

Geosci. Model Dev. Discuss., referee comment RC2  
<https://doi.org/10.5194/gmd-2022-60-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on gmd-2022-60

Anonymous Referee #2

---

Referee comment on "The Seasonal-to-Multiyear Large Ensemble (SMYLE) prediction system using the Community Earth System Model version 2" by Stephen G. Yeager et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-60-RC2>, 2022

---

This submission describes an extensive set of hindcasts from the CESM2 model that enable the performance of initialized predictions in the relatively unexamined multi-year range (out to 24 months in this case) to be extensively explored. Notably, performance over a broad range of Earth system components (atmosphere, sea ice, ocean and land including biogeochemistry) is addressed. The paper is very well organized and written, and criticisms are limited mainly to relatively minor details of description and presentation. Exceptions are items 7 and 17 below, which will require modest additional computation if the authors concur that acting on these recommendations will improve the paper. Overall however the authors are to be congratulated for this interesting and compelling documentation of SMYLE.

### Main comments

1) At line 47, suggest replacing "at least 10-years duration" with "up to 10-years duration" because some operational "decadal" systems have a 5-year range, but none that I'm aware of run for >10 years.

2) Suggest additionally referencing Boer et al. <https://doi.org/10.1007/s00382-013-1705-0> and Chikamoto et al. <https://doi.org/10.1038/s41598-017-06869-7> in the sentence starting on line 66, possibly as follows: "...volcanic activity (Hermanson et al. 2020), greenhouse gas forcing (Boer et al. 2013), or some combination thereof (Chikamoto et al. 2017)."

3) At line 72 should also reference Ilyina et al. <https://doi.org/10.1029/2020GL090695>

4) Line 128 states that forcing is applied cyclically from 1901-1920 to equilibrate the land state. Please go into a bit more detail about the total length of time this cyclic forcing was applied, in relation to expected equilibration times of land variables such as vegetation and soil carbon.

5) It's stated that the hindcasts cover 1970-2019. Presumably this is the period covered by the initialization times, and not the simulations themselves which would extend into 2021? If so please be explicit that 1970-2019 spans the initialization times.

6) Regarding "Potentially useful prediction skill ( $ACC > 0.5$ ) is seen for land precipitation over the southwestern US in DJF and MAM (lead month 1)" at lines 208-209, this really should say "southwestern North America" considering that the only DJF grid boxes  $> 0.5$  are in Mexico.

7) Regarding "A more rigorous analysis is needed to definitively demonstrate that SMYLE skill differences from DPLE are statistically significant and not likely explainable by chance" (lines 226-227), this could be done relatively easily by applying the random walk methodology of DelSole and Tippett, <https://doi.org/10.1175/MWR-D-15-0218.1>, where differences either in the anomaly pattern correlation or the RMSE between the 50 pairs of November-initialized hindcasts could be used as the basis for comparison.

8) At line 178, please provide a rationale for regridding to a 5x5 or 3x3 degree grid. (Also a small point, but I'm not sure that regridding to a coarser grid qualifies as "interpolation".)

9) Below line 360 it would be appropriate to reference Butler et al. <https://doi.org/10.1002/qj.2743> in relation to the influence of lid height on skill in forecasting the NAO. (For example could append as "...relative to these baseline SMYLE results, although a robust connection between atmospheric vertical resolution and NAO skill has not been demonstrated (Butler et al. 2015).")

10) Please replot Figs. 8d-i using the tick mark values in Fig. 9 which are better aligned with the experiment.

11) Are there any evident explanations or hypotheses for the strong seasonal dependence of skill in Figs. 8d-i, e.g. high SE US shelf NPP ACC in JJA and SON, and low CA current NPP ACC in DJF?

12) Should mention in the captions for Figs. 8-9 that shading and filled symbols indicate statistical significance.

13) The OceanSODA-ETHZ aragonite saturation dataset covers 1985 to 2018 according to Gregor and Gruber (2021), so presumably the skill results in Fig. 9 are specific to this period? Or does the verification period leave out the years before 1990 which are much more uncertain according to those authors? Please be explicit about this and any other deviations of verification periods from the 1970-2019 period covered by SMYLE.

14) In the captions to Figs. 10 and 11 suggest removing "(see text for details)" since the text doesn't provide any significant additional detail.

15) In Fig. 13 the lead times in the legends of the plots disagree with the lead times indicated in the caption.

16) Relating to Figs. 12 and 13, it would be interesting to have a sense of how the correlation and nRMSE values shown compare to values based on comparing OBS to FOSI.

17) Figure 14 shows JJASON (NH) and DJFMMA (SH) cyclone track densities regressed against annual mean Nino3.4 index. However, because ENSO typically peaks around December and frequently changes phase between about April and August, annual mean Nino3.4 is not a very good indicator of ENSO activity. In addition, this procedure introduces a seasonal disconnect in that January Nino3.4 is presumed to influence TC activity in the following November (for example). Suggest instead regressing JJASON track densities against JJASON Nino3.4, and DJFMAM track densities against DJFMAM Nino3.4, or else better justifying the original choice made. (Also please be explicit in the caption to Fig. 14 what is the timing of the Nino3.4 index.)

18) Tables 1 and S1 along with Fig. 15 imply that JJASON and DJFMAM TC predictions are made at 19-month lead time. However, for JJASON this implies initialization on 1 Nov, which in turn implies prediction of JJASO (not JJASON) at 19-month lead by the 24-month hindcast, and similarly for DJFMAM. Although this is a small point it should briefly be acknowledged somewhere (similarly to the 22-mon lead Nino3.4 forecasts in the caption to Fig. 4).

19) Regarding the RMSE scores, "RMSE" isn't defined anywhere, and when introducing RMSE in the text should briefly comment on the use of normalized RMSE and introduce the nRMSE notation. Also, is nRMSE defined such that predictions of climatology (zero anomaly) will yield values of 1? If so then briefly mentioning this will help the reader appreciate that nRMSE values  $<1$  indicate that the prediction is more skillful than a climatological prediction.

Technical corrections:

line 203: central America -> Central America

line 207: SAT hasn't been defined

lines 270 and 275: "1998" -> "1997" (as year of a strong El Nino)

line 361: Quasi-biennial -> Quasi-Biennial

line 510: Figs. S1, S2 -> Figs. S4, S5

line 520 vs 185 vs 591: is it "best track", "Best Track" or "BestTrack"?

line 528: should the 2 in  $kt^2$  be a superscript?

line 556: suggest "multi-year skill" -> "multi-year skill or potential skill"

line 561: suggest to remove "obviously"

line 923: ACC map gross -> ACC map for gross