

Prospects for decadal climate prediction in the Mediterranean region

Virginie Guemas^{1,2}, Javier García-Serrano^{1,3}, Annarita Mariotti⁴, Francisco
Doblas-Reyes^{1,5}, Louis-Philippe Caron^{1,6}

1 – Institut Català de Ciències del Clima, Barcelona, Spain

2 – Centre Nacional de Recherches Météorologiques, Groupe d'Etude de
l'Atmosphère Météorologique, Météo-France, UMR3589, Toulouse, France

3 – LOCEAN-IPSL, Université Pierre et Marie Curie, Paris, France

4 – National Oceanic and Atmospheric Administration, Climate Program
Office, Silver Spring, USA

5 – Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

6 – Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

Keywords: climate prediction, initialization, Atlantic Multidecadal Oscillation,
Mediterranean climate

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.2379

Abstract

The Mediterranean region stands as one of the most sensitive to climate change, both in terms of warming and drying. On shorter timescales, internal variability has substantially affected the observed climate and in the next decade might enhance or compensate long-term trends. Here we compare the multi-model climate predictions produced within the framework of the CMIP5 (Coupled Model Intercomparison Project Phase 5) project with historical simulations to assess the level of multi-year climate prediction skill in the Mediterranean region beyond that originating from the model accumulated response to the external radiative forcings. We obtain a high and significant skill in predicting 4-year averaged annual and summer mean temperature over most of the study domain and in predicting precipitation for the same seasons over northern Europe and sub-Saharan Africa. A lower skill is found during the winter season but still positive for temperature. Although most of this high skill originates from the model response to the external radiative forcings, the initialization contributes to the temperature skill over the Mediterranean Sea and surrounding land areas. The high and significant correlations between the observed Mediterranean temperatures and the observed AMO in the summer and annual means are captured by the CMIP5 ensemble which suggests that the added skill is related to the ability of the CMIP5 ensemble to predict the AMO. Such a link to the AMO seems restricted to Western Africa and summer means only for the precipitation case.

1. Introduction

The Mediterranean region is projected to be among the most heavily affected by the 21st century greenhouse gas induced climate change, with significant regional warming and

drying by the end of the century (e.g. Mariotti et al., 2008; Mariotti et al, 2013). On the shorter term, a few decades into the future, decadal variability of both internal and external origin, will largely determine the conditions to be experienced in the region, as already seen over the course of the 20th century (Mariotti and Dell'Aquila, 2012). A main question and one of great societal relevance, is to which extent will natural variability enhance or reduce externally forced changes and for how long. In view of the seemingly robust climate change signal in the Mediterranean region, projected external forcing constitutes a source of decadal predictability for this region. What remains to be understood is the degree of predictability of regional internal decadal variability and its impact on our ability to predict future changes a few years ahead.

Early potential predictability studies highlighted the Atlantic Ocean as the most promising region for decadal climate prediction (Griffies and Bryan, 1997; Boer, 2000; Pohlmann et al., 2004; Collins et al. 2006). The Atlantic Ocean exhibits strong multi-decadal variability which involves variations in the strength of the ocean thermohaline circulation associated with a large scale signature in sea surface temperatures (SST) referred to as the Atlantic multidecadal oscillation (AMO, Kerr, 2000; Knight et al, 2005, 2006). Much of the current quest to predict future decadal climate variations hinges on the potential predictability associated with the variations of the ocean gyres and the Atlantic meridional overturning circulation (AMOC). At the edge between seasonal forecasting and climate change projections, the decadal climate prediction exercise consists of initializing a climate model from an estimate of the observed climate state while also prescribing the changes in external radiative forcings, assuming these are known (Hawkins and Sutton 2009; Meehl et al., 2009; Murphy et al, 2010; Doblas-Reyes et al, 2011). Based on this approach, hindcast (i.e. retrospective climate prediction) studies (Smith et al, 2007, 2012; Keenlyside et al, 2008; Pohlmann et al., 2009) showed that the Atlantic Ocean stands out as a region of increased

forecast skill with initialization. The initialization allows for more accurate predictions of the AMO compared to coupled experiments accounting only for the model accumulated response to changes in external radiative forcings (García-Serrano et al, 2012), in particular in the multi-model ensemble climate hindcasts produced in the framework of the Climate Model Intercomparison Project Phase 5 – CMIP5 (Doblas-Reyes et al., 2013; García-Serrano et al., 2014).

A recent study using observational data and a regional climate model (Mariotti and Dell’Aquila, 2012) suggests a linkage between the Atlantic multidecadal oscillation/variability (AMO/AMV) and decadal climate variability in the Mediterranean region. Mediterranean SST exhibit AMO-like variability, which in turn is reflected in AMO-like evaporation variability (Fenoglio et al., 2013). There is also a linkage between the AMO and Mediterranean surface air temperatures which appears to be strong in summer, also present to a lesser extent in the transition seasons, but negligible in winter. The modeling results from Mariotti and Dell’Aquila (2012) suggest that atmospheric processes are the primary mechanism connecting the AMO with the Mediterranean rather than a direct influence of the Atlantic on the Mediterranean Sea. This study therefore suggests the potential for a certain degree of decadal climate (namely, temperature and evaporation) predictability in the Mediterranean region, to the extent that the AMO is predictable. In contrast, decadal variability of regional precipitation was found to be largely affected by the seasonal manifestations of the North Atlantic Oscillation, rather than the AMO. It is currently unknown whether there is any predictability in NAO variability. Hence the prospects for predicting regional precipitation variability, outside of what may be the externally forced long-term drying trend, are also unclear.

In this article, we aim at assessing the level of decadal climate prediction skill in the Mediterranean region beyond that originating from the model accumulated response to the

external radiative forcings, and the seasonal dependence of this skill, contrasting summer (defined as July to August throughout our study) and winter (December to February) values. We also aim at assessing to what extent the added-value from the initialization is originating from the predictability of the AMO and its linkage to the Mediterranean region. It should be noted that the focus of the present study is not on the AMO drivers (Ottera et al. 2010; Booth et al. 2012; Terray 2012; Zhang et al. 2013) but on its prediction skill and the predictability gained from simulating its teleconnections, in particular over the Mediterranean region as suggested by Matei et al (2012). These assessments rely on the CMIP5 multi-model ensemble (MME) climate hindcasts (Taylor et al., 2012) which are introduced in Section 2 together with the observational data used for verification and the skill scores used to measure the performance of this MME. Section 3 then illustrates the performance of the CMIP5 MME in predicting Mediterranean temperatures and precipitation, and compares this skill with the one from historical climate simulations which do not take into account the internal variability as a source of predictability and which contain potential accumulated errors in the model response to the external radiative forcings. The added-value from the initialization is then related to the ability of the MME to capture the linkage between the AMO and those regional climate variables. Section 4 and 5 provide respectively a discussion of major results and the conclusions from our study.

2. Methodology

2.1 Model data

The multi-model ensemble (MME) analyzed here includes the contributions to the CMIP5 project produced with the HadCM3, MRI-CGCM3, MIROC4h, MIROC5, CanCM4, CNRM-CM5, MPI-M, GFDL-CM2, CMCC-CM, IPSL-CM5 and EC-Earth v2.3 coupled

climate models. These contributions consist in 10-year long hindcasts initialized from an estimate of the observed climate state every 5 years over the period 1960-2006 between November, 1st and the following January, 1st. This ensemble is referred to as Init thereafter. Complete calendar years from January to December are considered as forecast years here. Two ranges of forecast years are analyzed in the following: from forecast years 2 to 5 and 6 to 9, which are classically used as target periods for analyses of climate predictions (Doblas-Reyes et al., 2013). The MME hindcast set gives 10 predicted values for a specific forecast range. The forecast fields are interpolated using a conservative method onto the grid of the observational dataset used for verification before any analysis. More details on the model experiments analyzed in this work are provided in Table 1.

A sister ensemble has been built from the “historical” coupled simulations and the first years of the climate projections following the CMIP5 RCP4.5 (Representative Concentration Pathway 4.5W/m², Taylor et al., 2012) scenario run with the same models. In these experiments, the simulated natural variability is not constrained to follow the contemporaneous observed one except through the effect of the external radiative forcings, which are exactly the same as in the Init MME. Additionally, errors in the response to the external radiative forcings could accumulate over time. This ensemble is referred to as NoInit thereafter. Both Init and NoInit CMIP5 experiments include observed time-dependent solar irradiance variations and volcanic aerosols until 2005 whereas these would not be known in a true prediction exercise.

The Init and NoInit ensembles do not contain the same number of members. We have checked that restricting those ensembles to the subsets allowing for the same number of members for each model in the two ensembles do not affect our main conclusions (not shown).

2.2. Observational data

For temperature verification, we have used a combination of land surface temperatures from the National Centers for Environmental Prediction (NCEP) GHCN/CAMS dataset (Fan and van den Dool, 2008), SST from the NCDC ERSST v3b dataset (Smith et al, 2008) and, north of 60°N, the GISTEMP dataset with 1200km decorrelation scale (Hansen et al 2010). This dataset covers the 1960-2012 period. For precipitation, a land-only monthly dataset is considered because of the sparse observational coverage over the oceans before 1979. We use the Global Precipitation Climatology Centre (GPCC) v5 data (Rudolf et al 2010) for the 1960-2012 period. For sea level pressure, we have used the HadSLP2 dataset (Allan and Ansell, 2006) for the 1960-2004 period.

2.3. Forecast skill assessment

In our study we focus on the analysis of the skill for the annual values and contrast the summer (JJA) and winter (DJF) values. We compute JJA, DJF and annual means for both the model experiments and observations. The performance of the MME is assessed for the average of the predicted values over forecast years or seasons 2 to 5 and 6 to 9. For the winter season, the average over forecast seasons 2 to 5 and 6 to 9 comprises the December values from years 1 to 4 and 5 to 8 respectively. When initialized from an estimate of the observed climate state, climate models drift towards their attractor, which leads to systematic errors in climate forecasts. These systematic errors depend on the forecast time and we therefore apply a drift correction to our multi-model ensemble (Goddard et al., 2012; Meehl et al., 2014). Anomalies are computed using the so-called “per-pair” methodology (Garcia-Serrano and Doblas-Reyes, 2012): the modeled or observed climatology is defined as a function of the

forecast time (multi-year or multi-season average), by averaging the predicted variable (temperature or precipitation) across the start dates taking into account only the predicted values for which observations are available at the corresponding date; a forecast-time-dependent climatology is computed for each individual model averaging all members. The climatology of each model is then subtracted from each hindcast to obtain anomalies over the whole predicted period for each member of each model:

$$X'_{msf} = X_{msf} - \frac{1}{(M-1)(S-1)} \sum_M \sum_S X_{msf} (O_{sf} \neq NA)$$

$$O'_{sf} = O_{sf} - \frac{1}{(S-1)} \sum_S O_{sf} (O_{sf} \neq NA)$$

where X_{msf} is the model variable at forecast time f , for start date s and member m , O_{sf} is the corresponding observation, X'_{msf} and O'_{sf} , their respectively anomalies, M is the number of member, S the number of start dates and NA refers to a missing value. Then, ensemble-mean anomalies are computed for each model averaging all its members and the multi-model ensemble-mean anomalies are computed giving an equal weight to each model, regardless of the number of ensemble members available for that model. The same method is applied to the observations to obtain anomalies over the whole verification period. All the climate prediction scores described below are computed from those anomalies.

Hindcast prediction skill is measured using the anomaly correlation coefficient (ACC), the mean square skill score (MSSS) or the Brier Skill Score (BSS). The ACC is computed by correlating the predicted MME anomalies with the observed ones across the start date dimension:

$$ACC_f = \frac{\frac{1}{S-1} \sum_S X'_{sf} O'_{sf}}{\sqrt{\left(\frac{1}{S-1} \sum_S X'^2_{sf} \right) \left(\frac{1}{S-1} \sum_S O'^2_{sf} \right)}}$$

where X'_{sf} is the MME mean anomaly for forecast time f and startdate s , and O'_{sf} is the corresponding observed anomaly. A t -test, after a Fisher-Z transformation, is performed to assess the significance level of an ACC or of a difference in ACC, accounting for the autocorrelation of the data when computing the degrees of freedom following von Storch and Zwiers (1999). Indeed, hindcasts initialized every 5 years cannot be considered as independent when attempting to predict decadal climate variability, hence our reduction of the degrees of freedom based on von Storch and Zwiers (1999). It is worth noting that only 9 or 10 observations (when verifying predictions for forecast years 6-9 or 2-5 respectively) can be used to compute the ACC or any other skill score, and those 9-10 data are reduced by about a factor 2 to obtain the number of effective independent data. Increasing the frequency of the start dates would not increase significantly the number of independent data since the additional hindcasts would still not be independent for analyses of decadal climate variability. Extending backward the period which is sampled by our start dates would be necessary to obtain more independent hindcasts and more robust estimates of the added-value from initialization in terms of skill scores. Obtaining accurate initial conditions is unfortunately harder to achieve as we extend the hindcast period back into the past.

The MSSS is computed as 1 minus the ratio of the mean square error (MSE) of the predicted per-pair MME anomalies over the MSE of a climatological forecast (Goddard et al, 2013):

$$MSSS_f = 1 - \frac{\frac{1}{S-1} \sum_s (X'_{sf} - O'_{sf})^2}{\frac{1}{S-1} \sum_s O'^2_{sf}}$$

A Fisher test is performed to assess the significance level of a MSSS accounting for the autocorrelation of the data following von Storch and Zwiers (1999). Since the MME is used in the forecast skill assessment through the ACC and the MSSS, externally forced signals and

any signal due to the initialization will stand out (that is only the robust patterns, consistent across members and models will stand out). In contrast, any “noise” will be to a large extent eliminated by the MME averaging operation.

The BSS is computed for the lower, middle and upper terciles:

$$BSS_{ft} = 1 - \frac{\frac{1}{S} \sum_s (p_{fst} - o_{fst})^2}{\bar{o}_t (1 - \bar{o}_t)}$$

where p_{sft} is the forecast probability for a given tercile t to occur, for forecast time f and startdate s , o_{sft} is its observed probability (0 or 1) and \bar{o}_t is its observed climatological probability. The model terciles and the model forecast probabilities are estimated from the multi-model ensemble of model mean anomalies which makes an ensemble of 11 values (models) for each start date, each variable and each forecast averages. This approach ensures equal representativeness for all models in a similar fashion as we operated for the deterministic skill measures (ACC and MSSS).

The skill gained by the initialization process is assessed by computing the difference between Init and NoInit ACC. Whenever this difference in skill is positive the initialization process increases the skill beyond what is achievable based on the integration of the model response to external radiative forcing only. For temperature, we compute additionally the skill scores after detrending the observed, Init and NoInit grid-point temperature by subtracting their respective global land mean averaged between 60°S and 60°N at land points and global ocean mean averaged between 60°S and 60°N at ocean points. We have checked that a linear detrending led to similar qualitative conclusions (not shown). The robustness of our conclusions on the added-value from the initialization is also assessed by computing the ratio between the root mean square error (RMSE) of Init over the RMSE of NoInit. A Fisher test is performed to assess the significance level of this ratio accounting for the autocorrelation of the data following von Storch and Zwiers (1999).

The linkage between the AMO and the Mediterranean temperature and precipitation is assessed via correlation maps of observed temperature and precipitation with the observed AMO index. The AMO index used here is defined as in Trenberth and Shea's (2006): the SST averaged over the North Atlantic (80°W - 0°E / EQ- 60°N) minus the global (60°S - 60°N) SST average. Previous research shows that this definition largely decontaminates the AMO index from the long-term warming (e.g. van Oldenborgh et al. 2009, 2012). This AMO index represents the differential warming of the North Atlantic Ocean with respect to the global oceans. The ability of the CMIP5 MME to capture the linkage between the AMO and the Mediterranean temperature and precipitation is assessed by computing similar correlation maps but for the predicted variables. The temperature is detrended prior to the computation of those correlation maps whereas the precipitation is not. The subsequent comparison between the AMO correlation patterns and the ACC patterns aims at highlighting any AMO derived predictability. By extending these analyses to the sea level pressure, we aim to highlight the atmospheric mechanisms by which the AMO predictability contributes to the Mediterranean temperature and precipitation skill and to which extent those are captured by our MME.

3. Results

3.1. Temperature skill and role of the AMO

The Init ACC maps (Figure 1, top row) show a high and significant skill over most of the study domain for the temperature averaged over forecast years 2 to 5 (left panel) and 6 to 9 (right panel). Though slightly lower than the ACC, the MSSS maps show a very consistent pattern of high and broadly significant skill (Figure 1, fourth row). Most of this high skill originates from the external radiatively forced climate signal since the differences in ACC between Init and NoInit are small (Figure 1, second row). None of the differences in ACC are

significant. However, as mentioned in section 2.3, only 9 (right panel) or 10 (left panel) data are used to compute the ACC, and those 9-10 data, only 5 years apart, cannot be considered as independent, hence a high significance threshold. Additionally, the skill in NoInit is already very high due to the long-term warming trend. Hence, any increase in ACC when capturing the superimposed internal variability or correcting the model response to the external radiative forcings prior to the initialization date is of minor magnitude compared to the NoInit ACC. Despite the lack of significance of the differences in ACC, those differences show an increase in ACC over the Eastern Atlantic and the Mediterranean Sea. To highlight the added-value from the initialization, we re-compute the maps of difference in ACC between Init and NoInit after detrending the grid-point observed and modeled (both Init and NoInit) temperature as described in section 2.3 (Figure 1, third row). It highlights the increase in ACC thanks to the initialization over the Eastern Atlantic and the Mediterranean Sea and a decrease over Northern Europe and Sub-Saharan Africa. The ratio between the Init RMSE and the NoInit RMSE (Figure 1, bottom row) exhibits again a similar pattern of improved performance with initialization over the Eastern Atlantic and the Mediterranean Sea and of decreased performance over Northern Europe and Sub-Saharan Africa. The probabilistic verification (Figure S2) shows broadly consistent features apart from the Sub-Saharan Africa where there is no clear signal. In agreement with the conclusions from the deterministic verification, we observe: 1) a larger BSS in Init than NoInit over the Eastern Atlantic for all terciles and both forecast averages except for the lower tercile and forecast years 6-9, 2) a larger BSS in Init than NoInit over the Mediterranean Sea for the middle tercile but a noisier response for the other two terciles, 3) a lower BSS in Init than in NoInit over the Northern Europe for the lower and upper terciles but a noisier response for the middle tercile.

The increase in ACC over the Eastern Atlantic and the Mediterranean Sea with initialization might be associated with their linkage to the AMO, the latter already known to

have an increased skill with initialization in the CMIP5 MME (Doblas-Reyes et al., 2013; Garcia-Serrano et al, 2014). The observed annual AMO index underwent a positive phase during the first decade of the hindcast period, followed by a negative phase that lasted about three decades and a final positive phase, stronger than the first positive phase (Figure 2 top row). Those multi-decadal oscillations produce a slight positive trend over the hindcast period. NoInit does not exhibit more than this slight positive trend at both forecast averages. A higher and significant skill in predicting the annual AMO index is obtained in Init which captures the multi-decadal oscillations, but also successfully reproduces the minimum in the 70s and the recent maximum at forecast years 2 to 5. The positive trend is less pronounced in the winter (Figure 2 second row) than in the annual mean AMO index. Accordingly a larger added-value of the initialization appears in winter: the difference between the Init and NoInit ACC is larger. By contrast, the positive trend is more pronounced in summer (Figure 2 bottom row) and the differences between Init and NoInit performance weaker. The difference between Init and NoInit performance decreases at forecast years 6 to 9 as compared to forecast years 2 to 5 (Figure 2, right column).

The correlation maps of observed temperature with the observed AMO index (Figure 3, top row) exhibit a rotated T-structure, with positive correlation covering the Eastern Atlantic, Northern Africa and the Mediterranean Sea while negative correlations appear over Europe. The maps for forecast years 2-5 (left panel) and 6-9 (right panel) are very similar even though they are based on different observed time windows ([1962-1965] to [2006-2009] for forecast years 2-5; [1966-1969] to [2006-2009] for forecast years 6-9) which confirm the robustness of this correlation pattern against sampling (which is not the case for sea level pressure for example as mentioned later). The pattern of correlations between the ensemble-mean temperature and the ensemble-mean AMO index is smoother than the observed pattern because of the MME averaging operator. Init captures the linkage between the AMO and the

Eastern Atlantic, Mediterranean Sea and northern North Africa (Figure 3, middle row), which is consistent with the increase in ACC with initialization in those regions. Init however exhibits negative correlations between the AMO index and the temperature between about 10°N and 20°N and positive correlations over Europe which do not appear in the observations, probably at the origin of the decrease in ACC with initialization in both regions. The AMO index exhibits a positive trend, related to its negative-to-positive phase transition described above, which might correlate well with the external radiatively forced warming signal. Subtracting the correlation between the NoInit AMO and temperature from the correlation between the Init AMO and temperature allows isolating the linkage between the temperature and the multidecadal AMO oscillations around the model accumulated response to the external radiative forcings, identifying the AMO-related signature coming from the initialization. Such differences between Init and NoInit correlations of the AMO index with temperatures (Figure 3, bottom row) share strong similarities with the differences between Init and NoInit ACC after detrending (see Figure 1, third row). The ability of the Init MME to capture the observed AMO variability seems to contribute substantially to its skill in predicting multi-year averaged temperature over our study domain.

In boreal summer, the Init ACC is similarly high and significant in most of the study domain (Figure 4, top row), consistent with the Init MSSS pattern, and mostly originating from the externally forced climate signal. The differences in ACC between Init and NoInit though still small, show a slightly different pattern than for the annual mean temperature, and less spatially coherent. The ratio of the Init RMSE over the NoInit RMSE provides consistent patterns of added-value from the initialization with the differences between Init ACC and NoInit ACC. There is still an increased skill over the Eastern Atlantic with initialization but the added-value over the Mediterranean Sea is less pronounced. There are also two additional areas where the initialization tends to increase the skill: over Eastern

Europe and sub-Saharan Africa. In those two regions, the correlation between the observed JJA AMO index and the observed JJA temperature has an opposite sign to the one for annual temperature: it becomes positive over Eastern Europe and negative over sub-Saharan Africa (Figure 5, top row). The latter is probably due to the positive link between the AMO and the West African monsoon (Mohino et al. 2011; van Oldenborgh et al. 2012), which negatively feeds back on local land-surface temperature (Giannini et al. 2003, 2005; Kucharski et al. 2012). The correlation between the Init JJA AMO index and the Init JJA temperature is positive over most of the domain except for the sub-Saharan Africa (Figure 5, middle row), as in the annual correlation maps (Figure 3, middle row). The Init and observed correlation maps are therefore more consistent over Eastern Europe and sub-Saharan Africa for the JJA means than for the annual ones, hence a larger added-value of the initialization in JJA over those two regions. The probabilistic verification (Figure S3) consistently shows an increased BSS with initialization in the Eastern Atlantic for the upper and middle terciles and in the Mediterranean region and over Eastern Europe for all the terciles and forecast averages.

The winter (DJF) stands out as the season for which the increase in ACC with initialization is the largest, especially over the oceans (Figure 6, second and third rows), and consistently the season for which the decrease in RMSE is the largest (Figure 6, bottom row). The skill, both deterministic (Figure 6) and probabilistic (Figure S4), remains improved over the Eastern Atlantic but it is also improved over the Eastern North Africa, Eastern Mediterranean and Eastern Europe, in terms of RMSE and ACC. The skill is strongly reduced though over Western Africa and Western Europe, especially for forecast years 2 to 5, and over the Arabian Peninsula. DJF is also the season for which the Init ACC (Figure 6, top row) and MSSS (Figure 6, fourth row) skill are the lowest, as well as the BSS for all terciles (Figure S4). The warming trend has indeed lower relative amplitude in this season compared to the amplitude of the internal natural variability (Figure S1). The correlation maps of

observed DJF temperature with the observed DJF AMO index rather exhibit an L-pattern with positive values over Eastern Atlantic, Northern Africa, Eastern Mediterranean Sea and the Arabian Peninsula and negative values over Europe (Figure 7, top row). This pattern is however not robust against the sampling (left and right panels) and slightly different from the one obtained by Mariotti and Dell'Aquila (2012) over a different period. The observed linkage between the DJF AMO and the DJF temperature is not properly captured by Init over Europe. Over the Mediterranean Sea, the signature is noisy (Figure 7, middle row). The increase in skill with initialization in those regions therefore does not seem to be related to the ability of the MME in predicting the AMO. However, Init is able to simulate the observed AMO correlation pattern over the Eastern Atlantic and Northern Africa reasonably well and over the Arabian Peninsula to some extent. Although boreal winter (DJF) is the season for which the initialization allows for the largest increase in skill in predicting the AMO index (Figure 2 middle row), there is little agreement between the pattern of differences in ACC skill Init-NoInit and the pattern of Init-NoInit differences in correlations of the temperature with the AMO index (Figure 7, bottom row). The ability of the MME to capture the observed winter AMO seems to contribute to the skill in predicting winter temperature only over confined regions such as the Egypt-Sudan area and Eastern Atlantic.

3.2. Precipitation skill and role of the AMO

As discussed in the introduction, previous studies have not highlighted any AMO-related precipitation variability in the Mediterranean. Hence the basis for decadal precipitation prediction is currently unclear. The precipitation patterns additionally exhibit less spatial coherence than the temperature patterns. We therefore face a double-challenge for precipitation forecast skill: the lack of knowledge about predictability sources and the spatial noise. Nevertheless, an exploratory analysis of decadal prediction skill is presented here to

potentially gain further insight in regional precipitation predictability including that from externally forced causes.

Despite the noisy ACC Init patterns for total precipitation (due to the low spatial coherence of precipitation events compared to temperature), significant ACC appear over Northern Europe and over sub-Saharan Africa for both forecast year ranges: 2 to 5 and 6 to 9 (Figure 8, top row). A high MSSS is also seen in those two regions, though not significant (Figure 8, top row). A high MSSS is also seen in those two regions, though not significant (Figure 8, third row) as well as a high BSS for the lower and upper terciles (Figure S5). The Init skill is however generally lower for precipitation than for temperature, mainly because the externally forced signal has lower relative amplitude compared to the internally generated variability (Figure S1). The high skill in northern Europe and sub-Saharan Africa originates mainly from the accumulated model response to the external radiative forcing since weak differences in ACC appear between Init and NoInit (Figure 8, second row), except for part of Western Africa and for the South-Eastern tip of the Arabian Peninsula. A larger BSS in Init than NoInit is also seen over part of Western Africa for the lower and upper terciles and for both forecast averages (Figure S5). The long-term warming has been shown to have an impact on the West African monsoon, with maximum amplitude over the western Sahel (Ting et al. 2009; Mohino et al. 2011; Smith et al. 2012). A larger added-value from the initialization (Figure 8, second and fourth rows) is obtained for precipitation than for temperature, because the signal from internal variability has larger relative amplitude compared to the externally forced signal (Figure S1), hence a larger room for improvement with initialization. The patterns of precipitation forecast skill are however much less spatially coherent than those of temperature hence we have a lower confidence in their physical robustness. The correlation maps of observed precipitation with the observed AMO index exhibit a tripolar pattern, with a positive band over northern Europe, a negative band extending from Western Europe toward the Arabian Peninsula and a positive band over

Northern Africa (Figure 9, top row). This tripolar pattern is captured by Init, although with a slight southward (northward) shift of the negative band over Western (Eastern) Europe (Figure 9, middle row). The correlations are higher and less noisy in Init than in the observations since the MME mean operator smoothes out any signal other than externally forced or persisting the initial conditions. The core of the positive correlation bands, i.e. over Northern Europe and sub-saharan Africa, which are well captured by Init corresponds to high ACC and MSSS in Init. In transitional areas between positive and negative correlation bands of precipitation with the AMO index (southern Europe and Mediterranean Sea), Init fails to capture accurately the linkage between precipitation and the AMO (slight shift of the patterns) and the ACC and MSSS are low in those areas. The only region where the Init-NoInit differences in correlation with the AMO index (Figure 9, bottom row) are consistent with the pattern of Init-NoInit differences in ACC skill (Figure 8, second row) locates over the westernmost part of North Africa.

In summer (JJA), high and significant ACC cover also most of Northern Africa and a few points in Northern Europe (Figure 10, top row). The pattern of Init skill is robust against the score considered but the MSSS (Figure 10, third row) and the BSS (Figure S6) show more modest performance. MSSS are significant over Egypt but this performance is not supported by the other skill scores. Although the differences in ACC between Init and NoInit are not significant, a spatially coherent increase in ACC with initialization appears over Western Africa (Figure 10, second row), together with a decrease in RMSE (Figure 10, fourth row). This low but positive added-value from the initialization in multi-year prediction skill over the West African monsoon region might be associated with the positive, although not statistically significant, skill of the Sahelian precipitation regime found by García-Serrano et al. (2013). The correlation maps of observed JJA precipitation with the observed JJA AMO index exhibit a similar tripolar pattern to the annual mean maps, but with a negative band

centered on the Mediterranean Sea (Figure 11, top row). The positive center of action over Africa is well captured by Init, including in terms of shape, which could explain its high skill in predicting JJA precipitation in this area (Figure 11, second row). As for annual means, the westernmost part of North Africa is the region showing the largest consistency between Init-NoInit differences in correlation with the AMO index (Figure 11, bottom row) and Init-NoInit differences in ACC skill (Figure 10, second row). This result is consistent with previous evidence on the predictor role of the AMO upon the Sahelian rainfall (Mohino et al. 2011; van Oldenborgh et al. 2012; García-Serrano et al. 2013).

The winter Init ACC patterns are particularly noisy, with a few points of significant ACC surrounded by negative ones spread over Northern Africa and a few positive ACC over Northern Europe (Figure S7). The MSSS show a significant skill over the Eastern Sahel which does not correspond to any positive ACC but with a decrease in BSS for all terciles and forecast averages (Figure S8). The correlation maps of observed DJF precipitation with the observed DJF AMO index exhibit a very different pattern from the summer one with rather a dipolar structure across the Western Mediterranean which Init fails to capture (Figure S9).

4. Discussion

As mentioned in section 3.1, the AMO fluctuations exhibit a recent phase transition from negative conditions to positive conditions, which shapes a positive trend. Note that this index is computed as the SST averaged over the North Atlantic minus the global SST average to minimize the effect of the long-term warming. For consistency, the temperature has been detrended by subtracting its global ocean mean at ocean grid points and its global land mean at land grid points prior to computing the correlation between the AMO and the grid-point temperature as explained in section 2.3. The correlation between the AMO and grid-point temperature anomalies shown in Figure 3, 5 and 7 might be affected by a residual warming signal in the grid-point temperature. Similar qualitative results were obtained with a linear detrending method (not shown) but such method does not filter out perfectly the warming signal either. Whether this residual warming and the warming trend in the AMO are related to a larger sensitivity to radiative forcings in our study domain than in other regions or whether it is related to internal variability remains unclear. Given the shortness of the available time-series, a proper assessment is highly challenging. For precipitation, we have chosen not to detrend because of the regional sensitivity to climate change and the uncertainties about whether those changes are attributable to climate change or not (Trenberth et al., 2007).

As explained in section 2.3, only 9 to 10 data can be used to compute the skill scores shown in this article and those 9 to 10 data are not independent since the CMIP5 retrospective predictions are initialized every 5 years and we focus on decadal climate variability. An extension of the period sampled by our start dates would allow for more independent hindcasts and more robust estimates of the added-value of initialization in terms of skill scores. However, an extension backward in time is highly challenging given the lack of accurate observational data to initialize the predictions. The denser observational network

that grows with time will allow for a forward extension in the coming decades and therefore for more robust estimates of the prediction skill scores. Likewise, 9 to 10 data are used to detrend time-series, which questions the accurate applicability of any method. This leads to consider that detrending methodologies represent an important source of uncertainty in decadal climate prediction.

Mariotti and Dell'Aquila (2012) suggested that the linkage between the AMO and the Mediterranean region relies mainly on atmospheric processes. To assess to what extent those atmospheric mechanisms are captured by the CMIP5 MME, we computed correlation maps of the AMO index with the sea level pressure in both the observations and the model data for annual means (Figure S10), summer (Figure S11) and winter (Figure S12) values. The pattern obtained for the observations is not robust against the sampling, contrary to the patterns obtained previously for temperature or precipitation, which suggests that the sea level pressure variability is dominated by other factors and we don't have enough data to isolate accurately the AMO influence on this variable. The MME fails to capture the observed patterns, but it remains unclear whether this failure is only an apparent failure due to a sampling issue or a real failure suggesting some room for improvement of the skill in predicting Mediterranean decadal climate variability through a better representation of the atmospheric mechanism linking the Atlantic to the Mediterranean region.

5. Conclusions

In this article, we have used the extensive CMIP5 database of multi-model ensemble decadal hindcasts to assess the state-of-the-art skill in predicting the Mediterranean temperature and precipitation and the reasons behind this skill. High and significant skill in predicting 4-year

averaged annual and summer mean temperature is found over most of our study domain. Still positive but slightly lower skill is found during the winter season. Most of the high skill originates from the model accumulated response to the external radiative forcings. The initialization contributes however to the temperature forecast skill over the Mediterranean Sea and surrounding land areas. The high and significant correlations between the observed Mediterranean temperatures and the observed AMO in the summer and annual means which are captured by the multi-model ensemble suggest that this added skill is related to the ability of the multi-model ensemble to predict the AMO. Although the winter season stands as the season with the largest increase in AMO forecast skill with initialization, the linkage between the AMO and the Mediterranean temperatures are poorly captured by the multi-model ensemble. The increased skill in annual mean temperature with initialization seems to originate from the winter for the Eastern Mediterranean and from the summer for the Western Mediterranean. The results for the annual means are more robust than the ones for the seasonal means though. High and significant skill is also found for annual and summer mean precipitation over northern Europe and sub-Saharan Africa, which is mostly explained by the model accumulated response to the external radiative forcings except over Western Africa in summer. The positive correlation between summer observed AMO and precipitation in the latter region are well captured by the multi-model ensemble which seems to contribute to the precipitation forecast skill. The AMO-related skill in summer precipitation does not seem to compensate for the weak signal in winter. The winter precipitation forecast skill is much lower and noisier than the summer and annual mean ones but improving the winter precipitation forecast skill would be highly relevant. Improving the precipitation forecast skill stands as a challenge though, given the current lack of knowledge about its predictability sources.

Acknowledgments

This work was supported by the EU-funded SPECS (FP7-ENV-2012-308378), NACLIM (FP7-ENV-2012-308299), QWeCI (FP7-ENV-2009-1-243964), CLIM-RUN (FP7-ENV-2010-1-265192) projects, the MICINN-funded RUCSS (CGL2010-20657) project, the Catalan Government and the NOAA grant NA10OAR4310208. Oriol Mula-Valls and Domingo Manubens-Gil are thoroughly acknowledged for the invaluable technical support.

References

Allan R.J., Ansell T.J. (2006) A new globally complete monthly historical mean sea level pressure data set (hadslp2). 1850-2004. *J Clim*, 19, 5816–5842.

Belluci A., Gualdi S., Masina S., Storto A., Scoccimarro E., Cagnazzo C., Fogli P., Manzini E., Navarra A., Decadal climate predictions with a coupled OAGCM initialized with oceanic reanalyses, *Clim. Dyn.*, 40,1483-1497, doi:10.1007/s00382-012-1468-z.

Boer G J, 2000, A study of atmosphere-ocean predictability on long timescales, *Clim. Dyn.*, 16, 469-472.

Booth BBB, Dunstone NJ, Halloran PR, Andrews T, Bellouin N (2012) Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature* 484:228-232.

Collins M, M Botzet, A F Carril, H Drange, A Jouzeau, M Latif, S Masina, O H Otteraa, H. Pohlmann, A Sorteberg, R Sutton and L Terray (2006) Interannual to Decadal Climate Predictability in the North Atlantic: A Multimodel-Ensemble Study. *J. Clim.*, 19, 1195–1203.

Delworth, T.L., A.J. Broccoli, A. Rosati, R.J. Stouffer, V. Balaji, J.A. Beesley, W.F. Cooke, K.W. Dixon, J. Dunne, K.A. Dunne, J.W. Durachta, K.L. Findell, P. Ginoux, A. Gnanadesikan, C.T. Gordon, S.M. Griffies, R. Gudgel, M.J. Harrison, I.M. Held, R.S. Hemler, L.W. Horowitz, S.A. Klein, T.R. Knutson, P.J. Kushner, A.R. Langenhorst, H.-C. Lee, S.-J. Lin, J. Lu, S.L. Malyshev, P.C.D. Milly, V. Ramaswamy, J. Russell, M.D. Schwarzkopf, E. Shevliakova, J.J. Sirutis, M.J. Spelman, W.F. Stern, M. Winton, A.T. Wittenberg, B. Wyman, F. Zeng and R. Zhang, 2006. GFDL's CM2 global coupled climate models - Part 1: Formulation and simulation characteristics. *J. Climate*, 19, 643-674.

Doblas-Reyes F J, A Weisheimer, T N Palmer, J M Murphy and D. Smith (2010) Forecast quality assessment of the ENSEMBLES seasonal-to-decadal Stream 2 hindcasts. ECMWF Tech Memo, 621, 45pp.

Doblas-Reyes, F. J., M. A. Balmaseda, A. Weisheimer, and T. N. Palmer, 2011: Decadal climate prediction with the European Centre for MediumRange Weather Forecasts coupled forecast system: Impact of ocean observations. *J. Geophys. Res.*, 116 (D19111), doi10.1029/2010JD015394.

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

Du, H., F.J. Doblas-Reyes, J. García-Serrano, V. Guemas, Y. Soufflet and B. Wouters, 2012. Sensitivity of decadal predictions to the initial atmospheric and oceanic perturbations. *Climate Dyn.*, doi:10.1007/s00382-011-1285-9.

Dufresne, J.-L., M.-A. Foujols, S. Denvil, A. Caubel, O. Marti, O. Aumont, Y. Balkanski, S. Bekki, H. Bellenger, R. Benshila, S. Bony, L. Bopp, P. Braconnot, P. Brockmann, P. Cadule, F. Cheruy, F. Codron, A. Cozic, D. Cugnet, N. de Noblet, J.-P. Duvel, C. Ethé, L. Fairhead, T. Fichefet, S. Flavoni, P. Friedlingstein, J.-Y. Grandpeix, L. Guez, E. Guilyardi, D. Hauglustaine, F. Hourdin, A. Idelkadi, J. Ghattas, S. Joussaume, M. Kageyama, G. Krinner, S. Labetoulle, A. Lahellec, M.-P. Lefebvre, F. Lefevre, C. Levy, Z.X. Li, J. Lloyd, F. Lott, G. Madec, M. Mancip, M. Marchand, S. Masson, Y. Meurdesoif, J. Mignot, I. Musat, S. Parouty, J. Polcher, C. Rio, M. Schulz, D. Swingedouw, S. Szopa, C. Talandier, P. Terray and N. Viovy, 2012. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Climate Dyn.*, submitted.

Fan Y, van den Dool H (2008) A global monthly land surface air temperature analysis for 1948-present. *J Geophys Res* 113:D01,103, DOI 10.1029/2007JD008470

Fenoglio-Marc L., A. Mariotti, G. Sannino, B. Meyssignac, A. Carillo, M.V. Struglia, M. Rixen (2013) Decadal variability of net water flux at the Mediterranean Sea Gibraltar Strait, *Global and Planetary Change*, 100, 1-10, doi:10.1016/j.gloplacha.2012.08.007.

Fyfe, J.C., W.J. Merryfield, V. Kharin, G.J. Boer, W.-S. Lee and K. von Salzen, 2011. Skillful predictions of decadal trends in global mean surface temperature. *Geophys. Res. Lett.*, 38, L22801, doi:10.1029/2011GL049508, 2011.

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

García-Serrano, J., F.J. Doblas-Reyes and C.A.S. Coelho (2012). Understanding Atlantic multi-decadal variability prediction skill. *Geophysical Research Letters*, 39, L18708, doi:10.1029/2012GL053283.

García-Serrano, J., F.J. Doblas-Reyes, R. J. Haarsma and I. Polo (2013) Decadal prediction of the dominant West African monsoon rainfall modes. *Journal of Geophysical Research - Atmospheres*, 118, 1-20, doi:10.1002/jgrd.50465.

García-Serrano, J., V Guemas and F.J. Doblas-Reyes (2014) Added-value from initialization in skilful predictions of Atlantic multi-decadal variability, submitted.

Giannini, A., R. Saravanan and P. Chang (2003): Oceanic forcing of Sahel rainfall on interannual to interdecadal timescales. *Science*, 302, 1027-1030.

Giannini, A., R. Saravanan and P. Chang (2005): Dynamics of the boreal summer African monsoon in the NSIPP1 atmospheric model. *Clim. Dyn.*, 25, 517-535.

Goddard, L., A. Kumar, A. Solomon, D. Smith, G. Boer, P. Gonzalez, V. Kharin, W. Merryfield, C. Deser, S. J. Mason, B. P. Kirtman, R. Msadek, R. Sutton, E. Hawkins, T. Fricker, G. Hegerl, C. A. T. Ferro, D. B. Stephenson, G. A. Meehl, T. Stockdale, R.

Burgman, A. M. Greene, Y. Kushnir, M. Newman, J. Carton, I. Fukumori, T. Delworth, 2013, A verification framework for interannual-to-decadal prediction experiments. *Clim. Dyn.*, 40, 245-272, doi: 10.1007/s00382-012-1481-2.

Gordon, C., C. Cooper, C. Senior, H. Banks, J. Gregory, T. Johns, J. Mithell, and R. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Center coupled model without flux adjustments. *Clim. Dyn.*, 16, 147–168.

Griffies S M and K Bryan, 1997, Predictability of North Atlantic multidecadal climate variability, *Sciences*, 275, 181-184.

Hasumi, H. and S. Emori, 2004: K-1 Coupled GCM (MIROC) Description. Technical report Center for Climate System Research, University of Tokyo, 34pp.

Hawkins E and R Sutton (2009) The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.* 90, 1095-1107. doi: 10.1175/2009BAMS2607.1

Hazeleger, W., et al., 2010: EC-Earth: A Seamless Earth System Prediction Approach in Action. *BAMS*, 91, 1357–1363.

Hazeleger, W., Guemas V., Wouters B., Corti S., Andreu-Burillo I., Doblus-Reyes F. J., Wyser K., Caian M., 2013, Multiyear climate predictions using two initialisation strategies. *Geophysical Research Letters*, doi:10.1002/grl.50355.

Kerr, R. A. (2000), A North Atlantic climate pacemaker for the centuries, *Science*, 288, 1984 – 1985.

Keenlyside N S, M Latif, J Jungclauss, L Kornbluh and E Roeckner (2008) Advanced decadal scale climate prediction in the North Atlantic sector. *Nature*, 453, 84–88, doi:10.1038/nature06921.

Knight J. R., R. J. Allan, C K Folland, M Vellinga and M. E. Mann, 2005, A signature of persistent natural thermohaline circulation cycles in observed climate, *Geophys. Res. Lett.*, 32, L20708, doi:10.1029/2005GL024233.

Knight J. R., Folland C. K., Scaife A.A., 2006: Climate impacts of the Atlantic Multidecadal Oscillation, *Geophys. Res. Lett.*, 33 (17), doi:10.1029/2006GL026242.

Kucharski, F., N. Zeng and E. Kalnay (2012): A further assessment of vegetation feedback on decadal Sahel rainfall variability. *Clim. Dyn.*, doi 10.1007/s00382-012-1397-x (in press).

Mariotti A and Dell'Aquila A, 2012: Decadal climate variability in the Mediterranean region: roles of large-scale forcings and regional processes. *Clim. Dynamics*, 38 (5-6), 1129-1145: doi: DOI 10.1007/s00382-011-1056-7.

Mariotti A, Zeng N, Lau K-M, 2002: *Euro-Mediterranean rainfall and ENSO - a seasonally varying relationship*. *Geophysical Research Letters*, 29 (12) 10.1029/2001GL014248.

Mariotti, A., N. Zeng, J. Yoon, V. Artale, A. Navarra, P. Alpert, and L. Z X Li, 2008: Mediterranean water cycle changes: transition to drier 21st century conditions in observations and CMIP3 simulations. *Environ. Res. Lett.*, 3, 044001 (8pp) doi:10.1088/1748-9326/3/4/044001.

Matei, D., H. Pohlmann, J. Jungclaus, W. Müller, H. Haak, and J. Marotzke, 2012 : Two tales of initializing decadal climate predictions experiments with the ECHAM5/MPI-OM model. *Journal of Climate* 25, 8502-8523.

Meehl G A et al. (2009) Decadal Prediction. *Bull. Amer. Meteor. Soc.*, 90 (10), 1467–1485.

Meehl G A et al (2014) Decadal Climate Prediction: An Update from the Trenches, *Bull. Amer. Meteor. Soc.*, doi: 10.1175/BAMS-D-12-00241.1.

Merryfield, W. J., W.-S. Lee, G. J. Boer, V. V. Kharin, J. F. Scinocca, G. M. Flato, R. S. Ajayamohan, J. C. Fyfe, Y. Tang, and S. Polavarapu, 2013. The Canadian Seasonal to Interannual Prediction System. Part I: Models and Initialization, *Monthly Weather Review*, doi:10.1175/MWR-D-12-00216.1.

Mohino, E., S. Janicot and J. Bader (2011): Sahelian rainfall and decadal to multidecadal SST variability. *Clim. Dyn.*, 37, 419-440

Murphy J, V Kattsov, N Keenlyside, M Kimoto, G Meehl, V Mehta, H Pohlmann, A Scaife and D Smith (2010) Towards Prediction of Decadal Climate Variability and Change *Procedia Environmental Sciences*, 1, 287–304.

Ottera OH, Bentsen M, Drange H, Suo L (2010) External forcing as a metronome for Atlantic multidecadal variability. *Nature Geosci* 3:688-694.

Pohlmann, H., M. Botzet, M. Latif, A. Roesch, M. Wild, P. Tschuck, 2004: Estimating the decadal predictability of a coupled AOGCM. *J. Climate*, 17, 4463-4472

Pohlmann, H., J. H. Jungclaus, A. Köhl, D. Stammer, J. Marotzke, 2009: Initializing decadal climate predictions with the GECCO oceanic synthesis: Effects on the North Atlantic. *J. Climate*, 22, 3926-3938. doi:10.1175/2009JCLI2535.1

Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton, 2000: The impact of new physical parametrizations in the Hadley Centre climate model - HadAM3. *Clim. Dyn.*, 16, 123–146, doi:10.1007/s003820050009.

Rudolf B, Becker A, Schneider U, Meyer-Christoffer A, Ziese M (2010) The new “GPCC Full Data Reanalysis Version 5” providing high-quality gridded monthly precipitation data for the global land-surface is public available since december 2010. Tech. rep., GPCC, URL gpcp.dwd.de

Sakamoto, T.T., Y. Komuro, M. Ishii, H. Tatebe, H. Shiogama, A. Hasegawa, T. Toyoda, M. Mori, T. Suzuki, Y. Imada-Kanamaru, T. Nozawa, K. Takata, T. Mochizuki, K. Ogochi, T. Nishimura, S. Emori, H. Hasumi and M. Kimoto, 2012. MIROC4h - a new high-resolution atmosphere-ocean coupled general circulation model. *J. Meteor. Soc. Japan*, in press.

Scoccimarro E., S. Gualdi, A. Bellucci, A. Sanna, P.G. Fogli, E. Manzini, M. Vichi, P. Oddo, and A. Navarra, 2011. Effects of tropical cyclones on ocean heat transport in a high resolution coupled general circulation model. *J. Climate*, 24, 4368-4384.

Smith D, A Cusack, A Colman, C Folland, G Harris, and J M Murphy (2007) Improved surface temperature prediction for the coming decade from a global climate model. *Science*, 317, 796–799.

Smith T, Reynolds R, Peterson T, Lawrimore J (2008) Improvements to NOAA’s historical merged land–ocean surface temperature analysis (1880–2006). *J Climate* 21:2283–2296, DOI : 10.1175/2007JCLI2100.1

Smith, D. M., A. A. Scaife and B. P. Kirtman (2012): What is the current state of scientific knowledge with regard to seasonal and decadal forecasting? *Environ. Res. Lett.*, 7, doi 10.1088/1748-9326/7/1/015602.

Swingedouw D., Mignot J., Labetoulle S., Guilyardi E. and Madec G., 2013, Initialisation and predictability of the AMOC over the last 50 years in a climate model., *Climate Dynamics* DOI: 10.1007/s00382-012-1516-8.

Tatebe, H., M. Ishii, T. Mochizuki, Y. Chikamoto, T. T. Sakamoto, Y. Komuro, M. Mori, S. Yasunaka, M. Watanabe, K. Ogochi, T. Suzuki, T. Nishimura, and M. Kimoto, 2012: Initialization of the climate model MIROC for decadal prediction with hydrographic data assimilation. *J. Meteor. Soc. Japan*, 90A, 275-294.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of cmip5 and the experimental design. *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-11-00094.1.

Terray L (2012) Evidence for multiple drivers of North Atlantic multi-decadal climate variability. *Geophys Res Lett* 39:L19712.

Ting, M., Y. Kushnir, R. Seager and C. Li (2009): Forced and internal twentieth-century SST trends in the North Atlantic. *J. Clim.*, 22, 1469-1481.

Trenberth K. E., P. D. Jones, P.G. Ambenje, R. Bojariu, D.R. Easterling, A.M.G. Klein Tank, D.E. Parker, J.A. Renwick, F. Rahimzadeh, M. M. Rusticucci, B.J. Soden and P.-M. Zhai, 2007, Observations: surface and atmospheric climate change. *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H. L. Miller, Eds Cambridge University Press, 235-336.

van Oldenborgh, G. J., L. A. te Raa, H. A. Dijkstra, S. Y. Philip (2009): Frequency- or amplitude-dependent effects of the Atlantic meridional overturning on the tropical Pacific Ocean. *Ocean. Sci.*, 5, 293-301.

van Oldenborgh, G. J., F. J. Doblas-Reyes, B. Wouters, and W. Hazeleger (2012): Decadal prediction skill in a multi-model ensemble. *Clim. Dyn.*, 38, 1263-1280.

Voltaire, A. E. Sanchez-Gomez, D. Salas y Méria, B. Decharme, C. Cassou, S. Sénési, S. Valcke, I. Beau, A. Alias, M. Chevallier, M. Déqué, J. Deshayes, H. Douville, E. Fernandez, G. Madec, E. Maisonnave, M.-P. Moine, S. Planton, D. Saint-Martin, S. Szopa, S. Tyteca, R. Alkama, S. Belamari, A. Braun, L. Coquart, F. Chauvin., 2012, The CNRM-CM5.1 global climate model : description and basic evaluation, *Clim. Dyn.*, DOI:10.1007/s00382-011-1259-y.

Watanabe, M., T. Suzuki, R. O'ishi, Y. Komuro, S. Watanabe, S. Emori, T. Takemura, M. Chikira, T. Ogura, M. Sekiguchi, K. Takata, D. Yamazaki, T. Yokohata, T. Nozawa, H. Hasumi, H. Tatebe and M. Kimoto, 2010. Improved climate simulation by MIROC5: mean states, variability, and climate sensitivity. *J. Climate*, 23, 6312-6335.

Yang, X., A. Rosati, S. Zhang, T. L. Delworth, R. Gudgel, R. Zhang, G. A. Vecchi, W. G. Anderson, Y.-S. Chang, T. DelSole, K. W. Dixon, R. Msadek, W. F. Stern, A. T. Wittenberg, and F. Zeng, 2013: A predictable AMO-like pattern in GFDL's fully-coupled ensemble initialization and decadal forecasting system. *Journal of Climate*, 26(2), doi:10.1175/JCLI-D-12-00231.1.

Yukimoto, S., et al., 2001: The New Meteorological Research Institute Coupled GCM(MRI-CGCM2) Model Climate and Variability. *Pap. Meteor. and Geophys.*, 51, 47–88.

Zhang R, Delworth TL, Sutton R, Hodson DLR, Dixon KW, Held IM, Kushnir Y, Marshall J, Ming Y, Msadek R, Robson J, Rosati AJ, Ting M, Vecchi GA (2013) Have aerosols caused the observed Atlantic multi-decadal variability?. *J Atmos Sci* 70:1135-1143.

List of Tables:

Model	Initialization Methodology	# of Init Members	# of NoInit Members
HadCM3 (Gordon et al. 2000; Pope et al. 2000)	Coupled anomaly assimilation : ERA-40 and ERA Interim atmospheric reanalyses, ocean observations	10	10
MRI-CGCM3 (Yukimoto et al. 2001; Tatebe et al, 2012)	Full-field assimilation	9	1
MIROC4 (Sakamoto et al. 2012)	Assimilation in the coupled model of ocean anomalies of gridded subsurface observations of T and S	3	3
MIROC5 (Hasumi and Emori 2004; Watanabe et al, 2010)		6	1
CANCM4 (Fyfe et al, 2011; Merryfield et al., 2013)	Coupled assimilation of the ERA40 and ERA Interim atmospheric reanalyses, observed SSTs and SODA and GODAS subsurface ocean T and S, beforehand adjusted to preserve T-S relationship	10	10
CNRM-CM5 (Voltaire et al, 2012)	Ocean T and S nudged (3-D Newtonian damping) toward NEMOVAR-COMBINE	10	6

	in the coupled model; NEMOVAR-COMBINE reanalysis produced using multivariate 3DVAR data assimilation in NEMO ocean model		
MPI-M (Jungclaus et al, 2006; Matei et al, 2012)	Nudging in the coupled model of T and S anomalies obtained from an ocean-only run forced with NCEP atmospheric reanalyses	10	3
GFDL-CM2 (Delworth et al. 2006; Yang et al, 2013)	Coupled assimilation of atmospheric reanalysis and ocean observations of 3-D T and S and SST	10	10
CMCC-CM (Scoccimarro et al. 2011; Bellucci et al., 2013)	Full field ocean initialization from 3 different realizations of CMCC-INGV ocean synthesis of T and S	3	1
IPSL-CM5 (Dufresne et al. 2012; Swingedouw et al., 2013)	Nudging in the coupled model of SST anomalies to Reynolds observations	6	3
EC-Earth v2 (Hazeleger et al. 2010; Du et al, 2012; Hazeleger et al.	Full-field initialization : ERA-40 and ERA Interim atmosphere/land reanalyses; NEMOVAR-S4 ocean reanalysis	10	11

2013)			
-------	--	--	--

Table 1: CMIP5 model experiments analyzed in this work. The number of ensemble members analyzed in the Init and NoInit MME are specified.

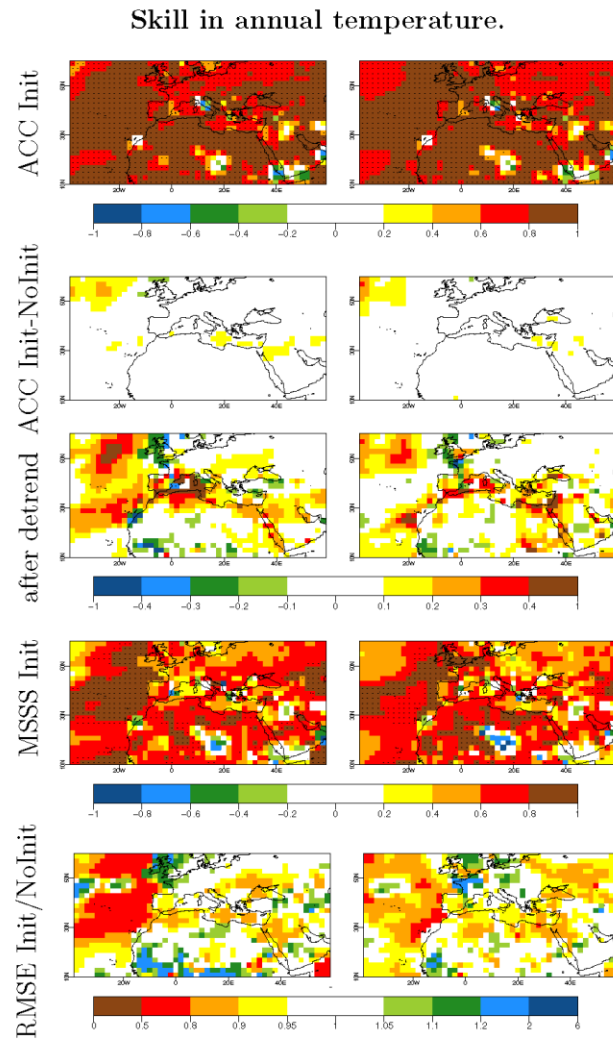


Figure 1: First row : Correlation between the ensemble-mean predicted and the observed temperature anomalies averaged along the forecast times ranging from the 2nd to the 5th years (left) and the 6th to the 9th years (right). Second row: Difference between the correlations in the decadal hindcasts and the equivalent in the historical simulations. Third row : Difference between the correlations in the decadal hindcasts and the equivalent in the historical simulations after detrending by subtracting the global mean SST to the grid-point SST and the global land temperature to the grid-point land temperature. Fourth row : Mean Square Skill Score of the ensemble-mean predicted anomalies. Fifth row : Ratio of the Root Mean Square Error (RMSE) of the decadal hindcasts with respect to the historical simulations. The black dots indicate the region where the correlation, the difference in correlation, the MSSS or the ratio of the RMSE is significant to the 95% level.

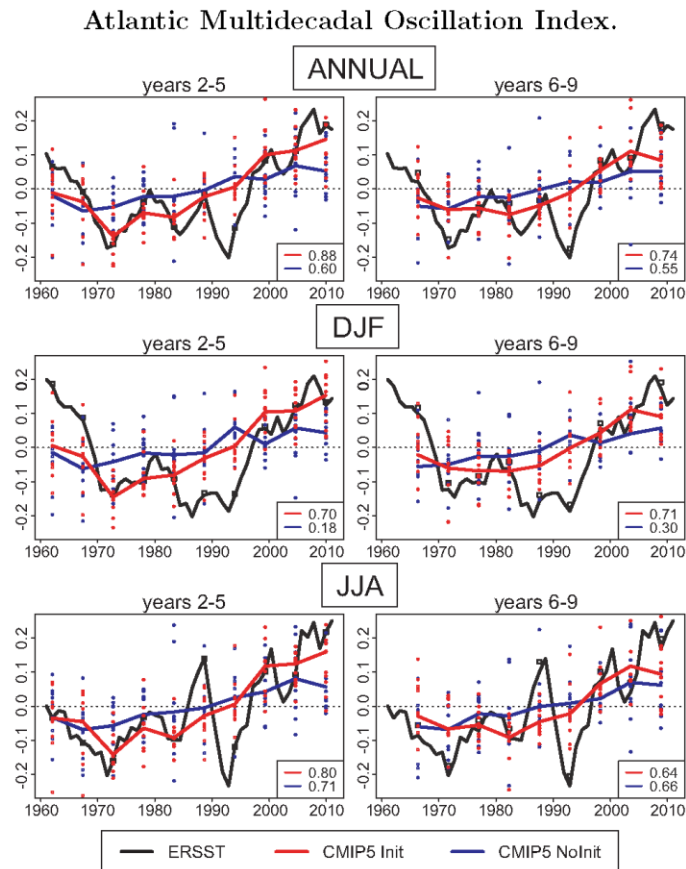


Figure 2: Atlantic Multidecadal Oscillation Index in the observations (black), in the decadal hindcasts (red), and in the historical simulations (blue) averaged across the forecast times ranging from the 2nd to the 5th years (left) and ranging from the 6th to the 9th years (right), annual mean in the top row, winter mean (DJF) in the middle, summer mean (JJA) in the bottom. The multi-model ensemble mean is indicated by a continuous line while the individual members by dots. The ACC skill is provided in the box in the right bottom corner of each panel.

Correlation annual temperature with AMO index

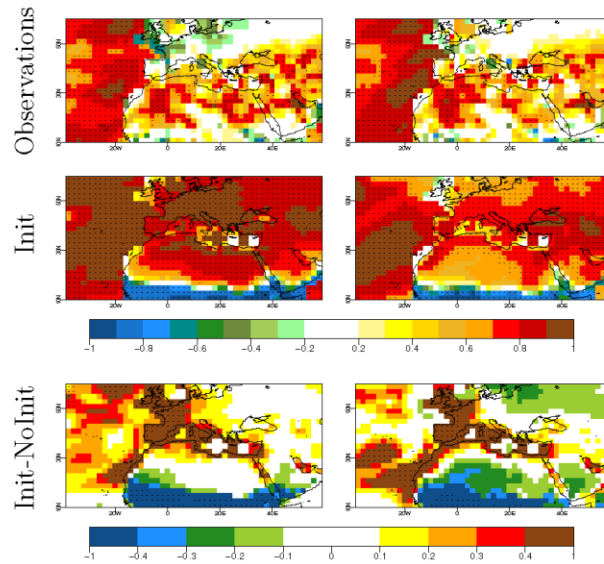


Figure 3: Correlation between the AMO index and the temperature anomalies from observational datasets (first row), and the ensemble mean decadal hindcasts (second row), averaged along the forecast times ranging from the 2nd to 5th years (left) and from the 6th to 9th years (right). Third row : difference between second row and its equivalent in NoInit. The temperature anomalies have been detrended by subtracting the global mean SST to the grid-point SST and the global land temperature to the land temperature prior to the computation of correlation. The black dots indicate the region where the correlation or difference in correlation is significant to the 95% level.

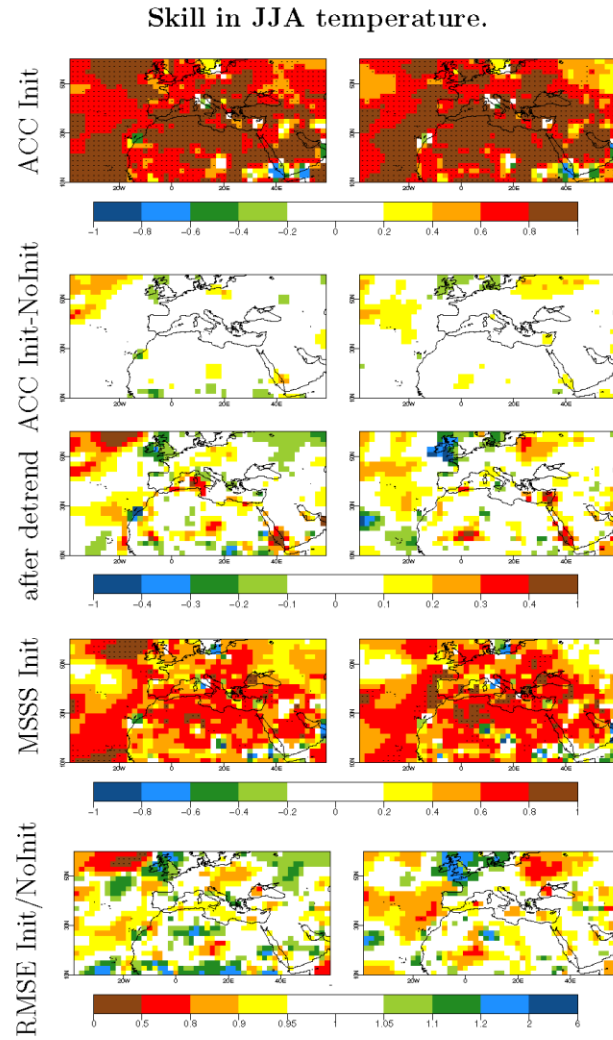


Figure 4: First row : Correlation between the ensemble-mean predicted and the observed summer (JJA) temperature anomalies averaged along the forecast times ranging from the 2nd to the 5th years (left) and the 6th to the 9th years (right). Second row: Difference between the correlations in the decadal hindcasts and the equivalent in the historical simulations. Third row : Difference between the correlations in the decadal hindcasts and the equivalent in the historical simulations after detrending by subtracting the global mean SST to the grid-point SST and the global land temperature to the land temperature. Fourth row : Mean Square Skill Score of the ensemble-mean predicted anomalies. Fifth row : Ratio of the Root Mean Square Error (RMSE) of the decadal hindcasts with respect to the historical simulations. The black dots indicate the region where the correlation, the difference in correlation, the MSSS or the ratio of the RMSE is significant to the 95% level.

Correlation JJA temperature with AMO index

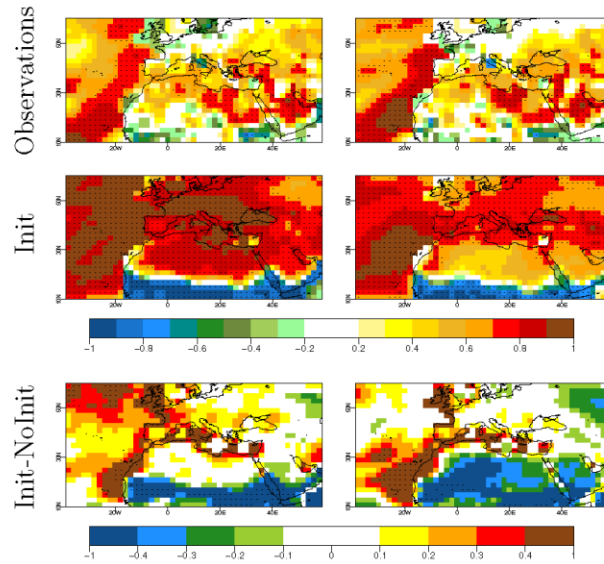


Figure 5: Correlation between the AMO index and the summer (JJA) temperature anomalies from observational datasets (first row), and the ensemble-mean decadal hindcasts (second row), averaged along the forecast times ranging from the 2nd to 5th years (left) and from the 6th to 9th years (right). Third row : difference between second row and its equivalent in NoInit. The temperature anomalies have been detrended by subtracting the global mean SST to the grid-point SST and the global land temperature to the land temperature prior to the computation of correlation. The black dots indicate the region where the correlation or difference in correlation is significant to the 95% level.

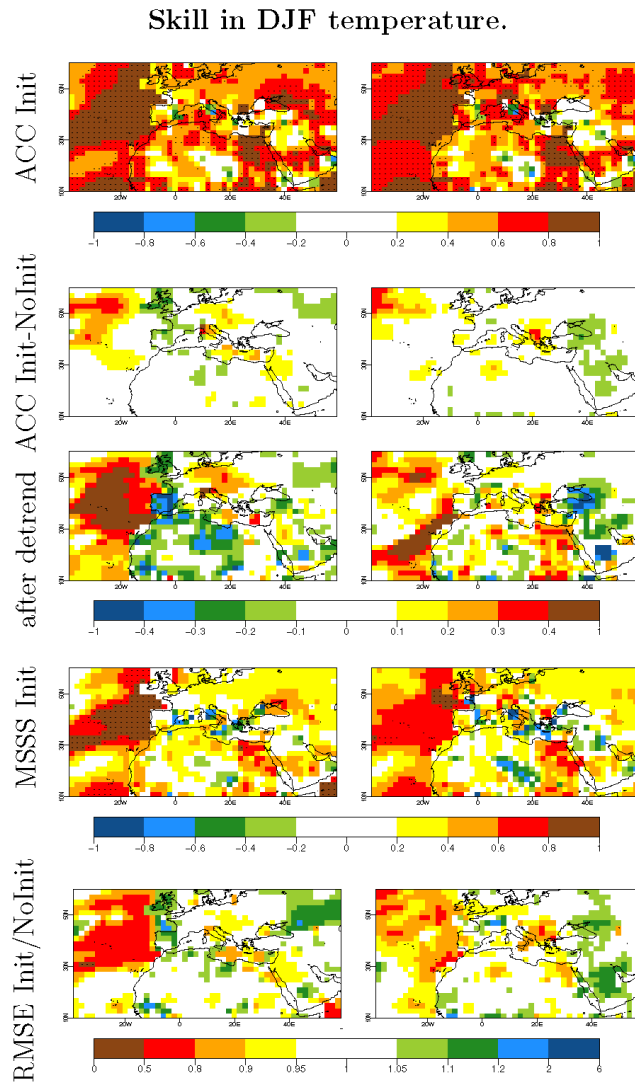


Figure 6: First row : Correlation between the ensemble-mean predicted and the observed winter (DJF) temperature anomalies averaged along the forecast times ranging from the 2nd to the 5th years (left) and the 6th to 9th years (right). Second row: Difference between the correlations in the decadal hindcasts and the equivalent in the historical simulations. Third row : Difference between the correlations in the decadal hindcasts and the equivalent in the historical simulations after detrending by subtracting the global mean SST to the grid-point SST and the global land temperature to the land temperature. Fourth row : Mean Square Skill Score of the ensemble-mean predicted anomalies. Fifth row : Ratio of the Root Mean Square Error (RMSE) of the decadal hindcasts with respect to the historical simulations. The black dots indicate the region where the correlation, the difference in correlation, the MSSS or the ratio of the RMSE is significant to the 95% level.

Correlation DJF temperature with AMO index

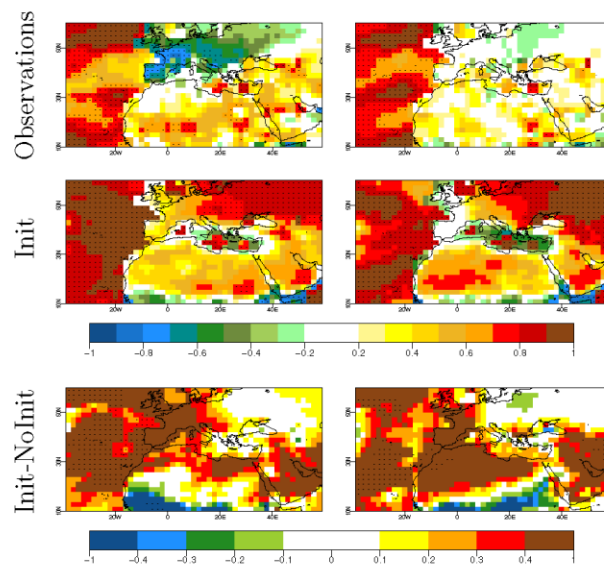


Figure 7: Correlation between the AMO index and the winter (DJF) temperature anomalies from the observational datasets (first row), and the ensemble-mean decadal hindcasts (second row), averaged along the forecast times ranging from the 2nd to 5th years (left) and from the 6th to 9th years (right). Third row : difference between second row and its equivalent in NoInit. The temperature anomalies have been detrended by subtracting the global mean SST to the grid-point SST and the global land temperature to the land temperature prior to the computation of correlation. The black dots indicate the region where the correlation or difference in correlation is significant to the 95% level.

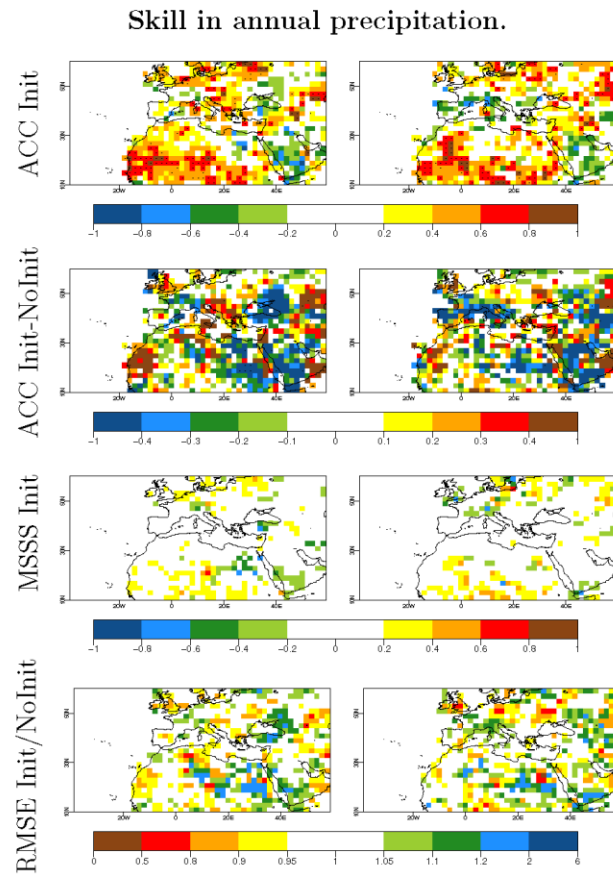


Figure 8: First row : Correlation between the ensemble-mean predicted and the observed precipitation anomalies averaged along the forecast times ranging from the 2nd to the 5th years (left) and the 6th to the 9th years (right). Second row: Difference between the correlations in the decadal hindcasts and the equivalent in the historical simulations. Third row : Mean Square Skill Score of the ensemble-mean predicted anomalies. Fourth row : Ratio of the Root Mean Square Error (RMSE) of the decadal hindcasts with respect to the historical simulations. The black dots indicate the region where the correlation, the difference in correlation, the MSSS or the ratio of the RMSE is significant to the 95% level.

Correlation annual precipitation with AMO index

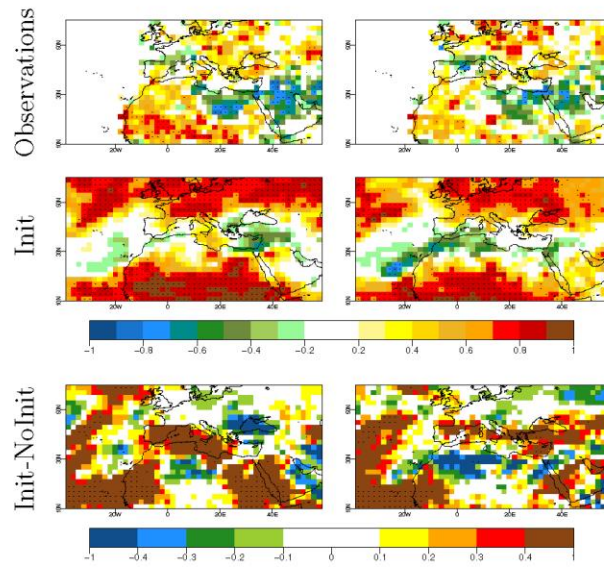


Figure 9: Correlation between the AMO index and the precipitation anomalies from the observational datasets (first row) and the ensemble-mean decadal hindcasts (second row), averaged along the forecast times ranging from the 2nd to the 5th years (left) and from the 6th to the 9th years (right). Third row : difference between second row and its equivalent in NoInit. The black dots indicate the region where the correlation or difference in correlation is significant to the 95% level.

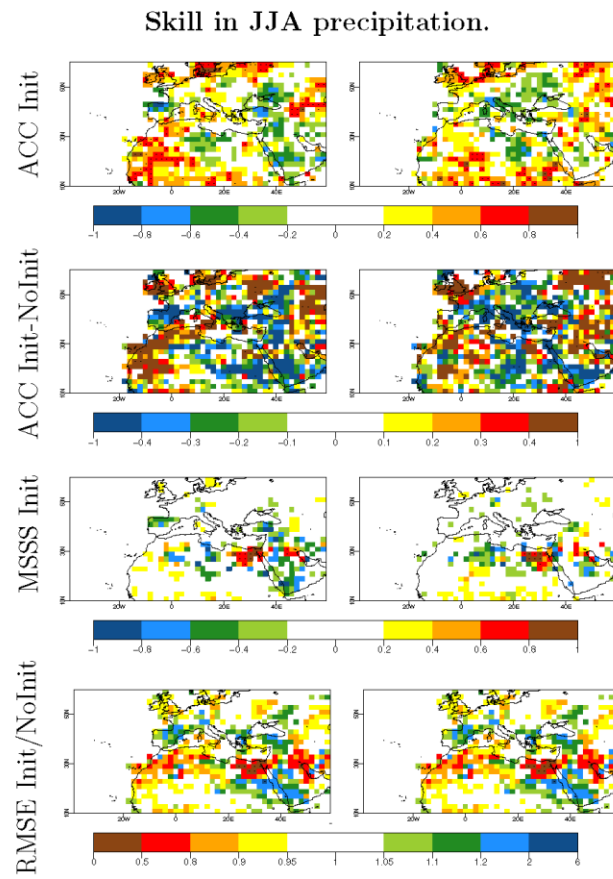


Figure 10: First row : Correlation between the ensemble-mean summer (JJA) predicted and the observed precipitation anomalies averaged along the forecast times ranging from the 2nd to the 5th years (left) and the 6th to the 9th years (right). Second row: Difference between the correlations in the decadal hindcasts and the equivalent in the historical simulations. Third row : Mean Square Skill Score of the ensemble-mean predicted anomalies. Fourth row : Ratio of the Root Mean Square Error (RMSE) of the decadal hindcasts with respect to the historical simulations. The black dots indicate the region where the correlation, the difference in correlation, the MSSS or the ratio of the RMSE is significant to the 95% level.

Correlation JJA precipitation with AMO index

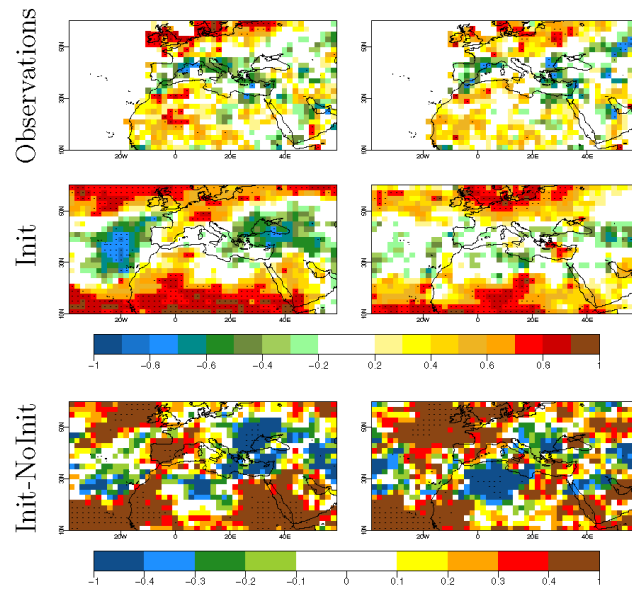


Figure 11: Correlation between the AMO index and the summer (JJA) precipitation anomalies from the observational datasets (first row), and the ensemble-mean decadal hindcasts (second row), averaged along the forecast times ranging from the 2nd to the 5th years (left) and from the 6th to the 9th years (right). Third row : difference between second row and its equivalent in NoInit. The black dots indicate the region where the correlation or difference in correlation is significant to the 95% level.