**Referee 2:**

The added value of initialization for long (multidecadal) climate outlooks remains unclear, so this analysis of 30-year initialized predictions using the EC-Earth3 model will be of considerable interest to the decadal prediction community. The manuscript is clear and concise, and the results are intriguing. I recommend this be published after some minor issues are addressed. Most importantly, I think an additional analysis is needed (perhaps in supplemental) that shows how skill varies with lead time when a common verification window is used.

**Author Response:** We thank the reviewer for taking time and providing very constructive feedback on our study. We have revised the manuscript by taking into account all the comments and suggestions provided by the referee. Below we provide a point-by-point response to each of the referee comments.

Minor Comments

L48-60: A recent study suggests that initialized internal variability could constrain surface climate variability for multiple decades (Deser et al., 2025, https://doi.org/10.1007/s00382-024-07553-z). It may be worth including such potential predictability examples in this intro section.

**Author Response:** Added a new sentence citing Deser et al. (2025) as: "Recent work by Deser et al. (2025) with idealised experiments further suggests that the initialized internal variability can constrain surface climate variability for multiple decades, highlighting the potential for extended predictability when key modes of variability are properly captured."

L62: Unclear what "model dependence in previous results" refers to exactly

**Author Response:** We agree that it was not fully clear. It was meant for model dependence compared to the model used by Düsterhus and Brune (2023). We modified the sentence to indicate this as: "Motivated by the prospect of multi-decadal predictability from model initialization and with the aim to understand the model dependence in previous study (Düsterhus and Brune, 2023), in this study we performed a set of initialized climate predictions with the EC-Earth3 model with a forecast horizon of 30 years".

L112: If hindcasts are 30 years long, why not examine FY21-30 and FY11-30, for consistency?

**Author Response:** The predictions are initialized from November and run for 30 years, however the final prediction year runs until October. In order to analyze annual means (from January to December), we have one less year for the analysis. We agree with the reviewer that this information was not written explicitly in the manuscript. We have now clarified in the revised version of the manuscript about the availability of 29 full prediction years for analysis purposes.

L121: Should mention that observed anomalies are computed in similar fashion (observed climatology matching months/years used for hindcast climatology).

**Author Response:** Agree. We added a new sentence indicating this as: "The observed and historical simulation climatologies are computed for the same temporal period used to define the hindcast climatology."

Fig. 1: Caption should state the field being analyzed. Methods should clarify how statistical significance is quantified.

**Author Response:** Agree. Added near-surface air temperature in the figure caption. Also added information on how the statistical significance is computed in the methods section.

L137: Authors have an opportunity here to directly address skill sensitivities to sampling and system design choices, so why not do that in Fig. S1? A first FY1-10 ACC plot could subsample the previous system, selecting only start dates used in Fig. 1 (this would isolate the effects of changes in initialization, I believe). A second FY1-10 ACC plot could show skill when all start dates are included from the previous system (which would reveal the effect of reduced hindcast sampling on skill).

**Author Response:** We have now included in supplementary materials figures using all startdates and every 5th start-date from the 10 year prediction system. We have also added some discussion related to this at lines 145-150 in the revised manuscript.

L142: I think this should read "1981-2020 for FY21-29"? As noted above, the use of 9-year average (instead of 10-year average) for FY21-29 introduces another avoidable difference. The differences in verification window are likely important, so why not include a supplemental figure that shows skill for a common verification window (1981-2020 for all leads)? Otherwise, the conclusions that can be drawn from this comparison are very limited.

**Author Response:** Thank you. The description of years is corrected now. We have now included a new supplementary figure (Figure S2) and also added related text in the revised manuscript.

L152: I think the conclusions about impact of initialization are somewhat shaky given the discrepancies in verification window (see above).

**Author Response:**

L187: I would delete "as it was assumed in the past", because I don't consider fast convergence to be a generally held assumption, particularly for the North Atlantic.

**Author Response:** Agree. The text is deleted.

L213-215: Repetitive with introduction.

**Author Response:** Agree, the text is now removed from the revised version of the manuscript.

Fig. 4: Is it surprising that the AMOC climatology in DP_30yrs is almost identical to that from the DP (10-year), even at short leads, despite the different initialization methods?

**Author Response:** Both DP_30yrs and DP are initialized exactly in the same way. To clarify this we have added the AMOC climatology of the previous DP system. Comparing AMOC45 in DP_30yrs and DP shows that the climatology is well constrained even with fewer start dates. Comparing with the previous system shows that the updated initial conditions have an impact initially but overall the current system suffers from the same problems and the AMOC climatology evolution is the same. We have added a short paragraph in the text on this.