try to focus on the comparison between G6multi and G6single and its relevant text is also updated in the manuscript.*

2. Moderate the strength of the conclusions in light of methodological limitations. Please consider toning down several arguments, given the following issues and limitations that may affect robustness and interpretation:

- Ensemble-size imbalance and implications for significance/robustness: The experimental design uses different ensemble sizes across key configurations (e.g., 7 members versus 4 members). This imbalance may influence the estimated mean responses and statistical power. If feasible, please provide an appropriate robustness assessment (e.g., a matched-size subset analysis, resampling/bootstrapping, or another clearly justified approach) and clearly state how this imbalance affects your inference.

➔ *We acknowledge the difference in ensemble sizes (G6multi: 7 members, G6single: 4 members). To ensure robustness, we performed a sensitivity test by sub-sampling 4 members from G6multi. We found that the climatological mean signals remained consistent, confirming that the imbalance did not significantly bias our primary findings. We have added a statement regarding this in Section 3 (Lines 292-295).*

"To account for the imbalance in ensemble sizes between G6multi (7 members) and G6single (4 members), a resampling analysis was conducted. Results using a matched 4-member subset of G6multi showed no statistically significant difference from the 7-member mean for the variables analyzed, suggesting the findings are robust to ensemble size."

- Sensitivity and limitations of the Granger-causality analysis: While Granger causality can be informative as a diagnostic tool, it is sensitive to assumptions and preprocessing choices. Please provide a clearer statement of its limitations with respect to establishing physical causality, and discuss the potential for confounding by common large-scale atmospheric drivers.

➔ *We have expanded Section 3.2 (Lines 341-345) to more explicitly state that Granger causality identifies predictive precedence rather than definitive physical causation. We now discuss the potential for large-scale atmospheric drivers to act as confounding variables.*

"Nevertheless, as Granger causality only tests for predictive precedence, the results may reflect statistical associations of predictive precedence due to shared external drivers and should not be interpreted as definitive physical causation between both variables. Specifically, shared external atmospheric drivers can influence both evaporation and precipitation, potentially confounding the identified causal links."

- Sensitivity of soil-moisture memory and coupling metrics: The soil-moisture memory metric and coupling indices are central to your mechanistic narrative (snow → melt timing → SM → LH → temperature/precipitation). However, these diagnostics can be sensitive to forecast drift and lead-time dependence. Please consider additional clarification and/or sensitivity analyses addressing these dependencies (e.g., lead-time stratification, drift treatment, or related robustness checks).

➔ *We appreciate the reviewer's insightful comment regarding the potential sensitivity of soil-moisture memory (SMM) and coupling metrics to forecast drift and lead-time. To address this, we have added clarification in Section 3.1 and 3.2 regarding how we treated the forecast data to minimize the impact of drift. Specifically, the SMM was calculated using concatenated time series of 1-month forecasts from each prior month's initialization, which helps mitigate the influence of long-term model drift that often intensifies at longer lead times.*

"In the calculation of the SMM in both seasonal forecast systems, to minimize the impact of systematic forecast drift, the JJA SM time series for each year are constructed by concatenating 1-month lead forecasts for each respective month, specifically June from the 1 May initialization, July from 1 June, and August from 1 July."

"The analysis is conducted using MJJA time series of 24-year forecast runs initialized on 1st March for each forecast experiment and ensemble member. The results exhibit a negligible discrepancy with the analysis of the JJA time series (not shown), which accounts for forecast drift and seasonality during the transitional period."

3. Title revision. Please revise the title for concision and to improve compliance with GMD conventions.

- The current title contains redundant phrasing; please simplify it.

- Please include the relevant model name(s) and version number(s) (e.g., the seasonal forecast system and land model), consistent with GMD guidance for model evaluation papers and related categories.

➔ *We change the title of this paper to "**Implementation of a multi-layer snow scheme in the GloSea6 seasonal forecast system: Impacts on land–atmosphere interactions and climatological biases**".*

4. Code and data availability. Please ensure that materials currently listed as "available upon request" are either deposited in a suitable permanent archive or replaced by a fully reproducible access pathway.

This should include explicit versioning, retrieval scripts/queries, and a clear description of any licensing constraints. If open release is restricted for any component, please describe how confidential or controlled access can be provided to others upon request, consistent with journal policy.

- ➔ *We have updated the "Data availability" and "Code availability". Timely-filtered ESA CCI SM product previously listed as "available upon request" are being prepared for deposition on Zenodo to provide a fully reproducible access pathway. In contrast, the GloSea5 and G6single retrospective forecasts are not publicly hosted, so that this is informed in the "Data availability".*

5. Figure accessibility and color-vision-deficiency (CVD) robustness. Several figures are visually crowded, with too many lines and markers, which reduces readability. In addition, some key figures rely on color combinations that may be difficult for readers with common forms of color-vision deficiency (notably red–green distinctions in line plots). Because these figures convey central results, please revise them so that the information remains clear without reliance on color cues (e.g., by reducing visual complexity and using line styles, marker shapes, and adequate luminance contrast).

- ➔ *We have revised Figures 1 and 2 to ensure accessibility for readers with color-vision deficiency (CVD). We replaced red-green color scales with CVD-robust palettes (e.g., blue-red) and used distinct line styles (solid vs. dashed) and marker shapes to distinguish experiments without relying solely on color.*