Reviewer 1

The manuscript "Comparing the seasonal predictability of the Tropical Pacific variability in EC-Earth3 at two horizontal resolutions" by Carréric et al. explores the forecast skills in ENSO in the EC-Earth model in two different horizontal resolutions. They explore what biases the models have that could explain the limitations in forecast skill. This study is of interest to the research community. I find the manuscript generally well worked out, but I do have a number of comments that should be addressed before publication. Although none of my comments require any major work, I do find some important discussions need to be addressed. I therefore recommend major revisions.

Thank you for your review and constructive comments. We have addressed them and provide a detailed response to each of them below (in violet). We have included the changes made to the improved manuscript.

Main points:

(*) Higher resolution = better: the study suggests that HR simulations are better, and it is indeed clear from the analysis that the forecast skill of the HR simulations is better than that of the SR simulation. The study also points out that both model simulations have similar biases and that these biases are likely the cause of the degraded forecast skills. The study also points out that the higher resolution is much more expensive to run. So, is the HR model really better, all things considered?

It is worth discussing that the lower resolution may not be as good, but it is cheaper and may therefore allow faster model development by improving model parameterisations. It will also allow to run more ensembles in different configurations, which could reduce model biases.

We agree with the reviewer that we can discuss the advantages of standard resolution further in the Discussion section. We have amended the penultimate paragraph of the section as follows:

“Conversely, standard resolution simulations are more economical to produce and store, which facilitates and accelerates the production and analysis of long and large ensembles of simulations. These can be leveraged to test model developments, refine the model tuning, produce sensitivity simulations designed to address specific scientific questions, constrain more accurately the predictable signals, and increase model complexity at a lower cost, all with the aim of improving our understanding of the climate system. The choice of modelling strategy remains dependent on the scientific question and the intended application, as all strategies are complementary (Bordoni et al., 2025, Baldissera Pacchetti et al. 2024, Katzav and Parker, 2015).”

We also reinforced our statement in the abstract, specifying that this improvement in high resolution may only be valid in certain specific configurations: “Small but statistically significant improvements are found in HR with respect to SR for predicting ENSO, skill improvements that cannot be generalized to other regions and initialization times.”

We also checked throughout the text where we used the term “better” in connection with the High-Resolution configuration. We used it mainly in the analysis of the Atlantic Niño teleconnection with the tropical Pacific. In order to remove any uncertainty regarding the use of the term “better,” we replaced it with “more realistic.”

Baldissera Pacchetti, M., J. Jebeile, and E. Thompson, 2024: For a Pluralism of Climate Modeling Strategies. *Bull. Amer. Meteor. Soc.*, **105**, E1350–E1364, https://doi.org/10.1175/BAMS-D-23-0169.1.

Bordoni, S., Kang, S.M., Shaw, T.A. *et al.* The futures of climate modeling. *npj Clim Atmos Sci* **8**, 99 (2025). https://doi.org/10.1038/s41612-025-00955-8

Katzav, J., Parker, W.S. The future of climate modeling. *Climatic Change* **132**, 475–487 (2015). https://doi.org/10.1007/s10584-015-1435-x

(*) Causality: The authors argue several times in the manuscript that certain biases in the model "cause" the forecast biases (e.g. line 440). While the statements are plausible, no evidence is presented about the causality. The authors should not assume causality, when no analysis or experiments are presented that allow them to do so.

We agreed with the reviewer that no evidence of causality is provided in the paper and remains beyond the scope of this research. We have tried not to formulate causality in an assertive manner, but simply to put forward possible assumptions based in particular on other research that we cite. However, as the reviewer points out, certain passages escaped us and suggest causality. We have modified the following sentences accordingly:

Line 324: It is less pronounced in EC-Earth3-HR, which may be related to the improved mean state, which reduces the impact of the coupled feedback bias due to the weak coupling. → It is less pronounced in EC-Earth3-HR, which may be related to the improved mean state, which would reduce the impact of the coupled feedback bias due to the weak coupling.

Line 329: These ENSO-related errors, which induce low SST predictive skill in the WEP regions, are linked to the erroneous spatial simulation of ENSO, which is itself linked to the mean state representation. EC-Earth3-SR shows particularly poor predictive skill in the WEP region compared to EC-Earth3-HR due to higher ENSO-related errors and mean state bias. Overall, a weak air-sea coupling, weaker in EC-Earth3-SR, prevents the model from simulating the correct ENSO development. → These ENSO-related errors, which induce low SST predictive skill in the WEP regions, are linked to the erroneous spatial simulation of ENSO, which is itself likely linked to the biased mean state representation. EC-Earth3-SR shows particularly poor predictive skill in the WEP region compared to EC-Earth3-HR due to higher ENSO-related errors and possibly due to a mean state bias. Overall, a weak air-sea coupling, weaker in EC-Earth3-SR, may prevent the model from simulating the correct ENSO development.

Line 425: The poorest skill in EC-Earth3-SR compared to EC-Earth3-HR in the WEP region seems linked to larger mean state biases, and in particular the "cold tongue bias", a common bias in GCMs (Bellenger et al., 2014; Guilyardi et al., 2020), which leads to larger ENSO-related errors. → The poorer skill in EC-Earth3-SR compared to EC-Earth3-HR in the WEP region, associated with larger ENSO-related errors, also seems linked to larger mean state biases, and in particular the "cold tongue bias", a common bias in GCMs (Bellenger et al., 2014; Guilyardi et al., 2020).

Lines 440: Finally, the stronger decrease of skill in EC-Earth3-SR is also related to a weaker air-sea coupling in EC-Earth3-SR than in EC-Earth3-HR, which prevents the model from simulating the correct ENSO development. → Finally, the stronger decrease of skill in EC-Earth3-SR is also likely related to a weaker air-sea coupling in EC-Earth3-SR than in EC-Earth3-HR, which prevents the model from simulating the correct ENSO development.

other points (in order as they appear in the text):

----

line 52 "... and that take overly low values ...": Not clear what this means.

We rephrased it as follows: “positive SST anomalies [...] whose amplitudes are underestimated compared to observations in the cold tongue region.”

----

line 87-90 "... only differ in the horizontal grid spacing... Differences between systems can be attributed to the change in process parametrization adapted for each resolution,": This is contradicting itself. The two models are not just different in resolution, but also differ in some of the parameterisations.

Indeed, thank you for noting it. We rephrased this sentence: “We use a Standard Resolution (SR) configuration based on an eddy-parametrised ocean, and a High Resolution (HR) version based on an eddy-permitting ocean, their atmospheric configurations being set up to match the respective ocean resolutions as closely as possible. Both systems are initialized following exactly the same strategy, so that forecast differences can be directly attributed to the change resolution, which involves both changes in grid spacing and parameterization choices.”

----

line 154 "... with g10 bias corrections ...": What is that?

The g10 bias correction is a bathythermograph correction scheme applied to the version of the EN4 dataset used in this study. Different bias correction schemes included in the EN4 datasets are published and the MetOffice recommends to specify which version is used (see https://www.metoffice.gov.uk/hadobs/en4/terms_and_conditions.html).

We rephrased the sentence to make it clearer: “the EN4 v4.2.1 ocean objective analysis (Good et al., 2013 - we used the version with g10 bias correction scheme (Gouretski and Reseghetti, 2010)).”

"Drift correction": Would it make sense to analyze this? Is it different between the two simulations? Less correction for the HR model?

To account for the drift effects on the forecasts, the following figure shows the mean bias (compared to the ESA SST CCI dataset) by lead month for EC-Earth3-SR (blue) and EC-Earth3-HR (orange) for the whole hindcast period (1990-2015). Biases tend to be more important in EC-Earth3-SR upon initialisation and to grow further with lead time. In principle, higher forecast biases would not necessarily imply worse skill, as long as they remain stationary in time, which is our working assumption for both systems. We added the mean bias for two separate periods: 1990-2003 (pink for EC-Earth3-HR and turquoise for EC-Earth3-SR) and 2004-2015 (red for EC-Earth3-HR and black for EC-Earth3-SR) to assess this assumption. For both systems and all ENSO indices, we see some non-stationary features in the forecast biases, which appear to be higher at the longest lead-times for the earlier period. This non-stationary behaviour is introducing some spurious signals in the drift corrected anomalies that are expected to degrade their forecast skill. But since the non-stationary effects are similar in both systems, we can assume that they do not explain the differences in skill. The non-stationary drift and its potential implications for our results are now discussed in the manuscript, in the “Analysis methodology” section and the Figure R1 added in Supplementary material.

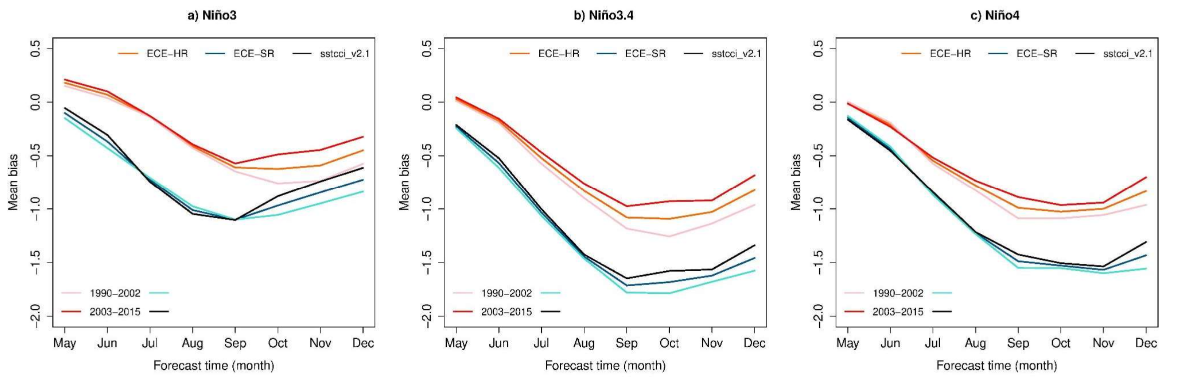

Figure R1: Mean bias of SST in the regions (a) Niño3 (5°S-5°N, 150°W-90°W), (b) Niño3.4 (5°S-5°N, 170°W-120°W) and (c) Niño4 (5°S-5°N, 160°E-150°W) as a function of forecasted month for (blue) EC-Earth3-SR and (orange) EC-Earth3-HR. The reference dataset is ESA SST CCI dataset.

----

figure 1: The labels of the rows could add the lead time in months (e.g., 2,4 and 6 mon).

We added the indication in lead time in the figure.

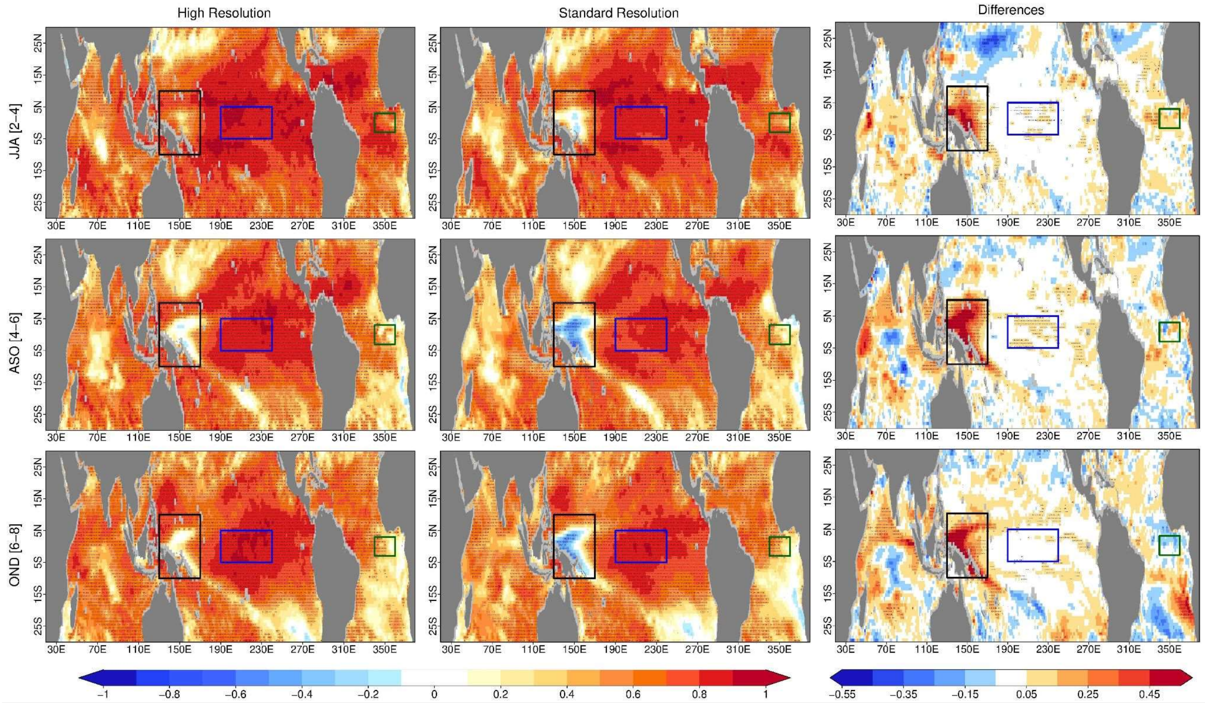

Figure R2: same as Figure 1 in the manuscript but with the indication of the lead time in months between brackets.

Figure 6 discussion: The text seems to suggest that at the start of the simulation all biases are zero? Then, why do the east Pacific SST biases are already present at the start in May?

The present study analysed monthly forecast outputs, the month of May being the entire first month for May initialisation (whose initial state is derived from the end of the 30th of April). Since the time step of the models is 45 and 15 minutes for SR and HR, respectively, the different biases are already developing in the first timesteps. They grow quickly and then appear on average during the first month. This rapid growth in forecast errors after initialisation has been studied in various studies, including one by the BSC on Arctic sea ice extent (Cruz-García et al. 2021) for instance.

Cruz-García, R., Ortega, P., Guemas, V. *et al.* An anatomy of Arctic sea ice forecast biases in the seasonal prediction system with EC-Earth. *Clim Dyn* **56**, 1799–1813 (2021). https://doi.org/10.1007/s00382-020-05560-4

----

line 318 "... during the development of El Niño ...": Not clear how one can see the development of ENSO in Fig. 7.

We agree that the sentence “The ENSO-related winds differ between EC-Earth3-SR and EC-Earth3-HR particularly during the development of El Niño events” is misleading and has been replaced by “The ENSO-related winds differ between EC-Earth3-SR and EC-Earth3-HR as El Niño events evolve with forecast time”. In Figure 7, the onset of the averaged El Niño event has already happened, as it tends to develop during the spring (March-June), even though the diversity of ENSO, which is not taken into account in our definition, leads to greater variability in the actual onset of events (Liu et al. 2025). Figure 7 shows the mean forecasted evolution of a typical El Niño event, leading to the peak season in late fall/early winter, which we only get a glimpse of in our simulation, which ends in December.

Liu, Y., Zhang, W., Jiang, F., Chen, H.-C., Jin, F.-F., & Hu, S. (2025). Diverse timing of El Niño onset linked to preconditioned recharge state and occurrence of westerly wind bursts. *Geophysical Research Letters*, 52, e2024GL113668. https://doi.org/10.1029/2024GL113668

"3.2.4 Summary": This is only the summary of the subsection, not the final summary. This is confusing; better rename.

We rename the section to “Summary of results on prediction skill in the WEP region”.

----

line 333 "The westward shift in ENSO SST and wind zones of influence directly impacts precipitation, ...": It can also be the other way around: The biases in precipitation cause the wind and SST to shift.

We have rewritten this passage to avoid any inference on causality: “Precipitation associated with ENSO events also shows a deficiency in model simulations, as the convergence zone does not move correctly toward the centre of the equatorial Pacific during El Niño events (Fig. 8). ENSO-related precipitation is located too far west reaching the first islands of the Maritime continent, particularly in EC-Earth3-SR.”

---- end -----

Reviewer 2

Review of "Comparing the seasonal predictability of the tropical pacific variability in EC-Earth3 at two horizontal resolutions" by Carreric et al. (2015)

General Assessment:

This manuscript investigates the impact of increased horizontal resolution in the EC-Earth3 model on the seasonal predictability of tropical Pacific sea surface temperatures (SSTs). The authors report an improvement in the SST anomaly correlation coefficients (ACC) in the western Pacific, comparing the standard resolution (SR) and the high resolution (HR) configuration. This enhancement is attributed to a better representation of the mean state and ENSO teleconnections. They also found the improvement of ENSO predictability results from the better presentation of Atlantic teleconnection. The findings are generally sound and contribute meaningfully to the ENSO and seasonal prediction communities. However, several points regarding the interpretation of results and discussion require modification prior to publication.

We thank the reviewer for the constructive comments. We believe we have addressed them and provide a detailed response to each of them below (in purple). We have revised the manuscript accordingly.

Major Comments:

1. **Significance of ENSO Predictability Improvement:** The assertion that ENSO predictability is significantly improved in the HR configuration is not strongly supported by the presented data. In Figure 1, the difference in SST ACC values between the SR and HR configurations in the central Pacific is marginal, ranging between 0.05 and 0.1. As shown in Figure 2, the ENSO index ACC already reaches 0.7 in the SR configuration. The observed increase to 0.75 or 0.8 in the HR configuration does not represent a substantial enhancement in practical forecast utility. Therefore, the conclusions regarding the magnitude of improvement should be tempered.

We acknowledge that the prediction skills of the model, in both configurations, are already high, as expected, in the Tropical Pacific. However, EC-Earth3-HR shows higher prediction skills than EC-Earth3-SR in Niño3 and Niño3.4 regions and, even if apparently small, these differences are statistically significant for most forecast years. To estimate the significance of the difference between the skill of the two model configurations, we used a two-sided test, at the 95 % confidence level, following Siegert et al., 2017. The test was specifically developed ”to assess correlation differences when the competing forecasting systems are strongly correlated with one another”.

Only in the Niño4 region, the westward most region where the standard ENSO indices are defined, the skill differences are not statistically better in EC-Earth3-HR than in EC-Earth3-SR throughout the forecast, an interesting result in itself, which led to a more in-depth analysis of the WEP region.

Siegert, S., O. Bellprat, M. Ménégoz, D. B. Stephenson, and F. J. Doblas-Reyes, 2017: Detecting Improvements in Forecast Correlation Skill: Statistical Testing and Power Analysis. *Mon. Wea. Rev.*, **145**, 437–450, https://doi.org/10.1175/MWR-D-16-0037.1.

2. **Discussion of the Spring Predictability Barrier:** The manuscript should more explicitly address the influence of the spring predictability barrier, a known limitation in ENSO forecasting where skill scores typically drop below 0.6 after spring. The current analysis utilizes hindcast data initialized in May, which crosses this barrier. Although lines 470-475 mention that November-initialized hindcasts show less noticeable differences in ACC between resolutions, it is recommended that these results be presented, particularly in a format similar to Figure 2. This would allow for a direct assessment of whether increasing model resolution mitigates the impact of the spring predictability barrier.

The Spring Predictability Barrier (SPB) is indeed a major challenge in ENSO forecasting. Our study analyses the results of forecast systems initialised in May, which, as noted by the reviewer, is less relevant because initialization happens within the barrier, not upstream of it.

The figure below shows the ACC for the equivalent forecast simulations initialized in November in the (a) Niño3 (5°S-5°N, 150°W-90°W), (b) Niño3.4 (5°S-5°N, 170°W-120°W) and (c) Niño4 (5°S-5°N, 160°E-150°W) regions as a function of forecasted month for (blue) EC-Earth3-SR and (orange) EC-Earth3-HR. The figure is analogous to Figure 2, but for November initialization (with the new colours as mentioned below).

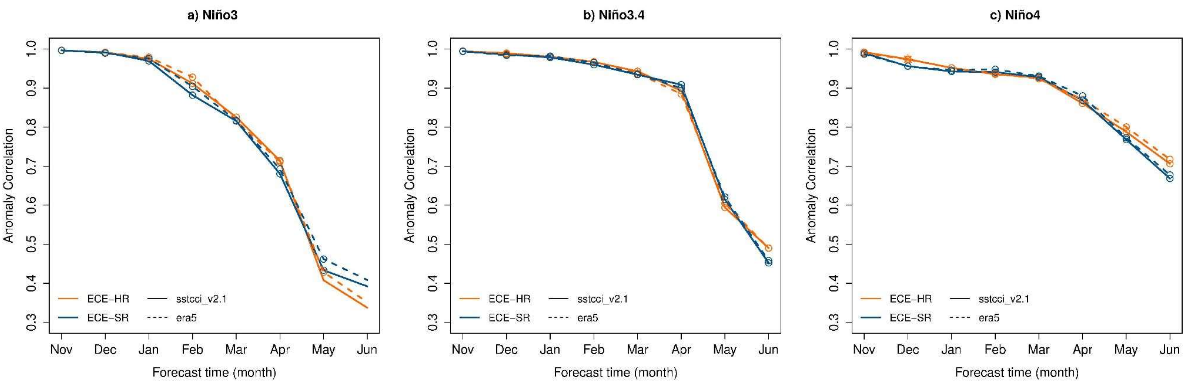

Figure R3: Same as Figure 2 of the manuscript but for an initialisation of the forecast systems in November: ACC in the (a) Niño3 (5°S-5°N, 150°W-90°W), (b) Niño3.4 (5°S-5°N, 170°W-120°W) and (c) Niño4 (5°S-5°N, 160°E-150°W) regions as a function of forecasted month for (blue) EC-Earth3-SR and (orange) EC-Earth3-HR. The reference datasets are ESA SST CCI dataset (plain line) and ERA5 dataset (dashed line). The hindcast period is 1990-2015 and the 20 members of each system are used. Open circles mean that the ACC is statistically significant at 95 % level of confidence.

The results clearly show that, when initialized in November, the higher-resolution configuration of EC-Earth does not mitigate the impact of the SPG on forecast skill. While extending the main article to fully analyse the November-initialized simulations is beyond the scope of this study, since it would roughly double its length, we have expanded the discussion in lines 470–475 to further address this specific result for November initializations (see below). We have also included the corresponding figure in the Supplement, to illustrate that the added value of increased resolution is strongly dependent on the initial state.

"However, the added value of resolution is less evident in other parts of the globe or when predictions are initialized in November. Indeed, for the November-initialized predictions, ENSO skill remains systematically lower in the EC-Earth3-HR configuration, especially during boreal spring months, indicating a stronger impact of the spring predictability barrier at the highest resolution (Supplementary Figure 2)."

Recommendation for Revision:

Based on the points above, it is suggested that the authors revise the abstract and conclusions to more accurately reflect the findings. A more appropriate characterization would be: "ENSO predictability shows a **slight** improvement with increased model resolution" and further analysis is required to determine the extent to which the spring predictability barrier is affected by this resolution increase.

As mentioned previously, the results are statistically significant even if the increase in the skill, already high in the region for this initialization time, is relatively low. This has been taken into account when rephrasing the sentence, which now reads: "Small but statistically significant improvements are found in HR with respect to SR for predicting ENSO, skill improvements that cannot be generalized to other regions and initialization times. Note that, as described in the reply to the other reviewer, we also specifically state that this improvement in EC-Earth3-HR is shown for a specific region (equatorial Pacific) and initialization time (May). The initial statement in the second paragraph conclusions has also been changed to: "The EC-Earth3-HR forecast system shows a small but statistically significant increase in predictive skill in the equatorial Pacific relative to EC-Earth3-SR; this modest improvement is consistent across regions (from the Western Equatorial Pacific through the canonical ENSO regions), and across variables (SST, zonal surface winds, precipitation, air surface temperature and mixed layer depth)."

Minor Comments:

- **Figure Presentation:** In Figures 2, 3, and 9, the use of more distinct color palettes for the ECE-HR and ECE-SR timeseries is recommended to enhance clarity.

We changed the colours and used blue and vivid orange (#FF6800) in the figures mentioned. We choose this new colour making sure that the figures are colourblind friendly. We show Fig2 as an example:

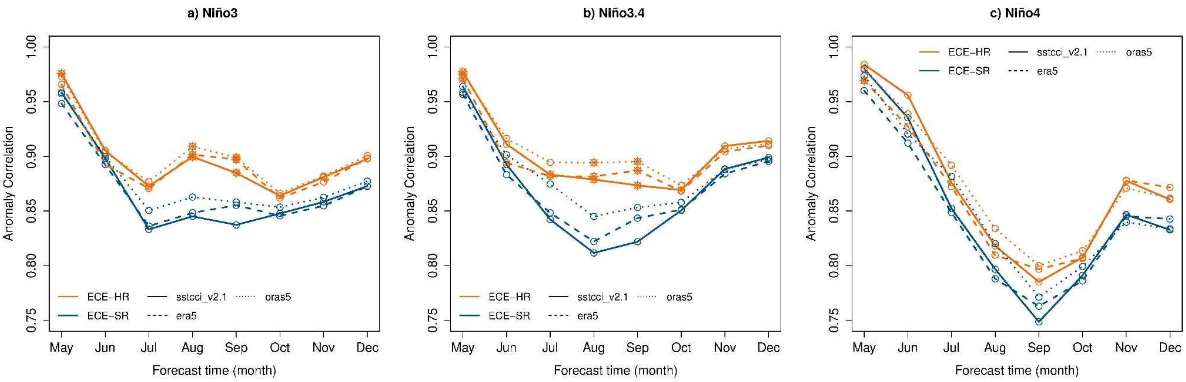

Figure R4: Same as Figure 2 of the manuscript but with new colours.

Typo:

- **Line 45:** The term "important model" is ambiguous.

The word "important" was removed from the sentence because it referred to both model errors and initialization errors, which was not clear from the wording and made it indeed ambiguous: "Dynamical seasonal forecasts, based on GCMs, are affected by model and initialisation errors, which can limit their applicability".