Clim Dyn (2013) 40:2359–2380
DOI 10.1007/s00382-012-1466-1

Decadal predictability of the Atlantic meridional overturning circulation and climate in the IPSL-CM5A-LR model

A. Persechino · J. Mignot · D. Swingedouw · S. Labetoulle · E. Guilyardi

Received: 2 November 2011 / Accepted: 18 July 2012 / Published online: 10 August 2012

Abstract This study explores the decadal potential predictability of the Atlantic Meridional Overturning Circulation (AMOC) as represented in the IPSL-CM5A-LR model, along with the predictability of associated oceanic and atmospheric fields. Using a 1000-year control run, we analyze the prognostic potential predictability (PPP) of the AMOC through ensembles of simulations with perturbed initial conditions. Based on a measure of the ensemble spread, the modelled AMOC has an average predictive skill of 8 years, with some degree of dependence on the AMOC initial state. Diagnostic potential predictability of surface temperature and precipitation is also identified in the control run and compared to the PPP. Both approaches clearly bring out the same regions exhibiting the highest predictive skill. Generally, surface temperature has the highest skill up to 2 decades in the far North Atlantic ocean. There are also weak signals over a few oceanic areas in the tropics and subtropics. Predictability over land is restricted to the coastal areas bordering oceanic predictable regions. Potential predictability at interannual and longer timescales is largely absent for precipitation in spite of weak signals identified mainly in the Nordic Seas. Regions of weak signals show some dependence on AMOC initial state. All the identified regions are closely linked to decadal AMOC fluctuations suggesting that the potential predictability of climate arises from the mechanisms controlling these fluctuations. Evidence for dependence on AMOC initial state also suggests that studying skills from case studies may prove more useful to understand predictability mechanisms than computing average skill from numerous start dates.

Keywords Decadal climate predictability · Atlantic meridional overturning circulation · Diagnostic and prognostic potential predictability · Ocean and climate dynamics

This paper is a contribution to the special issue on the IPSL and CNRM global climate and Earth System Models, both developed in France and contributing to the 5th coupled model intercomparison project.

A. Persechino (✉)
School of Ocean and Earth Science, University of Southampton, European Way, Southampton, Hampshire SO14 3ZH, UK
e-mail: a.persechino@noc.soton.ac.uk

J. Mignot · S. Labetoulle · E. Guilyardi
Institut Pierre-Simon Laplace/LOCEAN, Jussieu, Paris, France

D. Swingedouw
Institut Pierre-Simon Laplace/LSCE, Gif-sur-Yvette, CEA Saclay, Orme des Merisiers, France

E. Guilyardi
NCAS-Climate, University of Reading, Reading, UK

1 Introduction

In contrast to both weather and seasonal forecasts, decadal prediction is still a developing science. The ocean is assumed to be among the most predictable component of the climate system on the decadal timescale as it provides long-term climatic memory due to its large thermal inertia. However, there is still no clear understanding of the related predictability limits (Meehl et al. 2009; Hurrell et al. 2009; Solomon et al. 2011). Using control data from eleven state-of-the-art coupled climate models, Boer (2001) showed that the North Atlantic ocean exhibits the highest potential predictability at decadal timescales independently of the model used. Decadal variations of sea surface temperature (SST) in this region (often referred to as the Atlantic Multidecadal Oscillation, AMO) are thought to influence important climatic features, including rainfall over the

African Sahel, India and Brazil, Atlantic hurricanes and summer climate over Europe and America (e.g. Pohlmann et al. 2004; Sutton and Hodson 2005; Zhang and Delworth 2006; Knight et al. 2006; Dunstone et al. 2011). As there is evidence from climate models that the AMO is linked to the Atlantic Meridional Overturning Circulation (AMOC) (Knight et al. 2005), the AMOC has therefore been considered as a key target for the study of decadal potential predictability (e.g. Delworth and Mann 2000; Curry et al. 2003; Latif et al. 2004; Collins et al. 2006).

Two methods are commonly used to estimate potential predictability. In the diagnostic approach, the predictability (DPP for diagnostic potential predictability) is analysed by decomposing the variance of a climate variable into a long timescale component considered as potentially predictable, and an unpredictable noise component. Previous studies using this approach for both real and modelled systems include those of Rowell (1998), Boer (2001, 2004, 2011), Boer and Lambert (2008), and Hawkins et al. (2011). Following Boer et al. (2001) and using nine models participating in the first Coupled Model Intercomparison Project (CMIP1), Boer (2004) found potential predictability of surface air temperature predominantly over the high latitude oceans on multidecadal timescales, especially in the North Atlantic. On shorter timescales, he also found some hints of potential predictability in the tropical Atlantic, while predictability was overall limited over land. Boer (2011) extended this analysis to simulations taking into account different climate change scenarios and attempted to distinguish between internal and externally forced potential predictability. Using two different Atmosphere–Ocean General Circulation Models (HadCM3 and HadGEM1), Hawkins et al. (2011) identified the far North Atlantic in general, and the North Atlantic Current (NAC) region in particular, as regions with high potential predictability. They also found that the potential predictability estimates from the DPP approach for both AOGCMs have magnitudes comparable to the HadISST observations.

In the second approach, the predictability (PPP for prognostic potential predictability) is estimated prognostically, by re-running a climate simulation with slightly perturbed initial conditions (ICs). This approach does not compare to observations directly, and only assesses the ability of the modelled climate to reproduce itself given an amount of uncertainty on ICs, representing for example the atmospheric noise. These experiments are thus often called "perfect ensemble" experiments. Predictability studies using such an approach began with Griffies and Bryan (1997), who suggested that SST variation patterns in the North Atlantic are potentially predictable on multidecadal timescales. Similar studies further demonstrated potential predictability in AMOC variations and related oceanic and atmospheric fields at the decadal timescale (e.g. Collins 2002; Collins and Sinha 2003; Pohlmann et al. 2004; Collins et al. 2006; Hurrell et al. 2009; Msadek et al. 2010). For instance, the results from Msadek et al. (2010) indicate that the AMOC is potentially predictable up to 20 years, and, depending on the initial conditions, there are even some hints of potential predictability for more than 50 years into the future according to Collins and Sinha (2003). Both studies compute their predictability using slightly different definitions of root mean square error and consider that predictability is lost when the error reaches a certain threshold. Using the same model as Collins and Sinha (2003), Hermanson and Sutton (2009) follow a different approach and find lower predictability for the AMOC (with an average of only 5 years). It is therefore important to carefully define predictability and to use several metrics to better understand the limit and extent of predictable fields.

Although there have been an increasing number of studies on decadal predictability of the climate system in the last few years due to the impetus of the "near term" CMIP5 protocol (Taylor et al. 2009), the assessment of decadal climate predictability studies remains unclear as the level of predictability differs from one study to another (e.g. Meehl et al. 2009). This may be subject to model differences and uncertainties, as well as differences in the experimental protocol and metrics used. Here, we explore the decadal predictability of the AMOC and associated oceanic and atmospheric fields as they are represented in the IPSL-CM5A-LR model (Dufresne et al. 2012) under pre-industrial control conditions, using both DPP and PPP approaches. The aim of this study is to address the following questions: (1) Where do climate-related fields exhibit the strongest sensitivity to decadal AMOC fluctuations in the model? (2) Are specific changes in the AMOC potentially predictable and which observations of the ocean state are likely to be of greatest value to constrain predictions? (3) What is the predictability of the Atlantic climate in the model and how is it related to low-frequency AMOC variability?

After the description of the methods and the model in the next section, the control simulation is analysed in Sect. 3 to illustrate the impact of decadal AMOC fluctuations on the Atlantic Climate. In Sect. 4, the potential predictability of the AMOC is investigated using "perfect ensemble" experiments. Section 5 addresses the potential predictability of climate and its link with decadal AMOC variability. A summary and a discussion end the paper in Sect. 6.

2 Methods and experimental setup

2.1 Methods for measuring predictive skills

2.1.1 Diagnostic potential predictability (DPP)

The DPP approach uses the method of analysis of variance (Madden 1976; Rowell 1998) to examine the

low-frequency variability (considered to be at least potentially predictable) of a given variable. As an estimate of DPP, we used the non-biased estimation of potential predictability variance fraction (*ppvf*) from Boer (2004) that attempts to separate the long-term variability from internal variability. This metric is described in detail in Appendix 1. This approach is an easy and cheap statistical way to estimate the spatio-temporal predictability of climate-related fields. Its power lies in the fact that it only relies on a long control simulation.

2.1.2 Prognostic potential predictability (PPP)

The PPP approach consists in performing "perfect ensemble" experiments with a single coupled model by perturbing the ICs. The initial perturbation is supposed to represent atmospheric chaotic noise or uncertainty in the estimation of the climate state (e.g. Collins and Sinha 2003; Pohlmann et al. 2004; Msadek et al. 2010). This approach represents an estimate of the upper limit of predictability based on having a perfect model and near perfect knowledge of the current state of the climate system (principally the state of the ocean). Although this situation is never likely to be achieved in practice, this approach is useful in identifying explicitly the climate predictability over a specific climate trajectory.

Practically, both the spread and the correlation of the members of each ensemble are useful and important tools to quantify the reproducibility and thus predictability of the simulated fields. In this study, we thus consider two deterministic measures (following the Assessment of Intraseasonal to Interannual Climate Prediction and Predictability report, National Research Council 2010): the Ensemble Spread (ES) and Ensemble Correlation (EC). These two metrics are described in detail in Appendix 1. Note that EC alone does not indicate whether the forecast values are of the right magnitude (contrarily to ES). In the same way ES alone does not indicate the direction of the deviations (contrarily to EC). Thus, in this study, we explore the information given by both metrics, and consider that a variable is potentially predictable if it has a (low) statistically significant ES (below the saturation level which is defined as the control Root Mean Square Error RMSE, see Appendix 1 for further details) associated with a (high) statistically significant EC. By combining these two metrics, we are in good agreement with Hawkins et al. (2011) who claim that prediction skills should be measured using more than one metric. However, it has to be kept in mind that, as will be illustrated below, combining these two metrics might be too restrictive in some situations, and that information given by ES or EC alone should not be neglected.

Both these metrics have to be computed with respect to a target, a state that we wish to predict. At least two definitions of this target have been proposed in the literature: (1) the ensemble mean (e.g. as in Msadek et al. 2010, hereinafter M10) or (2) each individual member successively (e.g. as in Collins and Sinha 2003, hereinafter CS03). As demonstrated in Appendix 1, ES computed with each definition only differs by a factor of proportionality. Both definitions are thus equivalent for this metric. In contrast, no such relationship of proportionality could be found for EC. Here, we will therefore consider these two definitions to evaluate possible differences in their respective scores of PPP.

2.2 Model configuration and simulations

2.2.1 Brief model description

We use the coupled model IPSL-CM5A-LR (http://icmc.ipsl.fr/model-and-data/ipsl-climate-models/ipsl-cm5, Dufresne et al. 2012) in which the atmospheric general circulation model LMDZ5A (Hourdin et al. 2012) incorporating the ORCHIDEE land-surface model (Krinner et al. 2005) is coupled with the oceanic module NEMOv3.2 (Madec 2008), that includes the sea ice model LIM-2 (Fichefet and Maqueda 1997), and the oceanic bio-geochemistry model PISCES (Aumont and Bopp 2006). The coupling between oceanic and atmospheric models is achieved using OASIS3 (Valcke 2006). The atmosphere has a regular horizontal grid with 96 × 96 points corresponding to a resolution of 1.9° × 3.75°, and 39 vertical levels. The ocean model runs with an irregular grid of 182 × 149 points (ORCA grid) corresponding to a nominal resolution of 2°, enhanced over the Arctic and subpolar North Atlantic as well as around the Equator. There are 31 vertical levels for the ocean with the highest resolution in the upper 150 m.

2.2.2 Control integration

Our study is based on a 1000-year control simulation. The initial state was taken at the end of more than 400 years run in coupled mode, itself started from several hundreds of years of simulations of land and ocean carbon component separately to equilibrate the carbon pools (see Dufresne et al. 2012 for details). The simulation uses constant pre-industrial boundary conditions of tropospheric greenhouse gases and aerosol concentrations, and constitutes the pre-industrial control simulation of the IPSL-CM5A-LR model used for the CMIP5 exercises. The DPP approach only relies on this control integration.

2.2.3 "Perfect ensemble" experiments

The core of the PPP approach is a series of five ensemble experiments using the same code as the control integration

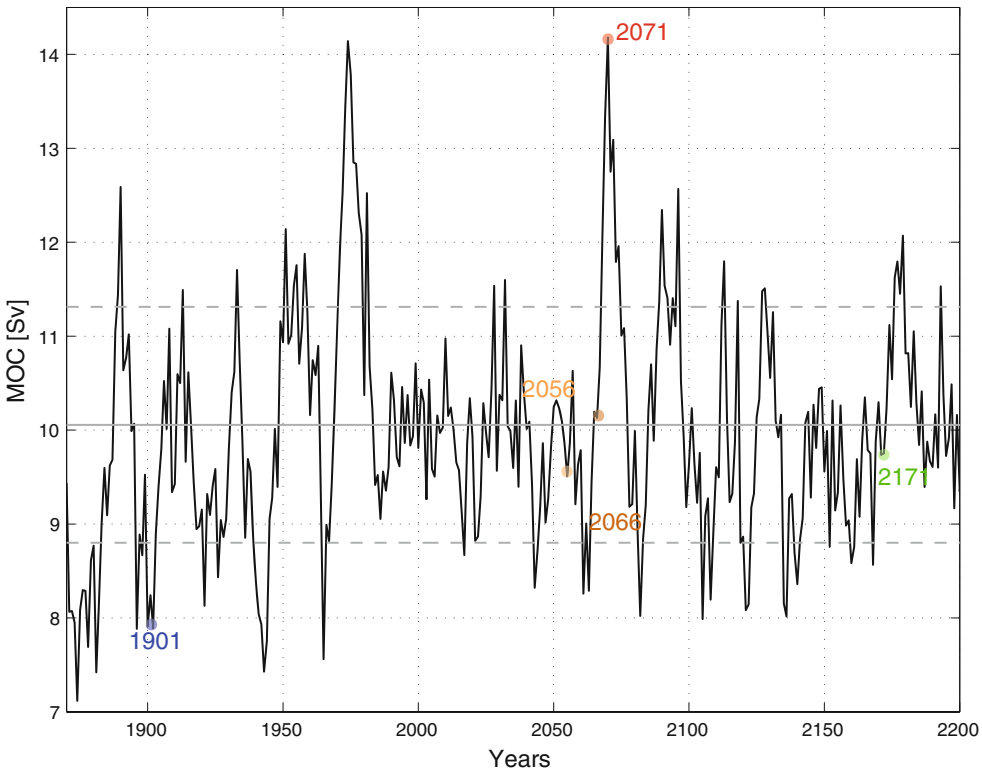

Fig. 1 Time series of AMOC index from year 1870 to 2200 with starting points of "perfect ensemble" experiments shown as *coloured points*. The 1000-year mean is shown as the *grey thick line* and the corresponding standard deviations are shown as the *grey dashed lines*

described above. Each ensemble starts from a different date of the control simulation and it includes 10 members, started from slightly different ICs and integrated for 20 years. ICs of the different members are obtained here by perturbing the SST from the control simulation with an anomaly chosen randomly for each grid point in the interval (−0.05 °C, 0.05 °C) with an equiprobable distribution for each value over this interval. This perturbation mimics a non-Gaussian white noise perturbation. No perturbation has been applied for the grid points under sea-ice cover. Figure 1 shows the five different starting dates of each ensemble experiment together with the time series of the AMOC index from years 1870 to 2200 in the control integration. One experiment starts from a year corresponding to relatively weak AMOC conditions (hereafter W, year 1901), one from intermediate conditions (hereafter I, year 2171), and one from strong conditions (hereafter S, year 2071). We have also chosen to start some experiments respectively 5 and 15 years before the large AMOC maximum in 2071 to investigate how far ahead this extreme value can be captured (15P and 5P, starting dates 2056 and 2066, respectively). Note that other choices could have been made and because of the limited number of starting dates, this experimental set up was not designed to draw robust conclusions about a possible predictability-dependence on the AMOC initial state. It could nevertheless give useful indications about it.

3 Fingerprints of AMOC variability

In the IPSL-CM5A-LR model, AMOC variability has been associated with a 20-year cycle described as a coupled mode driven by the subpolar region, and involving deep convection in the Nordic Seas, at the southern tip of Greenland, and south of Iceland (Escudier et al. 2012). Prior to the study of potential predictability in the AMOC, the regional impacts of AMOC variability are investigated in the control integration. To do so, we use regressions of 5-year moving averaged surface temperature and precipitation onto the 5-year moving averaged AMOC index when this latter leads by 10 years (lags at which regression coefficients are the strongest, Fig. 2). Despite some significant signals in the tropical Pacific, the main significant impacts are restricted to the North Atlantic surrounding regions. We therefore concentrate on this basin in the following.

3.1 Impacts on surface temperature

Figure 2a shows that the AMOC impact on temperature at the decadal timescale is dominant over the ocean, and in particular north of the NAC. Anomalously strong AMOC conditions are associated with significantly warm SST anomalies in the subpolar gyre, in both the eastern and southern branch of the subtropical gyre, and cold SST anomalies along the eastern coast of Greenland, south of

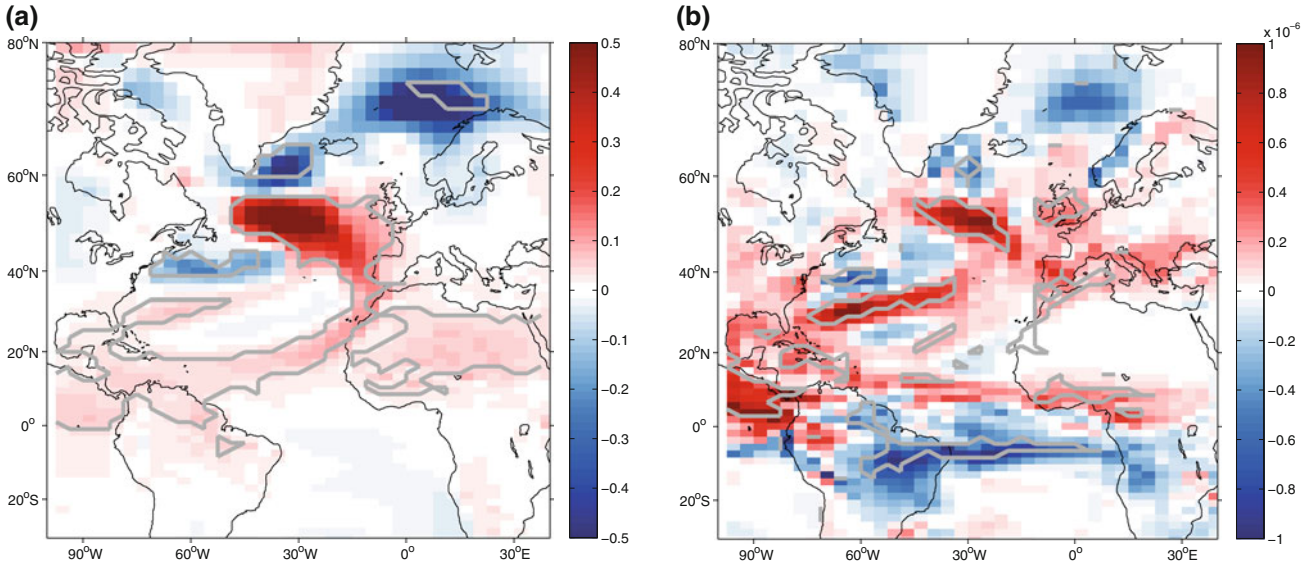

Fig. 2 Lagged regression of the 5-year moving average **a** surface air (sea) temperature at ground (sea) level (°C Sv^{-1}), and **b** precipitation (mm day^{-1} Sv^{-1}) onto the 5-year moving average AMOC index at the lag where regression coefficients are the strongest (i.e. when the AMOC leads by 10 years). Statistical significance of regression values has been tested using Student's *t* test, and Quenouille's (1952) method was used to calculate the effective degrees of freedom. The *grey contour* indicates 90 % confidence level for zero correlation

the Denmark Strait and in the Norwegian Sea, with a typical amplitude of about 0.5 °C Sv^{-1}. SST anomalies from the tropics to the subpolar regions in the Northern Hemisphere remain predominantly positive in constrast to the Southern Hemisphere where there are some hints of negative anomalies. Under anomalously strong AMOC conditions, an interhemispheric SST dipole pattern (although weak) therefore seems to emerge, as also identified in HadCM3 (Vellinga and Wu 2004). This pattern is also consistent with the AMO pattern in IPSL-CM5A-LR (Gastineau et al. 2012) as well as in an earlier version of the IPSL model (IPSL-CM4, Msadek and Frankignoul 2009). This suggests that as in other models (Kushnir 1994; Kerr 2000; Delworth and Mann 2000) an AMO-type response is associated with decadal AMOC fluctuations in IPSL-CM5A-LR. The SST pattern, identified in the latter, also resembles the observed AMO (e.g. Sutton and Hodson 2005), except for the localised significant negative anomalies in the high-latitudes of the North Atlantic. This result is consistent with previous modelling studies that found decadal AMOC fluctuations to be associated with an SST pattern resembling the observed AMO (Kushnir 1994; Kerr 2000; Delworth and Mann 2000).

As indicated above, decadal AMOC fluctuations have a much weaker impact over land. Anomalously strong AMOC conditions tend to be followed by significantly warmer conditions in Central America, subtropical Africa, and a few marine-influenced regions of Western Europe (with amplitude of anomalies up to +0.1 °C Sv^{-1}). Such links over land are consistent with previous studies (e.g. Pohlmann et al. 2004).

3.2 Impacts on precipitation

In terms of precipitation, the tropical Atlantic ocean clearly shows strong sensitivity to decadal AMOC fluctuations (Fig. 2b): stronger AMOC conditions are associated with significantly drier (wetter) southern (northern) tropics. This suggests a northward shift of the inter-tropical convergence zone (ITCZ) over the tropical Atlantic, as also identified in other climate models (e.g. Vellinga and Wu 2004; Swingedouw et al. 2009; Persechino et al. 2012). The ITCZ shift is also seen to extend to the Pacific Ocean, consistent with Xie et al. (2008) and Swingedouw et al. (2009). The strong sensitivity of tropical precipitation to AMOC fluctuations probably happens through the influence of SST anomalies identified earlier (Fig. 2a), consistent with the well-established strong coupling between the ocean and the atmosphere in this region (e.g. Chiang et al. 2008 and references therein). Significant precipitation anomalies are also found from the subtropics to the high-latitudes, largely resembling the corresponding SST anomalies.

The oceanic precipitation signal is again seen to leak over the adjacent continental areas, as for temperature. At midlatitudes, strong AMOC conditions are in particular associated with significantly wetter conditions over the British Isles (Fig. 2b). The signal identified over the tropical Atlantic also extends over the adjacent continents with significantly drier (wetter) conditions in the southern (northern) tropical regions of both America and Africa when the AMOC is increasing. This is consistent with several studies that already investigated the link between

decadal modulation of Sahelian rainfall, ITCZ shift and the AMO (Folland et al. 1986; Rowell et al. 1995; Zhang and Delworth 2006; Knight et al. 2006; Ting et al. 2009).

In view of these major climatic impacts of the AMOC, an important question remains whether AMOC fluctuations are potentially predictable. The ability to predict such fluctuations is now investigated using the PPP approach described in Sect. 2.2.3.

4 Potential predictability of AMOC fluctuations

Figure 3 shows the AMOC trajectories of each individual member, for each start date, together with the ensemble mean. At first sight, all ensemble means follow the initial control run relatively well, although with less variability due to the averaging effect. In particular, the extreme AMOC event at year 2071 is relatively well-captured (although underestimated in terms of amplitude) by both experiments starting 15 and 5 years before this peak.

4.1 Are changes in the AMOC potentially predictable?

4.1.1 Comparing the level of predictive skills to different definitions of metrics

Figure 4 shows ES of the AMOC index as a function of lead-time up to 2 decades for each experiment and both M10 (in grey) and CS03 (in black) definitions. This figure confirms the relation of proportionality mentioned earlier (in Sect. 2.1.2) between both definitions, with a factor of $\sqrt{[2\ M/(M-1)]}$ (M being the number of members, see Appendix 1 for further details). The last statistically significant lead-time before ES persistently exceeds the threshold is independent of the definition used, and represents the maximum lead-time of predictability as inferred from ES alone. Figure 5 shows EC computed between lead-time 1 year and varying lead-times, ranging 5–20 years (5 years corresponding to the minimum lead-time of predictability found from ES, Fig. 4). EC has generally higher scores for M10 than for CS03. Indeed, the ensemble mean (used in M10) is smoother and holds some information from each member, allowing higher correlations than one to one correlations among members (as used in CS03). However, in most cases, when EC is statistically significant (or not), it is generally also the case for the other definition. Note two exceptions (experiments W and I). However, from a predictability point of view, the statistical significance of EC at the lead-time at which ES saturates (information given in Fig. 4) is the same whichever the definition used. The main disadvantage of using EC_{CS03} is that too much weight could be given to an individual member that heavily diverges from the others, while EC_{M10} tends to average out extremes by the use of the ensemble mean. On the other hand, the latter can be seen as too lax as it involves a smoother baseline.

We showed that, overall, both definitions of EC and ES deliver similar messages, although EC_{CS03} seems to be slightly more severe than EC_{M10}. We therefore prefer to opt for the most cautious/severe definition, and will use the CS03 definition hereafter.

4.1.2 How far ahead is the AMOC potentially predictable?

Figure 6 shows a summary of results combining both ES_{CS03} (Fig. 4) and EC_{CS03} (Fig. 5) for the AMOC index. The predictive skill of each experiment is determined by the maximum lead-time at which ES saturates and its corresponding EC. Experiment S shows overall the highest PPP skill as its ES saturates at the longest lead-time and is associated with a high statistically significant EC (lower-right plot, Fig. 6); this experiment suggests a limit of predictive skill for the AMOC index of about 13 years. This result is consistent with a simple stochastic assumption for example (e.g. Frankignoul and Hasselmann 1977; Frankignoul 1985), which would predict that when starting from an extreme AMOC value, we expect most of the members to take the same direction towards a neutral state, thereby yielding high predictability. Nevertheless, the predictability timescale found here is longer than the persistence time estimated from the AMOC index autocorrelation function in a red noise framework (e.g. Frankignoul et al. 2002) which amounts to 4–5 years (not shown). This indicates that a simple autoregressive model provides low predictability for AMOC index behavior. Once back towards neutral (close to the mean) conditions (after about 13 years), experiments S indeed loses its predictive skills with a continually growing (decreasing) ES (EC) with lead-time. Similarly to S, experiment W is expected to have a similar predictive skill since it starts from an extreme state (more than one standard deviation σ away from the mean, Fig. 3). Although the EC associated with the maximum lead-time is statistically significant and high (0.74), ES however saturates twice as rapidly as in experiment S (at about 7 years, Fig. 6). This lower PPP skill could be explained by its starting date not being in such an extreme state as S; indeed, the starting value is superior (inferior) to 2σ in S(W). Alternatively, it might come from the dynamics itself, suggesting that the AMOC has more PPP skills when it starts from an anomalously strong overturning than from a weak one or a value close to its mean. The fact that the initial state corresponding to an anomalously strong AMOC is more predictable than those corresponding to a weak AMOC is in good agreement with several previous studies (e.g. CS03; Collins et al. 2006). Consistent with the idea that extreme states are associated with better predictive skills, both

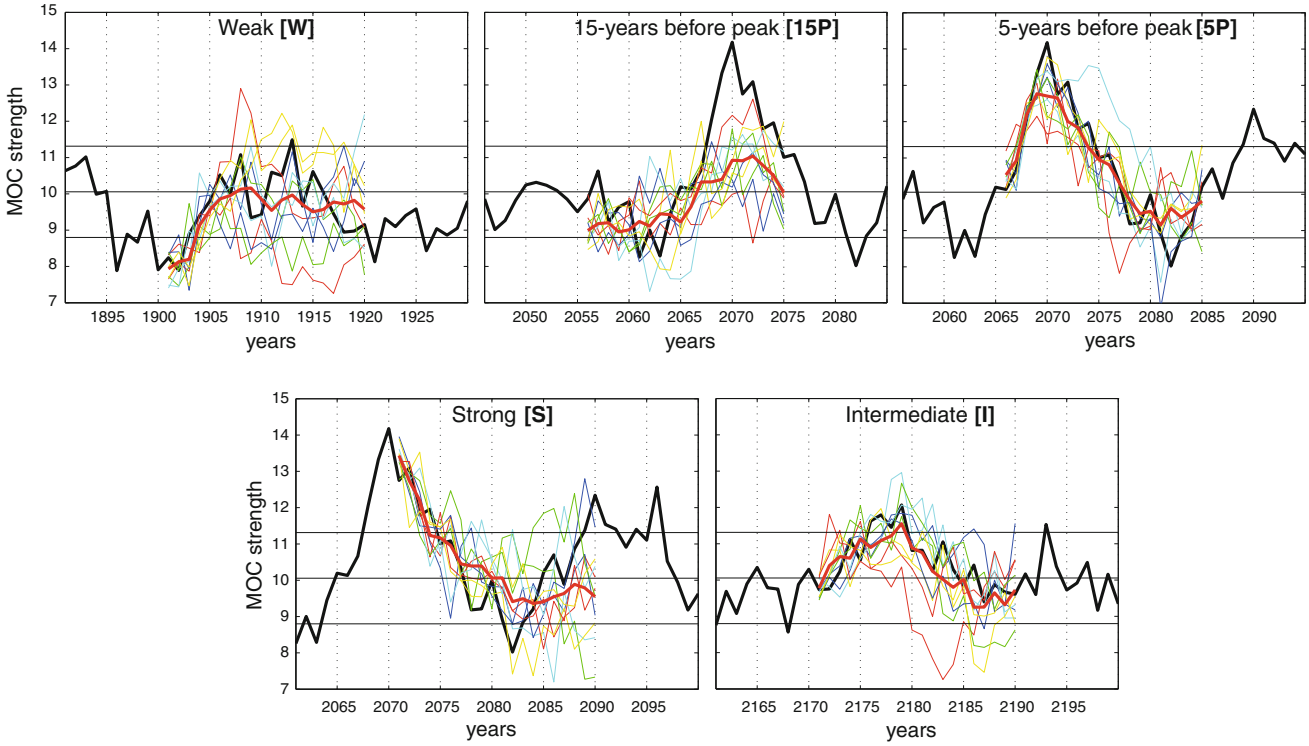

Fig. 3 "Plumes" of maximum-annual mean AMOC between 20 and 50°N from ensembles of the IPSL-CM5A-LR in which the initial conditions have been perturbed. Five ensembles are shown starting from different dates in the control simulation. The individual ensemble members are shown as *coloured lines*, the ensemble mean as the *red thick line*, and the control run as the *thick black line*. The *middle horizontal black line* is the mean AMOC, and both *upper* and *lower horizontal black lines* show standard deviations highlighting the range of variability of the AMOC

experiments I and 15P that start from neutral mean states have no predictive skills (as defined in Sect. 2.1.2); indeed EC is not significant for lead-times of 5 to about 15 years (Fig. 6). However, ES saturates after 5 and 7 years respectively. Based on ES only, this could still indicate a weak predictability. The above results suggest that predictability depends on the AMOC initial state, although the limited number of experiments limits the robustness of this claim.

Given the AMOC impact on climate (Fig. 2), the ability of the model to predict an extremum such as the one of the year 2071 (Fig. 1) could be of great interest. Such an ability is identified in experiment 5P, which shows the second highest predictive skills (after S) with a limit of predictive skill of about 8 years (that is after the peak has been captured, Fig. 6). In contrast, strictly speaking, experiment 15P has no predictive skills (as defined in Sect. 2.1.2). Nevertheless, this experiment still succeeds in capturing the peak of the year 2071 as seen in the plumes in Fig. 3, where most of the members exhibit a positive AMOC anomaly at 15 years lead-time. This feature is somewhat reflected in the statistically significant EC calculated for a lead-time longer than 15 years (i.e. when the peak is included). Although Fig. 3 shows that the amplitude of the peak is not well reproduced, there is some evidence for the ability of the model to capture an extreme AMOC event up to 15 years in advance. Note here that, despite the fact that ES saturates very rapidly and is not associated with a significant EC, EC alone still gives useful information about this ability to capture a peak. This underlines the importance of considering each metric (ES and EC) separately in addition to their combined information, in order to identify interesting features such as extreme events.

By averaging the maximum lead-time at which ES saturates for the five ensemble experiments, we found an average saturation level reached after 8 years. Note, however, that at this lead-time, the average EC amounts to 0.51 which is not significant at the 90 % level when considering the average number of degrees of freedom over each starting date (see Appendix 1). Indeed, Fig. 5 and Fig. 6 show that EC strongly depends on the starting date. For such a limited number of starting dates, it is thus of limited use for an estimation of the average predictive skill. It seems therefore reasonable to claim that, based on ES alone, the average predictive skills of the AMOC is of about 8 years in the IPSL-CM5A-LR model. Again, this lead-time is more than the persistence time of the AMOC

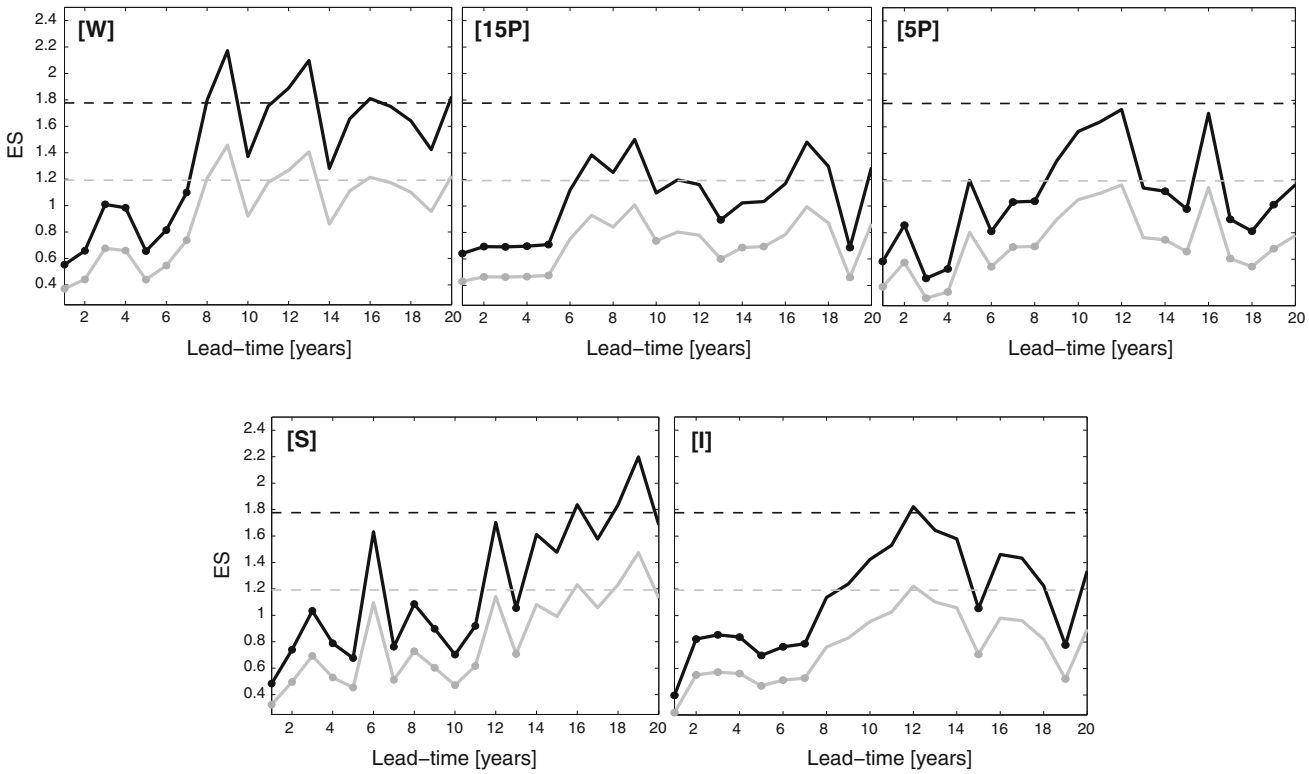

Fig. 4 ES of AMOC index for each of the five ensemble experiments for M10 (*grey line*) and CS03 (*black line*) definitions for lead-time up to 20 years. The threshold at which ES saturates (implying no potential predictability) is shown as the *black* (*grey*) horizontal dashed line for CS03 (M10). *Dots* indicate that ES is statistically smaller than the corresponding threshold at the 95 % level based on a F-test

index, confirming an important role of oceanic dynamics on the predictability of the AMOC.

Figure 6 also brings out some other interesting features worth pointing out. There is some evidence for both ES and EC not to be independent metrics, a decreasing (increasing) ES is generally associated with increasing (decreasing) EC. This claim is further supported in Appendix 1. There is also some apparent "return" of predictive skills for both experiments 15P and 5P. There is, indeed, some evidence for ES returning below the saturation level and recovering statistical significant a few years after saturation, with corresponding EC which also recovers significance. Note that this increasing of EC is relatively small in 5P (<0.1), compared to 15P (>0.5); the reason for the significant increase in this latter is certainly due to its ability to capture the extreme AMOC event present in the second decade (at least in terms of its presence and its sign). Although this apparent "return" of skill has already been pointed out by several studies, its origin still remains unclear. For example, Newman et al. (2003) suggest that this reflects variations of the actual noise rather than a true skill, while Hermanson and Sutton (2009) rather suggest that this might be a consequence of the use of a simple univariate measure to quantify predictability. Here, it is not to be excluded that this "return" of skill in the second decade could be related to the peak of energy at 20 years (as found in the control simulation, Escudier et al. 2012), which might increase correlation and then predictability for larger timescale. The origin of this phenomenon definitely merits further attention and should be the main focus of further studies.

4.2 An early warning system to predict extreme AMOC events?

Even though such events are rare and may be viewed as "surprises", providing an early warning system is extremely desirable considering their possible major climatic impacts. Results above showed clear evidence for the ability of the model to capture extreme AMOC events. More specifically experiment 15P gives hope for predicting such events earlier than the 8-year average predictive skills identified by the PPP approach. Given the lack of AMOC observations, we investigate here whether there exist monitorable precursors to such extreme events and whether they are themselves predictable. Note that these are likely to be strongly model-dependent.

As briefly mentioned in Sect. 3, Escudier et al. (2012) identified a 20-year cycle associated with the AMOC

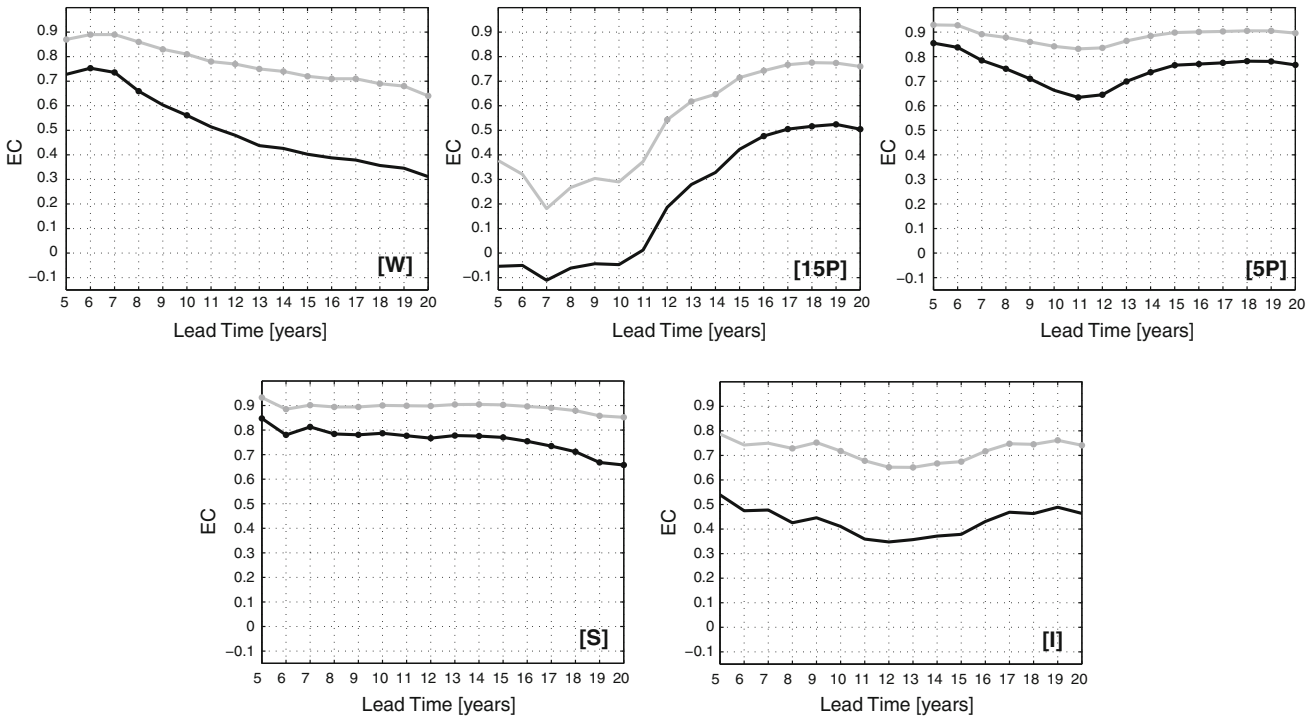

Fig. 5 EC of AMOC index (as calculated by the Fisher transformation) for each of the five ensemble experiments for M10 (*grey line*) and CS03 (*black line*) definitions for lead-time from 5 to 20 years. *Dots* indicate that EC is statistically significant at the 90 % confidence level using a one-tailed Student *t* test

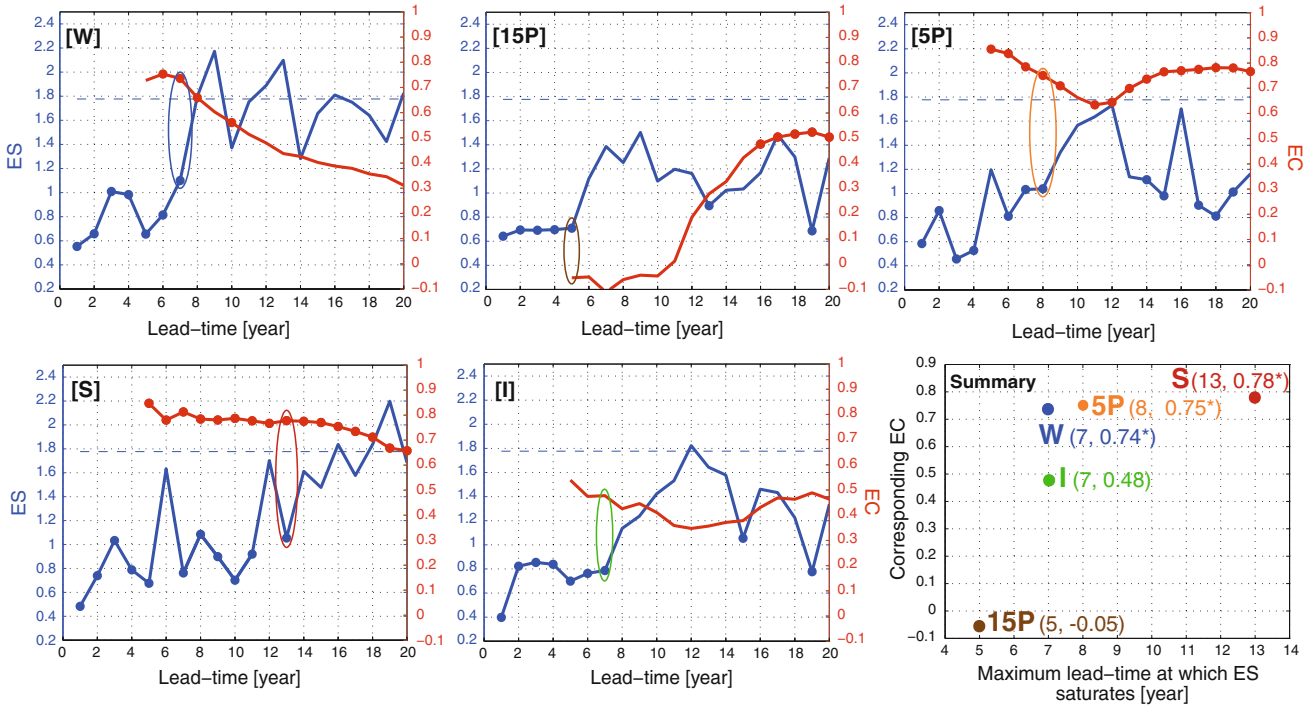

Fig. 6 Plots of results from CS03 definition showing ES (*blue line*) and EC (*red line*) against lead-times for each of the five ensemble experiments. *Dots* indicate statistical significance at the 95 % (90 %) confidence level for ES (EC). The summary plot shows the maximum lead-time at which ES saturates with its corresponding EC (as indicated by the *oval circles*) for each experiment (statistically significant ECs at the 90 % are marked with an asterisk)

variability in the IPSL-CM5A-LR control integration. Figure 7 shows a simplified schematic of the mechanism responsible for one half of this cycle. For example, starting from a positive temperature and salinity anomaly in the Labrador Sea, the surface mean currents along the subpolar gyre advect these anomalies of same sign eastward. The anomalous salty surface waters favour deep convection south of Greenland and Iceland along their path, thereby inducing an AMOC intensification after 9 years, and they eventually reach the Nordic Seas. There, the associated positive temperature anomaly associated with the salinity anomaly induces sea ice melting and an anomalous cyclonic circulation over the Nordic Seas, which strengthens the East Greenland Current (EGC), and in turn creates negative temperature and salinity anomalies in the Labrador Sea. This process lasts 10 years, and the second phase of the cycle begins. Escudier et al. (2012) found therefore evidence for the EGC intensity and water properties in the Labrador Sea to be precursors of changes in the model's AMOC, with a lead-time of about 14 years for the EGC and 11 years for the Labrador Sea salinity.

Using this apparent predictability in a practical way requires that a large change in the main identified precursors always lead to a corresponding change in the AMOC index. Figure 8 shows time series of the AMOC index, sea surface salinity (SSS) in the Labrador Sea and the EGC index (defined as the southward meridional transport across the Denmark Strait of waters with a salinity lower than 34 psu) in the control integration for each ensemble with the corresponding plumes superimposed. It is found that of the 6 identified "events" (represented as letters in Fig. 8), for which within 5 years at least one of the precursors changes by more than 2σ and the other one by at least 1.5σ, 5 are followed by an AMOC change of the correct predicted sign, of which 4 show a change larger than 1.5σ. This large change in AMOC occurs about 15 (13) years after a large change in EGC (SSS in the Labrador Sea). This result is consistent with the lead-times summarized in Fig. 7, and therefore illustrates the potential predictive role of these two variables. Large magnitude of change in precursors (around 0.9–1.2 Sv and 0.5–0.7 psu) therefore suggests the potential predictability of extreme AMOC events through observations of properties in the Labrador Sea and Denmark Strait. This also suggests that in the case of extreme AMOC events, there is the possibility for longer lead-time of predictability (13 or 15 years, depending on the predictor) than the average 8 years found earlier (see Sect. 4.1.2). Note that this longer lead-time has previously been discussed for the experiment 15P alone. Its ability to capture the peak 15 years later is indeed linked to the state of its EGC precursor, which is extreme at the beginning of the experiment (point C, Fig. 8).

Finally, hope for predictability of an extreme AMOC to go even beyond this suggested decadal lead-time could arise if these two precursors exhibit in turn some potential predictability skills. Indeed, both the EGC and SSS in the Labrador Sea have been found to have some robust predictability for lead-times up to 9 and 7 years, respectively (not shown). The possibility of predicting an extreme EGC event at least 9 years in advance gives hope for the predictability of an extreme AMOC event beyond 2 decades ahead, although this has not been tested prognostically.

We therefore found convincing evidence that extreme changes in the AMOC as seen in the IPSL-CM5A-LR model might be potentially predictable up to 2 decades ahead from the monitoring of its high-latitude Atlantic precursors. Hawkins and Sutton (2008) already found such a relationship with the HadCM3 model. If a comparable mechanism to the one identified in the IPSL-CM5A-LR model (Escudier et al. 2012) occurs in the real ocean, which remains to be demonstrated (encouraging elements can be found in Swingedouw et al. 2012), then the ability to predict AMOC fluctuations is promising for potential predictability of climate at multi-decadal timescales.

5 Spatio-temporal predictability of Atlantic climate

Potential predictability of both surface temperature and precipitation is now evaluated using and comparing both the DPP and PPP approaches. As mentioned in Appendix 1, a threshold for "useful" potential predictability is often hard to define in the DPP approach, as it only relies on a long control integration. On the other hand, it remains a cheap (in terms of computation time) and easy way to evaluate average predictive skills from long time series. This differs from the PPP approach which is much more expensive, but better evaluates the growth of perturbations in ICs and therefore the effective predictability within models. Given a choice of starting date, this approach can also illuminate the link between temperature and precipitation predictability and the AMOC.

5.1 Potential predictability of surface temperature

Figure 9 shows predictability maps of Atlantic surface temperature up to 1 and 2 decades as identified by both the DPP and PPP approaches in the IPSL-CM5A-LR model. For the former approach, the maps show the *ppvf* for 10 and 20-year means and are shown in Fig. 9a. For the PPP approach, regions combining surface temperature with both statistically significant EC and ES statistically smaller than the saturation level at the considered lead-time (i.e. regions potentially predictable as defined in Sect. 2.1.2) are shown in Fig. 9b as a function of the number of experiments for

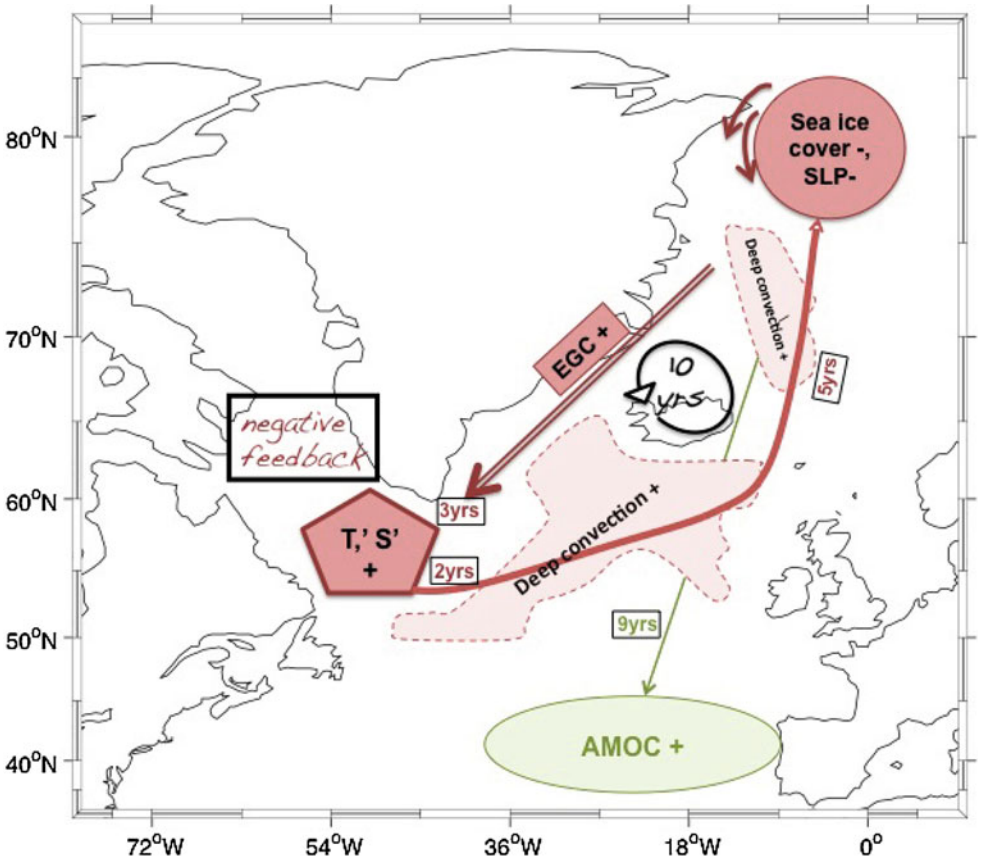

Fig. 7 Schematic view of mechanisms responsible for one half of the decadal AMOC cycle in IPSL-CM5A-LR. Items in red are actively involved in the 20-year cycle. T' stands for upper ocean temperature anomaly, S' for upper ocean salinity anomaly. EGC is the East Greenland Current and SLP the sea level pressure. Starting from a positive temperature and salinity anomaly, the signs in the *red boxes* indicate the sign of the correlation among items, and the *number* in the *square black boxes* the time lag in years. Items in *green* are periodically perturbed by the 10-year cycle but not actively taking part in its generation. The signs and the number of years denote correlation and time lags as above (Adapted from Escudier et al. 2012)

which these conditions are met. Note that only the lead-time 0–10 years and 10–20 years are considered respectively.

Over the ocean, the regions of highest (more than half of the experiments) predictive skills at both 1 and 2 decades identified by the PPP approach coincide to some extent with those of highest *ppvf* scores (for which 10–40 % of the variance is in the considered decadal band, Fig. 9a). These regions mainly include the convection sites (as identified by Escudier et al. 2012) together with the NAC path, and are in good agreement with results from the diagnostic multi-model predictability studies of Boer (2004) and Boer and Lambert (2008). The PPP approach also brought some hints of potential predictability (less than half of the experiments) for the two timescales in regions including the southeastern branch of the subtropical gyre and the tropics (more specifically the western deep tropics up to 1 decade extending to the northern western tropics up to 2 decades). These two regions are also identified by the DPP approach, although some discrepancies are present in the tropics; up to 2 decades, strongest signals are identified in the southern tropics rather than in the northern tropics. Interestingly, these regions of weak signals (i.e. the southeastern branch of the subtropical gyre and the tropics) are each identified in experiments including the extreme AMOC event of 2071, namely in experiment 15P and S mainly over 2 decades (see Appendix 2). Although it remains difficult to draw robust conclusions from the limited number of experiments, this suggests that an extreme AMOC event might favor the potential predictability of these regions. However, the reason for the weak scores in these regions in 5P remains to be clarified.

In general, in both approaches, potential predictability over land is less significant than over the ocean. It is found over the coastal areas bordering some of the potentially predictable oceanic regions (that mainly include the maritime-influenced regions of western Africa, the western coast of the Iberian peninsula, and the northern coasts of the British Isles and South America), and it seems to be favored by extreme AMOC events (see Appendix 2). The DPP approach identifies additional land areas located further away from the coast (i.e. in Europe, in both the African and South American continents). Note, however, that these additional land areas are regions of low *ppvf* values ($ppvf < 0.1$, Fig. 9a).

Finally, the evidence of a relationship between the potential predictability of surface temperature and the AMOC is due to the fact that the major regions identified as potentially predictable by both approaches, are also

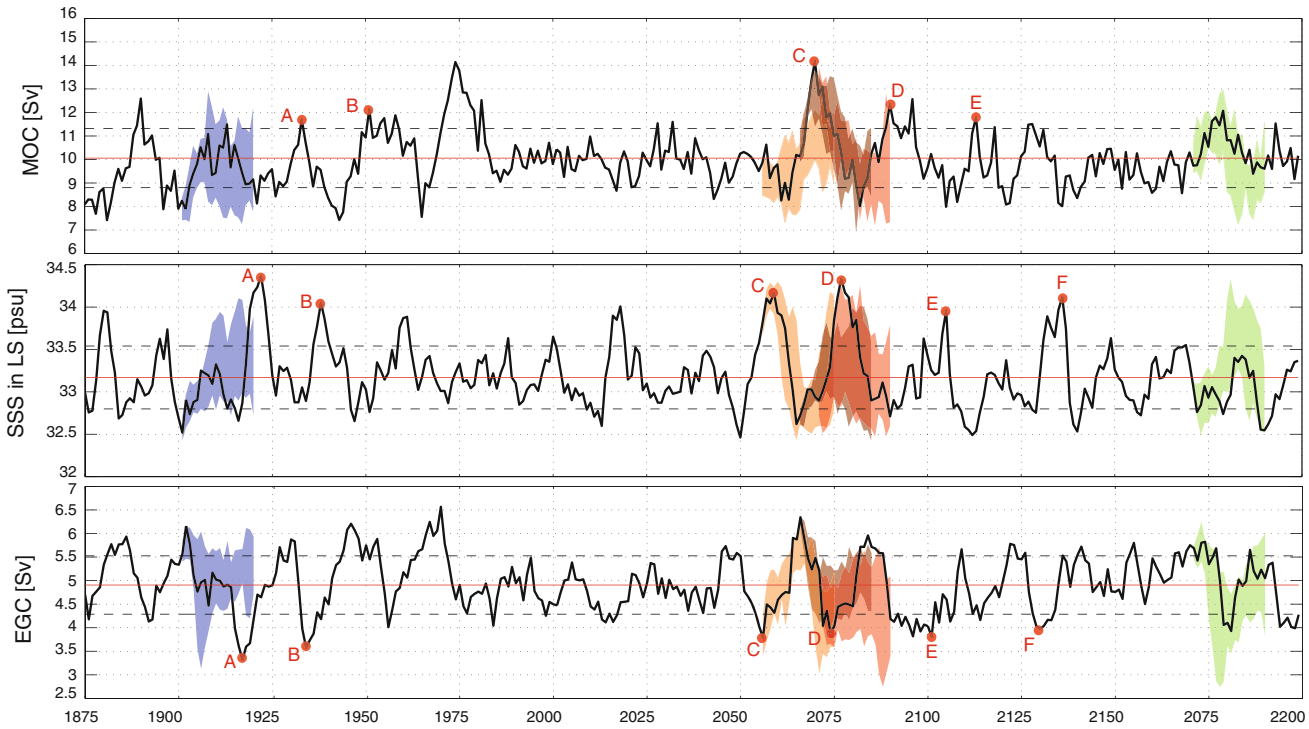

Fig. 8 Time series of AMOC index (*top panel*), SSS in Labrador Sea (*middle panel*), and East Greenland Current index (*bottom panel*). *Thick black line* is the control integration, *dashed lines* are the standard deviation, *red line* is the mean, and envelope of each experiment is shown as *coloured shading*. *Letters* (*A-B–C-D-E–F*) correspond to identified "events" (see text in Sect. 4.2 for further details)

remarkably similar to the regions significantly sensitive to decadal AMOC fluctuations (as shown in Fig. 2a).

5.2 Potential predictability of precipitation

Potential predictability of precipitation (Fig. 10) is considerably smaller than for surface temperature, in good agreement with the multi-model approach of Boer and Lambert (2008). Similar to the latter study, the Nordic Seas are the most prominent regions where precipitation seems to be predictable at both timescales. There are also some patches of predictability over the subpolar gyre in both approaches. Note that the DPP approach identifies additional regions (both oceanic and continental) mainly over the tropics (Fig. 9b). As for surface temperature, these additional regions have low *ppvf* values. Furthermore, as for regions of weak signals for surface temperature, regions identified by both approaches appear in experiments including the extreme AMOC state of the year 2071 (experiments 15P and S, see Appendix 2).

Similarly to surface temperature, the evidence for a link between an extreme AMOC event and predictability of precipitation in the above identified regions is due to the fact that they are also regions sensitive to decadal AMOC fluctuations (as shown in Fig. 2b). These results suggest the mechanisms responsible for climate predictability to be strongly linked to the mechanisms behind decadal AMOC variability. Note, nevertheless, that this link between regions potentially predictable and those sensitive to decadal AMOC fluctuations is less clear for precipitation than for temperature, and this could also explain the weaker PPP skills in precipitation in the tropical and subtropical regions, given our experimental set-up for the prognostic approach largely focused on specific AMOC events.

6 Summary and discussion

6.1 Potential predictability of the AMOC

The predictive skills of the AMOC index have been quantified by the prognostic (PPP) approach for five experiments starting from different AMOC initial states, using both the ensemble spread (ES) and the ensemble correlation (EC). In most cases, ES (EC) increases (decreases) with lead-time, and hence predictability is lost after a certain lead- time. In some cases, an apparent "return" of skill is detected a few years after saturation. This has to be interpreted carefully as it could simply reflect noise rather than predictability (e.g. Newman et al. 2003). Our experiments showed that it could nevertheless also be related to the large variability of the AMOC at the

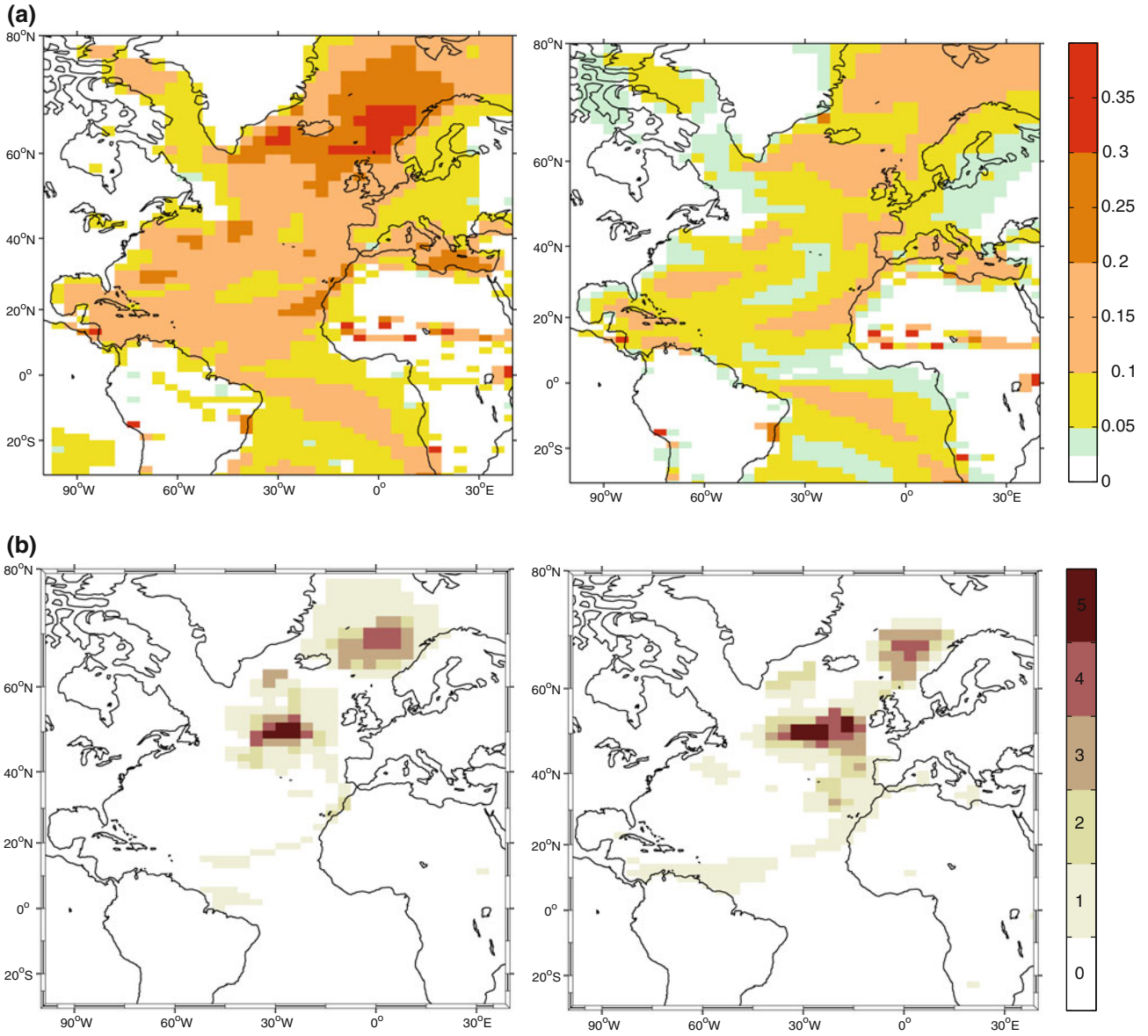

Fig. 9 Potential predictability of surface temperature in the Atlantic sector identified by: **a** the DPP approach showing maps of the internally generated decadal *ppvf* for 10-year (*left panel*) and 20-year (*right panel*) means in the unforced control climate of IPSL-CM5A-LR (the colored areas are significant at the 95 % level according to a F-test); **b** the PPP approach showing maps of the number of starting dates (out of 5) where grid points are potentially predictable (i.e. where it combines both statistically significant EC at the 90 % confidence level according to a student test and normalized ES smaller than saturation level at the 95 % level according to a F-test) up to one (*left panel*) and two (*right panel*) decades

20-year timescale found in the control simulation (Escudier et al. 2012). EC was found particularly useful to detect such features in the simulations.

It is difficult to determine average predictability skills in the perfect model experiments as it implies averaging skills over several starting dates which themselves have very different predictability skills. Nevertheless, it seems reasonable to claim that the modelled AMOC has an average predictive skills of 8 years in the IPSL-CM5A-LR model, when considering the average lead-time at which ES saturates. The corresponding EC averaged over all starting dates is not significant. Note that the AMOC index has also been found to have a persistence time (estimated from the AMOC index autocorrelation function in a red noise framework) of about 4–5 years, which is less than the average predictive skills found here. This suggests a role of the oceanic dynamics in this predictability. This average lead-time of predictability of the AMOC index found in the IPSL-CM5A-LR model is shorter than those identified in most similar published studies, for which the predictability

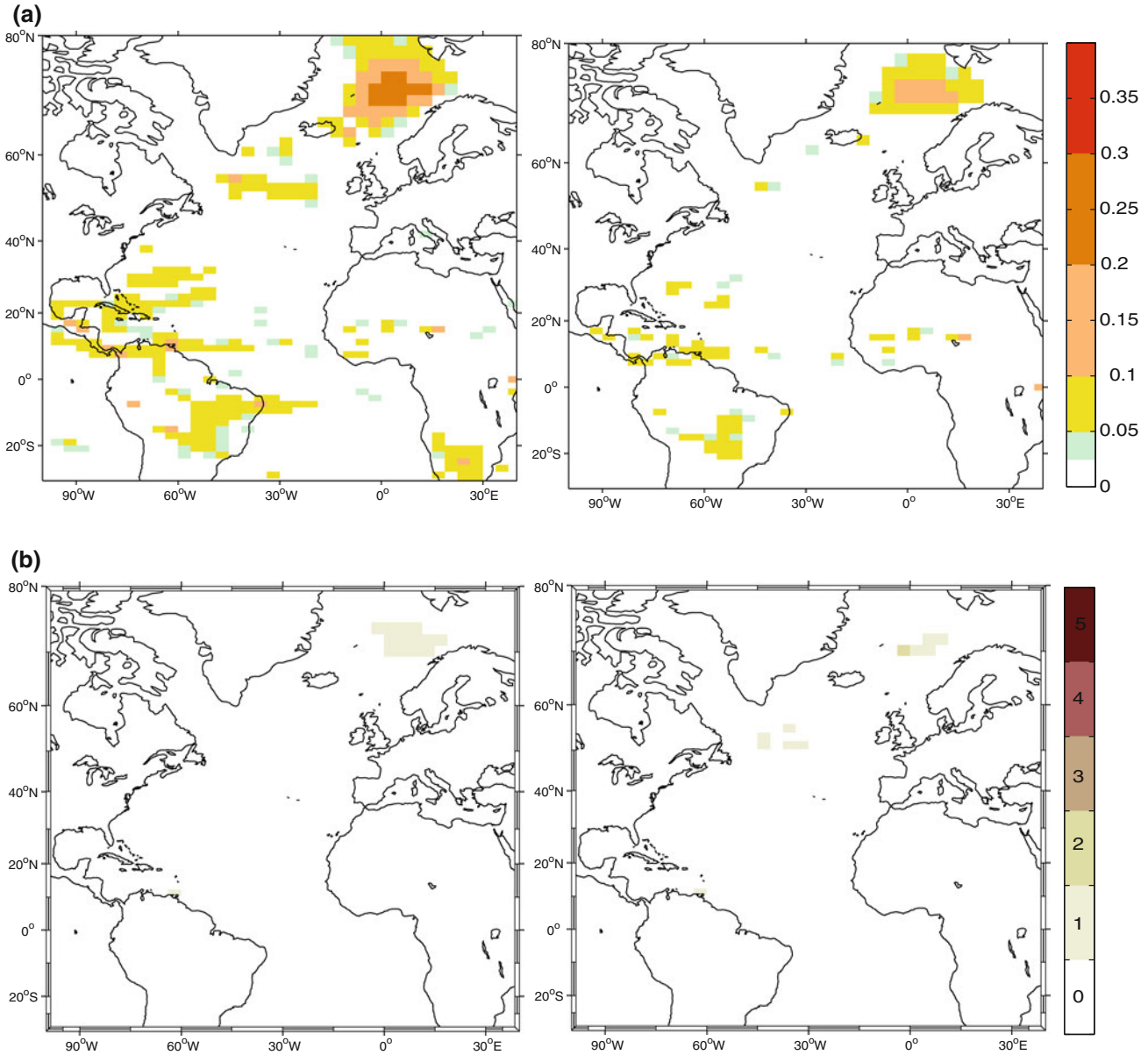

Fig. 10 Potential predictability of precipitation in the Atlantic sector as defined in Fig. 9

lead-time could reach 2 decades ahead (e.g. CS03; M10; Pohlmann et al. 2004; Collins et al. 2006). It is, however, somewhat in agreement with Teng et al. (2011) who found the AMOC to be predictable for only one decade in the CCSM3 model. Hermanson and Sutton (2009) identified a shorter lead-time in the HadCM3 model, with an average predictive skill of about 5 years. The IPSL-CM5A-LR model belongs to the middle-range of timescale of AMOC predictive skills identified so far in the literature. Such as comparison with existing studies should, however, be considered carefully because of the many differences in the experimental protocol used among predictability studies.

When considering the predictive skills of each ensemble experiment separately, there is evidence for predictive skills to depend on the AMOC initial state. Indeed, the highest skills have been found (in descending order) in the experiments starting (1) from a strong AMOC initial state (up to 13 years), (2) 5 years before a maximum peak (up to 8 years) and (3) from a weak AMOC initial state (up to 7 years). In contrast, no predictive skills (as defined in Sect. 2.1.2) have been found for experiments starting from an intermediate AMOC initial state and 15 years before a maximum peak. This is essentially because the ensemble correlation rapidly becomes insignificant. Based on ES alone, these starting date could be considered as being predictable 7 and 5 years ahead respectively. Nevertheless, generally, predictive skills have therefore been identified for experiments starting or nearly-starting from an extreme

AMOC state. The above results also suggest better predictive skills for an initial state corresponding to an anomalously strong AMOC than those corresponding to a weak AMOC, in good agreement with the perfect model studies of CS03 and Collins et al. (2006). However, the number of members (10) for each experiment is somewhat low to fully assess the robustness of such an impact of the AMOC initial state on its predictability. Note furthermore that, although no predictive skills (as defined from both EC and ES) have been identified in the experiment starting 15 years before a peak, this specific experiment still showed the ability of the model to capture relatively well an extreme AMOC event on a longer lead-time than the average one identified by the PPP approach.

In view of the major climatic impact induced by such extreme events, the development of an early warming system would be of great value. The present study shows that this is made possible through the monitoring of the high-latitude precursors of the AMOC in this model (which are the EGC and the upper-ocean properties in the Labrador Sea), which leads to an increase in predictive skills of extreme AMOC events up to 2 decades ahead. The perspective of an early warning system of such events thus motivates the monitoring of the EGC strength and water properties in the Labrador Sea. In this perspective, observation programs across e.g. the WOCE-AR7/A1 section for the Labrador Sea (http://cchdo.ucsd.edu/atlantic.html) and the East Greenland shelf and slope of south of Denmark Strait (Brearley et al. 2012), as well as the maintenance of mooring arrays in these areas, are likely to be of greatest added value to constrain prediction of the AMOC. Similar observational targets have also been pointed out by Hawkins and Sutton (2008) using the HadCM3 model.

6.2 Potential predictability of the North Atlantic climate

Changes in the AMOC have been found to have significant and widespread climate impacts. The prospect for predictability of decadal AMOC fluctuations is therefore promising for potential predictability of climate. This latter has been investigated using both diagnostic (DPP) and PPP approaches. They give overall very similar results, and strongly agree on the regions that exhibit the highest predictive skills. Some discrepancies, nevertheless, arise for regions where only some hints of predictability have been identified. Indeed, these regions are often larger in the DPP approach than in the PPP approach. In other words, the DPP estimation seems less discriminant. To strengthen the robustness of our results, note that the regions claimed to have some hints of predictability below are regions identified by both the DPP and PPP approaches.

The far North Atlantic (that includes the convection sites of the model and the NAC path) has been identified as the region exhibiting the highest predictive skills. Surface temperature is potentially predictable up to 2 decades in advance there, in good agreement with previous studies (e.g., Collins 2002; Boer 2004; Pohlmann et al. 2004; Hawkins et al. 2011; Branstator et al. 2012). Note that this ability to predict the North Atlantic subpolar gyre also gives hope for potential multi-year forecast of tropical storm and hurricane frequency (Smith et al. 2010). Some hints of potential predictability are also identified at this timescale in the subtropics (mainly over the southern part of the eastern branch of the subtropical gyre) and the tropics (mainly over the north western tropics). The predictability found in the latter region is clearly different from results of Pohlmann et al. (2004) in the ECHAM5-MPI/OM climate model, and also contrasts with Collins (2002), who found signals only up to the interannual timescales in the tropics in HadCM3. To some extent our result, however, agrees with Hawkins et al. (2011) who also found decadal predictability in the tropics in the HadCM3 model, but only up to 1 decade ahead and restricted to the southern tropics. Land areas display little potential predictability compared to oceans. Potential predictability at decadal timescales is generally restricted to the coastal areas bordering some of the oceanic regions identified above; they mainly include the coast of western Africa, the western coast of the Iberian peninsula, both the northern coast of the British Isles and South America. Signals over maritime Europe as identified by Boer and Lambert (2008) and Pohlmann et al. (2004) are not brought out as clearly in our study. Although potential predictability is largely absent for precipitation (as noted by Pohlmann et al. 2004; Boer and Lambert 2008; Boer 2011 in particular), there are some hints of potential predictability up to 2 decades over the convection sites of the Nordic Seas and the subpolar gyre.

Similarly to the AMOC, regions with weak but significant predictability (i.e. the tropics and subtropics for temperature, the Nordic Seas and subpolar gyre for precipitation) seem to depend at least partly on the AMOC state. Results suggests that extreme AMOC events might favor the potential predictability of regions of weak signals, as the latter are in most cases identified when the predicted time-period includes such events. Although the origin of a possible link between climate predictability and extreme AMOC still needs to be clarified, the likelihood for such a link is strengthened by the fact that regions identified as potentially predictable (for both surface temperature and precipitation) are also all strongly influenced by decadal AMOC fluctuations. This suggests that the mechanisms responsible for climate predictability are to some extent linked to the decadal AMOC variability. More research to understand the specific mechanisms that lead to predictability is, however, still needed. The present

study nevertheless underlines that the potential predictability of the AMOC could therefore lead to significant decadal predictability of climate (where the AMOC has a sufficiently strong impact), and may therefore be of economic and societal importance (e.g. Meehl et al. 2009).

6.3 Evaluation of predictive skills

Different definitions of predictability, different experimental protocols and metrics have often been used among the previous studies to evaluate predictive skills. It therefore remains difficult to estimate the weight of the metrics on the level of predictability found here in the IPSL-CM5A-LR model by comparing this level to those found in previous studies. Nevertheless, from a methodological point of view, our study still puts forward some interesting results regarding the evaluation of predictive skills.

Regarding the prognostic approach, we showed that combining ES and EC should be preferred in principle but it is sometimes difficult to apply in practice. For the evaluation of average predictive skills, EC was found insignificant. In the case of 15P, it reduced the quantification of predictability skill for weak lead-times but greatly helped to highlight the ability of the model to capture the late peak. We thus still claim that both metrics should be considered in parallel. Our results also suggest that considering either the ensemble mean of an experiment or each individual member as a baseline in the calculation of both metrics does not affect the overall results.

As already mentioned, both diagnostic and prognostic approaches generally brought out the same main features concerning both temperature and precipitation predictability. Marginal discrepancies concerned the regions of weak signals. Because of the difficulty to define a "useful" threshold of potential predictability in the DPP approach, the PPP approach allows more detailed analysis. It however relies on the subjective choice of starting dates, number of members and experiments. Despite the limited number of experiments starting with similar AMOC states, another aspect brought out by the PPP approach is that both the AMOC and some regions might have higher predictive skills under specific initial states, often when the predicted time period includes an extreme AMOC. This result needs to be confirmed by further work. Although reliable estimates of skill conditional on specific initial states are difficult to determine (due to the small sample for verification), more systematic experiments starting with similar initial states (i.e. weak, intermediate, strong, just before a peak) should therefore be undertaken. It could even be extended to further scenarios such as starting just after a peak. Note that, this dependence on initial states already exists with seasonal-to-interannual climate forecasts dependent on the phase of El-Niño Southern Oscillation (e.g. Chen et al. 2004), and it is expected to be the case with decadal predictions (Griffies and Bryan 1997). The present study suggests that forecasts starting from an extreme phase of natural internal variability can be more skillful than those starting from average conditions. In that sense, studying skill from case studies may prove more useful to understand predictability mechanisms than computing average skill from numerous start dates as done in most previous studies.

It is also important to bear in mind that here we have assessed the upper limit of both the AMOC and climate predictability as both perfect model and near perfect knowledge of the current state of the climate system are assumed. Indeed, climate models still have significant biases compared to observations, and their possible impacts on the level of predictability skill of a model cannot be ignored. As an illustration, Branstator et al. (2012) found that, using six state-of-the-art AOGCMs, the average lead-time of predictability for subsurface temperature (especially in the North Atlantic) varied considerably between the models highlighting how poorly the North Atlantic predictability must be represented in some, or perhaps all, of the six models. Therefore, bearing in mind the possible impact of the limitations of the IPSL-CM5A-LR model, its lack of deep convection in the Labrador Sea (Swingedouw et al. 2007) might well affect the effective level of predictability skill. This problem should be addressed in future work. The 20-year variability cycle in the subpolar North Atlantic in the model also greatly influences the present results and its occurrence in the real world further needs to be assessed. The possibility that lower predictability limits would arise in a real predictive system with this model cannot be ruled out (see Swingedouw et al. 2012). However, to the extent that both diagnostic and prognostic approaches are appropriate measures of skill, the present results give some indications as to where and to what extent skillful decadal forecasts might be possible.

Acknowledgments This work was supported by the "Gestion des Impacts du Changement Climatique" Programme (GICC) under the EPIDOM project funded by MEDDTL (French Ministery of Ecology and sustained development). The authors are grateful to Bablu Sinha, Roland Séférian and Lisl Weynans for their useful comments. We also thank the anonymous reviewers for their very useful comments.

Appendix 1: diagnostic potential predictability (DPP) approach

The DPP approach attempts to quantify the fraction of long-term variability (considered as predictable) as compared to the internal variability (considered as chaotic and unpredictable). The long-term variability that rises above this noise is deemed to arise from processes operating in

the physical system that are assumed to be, at least potentially, predictable. Boer (2004) defined the potential predictability variance fraction (*ppvf*) as an estimate of DPP. Here, we use its non-biased estimation (see Boer 2004 for further details) defined as:

$$ppvf = \frac{\sigma_N^2 - \frac{1}{N}\sigma^2}{\sigma^2} \quad (1)$$

where σ_N^2 represents the variance of N-year means, and σ^2 represents the interannual variance of the variable X considered. The *ppvf* varies between 0 and 1; a *ppvf* close to 0 implies no long-term variability and thus no potential predictability. Conversely, *ppvf* close to 1 implies large predictability. Statistical significance of *ppvf* is judged using a F-test at the 95 % confidence level. A threshold for "useful" potential predictability is however hard to define, as it is likely to be purpose and situation dependent. Nevertheless, it remains an easy statistical way to estimate the average predictive skill in a model.

Prognostic potential predictability (PPP) approach

Prognostic predictability studies consist in performing ensemble experiments with a single coupled model by perturbing the initial conditions (ICs) supposed to represent atmospheric chaotic noise or uncertainty in the present climate state. Ensemble Correlation (EC) and Ensemble spread (ES) are the two deterministic measures used here to quantify the predictability of the simulated climate.

Ensemble correlation (EC)

In the forecast framework, correlation addresses the question: "to what extent are the forecasts varying coherently with the observed variability?". In the "perfect ensemble" approach, the definition of "observed variability" differs from one study to another: it can refer to the variability of the ensemble mean (i.e. the average of all members for an individual ensemble) as in Msadek et al. 2010 (M10) or to the variability of an individual member as in Collins and Sinha 2003 (CS03). In other words, in the M10 approach, predictability skill is evaluated by correlating each member of the ensemble to the ensemble mean whereas in the CS03 approach each member is correlated to each other. If M is the number of members, we therefore obtain M (resp. M(M-1)/2) individual correlations for M10 (resp. CS03). Independently of the approach used, the formula for the individual correlation of any pairs p is:

$$r_p = \frac{\left[T\sum_{t=1}^{t=T} A_t B_t\right] - \left[\sum_{t=1}^{t=T} A_t \sum_{t=1}^{t=T} B_t\right]}{\sqrt{\left[T\sum_{t=1}^{t=T} A_t^2 - \left(\sum_{t=1}^{t=T} A_t^2\right)\right]\left[T\sum_{t=1}^{t=T} B_i^2 - \left(\sum_{t=1}^{t=T} B_t\right)^2\right]}} \quad (2)$$

where T is the number of years over which we want the correlation for, and A and B are the members forming the pair p. Once the individual correlations of all pairs have been calculated (M pairs for M10, M(M-1)/2 pairs for CS03), EC of the ensemble is computed as the mean of all individual correlations through a Fisher Transformation (Fisher 1921). We will consider the two definitions presented above to evaluate possible differences in their respective score of predictive skills. Statistical significance of the resulting EC is judged using a one-tailed Student's t-distribution test at the 90 % confidence level with degree of freedom corresponding to the average degree of freedom of all individual correlations. The degree of freedom of these latter takes into account the persistence in the two timeseries following Bretherton et al. (1999). In order to gain confidence in the estimation of the EC significativity, we also evaluated its significance by using the "field significance" approach (e.g. Livezey and Chen 1983). The statistical significance of EC obtained with this test is very similar to the ones obtained from the average degree of freedoom of all individual correlations.

Ensemble spread (ES)

ES or Root Mean Squared Error RMSE or again the Mean Squared Skill Score MSSS (as defined by the US CLIVAR working group on Decadal Predictability, http://clivar-dpwg.iri.columbia.edu) addresses the question: "how large are the typical errors in the forecast (among members) relative to those implied by baseline?". Consistently with EC, we consider the two definitions of the baseline which arise from the literature: for a given lead-time LT, ES of an ensemble of individual members i is defined respectively as:

$$ES_{M10}(LT) = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(X_i(LT) - \bar{X}(LT)\right)^2} \quad (3)$$

$$ES_{CS03}(LT) = \sqrt{\frac{2}{M(M-1)}\sum_{i=1}^{M}\sum_{j=i+1}^{M}\left(X_i(LT) - X_j(LT)\right)^2} \quad (4)$$

where we define: $\bar{X}(LT) = \frac{1}{M}\sum_{i=1}^{M} X_i(LT)$

We demonstrate below that there actually exists a relationship of proportionality between (3) and (4). Let consider the two following definitions of Mean Squared Error:

$$E_{M10} = \frac{1}{M}\sum_{i=1}^{M}\left(X_i - \bar{X}\right)^2 \quad (5)$$

$$E_{CS03} = \frac{2}{M(M-1)}\sum_{i=1}^{M}\sum_{j=i+1}^{M}\left(X_i - X_j\right)^2 \quad (6)$$

By expanding $(X_i - \bar{X})^2$ in (5) and after a few rearrangements we show that:

$$E_{M10} = \overline{X^2} - \bar{X}^2 \tag{7}$$

Then, if we introduce:

$$E = \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=1}^{M} (X_i - X_j)^2 \tag{8}$$

We show by a recurrence reasoning that:

$$E = 2E_{CS03} \tag{9}$$

By expanding $(X_i - X_j)^2$ in (8) and after a few rearrangements, we show that:

$$E = \frac{4}{M-1} (\overline{X^2} - \bar{X}^2) \tag{10}$$

By combining (7), (9) and (10), we obtain the following relationship:

$$E_{CS03} = \frac{2M}{M-1} E_{M10} \tag{11}$$

Therefore

$$ES_{CS03}(LT) = \sqrt{\frac{2M}{M-1}} ES_{M10}(LT) \tag{12}$$

And there exists a factor of proportionality $\sqrt{\frac{2M}{M-1}}$ between the ensemble spread of both CS03 and M10 definitions.

Generally, the trajectories of individual members diverge with time and thus ES increases with LT. When ES saturates at the control RMSE, we consider that there is no more potential predictability: the spread of the forecast is of similar magnitude as the natural spread of the modelled climate, and no predictability can be inferred. In CS03 (M10) the control RMSE is defined as $\sigma\sqrt{2}$ ($\sigma\sqrt{[(M-1)/M]}$), where σ is the standard deviation of the control integration. Statistical significance of ES as compared to the respective threshold (or control RMSE) is judged using a F-test at the 95 % confidence level. The maximum LT at which a variable is said to be potentially predictable is the last LT before ES persistently exceeds the threshold.

Relationship between ES and EC

We consider here centred and normalized (by the standard deviation) data in time t.

We consider the CS03 definition of ES and EC:

$$ES^2 = \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} (X_i(t) - X_j(t))^2 \tag{13}$$

$$EC = \frac{2}{M(M-1)} \sum_{i=1}^{M} \left(\sum_{j=i+1}^{M} corr(X_i(t), X_j(t)) \right) \tag{14}$$

where the discrete time correlation using centred and normalized data is:

$$\mathrm{corr}(X_i(t), X_j(t)) = \frac{1}{T} \sum_{t=1}^{T} X_i(t) X_j(t)$$

We consider the average of ES over the period of time T: $<ES^2>_T = \frac{1}{T}\sum_{t=1}^{T} ES^2(t)$

By expanding $(X_i(t) - X_j(t))^2$ in (13) and after a few rearrangments we can show that:

$$<ES^2>_T = \frac{1}{T} \sum_{t=1}^{T} \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} (X_i^2(t) + X_j^2(t) - 2X_i(t)X_j(t))$$

$$<ES^2>_T = \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} \frac{1}{T} \sum_{t=1}^{T} \left(X_i^2(t) + X_j^2(t) \right) - \frac{4}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} \frac{1}{T} \sum_{t=1}^{T} X_i(t) X_j(t)$$

Since the variables are centred and normalized:

$$\frac{1}{T} \sum_{t=1}^{T} X_i^2(t) + X_j^2(t) = 2$$

Hence, we obtain the following result:

$$<ES^2>_T = 2(1 - EC)$$

For the real case, where the data are not normalized and centred, which is more appropriate for ES estimation, no such simple relationship can be found analytically, but we hypothesize that ES and EC remain related. A few illustrations of such link are evidenced in the manuscript (see Sect. 4.1.2) and plead in favour of this hypothesis.

Appendix 2

We present the individual correlation maps for temperature (Fig. 11) and precipitation (Fig. 12) for each starting date in the PPP protocol. These individual results are aggregated in Figs. 9 and 10.

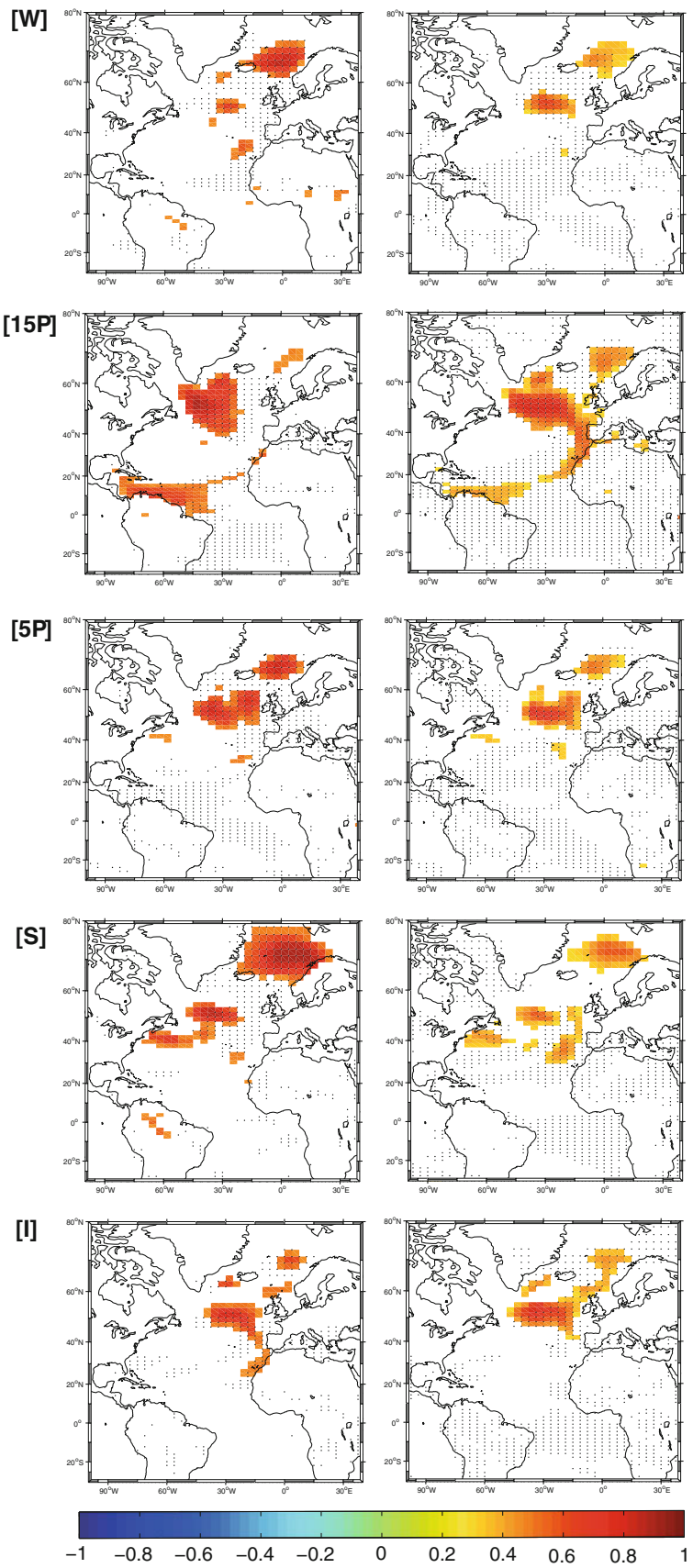

Fig. 11 Surface temperature. Colours represent EC computed as in CS03 for each starting date and years 1–10 (*left panels*), 1–20 (*right panels*) of each ensemble experiment. Areas where the correlation is not statistically significant at the 90 % level are shown in white. *Dots* represent grid points where the ES is statistically significantly smaller than the control RMSE at the 95 % level

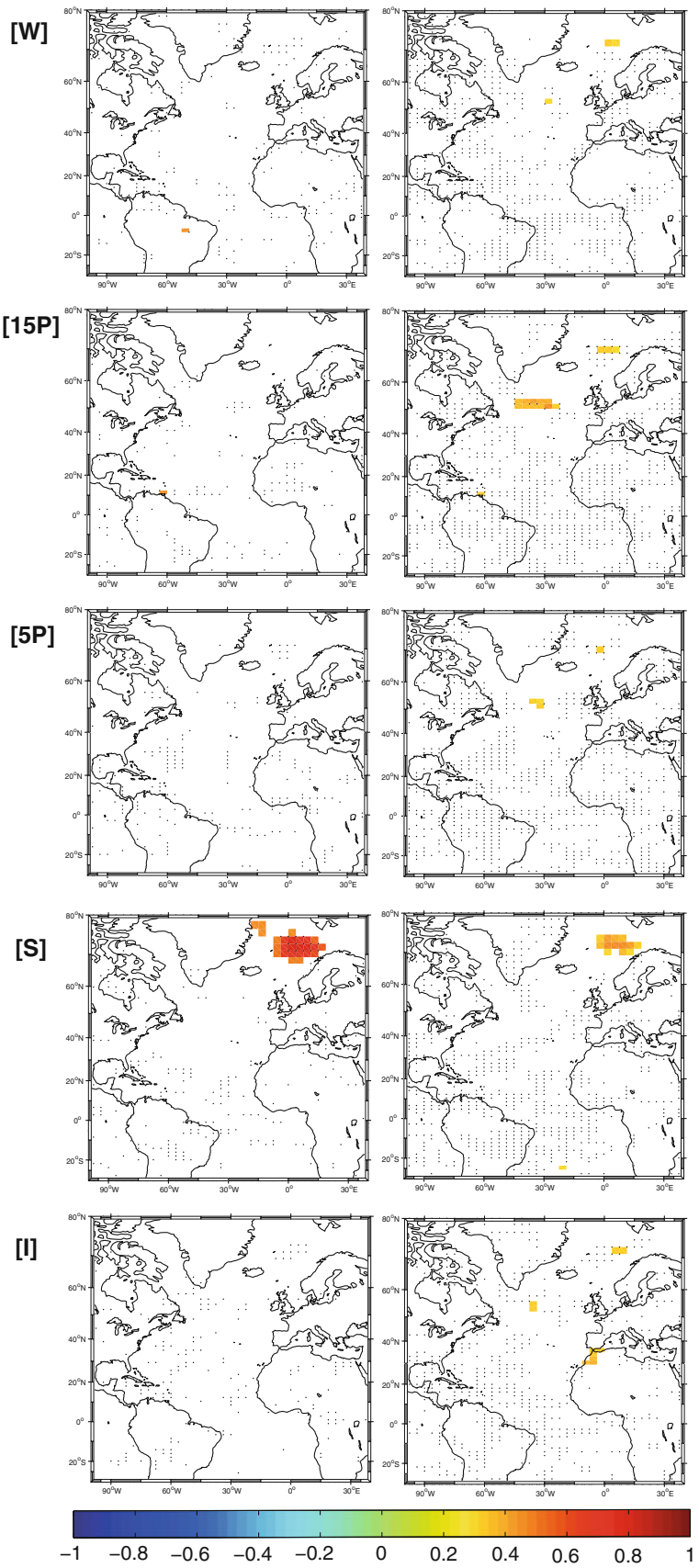

Fig. 12 As Fig. 11 for precipitation

References

Aumont O, Bopp L (2006) Globalizing results from ocean in situ iron fertilization studies. Global Biogeochem Cycles 20, GB2017, doi:10.1029/2005GB002591

Boer GJ (2001) Decadal potential predictability in coupled models. CLIVAR exchanges, vol. 129, no. 3

Boer GJ (2004) Long-timescale potential predictability in an ensemble of coupled climate models. Clim Dyn 23:29–44

Boer GJ (2011) Decadal potential predictability of twenty-first century climate. Clim Dyn 36:1119–1133. doi:10.1007/s00382-010-0747-9

Boer GJ, Lambert SJ (2008) Multi-model decadal potential predictability of precipitations and temperature. Geophys Res Lett 35:L05706. doi:10.1029/2008GL033234

Branstator G, Teng H, Meehl G, Kimoto M, Knight J, Latif M, Rosati A (2012) Systematic estimates of initial value decadal predictability for six AOGCMs. J Climate 25:1827–1846

Brearley JA, Pickart RS, Valdimarsson H, Jonsson S, Schmitt RW, Haine TWN (2012) The East Greenland boundary current system south of Denmark Strait. Deep Sea Res Part I 63:1–19. doi:10.1016/j.dsr.2012.01.001

Bretherton CS et al (1999) The effective number of spatial degrees of freedom of a time-varying field. J Clim 12:1990–2009

Chen D, Cane MA, Kaplan A, Zebiak SE, Huang D (2004) Predictability of El Niño over the past 148 years. Nature 428:733–736

Chiang JCH, Cheng W, Bitz CM (2008) Fast teleconnections to the tropical Atlantic sector from the Atlantic thermohaline adjustment. Geophys Res Lett 35:L07704. doi:10.1029/2008GL033292

Collins M (2002) Climate predictability on interannual to decadal timescales: the initial value problem. Clim Dyn 19:671–692

Collins M, Sinha B (2003) Predictable Decadal Variations in the Thermohaline Circulation and Climate. Geophys Res Lett 30(6). doi:10.1029/2002GLO16504

Collins M, Botzet M, Carril A, Drange H, Jouzeau A, Latif M, Otteraa OH, Pohlmann H, Sorteberg A, Sutton R, Terray L (2006) Interannual to decadal climate predictability in the North Atlantic: a multi-model ensemble study. J Clim 19:1195–1203

Curry R, Dickson B, Yashayaev I (2003) A change in the freshwater balance of the Atlantic Ocean over the past four decades. Nature 426:826–829

Delworth TL, Mann ME (2000) Observed and simulated multi-decadal variability in the Northern Hemisphere. Clim Dyn 16:661–676

Dufresne JL et al. (2012) climate change projections using the ISPL-CM5 earth system model: from CMIP3 to CMIP5. Clim Dyn (in revision)

Dunstone NJ, Smith DM, Eade R (2011) Multi-year predictability of the tropical Atlantic atmosphere driven by the high latitude North Atlantic Ocean. Geophys Res Lett 38(14):L14701. doi:10.1029/2011GL047949

Escudier R, Mignot J, Swingedouw D (2012) A coupled ocean-sea ice-atmosphere decadal variability mode in the North Atlantic in an AOGCM. Clim Dyn (accepted)

Fichefet T, Maqueda MAM (1997) Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. J Geophys Res 102:2609–2612

Fisher RA (1921) On the 'probable error' of a coefficient of correlation deduced from a small sample. Metron 1(4):3–32

Folland CK, Parker DE, Palmer TN (1986) Sahel rainfall and worldwide sea temperatures, 1901–85. Nature 320:602–607

Frankignoul C (1985) Sea surface temperature anomalies, planetary waves, and air-sea feedback in the middle latitudes. Rev Geo phys 23:357–390

Frankignoul C, Hasselmann K (1977) Stochastic climate models. Part II: application to sea-surface temperature variability and thermocline variability. Tellus 29:284–305

Frankignoul C, Kestenare E, Mignot J (2002) The surface heat flux feedback. Part II: direct and indirect estimates in the ECHAM4/OPA8 coupled GCM. Clim Dyn 19:649–655

Gastineau G, D'Andrea F, Frankignoul C (2012) Atmospheric response to the North Atlantic Ocean variability on seasonal to decadal timescales in IPSL-CM5. Clim dyn (accepted)

Griffies SM, Bryan K (1997) Predictability of North Atlantic multidecadal climate variability. Science 275(5297):181–184

Hawkins E, Sutton R (2008) Potential predictability of rapid changes in the Atlantic meridional overturning circulation. Geophys Res Lett 35:L11603. doi:10.1029/2008GL034059

Hawkins E, Robson J, Sutton R, Smith D, Keenlyside N (2011) Evaluating the potential for statistical decadal predictions of sea surface temperatures with a perfect model approach. Clim Dyn. doi:10.1007/s00382-011-1023-3

Hermanson L, Sutton RT (2009) Climate predictability in the second year. Philos Trans R Soc A 367:913–916

Hourdin F, Foujols MA, Codron F, Guemas V, Dufresne JL, Bony S, Denvil S, Guez L, Lott F, Ghattas J, Braconnot P, Marti O, Meurdesoif Y, Bopp L (2012) Climate and sensitivity of the IPSL-CM5A coupled model: impact of the LMDZ atmospheric grid configuration. Clim Dyn (in revision)

Hurrell J, Meehl G, Bader D, Delworth T, Kirtman B, Wielicki B (2009) A unified modeling approach to climate system prediction. Bull Am Meteor Soc 90:1819–1832. doi:10.1175/2009BAMS2752.1

Kerr RA (2000) A north Atlantic climate pacemaker for the centuries. 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:L20708. doi:10.1029/2005GL024233

Knight J, Folland CK, Scaife AA (2006) Climate impacts of the Atlantic multidecadal oscillation. Geophys Res Lett 33:L17706. doi:10.1029/2006GL026242

Krinner GN, et al. (2005) A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. Glob Biogeochem Cyc 19. doi:10.1029/2003GB002199

Kushnir Y (1994) Interdecadal variations in north Atlantic sea surface temperature and associated atmospheric conditions. J Clim 7(1):142–157

Latif M, Roeckner E, Botzet M, Esch M, Haak H, Hagemann S, Jungclaus J, Legutke S, Marsland S, Mikolajewicz U (2004) Reconstructing, monitoring, and predicting multidecadal-scale changes in the North Atlantic thermohaline circulation with sea surface temperature. J Clim 17:1605–1614

Livezey RE, Chen WY (1983) Statistical field significance and its determination by Monte Carlo techniques. Mon Weather Rev 111:46–59

Madden RA (1976) Estimates of the natural variability of time-averaged sea-level pressure. Mon Weather Rev 104:942–952

Madec G (2008) NEMO ocean engine, note du Pole de modelisation, Institut Pierre-Simon Laplace (IPSL)

Meehl GA, Goddard L, Murphy J, Stouffer RJ, Boer G, Danabasoglu G, Dixon K, Giorgetta MA, Greene A, Hawkins E, Hegerl G, Karoly D, Keenlyside N, Kimoto M, Kirtman B, Navarra A, Pulwarty R, Smith D, Stammer D, Stockdale T (2009) Decadal prediction: can it be skillful? Bull Am Meteor Soc 90:1467–1485

Msadek R, Frankignoul C (2009) Atlantic multidecadal oceanic variability and its influence on the atmosphere in a climate model. Clim Dyn. doi:10.1007/s00382-008-0452-0

Msadek R, Dixon KW, Delworth TL, Hurlin W (2010) Assessing the predictability of the Atlantic meridional overturning circulation and associated fingerprints. Geophys Res Lett 37:L19608. doi:10.1029/2010GL044517

National Research Council; Committee on Assessment of Intraseasonal to Interannual Climate Predition and Predictability (2010) Assessment of Intraseasonal to Interannual Climate Prediction and Predictability. National Academy of Science, Washington, p 192

Newman M, Sardeshmukh PD, Winkler CR, Whitaker JS (2003) A study of subseasonal predictability. Mon Weather Rev 131:1715–1732

Persechino A, Marsh R, Sinha B, Megann A, Blaker A, New A (2012) Decadal-timescale changes of the Atlantic overturning circulation and climate in a coupled climate model with a hybrid-coordinate ocean component. Clim Dyn 39(3):1021–1042

Pohlmann H, Botzet M, Latif M, Roesch A, Wild M, Tschuck P (2004) Estimating the decadal predictability of a coupled AOGCM. J Clim 17:4463–4472

Quenouille MH (1952) Associated measurements. Butterworths, p 242

Rowell DP (1998) Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. J Clim 11:109–120

Rowell DP, Folland CK, Maskell K, Ward MN (1995) Variability of summer rainfall over tropical North-Africa (1906–92) observations and modeling. Q J R Meteorol Soc 121:669–704

Smith DM, Eade R, Dunstone NJ, Fereday D, Murphy JM, Pohlmann H, Scaife AA (2010) Skilful multi-year predictions of Atlantic hurricane frequency. Nat Geosci. doi:10.1038/NGEO1004

Solomon A, et al. (2011) Distinguishing the roles of natural and anthropogenically forced decadal climate variability: implications for prediction US CLIVAR Decadal Predictability Working Group. Bull Am Meteor Soc 92(2). doi:10.1175/2010BAMS29621

Sutton RT, Hodson DLR (2005) Atlantic Ocean forcing of North American and European summer climate. Science 309:115–118

Swingedouw D, Braconnot P, Delecluse P, Guilyardi E, Marti O (2007) The impact of global freshwater forcing on the thermohaline circulation: adjustment of North Atlantic convection sites in a CGCM. Clim Dyn 28:291–305

Swingedouw D, Mignot J, Braconnot P, Mosquet E, Kageyama M, Alkama R (2009) Impact of freshwater release in the North Atlantic under different climate conditions in an OAGCM. J Clim 22:6377–6403

Swingedouw D, Mignot J, Labetoule S, Guilyardi E, Madec G (2012) Initialisation and predictability of the AMOC over the last 50 years in a climate model. Clim Dyn (in revision)

Taylor KE, Stouffer RJ, Meehl GE (2009) A summary of the CMIP5 experiment design. Lawrence Livermore National Laboratory Rep, p 32

Teng H, Branstator G, Meehl GA (2011) Predictability of the Atlantic overturning circulation and associated surface patterns in two CCSM3 climate change ensemble experiments. J Clim 24:6054–6076

Ting M, Kushnir Y, Seager R, LIC (2009) Forced and internal twentieth century SST trends in the North Atlantic. J Climate 22:1469–1481

Valcke S (2006) OASIS3 user guide (prism_2-5), technical report TR/CMGC/06/73, PRISM Report No 2. CERFACS, Toulouse, p 60

Vellinga M, Wu P (2004) Low-latitude freshwater influence on centennial variability of the Atlantic thermohaline circulation. J Clim 17:4498–4511

Xie SP, Okumura Y, Miyama T, Timmermann A (2008) Influences of Atlantic climate change on the tropical Pacific via the Central American Isthmus. J Clim 21:3914–3928

Zhang R, Delworth TL (2006) Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. Geophys Res Lett 33:L17712