TECHNICAL MEMORANDUM

633

Decadal climate prediction with the ECMWF coupled forecast system: Impact of ocean observations

F.J. Doblas-Reyes[1]*, M.A. Balmaseda, A. Weisheimer, T.N. Palmer

Research Department

(1) ICREA Research Professor at the Institut Català de Ciències del Clima (IC3), Doctor Trueta 203, 08005 Barcelona, Spain

* f.doblas-reyes@ic3.cat

October 2010

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

Abstract

The initialization of the ocean-atmosphere coupled models is a fundamental problem in dynamical decadal forecasting. Three ten-year ensemble decadal forecast experiments have been performed with the ECMWF coupled forecast system using an initialization strategy common in seasonal forecasting where both the ocean and the atmosphere use the best estimate that can be obtained in a re-analysis. One of the experiments initializes the ocean model using data from an ocean-only simulation forced with data from an atmospheric re-analysis. The other two take the ocean initial condition from a similar ocean-only run that in addition assimilates subsurface observations. This is the first time that such a set of experiments has been carried out. In all the experiments the coupled model drifts with forecast time from the realistic initial conditions towards the model climate. An analysis of the small drift differences found between the experiments suggests that in the tropics they are due to differences in latent heat and outgoing top of the atmosphere radiation in the first three forecast years. Instead, in the extratropics the initial-condition differences between the experiments persist for the whole duration of the simulations, associated to a persistent difference in latent heat flux between the experiments initialized with and without ocean data assimilation. In spite of the important drift, the three sets of decadal predictions show that the system is able to predict the interannual variability of the observed global and regional mean air temperature anomalies up to several years in the future. No significant benefit of the assimilation of ocean observations is found over the extratropics in terms of forecast quality, although there are some indications of the negative impact of the incorrect assimilation of the XBT profiles. The results illustrate the relevance of reducing the important model drift, increasing the sample size typically adopted in decadal forecasting and reducing the uncertainty in the reference ocean re-analyses. These three factors contribute to prevent these experiments from extracting more definite conclusions for accurately assessing the benefit of ocean data assimilation in terms of the forecast quality of decadal predictions.

1 Introduction

Climate-change projections and near-term climate prediction (also known as decadal prediction) attempt to satisfy a growing demand for climate information for this century. It is well established that, based on knowledge of the initial conditions, important aspects of regional climate are predictable up to a year ahead. Predictability at this time scale is primarily, though not solely, associated with the El Niño Southern Oscillation (ENSO). While climate forecasting is currently addressing the problem of climate prediction up to one year (e.g., Doblas-Reyes et al., 2009; Weisheimer et al., 2009), decadal prediction focuses on time scales of several years to a few decades (e.g., Smith et al., 2007).

There have been attempts to predict interannual climate variations using empirical models that take into account changes in boundary conditions, i.e. atmospheric composition and solar irradiance (e.g. Lean and Rind, 2009). As an alternative, dynamical decadal prediction explores the ability of the type of climate model employed in the Intergovernmental Panel for Climate Change (IPCC) assessments to predict regional climate changes in the near future by exploiting both initial-condition information and changes in boundary conditions. This approach aims to take advantage of the predictability of natural climate variability to make predictions. However, a critical question is to understand how far ahead the mean climate is predictable at regional spatial scales with some useful level of skill. Related to this is the question of the extent to which a better knowledge of the initial conditions of the climate system contributes to the quality of these forecasts. Beyond the first year of the forecast, little is known about the contribution of the initial conditions to the decadal prediction skill, how it changes between variables (temperature, precipitation, low-level winds) and between regions, whether it depends on the time of the year when the forecast is initialized, and what quality is required from ocean, atmosphere, soil and sea-ice initial conditions to find a positive impact on forecast quality.

The relative importance of the initial conditions in climate prediction is supposed to vary with the time scale, but has been assumed to be a continuous function that decreases with forecast time, becoming negligible after several decades (Hawkins and Sutton, 2009b). Ocean initial conditions are more relevant than variations in atmospheric composition in seasonal forecasting (Doblas-Reyes et al., 2006), except perhaps after an explosive volcanic eruption, while atmospheric composition has primary importance after several decades (Hawkins and Sutton, 2009b). For the time scales ranging between a few seasons to a couple of decades, previous work (Smith et al., 2007; Keenlyside et al., 2008; Pohlmann et al., 2009) has shown evidence that the initial state of the ocean can influence climate forecasts a decade or more ahead.

The strategy followed for the initialization of decadal predictions has so far been quite different from that used in seasonal forecasting. For instance, Smith et al. (2007) used the so-called anomaly initialization method, where ocean observations are assimilated in the form of anomalies into the coupled model taking into account the error covariance of the coupled model. In Keenlyside et al. (2008) only observed sea surface temperature (SST) anomaly information was used to initialize a coupled system. In seasonal forecasting, however, it is common practice the separate initialization of the ocean and the atmosphere, data assimilation being used to bring the state of each component of the coupled model close to the observed state. Balmaseda and Anderson (2009) showed that for seasonal forecasting this strategy works better than an approach equivalent to that used in Keenlyside et al. (2008).

In this study, we use the ECMWF coupled system to investigate how the initialization strategy used in seasonal forecasting behaves in decadal forecasting, and how it compares with an initialization strategy that uses a simpler estimate of the ocean state. The criteria for the assessment is the extent to which atmospheric and ocean variables are skillfully predicted in the forecast range from one to ten years, i.e. beyond the generally accepted limit of ENSO-related predictability. Problems related with sample and ensemble sizes in the context of the protocol set up in the next Coupled Model Intercomparison Experiment, known as CMIP5, for initialized dynamical decadal predictions are also discussed (http://www.clivar.org/organization/wgcm/references/Taylor_CMIP5.pdf).

A brief summary of the experiment follows in Section 2. The most relevant characteristics in terms of model drift and forecast quality results are given in Sections 3 and 4. The main conclusions are summarized in Section 5.

2 Description of the experiment

2.1 Experimental set-up

To address the key uncertainties at the source of decadal forecast error, such as uncertainties in the initial conditions and in model formulation (Palmer, 2000; Anderson et al., 2009), ensemble methods have been proposed. They involve not only running a single model several times with slightly different initial conditions, but also utilizing multi-model or perturbed-parameter approaches. In this paper, sets of ensemble re-forecasts have been carried out with the IFS/HOPE coupled system. The forecast system (Anderson et al., 2007) used the atmospheric IFS cycle 35r3 (Bechtold et al., 2008) with a horizontal truncation of T_L159 and 62 vertical levels extending up to 5 hPa. The ocean model has a horizontal resolution of 1°, with an equatorial refinement of 0.3°, and 29 levels in the vertical. The coupler OASIS2 is used to interpolate the fields exchanged once per day between the ocean and atmospheric grids. IFS uses a climatological annual cycle of five types of aerosol (sea salt, desert dust, organic matter, black carbon). The system includes the interannual evolution of global mean annual greenhouse trace gases (CO_2, CH_4, N_2O and CFCs) and anthropogenic aerosols, as well as interannual variations of total solar irradiance. No relaxation or flux correction was active during the forecast.

The re-forecasts were started once every five years over the period 1960 to 2005, i.e. in 1960, 1965, and so on. The atmosphere and land surface initialization was from the ERA-40 reanalysis (Uppala et al., 2005) for all start dates but for 2005, for which the operational ECMWF analyses were used. Each simulation started at 00 GMT on the 1st of November of each year and run for 120 months. This experimental setup is based on the decadal re-forecast experiment of the ENSEMBLES project (http://www.ecmwf.int/research/EU_projects/ENSEMBLES/exp_setup/stream2.html).

The baseline experiment uses ocean initial conditions from the ORA-S3 ocean re-analysis (Balmaseda et al., 2008), from which only one ensemble member has been used. All available observations of temperature, salinity and altimetric sea level anomalies are used in this re-analysis. The atmospheric fluxes are from the ERA40 reanalysis for the period January 1959 to June 2002 and ECMWF operational analysis thereafter. The SST is strongly relaxed to analyzed daily SST maps from the OIv2 SST (Reynolds et al., 2002) product from 1982 onwards. This experiment will be referred to henceforth as Assim. An alternative initialization consists in using data from an ocean simulation forced with the ECMWF atmospheric fluxes, but with no ocean data assimilated. This second experiment will be named NoOcObs. As the ocean model attractor inevitably differs from the attractor of the actual climate system, the lack of observed ocean data in NoOcObs will produce ocean initial states within the ocean model attractor, unlike in the case of the Assim experiment. Hence, in absence of error in the atmospheric model, the NoOcObs method would produce balanced ocean initial conditions. A third experiment has been performed using a re-analysis similar to ORA-S3, but where corrected XBT (eXpendable BathyThermograph) profiles according to Wijffels et al. (2008) and Ishii and Kimoto (2009) have been assimilated. The third experiment will be referred to as XBT-C.

2.2 Computation of the anomalies

Various measures of forecast quality have been used to assess the differences between the experiments. The scores include the anomaly correlation coefficient (ACC) and root mean square error (RMSE) of the ensemble mean. All forecast quality measures have used ERA40 before 2002 and ERA-Interim afterwards as the atmospheric reference dataset, except for precipitation, for which GPCP (Adler et al., 2003) was taken as the reference. For the ocean variables, the re-analysis performed with the corrected XBT profiles (ORA-XC henceforth) has been used over the period 1960-2005.

Every forecast quality measure has been computed taking into account the systematic error of the forecast systems. Forecast anomalies have been estimated in cross-validation mode by removing the mean model climate for the specific forecast period using the re-forecasts for which there are reference data available, as it is commonly done in seasonal forecasting. For instance, to obtain the anomalies of the average 6-10 year forecast period from the re-forecast initialized in November 1970, the model climate is estimated by averaging the data for the 6-10 year forecast period from all the re-forecasts for which there is reference data, except the re-forecast started in November 1970. This implies that data from the 1960, 1965, 1975, 1980, 1985, 1990 and 1995 re-forecasts (seven start dates) are used, because no reference data for the period 2005-2010 and 2010-2015 (i.e., the verifying dates of the re-forecasts with start dates in 2000 and 2005) were available. The anomalies for the reference dataset are estimated for the same calendar period. The upper panels of Figure 1 illustrate the process of a-posteriori removal of the drift, an estimate of which appears in the bottom row, for global-mean near-surface air temperature. The raw re-forecast values appear in the left panel, while the anomalies resulting after the drift estimate has been removed appear in the right panel. The reader should be aware that this linear method assumes that there is no relationship between the model drift and the anomalies.

This method of computing the anomalies is different from the approach that would be adopted in an operational context, where the anomalies would be computed using only past information. However, the shortness of the sample, with just ten re-forecasts available, prevents the authors from using a more robust computation of the anomalies. In other decadal prediction experiments prediction anomalies have been estimated using a model climate estimate from a set of simulations of the Twentieth Century climate that do not assimilate observations (e.g., Smith et al., 2007; Doblas-Reyes et al., 2010). This is not possible here because 1) there is no Twentieth Century control simulation available for our forecast system, and 2) even if a control simulation was available, as the coupled model is initialized from realistic initial conditions the model climate in the re-forecasts depends on the forecast time due to the unavoidable drift, making a unique model climate estimate from a long simulation inappropriate.

3 Mean state and model drift

Model inadequacy causes forecasts to drift away from the observed climate towards an imperfect model climate. Previous publications on decadal forecasting (Smith et al., 2007; Keenlyside et al., 2008; Hawkins and Sutton, 2009a) rarely illustrate the differences between the model and reference climates. Sometimes model drift is not described in detail because the initialization of the re-forecasts is carried out by assimilating observed anomalies into the model climate (Smith et al., 2007; Keenlyside et al., 2008), a method that is expected to reduce the model drift. Other approaches rely on the idea that the drift is small enough to not destroy the initial-condition information (Pohlmann et al., 2009). Here we consider that forecast drift is an important feature of the decadal forecasting problem when a realistic initialization is adopted and, hence, worth discussing.

Figure 1 shows the global-mean near-surface air temperature. The model and observed temperatures deviate from one another quickly in every re-forecast (Fig. 1a) and, regardless of the initialization dataset, the mean error reaches around 1 K after four years for global-mean temperature (Fig. 1c), while global SST error is around -0.4 K after a similar forecast time (not shown). The evolution of the forecast drift in the first few years is slightly different for the Assim/XBT-C and NoOcObs experiments, showing a sharper decline in the Assim and XBT-C experiments during the first two years, a forecast period after which the differences in the drift due to the initial conditions almost completely vanish (Fig. 1c). In terms of global-mean land temperatures, Assim and XBT-C are slightly warmer than NoOcObs, although both experiments have a cold bias (not shown). A cold bias is also found for the global-mean SSTs, so both land and ocean contribute to the cold global-mean bias. The three experiments reach a similar equilibrium state in terms of global-mean near-surface air temperatures over land and SSTs after the first five years of forecast.

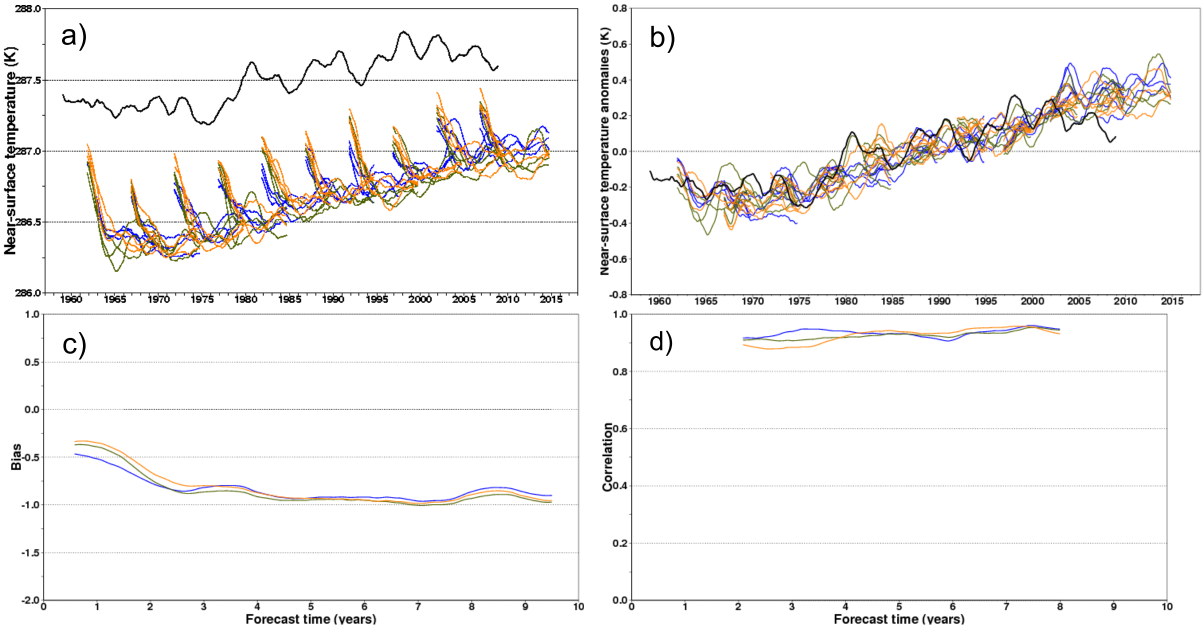

Figure 1: (a) Global-mean near-surface air temperature (K) for the ten three-member ensemble re-forecasts of the XBT-C (green), NoOcObs (blue) and Assim (orange) experiments. (b) Anomalies with respect to the corresponding climate over the period 1960-2005 are shown on the top right panel. Data from ERA40/ERA-Interim are shown in black. All time series have been smoothed out with a 24-month centred moving average that removes data for the first and last years of each time series. The bottom row shows the drift (K, c) and the ensemble-mean correlation (d) of the global-mean near-surface air temperature for the XBT-C (green), NoOcObs (blue) and Assim (orange) experiments. The drift and the correlation have been computed using ERA40/ERA-Interim and three-member ensemble re-forecasts for the period 1960-1995. A 12-month moving average has been applied to the drift estimates to illustrate the fast growth rate of the drift, while the correlation has been computed with a moving window of four-year averaged anomalies to retain the interannual variability that is beyond the ENSO typical frequency.

The spatial distribution of the drift is very similar in the three experiments. As an example, Figure 2 displays the mean drift of the boreal winter (December to February) near-surface air temperature for the forecast period two to five years. Re-forecasts with start dates within the period 1960-2000, which is the time interval with reference data available, have been used for the estimates. The experiments have in general air temperatures cooler than the reference, in agreement with Figure 1. This is particularly obvious over the tropical regions, where the cooling is slightly alleviated in Assim. This pattern is also found in summer, except for a strong warm bias over the equatorial eastern Atlantic, which is typical of both IPCC and seasonal forecast models (C. Caminade, personal communication). The differences in the mean drift between the three experiments are much smaller than the drift itself, which suggests that the type of initialization has a very small impact on the drift reduction. The drift similarity between the experiments is even higher for longer forecast periods. A broad similarity has also been observed for estimates of the interannual standard deviation of both experiments (not shown).

Figure 2 also shows the mean drift of the winter (December to February) ocean temperature averaged over the top 300 metres (a proxy for the upper ocean heat content) for the forecast period two to five years. The mean error has been computed with respect to the ORA-XC ocean re-analysis. The drift in the ocean temperature bears some similarity with the pattern found for near-surface air temperature, although there are some differences: the tropical cooling is not as widespread and the western

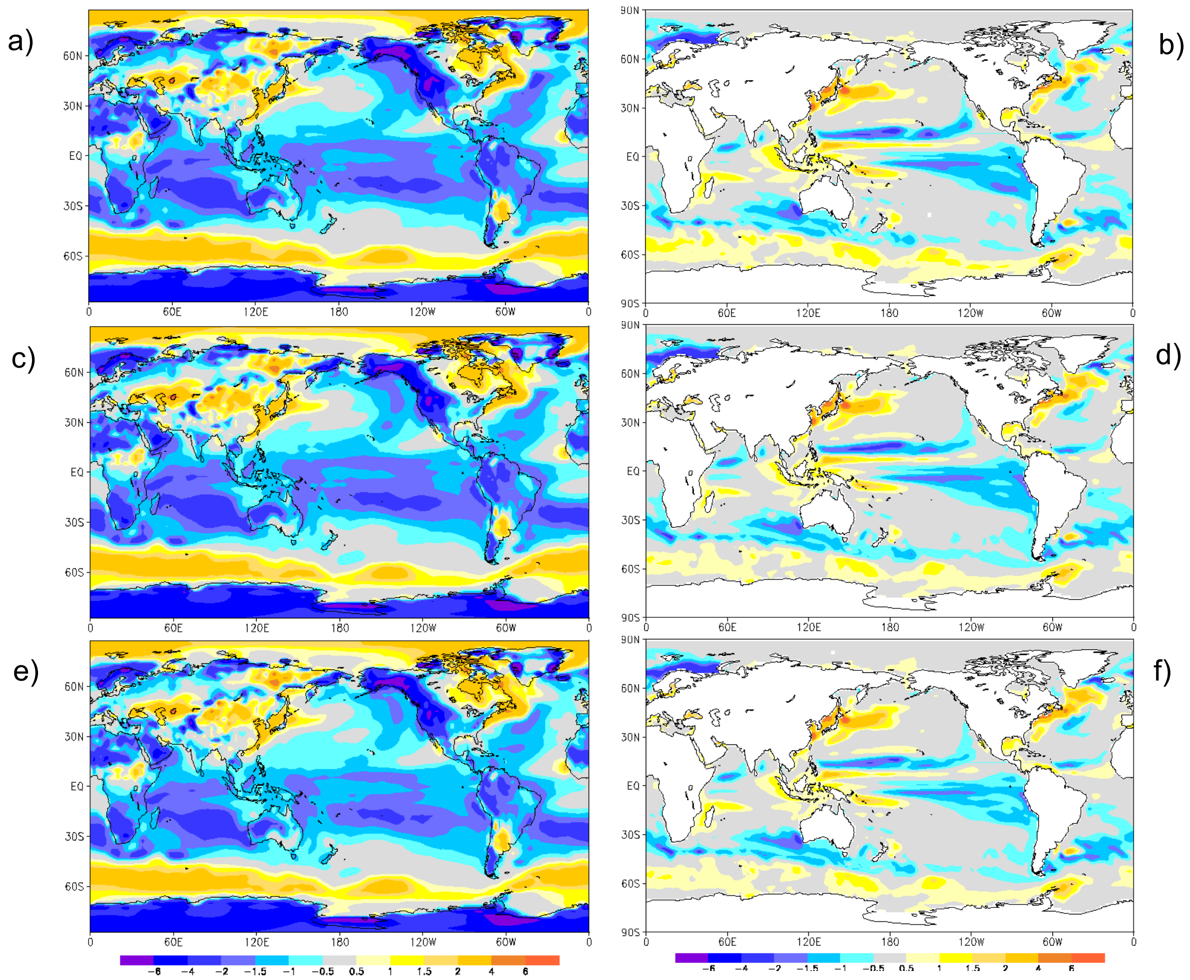

Figure 2: Winter (December to February) near-surface air temperature (left column, K) and ocean temperature averaged over the top 300 metres (right column, K) systematic error over the forecast period two to five years of the NoOcObs (top), XBT-C (middle) and Assim (bottom) experiments. The systematic error has been estimated with respect to ERA40/ERA-Interim (left) and ORA-XC (right). Three-member ensemble re-forecasts for the period 1960-2000 have been used.

equatorial Pacific is actually warmer than in the re-analysis resembling a La Niña pattern. There are also differences in the western boundary currents, which are warmer in the coupled model than in the verifying re-analysis with a pattern typical of the low-resolution systems used for climate-change projections (van Oldenboth et al., 2009). Overall, the large-scale patterns of drift in the upper ocean heat content suggest that there are errors in the ocean circulation and not just in the surface heat fluxes. As with the near-surface air temperature, the three experiments depict a similar degree of cooling with respect to the ocean re-analysis, but with some local differences. Small differences in the drift between the experiments (at least one order of magnitude smaller than the drift itself) can also be found for ocean variables.

To better illustrate the differences between the experiments in the upper ocean heat content drift, Figure 3 shows the drift of the global-mean and Southern Hemisphere upper ocean heat content for

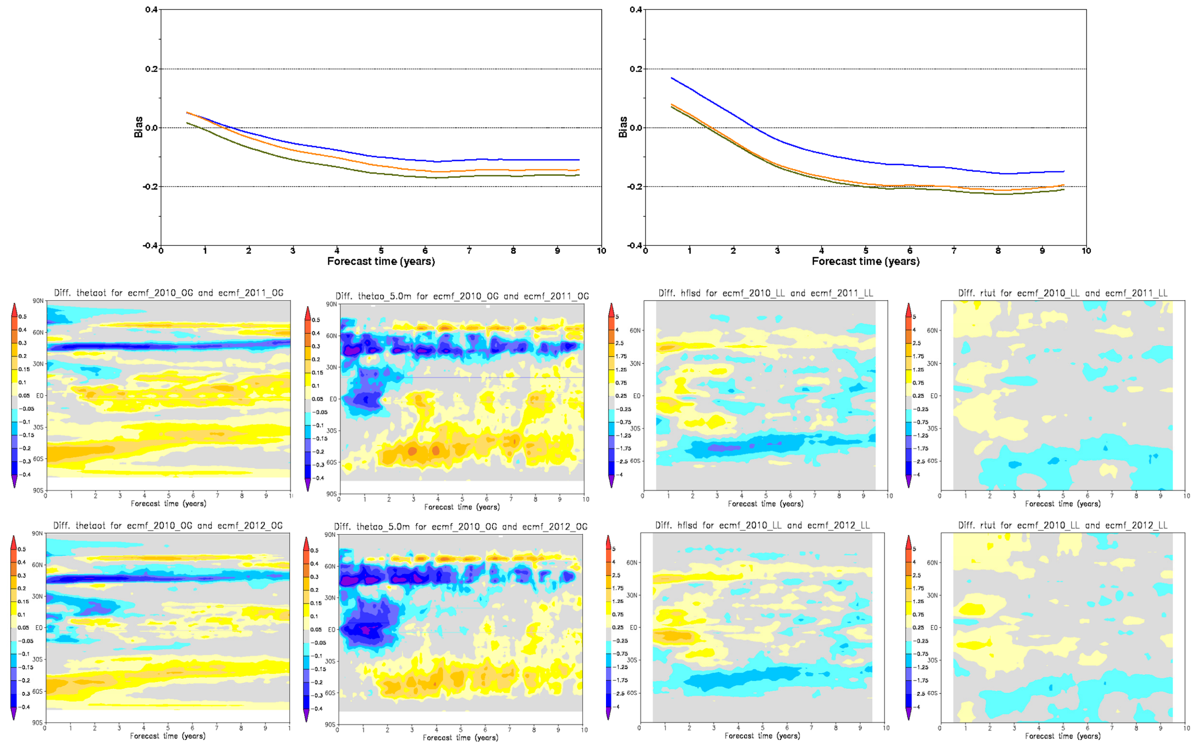

Figure 3 (Top row) Drift (K) with respect to ORA-XC of the global-mean (left) and Southern Hemisphere (south of 20°S, right) ocean temperature averaged over the top 300 metres for the ten three-member ensemble re-forecasts of the XBT-C (green), NoOcObs (blue) and Assim (orange) experiments. The time series have been smoothed out with a centred annual running mean that removes data for the first and last six months of each time series. (From the first left column to the right) Difference in zonal-mean climate of the NoOcObs and XBT-C (top row) and NoOcObs and Assim (bottom row) experiments as a function of latitude and forecast time for the ocean temperature averaged over the top 300 metres (K, first column), the SST (K, second column), the surface latent heat flux (W/m^2, third column) and top net radiation (W/m^2, fourth column). The fluxes have been smoothed out with a centred annual running mean which removes data for the first and last six months. All estimates have been computed using three-member ensemble re-forecasts for the period 1960-1995. Downward atmospheric fluxes are positive.

Assim and NoOcObs. The drift in the global-mean upper ocean heat content is cold after the first forecast year in the three experiments. The XBT-C drift is cooler than for the other two experiments. The drift diverges between Assim and NoOcObs as the forecast time increases. An examination of the ocean initial conditions showed that the upper ocean heat content of the Assim initial conditions is higher than in the NoOcObs initial conditions over the tropical band and, before 1980, also over the Northern Hemisphere. This suggests that the Assim experiment looses heat from the ocean faster on average than the other two experiments. The largest drift occurs over the Southern Hemisphere, where the upper ocean is warmer than the re-analysis in the first year, especially for NoOcObs. When the forecast time increases is up to 0.2 K cooler than the re-analysis for Assim and XBT-C, the difference with NoOcObs (which is warmer) being rather constant with forecast time.

These results point at a strong latitudinal dependence of the drift differences. Figure 3 shows that the differences between NoOcObs and XBT-C differences in the tropical band (30ºS-30ºN) are small and change sign after 2-to-4 years, while poleward of 30º the differences in the initial conditions are much more persistent and remain during the whole duration of the forecasts (although with a slight latitudinal sway). The differences between NoOcObs and Assim grow with time in the tropical band and are very similar to what is found for XBT-C outside of the tropics. For the differences of both XBT-C and Assim with NoOcObs, a northward shift with forecast time of the warm differences at 60ºS is found. This is the result of the initial warm differences shifting away from the Circumpolar Current and concentrating in the South Atlantic (not shown). At the same time, the initial cold difference at around 45°N, which is due to the different location of the Gulf Stream and Kuroshio in the re-analysis with and without ocean data assimilation, is compensated by a warm difference at around 65°N, coming from a warm core developing both sides of Iceland. It is important to emphasize the fact that these drift differences show dynamical properties as animations show that they propagate within the extratropics both zonally and meridionally. In contrast with the extratropics, the tropical band reaches a stationary state after a few years. Some of the extratropical differences reach the tropical area, and there, through westward wave propagation and coastal wave reflection, eventually reach the equator.

The differences in SST drift shown in Figure 3 are broadly consistent with the differences in ocean heat content, albeit with broader latitudinal structure and a smaller degree of latitudinal shift with forecast time. The SST drift differences have an impact on the atmosphere mainly via changes in the surface latent heat flux. The surface latent heat flux acts as a negative feedback canceling out the differences in SST (Fig. 3). Differences in surface shortwave radiation also play a role (not shown). As a consequence, some changes of the net flux at the top of the atmosphere (Fig. 3) have been found. In particular, a reduced outgoing global-mean flux (~0.5 W/m^2) is found in the tropical band for NoOcObs with respect to Assim and XBT-C during the first three years of the forecast. This could explain the slower decrease with forecast time of global-mean upper ocean heat content found for NoOcObs with respect to Assim and XBT-C. The top of the atmosphere energy loss in the initial four years of the Assim and XBT-C experiments with respect to NoOcObs takes place mostly over the west Pacific and the South Pacific Convergence Zone (SPCZ) and is related to an increased surface latent heat flux into the atmosphere. After the first two years both Assim and XBT-C have a higher outgoing radiation flux south of 45ºS (Fig. 3) that agrees with the decrease of the differences in ocean heat content as forecast time increases over the area.

A better example of how much the ocean mean state of both experiments can differ is illustrated in Figure 4, which shows the zonally integrated meridional velocity across the Atlantic basin at 36ºN, a proxy for the intensity of the Atlantic meridional overturning circulation (AMOC). The AMOC in the NoOcObs initial conditions is shallower and weaker (~20 Sv) than in the verifying analysis XBT-C (~24 Sv), in agreement with Balmaseda et al. (2007). The AMOC in the ocean re-analyses with ocean data assimilation also show a stronger interannual variability. The Assim and XBT-C simulations experience a transition after the first two forecast years towards a shallower (by ~500 m) AMOC cell, weakening the northward branch and strengthening the southward one. Such a drift is slower in NoOcObs, for which the AMOC cell is already shallower than in the ORA-S3 analysis at the beginning of the simulations. As for the atmospheric variables, the three experiments tend towards a similar equilibrium state. As a consequence of the shallowing of the meridional overturning cell and the reduction of its vertical gradient, the AMOC intensity, estimated as the maximum of the vertically integrated meridional transport across the Atlantic at 36ºN (the latitude where the maximum intensity occurs in the re-analyses), decreases (Fig. 4e) to around 10 Sv. Model results (e.g. Drijfhout et al., 2008; Lozier, 2010) suggest that whereas the interannual variability in the ocean overturning is largely driven by surface winds, variability on decadal and longer timescales (and probably the drift too) is primarily driven by buoyancy fluxes. Contributions to the buoyancy fluxes comprise fresh water forcing by precipitation, evaporation, runoff and sea-ice melting or formation, and thermal forcing by turbulent fluxes (sensible and latent heat), radiative fluxes and the latent heat of fusion associated with formation or melting of sea ice. It is likely that most of these processes, in particular the missing ones such as those related to sea-ice melting and formation or unresolved ocean eddies, are responsible for the drift of the AMOC in this model. The strong drift in the AMOC could explain part of the cooling drift over the North Atlantic and western Europe (Fig. 2), although concluding if the differences in AMOC drift between the experiments may lead to any difference in the upper ocean heat content drift will require additional experimentation.

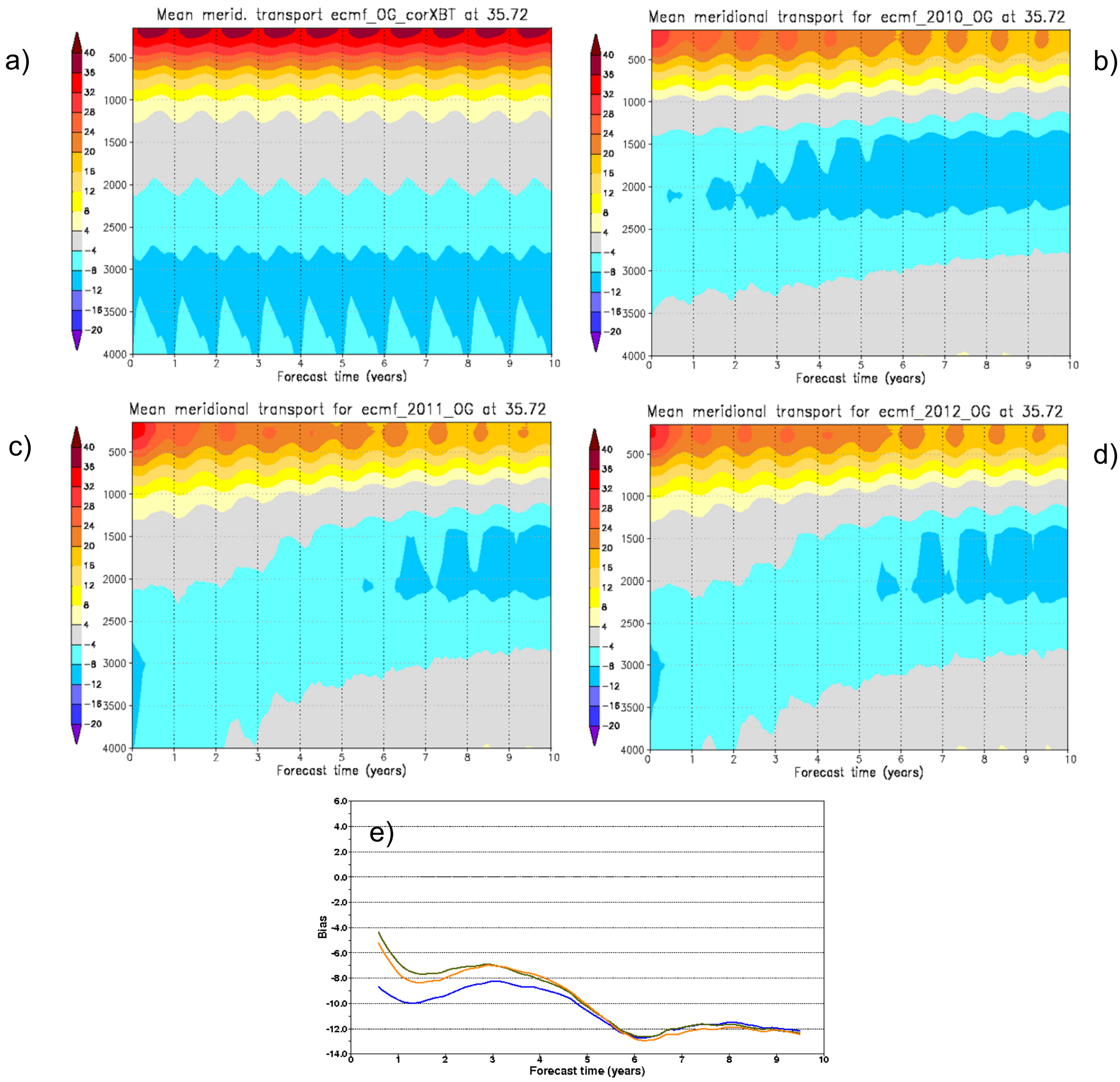

Figure 4 Zonally integrated meridional water velocity (10^3 m^2/s) across the Atlantic basin at 36°N for the ORA-XC ocean re-analysis (a), and the NoOcObs (b), XBT-C (c) and Assim (d) experiments as a function of depth and forecast time. The horizontal axis covers 120 months and represents the mean seasonal cycle of the re-analysis repeated ten times and the drift from the ten re-forecasts. The vertical axis starts at 150 metres. (e) Drift with respect to ORA-XC of the Atlantic meridional overturning circulation for the ten three-member ensemble re-forecasts of the Assim (green), NoOcObs (blue) and XBT-C (orange) experiments. The time series have been smoothed out with a centred annual running mean, which removes data for the first and last six months of each time series. All estimates have been computed using three-member ensemble re-forecasts for the period 1960-1995.

4 Forecast quality

4.1 Atmospheric variables

A subset of the re-forecasts has been used to estimate the forecast quality of the experiments. This is because there is not a complete reference dataset available to compare to the 2000 and 2005 start date simulations beyond 2009. In other words, at the time of writing there is no verification available for the tenth forecast year of the 2000 start date and the forecast period 5-10 of the 2005 start date.

The near-surface air temperature anomalies obtained using the verification data available to compute the model climate are shown in Figure 1b. No substantial differences between the re-forecasts of the three experiments are found at first sight. Every experiment reproduces the upward trend that is especially noticeable from 1975 onwards. The ensemble-mean correlation for four-year mean predictions computed using a centred moving window is shown in Figure 1d. Note that only a reduced range of forecast times is available because a four-year mean allows predictions for forecast periods ranging from months 25 to 96 (i.e. forecast time three to eight years). In spite of the non-negligible drift, the correlation is high and statistically significant with 95% confidence. When comparing the results for both experiments, NoOcObs has slightly higher correlation than Assim and XBT-C for the first four forecast years, the difference decreasing with forecast time. Contrary to the conventional dependence of forecast skill with lead time, in both experiments the correlation increases with forecast time from ~0.8 to 0.95, a feature that when the same calculation is carried out on the predictions described in Smith et al. (2007) and Keenlyside et al. (2008) also shows up (Doblas-Reyes et al., 2010). A possible reason for the correlation increase with forecast time is that the effect of the rising trend is stronger for longer forecast times because those samples contain less data prior to 1975. Unfortunately, the small sample typical of the CMIP5 experiments makes impossible the verification of this hypothesis.

In an attempt to explain the origin of the higher correlation of global-mean temperature in the NoOcObs experiment, Figure 5 shows the ensemble-mean correlation for near-surface temperature computed for the four-year winter averages covering the forecast period two to five years. Large areas with positive skill appear in both experiments. Assim and XBT-C have both higher skill than NoOcObs over the tropical band, especially the tropical Pacific, and some regions of the Southern Hemisphere. Results for other seasons are similar. A reason for the better forecast quality of the Assim and XBT-C experiments over the tropical band during the first half of the forecast period may reside in the better representation of the long-term temperature linear trend in Assim and XBT-C with respect to NoOcObs (not shown). A more realistic trend has also been observed in predictions of the tropical upper ocean heat content of both Assim and XBT-C. The tropical skill improvement of the Assim experiment is consistent with the results found in a seasonal re-forecast experiment (Balmaseda and Anderson, 2009), although in this occasion for longer lead times. Instead, several regions such as the Northern Hemisphere continental areas and the North Pacific show NoOcObs as having higher skill than the two other experiments. It is important to bear in mind that the differences in skill between the experiments should be considered in the context of the correlations being computed with very small samples.

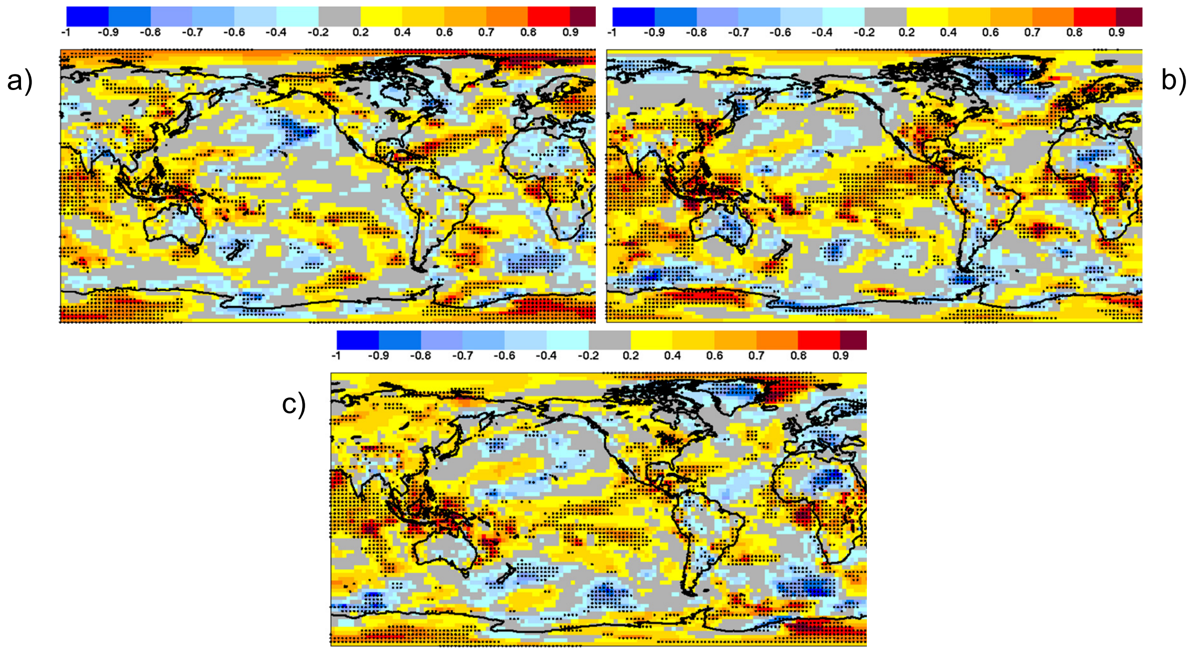

Figure 5 Ensemble-mean correlation for near-surface air temperature with respect to ERA40/ERA-Interim for winter (December to February) over the forecast period two to five years of the NoOcObs (a), XBT-C (b) and Assim (c) experiments. Three-member ensemble re-forecasts for the period 1960-2000 have been used. The black dots depict the grid points where the correlation is significantly different from zero with 95% confidence.

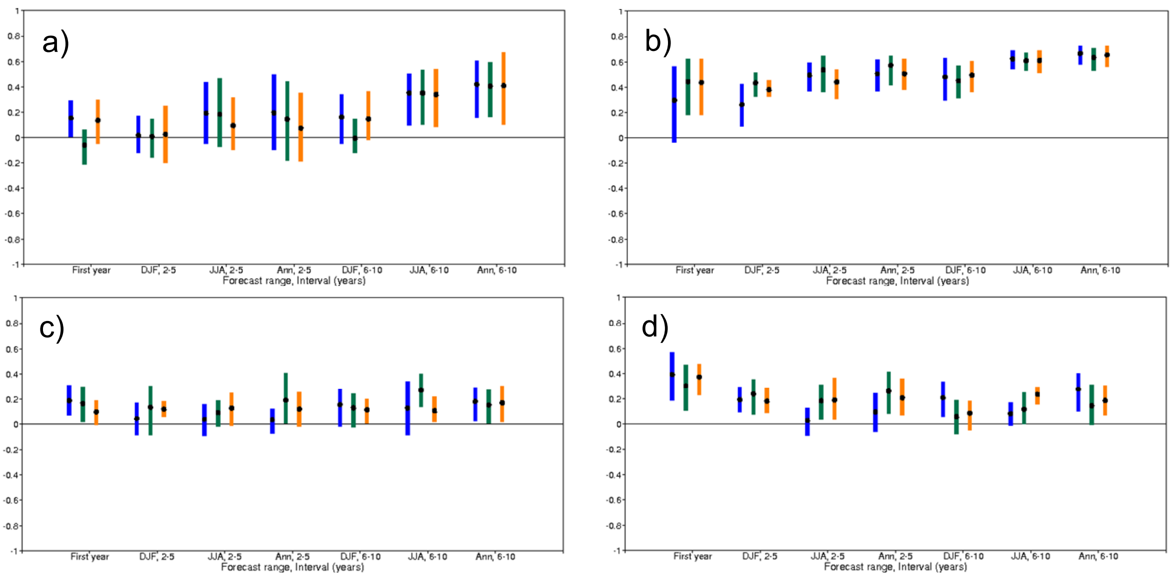

Figure 6 : Ensemble-mean correlation (top row) and perfect-model correlation (bottom row) for 850 hPa temperature over the Northern Hemisphere (north of 30°N, left column) and the tropical band (20°N-20°S, right column) for different forecast periods of the NoOcObs (blue bars), XBT-C (green bars) and Assim (orange bars) experiments. The pairs of bars correspond, from left to right, to the first calendar year of forecast (months 3-14, computed with re-forecasts for the period 1960-2005), the winter, summer and annual mean of the 2-5 year forecast period (computed with re-forecasts for the period 1960-2000), and the winter, summer and annual mean of the 6-10 year forecast period (computed with re-forecasts for the period 1960-1995). The black dots depict the sample values and the bars show the 95% confidence intervals. The estimates have been computed using ERA40/ERA-Interim and three-member ensemble re-forecasts.

A clearer picture of the differences in forecast quality between the three experiments for different forecast periods is shown in Figure 6, which depicts the anomaly correlation of 850 hPa temperature over the Northern Hemisphere and the tropical band for different forecast periods. To obtain each anomaly correlation coefficient, the spatial variance/covariance between the ensemble-mean and the corresponding reference is computed for each one of the re-forecasts available. The set of variances and covariances are then averaged over the set of start dates before the final correlation is computed. Confidence intervals for the scores have been computed using a bootstrap method, where the re-forecast/reference pairs were resampled with replacement 1,000 times (Lanzante, 2005; Jolliffe, 2007). The scores were then computed for each of the 1,000 samples, ranked and the intervals for specific confidence levels estimated (Doblas-Reyes et al., 2009). The reader should bear in mind that these correlations are lower than those obtained for global-mean variables due to the additional requirement of an adequate spatial distribution of the signal. In agreement with the results for near-surface air temperature, most cases display positive skill, with higher values for the tropical band than for the northern extratropics. Estimates for predictions of the first forecast year have been included to illustrate the impact of the time averaging, as well as the larger impact of the improvement of the initial conditions in the tropics with respect to the extratropics.

As in previous results, the experiments show a similar forecast quality over the northern extratropics and the tropical band. Similar results have been found for the Southern Hemisphere (not shown). Correlations for precipitation are much lower and in most cases non significant. The scores for the tropical region are almost all statistically significantly different from zero, while they only are for long forecast times over the northern extratropics. As in previous examples, the re-forecasts show larger skill for longer forecast times.

Figure 6 also contains information about the behaviour of the ensemble. The ratio between the spread, computed as the standard deviation of the ensemble members around the ensemble mean, and the RMSE has traditionally been used as a measure of the degree of calibration of the ensemble (Palmer et al., 2007). However, the small ensemble size of these experiments prevents the spread estimates from being robust enough to obtain meaningful results. Instead, the so-called perfect-model anomaly correlation has been used. This estimator measures the spread relative to the variability of the predictions and should not be interpreted as an upper-level of skill because it is model dependent. The perfect-model anomaly correlation is computed as the ensemble-mean anomaly correlation mentioned above, but in this instance taking one ensemble member as the reference. In other words, this estimator assumes that the reference is drawn from the same population as the re-forecasts (a hypothesis that is rarely true in an actual context), hence the use of the words "perfect-model". The set of variances/covariances are computed taking each one of the ensemble members as reference in turns and then averaging the other two, prior to the computation of the correlation.

The higher values found in Figure 6d for the first year suggest that the ensemble spread over the tropics is smaller at the beginning of the re-forecasts, the spread increasing with time, something that in most single-model forecast systems is a desirable feature due to their tendency to underforecast (Weigel et al., 2008). As mentioned above, the current experimental setup, which is shared with the one proposed for CMIP5, makes difficult to formulate conclusive statements about the ensemble spread due to the small sample. Larger ensemble sizes will be needed to address the question of an appropriate ensemble generation that would take into account the specific characteristics of decadal forecasting.

The perfect-model anomaly correlation has been used here as a measure of ensemble spread. However, it is sometimes also considered as a measure of the upper limit of the skill of a forecast system. This use might not be appropriate because, apart from this interpretation being valid only in the case of unbiased models, the perfect-model anomaly correlation could only be considered as an upper estimate of the skill in a stationary climate. This means that the perfect-model correlation can not take into account the effect of the long-term trends on skill, which can be a substantial contribution to the forecast quality. As an indication, the reader will note that in most cases in Figure 6 the anomaly correlation estimates are larger than the corresponding perfect-model anomaly correlations. This also suggests that some estimates of decadal predictability based solely on ensemble agreement measures (e.g. Boer and Lambert, 2008) might underestimate the actual skill.

4.2 Ocean variables

A substantial part of the skill in predicting near-surface air temperature is supposed to come from a skilful representation of the upper ocean heat content. Figure 7 displays anomalies of the global-mean upper ocean heat content from the three experiments. Anomalies for the XBT-C ocean re-analyses are also displayed in the figure. There are large differences between the three ocean re-analyses used to initialize the experiments, which indicate that there is a large uncertainty in the estimates of the ocean state (not shown). The re-analysis shows an upward trend after 1970 that is matched by the re-forecast anomalies. The main difference between the experiments is that the Assim predictions have a larger variability from one re-forecast to the next one in the early half of the period, to the point that they rarely overlap as the ensembles of the other two experiments do. This agrees with the behaviour of the ORA-S3 re-analysis, which has been affected by the error in the dropping rate of the XBTs. The re-analyses used as ocean initial conditions for Assim and XBT-C have different upper ocean heat content anomalies before 1990 (not shown), a feature that persists during the re-forecasts.

The ensemble-mean correlation of the global-mean upper ocean heat content with respect to ORA-XC is more similar for NoOcObs and XBT-C than for the Assim experiment (not shown). The low correlation for Assim is in contrast with what has been found for the global-mean near-surface air temperature (Fig. 1) and SST. A likely reason for such disagreement is the impact of using incorrect XBT data in ORA-S3 that, as an example, seems to have affected predictions such as the one started in 1980 (Fig. 7a). The differences between the experiments are also clear in the tropical upper ocean mean temperature re-forecasts (Fig. 7b). The higher skill of the upper ocean heat content in NoOcObs is also found over the tropics (Fig. 7c), although the correlation for the tropical band is sensibly lower than for the global average and not statistically significant with 95% confidence for most of the forecast range.

Previous studies (e.g. Collins et al., 2006) suggest that an accurate initialization of the AMOC could allow skilful predictions of the Atlantic multidecadal variability (AMDV) a few years in advance. However, past AMOC fluctuations have been poorly observed and a large uncertainty in the ocean analyses exists. This uncertainty is found even when ocean re-analyses carried out using the same ocean model are considered, as seen in Figure 8 where the AMOC from the XBT-C and NoOcObs re-analysis is shown. The uncertainty in the ocean reference implies that an assessment of the forecast quality of the AMOC predictions would necessarily give highly uncertain estimates. Hence, in this paper we focus on the ability of the experiments to simulate realistic AMOC anomaly variations.

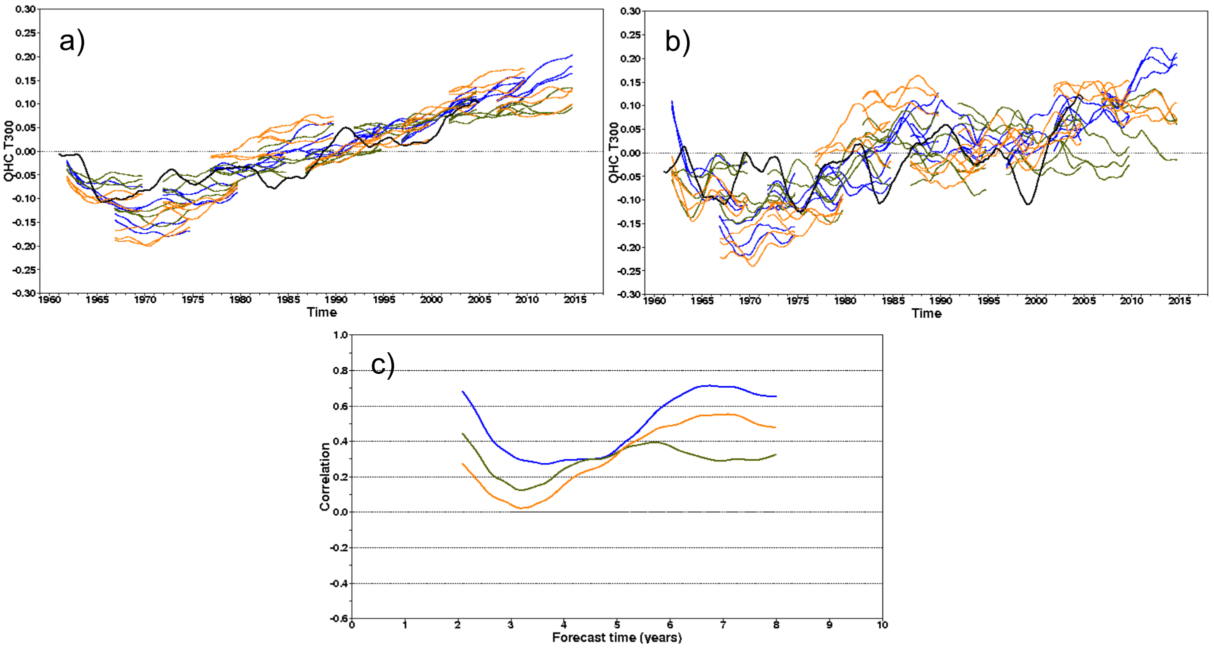

Figure 7 Anomalies of (a) global-mean and (b) tropical (20°N-20°S) ocean temperature (K) averaged over the top 300 metres for the ten three-member ensemble re-forecasts of the XBT-C (green), NoOcObs (blue) and Assim (orange) experiments. Anomalies are computed with respect to the corresponding climate over the period 1960-2005 (eight re-forecasts). Each re-forecast is illustrated with lines of a different colour. Anomalies from ORA-XC ocean re-analysis are shown in black solid lines. All time series have been smoothed out with a 24-month centred moving average that removes data for the first and last years of each time series. (c) Ensemble-mean correlation of the tropical ocean temperature averaged over the top 300 metres of the XBT-C (green), NoOcObs (blue) and Assim (orange) experiments and have been computed using ORA-XC data and three-member ensemble re-forecasts for the period 1960-1995. The correlation has been computed with a moving window of four-year averaged anomalies.

The XBT-C experiment has been initialized with ocean states that reproduce the AMOC variability to the best of our knowledge, while the ocean initial conditions used in the NoOcObs experiment underestimate both the mean AMOC intensity and its variability (Fig. 8a). Figure 8 shows the anomalies of the AMOC intensity and of an index of the Atlantic multidecadal oscillation (AMO), calculated as the average North Atlantic SSTs north of 10ºN. The AMOC re-forecast anomalies and those from the ORA-XC analyses show similar interannual oscillations. This coupled model shows a realistic interannual variability of the North Atlantic Oscillation (not shown), which is one of the mechanisms that govern the wind variability over the North Atlantic. However, this behaviour has to be put in the context of the strong AMOC drift described in the previous section. The ensemble-mean AMOC correlation (Fig. 8c) is always lower than 0.6, with large differences between the experiments although with positive values in the first few forecast years. On the basis of these results is virtually impossible to claim that one experiment is better than the other as the largest difference in correlation has a confidence of 85%, a low value even before taking into account the obvious serial correlation of the time series.

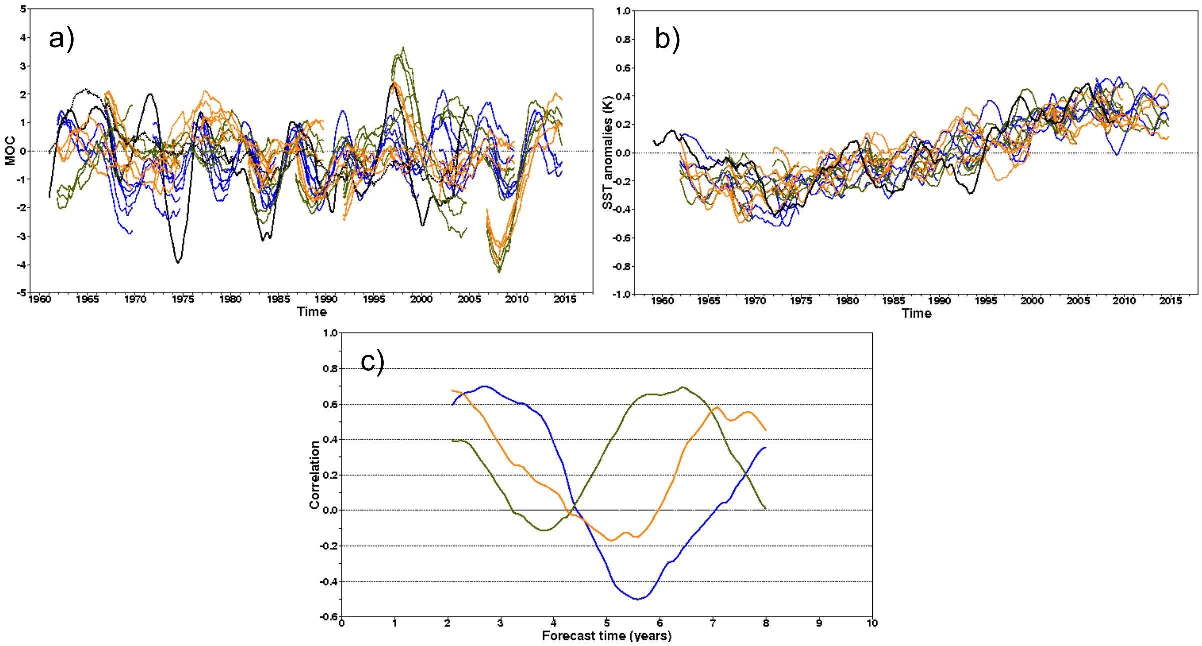

Figure 8 (a) Atlantic meridional overturning circulation intensity (Sv) and (b) North Atlantic (10°N-65°N) average SST (K) anomalies for the ten decadal three-member ensemble re-forecasts of the XBT-C (green), NoOcObs (blue) and Assim (orange) experiments. Anomalies are computed with respect to the corresponding climate over the period 1960-2005. Anomalies from ORA-XC (NoOcObs) ocean re-analysis are shown in black solid (dashed) lines on the (a) panel. Data from ERA40/ERA-Interim are shown in black in the (b) panel. All time series have been smoothed out with a 24-month centred running mean that removes data for the first and last years of each time series. (c) Ensemble-mean correlation of the Atlantic meridional overturning circulation intensity with respect to ORA-XC for XBT-C (green), NoOcObs (blue) and Assim (orange), which has been computed with three-member ensemble re-forecasts for the period 1960-1995 on four-year running mean anomalies.

As suggested by Dijkstra et al. (2006), some agreement is expected between the AMOC and AMO indices from the reference datasets, especially if the upward trend in the AMO is discounted. This supports the use of the AMO as a proxy indicator for the intensity of the AMOC, although the impact of global warming should be separated from the AMO interannual and decadal variability. Predictions of the AMO index (Fig. 8b) are quite skilful, especially because they reproduce the upward trend observed in the analysis since the mid 1970s. The ensemble-mean correlation of four-year average predictions is above 0.6 for all forecast periods and similar for all the experiments. A more pessimistic picture has been found with an alternative AMO index estimated as SST averaged over a northern (40-60°N, 60-10°W) and a southern (40-60°S, 50°W-0°) Atlantic box (Latif et al., 2006). In this case, the strong SST drift over the southern Atlantic Ocean (Fig. 2) might adversely affect the simulations and limit the reproducibility of that specific AMO index. In any case, it has proved difficult to find a correspondence between the AMO and AMOC estimates in these experiments.

5 Summary and discussion

The drift and forecast quality of three sets of decadal re-forecasts carried out with the IFS/HOPE coupled system have been analyzed. The three experiments are different in their initial conditions. The Assim experiment has been initialized with an ocean re-analysis that includes ocean data assimilation, the XBT-C experiment is initialized with data from an ocean re-analysis similar to the one used for

Assim but with an important correction in the XBT data assimilated while the NoOcObs experiment has been initialized with data from an ocean-only simulation where no subsurface ocean data have been assimilated. This is the first time that a similar set of decadal predictions has been carried out in an attempt to assess the relevance of ocean data assimilation in decadal prediction with realistic initialization.

As the re-forecasts are initialized with a realistic state, a sizeable drift develops during the forecast time in both ocean and atmospheric variables. There are many possible reasons for the existence of the drift, starting from the lack of balance between the radiation budget used to produce the ocean and atmospheric re-analyses, the intrinsic systematic errors of the atmospheric and ocean models, where the atmospheric model does not properly simulate the tropical Sc clouds and has excessively weak trade winds, while the ocean model does not correctly represent the most relevant eddies. We made efforts to reduce the model drift, and that was the reason for introducing changes in the cloud microphysics that reduced the atmospheric model cold bias be increasing the surface solar radiation, but much more remains to be done, including the inclusion of stochastic physics schemes. Furthermore, in a seamless climate prediction spirit, reducing the drift for decadal forecasting will benefit monthly and seasonal forecast systems, which are affected by a similar problem.

Although all experiments have a similar drift, cold over the tropical oceans and warm over certain areas of the northern continents and the southern oceans, there are differences of the order of tenths of a degree between Assim and NoOcObs. The differences in the drift tend to cancel out over the tropics in the first forecast years, although a slightly warmer tropical ocean develops in NoOcObs between 300 and 600 m. The tropical SST drift becomes virtually equal in the three experiments following an increased surface latent heat flux into the atmosphere over the west Pacific and the SPCZ in Assim and XBT-C with respect to NoOcObs, which seems to be linked to an increase in outgoing top net radiation. In contrast with the tropics, the extratropical drift differences persist for the whole duration of the simulations, the differences concentrating on specific basins as forecast time increases. This suggests that small-amplitude signals in the extratropical ocean initial conditions can have an impact a long time after the forecast is started. For instance, although the average top ocean temperature and SST drift differences in the southern and northern hemispheres cancel each other out to give a low global-mean drift, there is a regional impact between both hemispheres that affects the surface latent heat flux drift.

In spite of the model drift and the fact that several climate processes, such as those related to sea-ice formation, export and melting, are not represented in the model, the decadal prediction experiments described here show a positive forecast quality that can be statistically significant over several areas and that is comparable to that published previously. Positive correlation with observations is found for tropospheric air temperature and upper ocean heat content, the correlation increasing with forecast time in most cases. Precipitation does not show significantly positive skill beyond the first year.

The experiments show very similar forecast quality. In those cases where some differences appear, the differences are not statistically significant with a high confidence level. However, the experiments initialized with assimilated ocean observations (Assim and XBT-C) seem to be marginally more skilful over the tropics than NoOcObs in the first few forecast years. This happens in spite of the predictions of the top ocean averaged temperature being more skilful for NoOcObs than for the other two experiments. The ability to predict interannual variations of the AMOC is difficult to assess, particularly due to the uncertainty in AMOC estimates from ocean re-analyses. A visual exploration of the re-forecasts indicates that the anomalies in both experiments have common oscillations that resemble those observed in an independent ocean re-analysis and encourage a more in-depth analysis of the skill in predicting interannual variations of the AMOC.

The insignificant differences found between the three experiments might be disappointing, but present an important aspect. The fact that only small forecast quality differences are found between the experiments might be due to the important model drift that, as shown in the seasonal forecasting context, might prevent the model from making the most of the additional information available in the experiment initialized from re-analysis that use ocean data assimilation. The seasonal forecast community only found a positive impact of subsurface ocean observations in sufficiently skilful systems where model drift had been substantially reduced. The use of a a-posteriori linear bias-correction scheme is just a very simple approach to make the re-forecasts tractable, while it is clear that the interaction between the forecast anomalies and the drift can be highly non-linear. Additional experimentation with future versions of the forecast system presented here where the drift is reduced might help shedding light into the limiting role of the drift to assess the benefits of an improved observing system. Besides, the option of initializing the re-forecasts in anomaly mode (Smith et al., 2007) should also be explored.

The experiments described in this paper use the experimental setup defined in the ENSEMBLES project. This setup has been inherited by CMIP5 for its decadal prediction exercise. One of the main problems found with this setup is that the differences between experiments are so subtle and the uncertainty in the forecast quality estimates so large that it is difficult to extract significant conclusions with the short samples and small ensemble size considered. Unfortunately, both the sample and the ensemble sizes are, instead, limited by the important computing resources required to run even experiments of this size. Similar difficulties are likely to be found when this type of experiments will be used to determine how the decadal prediction forecast quality is improved when using initialization with respect to uninitialized predictions.

An additional issue is that the small sample size introduces a caveat in the anomaly computation. The model anomalies have been computed in cross-validation using the set of re-forecasts, a sub-optimal procedure that uses information from the future for certain re-forecasts in the estimation of the model climate. The method used computes the anomalies using a different model climate estimate for each forecast period to take into account the time-dependent nature of model drift, which is the main reason for these experiments to require long and expensive samples. As an alternative that could work even with small re-forecast samples, other systems (Smith et al., 2007) proposed using long-term simulations of 20th Century climate to estimate the model climate for the calculation of the anomalies. However, these long-term simulations do not allow taking into account any form of drift, which might happen even in systems that use flux correction terms (as some drift is present over land or on the atmospheric and ocean variability) or in systems using anomaly initialization. As a consequence, anomalies calculated with respect to the long climate simulations, even when they could be appropriate, might be biased and their information be misleading.

The results of the few initialized decadal forecast experiments carried out to date as well as the results shown here suggest that although there is some skill in predicting air temperature, the skill for other variables is rather limited. However, some gain in skill has been found of multi-year averages of atmospheric variables and ocean circulation from initializing with respect to uninitialized predictions (D. Smith, personal communication). As happened already in the field of seasonal forecasting, a reduction of the model drift and a better understanding of the processes at the origin of the interannual and decadal predictability should produce more skilful multi-year useful predictions in the future, as well as an increased benefit from a better informed initialization of the predictions.

Acknowledgements

This work was supported by the ENSEMBLES project (GOCE-CT-2003-505539). The authors acknowledge the significant contributions by Richard Forbes, Kristian Mogensen, Jean-Jacques Morcrette, Franco Molteni and Tim Stockdale. The manuscript has also benefited from fruitful discussions with David Anderson, Thomas Jung, James Murphy, Doug Smith and Noel Keenlyside.

References

Adler, R. F., G. J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, and E. Nelkin (2003), The version 2 vlobal precipitation climatology project (GPCP) monthly precipitation analysis (1979-present), *J. Hydrometeor*., **4**, 1147-1167.

Anderson, D. L. T., T. Stockdale, M. Balmaseda, L. Ferranti, F. Vitart, F. Molteni, F. J. Doblas-Reyes, K. Mogensen, and A. Vidard (2007), Development of the ECMWF seasonal forecast System 3, *ECMWF Technical Memorandum* No. 503, 56 pp, ECMWF, Reading, UK.

Anderson, D. L. T., F. J. Doblas-Reyes, M. Balmaseda, and A. Weisheimer (2009), Decadal variability: processes, predictability and prediction, *ECMWF Technical Memorandum* No. 591, 47 pp, ECMWF, Reading, UK.

Balmaseda, M. A., A Vidard and D. L. T. Anderson (2008), The ECMWF ocean analysis system: ORA-S3, *Monthly Weather Review*, ***136***, 3018–3034.

Balmaseda M., and D. L. T. Anderson (2009), Impact of initialization strategies and observations on seasonal forecast skill, *Geophys. Res. Lett*., ***36***, L01701, doi:10.1029/2008GL035561.

Balmaseda M. A., G. C. Smith, K. Haines, D. Anderson, T. N. Palmer, and A. Vidard (2007), Historical reconstruction of the Atlantic meridional overturning circulation from the ECMWF operational ocean reanalysis, *Geophys. Res. Lett*., **34**, L23615, doi:10.1029/2007GL031645.

Bechtold, P., M. Koehler, T. Jung, F. J. Doblas-Reyes, M. Leutbecher, M. J. Rodwell, F. Vitart, and G. Balsamo (2008), Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales, Quart. J. Royal Meteor. Soc. , 134, 1337-1351.

Boer G. J., and S. J. Lambert (2008), Multi-model decadal potential predictability of precipitation and temperature, Geophys. Res. Lett., 35, L05706, doi:10.1029/2008GL033234.

Collins, M., M. Botzet, A. F. Carril, S. Masina, H. Drange, O. H. Otteraa, A. Jouzeau, L. Terray, M. Latif, H. Pohlmann, A. Sorteberg, and R. Sutton (2006) Interannual to decadal climate predictability in the North Atlantic: A multimodel-ensemble study, *J. Climate*, **19**, 1195-1203.

Dijkstra, H. A., L. te Raa, M. Schmeits, and J. Gerrits (2006), On the physics of the Atlantic Multidecadal Oscillation, *Ocean Dyn*., **56**, 36-50, doi:10.1007/s10236-005-0043-0.

Doblas-Reyes, F. J., R. Hagedorn, T. N. Palmer, and J.-J. Morcrette (2006), Impact of increasing greenhouse gas concentrations in seasonal ensemble forecasts, *Geophys. Res. Lett*., **33**, L07708, doi:10.1029/2005GL025061.

Doblas-Reyes, F. J., A. Weisheimer, M. Déqué, N. Keenlyside, M. McVean, J. M. Murphy, P. Rogel, D. Smith, and T. N. Palmer (2009), Addressing model uncertainty in seasonal and annual dynamical seasonal forecasts, *Quart. J. Roy. Meteor. Soc*., **135**, doi:10.1002/qj.464.

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 Technical Memorandum* No. 621, 45 pp, ECMWF, Reading, UK.

Drijfhout, S., W. Hazeleger, F. Selten, and R. Haarsma (2008), Future changes in internal variability of the Atlantic Meridional Overturning Circulation, *Climate Dyn*., **30**, 407-419, doi:10.1007/s00382-007-0297-y.

Hawkins, E., and R. Sutton (2009a), Decadal predictability of the Atlantic Ocean in a coupled GCM: forecast skill and optimal perturbations using linear inverse modeling, *J. Climate*, **22**, 3960-3978.

Hawkins, E., and R. Sutton (2009b), The potential to narrow uncertainty in regional climate predictions, *Bull. Amer. Meteor. Soc*., in press.

Hermanson, L., and R. T. Sutton (2009), Climate predictability in the second year, *Phil. Trans. R. Soc.* **A, 367**, 913-916, doi:10.1098/rsta.2008.0181.

Ishii, M., and M. Kimoto (2009), Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections, *J. Oceanography*, **65**, 287-299, doi:10.1007/s10872-009-0027-7.

Jolliffe, I. T. (2007), Uncertainty and inference for verification measures, *Wea. Forecasting*, **22**, 137-150.

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

Lanzante, J. R. (2005), A cautionary note on the use of error bars, *J. Climate*, **18**, 3699-3703.

Latif, M., C. Böning, J. Willebrand, A. Biastoch, J. Dengg, N. Keenlyside, and U. Schweckendiek (2006), Is the thermohaline circulation changing?, *J. Climate*, **19**, 4631-4637.

Lean, J. L., and D. H. Rind (2009), How will Earth’s surface temperature change in future decades, *Geophys. Res. Lett*., **36**, L15708, doi:10.1029/2009GL038932.

Lozier, M. S. (2010), Deconstructing the conveyor belt, *Science*, **328**, 1507-1511, doi:10.1126/science.1189250.

Oldenborgh, G. J. van, S. S. Drijfhout, A. van Ulden, R. Haarsma, A. Sterl, C. Severijns, W. Hazeleger, and H. Dijkstra (2009), Western Europe is warming much faster than expected. *Climate of the Past*, **5**, 1-12.

Palmer, T. N. (2000), The prediction of uncertainty in weather and climate forecasting, *Rep. Prog. Phys*., **63**, 71-116.

Palmer, T. N., R. Buizza, R. Hagedon, A. Lawrence, M. Leutbecher, and L. Smith (2007), Ensemble prediction: A pedagogical perspective, *ECMWF Newsletter*, **106**, 10-17.

Pohlmann H., F. Sienz, and M. Latif (2006), Influence of the multidecadal Atlantic overturning circulation variability on European climate, *J. Climate*, **19**, 6062-6067.

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

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang (2002), An improved in situ and satellite SST analysis for climate, *J. Climate*, **15**, 1609–1625.

Smith D. M, S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy (2007), Improved surface temperature prediction for the coming decade from a global climate model, *Science*, **317**, 796–799, doi:10.1126/science.1139540.

Uppala, S. M., and 45 others (2005), The ERA-40 reanalysis, *Q. J. R. Meteorol. Soc.*, **131**, 2961-3012.

Weigel, A. P., M. A. Liniger, and C. Appenzeller (2008), Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?, *Quart. J. Roy. Meteor. Soc.*, **134**, 241-260.

Weisheimer, A., F. J. Doblas-Reyes, T. N. Palmer, A. Alessandri, A. Arribas, M. Déqué, N. Keenlyside, M. MacVean, A. Navarra, and P. Rogel (2009), ENSEMBLES - a new multi-model ensemble for seasonal-to-annual predictions: Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs, *Geophys. Res. Lett.*, **36**, L21711, doi:10.1029/2009GL040896.

Wijffels, S. E., J. Willis, C. M. Domingues, P. Barker, N. J White, A. Gronell, K. Ridgway, and J.A. Church (2008), Changing eXpendable BathyThermograph fallrates and their impact on estimates of thermosteric sea level rise, *J. Climate*, 21, 5657–5672.