We thank the reviewers for their careful reading of the paper and their positive comments. We welcome the reviewer's suggestions and have considered each of these in the revised version of the manuscript. We have enumerated each comment/suggestion followed by our reply, and provide a corrected version of the manuscript with changed/added text in red and removed text struck out. The line numbers referred to in our reply correspond to the marked version of the paper.

Reply to Reviewer #2:

1. Abstract L4: sounds like the hindcasts were started in 1961 and then run continuously until present.

   Corrected

2. P2 L31: "at the end" maybe be more specific?

   Corrected

3. P5 L5: "augmented" replace with more specific phrase, please.

   "augmented with" was changed to "merged with weekly"

4. P5 L16-18: I don't understand what initialization through response means. Are the carbon cycle components running during the assimilation phase and the initial state for each hindcast corresponds to their states at that point in the assimilation? Please clarify in the manuscript.

   This has been clarified in the second to last paragraph of section 3 "Forcing, initialization and ensemble generation".

5. P6 L9: Suddenly a subscript "e" appears in some equations, what is that? Not defining this makes it hard to follow the rest of the derivations.

   The text below equation (2) was modified to introduce the subscript e.

6. P8 L9: "secular" is a strange timescale (is it an astronomy term?), do you mean centennial?

   The expression "secular" refers to the long-term non-periodic variation of time series. To avoid confusion, we changed "secular" to "centennial or longer".

7. P9 L5: Why are you excluding the Arctic? The skill seems to come from initialization, but you'd think that it doesn't vary much under the ice, so that's a contradiction.

   The reviewer is correct that the potential predictability variance fraction in the Arctic mainly results from initialization, as seen in Fig. 3d-f. We have

changed the text to reflect this. As in sectors of the Southern Ocean and in the WSPNA and Labrador Sea regions, the uninitialized simulations display positive trends in the Arctic (Fig. 4c), whereas ORAS5 does not (Fig. 4b). These negative trends in ORAS5 are imprinted in the hindcasts (Fig. 4d) during the initialization process, which affect the predictable signal. As suggested by the reviewer, there is not much internal variability in the Arctic, therefore a large portion of the total variance result from the predictable variance derived from these negative trends. As a consequence, the potentially predictable variance fraction in the Arctic is high (Fig. 3a-c), and is largely attributed to initialization (Fig. 3d-f).

8. P9 L7: What is the relevance of mentioning the strong linear trends? What is the impact of them?

   This has been clarified in P10L1-3. The negative linear trends, which are strong in sectors of the Southern Ocean, the WSPNA and the Labrador Sea, result from the ocean reanalysis product used to initialize the forecasts. As explained in item 7 above, these trends drive the predictable variance attributed to initialization.

9. P9 L33: How can the errors be fully attributed to initialization and then in the same sentence also attributed to the response to external forcing?

   This is because initialization can affect the model response to external forcing. Note that the response to external forcing in the hindcasts is generally different to that in the uninitialized simulations (Eq. 1). We now emphasize this in P10L30-32. For the WSPNA and Labrador Sea regions, the erroneous trends in SST hindcasts are imprinted by the ocean reanalyses used for initialization. We show that these errors are "fully" attributed to initialization because the correlation skill satisfies $r_{XY} = r_i$ (see Fig. 5a-f). This is consequence of a mismatch in the forced responses of hindcasts and uninitialized simulations, since we found that $r_{YU} < 0$ which implies that $r_u = 0$ and so $r_{XY} = r_i$ (see Eqs. A16 and A17). In addition to the analysis provided in the text, we cite Sospedra-Alfonso and Boer (GRL 47, 2020) where this is discussed in detail.

10. P10 L35: "can potentially" - I feel you should have a little more certainty than this about the impacts of the Atlantic SST problem. Perhaps if sections 6 & 7 were swapped, you could discuss here what has been shown for skill over land.

    Although many studies such as those cited in section 6 have established a relationship between the North Atlantic SST and surface climate elsewhere, we prefer to be cautious as we have not fully established these relationships for CanESM5. We do expect this to be the case for CanESM5 and have therefore

changed "can potentially" to "are likely to". Also, we prefer to keep the order of sections 6 and 7, as we would like to assess decadal prediction skill in the ocean first (section 6), which is expected to contribute to the decadal prediction skill on land (section 7).

11. P11 L32: Perhaps it is worth having a note here that there are several papers link the Sahel precip to the AMV and that this is discussed at the end of Section 8 in the paper.

    Following the reviewer's suggestion, we have made a reference in the text to the discussion and bibliography cited at the end section 8

12. P12 L13: I think the description of the volcanic experiments needs to have a few more details and be placed in the methodology section of the paper.

    A brief description of the volcanic experiments has been included in the last paragraph of section 3 "Forcing, initialization and ensemble generation".

13. P12 L16: "volcanic forcing seems to be" I think the evidence presented is only strong enough to say "volcanic forcing could be". Initialisation seems to be quite important too!

    We agree with the reviewer and have changed the text to reflect the evidence presented here for the contribution of initialization and possibly volcanic forcing to precipitation skill.

14. P15 L14: "time pan"

    Corrected

15. P17 L11: "Strong warming" Where does the climate sensitivity of CanESM5 lie compared to CMIP5/6 estimates of the real world probable range?

    Meehl et al. (Sci. Adv. 2020) report that Earth system models participating in CMIP6 (CMIP5) have equilibrium climate sensitivity (ECS) ranging from 1.8 K to 5.6 K (2.1 K to 4.7 K), and transient climate response (TCR) ranging from 1.3 K to 3.0 K (1.1 K to 2.5 K). CanESM5 is towards the highest end with ECS = 5.6 K and TCR = 2.7 K. Historical warming trends in CanESM5 are also strong, as seen in Figs. 4c and 10c of the paper, and as reported in Figs. 25a and 26 by Swart et al (Geosci. Model Dev., 12, 4823–4873, 2019). For context, this information has been included in the first paragraph of section 7 "Predictability and skill of surface climate on land".

16. Figure 1: This color scale looks problematic for color blind people (https://www.color-blindness.com/coblis-color-blindness-simulator/). Additionally, if the color scale for Fig 2 was used here, it would be easy to flick between the figures to see how much of the potential predictability has been realised (this applies to the other Figures showing the same thing for other variables).

We followed the reviewer's suggestion and changed Figs. 1, 3, 4, 9, 10, 12, 14 and 15. Note that the results are the same and only the appearance of the figures have changed.

17. Figure 10: (b) Year 1 and (d) Year 2 look very similar. Year 1 is not shown in Fig 9, so perhaps use (b) Year 2 and (d) Years 2-5? This would help with backing up the conclusions in the text too by making the trends later in the forecast clearer.

    We changed Fig. 10 following the reviewer's suggestion.

18. Figure 14: Why is this years 2-4 and not years 2-5 like in other figures in the paper?

    We look at years 2-4 since the precipitation response to volcanic eruptions is expected to peak during this time window. This is noted in P13L17.

19. Figure 16: This is similar to Fig 15, would years 6-9 be more interesting?

    We are not sure about this question. Figure 16 shows the dependence of correlation skill on ensemble size and Fig. 15 shows maps of noise-to-predictable variance ratio for hindcasts and uninitialized simulations. Perhaps the reviewer is referring to Figs. 16 and 17. If so, note that the impact of initialization is much reduced for Year 6-9 (Fig. 13 f), which would make the discussion less meaningful. As pointed out in P14L22-35, the impact of initialization for the Year 2 annual precipitation forecast in Central South West Asia is detected for $\gtrsim 15$ ensemble members (Fig. 16d), whereas $\gtrsim 35$ are required for Year 2-5 forecasts (Fig. 17d).