

Response to Referee 1

We thank the referee for the careful review and positive assessment of the study. The comments help us clarify the benchmark and its interpretation. We respond below to each point and indicate the corresponding revisions.

General comments

1. Role of ERA5-Land, MERRA-2 and JRA-55

The current wording may indeed blur the role of the reanalysis products. ERA5-Land, MERRA-2 and JRA-55 are evaluated alongside CMIP6 as comparison products. They are not snow-state references or correction products. ERA5-Land is used separately only as the reference for the forcing diagnostics. We will revise the text to make these roles explicit.

2. Predominantly positive CMIP6 snow biases

The predominance of positive snow biases merits further discussion. In Sect. 5.2, we will consider cold-season temperature bias, snowfall amount and partitioning, and snow-model processes affecting persistence, compaction and melt. We will also make clear that a common positive bias does not necessarily reflect a common mechanism, particularly because forcing signatures differ among models.

3. Absolute versus relative snow-depth bias

Large absolute SD biases may partly reflect the spatial distribution of reference snow depth. We will therefore evaluate relative SD bias as a complementary sensitivity analysis. Because percentage bias becomes unstable where reference SD approaches zero, the calculation will be restricted to sufficiently snow-covered locations and tested for sensitivity to this criterion.

Specific comments

L71: "Spatial support"

We will replace "spatial support" with clearer wording, such as "spatial resolution and grid scale."

L76-77

We will state that the benchmark includes 21 CMIP6 historical simulations and three reanalysis products, and clarify the model-selection criteria.

L76

CMIP6 will be defined at first mention as the Coupled Model Intercomparison Project Phase 6.

L93-95

We will distinguish the benchmark from its reference datasets more clearly. The benchmark is the evaluation framework, and the MODIS-derived snow-cover and CHE snow-depth fields are the paired reference datasets used within it.

L109-111

We will reduce acronym use in the prose. SNC will be retained because it is used consistently in the benchmark and corresponds to the CMIP6 snow-cover variable. Elsewhere, we will use "air temperature", "precipitation" and "snowfall" in place of TAS, PR and PRSN where possible.

L111

We will define SNW more clearly and simplify its presentation because it is not an independent benchmark variable in the main analysis.

L114-119

All 21 models are included in the main snow-cover and snow-depth benchmark, whereas the forcing-signature analysis uses the seven models with the required forcing fields. We will make this distinction explicit.

L134

We will distinguish the common 2002-2014 benchmark period from the full temporal coverage of the source datasets in Table 1.

L213-214

We will clarify that the 10% SNC and 0.05 m SD values are diagnostic thresholds rather than universal physical thresholds, and provide sensitivity information for alternative choices.

L292-294

We will recheck the unusually large IPSL-CM6A-LR SD bias, including the units, missing values, masking, aggregation and spatial distribution. If confirmed, we will report it more clearly as an outlier and avoid assigning a mechanism that the available diagnostics do not support.

Figure 5

We will apply a common SD-bias colour scale across the reanalysis maps.

L349-351

We will revise the wording to state that, for many models, positive bias begins in November-December and persists into April.

Sections 4.3 and 4.4

We agree that placing the CMIP6 spatial co-organization analysis before the reanalysis comparison will give the Results a clearer progression.

Figure 8b

We will enlarge the arrows.

Figure 9

We will identify the seven-model forcing subset in the caption and model summary information.

Sect. 5.1

We will substantially shorten this section and remove material that repeats the Introduction.

L459-462

We will state more clearly how colder conditions, snowfall amount and partitioning, and snow compaction or melt can affect simulated snow cover and depth. We will retain these as possible contributors rather than formal attribution.

L538-541

We will quantify representative passive-microwave/CHE snow-depth uncertainties and compare them with the magnitude of the model biases. We will distinguish very large biases from differences of only a few centimetres.