Tellus (2005), 57A, 308–319
Printed in UK. All rights reserved

TELLUS

Actual and potential skill of seasonal predictions using the CNRM contribution to DEMETER: coupled versus uncoupled model

By JEAN-FRANÇOIS GUÉRÉMY*, MICHEL DÉQUÉ, ALAIN BRAUN and JEAN-PHILIPPE PIEDELIÈVRE, *Météo-France, Centre National de Recherches Météorologiques (CNRM), 42 av. Coriolis, F-31057 Toulouse Cedex, France*

(Manuscript received 31 March 2004; in final form 24 September 2004)

ABSTRACT

The skill of seasonal forecasts carried out with the CNRM general circulation model (GCM) is assessed over the winter and summer seasons during the period 1958–2001. Three types of forecasts are considered. Two of them, compatible with real-time forecasting, make use of the ocean–atmosphere coupled model and the sea surface temperature (SST) statistical model, while the third provides an upper limit of the skill taking the observed SSTs into account. The nine-member ensemble skill is evaluated with the help of anomaly correlation coefficients (deterministic skill) and economical values (probabilistic skill). The coupled model skill is larger than that of the statistical SST forced model, from both deterministic and probabilistic point of views. Moreover, the coupled forecast skill is not far from that of the observed SST forecasts. Over the tropics, the correlations (above 0.3) and economical values are significant. Over Europe, only observed SST forecasts reach the level of significance in both seasons. For the years of the largest observed normalized anomalies (tropical precipitation and temperature at 850 hPa over Europe), composites of the atmospheric response are constructed in a trial of understanding the nature of predictability in the GCM, in comparison with the observation. The teleconnection patterns between the tropics and the Northern Hemisphere are not well reproduced by the GCM. However, it shows a rather good potential skill in the sense that the years characterized by the largest internal ensemble consistency give an atmospheric response very similar to that obtained during the years for which the observed anomalies are the largest.

1. Introduction

The scientific basis of seasonal forecasting consists of the influence of slow evolving surface boundary fields on atmospheric variables. Among these boundary fields, sea surface temperature (SST) is the most important, especially over the tropics where this field presents high values together with large-scale slow evolving modes (Palmer and Anderson, 1994; Shukla, 1998). Over the tropics, the large-scale atmospheric responses to surface SST anomalies consist of divergent Kelvin wave circulations above the anomaly and Rossby wave circulations aside, linking that way tropics to mid-latitudes (Gill, 1980). Recently, the European Union PROVOST (PRediction Of climate Variations On Seasonal-to-interannual Time-scales) project has been devoted to the study of potential predictability by analysing ensembles of integrations of atmospheric models with specified observed SSTs over the four seasons during the 1979–1993 period (Brankovic and Palmer, 2000). Beyond the study of potential predictability, real-time seasonal forecasts produced by dynamical models require the knowledge of the SST field for the coming season. Two main approaches are considered for that purpose (i.e. predicting SSTs). The first approach is based on rather simple statistical schemes making use of persistence, while the second takes advantage of coupled ocean–atmosphere general circulation models (GCMs). The former has been presently chosen by Météo-France to produce its operational seasonal forecasts and has been assessed by earlier studies (e.g. Graham et al., 2000). The latter has been extensively used in the European Union DEMETER (Development of a European Multi-model Ensemble system for seasonal to inTERannual prediction) project (Palmer et al., 2004).

The aim of this paper is to comparatively evaluate the predictability of seasonal simulations carried out with ARPEGE-Climat (henceforth AC; CNRM GCM) by statistically predicted SSTs, and coupled with the OPA ocean model, over the summer and winter seasons of the 1958–2001 period. The PROVOST approach (not compatible with real-time forecasting) has also been

*Corresponding author.
e-mail: jean-francois.gueremy@meteo.fr

considered, because it provides an upper limit of the atmospheric predictability, under the assumption that coupling feedbacks play a negligible role. In Section 2 we present the experimental set-up. Scores are discussed in Section 3, focusing on tropical precipitation and temperature at 850 hPa over Europe. First, deterministic validation is considered with the help of anomaly correlation of the ensemble mean and, secondly, probabilistic validation is presented using the economical value (i.e. percentage of money saved with respect to a perfect forecast in the frame of a cost-loss model; Palmer et al., 2000) of a predicted variable being above a chosen threshold. Atmospheric response patterns to SST forcing are discussed in Section 4. Here, for a particular variable, two target regions (precipitation over the tropics and temperature at 850 hPa over Europe) are considered, for the observation and the three forecast approaches.

2. Experimental set-up

The DEMETER project (Palmer et al., 2004) is based on the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-yr Reanalysis (henceforth ERA-40; Simmons and Gibson 2000), which provides atmospheric surface fluxes in order to obtain oceanic initial conditions using forced ocean runs, initial atmospheric conditions and verification data. ERA-40 and DEMETER cover the period 1958–2001, that is 44 yr.

The CNRM coupled model is composed of the three following components: AC, OASIS and OPA. The atmospheric model is AC version 3 (Déqué et al., 1994; Déqué, 2001); it is a spectral atmospheric general circulation model (AGCM) using a linear T63 truncation, a reduced 128 × 64 Gaussian grid and 31 vertical levels. The dynamical core is the same as that of ECMWF [Integrated Forecast System (IFS)] with a semi-implicit, semi-Lagrangian, two time level discretization scheme. The physical parametrizations are partly inherited from the Météo-France operational forecast model. The oceanic model is OPA 8.0 (Madec et al., 1997); it is a finite-difference based oceanic general circulation model (OGCM), which solves the primitive equations with a non-linear equation of state on an Arakawa C-grid. The present configuration uses a rigid lid (i.e. free surface elevation set to zero, the surface pressure gradient being obtained by approximation). The northern point of convergence has been replaced by two poles located in Asia and North America to overcome the singularity at the North Pole. Its space resolution is roughly equivalent to a geographical mesh of $2^\circ \times 1.5^\circ$ (with a meridional resolution of 0.5° near the Equator). There are 31 vertical levels, with 10 levels in the top 100 m. Sea-ice cover is calculated from OPA surface temperature and salinity. The two GCMs are coupled with the OASIS 2.2 (Terray et al., 1995) coupler.

In order to obtain ocean initial conditions for the coupled forecasts, OPA has been modified to be integrated in forced mode, using the ERA40 daily surface fluxes (net energy flux, downward solar flux, water flux and wind stress) with the addition of daily wind stress perturbations (positive, zero and negative) provided by the ECMWF, both data sets being interpolated to the OPA grid. The forced integrations of OPA using a relaxation (time constant of 2 d) toward the ERA40 daily SSTs have been carried out to obtain three ocean initial states for each DEMETER starting date (four dates per year). These OPA forced integrations have been started at the beginning of each ERA40 main stream (i.e. 1 January 1958, 1 January 1973, 1 October 1986) starting with ocean initial conditions coming from an ocean spin-up run (8 yr with ERA15 averaged monthly fluxes). Initial conditions have then been obtained in the following manner. For the atmosphere, ERA40 atmospheric variables are interpolated to the AC grid, for each DEMETER starting date; continental subsurface variables are also interpolated, using a normalization for the soil wetness to keep the same saturation fraction for the two different soil models (IFS and AC). For the ocean, temperature perturbations (between the surface and 40 m) provided by the ECMWF are added to the three files created by the OPA forced integrations (see above), to finally obtain nine ocean initial conditions constituting each ensemble of coupled hindcasts.

As far as the SST statistical model is concerned, the methodology introduced by Déqué (2001) has been adopted. This model consists of a red noise process

$$\mathrm{SST}(m) = A\mathrm{SST}(m-1) + BWN(m), \tag{1}$$

where $\mathrm{SST}(m)$ is the SST anomaly of a given month, $\mathrm{SST}(m-1)$ is the SST anomaly of the previous month, and $WN(m)$ is a random component (white noise with mean equals zero and variance equals one). A and B are coefficients fitted by least-squares during the period 1950–1978. This autoregression has been computed with the first 20 principal components, A and B depending on the EOF number and on the calendar month.

Finally, only the winter and summer seasons have been considered in this work. Each forecast, among the three types, consists of an ensemble of nine members starting, respectively, in November and May (months two to four averaged to obtain a seasonal mean). For the observed SST simulations (hereafter called O) and the statistical SST predicted forecasts (hereafter called S), ensembles are built using the ERA-40 atmospheric analyses lagged by 24 h. Ensemble generation for the coupled forecasts (hereafter called C) has been described in the preceding paragraphs.

3. Scores

The first field that has to be considered is SST for experiments C and S. The statistical scheme used in experiment S has been designed with a zero bias constraint (see Section 2) while the C experiment surface temperature field has a bias shown in Fig. 1 for summer and winter. The absolute value of this bias is almost everywhere less than 1 K over the oceans, for both seasons. However, some specific regions are characterized by warm or cold biases due to known cloudiness deficiencies of the AC GCM.

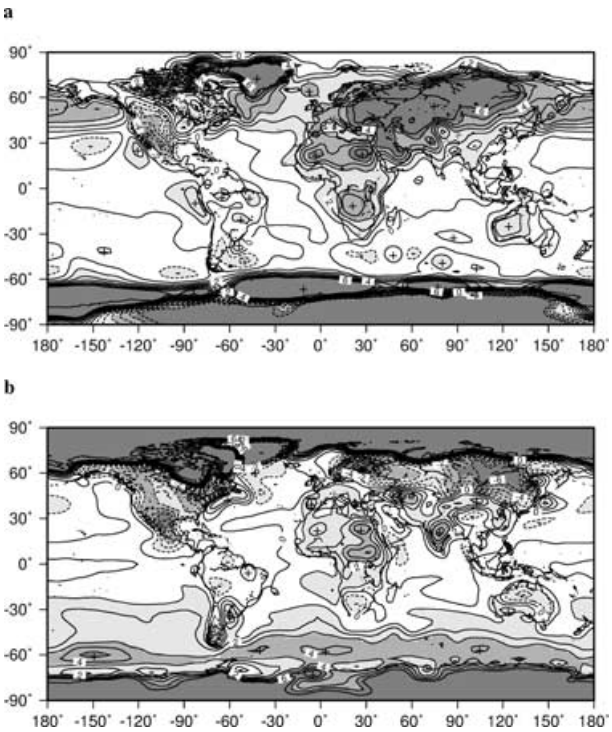

Fig 1. Coupled forecast surface temperature bias for JJA (a) and DJF (b). Contours every 1°C. A solid line represents positive values and dashed line negative values.

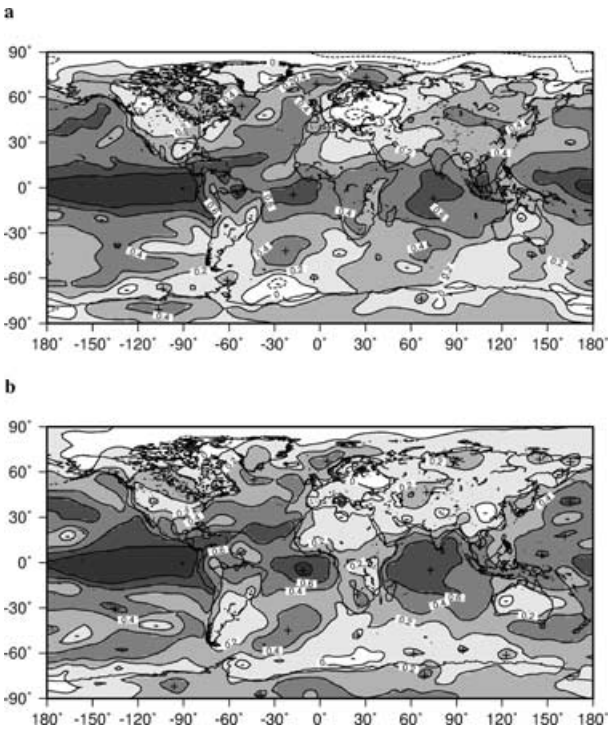

Fig 2. Surface temperature ACC (with respect to ERA-40) using (a) the coupled model and (b) the atmospheric model with statistical SST in DJF. Contours every 0.2. A solid line represents positive values and dashed line negative values.

Indeed, the eastern edges of tropical oceans and zonal bands around 60° in the summer hemisphere show a lack of clouds resulting in a warm bias, whereas both sides of the intertropical convergence zone together with the western coast of Central America and the eastern coast of Africa tend to be too cloudy resulting in a cold bias. The largest warm biases are located in the polar regions of the winter hemisphere; they are caused by the melting of sea ice because of the absence of relaxation toward the climatology in the deep ocean during the forced ocean run (see Section 2). Beyond the bias, it is more important for seasonal prediction to have the same sign of the SST anomaly. Figure 2 presents the global maps of surface temperature anomaly correlation coefficients (ACCs) of experiments C and S, with respect to the 'analysed' (ERA-40) surface temperature, for the winter season. Correlations are greater than 0.6 over most tropical oceans, with the largest values extending over the central Pacific Ocean, the El Niño region. It seems difficult to notice any significant difference between the two correlation patterns over the tropical oceans, except perhaps over the north-western Pacific, where C tends to be better than S (also true in summer, not shown).

In the following, the emphasis will be put on two variables: temperature at 850 hPa and precipitation (hereafter called T850 and Prec, respectively). For both variables, a target region will be considered: Europe (34°–76°N, 14°W–45°E) for the former and the tropics (zonal band between 30°S and 30°N) for the latter. T850 and Prec are typical variables chosen in seasonal forecasting studies (e.g. Doblas-Reyes et al., 2000); indeed, their seasonal anomalies affect the most the climate and human activities over the respective regions. Two types of scores will be presented: deterministic and probabilistic (e.g. Déqué, 2001). First, ACCs have been computed between predicted and analysed (ERA-40) variables in both time (at each grid point, over the ERA-40 time period) and space (for each year, considering the respective target regions); ERA-40 precipitation consists of a 24-h averaged field (hour 12 to hour 36 after the analysis). The advantage of this score versus the classical mean square error is the fact that it intrinsically removes the systematic errors of mean and standard deviations in its computation, therefore giving a potential skill. Secondly, the economical values of the considered forecasted variables have been calculated. This score, based on the cost-loss model (Palmer et al., 2000) with a two-category contingency table, has two advantages in the context of this paper. On one hand, it is less severe than the Brier score (Brier, 1950), which is the square difference between the forecast and observed probability, thus taking into account the systematic error in the standard deviation of the forecasted probability density. On the other hand, this score is directly oriented toward the end-users having their own cost-loss ratio with the possibility of letting vary the threshold defining the two categories of

Table 1. ACCs for the tropical precipitation (first two rows) and the temperature at 850 hPa over Europe (last two rows). Between parentheses, the 95% confidence interval of ACCs of three among the nine-member ensemble mean has been indicated. O, C and S denote observed SST, coupled and statistical SST modes, respectively

Mode	O	C	S
ACC JJA	0.29 (0.25 0.27 0.28)	0.27 (0.21 0.23 0.25)	0.16 (0.10 0.12 0.14)
ACC DJF	0.39 (0.35 0.36 0.37)	0.36 (0.30 0.31 0.33)	0.30 (0.23 0.25 0.27)
ACC JJA	0.12 (0.02 0.09 0.16)	−0.04 (−0.11 −0.02 0.05)	−0.01 (−0.07 0.00 0.07)
ACC DJF	0.08 (−0.05 0.06 0.16)	0.14 (−0.02 0.08 0.18)	−0.02 (−0.13 −0.02 0.08)

the contingency table. For both deterministic and probabilistic scores, a 15-yr running mean has been subtracted from T850 and Prec in order to avoid the influence of global change trend over the ERA-40 time period in our computations.

Table 1 presents ensemble mean ACCs of Prec over the tropics and T850 over Europe, for JJA and DJF and the three types of seasonal forecasts (O, C and S). ACCs have been computed over the ERA-40 time period as the mean covariance divided by the product of the mean standard deviations. The first value is the ACC of the nine-member ensemble mean, while between parentheses, the 95% confidence interval of ACCs of three among the nine-member ensemble mean (2.5%, 50% and 97.5% quantiles) has been indicated to provide information on the uncertainty of this score. Figure 3 shows the ensemble mean ACC global maps of Prec and T850 for the three types of seasonal forecasts (O, C and S).

Concerning tropical precipitation, it is interesting to notice that ACCs are larger in DJF compared to JJA. The larger skill in winter is due to the fact that, during this season, El Niño-type SST anomalies affect regions where the SSTs are the largest (equatorial western Pacific and the northern side of equatorial eastern Pacific), therefore producing the largest precipitation anomalies, giving rise to high values of ACC (Fig. 3). The largest ACCs (above 0.4, for the three types of forecasts) are recorded during five El Niño years (1972, 1982, 1997 in summer and winter, and 1965, 1994 in winter; for the winter season, DJF, the year is that of the month of December) and one La Niña year (1988 in summer and winter). Comparing ACCs from the three types

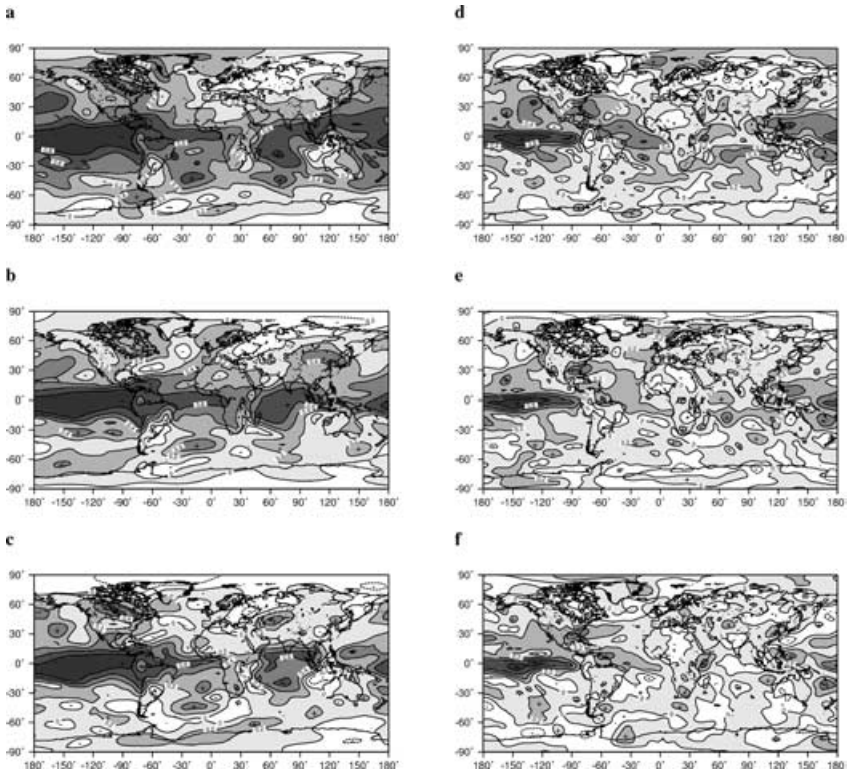

Fig 3. The 850-hPa temperature and precipitation ACC (with respect to ERA-40) using (a), (d) the atmospheric model with observed SST, (b), (e) the coupled model, and (c), (f) the atmospheric model with statistical SST in DJF. Contours every 0.2. A solid line represents positive values and dashed line negative values.

of seasonal forecasts, it clearly appears that C is close to the 'perfect' ocean forecast and that the gap between O and C is smaller than that between C and S. This result is an a posteriori proof that the coupled model is able to produce more realistic SST anomalies than that of the statistical model, even if it was not obvious looking at the SST ACCs (Fig. 2). The statistical model is unable to maintain the magnitude of the anomaly, despite keeping the right sign. For example, Prec ACCs over the eastern side of the Indian Ocean from the S forecast are weak compared to those from O and C (Fig. 3), with C tending to be even better than O. The last fact is an indication of the known importance of ocean–atmosphere feedbacks over that region during DJF in relation to the Madden and Julian oscillation (Madden and Julian, 1972). It is worth noting that the best winter ACCs (greater than 0.6) over continents are obtained in the south-east of the United States (US) for the three types of seasonal forecasts (Fig. 3). The ACC distribution intervals given in Table 1 show that the differences between the two types of seasonal forecasts are statistically significant, the intervals being disjoined; moreover, the dispersion in the ACC distribution is larger in C and S compared to O, as expected because of the different SST fields used in each member of S and C versus the unique observed SST field used in O.

As far as T850 over Europe is concerned, ACCs are, as expected, much lower than those of the tropical precipitation. Only O seasonal forecasts, in both winter and summer, and C in winter are characterized by significantly positive ACCs (considering the ACC distribution of three among the nine-member ensemble). The largest ACCs (above 0.4, for the three types of forecasts) are found during four El Niño years (1973, 1990, 1997 in winter and 1992 in summer). For the summer season, ACCs greater than 0.4 are recorded only for O forecasts (having an overall score of 0.12; see the discussion in the next section) during two additional El Niño years (1980, 1982) and five La Niña years (1975, 1978, 1985, 1996, 2001). Figure 3 clearly confirms that the largest ACCs are obtained by C forecasts over Europe during winter. Compared to O forecasts, this apparent better score (not that much better, if we refer to the ACC distribution shown in Table 1) is not caused by the right physical reasons, as explained in the following section. The larger distribution dispersion versus that of Prec, together with their similar magnitude from each type of forecast, a posteriori demonstrates the more chaotic nature of European T850 versus tropical Prec. Moreover, the larger distribution dispersion in winter versus summer is a good indication of the greater amount of transients affecting Europe during the former season.

As mentioned previously, probabilistic skill has been assessed with the help of the economical value. This score gives the percentage of money gained by an end-user taking the seasonal forecast into account, versus climatological and perfect forecasts (score respectively equals to 0 and 100). It is based on a cost-loss model with a two-category contingency table, defined by a chosen threshold f for a particular climate variable. Let E be the event corresponding to the occurrence of the climate variable being above the chosen threshold f, defining the first category (the second being defined by the non-occurrence of E). If an end-user can take an action that costs C but which avoids a loss L when the first category occurs, a comparison of the average over the considered time period (here 44 yr) of the money spent according to three different strategies can be made. The expense associated with each combination of action/inaction and occurrence/non-occurrence of E is given in the decision-model contingency table shown in Table 2. The first strategy of action corresponds to the use of the climatological forecast, that is taking action if C/L is less than the chosen threshold and never otherwise, giving rise to a mean expense $E1$. The second consists of the use of actual forecast, that is taking action only when the probabilistic forecast of the first category is above a threshold t (different from f, to be determined by the end-user when optimizing the strategy), causing a mean expense $E2$. The last corresponds to the virtual perfect forecast, that is taking action only when the first category occurs, giving rise to a mean expense $E3$. The economical value is defined as $V = (E1 - E2)/(E1 - E3)$. The difference between $E1$ and $E3$ is maximum for C/L equals f. Therefore, as pointed out by Déqué (2001), one can simplify the evaluation by keeping only the case of $C/L = f$ for which potential gain from the forecast is maximum. The choice of the probability threshold t has been done here by excluding the target year when optimizing the strategy; the threshold t is determined by maximizing the economical value V, for each threshold f. Whereas the choice of the thresholds f has been made following the same methodology as in Déqué (2001), that is for each grid point, forecast type and reference, by computing the average plus α times the standard deviation over the 44 yr, α being the deciles of a Gaussian distribution. In particular, $\alpha = 0$ corresponds to a forecast of a positive versus a negative anomaly. Additionally, a reliability correction has been applied in the computation of the economical value V (see Déqué, 2001, for more details).

Table 2. Decision-model contingency table: expense associated with each combination of action/inaction and occurrence/non-occurrence of a given climate event E

		Occurs	
		Yes	No
Take action	Yes	C	C
	No	L	0

Figure 4 presents the economical value V for the tropical precipitation as a function of the threshold f, in both summer and winter, for the three types of forecasts. First, all values are positive except for the driest threshold of S forecasts, in both summer and winter, and of C in summer. In order to investigate the significance of this score, a scrambling of the years has been carried out to obtain an empirical distribution of the economical value

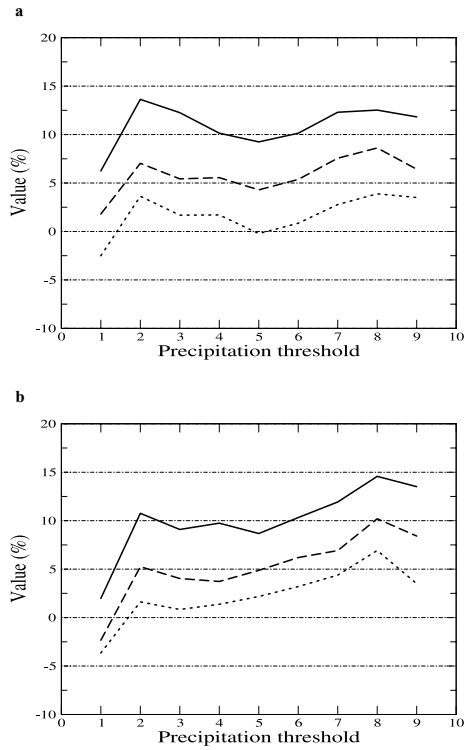

Fig 4. Economical value for the tropical precipitation as a function of the precipitation threshold in both JJA (a) and DJF (b). The observed SST forecast is denoted by the solid line, the coupled forecast by dashed line and the statistical SST forecast by the dotted line.

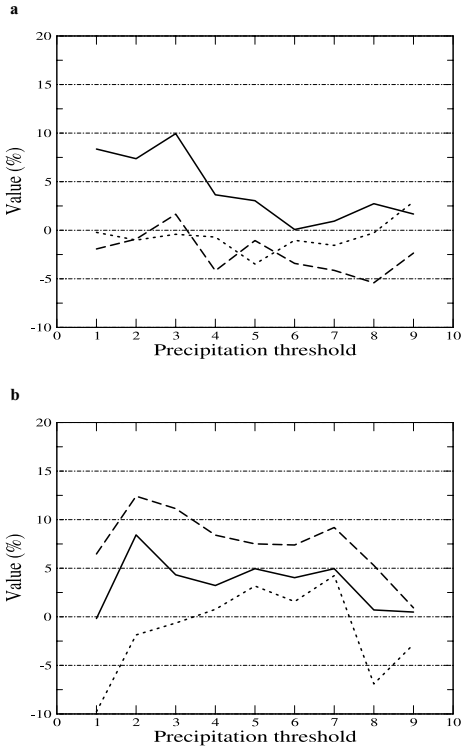

Fig 5. Economical value for the temperature at 850 hPa over Europe as a function of the temperature threshold in both JJA (a) and DJF (b). The observed SST forecast is denoted by the solid line, the coupled forecast by dashed line and the statistical SST forecast by the dotted line.

in the case of no skill (Déqué, 2001); the case of no skill corresponds to a value V equal to 0% (i.e. a forecast equivalent to the climatology). The upper boundary of the 95% confidence interval is below 1% in summer, 2.5% in winter, whatever the threshold and the type of forecasts (the largest value of the upper boundary being obtained for the eighth threshold), meaning that all positive economical values are statistically significant. Secondly, the hierarchy in the three types of forecasts is kept from that of the deterministic score. However, the gap of skill between C and O appears larger than that obtained using ACCs (see Table 1). This is due to the larger ensemble spread of C forecasts. This larger spread is intrinsically caused by the different SSTs for each member of the ensemble, but also by the ocean initialization strategy (see Section 2), which may exaggerate the spread, in particular for a small-size ensemble (nine members here).

Figure 5 presents the economical value V for T850 over Europe as a function of the threshold f, in both summer and winter, for the three types of forecasts. As expected, taking into account the deterministic scores (see Table 1), the economical values are much smaller than for the tropical precipitation. Moreover, only the O forecast shows a statistically significant score in summer on the cold side (thresholds 2, 3 and 5), while the C forecast is the only one that presents a statistically significant score in winter except on the coldest side (threshold 1), slightly cold (threshold 4) and warmest side (threshold 9). Nevertheless, as already mentioned, it will be shown, in the next section, that this apparent good result of the C forecast is not obtained for good physical reasons.

4. Atmospheric response patterns

The aim of this section is to investigate the physical causes giving rise to the scores described in Section 3, in particular to investigate why the coupled model gives better skill scores over Europe in DJF than the atmospheric model forced by observed SST. The methodology consists of three steps. First, covariances between SSTs and mean normalized anomalies of Prec and T850 over each respective region highlight forcing locations. Secondly, for the years with the largest observed normalized anomalies, composites of the atmospheric response are constructed to investigate the nature of predictability in the GCM in comparison to the observation. Thirdly, internal ensemble consistency is calculated (Feddersen, 2000); composites of Prec and T850 for the years of the largest consistency over the respective regions are compared to the previous ones in order to assess the GCM potential predictability.

Figure 6 shows the covariances between SSTs and mean normalized anomalies of T850 over Europe of the ERA-40 and the

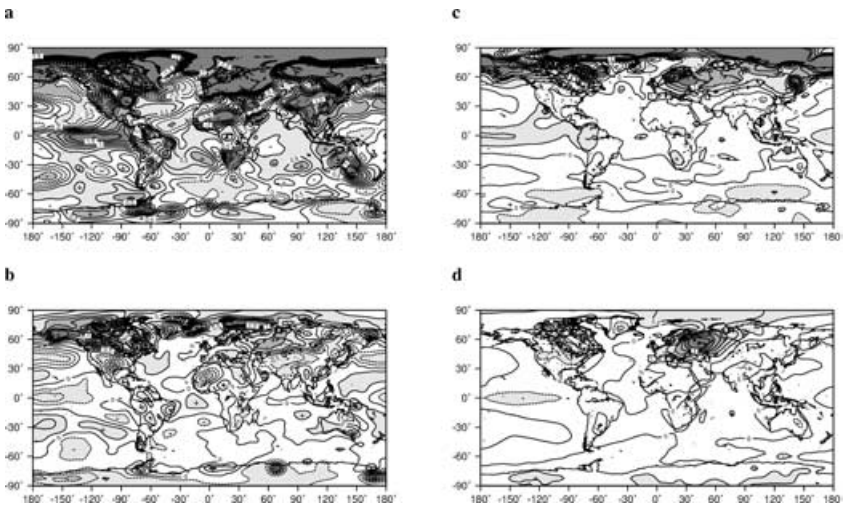

Fig 6. Covariances between SSTs and mean normalized anomalies of T850 over Europe of ERA-40 (a) and the three types of forecasts O (b), C (c) and S (d), in DJF. Contours every 0.25°C. A solid line represents positive values and dashed line negative values.

three types of forecasts for the winter season; the mean normalized anomalies of T850 over Europe have been computed as areal average seasonal mean anomalies divided by their standard deviation over the ERA-40 time period. These covariances correspond to the SST anomalies, related to mean normalized anomalies of T850 over Europe. For ERA-40, our reference, a well-known El Niño–La Niña SST anomaly (related respectively to positive–negative temperature anomalies over Europe) clearly shows up over the tropical Pacific; the northern Atlantic ocean is characterized by a meridional quadripole, the sign being opposite at 50°N between the ocean and the continent. For the same sign of European T850 anomaly, the O forecast presents an equatorial Pacific SST forcing opposite in sign compared to the reference, the Atlantic forcing being of the right sign but much less intense. This result means that the teleconnections between the Pacific Ocean and Europe are not working properly. The S forecast presents the same deficient equatorial Pacific SST forcing as O, while the Atlantic SST forcing is even more diffuse than for O. The C forecast is the only one that shows the right sign of the equatorial Pacific SST forcing. This apparent good result will be discussed in the following. As far as the summer season is concerned (not shown), the reference presents the same type of SST forcing pattern, that is an El Niño–La Niña SST anomaly pattern (opposite in sign to the winter situation, i.e. El Niño forcing corresponding to negative temperature anomaly over Europe; more precisely, the sign over western Europe being in fact opposite as it is over the rest of this continent during summer) and a less extended meridional quadripole (also opposite in sign) over the northern Atlantic. Regarding this SST forcing pattern, the three types of forecasts behave better than in winter.

Having investigated the SST forcing patterns knowing the target regions, the reverse exercise will be developed in the following. More precisely, considering the ERA-40 SST forcing patterns, composites of the atmospheric response have been constructed. For the seven largest and lowest mean normalized anomalies, mean atmospheric fields have been computed, keeping in that process 14 among the 44 yr of the considered period. In order to assess the linearity of forcing response, ERA-40 composites of surface temperature linked to T850 mean normalized positive, negative and total anomalies over Europe in winter are shown in Fig. 7. The three SST patterns are very similar, with the previously mentioned El Niño–La Niña SST anomaly in the Pacific and the northern Atlantic meridional quadripole. This result (i.e. quasi-linear relationship between the forcing and the response) constitutes a justification of the chosen composite approach. The difference between the composite linked to T850 mean normalized total anomalies (Fig. 7) and the covariance (Fig. 6) is only due to the weight factor applied to the considered field, plus or minus one in the case of the composite and the mean normalized anomaly itself in the case of the covariance; in the latter case, it increases the influence of the extreme values of the mean normalized anomalies.

Figure 8 presents the surface temperature composites of C and S linked to the ERA-40 T850 mean normalized anomalies over Europe, in winter. Compared to the observed SST pattern (Fig. 7), there is an overall good agreement; nevertheless, the C forecast shows a less intense El Niño-type anomaly, whereas the S forecast El Niño-type anomaly has a smaller meridional extension. Over the northern Atlantic Ocean, the meridional quadripole is badly reproduced by the C forecast, with the S forecast presenting a better shape. Having looked at the SST field, atmospheric response will be investigated considering successively precipitations, geopotential at 500 hPa (hereafter called Z500) and finally T850.

Figure 9 shows the precipitation composites of ERA-40, O, C and S linked to ERA-40 T850 mean normalized anomalies over Europe in winter. Concerning ERA-40, in relation to the El Niño-type SST anomaly (Fig. 7), a precipitation positive anomaly shows up over the equatorial Pacific. The O forecast anomaly is insufficiently intense around 120°W, certainly due to the incapacity of the present convection scheme to handle deep convective processes in this region. The S forecast anomaly presents the

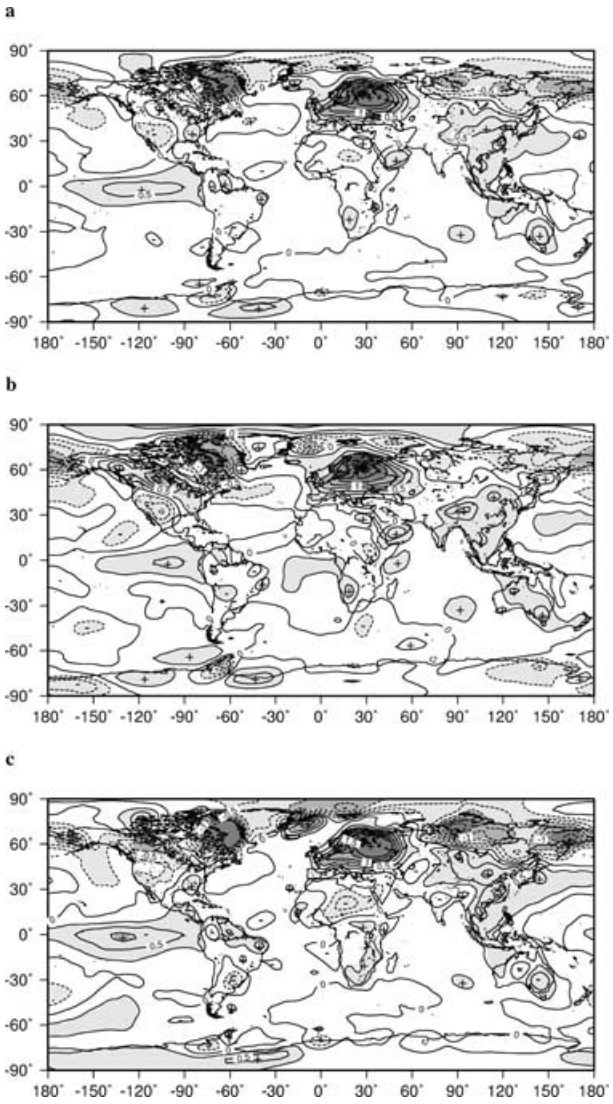

Fig 7. ERA-40 composites of surface temperature linked to T850 mean normalized (a) total, (b) positive and (c) negative (with a change of sign) anomalies over Europe in DJF. Contours every 0.25°C. A solid line represents positive values and dashed line negative values.

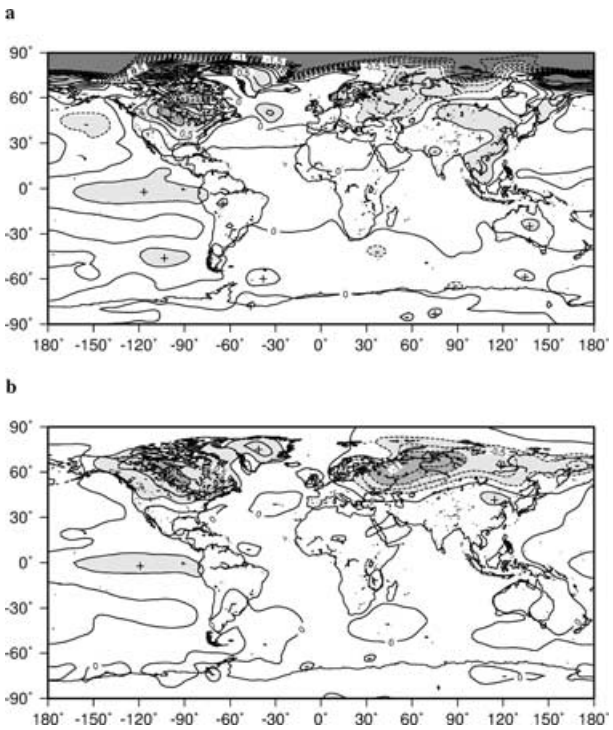

Fig 8. Surface temperature composites of C (a) and S (b) forecasts linked to ERA-40 T850 mean normalized anomalies over Europe in DJF. Contours every 0.25°C. A solid line represents positive values and dashed line negative values.

same type of discrepancy, while the anomaly produced by the C forecast is less intense over the whole equatorial Pacific, consistently with the less intense SST anomaly. In order to obtain a hint of the impact of equatorial deep convection on the mid-latitudes, Fig. 10 shows Z500 composites of ERA-40, O, C and S linked to ERA-40 T850 mean normalized anomalies over Europe in winter. The ERA-40 composite is characterized by a very nice Pacific North American (PNA) pattern, starting on both sides of the equator at around 120°W, and then extending over the US with a negative anomaly on the west, followed by a positive one on the eastern side. This pattern, which is almost identical to that obtained by Gill (1980) using a simple linear model, can be understood in terms of large-scale stationary Rossby waves.

The eastern US positive anomaly extends on the Atlantic and Europe, resulting in a nice North Atlantic Oscillation (NAO) pattern over the Atlantic Ocean. The O and S forecasts present almost the same pattern. On one hand, the first negative anomaly north of the equator is situated to the west of the ERA-40 location, consistent with the fact that the precipitation anomaly is too weak around 120°W (as indicated above). On the other hand, the second positive anomaly around 70°W is situated to the north of the ERA-40 location, due to a non-optimum choice of parameters in the gravity wave drag scheme and in the semi-Lagrangian advection scheme causing too zonal flow notably up and downstream of the Rockies. Finally, the western extension over Europe of the positive anomaly originating on the eastern side of the US is not present, here also due to the too zonal southerly flow simulated by the AC GCM. As a result of the overall less intense convection over the equatorial Pacific pointed out in the C forecast, there is, first, a better location on the west coast of the US of the first negative anomaly and, secondly, a less intense negative anomaly over the central Atlantic responsible for the appearance of a positive anomaly over Europe (tending also to slightly appear in the S forecast). T850 anomalies over Europe are closely related to those of Z500, the sign being the same, as presented in Fig. 11, where only ERA-40, O and C forecasts are depicted. Therefore, the C forecast presents the best pattern of T850 anomalies, with the same sign as ERA-40 but only over Western Europe.

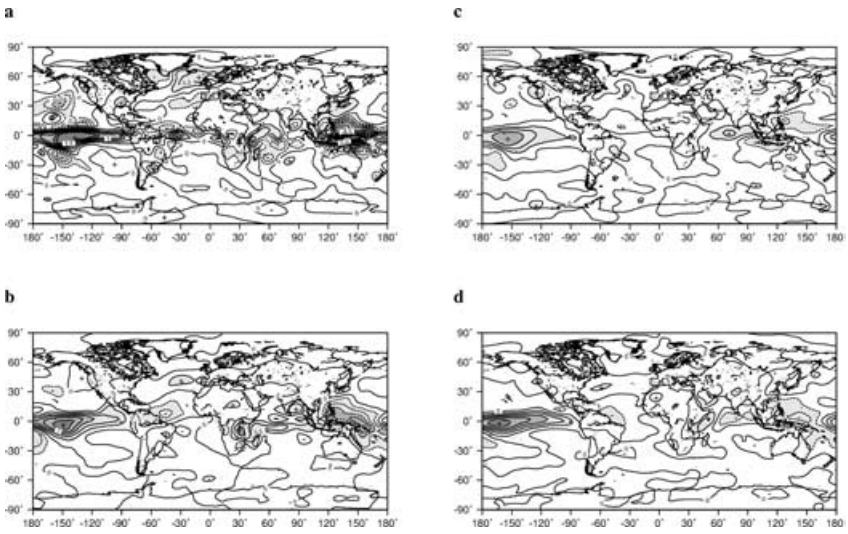

Fig 9. Precipitation composites of ERA-40 (a), O (b), C (c) and S (d) linked to ERA-40 T850 mean normalized anomalies over Europe in DJF. Contours every 0.25 mm d^{-1}. A solid line represents positive values and dashed line negative values.

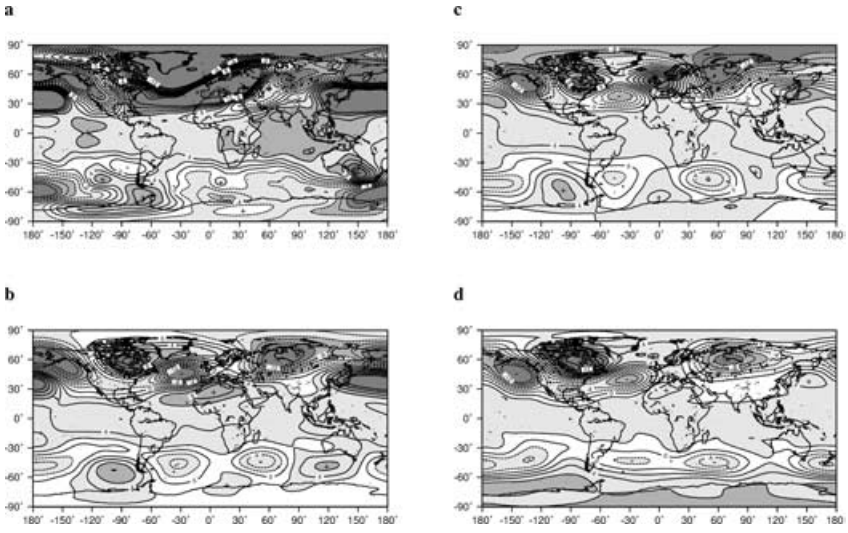

Fig 10. Z500 composites of ERA-40 (a), O (b), C (c) and S (d) forecasts linked to ERA-40 T850 mean normalized anomalies over Europe in DJF. Contours every 2 m. A solid line represents positive values and dashed line negative values.

This fact explains the apparent best T850 ACC over Europe (see Fig. 3), originally due to the less intense El Niño-type SST anomaly (Fig. 8) caused by a larger intra-ensemble dispersion. Interestingly, ERA-40 T850 anomalies over Europe (Fig. 11) show an opposition of sign between the Mediterranean area and the rest of the continent north of it, as was the case for the surface temperature (Fig. 7). This pattern is very similar to that obtained in earlier studies (Fraedrich and Müller, 1992).

Figure 12 shows the surface temperature composite of ERA-40 linked to ERA-40 T850 mean normalized anomalies over Europe, in summer. As indicated at the beginning of this section, one can see an El Niño–La Niña SST anomaly pattern (opposite in sign to the winter situation, i.e. El Niño forcing corresponding to negative temperature anomaly over Europe; more precisely, the sign over western Europe being in fact opposite as it is over the rest of this continent, zonal sign opposition versus meridional in winter, as can be seen in Fig. 11) and a less extended meridional quadripole (also opposite in sign) over the northern Atlantic, with the largest anomaly (twice larger than in winter) located around 50°N, being of same sign than over Europe (except, the western edge). In the same manner as for winter, the three types of forecasts present a less intense precipitation anomaly around and to the east of 120°W, which turns out to shift toward the west the mid-latitude stationary Rossby wave response (not shown). This wave pattern is characterized by a wavenumber of 4, while being 3 in winter. As a consequence of the westward shift over the eastern Pacific Ocean, the Rossby wave pattern is different from that of ERA-40 downstream over the US and the Atlantic sector. Certainly due to the large observed SST anomaly around 50°N (see Fig. 12), which is not reproduced in any of the predicted SST forecasts C and S, the O forecast shows a corresponding T850 anomaly close to the

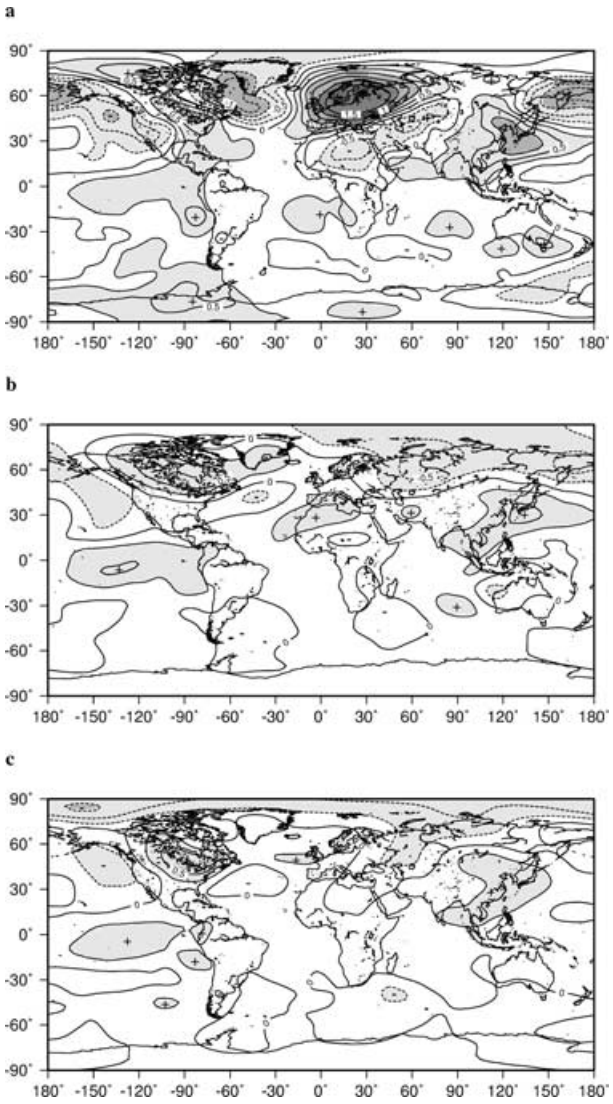

Fig 11. T850 composites of ERA-40 (a), O (b) and C (c) forecasts linked to ERA-40 T850 mean normalized anomalies over Europe in DJF. Contours every 0.25°C. A solid line represents positive values and dashed line negative values.

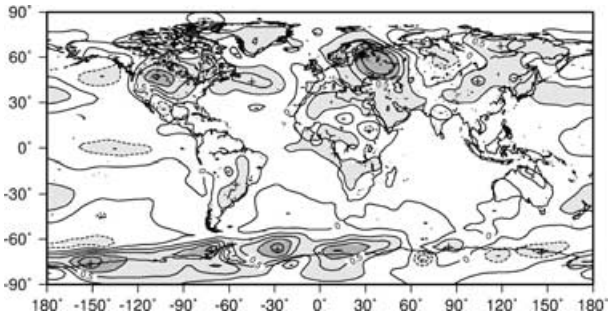

Fig 12. Surface temperature composites of ERA-40 linked to ERA-40 T850 mean normalized anomalies over Europe in JJA. Contours every 0.25°C. A solid line represents positive values and dashed line negative values.

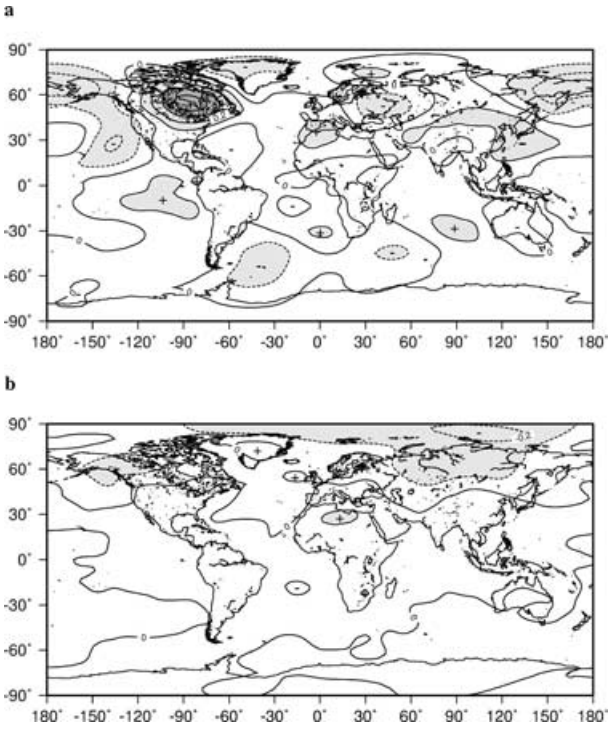

Fig 13. T850 composites of O (a) and C (b) forecasts linked to the largest consistency over Europe in DJF. Contours every 0.25°C. A solid line represents positive values and dashed line negative values.

analysed one. Because of this rather good behaviour over the Atlantic, the O forecast is able to produce a T850 anomaly close to that of ERA-40, but only over southern Europe.

Beyond the comparison of the GCM forecasts to the analyses, it is relevant in the context of seasonal forecasting to assess the capability of the GCM to compare to itself in the case of similar SST forcings (i.e. perfect model approach). The goal of this approach is to enable an a posteriori correction of the GCM under the prior condition of high internal ensemble consistency. It is this consistency condition that has been studied in the following. The GCM consistency has been computed following the work of Feddersen (2000). Having nine-member ensembles, the consistency has been computed as a mean ACC from the 36 possible pairs of the ensemble members. The larger the ACC, the higher the consistency will be.

Composites of Prec and T850 for the years of the largest consistency over the respective regions have been built. The 14 first years with the largest consistency (first tercile of the distribution) have been kept in order to be able to compare to previous composites discussed above. Figure 13 shows these composites for T850 linked to the largest consistency over Europe, for the O and C forecasts in winter. Interestingly, the patterns of Fig. 13 are very similar to those of Fig. 11, even if not exactly the same years have been used in each of the composites. To get a hint of

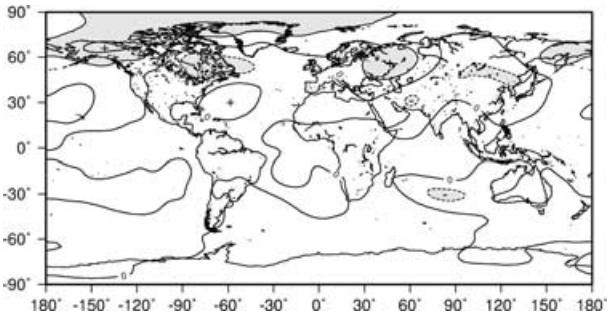

Fig 14. T850 composites of O forecasts linked to a medium consistency over Europe in DJF. Contours every 0.25°C. A solid line represents positive values and dashed line negative values.

the significance of this result, Fig. 14 presents the composite for T850 linked to a medium consistency over Europe (i.e. taking into account the 14 yr corresponding to the second tercile of the distribution), for the O forecast only in winter. The anomalies are much smaller than in Fig. 12. Thus, the atmospheric response pattern obtained during the years for which the consistency is the largest over Europe is very close to that obtained during the years for which the observed anomalies are the largest over Europe. This means that the GCM is able to respond the most consistently (i.e. showing its ability to extract a consistent signal out of the noise) over Europe to the SST forcing pattern giving rise to the largest observed anomalies over the same region. This result demonstrates a rather good GCM potential predictability, and further the relevance of the MOS approach to correct the GCM products. Such an approach has been recently investigated by Clark and Déqué (2003). For a considered season predicted by the GCM, it consists of the search of analogues in past GCM seasonal forecasts to obtain the corresponding ensemble of observed years as the actual forecast.

5. Conclusion

Assessing the predictability of seasonal simulations carried out with AC, it has been shown, as expected from earlier studies, that the skill comes from the ability of the GCM to respond properly to SST anomalies. In real-time forecasting, these anomalies can be predicted using either a statistical model or a coupled ocean–atmosphere model. The skill of the latter is larger than that of the former, from both deterministic and probabilistic point of views, according to the results obtained over the ERA-40 period (1958–2001). For a four-month forecast, the SST statistical model is not able to maintain the intensity of the anomalies. Moreover, the coupled forecast skill is not far from that of the observed SST forecasts, which one cannot carry out in real time, but which provide an upper limit of the skill. This is particularly true for the tropics. The mid-latitude ocean is affected both by the ocean model local errors and by the atmospheric remote errors, as seems to be the case for the Atlantic Ocean during summer.

The GCM atmospheric response to SST anomalies is obviously sensitive to the physics in the tropics and to the dynamics in the mid-latitudes. AC tends to shift the response to SST anomalies in the equatorial central Pacific toward the west and simulates a too zonal flow in the Northern Hemisphere around 50°N, which turns out to distort the teleconnection pattern over the Northern Hemisphere. However, AC shows a rather good potential skill in the sense that the years characterized by the largest internal ensemble consistency give an atmospheric response very similar to that obtained during the years for which the observed anomalies (tropical precipitation and temperature at 850 hPa over Europe) are the largest.

Our results highlight at least two areas of research needed to improve the skill of seasonal forecasts: first, the model itself, for which improvements are necessary in both the physics (diffusion and convection) and the dynamics (this work is presently in progress concerning AC); secondly, the model post-processing, for which robust statistical algorithms are to be designed to overcome the GCM systematic errors.

6. Acknowledgments

The authors are grateful to two anonymous reviewers for their useful comments and suggestions. This work was supported by the European Union Programme Energy, Environment and Sustainable Development under contract EVK2-CT-1999-00197.

References

Brankovic, C. and Palmer, T. N. 2000. Seasonal skill and predictability of ECMWF PROVOST ensembles. *Q. J. R. Meteorol. Soc.* **126**, 2035–2067.

Brier, G. W. 1950. Verification of forecasts expressed in terms of probabilities. *Mon. Wea. Rev.* **78**, 1–3.

Clark, R. T. and Déqué, M. 2003. Conditional probability seasonal predictions of precipitation. *Q. J. R. Meteorol. Soc.* **129**, 1–15.

Déqué, M. 2001. Seasonal predictability of tropical rainfall: probabilistic formulation and validation. *Tellus* **53A**, 500–512.

Déqué, M., Dreveton, C., Braun, A. and Cariolle, D. 1994. The ARPEGE/IFS atmosphere model: a contribution to the French community climate modelling. *Clim. Dyn.* **10**, 249–266.

Doblas-Reyes, F. J., Déqué, M. and Piedelièvre, J. P. 2000. Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Q. J. R. Meteorol. Soc.* **126**, 2069–2087.

Feddersen, H. 2000. Impact of global sea surface temperature on summer and winter temperatures in Europe in a set of seasonal ensemble simulations. *Q. J. R. Meteorol. Soc.* **126**, 2089–2109.

Fraedrich, K. and Müller, K. 1992. Climate anomalies in Europe associated with ENSO extremes. *Int. J. Climatol.* **12**, 25–31.

Gill, A. E. 1980. Some simple solutions for heat-induced tropical circulation. *Q. J. R. Meteorol. Soc.* **106**, 447–462.

Graham, R. J., Evans, A. D. L., Mylne, K. R., Harrison, M. S. J. and Robertson, K. B. 2000. An assessment of seasonal predictability using atmospheric general circulation models. *Q. J. R. Meteorol. Soc.* **126**, 2211–2240.

Madec, G., Delecluse, P., Imbard, M. and Levy, C. 1997. OPA, release 8, Ocean General Circulation Model reference manual. LODYC internal report (available from LODYC/IPSL, France).

Madden, R. A. and Julian, P. R. 1972. Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.* **28**, 1109–1123.

Palmer, T. N. and Anderson, D. L. T. 1994. The prospects for seasonal forecasting—a review paper. *Q. J. R. Meteorol. Soc.* **120**, 755–793.

Palmer, T. N., Brankovic, C. and Richardson, D. S. 2000. A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Q. J. R. Meteorol. Soc.* **126**, 2013–2033.

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M. et al. 2004. Development of a European multi-model ensemble system for seasonal to inter-annual prediction (DEMETER). *Bull. Am. Met. Soc.* **85**, 853–872.

Shukla, J. 1998. Predictability in the midst of chaos: a scientific basis for climate forecasting. *Science* **282**, 728–731.

Simmons, A. J. and Gibson, J. K. 2000. The ERA-40 Project Plan. ERA-40 Project Report Series. ECMWF, Shinfield Park, Reading, RG2 9AX, UK.

Terray, L., Sevault, E., Guilyardi, E. and Thual, O. 1995. OASIS 2.0, user's guide and reference manual. CERFACS Technical report (available from CERFACS, France).