

**Review of "Assessment of a full-field initialised decadal climate prediction system with the CMIP6 version of EC-Earth", by Bilbao et al.**

**Review 1: Reply to Dr. Panos Athanasiadis:**

**Specific Comments**

1. Line 5: It would be helpful to be admitted / clarified that the realistic initialization contains part of the externally forced trends as, for example, the oceans get warmer with global warming. Yes, there are also aerosols and CO<sub>2</sub> which modify radiation and clouds during the simulations, but the warming signal is also contained in the initialized ocean state (progressively warmer).

*Reply: We agree with the reviewer's comment. Certainly the initial conditions include a response to the external forcings, and can even correct part of the forced response that is misrepresented by the models and/or the forcings. To be more precise we have rewritten the sentence to simply say that most of the skill comes from the external radiative forcings.*

2. Line 6: "gets" → is

*Reply: corrected.*

3. Line 13: "in the surface" → at the surface..... the subsurface layer,

*Reply: corrected.*

4. Line 47-50: On this point, there is also another recent study using DCPA-A (Athanasiadis et al., 2020) that shows comparable (even higher) skill for the NAO using CESM-DPLE.

*Reply: The paragraph has been adjusted and the reference included.*

5. Line 61: "is initialisation" → is the realistic initialisation of the ocean state (or of the Earth system, if the authors prefer).

*Reply: suggestion accepted.*

6. Line 65: "especially in" → especially in the deep ocean and before modern instruments (such as ARGO floats) were introduced.

*Reply: suggestion accepted.*

7. Line 66: What is the meaning of the word "exclusively" in this sentence? Initial states are built from observations.

*Reply: The term "exclusively" has been removed.*

8. Line 95: "that take" → which take

*Reply: corrected.*

9. Line 115: “10 member” → 10-member

*Reply: corrected.*

10. Line 154: Has the word “cmorisation” been defined earlier? Perhaps it would be best to keep the “CMOR” part in capital letters.

*Reply: corrected.*

11. Line 155: “data was...” → data were systematically checked for their quality with...

*Reply: suggestion accepted.*

12. Line 174: I expect that the drift cannot affect equally all predictions (initialized in different years with different states, closer to or further from the model climatology). The drift is defined as the average tendency over many years, is not that so?

*Reply: The ‘mean drift correction’ that we apply assumes that drift is the same in all the forecasts, which may be a suitable approximation for certain variables. However, we agree with the reviewer’s comment that drift is unlikely equal in all predictions, in fact, in section 3.3 we highlight how this is not the case for the AMOC and SPGSI. An inefficient drift removal may compromise the skill evaluation. In literature several drift correction methods have been proposed, but to date there is no clear advantage to using a particular method. We have rephrased the sentence to acknowledge that the underlying assumption (i.e. the insensitivity of the drift to the initial state) might not always hold.*

13. Line 199: “persisting it” → making it persist

*Reply: corrected.*

14. Line 219: “Equator-60” → Equator–60 (not hyphen but en dash).

*Reply: corrected.*

15. Line 248: What do you mean by “phases”? The Reviewer guesses what the authors might mean. Please take into account the common use of “phase” as a verb (<https://www.merriamwebster.com/dictionary/phase>) and expand this sentence accordingly.

*Reply: The verb has been changed to ‘puts in phase’.*

16. Line 249: “equivalent” → comparable / similar

*Reply: corrected.*

17. Line 259: "influence of" → influence of the unpredictable part of

*Reply: corrected.*

18. Line 261: Here and elsewhere (where a similar expression is used as an adjective) use "lowfrequency".

*Reply: corrected.*

19. Line 295: "is for the some" → is for some

*Reply: corrected.*

20. Line 319: "5f" is a reference to "Fig. 5f"? Please follow the instructions for authors of this journal – in any case, all references to figures should follow a standard way (same throughout the text).

*Reply: corrected.*

21. Line 344: "ranges(Figure" (add space)

*Reply: corrected.*

22. Line 351: "to aid" → so as to aid

*Reply: corrected.*

23. Line 356: "feature" → behaviour / relationship

*Reply: corrected.*

24. Line 359: "evolve" → evolves (singular)

*Reply: corrected.*

25. Line 363: Why should that be? Same model => same attractor.

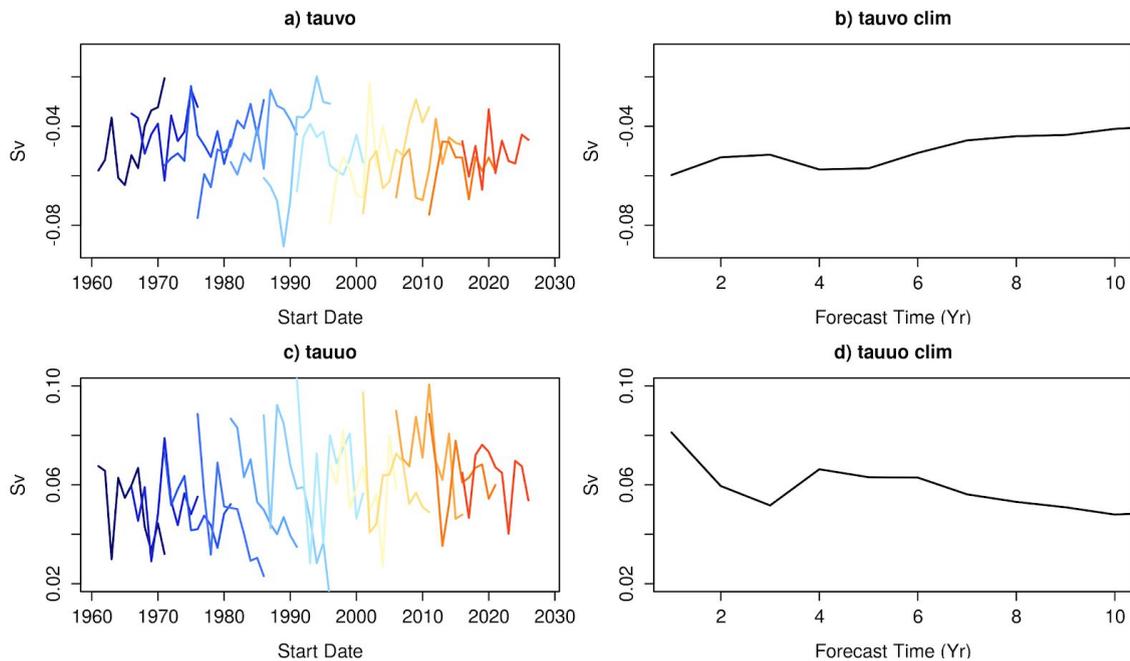
*Reply: What we mean to say is that the model might have more than 1 attractor, which seems to be the case given the existence of two different states of Labrador Convection. We have rephrased for clarity.*

26. Line 389: The Reviewer questions the idea that PRED can reach (or come in the neighborhood of) the model's AMOC attractor in just 10 years.

*Reply: We agree with the reviewer that ten years may be insufficient time to reach the model attractor(s), in particular for the AMOC. We have rephrased the sentence taking it into account.*

27. Line 392: But could not it be that a surface-wind bias (likely associated with a bias in Greenland blocking frequency) favors the formation of sea ice in the Lab. sea, which subsequently blocks heat and moisture surface fluxes? If the authors agree that this is a plausible explanation, at least in part, the Reviewer would suggest to take a look at surface wind biases in that area.

*Reply: To answer the reviewers comment we have looked at the surface wind stress over the Labrador Sea (See Supporting Figure 1). The plot shows that the drift in PRED is too small in comparison with the one in other variables and therefore seems unlikely that the wind is responsible for the very rapid sea ice growth.*



*Supporting Figure 1. FMA Windstress in the Labrador Sea in PRED for a) the meridional direction and b) zonal direction. Ensemble mean forecasts (10 members) of PRED are shown from blue to red every 3 startdates. Panels b) and d) are the climatological values as a function of forecast time.*

28. Line 414: What does “their” refer to?

*Reply: corrected.*

29. Line 421: Here and elsewhere, please make “3” a superscript (exponent). Also, add some small spaces between values/numbers and units.

*Reply: corrected.*

30. Line 428: “simulation” → simulations

*Reply: corrected.*

31. Line 486: “varibility” → variability

*Reply: corrected.*

32. Line 491: Speaking of an “effect”, is this positive, or negative? What kind of effect?

*Reply: We refer to a negative effect on its regional skill. It’s been rephrased to clarify it.*

33. Line 498: Speaking of different members exhibiting different mean states, it is likely that the AMOC has a degree of non-stationarity and not necessarily a uni-modal distribution. If that is so, then the multi-member time average may not correspond to any real attractor (preferred state).

*Reply: We agree with the reviewer’s assessment. The fact that the historical ensemble mean might not represent a preferred state is now mentioned in the sentence.*

34. Line 501: “brings the predictions apart from” → carries the predictions away from

*Reply: suggestion accepted.*

35. Line 517: “prone” → likely A method is prone to errors, instead, the errors themselves are not “prone” to occur.

*Reply: corrected.*

FIGURE 2: In the caption please change the sentence referring to the hatching – what are significant are the ACC values, not the areas themselves. Also write: “Points with missing values...” as the masking is applied to an area. From a scientific view point: the lack of predictive skill in the subpolar gyre (south of Greenland) is an indication of likely poor NAO skill (see Athanasiadis et al., 2020). Have the authors assessed the NAO skill for this set of hindcasts? If the NAO skill is indeed poor, perhaps it would be fair and worth mentioning this possible connection.

*Reply: The caption of Figure 2 has been updated as suggested. We have looked at the NAO and we have low insignificant skill. The possible link between the low NAO skill and that of the SPNA OHC is now mentioned at the end of section 3.3.*

FIGURE 5: This, but also other figures, should be expanded so as to best use the available space. The overall figure width should, however, remain a bit narrower than the width of the main text (plenty of space until there). From a scientific view point: Why does HIST ensemble mean exhibit such a weak ENSO variability, in contrast to PRED, for the 1st year of the predictions? Arguably, because ENSO events are mainly out of phase across the HIST ensemble (as expected). The authors may want to mention this rather trivial explanation.

*Reply: The figures have been changed as suggested. The reviewer is correct, we have added a comment in the text.*

FIGURE 11: “Scatterplot diagram” → Scatter plot.

*Reply: corrected.*

**Review 2: Reply to Dr. Steve Yeager:**

Specific Comments:

My specific recommendations for improvement:

1) Much of the paper elaborates on the negative effects of an “initialisation shock” in the subpolar Atlantic, and this term is even included in the abstract. While the authors offer a definition of what this phrase means (“abrupt changes that occur soon after initialisation as a result of the adjustment of the climate model to the initial state”), I felt that the precise meaning of this term (and its usefulness for understanding system behavior) faded as I read. Certainly, there is a pathological adjustment to initialization going on in this system, but the distinction between shock and drift is not clear, nor is it clear that the initial shock (enhanced Labrador Sea convection) causes the longer term drift (towards reduced convection and sea ice expansion, AMOC decline, etc.). Is the shock really the essential problem in EC-Earth, or is it the drift towards ice-covered Labrador Sea? I suspect the latter is the more fundamental problem. I recommend a reconsideration of the phraseology used throughout.

*Reply: This is a really good point. We agree that both the initialization shock and the mean drift are closely related in our predictions, and that is not possible to disentangle from our analysis if the problem is caused by the processes behind the initial adjustment or by those related to the long-term drift, which might be related. We have tried to improve the clarity of their definitions in the introduction and explain how they might relate with one another. We also specify now that initial adjustments or shocks, when they occur systematically across start dates, can be regarded as the initial stage of the model drift, and even condition its later evolution. We have also carefully revised the rest of the paper to mention both the shock and the drift as the ultimate causes of the lack of skill in the central SPNA.*

2) Related to above, the skill improvement with lead time for NASPG-OHC300 (Figs. 5k and S1k) is interpreted as reflecting initialization shock behavior. However, the later figures (in particular, Fig. 7) make me question whether the relatively high skill for later lead times (e.g., LY7-10) is real skill. I note that HIST\_NoConv exhibits a reasonably high correlation with RECON for Western SPNA-OHC300 (Fig. 7c) which is almost certainly spurious—it appears to relate to a post-1990s spinup of the NASPG in those members (Fig. 6b) which in turn appears related to a transition from fully ice-covered Lab Sea to only partially-covered Lab Sea, with associated increase in convection (Fig.7). This mechanism for reproducing the late 20th century warming of the SPG is unequivocally unrealistic, even though it might yield higher correlation scores for NASPG-OHC300 than HIST itself (could you check this?). At long lead times, PRED seems to show similar behavior as HIST\_NoConv (as noted in the text, but also in terms of Lab Sea transition from ice-covered\_no-convection to partially-ice-covered\_some convection), suggesting that the better NASPG-OHC300 “skill” at long lead times is a spurious artifact of an unrealistic warming mechanism. If true, this changes the interpretation of what is happening in the prediction system (i.e., it is not “initialization shock” followed by skill recovery via better representation of real mechanisms).

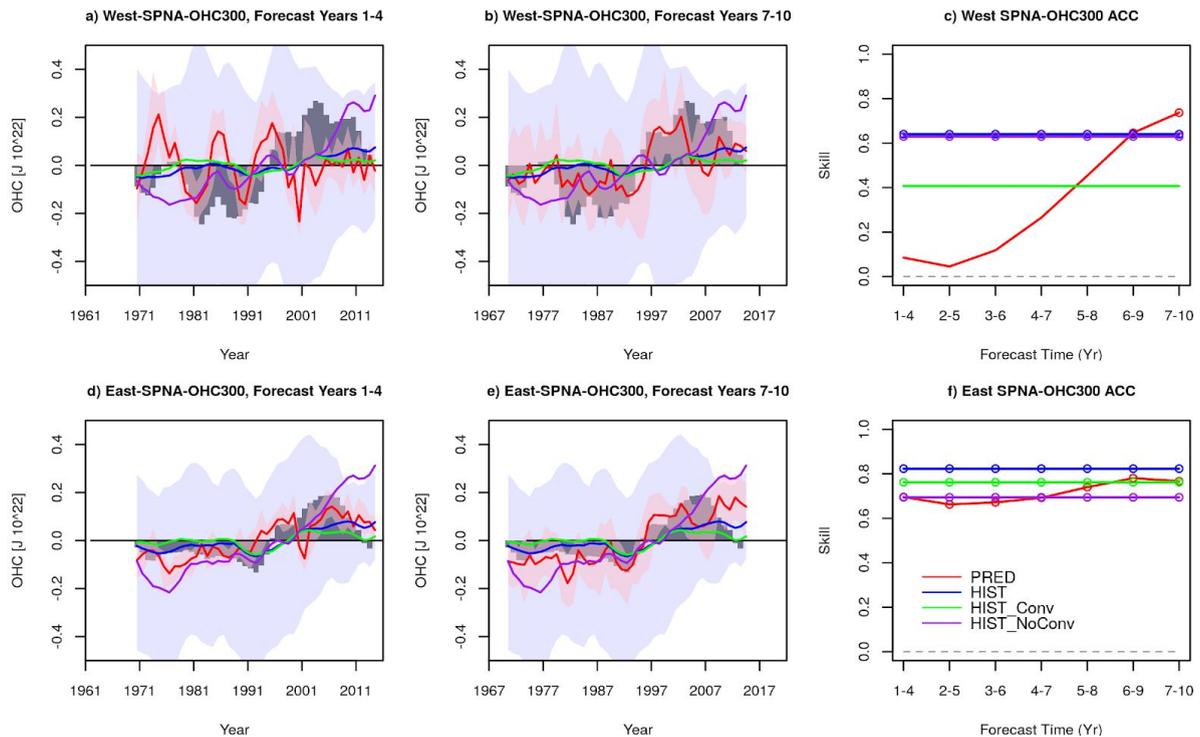
If not true, how do the authors explain the increase in NASPG-OHC300 skill with lead time (Fig. 5k)?

*Reply: This is a really interesting hypothesis, which we had not thought of. To check if the hypothesis is true we have looked in more detail at the East and West SPNA OHC in the upper 300m (see Supporting Figure 1). As suggested, we have divided the historical ensemble into those members which exhibit convection (Hist\_Conv, 7 out of 10 simulations) and those with suppressed convection (Hist\_NoConv, 3 out of 10). The reviewer correctly pointed out that the historical members with no convection have higher ACC values than those with convection (Figure 1c), at least in the West-SPNA (note that to avoid differences in skill due to differences in the verification period we have set a common verification period: 1971-2008). This is because the members with no convection show a long-term warming trend which happens to coincide, in large part, with the observed one, even if it happens for the wrong reasons (that is the melting of Labrador Sea ice allowing for open ocean convection) .*

*Even though PRED at the later forecast times reaches comparable ACC values to Hist\_NoConv, these do not seem to be explained by the same mechanism explaining the spurious OHC300 warming in Hist\_NoConv (Figure 1c). Indeed, the timeseries in Supporting Figure 1b shows that the large improvement in skill at the longest forecast range with respect to the first forecast years comes from a good representation of the decadal variability, including the quick transition from cold to warm OHC anomalies that occurred during the mid-90s, with a radically different long-term evolution than in Hist\_NoConv. The lack of skill in PRED at the initial forecast times comes from a poor representation of the observed inter-annual variability, which in the forecast shows some ‘spikes’ (or abrupt transitions) that might well result from the initialization shocks.*

*The respective plots for the eastern SPNA OHC300 can help explain the origin of the skill recovery at forecast years 7-10 over the whole SPNA. Indeed, they show that the eastern side of the region has rather constant predictive skill, comparable in PRED and both Historical ensembles, which implies that most of the skill might be forced. In terms of the mean gyre circulation, the eastern SPNA is upstream of the western side, and therefore the mean flow might be advecting the (skilfully) forecasted anomalies from the eastern into the western region, eventually substituting the unrealistic OHC anomalies generated in the first forecast years by the labrador convection collapse. If we take into account that the eastern SPNA maintains a similar level of skill all along the forecast, and that the western SPNA recovers it at the very end, we could then explain the increase in skill with lead time for the whole SPNA in Figure 5k. This is, of course, just one plausible hypothesis, but exploring it further would require extending the paper in a new direction, something that we would prefer not to do given that it is already lengthy and dense.*

*A reduced version of Supporting Figure 1 has been included in the Supplement, and is now discussed in the text to explain our hypothesis behind the whole SPNA skill recovery.*



**Supporting Figure 1:** Timeseries and ACC skill of the West and East SPNA-OHC300 anomalies (with respect to 1971-2018). The timeseries and ACC have been computed for the common period for all forecast ranges (i.e. 1970-2018). The first two columns show the observed (grey bars) and predicted (PRED in red, HIST in blue, HIST\_Conv in green and HIST\_NoConv in purple) timeseries for the 1-4 and 7-10 forecast years respectively. The third column shows the ACC for PRED (red), HIST (blue), HIST\_Conv (green) and HIST\_NoConv (purple). Statistically significant ACC values (at the 95% confidence level) are shown as empty circles.

3) Figures 8 and 10 have many small thin lines of various colors and hues that are very hard to distinguish (this reviewer is slightly color blind). Can a revised version be developed that is easier to see, particularly Fig. 8? I recognize that “easy to see” is quite subjective, and that these figures contain lots of information that is hard to display any other way. Perhaps the answer is “the figures are as clear as they can reasonably be” and I am in a small minority that has trouble viewing them, but if others (reviewers, coauthors, colleagues) also have difficulty with these figures then please make an effort to improve them.

*Reply:* The resolution and quality of the figures have been improved. In figures 8 and 10 the profiles now go down to 500m rather than 800m to allow for a better visualization of the near-surface differences. And in Figure 8 we now include two rows, one comparing PRED with RECON, and another comparing PRED with HIST, which reduces the amount of lines, and allows to see better the differences between RECON and PRED.

Additional Comments (by line number)

63: This is not a complete sentence.

*Reply: corrected.*

80: The meaning of “biases in the predictions” is not clear. Model mean bias is to be expected when using anomaly initialization. Do you mean “time-dependent biases in the predictions” (i.e. drift)?

*Reply: We have rephrased it now as “skill degradation in the predictions”.*

111: ORCA has not been defined

*Reply: Added information.*

124: There is no mention of how the land model component is initialized—can you please clarify?

*Reply: Added information of the land model component (HTESSEL) and the initialisation.*

205-207: Since sentence paragraphs are not advisable.

*Reply: the sentence has been deleted as it was not necessary.*

243: “signal” instead of “trend” to avoid awkward phrasing?

*Reply: suggestion accepted.*

264: “associated with” instead of “to”

*Reply: corrected.*

271: It would help to interpret Fig. 3 if the breakdown of MSSS into correlation and conditional bias terms were given explicitly (perhaps in section 2.3), and the corresponding relationships between Fig. 3 panels clarified (e.g., is panel a = panel d + panel g?).

*Reply: The description of the MSSS in section 2.3 has been extended. We now include two equations, the one we used for the computation, that it's taken from Goddard et al. (2013; Eq. 5), and a more compact version that represents the numerator as the difference between two terms, one based on ACCs and another on conditional biases. Both equations are reproduced below. The re-arrangement in Equation 2 has allowed us to see that the middle and bottom plots in the former version of Figure 3 did not represent direct contributions to the MSSS. It is actually the differences in the squared ACCs/conditional biases that determine the final MSSS, and therefore those are the quantities that we now represent in the middle and bottom rows of Figure 3 to guide the interpretation of the MSSS. If we disregard the denominator (which is just a scaling factor with no impact on the sign to produce a skill metric that goes from -1 to 1), we can interpret the values in the upper row as the difference between the values on the second and the third row. The discussion on Figure 3 has been modified according to the changes.*

Also, because positive values in the difference between the squared ACC in PRED and the squared ACC in HIST do not necessarily correspond to a beneficial effect of initialization on skill (e.g. if the ACC in PRED is negative, and positive in HIST) we have decided to keep the plots on the differences in ACC from the former figure, which have been placed as a third row in Figure 2.

286: Missing "(Figure"

*Reply: corrected.*

302: There also seems to be noteworthy skill in the western tropical Pacific which should not be ignored.

*Reply: added to the text.*

315: I'm confused by this statement. Since both PRED and HIST show  $SER < 1$  in the first few months (Fig. 5c), aren't they both overconfident (under-dispersed)?

*Reply: This has been corrected.*

Fig. 5: It's unclear from the caption whether purple line (persistence forecast) is an ACC or MSSS score.

*Reply: It has been indicated in the caption that the purple line refers to the persistence based on the ACC.*

326, 340: It's not clear to me that the HIST spread is "excessive" and "too large" (although it is certainly larger than PRED) since I'm unsure how the concept of reliability applies to uninitialized ensembles that aren't expected to be able to predict internal variability.

*Reply: The spread-error-ratio is a measure of reliability that has been typically applied both to initialised and non-initialised forecasts (Ho et al. 2013; Robson et al. 2018). It evaluates if the typical distance between ensemble members is comparable to the typical distance between the individual members and the observations. Both terms are small at the beginning of an initialised forecast (because of initialization itself) and are expected to grow as the forecast progresses, although their ratio could vary, and converge to the one in the uninitialised experiments. That's why we show them both.*

*References:*

*Ho CK, Hawkins E, Shaffrey L, Bröcker J, Hermanson L, Murphy JM, Smith DM, Eade R (2013) Examining reliability of seasonal to decadal sea surface temperature forecasts: the role of ensemble dispersion. Geophys Res Lett 40(21):5770–5775*

*Robson, J., Polo, I., Hodson, D. L. R., Stevens, D. P., and Shaffrey, L. C.: Decadal prediction of the North Atlantic subpolar gyre in the HiGEM high-resolution climate model, Climate Dynamics, 50, 921–937, <https://doi.org/10.1007/s00382-017-3649-2>, 2018.*

360: “black” should be “green”?

*Reply: corrected.*

396: Fig. 7f is mislabelled as “e)”

*Reply: corrected.*

Fig. 8: I find it very hard to make out the relevant details in this figure even after magnifying to 400%. Can you devise a better graphic that is more legible for color challenged individuals? Same comment applies to Fig. 10. One simple option might be to just plot upper 400m to magnify the key region of interest. Another might be to plot as differences from HIST.

*Reply: To improve the visibility of the relevant details we have increased the size of the figure, zoomed it to the upper 500m, and duplicated the panels to compare separately PRED and RECON (for which the lines are closer to each other), and PRED and HIST.*

431: Please double check the sign of the restoring freshwater fluxes. Fig. 8 suggests that RECON is saltier at the surface than HIST (less stratified by salinity) which implies that a positive SALT flux (ie, negative freshwater flux) is used in the restoring.

*Reply: This has been corrected. The freshwater fluxes are defined in the model as going out of the ocean, while the heat fluxes into the ocean. The figure has been modified so both fluxes are into the ocean and the text adjusted.*

451: Incorrect reference to figure 10 within this sentence.

*Reply: corrected.*

Fig. 11: I think the last sentence of caption should be “dark green cross”?

*Reply: corrected.*

501: Here and elsewhere, the distinction between “initialization shock” and model “drift” could be clarified. (also, what is the “expected trajectory”? a skillful one? one towards the model mean climatology?)

*Reply: The sentence and all the other mentions to the drift and initialization shock have been rewritten according to this comment and the first one.*

### **Review 3: Reply to Prof. Gerard Meehl:**

#### **Specific comments:**

Line 42: Researchers at NCAR have documented the prediction of aspects of the IPO (e.g. Meehl et al. 2016) and have noted that the response to volcanic eruptions could explain in part why there is less overall predictability of the IPO compared to AMV (Meehl et al., 2015).

*Reply: These articles are now cited and have been added accordingly to the list of references.*

Line 83: A paper that should be referenced here that was important for documenting one of the main methods of bias adjustment that has subsequently been used is Doblas-Reyes, et al. 2013 (already in the reference list).

*Reply: The paper is now cited also in this part of the manuscript.*

Lines 279-281: The authors need to explain more clearly what a negative value of MSSS means. They say in passing that PRED has lower ACC values than HIST, but more explanation would be helpful for the reader to interpret this important result which produces strikingly large areas of negative values in Fig. 3.

*Reply: A more detailed description of MSSS, what it represents, and the equations used to compute it has been included in the methods section (subsection 2.3). We have also changed Fig. 3 to make it more easily interpretable, and adjusted the corresponding discussion accordingly.*

Lines 370-372: The authors note a very interesting feature in that their model drifts differently in two different periods. They should elaborate a bit more about this potentially very important aspect of their simulations that has profound implications for assessing prediction skill.

*Reply: We have expanded the paragraph to discuss more at depth these non-stationary drifts and the need to address them with better drift correction techniques.*

# Assessment of a full-field initialised decadal climate prediction system with the CMIP6 version of EC-Earth

Roberto Bilbao<sup>1</sup>, Simon Wild<sup>1</sup>, Pablo Ortega<sup>1</sup>, Juan Acosta-Navarro<sup>1</sup>, Thomas Arsouze<sup>1</sup>, Pierre-Antoine Bretonnière<sup>1</sup>, Louis-Philippe Caron<sup>1</sup>, Miguel Castrillo<sup>1</sup>, Rubén Cruz-García<sup>1</sup>, Ivana Cvijanovic<sup>1</sup>, Francisco Javier Doblas-Reyes<sup>1,2</sup>, Markus Donat<sup>1</sup>, Emanuel Dutra<sup>3</sup>, Pablo Echevarría<sup>1</sup>, An-Chi Ho<sup>1</sup>, Saskia Loosveldt-Tomas<sup>1</sup>, Eduardo Moreno-Chamarro<sup>1</sup>, Nuria Pérez-Zanon<sup>1</sup>, Arthur Ramos<sup>1</sup>, Yohan Ruprich-Robert<sup>1</sup>, Valentina Sicardi<sup>1</sup>, Etienne Tourigny<sup>1</sup>, and Javier Vegas-Regidor<sup>1</sup>

<sup>1</sup>Barcelona Supercomputing Center, Jordi Girona 29, 08034, Barcelona, Spain

<sup>2</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010, Barcelona, Spain

<sup>3</sup>Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

**Correspondence:** Roberto Bilbao (roberto.bilbao@bsc.es)

**Abstract.** In this paper we present and evaluate the skill of ~~the~~an EC-Earth3.3 decadal prediction system contributing to the Decadal Climate Prediction Project - Component A (DCPP-A). This prediction system is capable of skilfully simulating past global mean surface temperature variations at interannual and decadal forecast times as well as the local surface temperature in regions such as the Tropical Atlantic, the Indian Ocean and most of the continental areas, although most of the skill comes from the representation of the external radiative forcings. A benefit of initialisation in the predictive skill is evident in some areas of the Tropical Pacific and North Atlantic Oceans in the first forecast years, an added value that ~~gets~~is mostly confined to the south-east Tropical Pacific and the eastern Subpolar North Atlantic at the longest forecast times (6-10 years). The central Subpolar North Atlantic shows poor predictive skill and a detrimental effect of ~~the initialisation due to the occurrence of an initialisation shock, itself related to a~~initialisation that leads to a quick collapse in Labrador Sea convection ~~by the third forecast year that leads to a rapid,~~followed by a weakening of the Atlantic Meridional Overturning Circulation (AMOC) and excessive local sea ice growth. The shutdown in Labrador Sea convection responds to a gradual increase in the local density stratification in the first years of the forecast, ultimately related to the different paces at which surface and subsurface temperature and salinity drift towards their preferred mean state. This transition happens rapidly ~~in~~at the surface and more slowly in the subsurface, where, by the tenth forecast year, the model is still far from the typical mean states in the corresponding ensemble of historical simulations with EC-Earth3. Our study thus highlights the ~~importance of the Labrador Sea for initialisation, the relevance of reducing model bias by model tuning or, preferably, model improvement when using full-field initialisation, and the need to identify~~Labrador Sea as a region that can be sensitive to full-field initialisation and hamper the final prediction skill, a problem that can be alleviated improving the regional model biases through model development and by identifying more optimal initialisation strategies.

## 1 Introduction

Interest in seasonal-to-decadal climate predictions has grown in recent years due to their potential to provide relevant climate information for decision making in different socio-economic sectors (e.g. Suckling, 2018; Solaraju-Murali et al., 2019; Merryfield et al., 2020). Scientifically, climate predictions have provided novel ways of evaluating and comparing climate simulations with observations and improve our understanding of the intrinsic predictability of the climate system, including the key mechanisms operating at interannual-to-decadal timescales.

On these time-scales a large part of the predictable signal of climate variations during the observational period is attributable to changes in external radiative forcings (i.e. changes in the climate system energy balance), which can be of natural (e.g. solar irradiance, volcanic aerosols) or anthropogenic origin (e.g. greenhouse gas concentrations, land use changes and anthropogenic aerosols). For example, at the global scale most of the surface temperature changes can be explained by the warming trend caused by the increasing atmospheric greenhouse gas concentrations, which is partly compensated by a parallel increase in anthropogenic aerosols (e.g. Bindoff et al., 2013), and the sporadic cooling episodes that followed the major volcanic eruptions of Mt Agung (1963), El Chichon (1982) and Pinatubo (1991) (e.g. Ménégoz et al., 2018; Hermanson et al., 2020). Estimates of past changes in these radiative forcings are prescribed as boundary conditions to drive the so called historical climate simulations, which investigate the influence of the forcings on the recent climate variations. These experiments are continued into the future as climate projections with imposed anthropogenic radiative forcings that follow different theoretically derived socio-economic emission scenarios (O'Neill et al., 2016).

The other main source of predictability originates from internal variability, in particular in the slowly evolving components of the climate system (i.e. the ocean) (e.g. Meehl et al., 2009). ~~Beside~~Besides being driven with external radiative forcings, climate predictions are also initialised from the observed state to put the model in phase with observed internal variability. Predictive skill of real-time forecast systems is assessed by producing retrospective predictions (or hindcasts), that are then contrasted with observations. At seasonal to annual timescales, hindcasts show high levels of predictive skill for El Niño-Southern Oscillation (ENSO) (e.g. Barnston et al., 2019). On decadal timescales, many climate models have also shown the capacity to skillfully predict the Atlantic Multidecadal Variability (AMV) (e.g. García-Serrano et al., 2015), and to a lesser extent the Interdecadal Pacific Oscillation (IPO) (~~e.g. Mochizuki et al., 2009; Chikamoto et al., 2015~~)(e.g. Mochizuki et al., 2009; Chikamoto et al., 2015; Meehl et al., 2009). The North Atlantic Ocean, and more precisely the Subpolar Gyre, has been identified as a region where different retrospective prediction systems skillfully predict the evolution of sea surface temperatures (SST) and upper ocean heat content (OHC) (e.g. Pohlmann et al., 2009; Keenlyside et al., 2008; Robson et al., 2018; Yeager et al., 2018), although these same systems show limited capacity to predict the climate of the neighboring continents, which might be related to an inaccurate representation of key teleconnection mechanisms (e.g. Goddard et al., 2013; Doblas-Reyes et al., 2013). Encouragingly, a recent ~~paper~~study by Smith et al. (2020) has shown that decadal predictions contributing to CMIP6 can be skillful at predicting the low-frequency variations of the North Atlantic Oscillation (NAO), the leading mode of the winter atmospheric circulation in the Northern

Hemisphere (Hurrell, 1996), when considering a large multi-model ensemble. ~~The study concludes~~ Similarly promising results for predicting the NAO at multi-annual time-scales has also been shown for the Decadal Prediction Large Ensemble from NCAR (Athanasiadis et al., 2020). Smith et al. (2020) also conclude that the NAO and the related climate signals over Europe might be more predictable than models suggest, and that large ensembles of predictions are necessary ~~to circumvent an inherent problem of since~~ current forecast systems ~~-, the fact that they~~ can strongly underestimate the predictable ~~signals~~ signal (Scaife and Smith, 2018).

To reinforce the inter-comparability of the results and allow for the exploitation of large multi-model ensembles, the decadal climate prediction community has promoted the development of coordinated climate prediction exercises. The Decadal Climate Prediction Project (DCPP; Boer et al., 2016), as part of the Coupled Climate Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016), and building upon CMIP5 (Taylor et al., 2012) and the efforts of previous projects (e.g. SPECS, ENSEMBLES), provides such a framework for addressing different aspects and current knowledge gaps of decadal climate prediction. DCPP-A is the main component and consists of an ensemble of decadal hindcasts, initialised at yearly intervals from 1960 up to 2018 using prescribed CMIP6 external radiative forcings.

A crucial step to maximise the skill of decadal predictions is ~~initialisation~~ the realistic initialisation of the climate system. It is of major importance to produce physically consistent initial conditions that reflect as faithfully as possible the observed state of climate. In particular, ~~for~~ the 3-dimensional ocean temperature and salinity fields, which determine the basin-wide density gradients and through them the large-scale ocean circulations and transports, which are essential for predictability on decadal timescales (~~e.g., Meehl et al., 2014; Brune and Baehr, 2020~~) (e.g. Meehl et al., 2014; Brune and Baehr, 2020). However, observational records are sparse in time and space, especially in the deep ocean and before modern instruments (such as ARGO floats) were introduced, which poses a challenge to accurately constrain the initial state ~~exclusively~~ from observations. For this reason, a typical approach in climate prediction is to use initial conditions from ocean and atmosphere reanalysis. These are produced with data assimilation techniques that ensure a dynamically consistent estimation of the climate state that takes into account observational uncertainties.

Due to structural errors in climate models and biases in their climatologies, when initialised from the observed state, predictions suffer from initial shocks and drifts (e.g. Magnusson et al., 2012; Sanchez-Gomez et al., 2016; Kröger et al., 2018; Meehl et al., 2014). ~~Initial shocks are~~ We refer to as initial shocks to abrupt changes that occur soon after initialisation as a result of the adjustment of the climate model to the initial state ~~-, while the drift represents and/or to incompatibilities between the initial conditions of the different components, while with model drift we refer to~~ the mean evolution of the forecasts towards ~~the an~~ imperfect mean model climate~~-, which is tightly linked to how systematic model errors develop (Sanchez-Gomez et al., 2016). When their occurrence is consistent across start dates, initial shocks (which are not present in all systems) are part of the drift and can even condition its development.~~ Two main approaches of initialisation are often used; ‘full-field initialisation’ which uses directly observational estimates to initialise the model (~~e.g., Pohlmann et al., 2009~~) (e.g. Pohlmann et al., 2009), and ‘anomaly initialisation’ that imposes the observational estimate anomalies on the model climatology (~~e.g., Smith et al., 2007; Keenlyside et~~ No clear advantage of one approach with respect to the other has been found in terms of forecast quality (e.g. Magnusson et al., 2012; Weber et al., 2015; Boer et al., 2016). The latter was specifically designed to reduce the forecast drift, as it implies

initialising from a state closer to the model climatology. However, incompatibilities between the imposed anomalies and the typical model variability have been shown to cause dynamical imbalances leading to biases skill degradation in the predictions (e.g., Magnusson et al., 2012; Hazeleger et al., 2013; Volpi et al., 2017) (e.g. Magnusson et al., 2012; Hazeleger et al., 2013; Volpi et al., 2017). The occurrence of forecast drifts and biases compromises the quality of the predictions, a problem that can be partly circumvented by correcting the predictions a posteriori, for example by computing forecast-time dependent anomalies (e.g., Meehl et al., 2014; Goddard et al., 2013; Choudhury et al., 2017).

With the objective of reducing initial shocks, several decadal forecast centres consider the production of in-house assimilation experiments with the same model or model components used for the forecasts from which the initial states are derived. The simplest and commonly used assimilation framework consists in producing assimilation runs with individual model components (referred to as 'weakly coupled'), typically of the ocean model, since it is the most important for the predictability on decadal timescales (e.g., Sanchez-Gomez et al., 2016; Servonnat et al., 2015) (e.g. Sanchez-Gomez et al., 2016; Servonnat et al., 2015). This method may benefit the initialisation of an individual model component, however initialisation shocks may occur due to incompatibilities among the initial conditions. For this reason, many forecast centres are moving towards fully coupled assimilation (referred to as 'strongly coupled'), which is more technically challenging but assures physical consistency of the initial conditions of all the components, among other advantages (e.g., Brune and Baehr, 2020) (e.g. Brune and Baehr, 2020). For assimilation, a range of different techniques have been used to produce the reconstructions, from classical nudging approaches based on Newtonian relaxation (e.g., Sanchez-Gomez et al., 2016; Servonnat et al., 2015) (e.g. Sanchez-Gomez et al., 2016; Servonnat et al., 2015) to more complex and computationally expensive methods like the Ensemble Kalman Filter approach (e.g., Counillon et al., 2014; Dai et al., 2020) that (e.g. Counillon et al., 2014; Dai et al., 2020), which take into account aspects of the observational uncertainty.

The aim of this paper is to present and analyse a decadal prediction system within the EC-Earth3 model contributing to the CMIP6 DCP-A. The paper is organized as follows: section 2 provides a description of the EC-Earth3 forecast system, the initialisation approach considered, the skill evaluation metrics and the observational datasets used. In section 3, we characterize the predictive capacity for the surface temperatures and investigate the importance of the initialisation on surface temperatures, upper ocean heat content and several interannual-to-decadal indices of climate variability, followed by an analysis of the predictive skill in the North Atlantic. This section illustrates that the low skill in the Central Subpolar North Atlantic appears to be related to a strong initial shock and the subsequent model drift. The final section summarizes the key results and conclusions of this study.

## 2 Data and Methodology

### 2.1 EC-Earth3 Decadal Forecast System

All experiments analysed in this study were performed with the CMIP6 version of the EC-Earth version 3 Atmosphere-Ocean General Circulation Model (AOGCM) in its standard resolution (Döscher and the EC-Earth Consortium, in prep.). Its atmospheric component is the Integrated Forecast System (IFS) from the European Centre for Medium-Range Weather Forecasts (ECMWF), cycle cy36r4, with a T255 horizontal resolution (grid size approximately 80km) and 91 vertical levels. The land

surface scheme HTESSEL (Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land, (Balsamo et al., 2009)) is integrated in IFS and the vegetation fields are prescribed and have been derived from an EC-Earth historical simulation coupled with the LPJ-GUESS dynamic vegetation model (LPJGuess; Smith et al., 2014). The ocean component is the version 3.6 of the Nucleus for European Modelling of the Ocean (NEMO; Madec and the NEMO Team, 2016), which is itself composed of the ocean model OPA (Ocean PARallelise) and the Louvain-La-Neuve sea ice model (LIM3; Rousset et al., 2015), both run with an ORCA1 ~~horizontal resolution (ea-configuration (a tripolar grid defined for NEMO with 1° horizontal~~ nominal resolution) and 75 vertical levels. The atmospheric and oceanic components are coupled through OASIS (Craig et al., 2017). ~~The vegetation fields are prescribed and have been derived from an EC-Earth historical simulation coupled with the LPJ-GUESS dynamic vegetation model (LPJGuess; Smith et al., 2014).~~

Our decadal prediction system follows the CMIP6 DCP-A protocol (Boer et al., 2016) and therefore consists of ~~10-member~~ 10-member ensembles of 10-year long predictions initialised every year in November from 1960 to 2018 (referred to as PRED hereafter). To determine the impact of initialisation, PRED is compared with an ensemble of 15 CMIP6 historical simulations (1960-2015) (Eyring et al., 2016) continued with the SSP2-4.5 scenario simulations (2015-2100) (O'Neill et al., 2016) and performed with the same model version as PRED. These experiments (referred to as HIST hereafter) correspond to the CMIP6 members (2,7,10,12,14,16-25), all the ones which were performed at the Barcelona Supercomputing Center (BSC). PRED and HIST are both performed with prescribed CMIP6 radiative forcing estimates (i.e., solar irradiance, and green house gas, anthropogenic aerosol and volcanic aerosol concentrations) for the historical period (1960-2014) and the SSP2-4.5 scenario on the subsequent years (Eyring et al., 2016).

In PRED, the different components (atmosphere, land, ocean and sea ice) have been initialised using full-field initialisation. The atmospheric and land initial conditions have been interpolated from ERA-40 reanalysis outputs (Uppala et al., 2005) for the period 1960-1978 and from ERA-Interim (Dee et al., 2011) for the period 1979-2018. The ERA-Interim land surface fields were replaced by the ERA-Interim/Land offline land surface reanalysis (Balsamo et al., 2015) driven by ERA-Interim surface fields and bias-corrected using precipitation from the Global Precipitation Climate Project (GPCP, Adler et al., 2018). The ocean and sea ice initial conditions come from a NEMO-LIM reconstruction, forced at the surface with fluxes from the Drakkar Forcing Set v5.2 (DFS5.2 Brodeau et al., 2010) up to 2015 and with ERA-Interim (Dee et al., 2011) thereafter. In this reconstruction, ocean temperature and salinity fields from the ORAS4 reanalysis (Mogensen et al., 2012) are assimilated through a standard surface nudging approach (e.g., Servonnat et al., 2015)(e.g. Servonnat et al., 2015), using temperature and salinity restoring coefficients of  $-40W/m^2/K$  and  $-150mm/day$ , respectively. Even if no direct assimilation of sea ice products is performed, the atmospheric surface fluxes combined with the surface ocean temperature nudging, are sufficient to bring the initial sea ice state close to observations (e.g., Guemas et al., 2014)(e.g. Guemas et al., 2014). Below the mixed layer, a Newtonian relaxation term is also applied to assimilate 3D ORAS4 temperature and salinity fields. For this, a relaxation timescale that increases monotonically with depth is used, which takes approximate values of 10 days at 1000m, 100 days at 3000m and 330 days at 5000m. Subsurface relaxation is applied everywhere except within the 3°S–3°N band to prevent spurious vertical velocity effects (Sanchez-Gomez et al., 2016).

155 To generate the 10 members of PRED different strategies are followed depending on the model component. The ensemble spread for the atmospheric initial conditions is generated by adding infinitesimal random perturbations to the 3-dimensional temperature field. For the ocean and sea ice initial conditions, five different reconstructions are performed following the nudging strategy previously described, each assimilating one of the 5 members of ORAS4. The 5 ocean and sea-ice states generated are combined with two different atmospheric initial conditions each to produce the 10 ensemble members.

160 All the simulations completed in this study were performed in the supercomputer Marenostrum IV, hosted at the BSC, using the Autosubmit workflow manager (Manubens-Gil et al., 2016), a Python toolbox specifically developed at the BSC to facilitate the production of experimental protocols with EC-Earth. This toolbox can easily handle experiments with different members, different start dates and different initial conditions. Autosubmit is hosted in the Gitlab repository of the BSC Earth Sciences Department (<https://earth.bsc.es/gitlab/es/autosubmit>). The scripts to run the model and all subsidiary tools are also included  
165 in the Gitlab repository under version control, and the tool that generates the perturbations saves the seed employed for each member, both contributing to guarantee the reproducibility of the experiments within the maximum fidelity permitted by the model.

The raw model outputs were formatted following the Climate Model Output Rewriter (CMOR)/CMIP6 conventions to ensure efficient use and dissemination with the scientific community. This was done with 'ece2cmor' ([https://github.com/EC-](https://github.com/EC-Earth/ece2cmor3)  
170 [Earth/ece2cmor3](https://github.com/EC-Earth/ece2cmor3)), a Python tool for post-processing and ~~emorisation~~ [CMORisation](#) developed for EC-Earth3 which was implemented in the Autosubmit workflow. After re-formatting, the model data ~~was~~ were systematically quality checked [for their quality](#) with various tools to ensure no missing files and scientific validity. Both PRED and HIST experiments are published on the BSC data node of the Earth System Grid Federation (ESGF) where they are publicly available.

## 2.2 Observational Data for Verification

175 Various datasets are used as reference for estimating the forecast quality of the two EC-Earth3 ensembles HIST and PRED. To evaluate surface temperature we use the gridded temperature anomaly products NASA GISTEMPv4 (Lenssen et al., 2019) and the Met Office HadCRUT4 (Morice et al., submitted). Both datasets combine near-surface air temperature (SAT) over land and sea surface temperature over the ocean (SST). For indices related to SST only, we use the Met Office HadISSTv3 (Kennedy et al., 2011). For upper-ocean heat content we use the Met Office EN4.2.1 gridded ocean temperature (Good et al., 2013). For  
180 comparing spatial fields with observations, EC-Earth3 predictions and historical simulations are re-gridded to the observational grid in the case of the surface temperature variables corresponding to a 2°x2° regular grid for NASA GISTEMP4 and a 5°x5° regular grid for HadCRUT4. Ocean heat content is re-gridded to a 2°x2° regular grid. Model simulations are masked in regions where and when observations have missing values. The regions with missing values in observations remain similar over the investigated period, especially for NASA GISTEMPv4.

## 185 2.3 Forecast Drift Adjustment and Verification Metrics

In the full-field initialisation approach, models are initialised close to the observed state and as the forecasts progress they experience a drift towards the (imperfect) model attractor. This drift needs to be corrected to prevent systematic errors in the

prediction. To avoid drift-related effects, the evaluation of climate predictions against observations is performed in the anomaly space (e.g., Meehl et al., 2014) (e.g. Meehl et al., 2014). In this paper we use the 'mean drift correction' method which consists  
 190 in computing the anomalies relative to the forecast time-dependent climatology, ~~which for simplicity assumes~~. This implies that the drift is assumed to be insensitive to the background climate state (e.g., Garrea-Serrano and Doblas-Reyes, 2012; Goddard et al., 2013) (e which might not always hold.

Observed and HIST anomalies are computed with respect to their climatologies, ~~in~~. In the case of HIST it is computed from the ensemble mean. All climatologies are computed for the common reference period 1970-2018. This is the longest  
 195 period for which predictions at all forecast years (1 to 10) can be produced, and thus allows us to compute a climatology that is consistent across the whole forecast range. For forecast quality assessment purposes, we use the longest period available for each forecast year (e.g. 1961-2018 for forecast year one, and 1970-2018 for forecast year ten), to produce in each case the most accurate estimate of the predictive skill. This implies that the skill of PRED and HIST may change with the forecast time as the verification period changes.

200 To measure the forecast quality we use the Anomaly Correlation Coefficient (ACC) and the Mean Square Skill Score (MSSS). The ACC measures the linear association between the predicted mean and the observations, but is insensitive to the scaling. ~~To determine the impact of the initialisation we compute~~ The MSSS metric takes into account differences in magnitude, as it compares the mean-square errors of a given forecast with those from a benchmark prediction (e.g. persistence), both evaluated against observations.

205 The impact of initialisation on predictive skill is first assessed by computing the ACC differences between the decadal hindcasts and historical simulations. ~~To determine whether these differences are statistically significant we follow~~ These differences are deemed to be statistically significant according to the methodology developed by Siegert et al. (2017), which corrects for the independence assumption when two forecasts are strongly correlated. ~~In addition to the differences in ACC, to investigate the skill associated with the initialisation, we use the Mean Square Skill Score ( $MSSS = 1 - (MSE\_PRED/MSE\_HIST)$ );~~  
 210 ~~where MSE\_PRED and MSE\_HIST are the mean square errors between observation and~~ The MSSS is used as well to evaluate the impact of initialization. To do so, we compute it for PRED, using HIST as the reference forecast to beat. Following Equation 5 from Goddard et al. (2013):

$$MSSS(P, H) = \frac{ACC_P^2 - [ACC_P - (\sigma_P/\sigma_O)]^2 - ACC_H^2 + [ACC_H - (\sigma_H/\sigma_O)]^2}{1 - ACC_H^2 + [ACC_H - (\sigma_H/\sigma_O)]^2} \quad (1)$$

215 where  $ACC_P$  and  $ACC_H$  are respectively the ACC for PRED and HIST, and  $\sigma_O$ ,  $\sigma_P$  and  $\sigma_H$  are the standard deviations for the observations, PRED and HIST, respectively), which is described in detail in Goddard et al. (2013). The MSSS is a skill metric that allows us to compare the performance of PRED with respect to HIST. In the absence of a mean bias the MSSS can be divided into two terms, the correlation and the conditional bias, which provide different information. While the correlation takes into account the interannual variability and sign of linear trend, it is scaling invariant (i.e. independent of the signal amplitude). The conditional bias does consider the magnitude (amplitude) of the timeseries and the linear

220 ~~trend (Goddard et al., 2013). The MSSS is thus a skill measure that depends on the correlation minus a penalisation for the (conditional) bias. The~~ respectively.

A positive MSSS value indicates that PRED is more accurate than HIST while a negative value indicates the opposite, however caution is recommended for its interpretation as MSSS is not symmetric around zero and positive and negative values of the same magnitude are not comparable. The statistical significance of the MSSS is estimated using a random walk test  
225 following the methodology developed by DelSole and Tippett (2016).

~~To evaluate the predictive skill of~~ The terms in brackets in Equation 1 represent the conditional bias of the predictions (CB). We can therefore rewrite the numerator of Eq. 1 in terms of the difference between the squared ACC values in PRED and HIST we compare them to a persistence forecast. The persistence forecast was computed by taking the monthly or annual anomaly at the time of initialisation and persisting it for all forecast times (e.g. Yeager et al., 2018). Therefore the correlation skill for the  
230 ~~persistence forecast is equivalent to a lag auto-regressive model~~ ( $\Delta ACC$ ), and the difference between the respective squared CB values ( $\Delta CB$ ):

$$MSSS(P, H) = \frac{[ACC_P^2 - ACC_H^2] - [CB_P^2 - CB_H^2]}{1 - ACC_H^2 + CB_H^2} = \frac{\Delta ACC^2 - \Delta CB^2}{1 - ACC_H^2 + CB_H^2} \quad (2)$$

These final two terms in the numerator, which compare different characteristics of the forecasts, will be later used to aid the interpretation of the MSSS results. While the correlation term focuses on phase variability disregarding the signal amplitude, the conditional bias term reflects both the magnitude (amplitude) of the respective timeseries as well as their linear trends (Goddard et al., 2013).  
235

The spread-error-ratio (SER; Ho et al., 2013) has been used in addition as an indicator of the forecast reliability, which is defined as the ratio between the mean intra-ensemble standard deviation and the root-mean-square error of the forecast ensemble mean. When the SER is greater (lower) than one, the ensemble is over-dispersed (under-dispersed) and the predictions  
240 will be under-confident (over-confident).

For data retrieval, loading, processing and calculation of verification measures 'startR' and 's2dverification' (Manubens et al., 2018) R libraries have been used, both developed at the BSC.

## 2.4 Climate Indices and Diagnostics

~~Several climate indices are used to assess the ability to predict both global and regional multi-annual variability.~~

245 Global mean surface temperature (GMST) is derived by blending SST over ocean and SAT temperatures over land. This allows for a consistent comparison with the aforementioned observational datasets NASA GISTEMPv4 and HadCRUT4. The use of GMST over global surface air temperature at 2m height (GSAT) is particularly favourable when assessing long-term climate trends ~~(e.g., Richardson et al., 2018)~~ (e.g. Richardson et al., 2018).

In the Pacific ocean, to distinguish between seasonal-to-interannual and decadal variability, we look at the ENSO and IPO  
250 respectively. For ENSO we use the NINO3.4 index, which is defined as the area weighted average over the region: 5°N-5°S and 170°W-120°W. For the IPO we use the Tripolar Pacific Index (Henley et al., 2015), which corresponds to the average of SST

anomalies over the central equatorial Pacific (region 2: 10°S–10°N, 170°E–90°W) minus the average of the SST anomalies in the Northwest (region 1: 25°N–45°N, 140°E–145°W) and Southwest Pacific (region 3: 50°S–15°S, 150°E–160°W). To describe the decadal variability over the Atlantic basin, we use the AMV definition from Trenberth and Shea (2006). The AMV is calculated as the spatial average of SST anomalies over the North Atlantic (~~Equator-60°E-100°E~~Equator-60°E-100°E and 80°–0°W) minus the spatial average of SST anomalies averaged from 60°S to 60°N (Trenberth and Shea, 2006; Doblus-Reyes et al., 2013). In addition to the AMV, we also compute the Subpolar North Atlantic (50–65°N, 60–10°W) ocean heat content in the upper 300m (referred to as SPNA-OHC300 hereafter). Since the IPO, AMV and SPNA-OHC300 are decadal modes of variability, we filter out part of the interannual variability by considering four-year temporal averages along the forecast time (i.e. forecast years 1-4, 2-5, 3-6...) for these indices.

Likewise, density has been computed using the International Equation of State of seawater (EOS-80) referred to the level of 2000 dbar (sigma-2), a level that represents better the deep water properties. In addition, the contributions of temperature (sigma-T) and salinity (sigma-S) to density were computed using the thermal expansion and haline contraction coefficients, which were themselves estimated as the density change (in the EOS-80 equation) associated with a small increase in temperature (0.02 °C) and salinity (0.01 psu), respectively.

All ocean diagnostics have been computed using 'Earthdiagnostics', a python-based package developed at the BSC.

### 3 Results

#### 3.1 Characterising the Predictive Capacity of Surface Temperature

##### 3.1.1 Global Mean Surface Temperature

First we compare the predicted GMST evolution for different forecast periods (Figure 1), estimated by combining SAT over land and SST over the ocean (cf. Data and Methodology). PRED reproduces the observed variability and shows very high ACC skill values: 0.96, 0.97 and 0.95 for forecast year one, years 1-5 and years 6-10 respectively (similar values are obtained when comparing to other observational products like HadCRUT4). As expected, HIST does not capture most of the interannual variability (Figure 1), since it is largely of internal origin and therefore averages out by construction. Nevertheless, HIST shows very high skill (0.94, 0.96 and 0.95 for forecast year one, years 1-5 and years 6-10 respectively) associated with the global warming trend and the cooling episodes in response to the volcanic eruptions of Agung (1963), El Chichon (1982) and Pinatubo (1991). The differences in ACC skill between PRED and HIST are not statistically significant, indicating that the high skill of PRED is mainly associated with the external forcings. When the simulations are detrended (a simple attempt to remove the warming ~~trend~~signal) PRED shows higher ACC skill values than HIST, revealing the benefit of the initialisation, especially for the earlier forecast years (ACC values 0.75 and 0.5 in forecast year one for PRED and HIST, respectively).

Comparing the intra-ensemble spread of PRED and HIST (shown by the Box-Whisker for PRED and shading for HIST in Figure 1) shows that PRED has considerably smaller spread even on the longer forecast times. For example the mean intra-ensemble standard deviation of PRED is 0.05K while for HIST is 0.20K for the first forecast year. This is probably due to

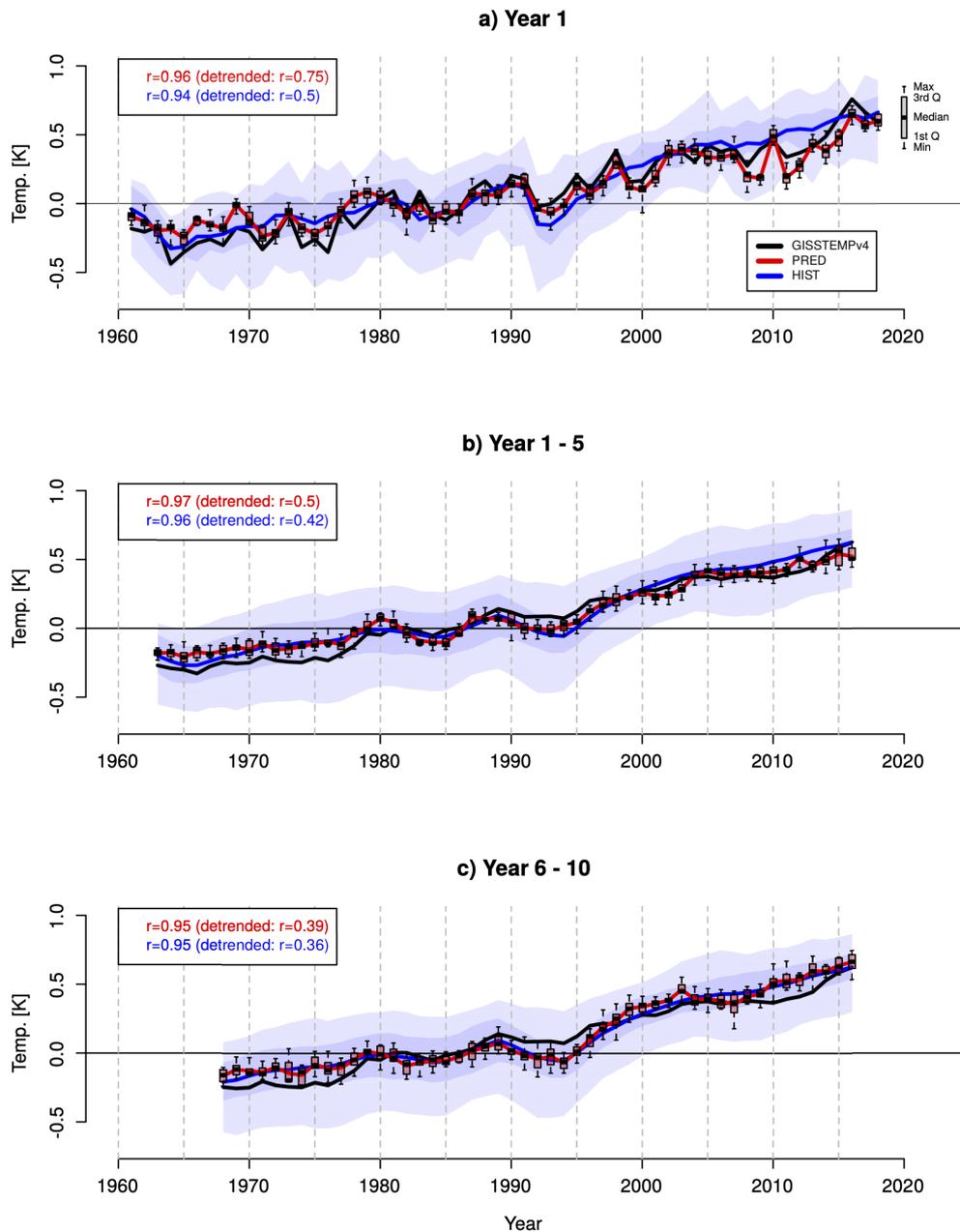
the initialisation of PRED, which puts in phase the simulated internal and observed variability and may also help to correct systematic errors in model response to external forcing (e.g., [Doblas-Reyes et al., 2013](#)) (e.g., [Doblas-Reyes et al., 2013](#)). This difference in spread remains equivalent-comparable when the ensemble size of HIST is reduced to 10 members, i.e. the same ensemble size of PRED.

### 3.1.2 Added Value of Initialisation at the Regional Scale

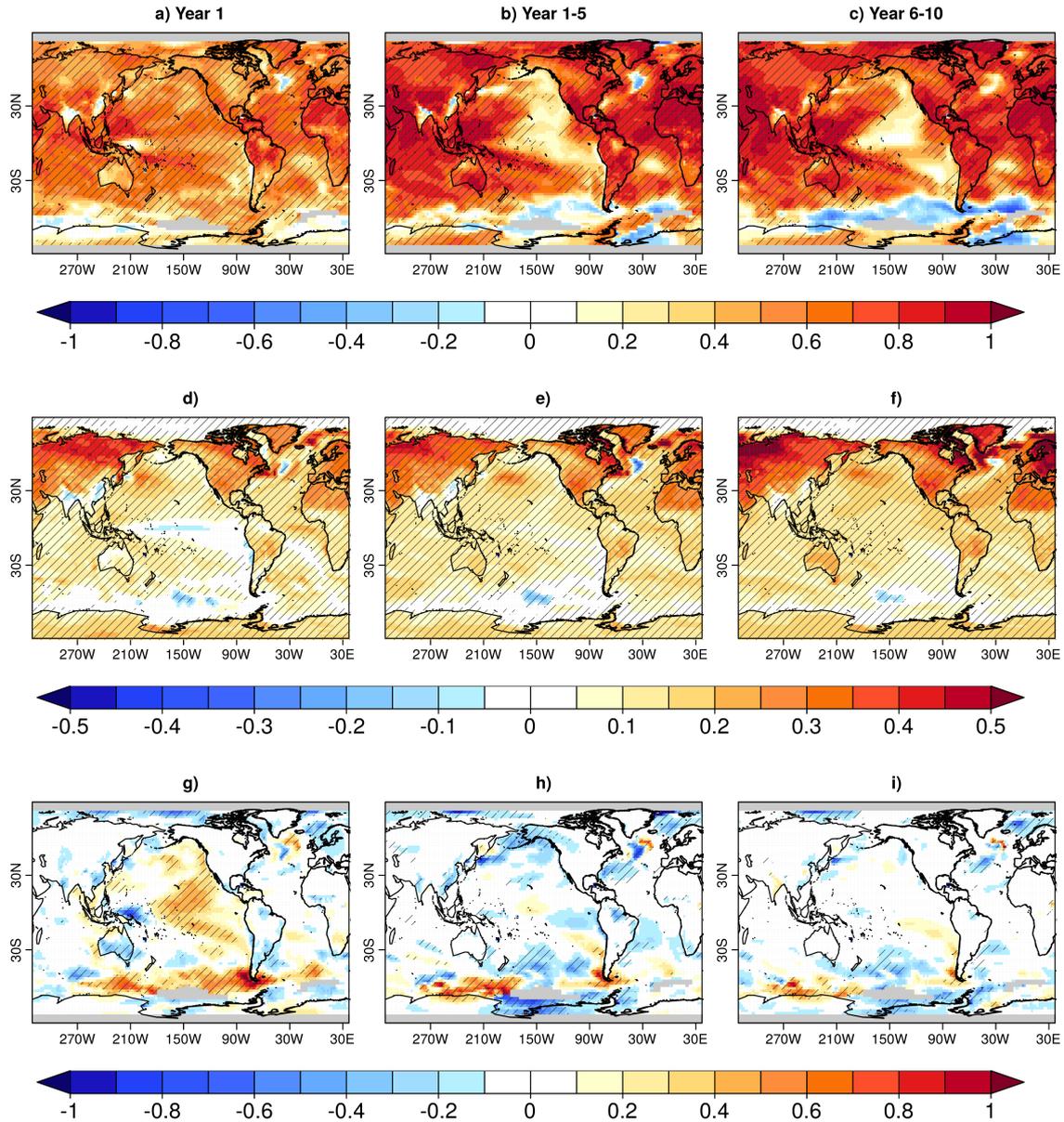
At the regional level, PRED shows high skill at predicting surface temperature at different forecast ranges (Figure 2a-c), as expected from the presence of long-term warming trends (Figure 2d-f). In the first forecast year, most regions show significant skill ([2a](#)), with a few exceptions such as the central Subpolar North Atlantic, ~~and~~ some regions of Asia, Australia and the Southern Ocean, where the simulated trends are small and mostly not statistically significant (Figure 2d). ~~In~~ By contrast, the Eastern Pacific shows no significant trends but does show significant skill in the first forecast year associated with the initialisation of ENSO. On longer time-scales (forecast ranges 1-5 and 6-10 years) PRED also shows significant skill in many regions worldwide with greater ACC values compared to forecast year one (Figure 2b and c). This is probably a consequence of considering 5-year averages for validation, which reduces the influence of the unpredictable part of interannual variability. There is however an important degradation of the skill in some regions for these forecast ranges, in particular in the Eastern Tropical Pacific where the model might not be representing the correct ENSO low-frequency-low-frequency variability, and in the North Pacific where generally low levels of skill have been related to model biases in ocean mixing processes (Guemas et al., 2012). Comparing the forecast periods for years 1-5 and 6-10, two major differences are apparent. First, in the Southern Ocean, skill degrades dramatically with forecast time, which is probably associated ~~to~~ with the development of a warm bias due to the incorrect representation of cloud feedbacks in the region (Hyder et al., 2018). Second, the central Subpolar North Atlantic seems to exhibit an opposite change in skill, from negative ACC values during the first 5 forecast years to positive but insignificant ACC values in the five following years, which might reflect the recovery from an initial shock-adjustment that affects the North Atlantic. This shock-adjustment might be responsible for the strong negative trends over the regions in forecast years 1-5, which are substantially reduced in forecast years 6-10 (Figure 2e,f). This will be further discussed in 3.3.

To determine whether there is a benefit of ~~the initialisation~~, initialisation on the surface temperature skill we compute the difference in ACC between PRED and HIST (Figure 2g-i), as well as the MSSS of PRED ~~considering using~~ HIST as the baseline reference forecast (Figure 3a-c). ~~To Also, to~~ aid in the interpretation of the MSSS, ~~we also show the ACC difference between values, we show the two terms that determine its sign, the first being the difference between the squared ACC values of PRED and HIST (Figure 3d-f) and the conditional bias difference, and the second being the difference between their squared conditional biases~~ (Figure 3g-i). The color scales in all panels have been adjusted so that red colors represent a contribution to improved skill from initialization (i.e. the colorbar was inverted for the conditional bias plots).

In the first forecast year, ~~the MSSS shows both ACC differences (Figure 2g) and MSSS (Figure 3a) show~~ added value from initialisation ~~mostly~~ in the Pacific Ocean, the neighboring region of the Southern Ocean, ~~and~~ the eastern Subpolar North Atlantic ~~and the Atlantic Subtropical areas (Figure 3a). An inspection of the ACC and conditional bias difference maps (PRED-HIST) reveals that for~~, with MSSS also showing positive impact of initialization in the Subtropical Atlantic. The MSSS



**Figure 1.** GMST annual mean anomaly timeseries (K) for PRED, HIST (historical+SSP2-4.5) and GISTEMPv4 for a) forecast years 1, b) forecast years 1-5 and c) forecast years 6-10. The anomalies cover the period 1961-2018 and are referenced to the 1971-2000 mean. Multi-year means (panels b and c) are plotted on the central year (e.g.,  $\bar{y}=2000$  represents the values from 1998-2002 in b and c). For a fair comparison with observations, PRED and HIST have been masked where and when GISTEMPv4 has missing values. The intra-ensemble spread for PRED and HIST is represented by the box-and-whisker plots and shading respectively. The ACC for PRED and HIST is shown in the top left part of each panel, including in brackets the ACC after removing a linear trend from the timeseries. All ACC values are statistically significant at the 95% level.



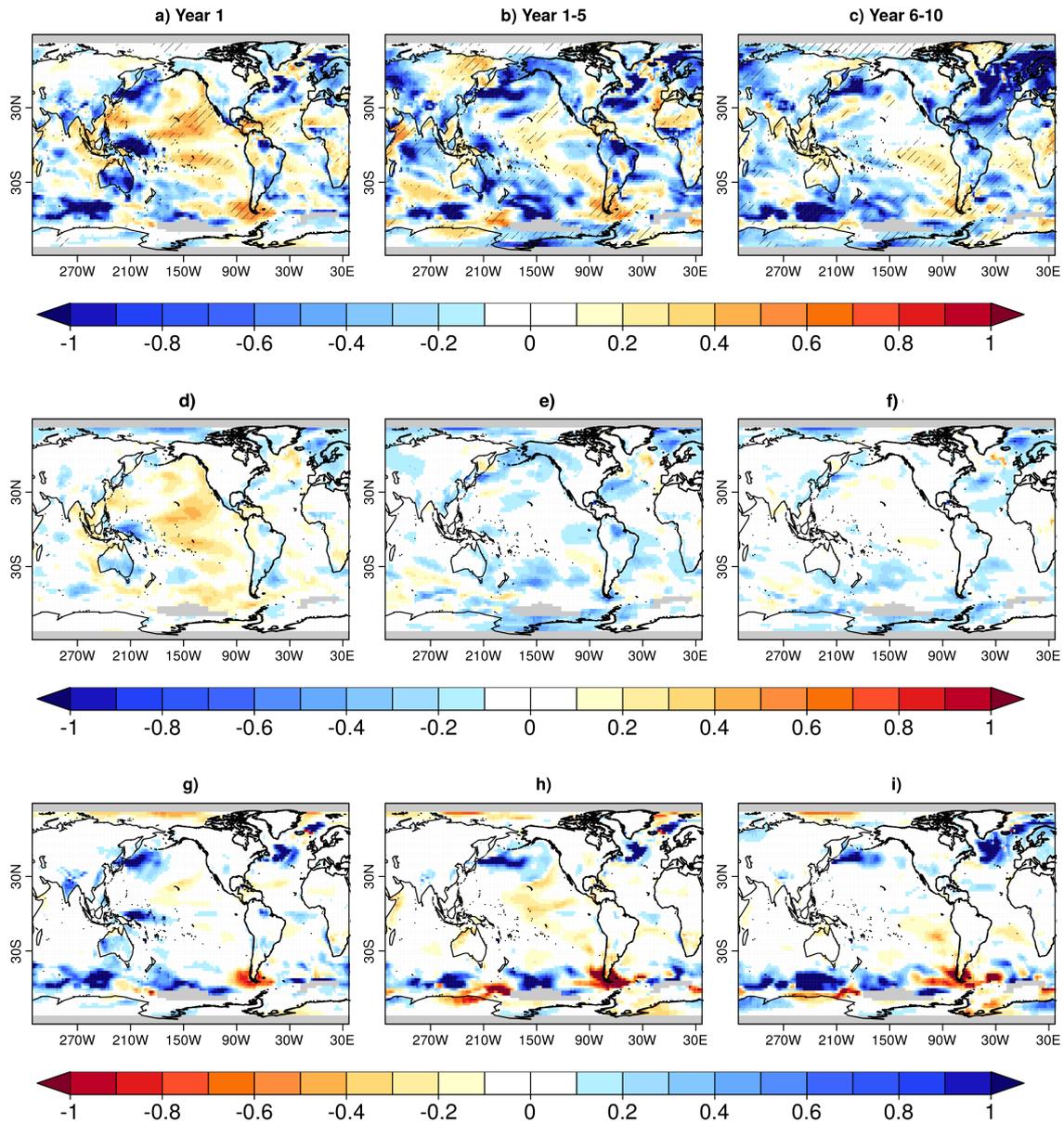
**Figure 2.** Top row: Anomaly correlation coefficient (ACC) for the annual surface temperature (SAT and SST blend) in PRED for a) forecast years one a) 1, b) forecast years 1-5 and c) forecast years 6-10. The ACC is computed between the model ensemble mean of the blended SAT (over land) and SST (over the ocean) fields and GISTEMPv4. Hatching indicates areas that are significant at the 95% confidence level. Missing Grid points with missing values in observations are masked in grey. Bottom-Middle row: d-f) surface temperature linear trends [in K/decade] for-in PRED for d) the same forecast years one, e-in the top row. Bottom row: g-i) forecast years 1-5-ACC difference between PRED and f) HIST in the same forecast years 6-10as above. Hatching-In all rows, hatching indicates a trend that is statistically different from zero-significance at the 95% confidence level. Both ACC and trends are computed for annual mean values of all available years for the respective forecast period referenced to the 1970-2018 climatology.

terms reveal that in the first forecast year, the positive MSSS values (~~which are~~ indicative of added value of initialisation) are mostly associated with ~~increased ACC values in PRED (i.e. the hindcasts reproduce better the observed variability)~~the squared  
320 ACC term (Figure 3d), while the ~~Subtropical Atlantic is the only region where a positive MSSS is associated with a reduction in the conditional bias (i.e. a better representation of the linear trend).~~

~~MSSS values squared conditional biases contribute mostly to negative MSSS skill, in particular over the whole SPNA region and the Northern North Pacific (Figure 3g). These negative contributions of the squared conditional bias term are also present at forecast years 1-5, in which both the ACC and MSSS differences become predominantly negative~~for forecast years 1 to  
325 5 (Figure 3b, e,h) and, supporting better predictive capacity in HIST than in PRED and therefore indicating that only a few ~~regions benefit from the initialisation~~areas, such as the ~~Tropical Northeastern SPNA and the tropical Pacific (Figure 3b)~~due to a reduction of the conditional bias (Figure 3h). ~~The rest of the Pacific and most of the Atlantic basin show negative MSSS values (Figure 3b). In these regions, PRED has generally lower ACC values than HIST, and a larger conditional bias, are benefiting from initialisation. While the skill improvement in the Northeastern SPNA is clearly due the higher ACC in PRED (Figure 2h and Figure 3e), the improvement in the Tropical Pacific is due to a reduction of the squared conditional bias of PRED with respect to HIST (Figure 3e,h). At longer forecast times (6-10 years), positive added value from initialization remains very limited. Positive MSSS values are almost exclusive to the Eastern tropical South Pacific and tropical South Atlantic, where the conditional bias of the forecast is reduced (Figure 3i), indicating a better representation of the long-term trend Pacific and the Southeaster Atlantic, due to a reduction of the squared conditional bias in PRED with respect to HIST .By contrast, negative MSSS values span across most of the North Atlantic, and reach the Eurasian continent (Figure 3c), mostly associated with a larger (Figure 3i). In the Northeastern SPNA, although the ACC is greater in PRED (Figure 2i and Figure 3f), the squared conditional bias in PRED compared to HIST is considerably larger than in HIST, leading to very negative MSSS values (Figure 3i). The ACC differences are also predominantly negative in the Subpolar North Atlantic, a region in which many prediction systems exhibit prolonged added value of initialisation (e.g., Robson et al., 2018; Yeager et al., 2018). This~~  
330 suggests, which suggest that some regional key physical processes (e.g. ,the gyre and overturning ocean circulations) might not be well represented in EC-Earth3 predictions.

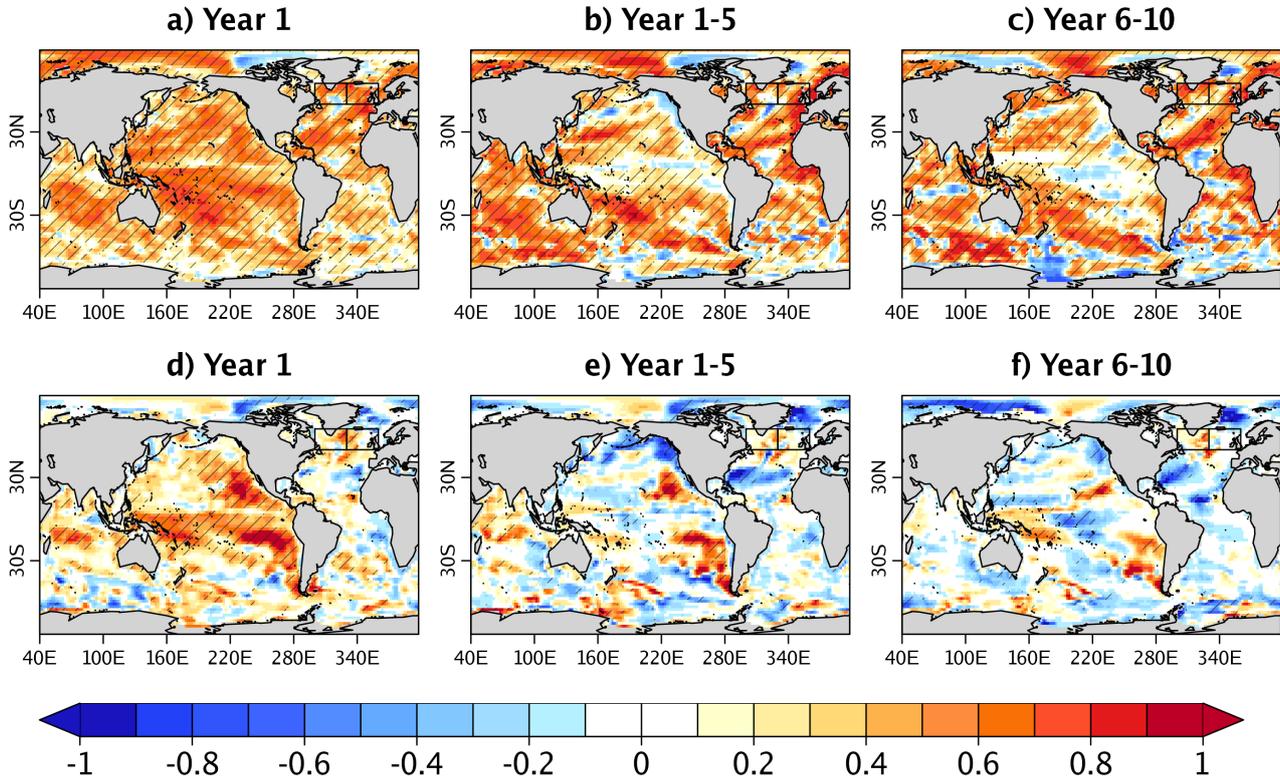
To complement the analysis of surface temperature, we also consider the upper ocean heat content, a quantity that better represents the thermal inertia of the ocean and a source of decadal variability and predictability for the surface climate (e.g., Meehl et al., 2014; Yeager et al., 2018)(e.g. Meehl et al., 2014; Yeager et al., 2018). As previously shown for sur-  
345 face temperature, in the first forecast year high and significant ACC values are obtained in all major basins for the ocean heat content in the upper 300m (referred to as OHC300 hereafter; Figure 4a). A region of negative skill values is evident over the central Subpolar North Atlantic as for the surface temperatures (Figure 2a). Forecast years 1-5 and 6-10 show that the skill in the Tropical and Eastern Pacific is lost as is for ~~the~~ some regions in the Atlantic and Pacific sectors of the Southern ocean (Figure 4b and c). As for surface temperature, the skill in the central Subpolar North Atlantic moderately improves in forecast  
350 years 6-10 with respect to forecast years 1-5.

Comparing the ACC difference between the PRED and HIST ensembles reveals that the initialisation increases the forecast skill of OHC300 in the eastern Subpolar North Atlantic in all the three forecast times (and ranges) considered (Figure 4d-f). A



**Figure 3.** Annual Top row: MSSS of the annual mean surface temperature (SAT and SST blend) skill comparison and bias differences between in PRED and using HIST, computed with the GISTEMPv4 observations. The top row shows the MSSS of PRED with the HIST as the reference forecast (see section 2.3) for forecast years a) forecast year one, b) forecast years 1-5 and c) forecast years 6-10. Hatching indicates where the MSSS is significant using a random walk test (see section 2.3). The middle row shows its sign is determined by the ACC difference between PRED the terms in the second and HIST for d third rows (see Equation 2) forecast year one, e) forecast years 1-5 and f) forecast years 6-10. Hatching indicates a significant Middle row: difference in ACC between the squared ACC values in PRED and HIST based on for the methodology developed by Siegert et al. (2017) (see section 2.3) at the 95% confidence level same forecast years. The bottom Bottom row shows the difference in absolute values of between the squared conditional bias between biases in PRED and HIST for g) the same forecast year one, h) forecast years 1-5 and i) forecast years 6-10. Note that the color scale is reversed with respect to the ones in other rows so that positive values contribute to improved skill from initialization. GISTEMPv4 is used as the observational reference for all calculations. Annual mean anomalies are computed masking PRED and HIST with the GISTEMPv4 missing values (masked in grey) and

result that is consistent with other forecast systems (e.g., Robson et al., 2018; Yeager et al., 2018) (e.g. Robson et al., 2018; Yeager et al., 2018). The Pacific Ocean shows significantly improved skill from initialisation basin wide in the first forecast year (i.e. ENSO), but for forecast years 1-5 and 6-10 the added value of initialisation is limited to parts of the Eastern subtropical Pacific and the Western tropical Pacific. The initialisation improves the skill in most of the Indian Ocean at all forecast times considered (although it is not statistically significant for the forecast range 6-10), consistent with previous studies showing the high skill of decadal predictions in this region (Guemas et al., 2013).



**Figure 4.** Upper 300m OHC ACC of PRED computed with the EN4 observations for a) forecast year one, b) forecast years 1-5 and c) forecast years 6-10. The impact of initialisation is shown as the difference in ACC between PRED and HIST for d) forecast year one, e) forecast years 1-5 and f) forecast years 6-10. In panels a-c, the hatching indicates the statistical significance of the correlation at the 95% confidence level. For panels d-f hatching indicates the regions where the difference in correlation between HIST and PRED are statistically significant at 95% confidence level. The black boxes in the North Atlantic delimit the region over which the SPNA indices have been defined. It also includes the central boundary used to distinguish between its western and eastern side.

### 3.2 Skill for the Main Ocean Modes of Variability

360 We further evaluate the predictive capabilities of the EC-Earth3 PRED experiment by considering the skill for predicting several modes of interannual to decadal variability (Figure 5). In the Pacific Ocean, ENSO is the main mode of climate variability on seasonal to interannual time-scales, and can help the predictive capacity worldwide through its well-known climate impacts (e.g., McPhaden et al., 2006; Yuan et al., 2018)(e.g. McPhaden et al., 2006; Yuan et al., 2018). Figure 5a shows that PRED captures the year-to-year evolution of the observed ENSO, while HIST does not since ENSO events are out of  
365 phase across the HIST ensemble and average out. This is confirmed by the high ACC values during the first four forecast months in PRED (ACC>0.9), which are followed by a typical loss of skill that many dynamical forecast systems experience during the spring season (i.e. the spring barrier; Webster and Yang, 1992) and by a recovery in summer through the next winter, in which ACC values remain positive and significant (Figure 5b). Added value for ENSO due to initialisation is evident up to the second forecast year, as indicated by the positive MSSS values and the statistically significant difference  
370 in ACC values between PRED and HIST (Figure 5b). The ~~spread-error-ratio, however, lack of skill in HIST is expected, because ENSO is barely influenced by the external forcings and the ensemble mean averages the phases of the individual members out. In the predictions, the spread-error-ratio~~ reveals that the ENSO predictions ~~are overconfident for the first few months~~tend to be overconfident in the first two years, in contrast to HIST which tends to be underconfident. On decadal time-scales, the dominant mode of climate variability in the Pacific basin is the IPO. Figure 5e shows that neither PRED  
375 nor HIST are capable of skillfully predicting it, a lack of skill that has been documented in many other prediction systems (e.g., Doblas-Reyes et al., 2013)(e.g. Doblas-Reyes et al., 2013). Nonetheless, initialisation does seem to improve the reliability of the IPO ~~5f~~(Figure 5f).

In the Atlantic Ocean, the AMV is the dominant mode of decadal climate variability, and has been linked to several climate impacts over Europe, North America and the Sahel (Zhang and Delwoth, 2006; Sutton and Dong, 2012; Ruprich-Robert et al.,  
380 2017, 2018) and to Atlantic tropical cyclones (e.g. Caron et al., 2015, 2018). Both PRED and HIST are capable of skillfully predicting the AMV, and do better than a persistence forecast (except for the forecast range 1-4 years in HIST), as shown by the ACC (Figure 5h). PRED however, is consistently better than HIST as shown by the MSSS, even though the ACC differences are not statistically significant. The spread-error-ratio shows that initialisation improves the reliability of the AMV predictions at all forecast ranges (Figure 5i), since the historical simulations are overdispersive probably due to excessive intra-ensemble  
385 spread as previously described in section 3.1.1.

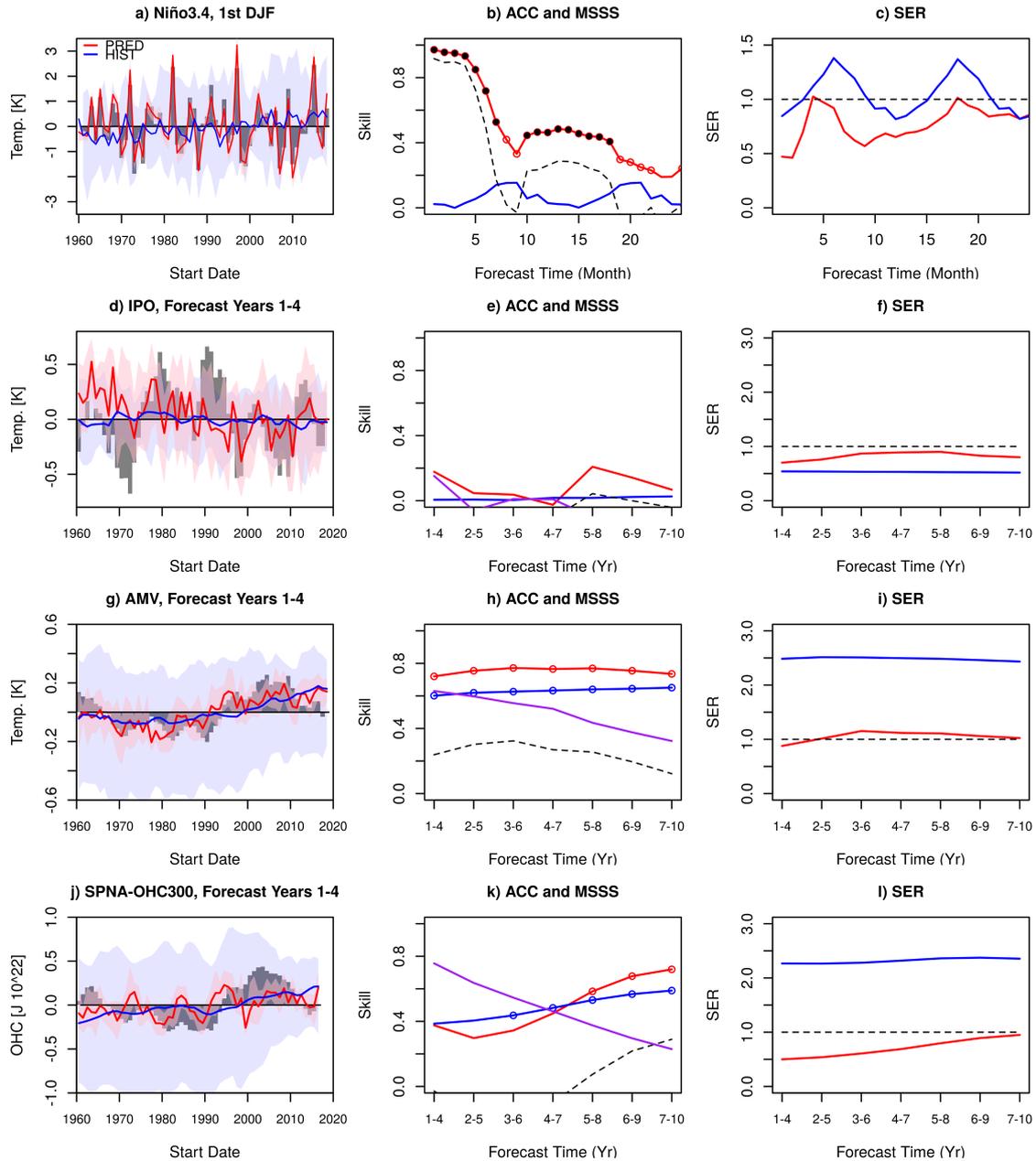
Since the Subpolar North Atlantic has been shown to be a region where forecast systems exhibit skill on decadal times scales, we analyse the SPNA-OHC300 index (see Section 2.4). ~~PRED~~As for the AMV, the initialisation improves the reliability of the PRED and shows the HIST intra-ensemble spread may be too large and therefore under-confident (Figure 5l). In terms of predictive capacity, PRED exhibits a lack of skill ~~in this~~for the SPNA-OHC300 index up to forecast range 4-7, with significant  
390 ACC values emerging for longer forecast ranges, coinciding with the time in which the system outperforms persistence. ACC values in the HIST ensemble (which are not statistically different from those in PRED) also increase with forecast time. In HIST, this is due to the fact that the skill for each forecast range is computed over a different verification period, the same one

used for PRED, which used for each forecast range the longest period available. For the longest forecast ranges [e.g. 7-10], the first start dates cannot be used, excluding some of the earliest years [e.g. 1960-1966], for which the warming trend was less prominent, thus producing an artificial increase in skill for longer forecast times. Repeating the calculations for PRED over a common verification period to all forecast ranges (i.e. 1970-2018) reveals that lower skill values are still present for the first forecast years (see Supplementary Figure 1), which suggests that the differences in skill with forecast time are not due to differences in the verification period but to other causes, ~~for example some potential initialisation shocks from which the system is recovering some years later. This possibility.~~ If we determine separately the skill in OHC300 in the Western and Eastern SPNA (see Supplementary Figure 2), we can then see that in the Western SPNA the skill in PRED is initially poor and then gradually increases reaching significant values in the last forecast years, while the Eastern SPNA maintains a high and relatively stable predictive skill. The poor initial predictive skill in the western SPNA might arise from a potential initialisation shock, a possibility that is discussed in the next sub-section. ~~As for the AMV, the initialisation improves the reliability of the PRED and shows the HIST intra-ensemble spread may be too large and therefore under-confident (Figure 5)~~ The final skill recovery in the region might be related to the arrival of OHC anomalies from the eastern SPNA, which are slowly advected by the mean gyre circulation.

### 3.3 Understanding the Limited Predictive Skill in the Subpolar North Atlantic

In the previous section we have shown an overall detrimental effect of initialisation in the EC-Earth3 predictions over some regions of the North Atlantic at all forecast ranges (Figure 3), leading to lack of predictive skill in the specific case of the central Subpolar North Atlantic as shown by the ACC maps of surface temperature (Figure 3) and upper ocean heat content (Figures 4 and 5). Decadal variability in the Subpolar North Atlantic is highly influenced by the changes in the ocean circulation, both meridional and barotropic (e.g., Ortega et al., 2017)(e.g. Ortega et al., 2017). To understand the role of the ocean circulation we analyse the evolution of the Atlantic Meridional Overturning Circulation at 45°N (defined as the overturning stream-function value at 45°N and at 1000m depth; referred to as AMOC45N hereafter) and North Atlantic Subpolar Gyre Strength Index (NASPG, defined as the regional average of the barotropic stream-function in the Labrador Sea [52-65°N, 58-43°W], multiplied by minus one to make the values positive so as to aid the comparison) in PRED and HIST (Figure 6a and b). Additionally we include the ocean-only reconstruction from which the initial conditions are obtained (referred to as RECON hereafter) to determine how the predictions depart from the initial conditions. Actual model values are used to illustrate how the forecast drift develops. The mean forecast drift is also shown for completeness, estimated as the climatological value as a function of forecast time (Figure 6c and d). Figure 6 shows that decadal changes in the AMOC and NASPG are highly correlated (e.g. R=0.8 in RECON), a ~~feature relationship~~ that has been shown in previous studies (Ortega et al., 2017).

Comparing PRED and RECON allows us to identify several interesting features. In the first forecast year the predicted AMOC45N is of equal value with respect to RECON, while for the NASPG index the predicted values tend to be weaker (Figure 6). As the ~~forecast forecasts~~ evolve and the model transitions towards its free-running attractor both indices diverge from RECON and experience a pronounced weakening. By forecast year ten, the indices in PRED reach a weaker mean state than in HIST (~~black-green~~ dashed lines in Figure 6c and d, respectively). These differences between PRED and HIST suggest

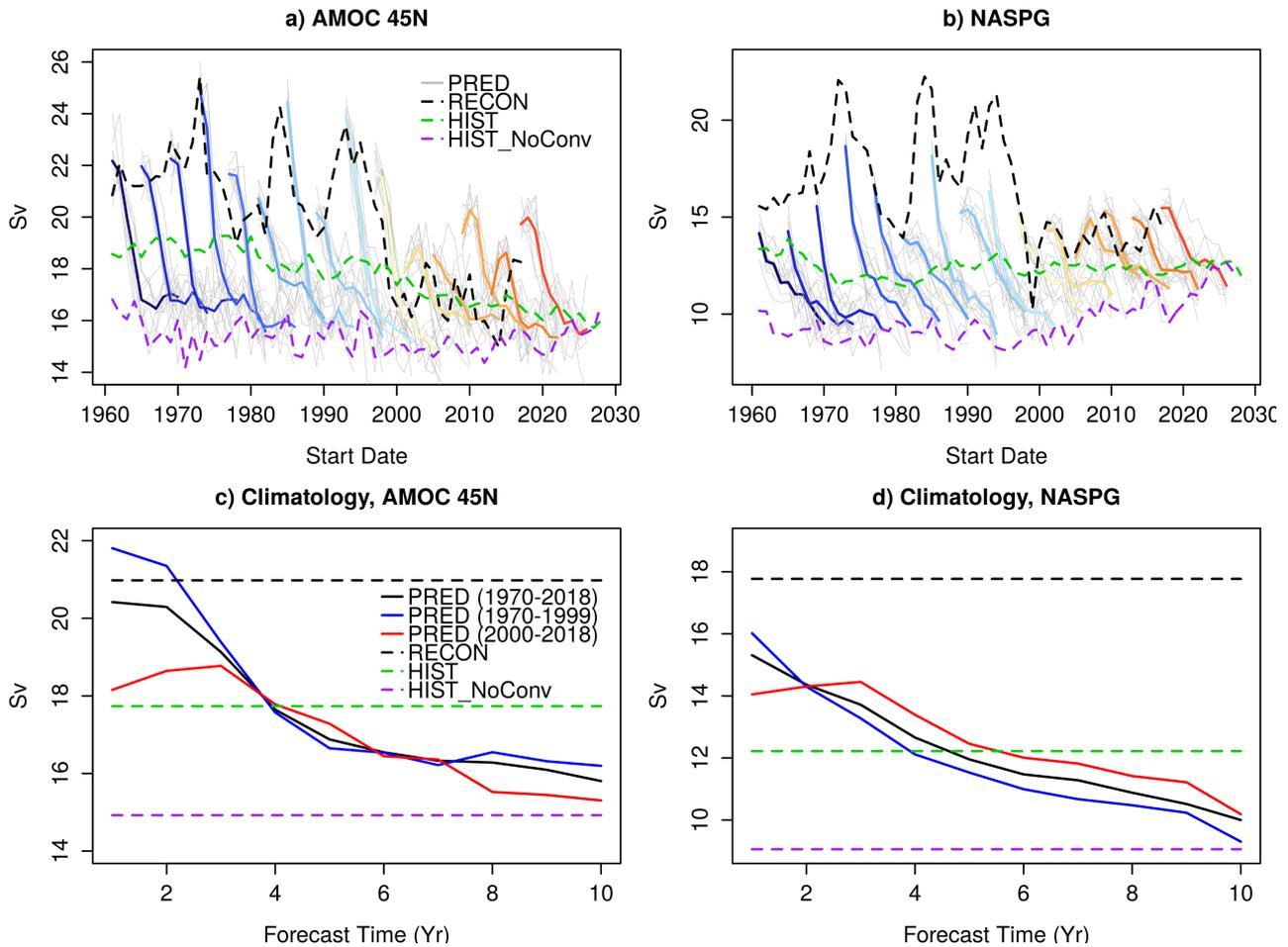


**Figure 5.** Skill of the selected modes of ocean variability: a-c) ENSO (Niño3.4 index), d-f) IPO, g-i) AMV and j-l) SPNA-OHC300. The first column shows the observed (grey bars) and predicted (PRED in blue, HIST in red) time series of the indices. The ensemble means are represented with lines and the ensemble spread with coloured envelopes. The first column shows the ENSO index for the first winter (DJF), while for the other indices the average of the first 4 forecast years is shown. The second column shows the ACC of PRED (blue) and HIST (red), the MSSS of PRED considering HIST the baseline prediction (black dashed line) and the ACC of a persistence forecast (purple). Statistically significant ACC values (at the 95% confidence level) are shown as empty circles. ACC differences that are statistically significant (at the 95% confidence level) between the PRED and HIST are shown as filled circles. The third column shows the spread-error-ratio of PRED (blue) and HIST (red).

either that the forecasts in PRED need to be run for longer to reach its attractor (~~HIST~~) (e.g., ~~Sanchez-Gomez et al., 2016~~) i.e. HIST (e.g. Sanchez-Gomez et al., 2016), or that more than one model attractor exists.

For both indices, we also note a clear difference in the way the forecast transitions to the model attractor before and after  
430 year 2000. In the first 30 start dates, the AMOC45N and NASPG in PRED start at stronger values than HIST (c.f. RECON  
values in Figure 6a and b), and the individual predictions exhibit a fast decline that surpasses the HIST mean state. In year  
1995 of RECON, both indices experience a sharp decrease and eventually stabilise around a substantially lower mean state, a  
transition that has been shown to be partly predictable in previous studies (Robson et al., 2012; Yeager et al., 2012; Msadek  
et al., 2014). Due to this ~~lower-weaker~~ initial state, all predictions after the year 2000 start much closer to the HIST mean  
435 state ~~and the PRED attractor~~. In consequence, for this period the drift in PRED is smoother. The fact that there are two distinct  
periods in which the model drifts in different ways (Figures 6c and d) may compromise the applicability of the drift correction  
methods used to compute the forecast anomalies, which assumes a stationary forecast drift. This is particularly evident for  
the AMOC45N, which shows important differences in the PRED climatologies during the ~~first three~~ three first forecast years  
when the climatologies are computed for the time periods preceding and following the year 2000 (Figure 6c red and blue lines,  
440 respectively). ~~These period-sensitive climatological differences are less pronounced for the NASPG~~ Thus, in the case of the  
AMOC and only within the three first forecast years, applying the standard mean drift correction leads to an underestimation  
of its intensity over the first period, and an overestimation over the second period within the three first forecast years. In light  
of this problem refining the current drift correction techniques to account for this sensitivity to the period and/or initial state  
considered, or exploring other alternative statistical drift models to recalibrate the predictions (e.g. Nadiga et al., 2019) might  
445 help to better estimate the true prediction skill. Interestingly, other variables like the NASPG seem to be less affected by  
non-stationary drifts (Figure 6d).

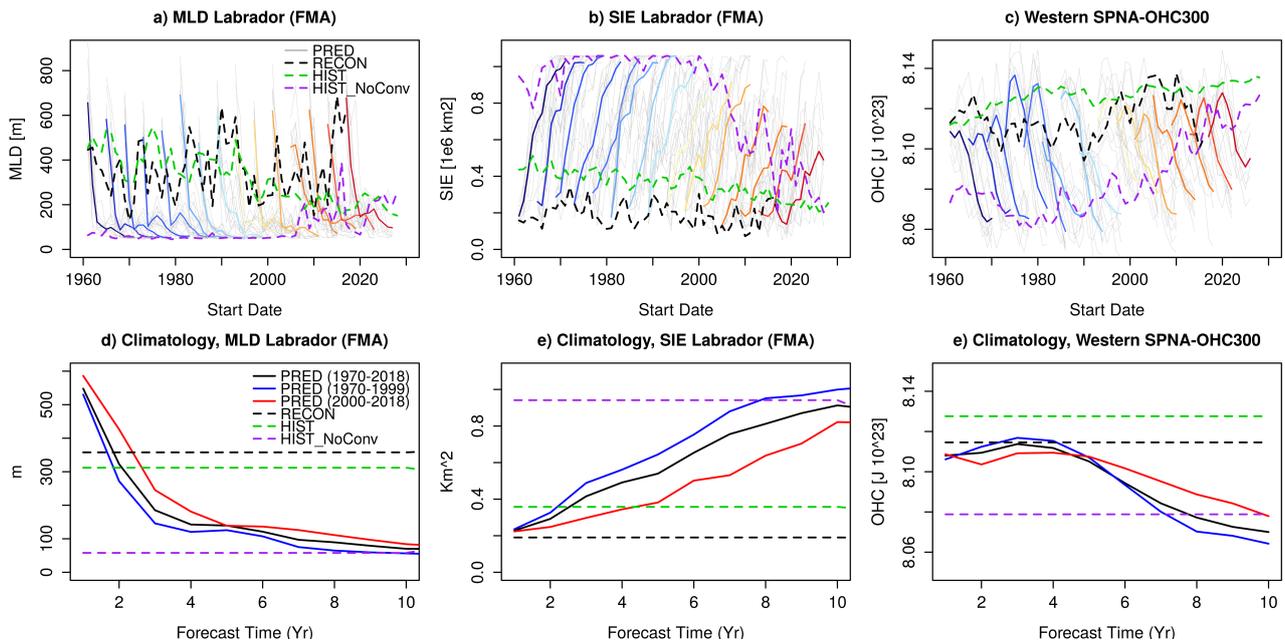
To understand why the AMOC45 and NASPG are not stabilising in the predictions around the mean HIST state, we focus  
on the Labrador Sea. The Labrador Sea is a key region of deep water formation, in which climate models show limitations to  
represent realistic oceanic convection, which can happen too often, too deep, or can be completely absent in some cases (Heuzé,  
450 2017). Figure 7a shows the mixed layer depth (MLD) evolution in the Labrador Sea, a proxy for the convection activity in this  
region. The MLD index is computed as the average of February-March-April, the months with the deepest mixing. In PRED,  
MLD systematically collapses within the first three forecast years, which is in stark contrast with the typical behaviour in  
the HIST ensemble, in which deep convection happens regularly. In the HIST ensemble mean, Labrador convection remains  
active throughout the whole period although it exhibits a long-term weakening trend, consistent with the increase in local  
455 stratification caused by the externally forced ocean surface warming. The Labrador MLD index also allows us to identify three  
HIST members with a distinct evolution from the rest, characterised by no convection during most of the historical period with  
slight increase from 2005 onward (purple line in Figure 7a). These simulations have a remarkable similarity with the state  
towards which PRED appears to be drifting. The ensemble mean of these three HIST members is also compatible with the  
AMOC45N and NASPG states at the end of the forecasts (purple lines in Figure 6), suggesting that the attractor towards which  
460 PRED ~~evolves-converges~~ is associated with a suppressed Labrador Sea convection state. Note also that in the first forecast  
year of PRED the Labrador Sea MLD is stronger than in RECON. All of the above suggests the existence of an initial ~~shock~~



**Figure 6.** Evolution of the a) AMOC45N and b) NASPG in the raw forecasts, historical ensemble and reconstruction. Ensemble mean forecasts (10 members) of PRED are shown from blue to red every 3-4 startdates, with individual ensemble members shown in grey. The ensemble mean RECON (5 members) is shown with the black dashed line. The ensemble mean of all HIST (15 members) is shown in green, and the ensemble mean of the HIST members that do not exhibit convection are shown in purple. Panels c) and d) show the PRED climatological values as a function of forecast time for the AMOC45N and NASPG, respectively. Three time periods are considered for PRED: in black the climatology is for the period of 1970-2018, in red for the period 1970-2000 and in blue for 2000-2018. The black, green, and purple dashed lines indicate the climatology computed over the 1970-2018 period for RECON, all HIST members, and HIST members that do not show convection, respectively.

adjustment in PRED, which initially boosts convection and subsequently brings the model towards a non-convective mean state.

Other key indices are also affected by the Labrador Sea convection collapse in PRED. For example, we see that sea ice grows to occupy the whole Labrador Sea as soon as convection ceases (Figure 7b and e). Like for the MLD, the sea ice extent of the HIST members with no convection is remarkably similar, while convection in the other members keeps a relatively reduced sea ice coverage. The western SPNA-OHC300 (50–65°N,60–30°W) also seems to experience an initialisation shock-abrupt initial change as shown in Figure 7c; the PRED climatological value at forecast time 1 year is lower than in the RECON climatology (Figure 7f). In forecast years 2-3, this index tends to increase, approaching the HIST mean state, which is higher than in RECON. However, this trajectory changes drastically after forecast year 3 (Figure 7f), and a quick cooling begins towards the no-convection HIST state. This sudden change could be explained by a delayed response to the convection collapse in the predictions, which is expected to drive a weakening of the SPG intensity by decreasing the density of its inner core and its associated geostrophic current (Levermann and Born, 2007). For all these indices we note again that their climatological drifts seem non-stationary (Figure 7d, e and f), and that predictions started after the year 2000 might not be well bias corrected. This, together with the effect of the strong initial shock, could explain the low (negative) predictive skill in the western SPNA region (Figures 2 and 4), a region that is key for the North Atlantic Oscillation (e.g.?) (e.g. Athanasiadis et al., 2020) and that in our model might be hampering its prediction skill, which is also low (not shown).



**Figure 7.** The same as in Fig. 6 but for the MLD in the Labrador Sea February-March-April, the SIE in the Labrador Sea during the same months and the western SPNA-OHC300 annual mean.

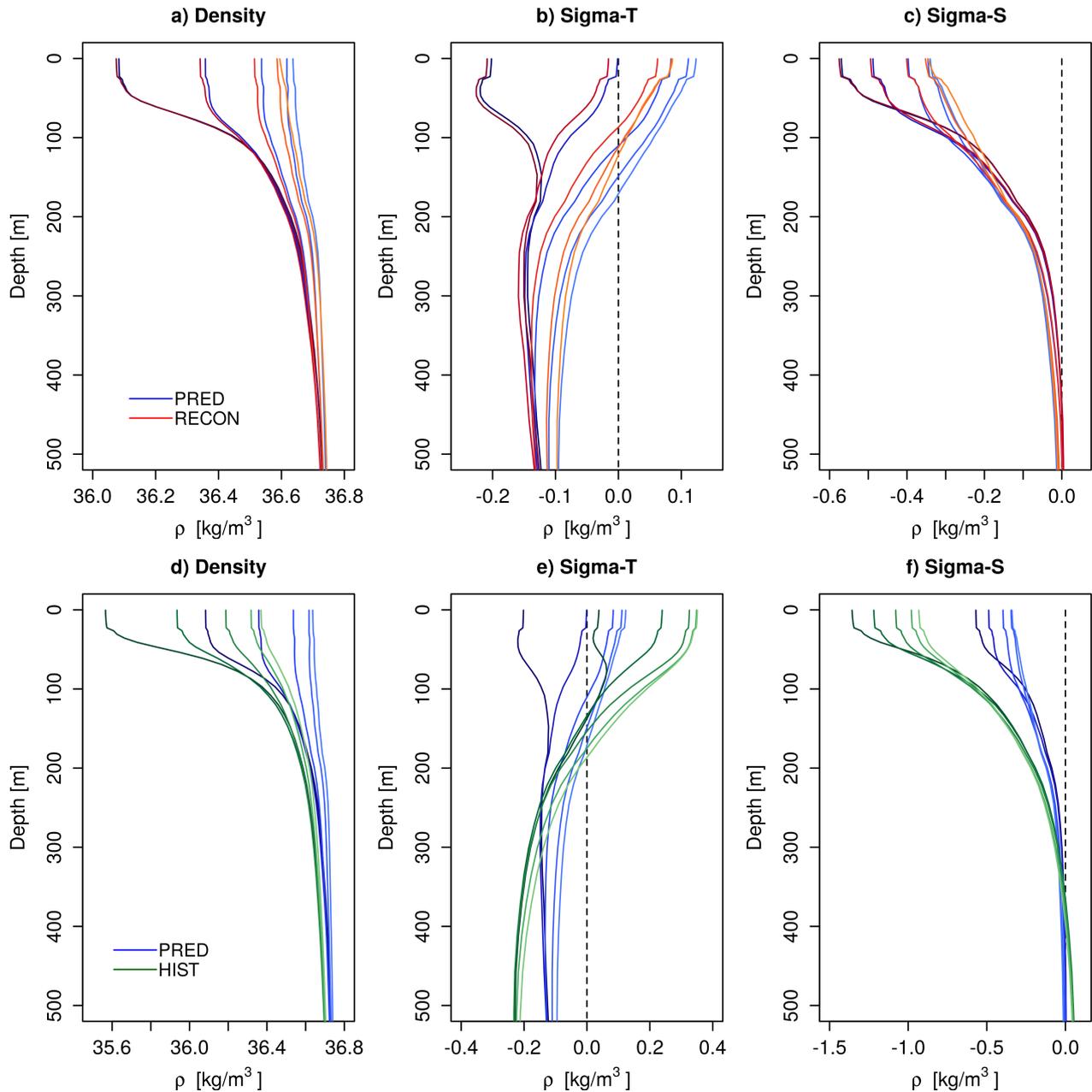
### 3.4 Insights on the Labrador Sea Initialisation Shock and Drift

Full-field initialisation can sometimes produce strong initialisation shocks and drifts, as the climate model adjusts from an initial state that might be substantially far from its attractor (e.g., ~~Sanchez-Gomez et al., 2016~~) (e.g. [Sanchez-Gomez et al., 2016](#)). This section focuses on Labrador Sea Convection, for which Figure 7a shows a clear ~~adjustment~~ [readjustment](#) marked by an initial increase and a subsequent decline. Both aspects of the predicted Labrador Sea evolution are investigated separately. We focus on the preconditioning role of Labrador Sea density stratification on convection, and investigate the role of temperature and salinity, two variables that might be experiencing a different initialisation adjustment and forecast drift over the region.

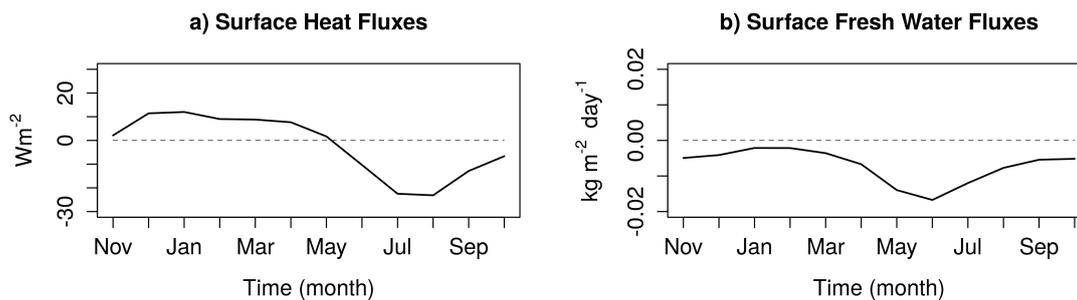
The initial enhancement of convection is explored in Figure 8, describing the evolution of stratification in the Labrador Sea the first five months of the forecast (Nov-Mar). At the time of initialisation (Nov), the density profiles of PRED and RECON are almost identical (dark red and blue lines in Figure 8a). Differences start to emerge in the subsequent forecast months, in which ~~their~~ [the](#) density stratification weakens at a different pace, with PRED becoming more weakly stratified and therefore more favorable to deep convection. HIST (green lines in [8aFigure 8d](#)) also shows a similar tendency to reduce density stratification from November to March, although the Labrador Sea density remains more strongly stratified than in PRED and RECON, which would explain why convection is also weaker. By considering the temperature and salinity contributions to density stratification (sigma-T and sigma-S, [figure 8band Figure 8b,c](#)) we find that even though the overall density structure is dominated by salinity, with temperature largely opposing the mean density stratification, the major differences between PRED and RECON occur in the sigma-T profile and are more notable at the surface. During the first 5 months of the forecast, sigma-T fully accounts for the differences between PRED and RECON in Labrador Sea density (e.g.  $-0.038 \text{ kg/m}^3\text{m}^3$  at the surface by March), with virtually no differences arising in sigma-S ( $0.005 \text{ kg/m}^3\text{m}^3$ ), which fails to counterbalance the temperature driven changes. As a result, the destabilising role of temperature on density stratification in the deep convection months is stronger in PRED than in HIST ~~;~~ [\(Figure 8e,f\)](#), promoting deeper convection.

To understand why Labrador Sea stratification diverges from RECON to PRED in the first forecast months we inspect the local surface restoring fluxes in the former, which, on average, are indicative of systematic model biases in the ocean component. In RECON, the heat flux restoring term is consistently positive and contributes thus to maintain a warmer surface in these months of deep convection (Figure 9). These fluxes are not present in PRED because the ~~simulation~~ [simulations](#) are fully coupled, which will quickly adjust to a new free-running state with a colder upper Labrador Sea, explaining in this way the relative surface cooling (and associated weakening of density stratification) with respect to RECON (Figure 8). [SimilarlyBy](#) [contrast](#), the freshwater fluxes from the salinity restoring term are ~~also positive~~ [negative](#) in the Labrador Sea, and contribute to keep a  ~~fresher (and lighter)~~ [saltier \(and denser\)](#) surface in RECON than in PRED. Its effect, however, appears to be small in magnitude as no remarkable differences emerge in sigma-S between RECON and PRED.

After better understanding the process behind the initial shock in the first winter, we now investigate the origin of the weakening in the Labrador Sea convection after the first forecast year of PRED. Again we analyse the evolution of the Labrador Sea density profile in PRED, but for each convective season (FMA) as a function of the forecast year (Figure 10). The corresponding profiles for RECON and the ensemble members of HIST are included to contextualise the predictions. The sigma-T



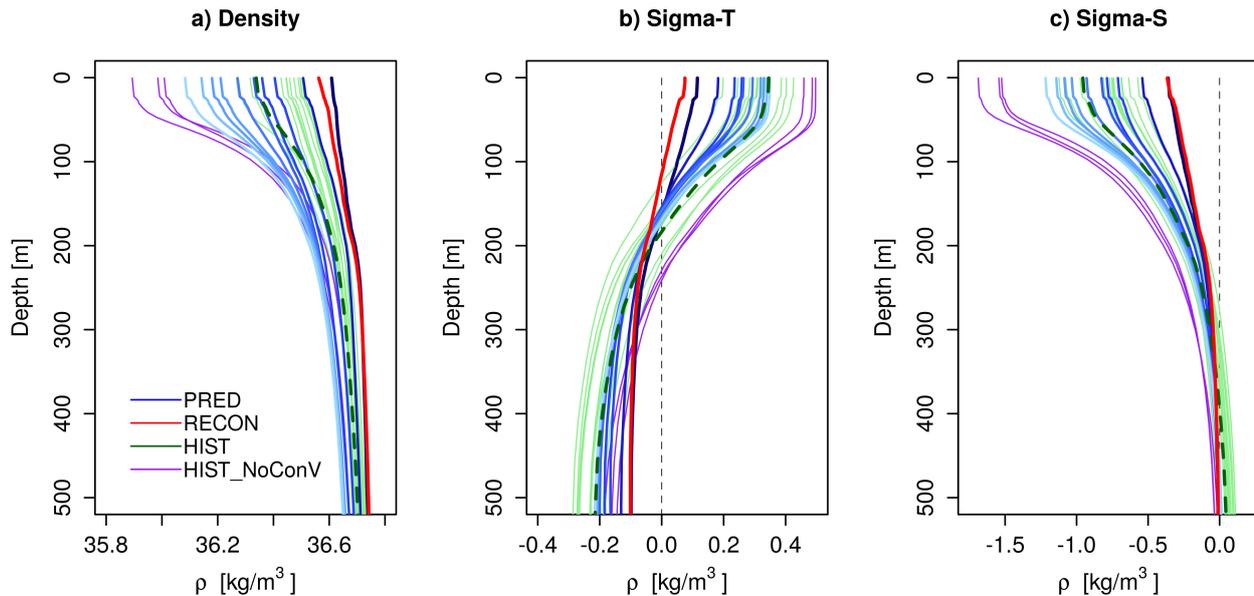
**Figure 8.** Labrador sea a) density, b) sigma-T and c) sigma-S climatological profiles in the first five forecast months (November to March) of PRED, and the equivalent calendar months of HIST and RECON. Panels d-f are the same as in panels a-c but for PRED and HIST (different panels are used to increase visibility). The colour intensity of the profiles from dark to light refers to increasing the forecast month.



**Figure 9.** Labrador Sea [52-65°N, 58-43°W] monthly climatology of the nudging correction fluxes in RECON of a) heat and b) freshwater. [In both cases the fluxes are defined from the atmosphere into the ocean.](#)

and sigma-S profiles are also shown to disentangle the contributions from temperature and salinity to density. After the first forecast FMA (darkest blue line in Figure 10), for which we showed a decrease in stratification that favoured deeper convection with respect to RECON, the density stratification becomes increasingly stronger with forecast time. This evolution is explained  
 515 by the changes in salinity (Figure 10c), as temperature contributes to decrease stratification at all forecast ranges (Figure 10b). In the second forecast FMA, density stratification in PRED is already stronger than in RECON (red line in Figure 10a). By the third forecast FMA it becomes stronger than in most of the HIST members with active convection (green lines), and by the sixth FMA it is already higher than in all of them. Interestingly, the stratification of sigma-S is not particularly different in PRED than in the HIST ensemble members with convection, which suggests once again that the counterbalancing effect of  
 520 sigma-T is important to understand the absence of convection in the forecasts. By the tenth (last) forecast FMA (lightest blue line in Figure 10) the density stratification is remarkably similar to that in the HIST ensemble members without convection. This may suggest that PRED is stabilising around this particular HIST state. However, this hypothesis is contradicted by the vertical profiles of sigma-T and sigma-S, that in the final forecast FMA appear to be more comparable to the HIST members with convection. It is therefore possible that the forecast drift is bringing PRED to a different equilibrium state than in HIST.

525 To investigate the model drift in the Labrador Sea and how it affects its stratification [figure-10 shows the we use](#) scatter-plots of the climatological Labrador Sea FMA temperature and salinity both at the surface and at 500m of depth (Figure 11). At the surface, the mean temperature and salinity for the first forecast FMA remain close to those in RECON, as well as to the values in several HIST members with active convection, all placed along the same isopycnal ( $27kg/m^3$ ). With increasing forecast time, PRED drifts towards a state with lower temperature and fresher conditions, the same one of the HIST members with no  
 530 convection. During this transition, the surface in PRED also becomes lighter, contributing to increase the stratification in the region. Important differences are observed in the subsurface (i.e. 500 m). For example, unlike for the surface, all the HIST members (i.e. the convective and non-convective ones) show rather similar climatological T,S values, roughly aligned along the same isopycnal ( $27.25kg/m^3$ ). PRED starts in this case far away from the HIST state (Figure 11c), although with similar density conditions. The main difference with respect to the surface is that with the subsequent forecast years, the subsurface

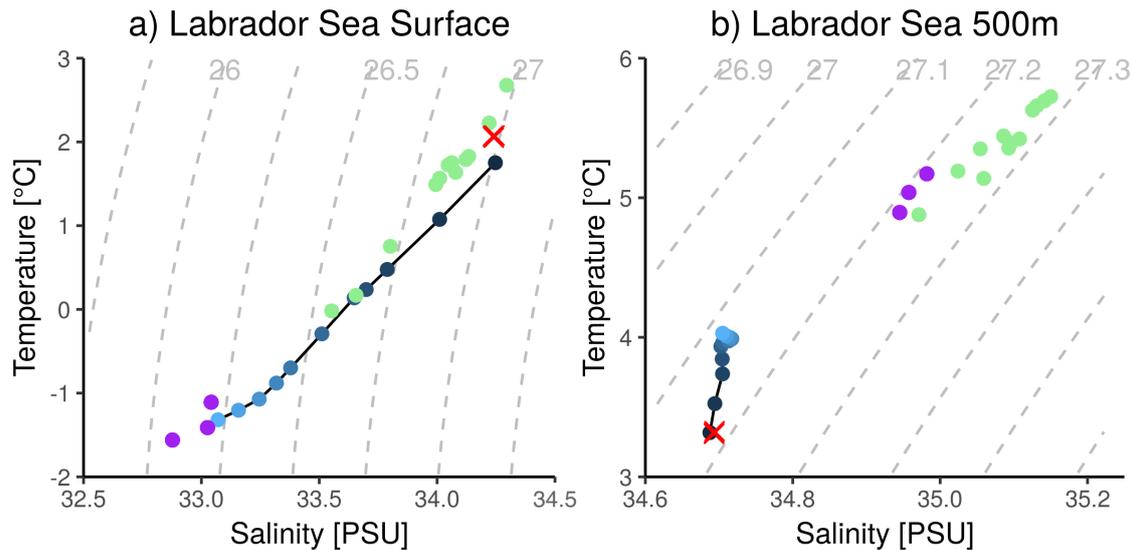


**Figure 10.** Labrador sea a) density, b) sigma-T and c) sigma-S climatological profiles for the convection season (February-March-April) in PRED, RECON and HIST. In PRED, the intensity of the blue lines is used to represent the changing forecast time, with the darkest blue line corresponding to the first forecast year, and the lighter blue line to the tenth. The HIST members have been divided into two sub-ensembles, those with and without convection in the Labrador Sea, HIST (green lines) and HIST-NoConv (purple lines). The green dashed line is the HIST ensemble mean using all members.

535 does not converge towards the typical mean HIST states. By the sixth forecast FMA the mean T,S value appears to stabilise  
 along a weaker isopycnal ( $27.1\text{kg}/\text{m}^3$ ). This suggests that the forecast drift has brought the model to a different equilibrium  
 state, at least in the Labrador Sea. The shock and the later drift may be caused by RECON being far from the EC-Earth3  
 model climate state in the Labrador Sea subsurface. In particular in terms of temperature and salinity (Figure 11), as the mean  
 density profiles are rather comparable due to the compensation between the temperature and salinity contributions (Figure 10).  
 540 Similar T,S diagrams, using HIST as a baseline, will be used in the future when evaluating the suitability of different ocean  
 reconstructions to initialise our next decadal prediction systems.

#### 4 Summary and Conclusions

In this paper we have presented and evaluated the predictive skill of a decadal forecast system with EC-Earth, based on full-  
 field initialisation, that contributes to the Decadal Climate Prediction Project component A (DCPP-A). The main findings of  
 545 the skill assessment are as follows:



**Figure 11.** Scatterplot diagram Scatter plot between the climatological Labrador Sea temperature and salinity during the convection season (February-March-April) both a) at the surface and b) at 500m. Blue dots of different intensity represent the climatological PRED values as a function of forecast year, the red cross represents RECON and the green (purple) dots the HIST members with (without) active Labrador Sea convection. Isopycnals are represented by dashed grey lines in the background. The dark green line shows the HIST ensemble mean using all members.

- In agreement with other decadal forecast systems (e.g., Yeager et al., 2018; Robson et al., 2018) (e.g. Yeager et al., 2018; Robson et al. 2018) EC-Earth3 is able to skillfully simulate the global mean surface temperature at short (forecast year one) and long forecast times (forecast years 6-10), with a large part of the skill arising from changes in the external forcings.
- Comparing different skill metrics (i.e. anomaly correlation coefficient and mean square skill score; ACC and MSSS) in the predictions and in an ensemble of historical simulations we have shown a beneficial effect of initialisation. In the first forecast year, surface temperature anomalies in regions like the Tropical Pacific, the eastern Subpolar North Atlantic and the Southern ocean show added value from initialisation in the predictions. At longer forecast times only a few localised regions show improvements in terms of MSSS due to initialisation, exemplified by the Eastern Equatorial Pacific and the Equatorial Atlantic. ACC differences show more limited improvements, from which we highlight a narrow band in the eastern Subpolar North Atlantic.
- The added value of initialisation is more easily discernible when considering both vertically and regionally integrated ocean quantities. For example, skill maps of the upper 300m ocean heat content (OHC300), which is more persistent than surface temperature as it is less affected by atmospheric perturbations, show larger areas of improved skill both in the Pacific and Atlantic oceans. Likewise, skill metrics are systematically better in the initialised predictions for the Atlantic Multidecadal Variability, although the improvements are not statistically significant.

- Another beneficial effect of initialisation is the reduction of the ensemble spread in the predictions with respect to the historical simulations, at least for the variables and indices analysed. The spread of the predicted anomalies is therefore better constrained at all forecast times.
- In contrast with other studies, the central Subpolar North Atlantic is a region of poor forecast skill in the EC-Earth3 forecast system. Both SST and the OHC300 show a detrimental effect of initialisation on its regional skill in the first 5 forecast years, which could be explained by an initialisation shock and the related long-term drift.

565

570

575

To investigate this potential shock we have further explored the forecast evolution in a selection of key ocean variables controlling multidecadal variability in the North Atlantic. The analysis showed that Labrador Sea convection collapses by forecast year 3 in the predictions, leading to a rapid weakening of the Atlantic Meridional Overturning Circulation (AMOC) and the Subpolar Gyre Circulation. This causes a cooling tendency of the western SPNA and a local expansion of sea ice, which occupies the entire Labrador Sea by forecast year 10. Although a similar state of suppressed convection is found in 3 out of 15 of the historical experiments, the mean of the historical ensemble (which in ~~this-our~~ case might not exactly correspond to a preferred model state) exhibits higher AMOC and subpolar gyre strength values, regular convection in the Labrador Sea and a more realistic sea ice extent. This suggests that the Labrador Sea convection collapse and subsequent North Atlantic changes are associated with an initialisation shock and drift that brings the predictions apart from their expected trajectory as represented by the historical ensemble mean.

580

585

We have further related the Labrador Sea convection collapse to the evolution of local density stratification and the separate contributions from temperature and salinity. During the first three forecast years, the Labrador density profile becomes more strongly stratified than in most of the historical members with active convection, following an intense surface freshening. This increase in stratification continues with forecast time, approaching but not reaching the strong density stratification levels from the three historical members without convection. To assess if the forecasts actually drift to an attractor characterised by these three historical members, we have additionally evaluated the climatological temperature and salinity in the region as a function of forecast time at the surface and 500 m. At the surface, the predictions start with mean temperature and salinity conditions within the range of those in the historical members with active convection, and by the end of the forecast they approach the typical state of the members without convection. At the subsurface, however, the forecasts remain far from either of the typical historical states, stabilising at forecast year 10 around a different (and lighter) attractor.

590

These results thus highlight the risk of initialising a sensitive region for decadal prediction, such as the Labrador Sea, too far from its preferred model state. A problem that, in this case, could have been minimised by applying a weaker nudging in the subsurface when producing the reconstruction that provided the ocean and sea ice initial conditions. Our findings also underline the importance of reducing as much as possible the mean model biases in the Labrador Sea, in particular at the subsurface. The problems herein described are particularly important when considering full-field initialisation (e.g., Magnusson et al., 2012; Smith et al., 2013) (e.g. Magnusson et al., 2012; Smith et al., 2013), an approach in which ~~shocks~~ initial imbalances of this kind are more ~~prone~~ likely to occur. In this sense, anomaly initialisation emerges as a potential alternative to minimise the drifts and, more importantly here, to minimize the occurrence of initialisation shocks. However, as

595 previous studies have shown (e.g., [Volpi et al., 2017](#)) ([e.g. Volpi et al., 2017](#)), this approach is not exempt from problems, and does not prevent initial model imbalances from happening, whose effect on Labrador Sea convection remains unknown. A complementary alternative is to devote new efforts in climate model development to reduce the model biases over the region, to thus reduce the mismatches with the observation-constrained products used for initialisation. Indeed, an appropriate model tuning in the Labrador Sea would benefit decadal prediction in two ways. First, by improving in the model realism in a source  
600 region of decadal skill, and the second, by helping to prevent or reduce problems associated with initialisation.

*Author contributions.* R.B. and S.W. led the analysis. R.B. and P.O. prepared the manuscript with contributions from all co-authors. M. D., P. O., Y. R.-R. and F.J. D.-R. contributed to the discussion and interpretation of the results. E. T., P. E. and M. C., made critical contributions to the development of the model version and workflow manager used to produce the experiments. J. A.-N., R. B., L.P.C. R. C.-G., V. S., E. T. and E. D. contributed to create the initial conditions for the decadal predictions. R. B. performed the decadal climate predictions and A.  
605 R., E. M.-C., I. C. the historical simulations. P.-A. B. and A. R. contributed to post-processing the model outputs. T. A., S. L.-T. and J. V.-R. developed the code to compute key ocean diagnostics. A.-C. H. and N. P.-Z. developed the analysis tools.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* The work in this paper was supported by the European Commission H2020 projects EUCP (776613), APPLICATE (727862) and [INTAROS \(727890\)](#) and [PRIMAVERA \(641727\)](#), a Spanish project funded by the Spanish Ministry of Economy (CLINSA, CGL2017-85791-R), Industry and Competitivy, a FRS-FNRS/FWO funded Belgian project (PARAMOUR, EOS-30454083) and an ESA contract (CMUG-CCI3-TECHPROP). The climate simulations analysed in the paper were performed using the internal computing resources available at MareNostrum and additional resources from PRACE (HiResNTCP, project 3: 2017174177) and the Red Española de Supercomputación (AECT-2019-2-0003 and AECT-2019-3-0006 projects) as well as the technical support provided by the Barcelona Supercomputing Center. In addition, several coauthors have been supported by personal grants: Y.R.-R, E.T. and S. W. received funding from the European  
615 Union Horizon 2020 research and innovation programme (Grant Agreements 800154, 748750, and 754433, respectively); I.C. was supported by Generalitat de Catalunya (Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement) through the Beatriu de Pinós programme; J. A.-N. was supported by the Spanish Ministry of Science, Innovation and Universities through a Juan de la Cierva personal grant (FJCI-2017-34027); R. C.-G. was funded by the Spanish Ministry of Education, Culture and Sports with an FPU Grant (FPU15/01511); and M.D. and P.O. were funded by the Spanish Ministry of Economy, Industry and Competitivy through the Ramon y Cajal grants RYC-  
620 2017-22964 and RYC-2017-22772.

## References

- Adler, R., Sapiano, M., Huffman, G., Wang, J.-J., Gu, G., Bolvin, D., Chiu, L., Schneider, U., Becker, A., Nelkin, E., and et al.: The Global Precipitation Climatology Project (GPCP) Monthly Analysis (New Version 2.3) and a Review of 2017 Global Precipitation, *Atmosphere*, 9, 138, <https://doi.org/10.3390/atmos9040138>, 2018.
- 625 Athanasiadis, P. J., Yeager, S., Kwon, Y.-O., Bellucci, A., Smith, D. W., and Tibaldi, S.: Decadal predictability of North Atlantic blocking and the NAO, *npj Climate and Atmospheric Science*, 3, 20, <https://doi.org/10.1038/s41612-020-0120-6>, 2020.
- [Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, \*Journal of Hydrometeorology\*, 10, 623–643, <https://doi.org/10.1175/2008JHM1068.1>, 2009.](#)
- 630 Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T., and Vitart, F.: ERA-Interim/Land: a global land surface reanalysis data set, *Hydrology and Earth System Sciences*, 19, 389–407, <https://doi.org/10.5194/hess-19-389-2015>, 2015.
- Barnston, A. G., Tippett, M. K., Ranganathan, M., and L'Heureux, M. L.: Deterministic skill of ENSO predictions from the North American Multimodel Ensemble, *Climate Dynamics*, 53, 7215–7234, <https://doi.org/10.1007/s00382-017-3603-3>, 2019.
- 635 Bindoff, N. L., Stott, P., AchutaRao, K., Allen, M., Gillett, N., Gutzler, D., Hansingo, K., G. Hegerl, Y. H., Jain, S., Mokhov, I., Overland, J., Perlwitz, J., Sebbari, R., and Zhang, X.: Detection and Attribution of Climate Change: from Global to Regional., in: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., Cambridge University Press, 2013.
- 640 Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R.: The Decadal Climate Prediction Project (DCPP) contribution to CMIP6, *Geoscientific Model Development*, 9, 3751–3777, 2016.
- Brodeau, L., Barnier, B., Treguier, A.-M., Penduff, T., and Gulev, S.: An ERA40-based atmospheric forcing for global ocean circulation models, *Ocean Modelling*, 31, 88–104, <https://doi.org/10.1016/j.ocemod.2009.10.005>, 2010.
- 645 Brune, S. and Baehr, J.: Preserving the coupled atmosphere–ocean feedback in initializations of decadal climate predictions, *WIREs Climate Change*, 11, e637, <https://doi.org/10.1002/wcc.637>, 2020.
- Caron, L.-P., Hermanson, L., and Doblas-Reyes, F. J.: Multiannual forecasts of Atlantic U.S. tropical cyclone wind damage potential, *Geophysical Research Letters*, 42, 2417–2425, <https://doi.org/10.1002/2015GL063303>, 2015.
- Caron, L.-P., Hermanson, L., Dobbin, A., Imbers, J., Lledó, L., and Vecchi, G. A.: How Skillful are the Multiannual Forecasts of Atlantic Hurricane Activity?, *Bulletin of the American Meteorological Society*, 99, 403–413, <https://doi.org/10.1175/BAMS-D-17-0025.1>, 2018.
- 650 Chikamoto, Y., Timmermann, A., Luo, J.-J., Mochizuki, T., Kimoto, M., Watanabe, M., Ishii, M., Xie, S.-P., and Jin, F.-F.: Skillful multi-year predictions of tropical trans-basin climate variability, *Nature Communications*, 6, 6869, <https://doi.org/10.1038/ncomms7869>, 2015.
- Choudhury, D., Sen Gupta, A., Sharma, A., Mehrotra, R., and Sivakumar, B.: An Assessment of Drift Correction Alternatives for CMIP5 Decadal Predictions, *Journal of Geophysical Research: Atmospheres*, 122, 10,282–10,296, <https://doi.org/10.1002/2017JD026900>, 2017.
- 655 Counillon, F., Bethke, I., Keenlyside, N., Bentsen, M., Bertino, L., and Zheng, F.: Seasonal-to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System Model: a twin experiment, *Tellus A: Dynamic Meteorology and Oceanography*, 66, 21074, <https://doi.org/10.3402/tellusa.v66.21074>, 2014.

Craig, A., Valcke, S., and Coquart, L.: Development and performance of a new version of the OASIS coupler, OASIS3-MCT\_3.0, *Geoscientific Model Development*, 10, 3297–3308, <https://doi.org/10.5194/gmd-10-3297-2017>, <https://gmd.copernicus.org/articles/10/3297/2017/>, 2017.

660

Dai, P., Gao, Y., Counillon, F., Wang, Y., Kimmritz, M., and Langehaug, H. R.: Seasonal to decadal predictions of regional Arctic sea ice by assimilating sea surface temperature in the Norwegian Climate Prediction Model, *Climate Dynamics*, 54, 3863–3878, <https://doi.org/10.1007/s00382-020-05196-4>, 2020.

Dee, D. P., Uppala, S. M., Simmons, A. J., and et al: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, 137, 553–597, <https://doi.org/10.1002/qj.828>, 2011.

665

DelSole, T. and Tippett, M. K.: Forecast Comparison Based on Random Walks, *Monthly Weather Review*, 144, 615–626, <https://doi.org/10.1175/MWR-D-15-0218.1>, 2016.

Doblas-Reyes, F. J., Andreu-Burillo, I., Chikamoto, Y., García-Serrano, J., Guemas, V., Kimoto, M., Mochizuki, T., Rodrigues, L. R. L., and van Oldenborgh, G. J.: Initialized near-term regional climate change prediction, *Nature Communications*, 4, 1715, <https://doi.org/10.1038/ncomms2704>, 2013.

670

Döscher, R. and the EC-Earth Consortium: The EC-Earth3 Earth System Model for the Climate Model Intercomparison Project 6, in prep.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development*, 9, 1937–1958, 2016.

García-Serrano, J. and Doblas-Reyes, F. J.: On the assessment of near-surface global temperature and North Atlantic multi-decadal variability in the ENSEMBLES decadal hindcast, *Climate Dynamics*, 39, 2025–2040, <https://doi.org/10.1007/s00382-012-1413-1>, 2012.

675

García-Serrano, J., Guemas, V., and Doblas-Reyes, F. J.: Added-value from initialization in predictions of Atlantic multi-decadal variability, *Climate Dynamics*, 44, 2539–2555, <https://doi.org/10.1007/s00382-014-2370-7>, 2015.

Goddard, L., Kumar, A., Solomon, A., Smith, D., Boer, G., Gonzalez, P., Kharin, V., Merryfield, W., Deser, C., Mason, S. J., Kirtman, B. P., Msadek, R., Sutton, R., Hawkins, E., Fricker, T., Hegerl, G., Ferro, C. A. T., Stephenson, D. B., Meehl, G. A., Stockdale, T., Burgman, R., Greene, A. M., Kushnir, Y., Newman, M., Carton, J., Fukumori, I., and Delworth, T.: A verification framework for interannual-to-decadal predictions experiments, *Climate Dynamics*, 40, 245–272, <https://doi.org/10.1007/s00382-012-1481-2>, 2013.

680

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, *J. Geophys. Res.*, 118, 6704–6716, <https://doi.org/10.1002/2013JC009067>, 2013.

Guemas, V., Doblas-Reyes, F. J., Lienert, F., Soufflet, Y., and Du, H.: Identifying the causes of the poor decadal climate prediction skill over the North Pacific, *Journal of Geophysical Research: Atmospheres*, 117, <https://doi.org/10.1029/2012JD018004>, 2012.

685

Guemas, V., Corti, S., García-Serrano, J., Doblas-Reyes, F. J., Balmaseda, M., and Magnusson, L.: The Indian Ocean: The Region of Highest Skill Worldwide in Decadal Climate Prediction\*, *Journal of Climate*, 26, 726–739, <https://doi.org/10.1175/JCLI-D-12-00049.1>, 2013.

Guemas, V., Doblas-Reyes, F. J., Mogensen, K., Keeley, S., and Tang, Y.: Ensemble of sea ice initial conditions for interannual climate predictions, *Climate Dynamics*, 43, 2813–2829, <https://doi.org/10.1007/s00382-014-2095-7>, 2014.

690

Hazeleger, W., Guemas, V., Wouters, B., Corti, S., AndreuBurillo, I., DoblasReyes, F. J., Wyser, K., and Caian, M.: Multiyear climate predictions using two initialization strategies, *Geophysical Research Letters*, 40, 1794–1798, <https://doi.org/10.1002/grl.50355>, 2013.

Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., and Folland, C. K.: A Tripole Index for the Interdecadal Pacific Oscillation, *Climate Dynamics*, 45, 3077–3090, <https://doi.org/10.1007/s00382-015-2525-1>, 2015.

Hermanson, L., Bilbao, R., Dunstone, N., Ménégos, M., Ortega, P., Pohlmann, H., Robson, J. I., Smith, D. M., Strand, G., Timmreck, C., Yeager, S., and Danabasoglu, G.: Robust Multiyear Climate Impacts of Volcanic Eruptions in Decadal Prediction Systems, *Journal of Geophys-*

695

- ical Research: Atmospheres, 125, e2019JD031739, <https://doi.org/10.1029/2019JD031739>, e2019JD031739 10.1029/2019JD031739, 2020.
- Heuzé, C.: North Atlantic deep water formation and AMOC in CMIP5 models, *Ocean Science*, 13, 609–622, <https://doi.org/10.5194/os-13-609-2017>, 2017.
- 700 Ho, C. K., Hawkins, E., Shaffrey, L., Bröcker, J., Hermanson, L., Murphy, J. M., Smith, D. M., and Eade, R.: Examining reliability of seasonal to decadal sea surface temperature forecasts: The role of ensemble dispersion, *Geophysical Research Letters*, 40, 5770–5775, <https://doi.org/10.1002/2013GL057630>, 2013.
- Hurrell, J. W.: Influence of variations in extratropical wintertime teleconnections on northern hemisphere temperature, *Geophysical Research Letters*, 23, 665–668, <https://doi.org/10.1029/96GL00459>, 1996.
- 705 Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory, J. M., Wood, R. A., Meijers, A. J. S., Mulcahy, J., Field, P., Furtado, K., Bodas-Salcedo, A., Williams, K. D., Copsey, D., Josey, S. A., Liu, C., Roberts, C. D., Sanchez, C., Ridley, J., Thorpe, L., Hardiman, S. C., Mayer, M., Berry, D. I., and Belcher, S. E.: Critical Southern Ocean climate model biases traced to atmospheric model cloud errors, *Nature Communications*, 9, 3625, <https://doi.org/10.1038/s41467-018-05634-2>, 2018.
- Keenlyside, N. S., Latif, M., Jungclauss, J., Kornbluh, L., and Roeckner, E.: Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, 453, 84–88, <https://doi.org/10.1038/nature06921>, 2008.
- 710 Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., and Saunby, M.: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties, *J. Geophys. Res.*, 116, <https://doi.org/10.1029/2010JD015218>, 2011.
- Kröger, J., Pohlmann, H., Sienz, F., Marotzke, J., Baehr, J., Köhl, A., Modali, K., Polkova, I., Stammer, D., Vamborg, F. S. E., and Müller, W. A.: Full-field initialized decadal predictions with the MPI earth system model: an initial shock in the North Atlantic, *Climate Dynamics*, 51, 2593–2608, <https://doi.org/10.1007/s00382-017-4030-1>, 2018.
- 715 Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the GISTEMP Uncertainty Model, *Journal of Geophysical Research: Atmospheres*, 124, 6307–6326, <https://doi.org/10.1029/2018JD029522>, 2019.
- Levermann, A. and Born, A.: Bistability of the Atlantic subpolar gyre in a coarse-resolution climate model, *Geophysical Research Letters*, 34, <https://doi.org/10.1029/2007GL031732>, 2007.
- 720 Madec, G. and the NEMO Team: NEMO ocean engine, Tech. Rep. 27, Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL), 2016.
- Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., and Stockdale, T.: Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors, p. 28, <https://doi.org/10.21957/7j9qrvjy>, 2012.
- 725 Manubens, N., Caron, L.-P., Hunter, A., Bellprat, O., Exarchou, E., Fučkar, N. S., Garcia-Serrano, J., Massonnet, F., Ménégoz, M., Sicardi, V., Batté, L., Prodhomme, C., Torralba, V., Cortesi, N., Mula-Valls, O., Serradell, K., Guemas, V., and Doblas-Reyes, F. J.: An R package for climate forecast verification, *Environmental Modelling Software*, 103, 29 – 42, <https://doi.org/https://doi.org/10.1016/j.envsoft.2018.01.018>, 2018.
- Manubens-Gil, D., Vegas-Regidor, J., Prodhomme, C., Mula-Valls, O., and Doblas-Reyes, F. J.: Seamless management of ensemble climate prediction experiments on HPC platforms, in: 2016 International Conference on High Performance Computing Simulation (HPCS), pp. 895–900, 2016.
- 730 McPhaden, M. J., Zebiak, S. E., and Glantz, M. H.: ENSO as an Integrating Concept in Earth Science, *Science*, 314, 1740–1745, <https://doi.org/10.1126/science.1132588>, 2006.

- Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M. A., Greene, A. M.,  
735 Hawkins, E., Hegerl, G., Karoly, D., Keenlyside, N., Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer,  
D., and Stockdale, T.: Decadal Prediction: Can It Be Skillful?, *Bulletin of the American Meteorological Society*, 90, 1467–1486,  
<https://doi.org/10.1175/2009BAMS2778.1>, 2009.
- Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., Corti, S., Danabasoglu, G., Doblas-Reyes, F., Hawkins, E.,  
Karspeck, A., Kimoto, M., Kumar, A., Matei, D., Mignot, J., Msadek, R., Navarra, A., Pohlmann, H., Rienecker, M., Rosati, T., Schneider,  
740 E., Smith, D., Sutton, R., Teng, H., van Oldenborgh, G. J., Vecchi, G., and Yeager, S.: Decadal Climate Prediction: An Update from the  
Trenches, *Bulletin of the American Meteorological Society*, 95, 243–267, <https://doi.org/10.1175/BAMS-D-12-00241.1>, 2014.
- [Meehl, G. A., Teng, H., Maher, N., and England, M. H.: Effects of the Mount Pinatubo eruption on decadal climate prediction skill of Pacific  
sea surface temperatures, \*Geophysical Research Letters\*, 42, 10,840–10,846, <https://doi.org/10.1002/2015GL066608>, 2015.](https://doi.org/10.1002/2015GL066608)
- 745 [Meehl, G. A., Hu, A., and Teng, H.: Initialized decadal prediction for transition to positive phase of the Interdecadal Pacific Oscillation,  
\*Nature Communications\*, 7, 11 718, <https://doi.org/10.1038/ncomms11718>, <https://doi.org/10.1038/ncomms11718>, 2016.](https://doi.org/10.1038/ncomms11718)
- Ménégoz, M., Bilbao, R., Bellprat, O., Guemas, V., and Doblas-Reyes, F. J.: Forecasting the climate response to volcanic eruptions: prediction  
skill related to stratospheric aerosol forcing, *Environmental Research Letters*, 13, 064 022, <https://doi.org/10.1088/1748-9326/aac4db>,  
2018.
- 750 Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A. S., Danabasoglu, G., Dirmeyer, P. A., Doblas-Reyes, F. J.,  
Domeisen, D. I. V., Ferranti, L., Ilynia, T., Kumar, A., Müller, W. A., Rixen, M., Robertson, A. W., Smith, D. M., Takaya, Y., Tuma, M.,  
Vitart, F., White, C. J., Alvarez, M. S., Ardilouze, C., Attard, H., Baggett, C., Balmaseda, M. A., Beraki, A. F., Bhattacharjee, P. S., Bilbao,  
R., de Andrade, F. M., DeFlorio, M. J., Díaz, L. B., Ehsan, M. A., Fragkoulidis, G., Grainger, S., Green, B. W., Hell, M. C., Infanti, J. M.,  
Isensee, K., Kataoka, T., Kirtman, B. P., Klingaman, N. P., Lee, J.-Y., Mayer, K., McKay, R., Mecking, J. V., Miller, D. E., Neddermann,  
755 N., Justin Ng, C. H., Ossó, A., Pankatz, K., Peatman, S., Pegion, K., Perlwitz, J., Recalde-Coronel, G. C., Reintges, A., Renkl, C., Solaraju-  
Murali, B., Spring, A., Stan, C., Sun, Y. Q., Tozer, C. R., Vigaud, N., Woolnough, S., and Yeager, S.: Current and emerging developments  
in subseasonal to decadal prediction, *Bulletin of the American Meteorological Society*, <https://doi.org/10.1175/BAMS-D-19-0037.1>, 2020.
- Mochizuki, T., Ishii, M., Kimoto, M., Chikamoto, Y., Watanabe, M., Nozawa, T., Sakamoto, T. T., Shiogama, H., Awaji, T., Sugiura, N.,  
Toyoda, T., Yasunaka, S., Tatebe, H., and Mori, M.: Pacific decadal oscillation hindcasts relevant to near-term climate prediction, *Proc.*  
760 *Natl. Acad. Sci. USA*, 107, 1833–1837, <https://doi.org/10.1073/pnas.0906531107>, 2009.
- Mogensen, K., Balmaseda, M. A., and Weaver, A. T.: The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean  
analysis for System 4', Tech. Memo. 668. ECMWF: Reading, UK, 2012.
- Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an  
ensemble of observational estimates: the HadCRUT4 dataset, *J. Geophys. Res.*, submitted.
- 765 Msadek, R., Delworth, T. L., Rosati, A., Anderson, W., Vecchi, G., Chang, Y.-S., Dixon, K., Gudgel, R. G., Stern, W., Wittenberg, A., Yang,  
X., Zeng, F., Zhang, R., and Zhang, S.: Predicting a Decadal Shift in North Atlantic Climate Variability Using the GFDL Forecast System,  
*Journal of Climate*, 27, 6472–6496, <https://doi.org/10.1175/JCLI-D-13-00476.1>, 2014.
- [Nadiga, B. T., Verma, T., Weijer, W., and Urban, N. M.: Enhancing Skill of Initialized Decadal Predictions Using a Dynamic Model of Drift,  
\*Geophysical Research Letters\*, 46, 9991–9999, <https://doi.org/10.1029/2019GL084223>, 2019.](https://doi.org/10.1029/2019GL084223)

- 770 O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Krieglner, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, *Geoscientific Model Development*, 9, 3461–3482, <https://doi.org/10.5194/gmd-9-3461-2016>, 2016.
- Ortega, P., Robson, J., Sutton, R. T., and Andrews, M. B.: Mechanisms of decadal variability in the Labrador Sea and the wider North Atlantic in a high-resolution climate model, *Climate Dynamics*, 49, 2625–2647, <https://doi.org/10.1007/s00382-016-3467-y>, 2017.
- 775 Pohlmann, H., Jungclaus, J. H., Köhl, A., Stammer, D., and Marotzke, J.: Initializing Decadal Climate Predictions with the GECCO Oceanic Synthesis: Effects on the North Atlantic, *Journal of Climate*, 22, 3926–3938, <https://doi.org/10.1175/2009JCLI2535.1>, 2009.
- Richardson, M., Cowtan, K., and Millar, R. J.: Global temperature definition affects achievement of long-term climate goals, *Environmental Research Letters*, 13, 054004, <https://doi.org/10.1088/1748-9326/aab305>, 2018.
- Robson, J., Polo, I., Hodson, D. L. R., Stevens, D. P., and Shaffrey, L. C.: Decadal prediction of the North Atlantic subpolar gyre in the  
780 HiGEM high-resolution climate model, *Climate Dynamics*, 50, 921–937, <https://doi.org/10.1007/s00382-017-3649-2>, 2018.
- Robson, J. I., Sutton, R. T., and Smith, D. M.: Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s, *Geophysical Research Letters*, 39, 1–6, <https://doi.org/10.1029/2012GL053370>, 2012.
- Rousset, C., Vancoppenolle, M., Madec, G., Fichefet, T., Flavoni, S., Barthélemy, A., Benschila, R., Chanut, J., Levy, C., Masson, S., and Vivier, F.: The Louvain-La-Neuve sea ice model LIM3.6: global and regional capabilities, *Geoscientific Model Development*, 8, 2991–  
785 3005, <https://doi.org/10.5194/gmd-8-2991-2015>, 2015.
- Ruprich-Robert, Y., Msadek, R., Castruccio, F., Yeager, S., Delworth, T., and Danabasoglu, G.: Assessing the Climate Impacts of the Observed Atlantic Multidecadal Variability Using the GFDL CM2.1 and NCAR CESM1 Global Coupled Models, *Journal of Climate*, 30, 2785–2810, <https://doi.org/10.1175/JCLI-D-16-0127.1>, 2017.
- Ruprich-Robert, Y., Delworth, T., Msadek, R., Castruccio, F., Yeager, S., and Danabasoglu, G.: Impacts of the Atlantic Multidecadal Variability on North American Summer Climate and Heat Waves, *Journal of Climate*, 31, 3679–3700, <https://doi.org/10.1175/JCLI-D-17-0270.1>,  
790 2018.
- Sanchez-Gomez, E., Cassou, C., Ruprich-Robert, Y., Fernandez, E., and Terray, L.: Drift dynamics in a coupled model initialized for decadal forecasts, *Climate Dynamics*, 46, 1819–1840, <https://doi.org/10.1007/s00382-015-2678-y>, 2016.
- Scaife, A. A. and Smith, D.: A signal-to-noise paradox in climate science, *npj Climate and Atmospheric Science*, 1, 28,  
795 <https://doi.org/10.1038/s41612-018-0038-4>, 2018.
- Servonnat, J., Mignot, J., Guilyardi, E., Swingedouw, D., Séférian, R., and Labetoulle, S.: Reconstructing the subsurface ocean decadal variability using surface nudging in a perfect model framework, *Climate Dynamics*, 44, 315–338, <https://doi.org/10.1007/s00382-014-2184-7>, 2015.
- Siegert, S., Bellprat, O., Menegoz, M., Stephenson, D. B., and Doblas-Reyes, F. J.: Detecting Improvements in Forecast Correlation Skill: Statistical Testing and Power Analysis, *Monthly Weather Review*, 145, 437–450, <https://doi.org/10.1175/MWR-D-16-0037.1>, 2017.  
800
- Smith, B., Wärlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S.: Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model, *Biogeosciences*, 11, 2027–2054, <https://doi.org/10.5194/bg-11-2027-2014>, 2014.
- Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., and Murphy, J. M.: Improved surface temperature prediction for the  
805 coming decade from a global climate model, *Science*, 317, 796–799, <https://doi.org/10.1126/science.1139540>, 2007.
- Smith, D. M., Eade, R., and Pohlmann, H.: A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction, *Climate Dynamics*, 41, 3325–3338, <https://doi.org/10.1007/s00382-013-1683-2>, 2013.

- Smith, D. M., Scaife, A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Bilbao, R., Borchert, L., Caron, L.-P., Counillon, F., Danabasoglu, G., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Estella-Perez, V., Flavoni, S., Hermanson, L., Keenlyside, N., Kharin, V., Kimoto, M., Merryfield, W. J., Mignot, J., Mochizuki, T., Modali, K., Monerie, P.-A., Muller, W. A., Nicoli, D., Ortega, P., Pankatz, K., Pohlmann, H., Robson, J., Ruggieri, P., Sospedra-Alfonso, R., Swingedouw, D., Wang, Y., Wild, S., Yeager, S., Yang, X., and Zhang, L.: North Atlantic Climate far more predictable than models imply, *Nature*, 584, Published Online, 2020.
- 810 Solaraju-Murali, B., Caron, L.-P., Gonzalez-Reviriego, N., and Doblas-Reyes, F. J.: Multi-year prediction of European summer drought conditions for the agricultural sector, *Environmental Research Letters*, 14, 124 014, <https://doi.org/10.1088/1748-9326/ab5043>, 2019.
- 815 Suckling, E.: Seasonal-to-Decadal Climate Forecasting, pp. 123–137, Springer International Publishing, Cham, [https://doi.org/10.1007/978-3-319-68418-5\\_9](https://doi.org/10.1007/978-3-319-68418-5_9), 2018.
- Sutton, R. T. and Dong, B.: Atlantic Ocean influence on a shift in European climate in the 1990s, *Nature Geoscience*, 5, 788–792, <https://doi.org/10.1038/ngeo1595>, 2012.
- Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, *Bulletin of the American Meteorological Society*, 93, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>, 2012.
- 820 Trenberth, K. E. and Shea, D. J.: Atlantic hurricanes and natural variability in 2005, *Geophys. Res. Lett.*, 33, L12 704, <https://doi.org/10.1029/2006GL026894>, 2006.
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, I., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis., *Q. J. R. Meteorol. Soc.*, 131, 2961–3012, <https://doi.org/10.1256/qj.04.176>, 2005.
- 825 Volpi, D., Guemas, V., and Doblas-Reyes, F. J.: Comparison of full field and anomaly initialisation for decadal climate prediction: towards an optimal consistency between the ocean and sea-ice anomaly initialisation state, *Climate Dynamics*, 49, 1181–1195, <https://doi.org/10.1007/s00382-016-3373-3>, 2017.
- 830 Weber, R. J. T., Carrassi, A., and Doblas-Reyes, F. J.: Linking the Anomaly Initialization Approach to the Mapping Paradigm: A Proof-of-Concept Study, *Monthly Weather Review*, 143, 4695–4713, <https://doi.org/10.1175/MWR-D-14-00398.1>, <https://doi.org/10.1175/MWR-D-14-00398.1>, 2015.
- 835 Webster, P. J. and Yang, S.: Monsoon and ENSO: Selectively Interactive Systems, *Quarterly Journal of the Royal Meteorological Society*, 118, 877–926, <https://doi.org/10.1002/qj.49711850705>, 1992.
- Yeager, S., Karspeck, A., Danabasoglu, G., Tribbia, J., and Teng, H.: A decadal prediction case study: Late twentieth-century north Atlantic Ocean heat content, *Journal of Climate*, 25, 5173–5189, <https://doi.org/10.1175/JCLI-D-11-00595.1>, 2012.
- 840 Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., Karspeck, A. R., Lindsay, K., Long, M. C., Teng, H., and Lovenduski, N. S.: Predicting Near-Term Changes in the Earth System: A Large Ensemble of Initialized Decadal Prediction Simulations Using the Community Earth System Model, *Bulletin of the American Meteorological Society*, 99, 1867–1886, <https://doi.org/10.1175/BAMS-D-17-0098.1>, 2018.
- Yuan, X., Kaplan, M. R., and Cane, M. A.: The Interconnected Global Climate System—A Review of Tropical–Polar Teleconnections, *Journal of Climate*, 31, 5765–5792, <https://doi.org/10.1175/JCLI-D-16-0637.1>, 2018.

845 Zhang, R. and Delwoth, T. L.: Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes, *Geophys. Res. Lett.*, 33, L17 712, <https://doi.org/10.1029/2006GL026267>, 2006.