Review for the manuscript: "On the reliability of seasonal snow forecasts" by Vorobeva E., Orsolini Y., de Rosnay P., Day J., Senan R., Decremer D., Vitart F.

The study aims to evaluate the reliability of seasonal forecasts of SWE in the Northern Hemisphere produced by the ECMWF model using two reference datasets. In addition to the general interest of the topic to a wide range of audience, the difference of forecast reliability between regions and verification datasets brings a specific added value to the understanding and evaluation of this type of forecasts considering variability, model representation and uncertainties in the observations or reference datasets.

Considering the general target audience of the manuscript and journal, I've found section 2.1 very well presented and informative for readers not much familiar with ensemble prediction skill evaluation.

A: We thank the anonymous reviewer for taking their time to review our manuscript. Please find a detailed reply to your comments below.

I suggest mentioning here that examples of the reliability diagram for different categories are available in the supplement material as part of the results discussion. I find it very informative to "see" the association of each category to an example diagram.

A: We have added a sentence in section 2.1 that mentions that examples of diagrams for different categories are available later in the text. Following a suggestion expressed by R1, Figures with individual reliability diagrams have been moved into the main text.

It is not clear why this particular set of snow hindcasts were used ? Is it because it's a configuration close to a future evolution of SEAS5 ? Considering that SEAS5 is an operational product, and widely used, why not also include the evaluation of SEAS5, e.g. the results of Fig 3 but for SEAS5, just as supplement, but that some statement could be made like : the hindcats used in this study have a similar reliability as SEAS5 ? I understand if adding this extra analysis is too much work, and it's not essential for the paper.

A: Thank you for the good comment. Indeed, the set of snow hindcasts shown in the paper is close to a future evolution of the current SEAS5. The current operational forecasts system is in use since 2017 and is based on IFS cycle 43r1 where initialization of the snow fields is performed using 43r1 land surface model driven by ERA-Interim (like ERA-Interim land, Description of SEAS5-v20171101 C3S contribution - Copernicus Knowledge Base - ECMWF Confluence Wiki). ERA-Interim was succeeded by ERA5 in 2019 (Global reanalysis: goodbye ERA-Interim, hello ERA5 | ECMWF). As mentioned in the manuscript, this study is carried out in the framework of the CERISE (CopERnIcus climate change Service Evolution) project that provides us with access to the hindcasts produced using the latest, state-of-the-art ECMWF model configuration, more recent

than SEAS5. We use this opportunity to show how reliable snow hindcasts are at their current state. Carrying out evaluation of hindcasts produced by SEAS5 would require data acquisition, re-calculation, new round of analysis and visualization. We believe such comparison should be left for future studies that focus either on the improvements in snow forecast reliability between different generations of SEAS, similar to Manzanas et al. (2022) or to a multi-model comparison of seasonal forecast models under the CDS.

It is mentioned that the forecast model includes a new multi-layer snow scheme, but it is initialized from ERA5 (has a single snow layer). How is the mapping done from ERA5 snow to the initial conditions of the multilayer snow scheme?

A: The forecast model uses a multi-layer (ML) snow scheme, whereas ERA5 provides a bulk single-layer (SL) snow analysis. Therefore, an initialization procedure is applied at the start of each forecast to reconstruct vertically resolved snow profiles from the ERA5 bulk snow state. Using the analyzed snow depth/SWE, skin temperature, and upper-soil temperature, the snowpack is diagnostically partitioned into multiple layers, while conserving total snow mass, depth, and cold content. Vertical profiles of snow temperature, density, and liquid water content are then generated consistently with the surface and soil thermal state, following the approach described in section 4.1 of Arduini et al. (2019). We have now clarified this in the manuscript and added a reference.

In fig.4 and 5 it would be relevant to the results discussion to add the hindcasts SWE anomalies (distribution in fig 4 and timeseries in fig 5), as it would further document the previous results of the reliability.

A: Thank you for the comment. Figures 4 and 5 were merged into new figure 7, and hindcast SWE anomalies were added as suggested.

Line 112: “based on the i.i.d. bootstrap” it’s not clear the meaning of i.i.d. at this stage

A: We have clarified abbreviation as “i.i.d. (Independent, Identical Distributed Data) bootstrap”.