Decadal climate prediction with a refined anomaly initialisation approach

Article

Accepted Version

Volpi, D., Guemas, V., Doblas-Reyes, F. J., Hawkins, E. ORCID: https://orcid.org/0000-0001-9477-3677 and Nichols, N. K. ORCID: https://orcid.org/0000-0003-1133-5220 (2017) Decadal climate prediction with a refined anomaly initialisation approach. Climate Dynamics, 48 (5). pp. 1841-1853. ISSN 1432-0894 doi: 10.1007/s00382-016-3176-6 Available at https://centaur.reading.ac.uk/66231/

Published version at: http://dx.doi.org/10.1007/s00382-016-3176-6
To link to this article DOI: http://dx.doi.org/10.1007/s00382-016-3176-6

Publisher: Springer

Clim Dyn manuscript No.
(will be inserted by the editor)

Decadal climate prediction with a refined anomaly initialisation approach

Danila Volpi · Virginie Guemas · Francisco Doblas-Reyes · Ed Hawkins · Nancy Nichols

Received: date / Accepted: date

Abstract In decadal prediction, the objective is to exploit both the sources of predictability from the external radiative forcings and from the internal variability to provide the best possible climate information for the next decade. Predicting the climate system internal variability relies on initialising the climate model from observational estimates. We present a refined method of anomaly initialisation (AI) applied to the ocean and sea ice components of the global climate forecast model EC-Earth, with the following key innovations: (i) the use of a weight applied to the observed anomalies, in order to avoid the risk of introducing anomalies recorded in the observed climate, whose amplitude does not fit in the range of the internal variability generated by the model; (ii) the anomaly initialisation of the ocean density, instead of calculating it from the anomaly initialised state of tem-

---

D. Volpi · Virginie Guemas · Francisco Doblas-Reyes
Institut Català de Ciències del Clima, Barcelona, Spain
Tel.: +34-935679977
E-mail: danila.volpi@ic3.cat

D. Volpi
Department of Mathematics and Statistics, University of Reading, Reading, UK

V. Guemas
Centre National de Recherches Météorologiques/Groupe d'Etude de l'Atmosphère Météorologique, Météo-France, CNRS, Toulouse, France

F. J. Doblas Reyes
Instituciò Catalana de Recerca i Estudis Avançats, Barcelona, Spain
Barcelona Supercomputing Center, Barcelona, Spain

E. Hawkins
National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK

N. K. Nichols
Department of Mathematics and Statistics, University of Reading, Reading, UK

perature and salinity. An experiment initialised with this refined AI method has been compared with a full field and standard AI experiment. Results show that the use of such refinements enhances the surface temperature skill over part of the North and South Atlantic, part of the South Pacific and the Mediterranean Sea for the first forecast year. However, part of such improvement is lost in the following forecast years. For the tropical Pacific surface temperature, the full field initialised experiment performs the best. The prediction of the Arctic sea-ice volume is improved by the refined AI method for the first three forecast years and the skill of the Atlantic Multidecadal Oscillation (AMO) is significantly increased compared to a non-initialised forecast, along the whole forecast time.

Keywords Decadal climate prediction · Full field initialisation · Refined Anomaly initialisation

1 Introduction

Decadal prediction aims at providing interannual to decadal climate information socially relevant to implement suitable strategies for adaptation. Decadal predictions have been shown to provide more skill than climate projections thanks to their initialisation from observational data, which allows the climate predictability arising from internal variability to be exploited (Doblas-Reyes et al, 2013). However, the choice of the most suitable technique to initialise the climate system is controversial and several techniques are currently explored. Full field initialisation (FFI) makes use of the best estimate of the observed climate system (Pohlmann et al, 2009), but model error causes the drift of the prediction towards the model attractor (Smith et al, 2013). Distinguishing between the climate signal to be predicted and the model drift is a challenging task. The application of a-posteriori bias correction has the risk of removing part of the variability signal one aims at predicting. With the aim of reducing the drift, the anomaly initialisation (AI) assimilates the observed anomaly variables[1] onto an estimate of the model mean climate (Smith et al, 2008).
Previous studies (Smith et al, 2013; Hazeleger et al, 2013; Bellucci et al, 2014) showed that the differences in skill of the two techniques at interannual time scales are small and limited to specific regions. Volpi et al (2015) showed that the AI allows for reducing the drift but some residual drift is still present. It allows for increasing the skill for sea ice, AMO and the Pacific Decadal Oscillation (PDO) compared to full field initialisation. In this work, we explore the possibility of refining further the anomaly initialisation technique used in Volpi et al (2015) to try to obtain a better skill. The use of the standard AI technique involves the risk of introducing anomalies recorded in the observed data whose amplitude does not fit in the range of the internal variability generated by the model. Figures S1 and S2 of the Supplementary Material show how this can affect the prediction of the signal. Some further examples of this issue are shown in Section 2.3. The first idea developed in this work consists in scaling the observed anomalies in order to take into account the different amplitudes of the observed versus the model variability.

---

[1] The anomaly of a field is defined as its deviation from the mean state (climate), calculated over a period of at least 30 years (according to the World Meteorological Organisation definition)

The second idea implemented aims at providing the most suitable initialisation for the density variable which plays a crucial role in the ocean circulation. In fact, the ocean variability on decadal timescales is mainly driven by changes in temperature and density. On one hand, temperature has a key role in the heat fluxes, and on the other hand, the density gradients drive the thermohaline circulation (Broecker, 1997). When implementing anomaly initialisation or anomaly nudging, density is often not directly assimilated. This is the case for DePreSys (Smith et al, 2007), the CNRM-CM5.1 (Germe et al, 2014), the MPI-OM (Matei et al, 2012) and the EC-Earth (Hazeleger et al, 2013) forecast systems. Instead, it is computed by the model from the assimilated temperature and salinity fields through a non-linear relation. Section 2.4 will describe an alternative method to initialise the temperature and density variables instead of the temperature and salinity variables initialised in the standard method. Section 2.1 and 2.2 describe respectively the model and the hindcast set-ups. The skill of the hindcasts initialised with the improved AI method is shown and compared to both an FFI and a standard AI set of hindcasts in section 3. The discussion and the conclusions are in section 4.

2 Methodology

2.1 Climate model

The model in use is the coupled general circulation model EC-Earth version 2.3 (Hazeleger et al, 2010). The atmospheric component is based on the European Centre for Medium-Range Weather Forecasts integrated forecasting system (IFS cy31r1), with 62 vertical levels and a TL159 horizontal resolution. The ocean component is the NEMO model version 3.2 (Madec, 2008; Ethe et al, 2006), with ORCA1 configuration (about 1 degree with enhanced tropical resolution) and 42 vertical levels. The sea-ice component is LIM2 (Fichefet and Maqueda, 1997; Goosse and Fichefet, 1999) directly embedded into NEMO. The atmospheric and ocean components are coupled via OASIS3 (Valcke, 2006). Information on the atmospheric chemistry and the dynamic vegetation are prescribed from observations. The atmospheric top is at 5 hPa, so the lower stratosphere is resolved.

2.2 Reference simulations: the NOINI and the FFI hindcasts

The benchmark hindcasts of this work are the FFI experiment of Du et al (2012) and an uninitialised model experiment, i.e. a historical simulation (Guemas et al, 2013). They were both part of the CMIP5 exercise. In the FFI experiment, all the variables from each model component are initialised by replacing the model state at the initialisation time with observational estimates (reanalysis). The atmosphere and land surface initial conditions are taken from the ERA-40 reanalysis (Uppala et al, 2005) for start dates before 1989 and ERA-Interim (Dee et al, 2011) afterwards. The ocean initial conditions are taken from the 3D-Var five-member ensemble ocean re-analysis NEMOVAR-ORAS4 (Mogensen et al, 2012), while the sea-ice initial conditions are produced with a simulation using NEMO v2.0 coupled to LIM2 driven by DFS4.3 ocean forcing data (Brodeau et al, 2009). The DFS4.3

forcing data are derived from ERA40 data with tropical surface air humidity, Arctic sea surface temperature and global wind field corrections based on high-quality observations.
The observed volcanic and anthropogenic aerosol load and greenhouse gas concentrations are prescribed using observed values up to 2005. After that date the RCP4.5 scenario was used, as well as a background solar irradiance level and a constant background volcanic aerosol load. Every two years between 1960 and 2004, on the 1st of November, a set of 5 new simulations were started and run for 5 years. The 5 members ensemble is generated from atmosphere initial perturbations based on singular vectors (Magnusson et al, 2008), which are added at the initial time to all the prognostic variables except for humidity (Du et al, 2012).
The uninitialised experiment, called NOINI, is a 3-member historical simulation up to 2005, and simulations following the representative concentration pathways 4.5 (RCP4.5) after 2006 produced in the framework of CMIP5. In the NOINI experiment, the internal variability is not in phase with the observed variability since each member has been initialised in 1850 from a different date of a pre-industrial control simulation. The NOINI experiment as well as all the experiments implemented in this study, employs the same external radiative forcing as described for the FFI.

2.3 Weighted anomalies

As mentioned in the Introduction, the variability of the model and the observations can have different amplitudes. An example is shown in figure 1 that illustrates the strength of the meridional overturning stream function averaged vertically and meridionally (30°-40°N band and 1-2 km depth). The model, shown in red, is the historical simulation described in section 2.1 (NOINI). Its meridional overturning transport is roughly 50% weaker than the reanalysis NEMOVAR-ORAS4 (blue). Moreover its decadal variability is substantially less pronounced.
As another example of the difference in amplitude of the model and observed variability, figure 2 illustrates the variability of the barotropic stream function calculated as the horizontal transport integrated vertically. The maps of the left column show NEMOVAR-ORAS4 data, while the ones of the right column are from NOINI of the model EC-Earth. The rows represent respectively January, May and September. Independently from the month considered, EC-Earth has a weaker variability than NEMOVAR-ORAS4 in the tropical band and in the North Pacific, but it has a stronger variability in the South Atlantic and South Pacific.
~~Introducing anomalies outside the model internal variability range could cause extreme events, for example, triggering an intense El Niño or stopping the thermohaline circulation (Sanchez-Gomez et al, 2015) ~~. To avoid introducing anomalies that are outside the model internal variability range~~such undesirable consequences~~, the first modification in the initialisation proposed in this work consists in weighting the observed anomalies to make their amplitude more consistent with the simulated variability. As a first attempt of weighting, we measure the model variability amplitude with the standard deviation, and we calculate the weight as the ratio between the standard deviation of the model anomalies and the standard deviation of the observed anomalies computed along the 1971-2000 period.

2.4 Density initialisation

The need for a proper initialisation of the density arises from the sensitivity of some areas, such as the North Atlantic, to the density anomalies. The density is not a prognostic variable, it is calculated at the initial time from the initialised values of temperature and salinity. It follows that in the standard AI method, the density is calculated from the values of temperature and salinity obtained by placing the observed temperature and salinity anomalies onto the model climatology. Such an estimate of the density is different from the value that would be obtained if the density was anomaly initialised. This happens because the equation of state of the density (that we will call $g(T, s)$) is non-linear and therefore the function composition[2] of g and AI is not commutative as shown from the inequality 1. Let us call $AI(x)$ the anomaly initialisation equation (Carrassi et al, 2014) applied to any variable x (x in this case will be the ocean temperature T, salinity s, or density ρ). Thus, we define x^a the anomaly initialised state after applying AI to x, $AI(x) = x^a$ (therefore $AI(\rho^o) = \rho^a$, where the superscript ^{o} indicates the observation). We call $g(T^o, s^o) = \rho^o$ and $g(T^a, s^a) = \rho^{standard}$ the equation of state of density calculated respectively from the observed ocean temperature and salinity, and from the T and s state after applying AI. $\rho^{standard}$ is the density used in the standard anomaly initialisation implementation.

$$\begin{aligned} g \circ AI &\neq AI \circ g \\ g \circ [AI(T^o, s^o)] &\neq AI \circ [g(T^o, s^o)] \\ g(T^a, s^a) &\neq AI(\rho^o) \\ \rho^{standard} &\neq \rho^a. \end{aligned} \tag{1}$$

As shown in inequality 1 the standard density $\rho^{standard}$ used in the classical anomaly initialisation implementation is different from the density ρ^a obtained by applying AI to the observed density. In a study of the DePreSys decadal prediction system, Robson (2010) suggested the errors in the assimilated density anomalies (i.e. the use of $\rho^{standard}$ instead of ρ^a) as responsible for the rapid warming of the hindcasts in the sub-polar gyre region in the Atlantic at the beginning of the 1990s.

To illustrate the order of magnitude of the difference in density introduced by anomaly initialising temperature and salinity, Figure 3 shows the ratio between the root mean square difference of the density $\rho^{standard}$ and the density ρ^a, over the root mean square anomalies (standard deviation) of the observed density ρ^o. In this map, the dark blue areas are the ones where the difference in the density initial value is three times or even more (the maximum ratio reaches the value of 6) the observed anomalies. The regions that are affected the most by such difference are the Arctic, in particular along the sea ice edge, the North Atlantic, the Mediterranean Sea and some regions in the Antarctic. In other words those are the areas with the highest non-linearity of g.

The method implemented and tested in this work consists in applying the weighted anomaly initialisation to density and temperature, and to find the salinity s which

---

[2] The function composition is the application of one function on top of another function and it is indicated with the symbol $\circ$

produces the ideal density ρ^a through $g(T^a, s)$. Since the density function $g(T, s)$ is not invertible, a bisection algorithm has been applied as explained in the supplementary material.

2.5 The anomaly initialised simulations

The hindcasts initialised with standard AI are the ones from Volpi et al (2015), with anomaly initialisation in all variables of the ocean and the sea-ice components (and referred to as OSI-AI hereafter). The land and the atmosphere components are initialised as in FFI. The hindcasts have been initialised on the 1st of November and are comprised of a set of 5-member simulations, 5-years long to moderate the computing time. The choice of having start-dates every two years is a good compromise between the computational cost and the statistical robustness of the results. The hindcasts initialised with the improved AI method have an analogous experimental set-up and will be called ρ-OSI-wAI (density, ocean and sea-ice weighted anomaly initialisation) hereafter.

2.6 Skill assessment

The metrics that we used to evaluate the skill of the hindcasts are the anomaly correlation (AC) and the Root Mean Square Error (RMSE) as a function of the forecast time f applied to the ensemble mean forecast anomalies. The forecast anomalies are calculated by subtracting the forecast climatology from each hindcast. In order to implement a fair comparison between the different experiments we have applied the same bias correction to all of them. In fact, there is still a residual drift present after applying anomaly initialisation. The forecast climatology at each grid point depends on the forecast time. It is estimated by averaging the hindcast variable across the starting dates and the members using only hindcast values for which observations are available at the corresponding dates. This data selection process is referred to as per-pair (García-Serrano and Doblas-Reyes, 2012). The implementation of the per-pair method guarantees the use of all the observational data available with a consistent estimation of the model and reference climatologies. Let call $X_{m,d,f}$ a model variable at forecast time f, start date d and member m. M is the total number of member and D the total number of start dates, that in this work is 23. $O_{d,f}$ is the corresponding observation. The forecast climatology is given by:

$$\begin{aligned} \bar{X}_{m,f} &= \frac{1}{(M-1)(D-1)} \sum_M \sum_D X_{m,d,f}(O_{d,f} \neq NA) \\ \bar{O}_f &= \frac{1}{D-1} \sum_D O_{d,f}(O_{d,f} \neq NA) \end{aligned} \tag{2}$$

when NA refers to a missing value. The difference between the observed and the model forecast climatology is the bias and section 3.1 looks at the drift defined as the evolution of such bias with forecast time.

The anomaly correlation is defined as:

$$AC(f) = \frac{\sum_{d=1}^{D} [x_{d,f}]' [o_{d,f}]'}{\sqrt{\sum_{d=1}^{D} [x_{d,f}]'^2 \sum_{d=1}^{D} [o_{d,f}]'^2}} \tag{3}$$

The root mean square error is given by:

$$RMSE(f) = \sqrt{\frac{\sum_{d=1}^{D} [[x_{d,f}]' - [o_{d,f}]']^2}{D}} \tag{4}$$

In equation 3, $x_{d,f}$ indicates the hindcast ensemble mean (for example ensemble mean global mean temperature) as a function of the forecast time f and the start date d, and D is the number of start dates. Note that $'$ indicates the model or observed per-pair anomalies. The confidence interval is calculated with a t-distribution for the AC, and with a χ^2 distribution for the RMSE. The dependence between the hindcasts is accounted for in the computation of the confidence interval using Von Storch and Zwiers (2001) formula. The confidence interval also takes into account the trend, that is not removed in the computation of the skill. The skill scores are computed after applying a one-year running mean in order to filter out seasonal climate variability and focus on interannual prediction skill, except for the PDO which is calculated with annual mean values.

3 Results

3.1 Forecast biases and drift

Figure 4 shows the bias of the experiments along the forecast time. Its evolution (along the forecast time) is the drift. The SST drift (figure 4a) in NOINI is negligible because the initial state of NOINI is a random state within the range of the possible states of the model climate and therefore it is the most balanced with the model climate. Its bias is negative along the whole forecast time, consistent with the strong cold tropical bias of the model (not shown). Moreover, figure 4a shows the overshoot of FFI (red line) that jumps to too high temperatures in only a few months and drops quickly towards too low temperatures as compared to the observations (as the bias gets negative) and even to temperatures lower than the NOINI ones. This is due to the difference in timescales between the drift toward a warm bias in the Southern Ocean (a few months only) and the drift toward a cold state in the tropical band and the Northern hemisphere (a few years). FFI has the strongest drift because its initial state corresponds to the observed state and it is the furthest from the model climate. These results are consistent with Hazeleger et al (2013). The AI method does not remove the bias of the model from the initial state of the system. Consequently, the drift of both ρ-OSI-wAI and OSI-AI are largely reduced with respect to FFI, and the overshoot is avoided in both cases. The drift is further reduced in ρ-OSI-wAI compared to OSI-AI. The bias for the Arctic sea ice area (figure 4b) of the AI experiments is very close to the NOINI bias along the whole forecast time and there is no drift. This is not the case for the FFI, for which the bias in winter is still present after 5 forecast years although reduced compared with the first year.

3.2 Sea surface temperature

Figure 5 shows the improvements in SST skill of the refined AI technique (ρ-OSI-wAI) over the FFI (first panel) and the OSI-AI experiments (second panel), for the first forecast year, measured as the ratio of their RMSE. While the refined method improves the skill in the Labrador Sea and in the Weddell Sea with respect to the FFI experiment, it also degrades the skill in the Bering Sea, the tropical Pacific and the Indian Ocean (left panel figure 5). The added value of the anomaly weighting and the density initialisation over the standard AI technique is seen in the the northern part of the North Atlantic, part of the North Pacific and the Southern ocean. The improved sectors of the Mediterranean Sea, the Labrador and the Gin Seas correspond to the region highlighted in figure 3 as being sensitive to the density error. The following sections will explore, through the study of the thermohaline circulation and the main modes of variability, the sources of such improvements in skill.

3.3 Predicting the ocean heat content

Figure 6a shows the anomaly correlation of the ocean heat content as a function of forecast time for the four experiments. The refined AI method (green line) shows an improvement in skill with respect to NOINI, although its correlation is lower than the other initialised experiments (FFI in red and OSI-AI in purple). The skill of the three initialised experiments degrades with forecast time toward the skill of NOINI which is nearly constant. The RMSE plot (6b) illustrates the robustness of the conclusions drawn from the AC results. The supplementary material shows that the improvement in skill of the global ocean heat content does not come from the North Atlantic sector, where the best skill is obtained by the NOINI experiment (figure S5).

3.4 Predicting the thermohaline circulation

The correlation of the three initialised experiments (ρOSI-wAI, OSI-AI and FFI) for the AMOC index, calculated as in Figure 1, drops below the NOINI skill after the first forecast year (Figure 7a) and the ACs confidence interval cross the zero line during the second forecast year, which means that the skill is not significant any more. This is consistent with the results of the North Atlantic sub-polar and sub-tropical gyres shown in figure S6 and S7 of the Supplementary Material. While the correlation shows minor differences between the initialised experiments at the beginning of the forecast time, the RMSE plot (figure 7b) shows a higher RMSE of the refined AI method than the other initialised experiments at the beginning of the forecast, but a lower RMSE and a higher correlation a the end of the forecast.

3.5 Predicting the sea ice cover

The performance for the sea ice cover is validated against the HistDfsNudg sea ice reconstruction (Guemas et al, 2014), which has also been used for the initialisation.

For the Arctic sea-ice area, the forecast skill is improved for all the initialised experiments over NOINI during the first one to two forecast years. ρ-OSI-wAI is the experiment that has the highest correlation (figure 8a) and the smallest RMSE (figure 8b) during the first two forecast years, followed by OSI-AI and FFI. The results are less conclusive in the second half of the forecast. For the Arctic sea-ice volume (figure 8c and d), the skill of the experiments exhibit two types of behaviours: the anomaly initialised experiments (both ρ-OSI-wAI and OSI-AI) with the highest correlation and the smallest RMSE, both improving over the NOINI experiment for the first three forecast years, and the NOINI and FFI experiments with the lowest correlation and the largest RMSE.

3.6 Impact on some modes of climate variability

The Atlantic multidecadal oscillation (AMO) is a multidecadal climate variability pattern consisting in alternating phases of warm and cold sea surface temperature over the North Atlantic (Deser et al, 2010). It is thought to be the surface fingerprint of the thermohaline circulation (Kerr, 2000; Knight et al, 2005) and is calculated as the difference between the mean SST anomalies in the North Atlantic and the global mean SST anomalies between 60°S and 60°N following the definition of Trenberth and Shea (2006). Previous studies have shown that the predictive skill for AMO can be improved by initialisation (Meehl et al, 2014) The positive impact of the initialisation for the AMO index persists for the whole forecast time (figure 9). There is also a substantial improvement of ρ-OSI-wAI compared to FFI at every forecast time except for a few months in year 5 in which the skill of the two experiments are very close. ρ-OSI-wAI seems also to perform better than OSI-AI, although the skill of the two experiments are close and for a few months during the second year OSI-AI has larger skill. The improvements of the refined AI method over NOINI are significant along the whole forecast period (except for some months in year 3), whereas the difference between FFI and NOINI is significant for the first forecast year only. The RMSE results are consistent with what is shown in the correlation plot.

In addition, we focus on the Pacific Decadal Oscillation (PDO), the most long-lived sea surface temperature mode in the Pacific. The PDO is defined as the leading principal component of the Pacific annual SST variability calculated in the domain $20°\text{N} - 65°\text{N}$ (Mantua et al, 1997). The observed dominant EOF has been calculated from the detrended observed anomalies and then the model anomalies have been projected onto the observed EOF. The PDO is known to impact the North Pacific and North American climates and it has also been linked to variations in surface air temperature, snowpack, precipitation and marine ecosystems (Mantua et al, 1997; Anderson et al, 2009). For the PDO index (figure 10), there is an improvement in skill of ρ-OSI-wAI as well as OSI-AI and FFI, relative to NOINI for the first forecast year, but this improvement is not significant. This is consistent with the improvement seen in the North Pacific SST shown in section 3.2 from the refined AI initialisation method relative to the standard AI. The correlation (figure 10) after the first forecast year drops for all the experiments.

3.7 Regional behaviour of the initialisation techniques

Figure 11 shows the minimum SST RMSE across all the experiments respectively for the first forecast year (left panel) and the average of the years two to five (right panel). Each grid point takes the colour of the experiment that has the minimum SST RMSE. The black dots appear in those regions where the minimum RMSE differs from the second minimum RMSE by more than 0.05 K. There are a few areas where the NOINI experiment has the lowest RMSE during the first forecast year in the Southern hemisphere, probably due to the lack of observations that does not allow for good initialisation or robust verification. The FFI experiment has the lowest RMSE in the tropical Pacific and the ENSO region. This could lead to the conclusion that the initialisation of the mean state in the tropical region cannot be neglected and therefore the FFI might be preferred to the AI technique. Another possible cause of the poor performance in the tropical region of the AI might be the fact that the model and the observations reproduce a similar variability but in slightly different geographical positions. This would imply that when applying the anomaly initialisation, the observed anomalies are introduced in a shifted position with respect to the position where the model would simulate the corresponding anomalies. In most parts of the Atlantic and some parts of the Pacific, the ρ-OSI-wAI experiment performs the best.
When averaging the forecast years $2-5$, the benefits of ρ-OSI-wAI are still shown in some parts of the Arctic region, around Europe and in some regions of the tropical band. The areas of the tropical Pacific and Atlantic are still best predicted by FFI, although the regions where NOINI has the lowest RMSE have increased. Similar results are found when computing the maximum correlation for each grid point (not shown).

4 Summary and conclusions

With the aim of improving the prediction skill on decadal time scales, this work has introduced a new anomaly initialisation (AI) method (ρ-OSI-wAI) that tackle some of the limitations of the classical AI technique. The innovations implemented are:

- the weighting of the observed anomalies by the ratio between the amplitudes of the model and observed variabilities, to avoid the risk of introducing anomalies that are outside the range of the model variability in the initial state
- the anomaly initialisation of the ocean density, instead of calculating it from the anomaly initialised state of temperature and salinity.

We have justified the need for such refinements and illustrated the implementation of the new technique in the Methodology section. In the Results section we have tried to evaluate the effect of the refinements on the predictions through the skill assessment of a set of variables that have been compared with experiments initialised with classical techniques (full field initialisation FFI, classical anomaly initialisation OSI-AI and with a free run -NOINI-). Although the lack of resources did not allow the weight of the variability amplitude and the density correction to be tested separately, the combination of these two innovations improves the skill globally compared to the other classical methods of initialisation presented in this work. In particular the refined method:

- allows the drift of sea surface temperature (SST) to be further reduced with respect to the FFI and the standard AI.
- allows for a higher skill than the other methodologies presented in this study in the Arctic sea-ice area (first two forecast years) and volume (three forecast years), although the improvements are not statistically significant.
- improves the Pacific Decadal Oscillation skill over the first forecast year with respect to the other methodologies presented in this study, but the improvements are not significant.
- increases the SST skill over the standard AI method for forecast year 1 in the Labrador Sea, the Mediterranean, part of the North Pacific and the Southern ocean.

The Mediterranean, the Labrador Sea and the Southern Ocean, where the refined AI method improves the forecast quality over the standard ocean and sea ice AI implementation, are also some of the areas with high density difference with a standard AI technique at the initial time. This relation suggests a potential attribution to the density anomaly initialisation for the improvements in these regions. It might not be then by chance that the skill of the Atlantic Multidecadal Oscillation is significantly improved by the refined AI method compared to a historical simulation along the whole forecast time. In comparison, a full field initialisation technique allows for a significant improvement only during the first forecast year while a standard ocean and sea ice AI only for the first 2 forecast years. The large density differences between the standard and refined AI methods in key areas where ocean dynamics might play a key role for the decadal predictability would suggest a larger impact of this correction on the skill. The relatively small differences in skill found point towards the need of a further understanding of how to best implement this approach in current models, with coarse resolution and substantial systematic errors. However, the weighting of the observed anomalies as it is implemented has some limitations. The use of the standard deviation as a measure of the model variability amplitude is fully representative of this variability only when the distribution of the anomalies is Gaussian and the sample size is large enough to allow for a robust estimate, which is generally not the case for the variables of the climate system. Further efforts could be ~~inverted~~ invested into refining the weight implementation and further enhancing the skill of the predictions. The other open issue to address is the geographical shift between the model and the observed variability, that could be the cause of the loss in skill of the anomaly initialised predictions in the tropical region.

Acknowledgements The authors acknowledge funding support for this study from the SPECS (ENV-2012-308378) project funded by the Seventh Framework Programme (FP7) of the European Commission and the PICA-ICE (CGL2012-31987) project funded by the Ministry of Economy and Competitiveness of Spain. EH was also funded by the UK Natural Environment Research Council. DV gratefully acknowledges financial support from the University of Reading. The authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the Red Española de Supercomputación through the Barcelona Supercomputing Center.

References

Anderson DLT, Doblas-Reyes FJ, Balmaseda M, Weisheimer A (2009) Decadal variability: processes, predictability and prediction. ECMWF Technical Memorandum 591, URL http://www.ecmwf.int/publications/library/do/references/show?id=89132

Bellucci A, Haarsma R, Gualdi S, Athanasiadis PJ, Caian M, Cassou C, Fernandez E, Germe A, Jungclaus J, Kroger J, Matei D, Muller W, Pohlmann H, Salas-M*é*lia D, Sanchez E, Smith D, Terray L, Wyser K, Yang S (2014) An assessment of a multi-model ensemble of decadal climate predictions. Clim Dyn 44:2787–2806, DOI 10.1007/s00382-014-2164-y

Brodeau L, Barnier B, Treguier A, Penduff T, Gulev S (2009) An era40-based atmospheric forcing for global ocean circulation models. Ocean Model 31:88–104, DOI 10.1016/j.ocemod.2009.10.005

Broecker WS (1997) Thermohaline circulation, the achilles heel of our climate system: Will man-made co2 upset the current balance? Science 278(5343):1582–1588, DOI 10.1126/science.278.5343.1582

Carrassi A, Weber RJT, Guemas V, Doblas-Reyes FJ, Asif M, Volpi D (2014) Full-field and anomaly initialization using a low-order climate model: a comparison and proposals for advanced formulations. Non linear Processes in Geophysics 21:521–537, DOI 10.5194/npg-21-521-2014

Dee DP, Uppala SM, Simmons AJ, P Berrisford PP, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Holm EV, Isaksen L, K*ȧ*llbergc P, K*ö*hler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thpaut JN, Vitart F (2011) The era-interim reanalysis: configuration and performance of the data assimilation system. Quart J Roy Meteor Soc 137:553–597

Deser C, Alexander MA, Xie SP, Phillips AS (2010) Sea surface temperature variability: Patterns and mechanisms. Annual Review of Marine Science 2:115–143, DOI 10.1146/annurev-marine-120408-151453

Doblas-Reyes F, Andreu-Burillo I, Chikamoto Y, García-Serrano J, Guèmas V, Kimoto M, Mochizuki T, Rodrigues L, van Oldenborgh G (2013) Initialized near-term regional climate change prediction. Nature Communications 4(1715), DOI 10.1038/ncomms2704

Du H, Doblas-Reyes FJ, García-Serrano J, Guèmas V, Soufflet Y, Wouters B (2012) Sensitivity of decadal predictions to the initial atmospheric and ocean perturbations. Clim Dyn DOI 10.1007/s00382-011-1285-9

Ethe C, Aumont O, Foujols MA, Levy M (2006) Nemo reference manual, tracer component: Nemo-top. preliminary version. Note du Pole de modlisation, Institut Pierre-Simon Laplace (IPSL) France 28:1288–1619

Fichefet T, Maqueda MAM (1997) Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. J Geophys Res 102:12,609–12,646

García-Serrano J, Doblas-Reyes F (2012) On the assessment of near-surface global temperature and north atlantic multi-decadal variability in the ensembles decadal hindcast. Clim Dyn DOI 10.1007/s00382-012-1413-1

Germe A, Chevallier M, Salas-Mélia D, Sanchez-Gomez E, Cassou C (2014) Interannual predictability of arctic sea ice in a gloabl climate model: regional contrasts and temporal evolution. Clim Dyn DOI 10.1007/s00382-014-2071-2

Goosse H, Fichefet T (1999) mportance of ice-ocean interactions for the global ocean circulation: a model study. J Geophys Res 104:23,337–23,355

Guemas V, Doblas-Reyes F, Andreu-Burillo I, Asif M (2013) Retrospective prediction of the global warming slowdown in the past decade. Nature Clim Change 3:649–653

Guemas V, Doblas-Reyes F, Mogensen K, Keely S, Tang Y (2014) Ensemble of sea ice initial conditions for interannual climate predictions. Clim Dyn 43:2813–2829, DOI 10.1007/s00382-014-2095-7

Hazeleger W, Severijns C, Semmler T, Stefânescu S, Yang S, Wang X, Wyser K, Dutra E, Baldasano JM, Bintanja R, Bougeault P, Caballero R, Ekman AML, Christensen JH, van den Hurk B, Jimenez P, Jones C, Kållberg P, Koenigk T, McGrath R, Miranda P, van Noije T, Palmer T, Parodi JA, Schmith T, Selten F, Storelvmo T, Sterl A, Tapamo H, Vancoppenolle M, Viterbo P, Willân U (2010) Ec-earth: a seamless earth-system prediction approach in action. Bull Amer Meteorol Soc 91(10):1357–1363, DOI 10.1175/2010BAMS2877.1

Hazeleger W, Guemas V, Wouters B, Corti S, Andreu-Burillo I, Doblas-Reyes FJ, Wyser K, Caian M (2013) Multiyear climate predictions using two initialization strategies. Geophys Res Lett 40(9):1794–1798, DOI 10.1002/grl.50355

Kerr RA (2000) A north atlantic climate pacemaker for the centuries. Science 288(5473):1984–1985, DOI 10.1126/science.288.5473.1984

Knight JR, Allan RJ, Folland CK, Vellinga M, Mann ME (2005) A signature of persistent natural thermohaline circulation cycles in observed climate. Geophys Res Lett 32(20), DOI 10.1029/2005GL024233

Madec G (2008) Nemo ocean engine. Note du Pole de modlisation, Institut Pierre-Simon Laplace (IPSL) France 27:12881619

Magnusson L, Leutbecher M, Kallen E (2008) Comparison between singular vectors and breeding vectors as initial perturbations for the ecmwf ensemble prediction system. Mon Wea Rev 134:4092–4104

Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC (1997) A pacific interdecadal climate oscillation with impacts on salmon production. Bull Amer Meteorol Soc 78:1069–1079

Matei D, Pohlmann H, Jungclaus J, Müller W, Haak H, Marotzke J (2012) Two tales of initializing decadal climate prediction experiments with the echam5/mpi-om model. J Clim 25:8502–8523, DOI 10.1175/JCLI-D-11-00633.1

Meehl GA, 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, Pohlmann H, Rienecker M, Rosati T, Schneider E, Smith D, Sutton R, Teng H, van Oldenborgh GJ, Vecchi G, 1 SY (2014) Decadal climate prediction: an update from the trenches. Bull Amer Meteorol Soc 95:243–267, DOI 10.1175/BAMS-D-12-00241.1

Mogensen KS, Balmaseda MA, Weaver A (2012) The nemovar ocean data assimilation as implemented in the ecmwf ocean analysis for system4. ECMWF Technical Memorandum 657, in preparation

Pohlmann H, Jungclaus J, Köhl A, Stammer D, Marotzke J (2009) Initializing decadal climate predictions with the gecco oceanic synthesis: Effects on the north atlantic. J Clim 22:3926–3938

Robson JI (2010) Understanding the performance of a decadal prediction system. PhD thesis, University of Reading, dept. of Meteorology

Sanchez-Gomez E, Cassou C, Ruprich-Robert Y, Fernandez E, Terray L (2015) Drift dynamics in a coupled model initialized for decadal forecasts. Clim Dyn pp 1–22, DOI 10.1007/s00382-015-2678-y

Smith DM, Cusack S, Colman AW, Folland CK, Harris GR, Murphy JM (2007) Improved surface temperature prediction for the coming decade from a global climate model. Science 317:796–799

Smith DM, Eade R, Pohlmann H (2013) A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. Clim Dyn DOI 10.1007/s00382-013-1683-2

Smith T, Reynolds R, Peterson T, Lawrimore J (2008) Improvements to noaa's historical merged land-ocean surface temperature analysis (1880-2006). J Clim 21:2283–2296

Trenberth KE, Shea DJ (2006) Atlantic hurricanes and natural variability in 2005. Geophys Res Lett 33(L12704), DOI 10.1029/2006GL026894

Uppala SM, Kȧllberg PW, Simmons AJ, Andrae U, Bechtold VDC, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Berg LVD, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hlm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, Mcnally AP, Mahfouf J, Morcrette J, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J (2005) The era-40 reanalysis. Quart J Roy Meteor Soc 131:2961–3012

Valcke S (2006) Oasis3 user guide. PRISM Support Initiative Report 3:64

Volpi D, Guemas V, Doblas-Reyes FJ (2015) Comparison of full field and anomaly initialisation for decadal climate prediction: towards an optimal consistency between the ocean and sea-ice anomaly initialisation state. Clim Dyn Submitted: CLDY-D-15-00212

Von Storch H, Zwiers F (2001) Statistical analysis in climate research. Cambridge University Press

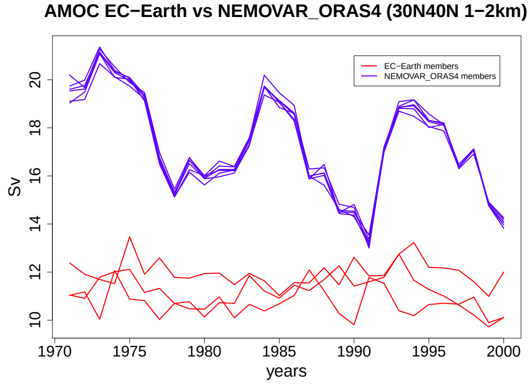

Fig. 1 Comparison of the Atlantic meridional overturning stream function averaged in the 30°-40°N band and 1-2 km depth, generated by NEMOVAR-ORAS4 (in blue) and the 3-member historical simulation performed with EC-Earth v2.3 -NOINI- (in red).

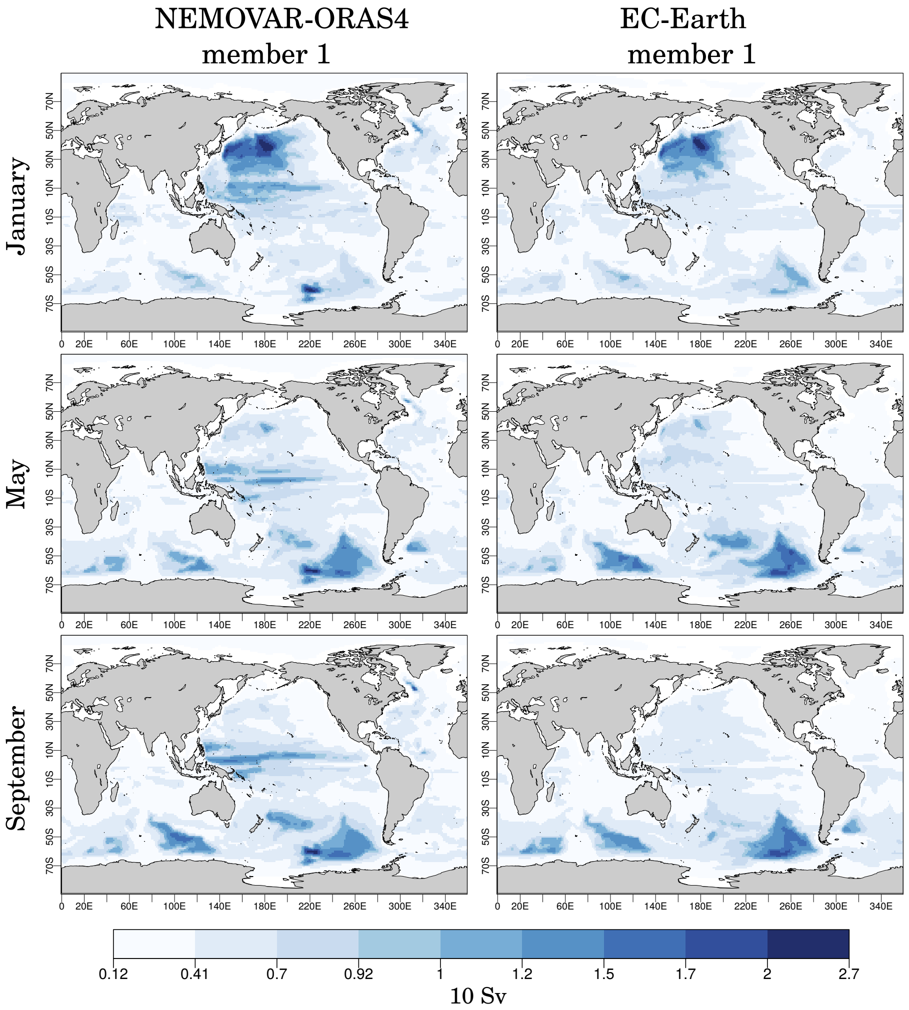

Fig. 2 Standard deviation of the horizontal barotropic stream function calculated as the ocean counterclockwise horizontal transport integrated vertically. Left: one member of NEMOVAR-ORAS4, right: one member of the historical simulation performed with EC-Earth historical simulation (NOINI). The rows represent respectively January, May and September. The standard deviation for each calendar month has been calculated over the 1971-2000 period after removing the annual cycle.

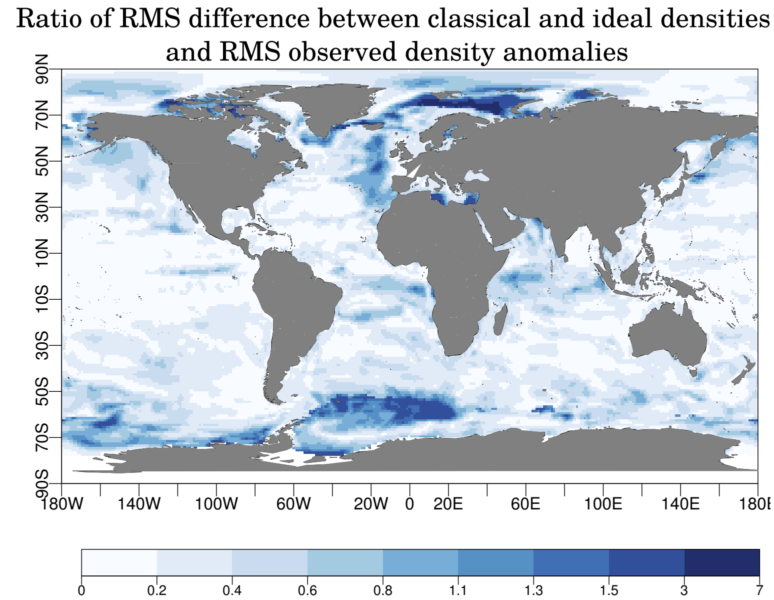

Fig. 3 Ratio between the root mean square difference between $\rho^{standard}$ and ρ^{a} over the standard deviation of the observed anomalies (i.e. the anomalies of ρ^{o}) from NEMOVAR-ORAS4 at sea surface, calculated from the initial conditions of November between 1960 and 2004.

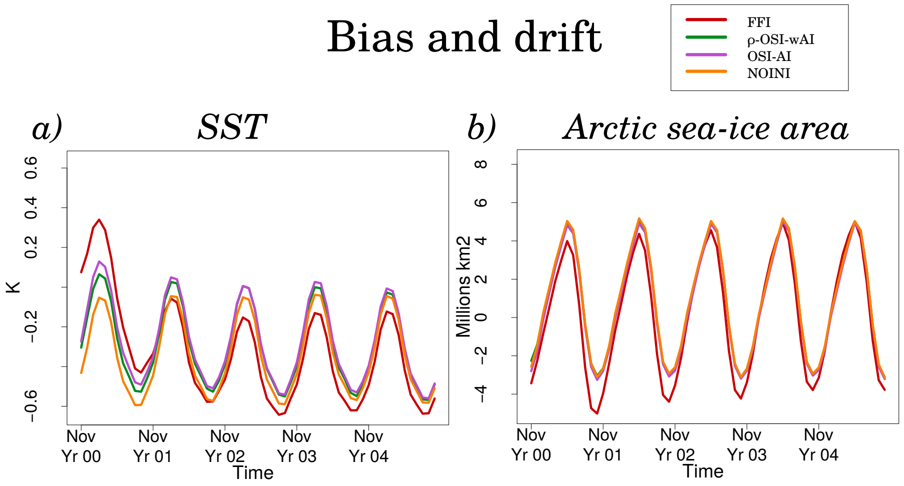

Fig. 4 Drift of a) Mean SST between 60°S and 65°N calculated with ERSST reference, b) Arctic sea-ice area calculated with the HistDfsNudg sea ice reconstruction as reference. FFI in red, ρ-OSI-wAI in green, OSI-AI in purple and NOINI in orange.

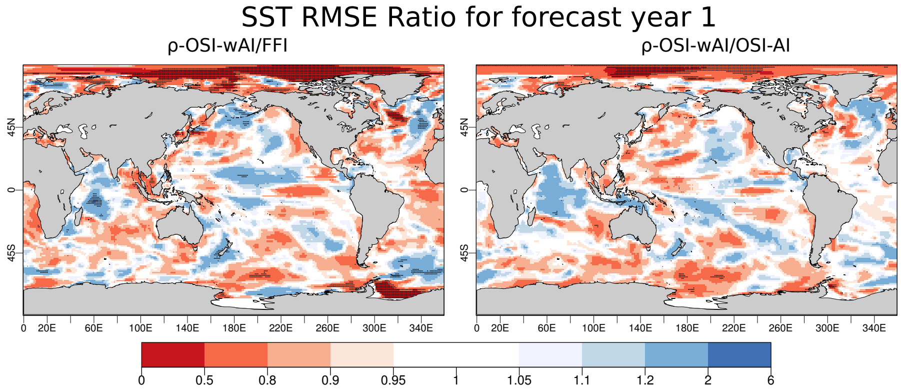

Fig. 5 Ratio of sea surface temperature RMSE maps for the first forecast year, calculated against ERSST data: the first panel is the ratio between ρOSI-wAI/FFI, the second panel between ρOSI-wAI/OSI-AI. When the ratio is smaller than 1 (red, yellow areas) the ρOSI-wAI experiment has smaller RMSE, i.e. improves the skill of the prediction. Vice versa, when the ratio is larger than 1 (region in blue) the skill is degraded. The black dots over the colours indicates where the RMSE is 95% significant according to a Fisher test.

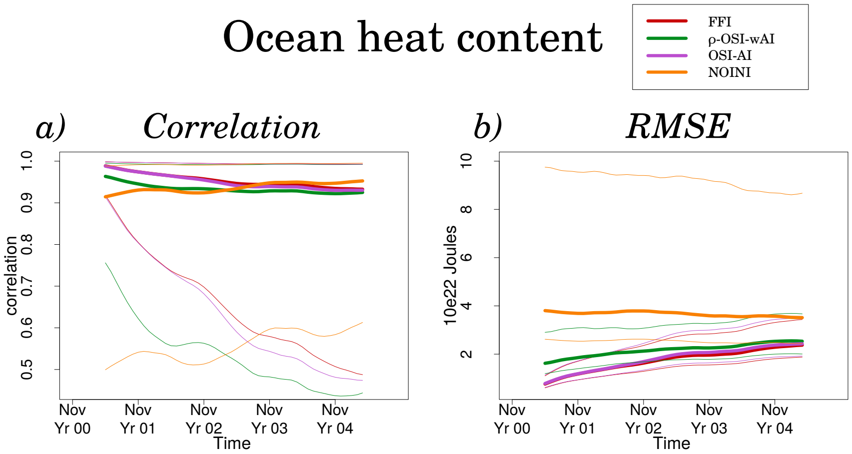

Fig. 6 Correlation and root mean square error for the global mean ocean heat content of the whole water column, with respect to NEMOVAR-ORAS4. Red for FFI, green for ρ-OSI-wAI, purple for OSI-AI and orange for NOINI. The thin lines represent the 95% confidence interval obtained with a t-distribution for the correlation and a χ^2 distribution for the RMSE. The dependence between the hindcasts is accounted for in the computation of the confidence interval using the Zebiak (1995) and Von Storch and Zwiers (1999) formula.

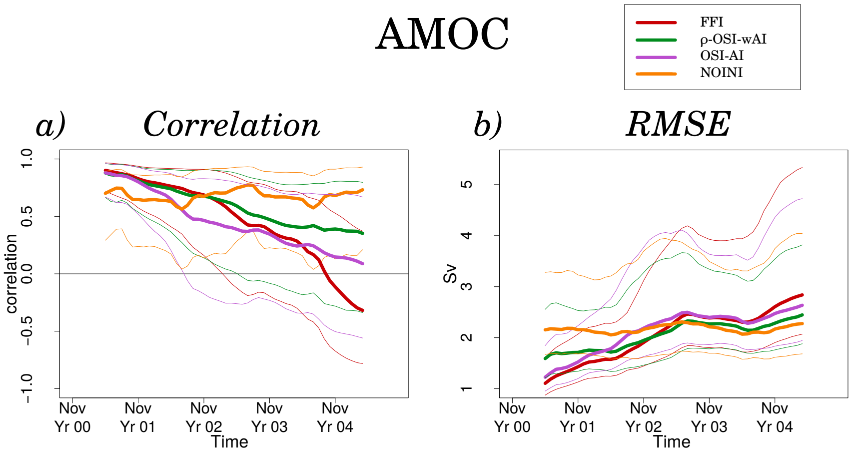

Fig. 7 Correlation and root mean square error for the Atlantic meridional overturning stream function averaged in the $40-55°$N band and 1-2 km depth with respect to NEMOVAR-ORAS4. Red for FFI, green for ρ-OSI-wAI, purple for OSI-AI and orange for NOINI. The thin lines represent the 95% confidence interval obtained with a t-distribution for the correlation and a χ^2 distribution for the RMSE. The dependence between the hindcasts is accounted for in the computation of the confidence interval using the Zebiak (1995) and Von Storch and Zwiers (1999) formula.

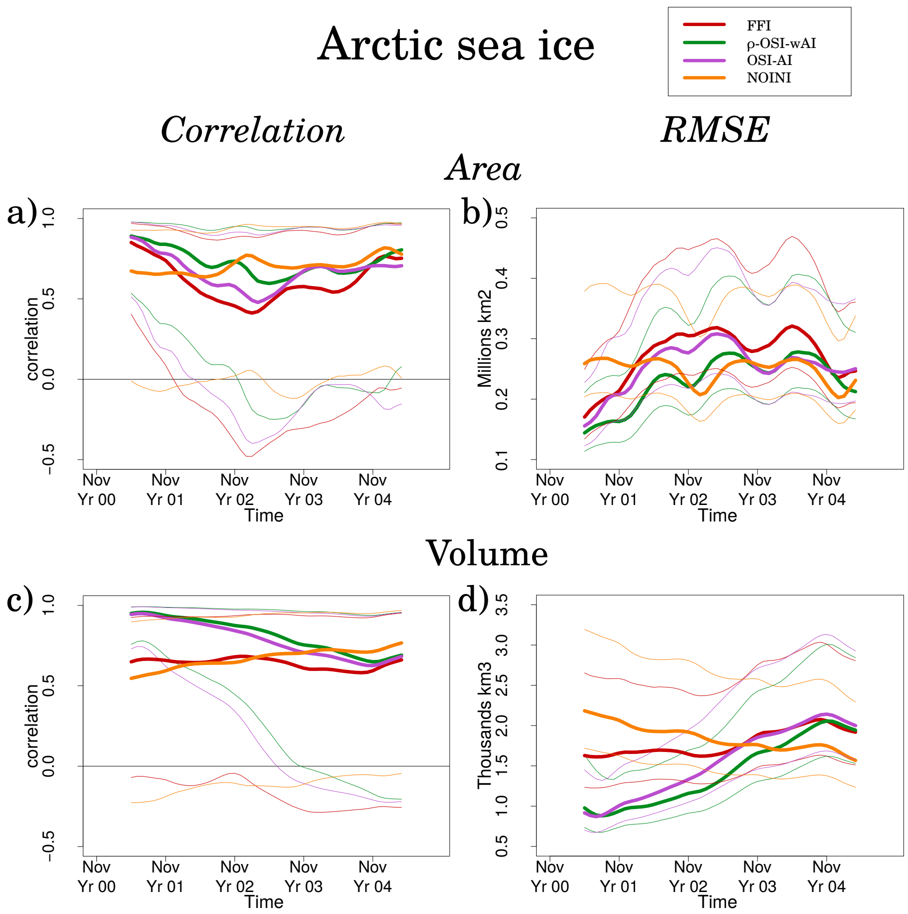

Fig. 8 Correlation and RMSE of Arctic sea-ice area (a-b) and sea-ice volume (c-d). The reference data is the HistDfsNudg sea ice reconstruction. Red for FFI, green for ρ-OSI-wAI, purple for OSI-AI and orange for NOINI. The thin lines represent the 95% confidence interval as in the previous figures.

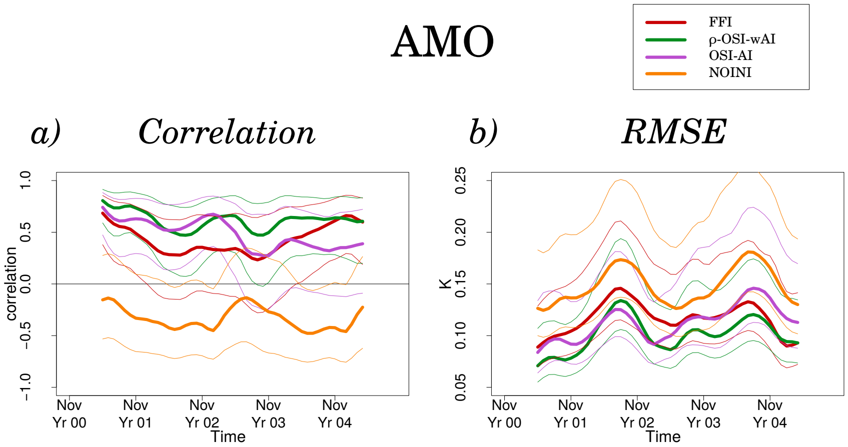

Fig. 9 Atlantic multidecadal oscillation a) correlation and b) RMSE with respect to ERSST data. Red for FFI, green for ρ-OSI-wAI, purple for OSI-AI and orange for NOINI. The thin lines represent the 95% confidence interval as in the previous figures.

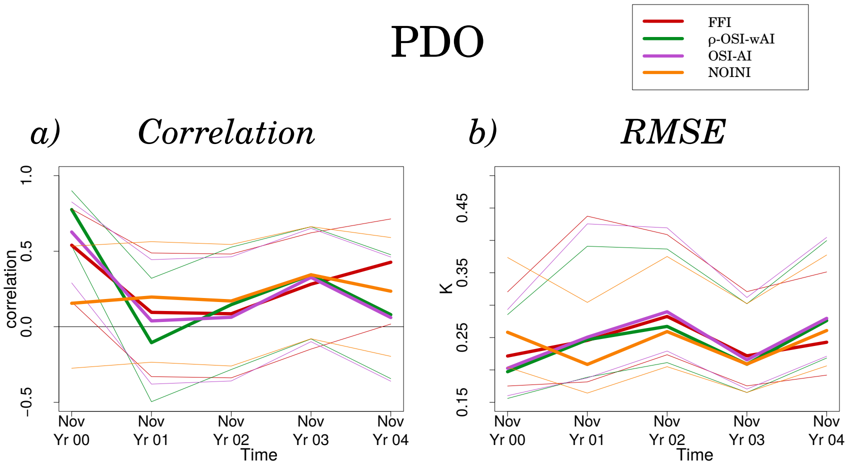

Fig. 10 Pacific decadal oscillation (20N-65N) a) correlation and b) RMSE with respect to the ERSST data. Red for FFI, green for ρ-OSI-wAI, purple for OSI-AI and orange for NOINI. The thin lines represent the 95% confidence interval as in the previous figures.

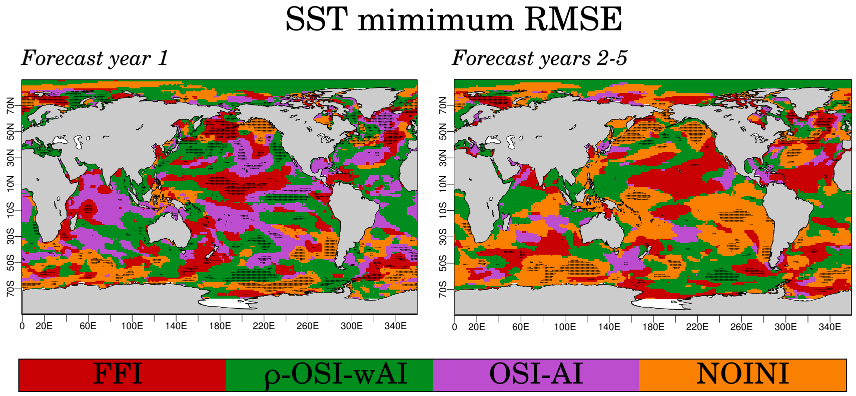

Fig. 11 Minimum RMSE of SST respectively for the forecast year 1 (left panel) and 2-5 (right panel). Each grid point takes the colour of the experiment with the smaller RMSE over the first forecast year on the left and over the forecast years 2-5 on the right. The black dots indicate the regions where the minimum RMSE differs from the second minimum RMSE for more than 0.05 K. In red the FFI experiment, in green the ρ-OSI-wAI, in purple the OSI-AI and in orange the NOINI experiment.