Q. J. R. Meteorol. Soc. (2001), **127**, pp. 2761–2786

Empirical forecasts of 850 hPa air temperature anomalies over the North Atlantic

By E. SÁNCHEZ GÓMEZ, F. ÁLVAREZ GARCÍA and M. J. ORTIZ BEVIÁ*
Universidad de Alcalá, Madrid, Spain

(Received 12 October 2000; revised 31 May 2001)

SUMMARY

The empirical forecasts presented here were developed as a benchmark for dynamical forecasts provided by uncoupled and coupled general-circulation models. The predictand field is the variability of 850 hPa air temperature anomalies over the North Atlantic basin at timescales greater than eight months. The predictor field is chosen among different datasets (sea surface temperature anomalies, the predictand field itself, etc.). Some of the experiments make use of a single predictor field, whereas others use several. The performance of the forecasts, measured by the correlation between observed and forecast fields, is assessed against the skill of forecasts performed on a simple persistence basis.

The statistical scheme used to make the forecast is based on singular-value decomposition (SVD) of the cross-covariance matrix between the predictor and the predictand field at the lead requested for the forecast. Prior to the SVD calculations, we remove with a filter the variability of predictor and predictand fields at timescales below eight months. Forecasts of air temperature anomaly have been made over the period from 1970 to 1998, at different lead times. The highest skill is achieved when the predictors are computed from anomalies of sea surface temperature over roughly the same domain as the predictand. The skill of the forecast is important south of 45°N and is broadly consistent through the three decades. About 56% of the variance of the eight-month filtered predictand field (36% of the variance of the original field) can be forecast nine months ahead. The best values of the forecast skill appear in two regions, one in the subtropics and the other around the Iberian peninsula. Only in this last region, and also in another region over the USA, does the skill of the forecasts beat the skill of persistence.

In order to identify the origin of the skill we study its dependence on the target season. We also investigate its dependence on the phase of the El Niño–Southern Oscillation (ENSO) and the North Atlantic Oscillation. The skill is, in first place, related to the ability of the scheme at capturing the variability at biennial and ENSO-related timescales.

KEYWORDS: Long-range forecasting Mid-latitude predictability Statistical methods

1. INTRODUCTION

Not long ago, empirical models were the only tool available for forecasting anomalies of the atmospheric circulation at lead times greater than two months. Basically, all empirical forecasts take advantage of the relationship between large-scale spatial patterns of the anomalies of some climatic variable and certain timescales of variability, as pointed out in the seminal work of Namias (1953). Empirical forecasts of the state of El Niño–Southern Oscillation (ENSO) show skills that are above those of some intermediate models and are comparable with the skills of some coupled models (Barnston and Ropelewski 1992; Penland and Magorian 1993). The skill of empirical forecasts for the mid latitudes is only modest (Davis 1976; Barnett *et al.* 1984; Barnett and Preisendorfer 1987; Shabbar and Barnston 1996; Johansson *et al.* 1998; Vautard *et al.* 1998). Also, they offer no prospect for dramatic improvements similar to the potential offered by coupled general-circulation models (GCMs). Some improvement can, nevertheless, arise from the use of new statistical techniques benefiting from longer and better versions of existing datasets, or from the incorporation of new observations as additional predictors.

The empirical forecasts presented here were intended as a benchmark to assess the GCM forecast skill in a European Union funded project entitled 'The prediction of climate variability at seasonal timescales (PROVOST)' (Palmer and Shukla 2000). Empirical forecasts can also contribute to the improvement of forecast skill of coupled models (Webster and Palmer 1997). They can highlight relationships among the predictors and

* Corresponding author: Departamento de Física, Universidad de Alcalá, C. Barcelona, Km. 33.6, Alcalá de Henares, Madrid, Spain. e-mail: mjose.ortiz@uah.es
© Royal Meteorological Society, 2001.

predictand fields. They can also improve forecasts by correcting model bias for some specific variable (Feddersen *et al.* 1999). Finally, they can contribute to extending the lead time of the dynamical forecasts, which may be limited by the available computing resources. In the empirical forecast scheme presented here, we incorporate both better predictor and predictand fields and a statistical procedure that, although more than twenty years in use as an analytical tool, has not been employed for climatic forecasting until the start of the PROVOST project.

The main spatial and temporal scales of the interannual variability of the predictand field are those recognizable in the evolution of any other atmospheric variable, and can be characterized by the state of the North Atlantic Oscillation (NAO). Although the NAO was identified long ago (Walker and Bliss 1932), the interest of the scientific community in it has been reawakened by a number of observational and modelling studies (see, for instance, Rodwell *et al.* (1999)). The connection between the NAO and the ENSO (the major interannual signal at planetary scale), as investigated by van Loon and Rogers (1977), has also received renewed attention.

The basis for long-range weather forecasting in the North Atlantic was laid out by Ratcliffe and Murray (1970), who pointed out the lagged correspondence between the North Atlantic sea surface temperature and the European pressure field. On the other hand, the predictability of the atmospheric flow in the North Atlantic has been targeted by numerous modelling studies. For instance, Palmer and Sun (1985) have shown that (i) a strong ENSO signal results in enhanced predictability in our domain of interest, and also that (ii) anomalously warm sea surface temperatures (SSTs) in the western North Atlantic over a large region centred to the south and south-east of Newfoundland act to shift the jet stream and storm tracks to the north, producing anomalously high surface pressure over the North Atlantic.

The empirical forecasts presented here take advantage of the availability of newer oceanic and atmospheric datasets. Longer and better datasets of the atmospheric circulation have progressively been made available through the NCAR*–NCEP† reanalysis project (Kalnay *et al.* 1996) for the three years of the PROVOST project. An SST dataset was built for the purposes of this forecast by combining the SST data of the Comprehensive Ocean and Atmosphere Datasets (COADS release 2) covering only up to 1993, with the Integrated Global Observation Services System (IGOSS) data with observations up to 1999. Details of the datasets used for the predictand and for the predictor fields are given in section 2.

A comparative assessment of the advantages of the statistical technique used in the present work (singular-value decomposition (SVD)) can be found in the papers by Bretherton *et al.* (1992) and Navarra (1993). A summary of the technique and of our forecast scheme is given in section 3.

In section 4 we present the layout of the forecast experiments. Then we describe in detail the results of the experiment that scores the highest skill; this is the one whose predictor is the SST anomalous field in the North Atlantic. Additional experiments, which were carried out in the hope of improving the forecasts in higher latitudes, are only reviewed briefly.

In section 5, the results are analysed and discussed. Forecasts were produced for the three decades from 1970 through to 1998. The performance of our scheme at forecasting a significant cooling in the mid 1970s and a persistent warming in the first part of the 1990s is presented as a case study. The dependence of the skill on the targeted season, and on the ENSO or NAO phase, is also considered. Finally, the singular vectors used in

* National Center for Atmospheric Research.
† National Centers for Environmental Prediction.

the forecast are clustered into a few classes. An analysis of the most important among these vectors, and their timescales of variability, allows further insight into the sources of the forecast skill.

2. Data and data treatment

In the forecasts presented here, the anomalous atmospheric circulation is represented by the 850 hPa air temperature anomalies (the predictand field). With respect to other atmospheric variables, this field has several advantages, apart from its interest for the public. Rain, an obvious candidate in this sense, was discarded because it is not very well simulated by GCMs. Air temperature at 850 hPa has the advantage of being closely connected to sea-level air temperature, without being sensitive to the different topographies of the GCMs. Therefore, air temperature anomalies became a common target for all the forecasts in the PROVOST project. In the case of the empirical models, air temperature at 850 hPa was also a convenient variable because it is known to be well represented by the NCAR–NCEP reanalysis.

The choice of the predictand domain was based on two factors. On the one hand, our forecasts needed to be relevant for the European climate. Additionally, the rationale that underlies all empirical forecasts is that the ocean, with its slow timescales, constitutes the principal source of long-term atmospheric predictability. Therefore, the domain was originally chosen as the North Atlantic basin, from 90°W to 10°E, and from 20°N to 70°N. Later, to improve the forecasts, the northern boundary was moved up to 90°N. The region selected includes large part of coastal Europe to the east, and part of the US and Canadian coast to the west. It also includes the Caribbean, a region of tropical–extratropical interactions.

Monthly-mean air temperature anomalies at 850 hPa were derived from the NCAR–NCEP reanalysis dataset* (Kalnay *et al.* 1996) for the period 1958–1998. The anomalies were computed with respect to the 41-year climatology.

Originally, our forecasts considered only one predictor field, the SST anomaly. This is a natural predictor for atmospheric variables if one wants to compare empirical predictions with those obtained with forced atmospheric models. As mentioned above, we built the predictor dataset for the required period (1950–1999) by merging the COADS (1950–1993) (Woodruff *et al.* 1987) and the IGOSS† (1982–1999) datasets (Reynolds and Smith 1994). Both datasets are global, but their grid resolutions are different (2° × 2° for the COADS against 1° × 1° for the IGOSS dataset) and the grid points do not coincide; for the new dataset, the IGOSS observations were interpolated to the COADS grid. Additionally, some points from the COADS grid were discarded when quality was lacking in the neighbouring IGOSS observations (mostly near the coasts). From these monthly values, anomalies were computed in the same way as for the predictand field. This SST field can be updated continuously, and therefore we can supply forecasts for the next year without having to wait for the release of analysed fields (COADS, HadISST‡ or IGOSS). However, the updated SST field is quite similar to the available analysed SST fields, as indicated by Fig. 1, which compares the root-mean-squares of our SST field and of the HadISST data (Rayner *et al.* 2000).

* This dataset was provided by the NOAA–CIRES Climate Diagnostic Center, Boulder, Colorado, USA, from their web site at http://www.cdc.noaa.gov/

† These data were downloaded from the IGOSS web site http://ingrid.ldgo.columbia.edu

‡ This dataset was provided by the Hadley Centre for Climate Prediction and Research, Met Office, Bracknell, Berkshire, UK.

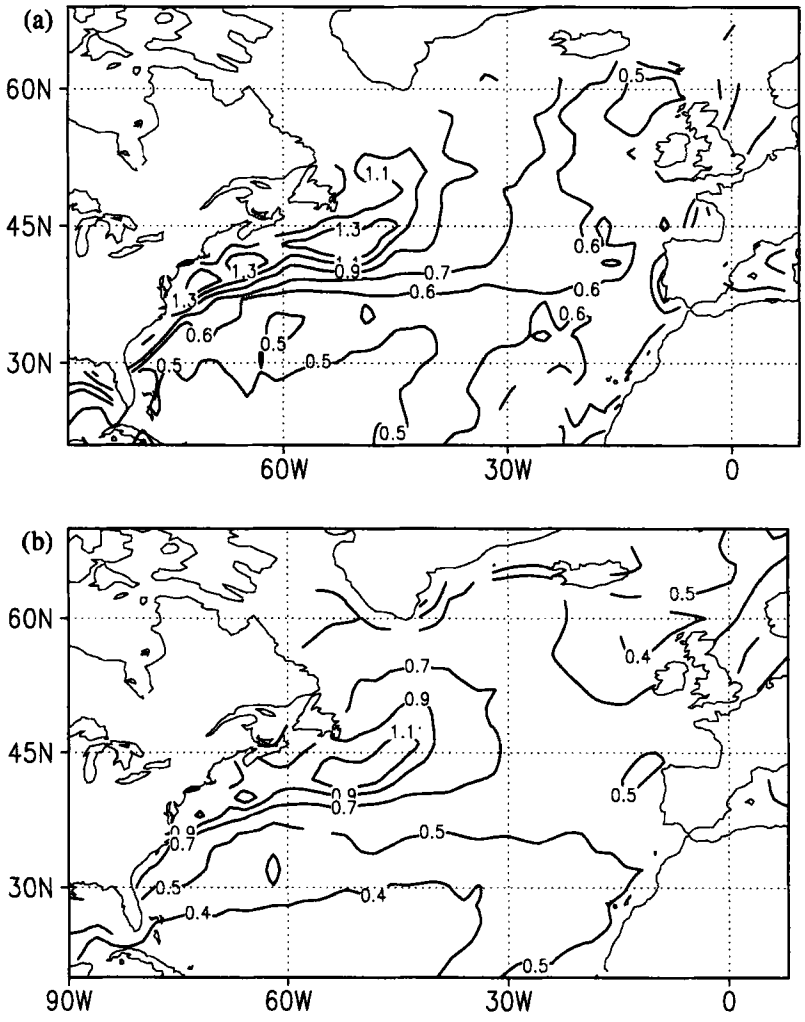

Figure 1. (a) The root-mean-square of the sea surface temperature (degC) obtained by merging the IGOSS and COADS datasets (1950–1999). (b) The root-mean-square of the sea surface temperature from the HadISST dataset for the same period. See text for further information.

Once a predictor variable has been selected, the choice of the most appropiate spatial domain has to be decided. A global domain has the advantage of focusing on the large-scale patterns, that are associated with longer timescales. But it can also result in a decrease in the forecast skill if many of the time series in the predictor field are not actually related to the predictand.

Several forecast experiments have been carried out, and they can be classified into three groups according to the predictor variable. In the first set, the SST anomaly has been used as a predictor. Differences between the experiments of this set lie in the predictor domain. In a first case, the predictor domain was chosen to coincide roughly with that of the predictand (90°W–10°E, 20°N–90°N). In another case, the SST anomaly in the North Pacific (120°E–130°W, 20°N–90°S) was used as the predictor field. In the second set of experiments various atmospheric fields, in the same domain as the predictand (North Atlantic basin), played the role of predictor field. Among them we have the predictand field itself (autoprediction), the sea-level pressure (SLP)

anomaly and the 700 hPa dynamic height anomaly (considered a good predictor in previous empirical forecasts for northern Europe (Johansson *et al.* 1998)). The last set of experiments used a multiple-predictor scheme, combining two of the data fields used in the previous single-predictor experiments.

To study the origin of the skill, we also considered two climate indices. The Niño3 index was obtained by averaging SST anomalies in the Niño3 region (150°W–90°W, 5°S–5°N). The NAO index was calculated from monthly values of SLP of the NCEP reanalysis dataset (previously filtered to remove the variability below eight months) as the difference between Lisbon (Portugal) and Stykkisholmur (Iceland), according to Hurrell (1996).

Another NAO index, computed from air temperature anomalies averaged over the region (25°W–2.5°W, 37.5°N–45°N) minus those of the region (25°W–12.5°W, 62.5°N–70°N), is also used in this paper. This index characterizes an important climatic signal in the forecast field.

3. Statistical model and forecast scheme

We can use several statistical methods to study the coupled variability between two fields. Compared with some of them (such as canonical correlation analysis) SVD is less biased towards the dominant modes of variability yielded by each field separately. Instead, SVD extracts the coupled features between the fields, even when these represent a small part of the total variance of each one (Bretherton *et al.* 1992).

Let x_j and y_j be, respectively, the values of the predictand and the predictor field at the point j of the spatial grid and at time t ($t = 1, \ldots, N_t$). For a forecast at a given lead l, we use the SVD of the cross-covariance matrix between predictor and predictand field at lag l.

This cross-covariance matrix can be written in the form

$$c_{ij}(l) = \sum_{\tau=l+1}^{N_t} y_i(\tau - l)x_j(\tau) = \sum_{k=1}^{p} \sigma_k u_{ik} v_{kj}, \tag{1}$$

where u_{ik} and v_{kj} are the components of the spatial singular vectors ($\mathbf{u}_k$ and $\mathbf{v}_k$) of the predictand and the predictor field, respectively, and σ_k are the singular values of the covariance matrix between the fields at lag l associated with a singular mode k. Further details about the statistical procedure have been given by Navarra (1993) and Wallace *et al.* (1992). Because the patterns can be ordered decreasingly by their corresponding singular value, yielding the amount of the coupled covariance explained at lead l, a reduction in the number of degrees of freedom is then straightforward. Through the projection of each field onto the corresponding vector we obtain the empirical coefficients ($a_k(t)$, $b_k(t)$) of the time evolution of each field. In our predictive scheme, the predictor evolution in time is related to that of the predictand by these empirical coefficients:

$$a_k(t) = c_k b_k(t - l), \tag{2}$$

and, therefore, the relationship between the predictor and predictand field can be written as

$$x_{mj}(t) = \sum_{k=1}^{p} c_k b_k(t - l) v_{kj}, \tag{3}$$

where the c_k coefficients are obtained by least-square minimization through the whole length of the training sample and x_{mj} is the forecast value for the x_j field at time t.

4. FORECAST EXPERIMENTS: LAYOUT AND SKILL

The scheme followed here for the forecast is represented as a diagram in Fig. 2. It is a combination of: (i) a preliminary filtering in time; (ii) the procedure described in the previous section; (iii) a spatio-temporal filtering; and (iv) the final forecasting stage, as illustrated in Fig. 3.

(i) Prior to each forecast, predictor and predictand time series are filtered removing from the training sample the variability at timescales below eight months. The time filter employed is a convolution filter. If r is the smoothing function and s_k the time series to be filtered, the discrete convolution of these functions is

$$s'_j = (r * s)_j = \sum_{k=-\frac{L}{2}+h(j)}^{\frac{L}{2}+h(j)} s_{j+k} r_k, \tag{4}$$

where

$$h(j) = \frac{L}{2} - (1-j) \qquad 1 \leqslant j \leqslant \frac{L}{2},$$

$$h(j) = 0 \qquad \frac{L}{2} < j < N - \frac{L}{2},$$

$$h(j) = -\frac{L}{2} + (N-j) \qquad N - \frac{L}{2} \leqslant j \leqslant N,$$

and $j = 1, \ldots, N$, L is the width of the smoothing function, and h (non-zero only near the borders) allows the window to become asymmetric there. In this way we avoid including any information about the future in the training sample. Predictor and predictand fields are split into a training sample and a validation period. To make a real forecast, the time filter is applied exclusively to the training interval. The need for such filtering is indicated in Fig. 4; in the averaged atmospheric spectra there is an important and noisy variability at timescales below eight months, unmatched in the oceanic averaged spectra. Our interest in forecasting at lead times of at least two seasons makes such filtering advisable so as to focus on the longer temporal and spatial scales. As mentioned in section 2, other atmospheric fields have been used as predictors—even the predictand itself; although the timescales of variability of the predictor and predictand fields are similar in this case, they have also been filtered in a similar way for the sake of methodological consistency.

(ii) In a statistical forecast, the parameters of the model are determined from the training sample. However, the assumptions required by the classical theory of estimation (i.e. linearity) might not be a satisfactory hypothesis for long records of data such as ours, which actually arise from nonlinear dynamics. It may prove convenient, then, to regard the time series as 'piecewise linear', in which case the training sample should not be too long. In the present work, a number of experiments have been carried out by varying the length of the training sample from 30 years to 6 years (see subsection 4(c)). The results point to a 12-year period as being the optimal training sample size. Validation is carried over the whole ensemble of forecast values, that is from the first forecast in January 1970 to the last one in December 1998.

(iii) The spatio-temporal filtering makes use of the expansion of the predictor and predictand fields in the basis provided by the singular vectors of each field. By truncation, we retain only a few singular vectors that together are supposed to explain most of the joint variability of both fields. To determine the optimal number of vectors to be retained, the most common and simple criterion is based on the

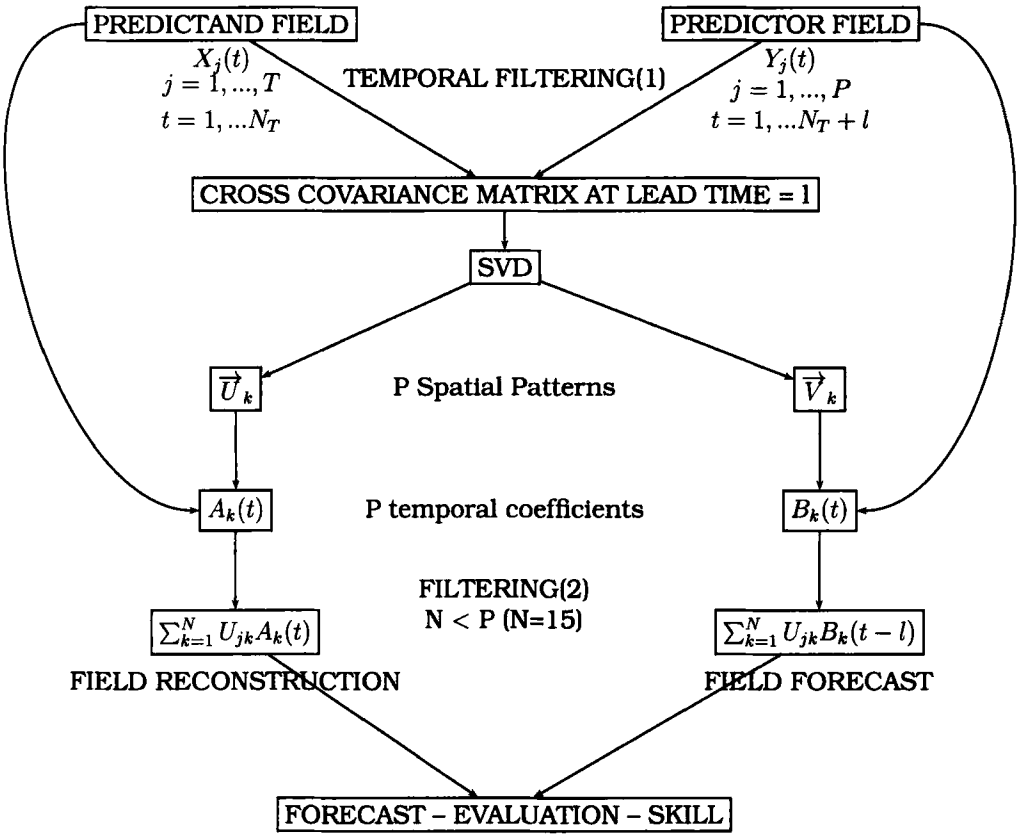

Figure 2. Representation of the forecast scheme described in the text. The procedure is repeated whenever a forecast at time t is issued.

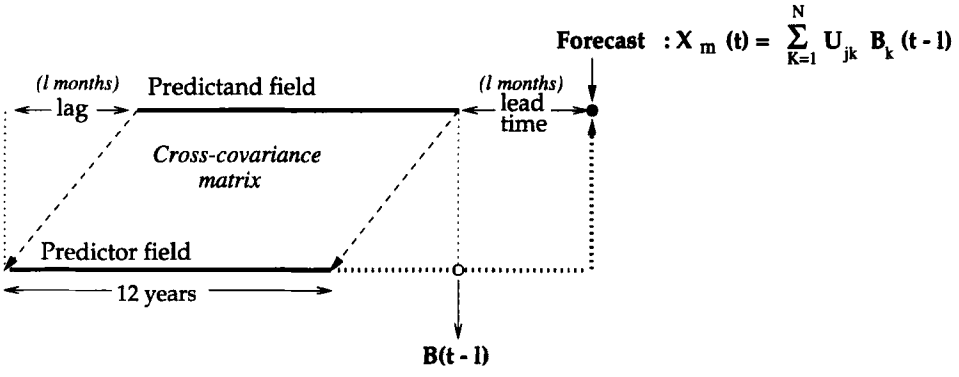

Figure 3. Schematic representation of the fourth stage in the forecast process. This illustrates how the training period is used to produce the forecasts.

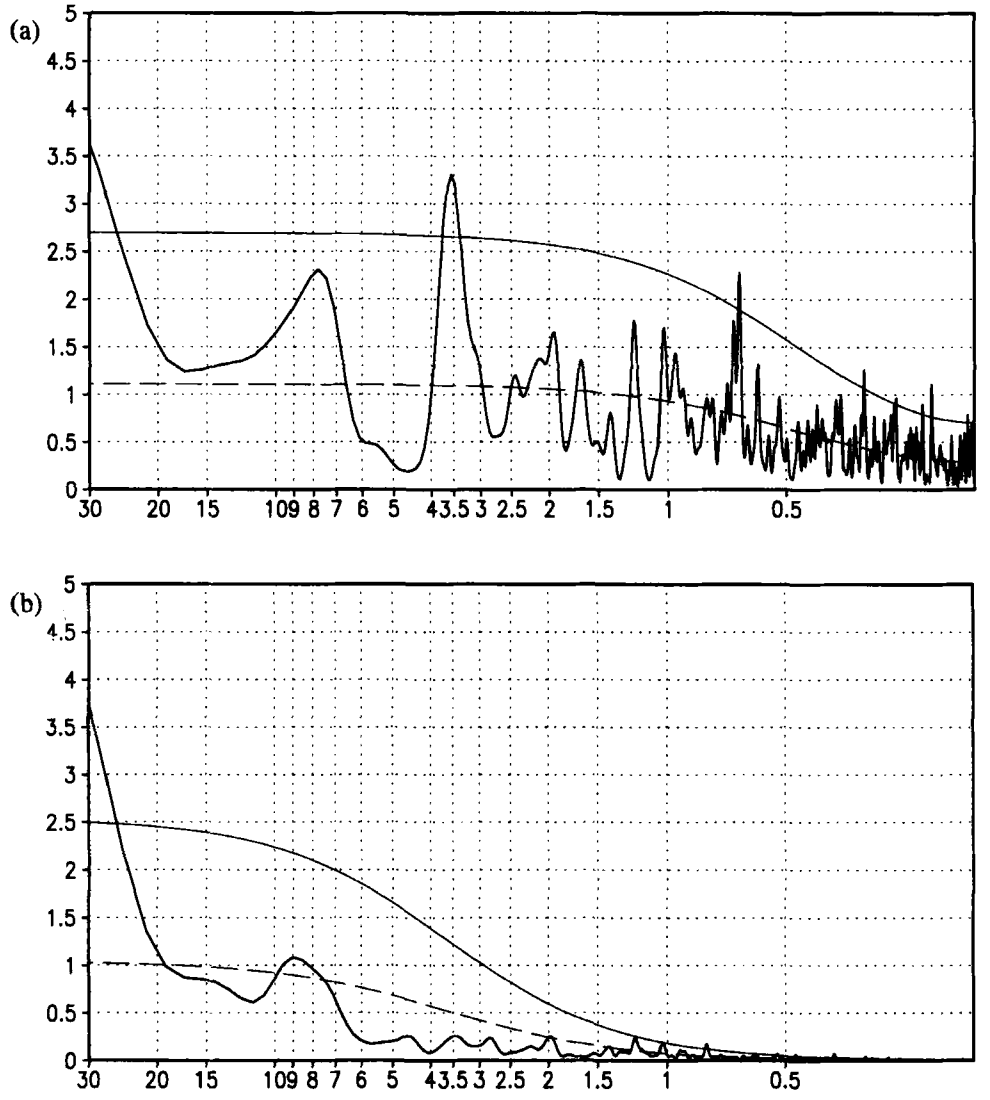

Figure 4. (a) The power spectra of the 850 hPa air temperature anomalies averaged over the band (90°W–10°E, 35°N–65°N). (b) The power spectra of the sea surface temperature anomalies averaged over the same domain. The horizontal axis is labelled in years.

behaviour of the variance of the field reconstructed with different numbers of vectors. The truncation order n should be such that an increase in the number of vectors used in the reconstruction does not yield a significant increase in the explained variance. In the present work, we also take into account how well major climatic signals are represented in the filtered field. Figure 5 shows the fraction of variance explained by the predictand field as a function of the number of singular vectors retained, and the correlation between the NAO index computed from the unfiltered predictand field and the NAO index from the filtered predictand field as a function of the number of singular vectors used in its reconstruction.

In all of the forecasts, the expansion has been truncated after the 15th term, for both predictor and predictand fields ($n = 15$). The forecast, issued at time t_i, will then consist

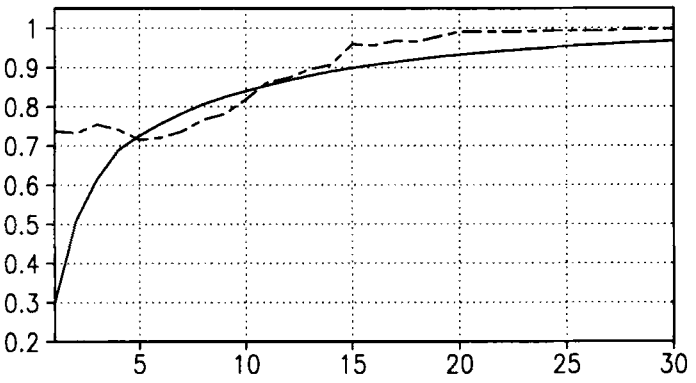

Figure 5. The cumulative variance explained by the predictand field as a function of the number of singular vectors retained (full line), and the correlation between the North Atlantic Oscillation (NAO) index and the index calculated from different anomalous temperature-field reconstructions, against the number of singular vectors retained for the reconstruction (dashed line). The NAO index is computed as the difference between the 850 hPa air temperature anomalies averaged over the regions (25°W–2.5°W, 37.5°N–45°N) and (25°W–12.5°W, 62.5°N–70°N).

of 15 values for the 15 coefficients needed to calculate the predictand field at time $t_i + l$, together with the retained singular vectors.

Finally, it is worth mentioning that no predictor interfield weighing has been carried out in any of the forecasts. The respective number of grid points of SST and 850 hPa air temperature determines their natural weights.

(iv) The validity of our forecast procedure is assessed in terms of the correlation between the twice-filtered observed field and the forecast one. This correlation S_r is a measure of the forecast skill and is defined by:

$$S_r = \frac{\sum_{i=1}^{M} x_o(t_i + l) x_m(t_i + l)}{\sum_{i=1}^{M} x_o^2(t_i + l) \sum_{i=1}^{M} x_m^2(t_i + l)}, \tag{5}$$

where x_o is the twice-filtered predictand field obtained from the observations, x_m the forecast field, l is the lead time, and summation extends over the M forecasts issued for each experiment. Values of the skill must be above the zero-confidence interval—this can be roughly estimated at the 95% significance level by $S_s = 2/\sqrt{M}$ (Jenkins and Watts 1968), for values of M (the number of forecast experiments considered) greater than 20. We call this threshold value S_s the significant skill. Also, because the number of spatial patterns used to forecast is not small, the problem of an artificial skill exclusively due to the large number of predictors has to be considered. This artificial skill has been estimated from a big number (100) of Monte Carlo forecast experiments. For each of these, a different synthetic predictor field has been used. These fields are a 'proxy' for the North Atlantic SST anomalies, assuming that the evolution in time of this variable can be modelled by a Langevin equation (Frankignoul and Hasselmann 1977):

$$\frac{dy_t}{dt} = -\beta y_t + w_t \tag{6}$$

where w_t is a driving noise, supposed to be white, that represents the atmospheric forcing. The forecast experiments have been carried out along the lines of the real

forecast experiments presented above. Each of them uses the same value of the feedback parameter β at each grid point (estimated from the observations), but uses a different realization of the forcing noise. The average skill of these experiments provides an estimation of the artificial skill S_a. When presenting the skill of the real forecast experiment, regions with values below this artificial skill at the 95% confidence limit have been shaded. In any case, it is accepted that a forecast may be useful only if values of the skill exceed 0.6.

To assess the relevance of the forecasts, at the present 'state of the art' for the public, we estimated the similarity between the annual means of the air temperature field at one location and the annual mean of the forecasts at the same place. This similarity was measured by the 'linear error in probability space' (LEPS) scores (Potts *et al.* 1996). For a start, to compute the LEPS scores between two fields, the observed and forecast situations are reduced to three categories (above, normal or below). The agreement or disagreement in each year between the observed and the forecast categories is given by their product and is accumulated in a contingency table. Afterwards a (3 × 3) matrix **T** is built from this table. This matrix is multiplied by a scoring matrix

$$\mathbf{A} = \begin{bmatrix} 1.35 & -0.15 & -1.20 \\ -0.15 & 0.30 & -0.15 \\ -1.20 & -0.15 & 1.35 \end{bmatrix}$$

that weights more heavily the anomalous situation that we are interested in predicting, as compared with the normal ones. The resulting score is divided by the score obtained for perfect forecasts for the same year (Vautard *et al.* 1998). The mathematical expression for the LEPS score is:

$$S = \frac{\sum_{i,j=1}^{3} a_{ij} T_{ij}}{\sum_{i=1}^{3} a_{ii} Q_i}, \tag{7}$$

where a_{ij} and T_{ij} are the components of the matrices **A** and **T**, and Q_i denotes the total number of observed occurrences of the category i which quantifies the value of the forecasts in terms of information theory. LEPS scores can take values from −1 (for a totally wrong forecast) to +1 (for a totally correct one).

(*a*) *Forecasts with the North Atlantic SST as the predictor field*

Forecasts of the filtered 850 hPa air temperature anomalies over the North Atlantic at different leads have been produced for three decades starting in January 1970, using the SST anomaly as the predictor field. The skill of these forecasts at a nine-month lead for each of the decades, is represented in Figs. 6(a), (c) and (e). Figure 6(c) shows how the skill worsens considerably during the 1980s. Nevertheless, over the three decades values of the skill remain useful (above 0.6) over two regions, one in the subtropics and the other near the Iberian peninsula. Another key region in our domain is the area around Iceland where values of the skill, though positive, are useless (except for the 1980s decade, when the skill over Iceland is quite good). Additionally, we consider the region of lowest skill in the north-east, near the Labrador sea, to characterize the forecasts. The differences between the forecast skill achieved with our technique and the one produced under the assumption of persistence are also shown in Figs. 6(b), (d) and (e).

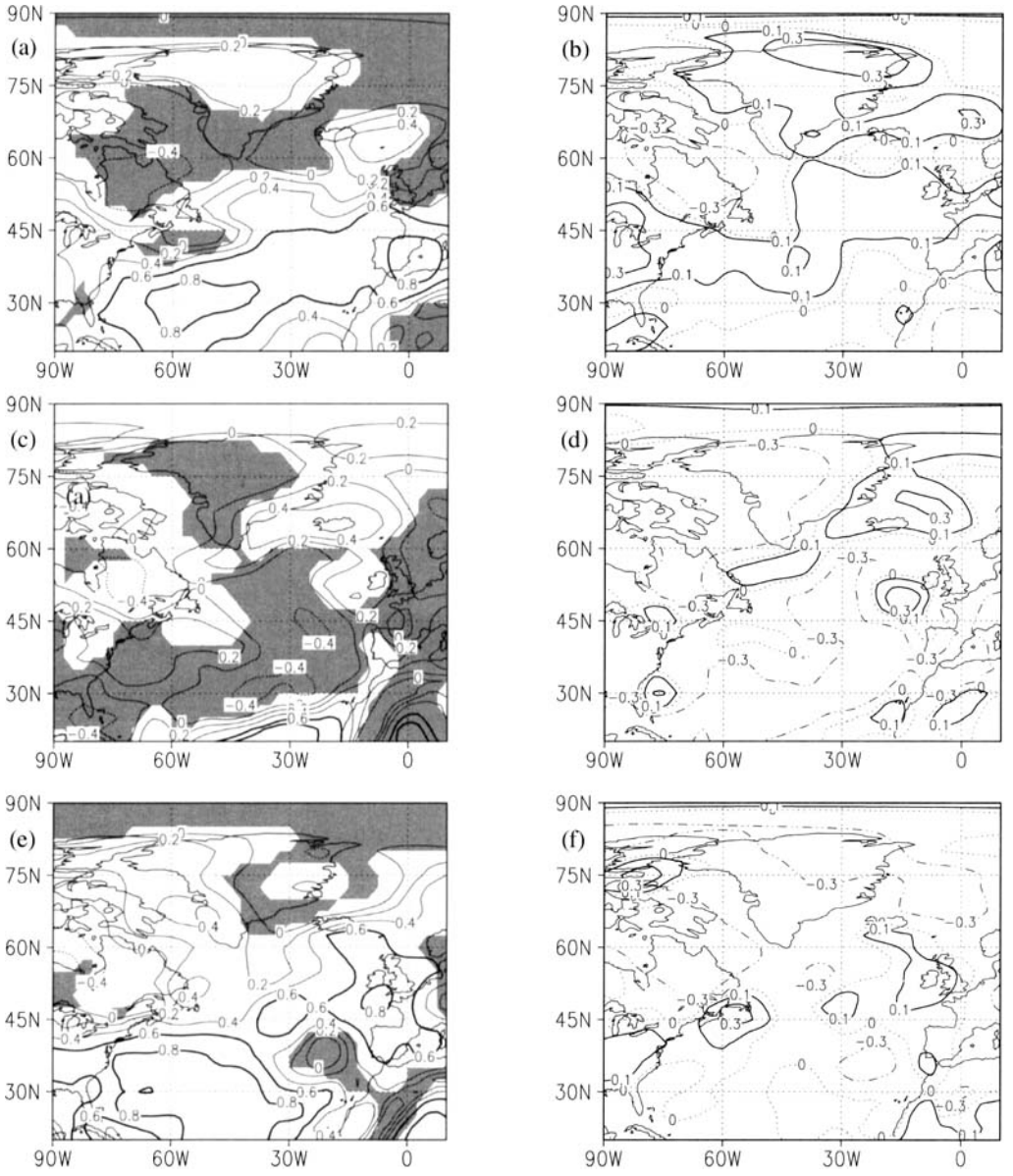

Figure 6. The spatial distribution of the correlation skill of the forecasts of 850 hPa air temperature anomalies over the North Atlantic region during (a) the 1970s, (c) the 1980s and (e) the 1990s (the regions where skill is below the artificial skill are shaded). As control measure the difference between the skill obtained with our model and the skill of persistence is presented in (b), (d) and (f) for the same decades (persistence is defined as the predictand autocorrelation at the lag time required for the forecast).

The spatial distribution of the skill for the entire forecast period 1970–1998 is displayed in Fig. 7(a) against the difference between it and the skill of the persistence forecasts (Fig. 7(b)). The spatial pattern in Fig. 7(a) resembles the first heterogeneous correlation map at a nine-month lead (not shown) calculated from the first coefficient ($b_1(t)$) in the expansion of the predictor field and the whole predictand field (Wallace *et al.* 1992; Bretherton *et al.* 1992). This map indicates those grid points where the SST

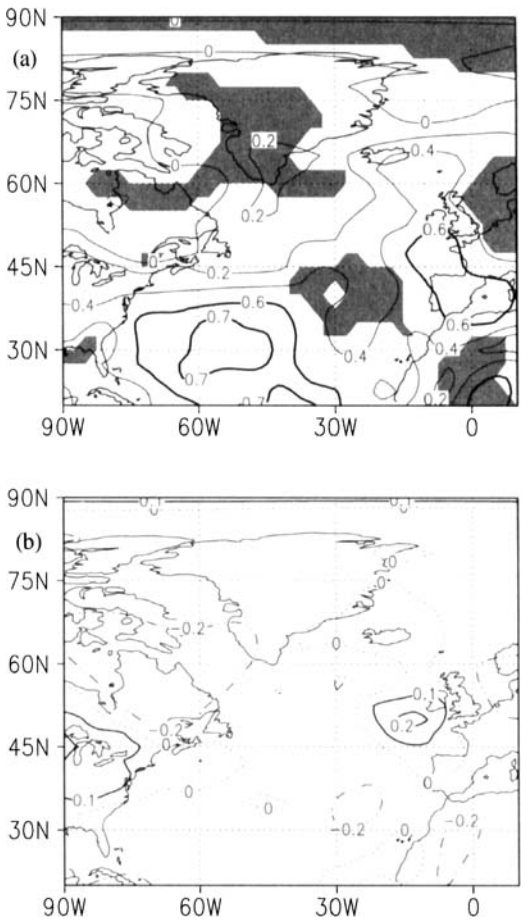

Figure 7. (a) The spatial distribution of the correlation skill of the forecasts of 850 hPa air temperature anomalies over the North Atlantic region for the entire forecast period (1970–1998) (regions whose skill is below the values estimated for the artificial skill are shaded). (b) The difference between the skill obtained with our model and the skill persistence (persistence is defined as the predictand autocorrelation at the lag required for the forecast).

field is more skilful at predicting the air temperature anomalies. The pattern in Fig. 7(a) is also similar to the spatial distribution of the predictand field variance. The regions where the skill is useless (the north and north-west of the domain) are connected with the short timescales that have been removed with the preliminary filter.

Values of the pattern in Fig. 7(a) do not differ greatly from those obtained in a cross-validation experiment. In this experiment, all the available years of observation except the forecast one were included in the training sample. Consequently, the filter has been modified to avoid including information about this year. The cross-validation skill and its persistence control appears in Fig. 8. The scores of the cross-validation skill are higher since both future and past information is used in the forecast.

Figure 9 shows the dependence of the skill averaged over the four characteristic domains on the forecast lead. These domains are selected in the following way: Domains I (75°W–30°W, 20°N–40°N) and II (30°W–0°W, 35°N–50°N) are the regions where the forecasts perform best, centred on Bermuda and around the Iberian peninsula, respectively. Domain III (25°W–10°W, 60°N–70°N) is a region around Iceland, included

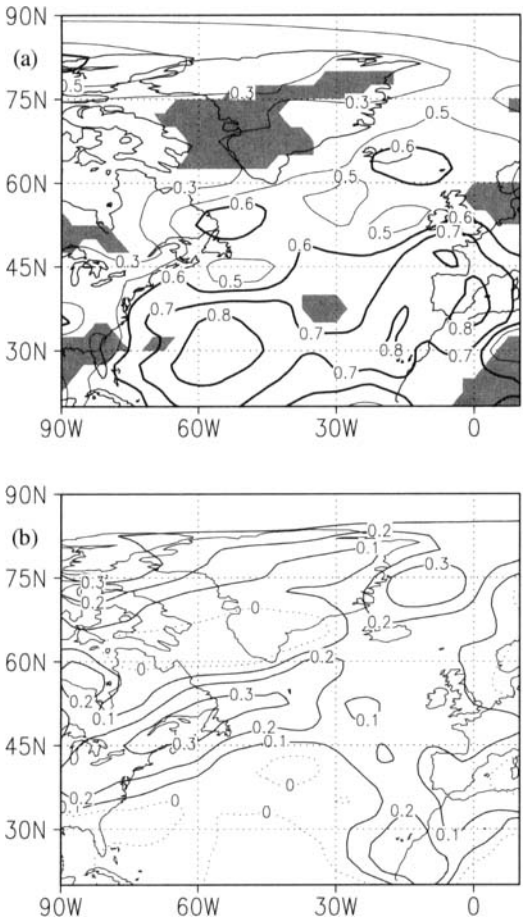

Figure 8. As Fig. 7, but for the cross-validation experiment (see text). In this experiment all the available observations, except the year that includes the targeted month, were used as the training sample (regions where the skill is not significantly different from the artificial skill are shaded).

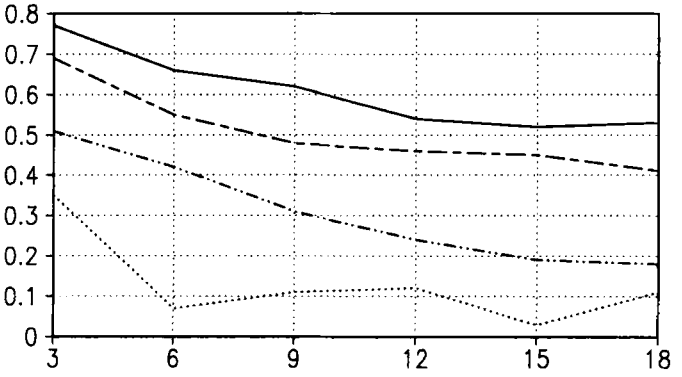

Figure 9. The dependence of the skill averaged over the four spatial domains on the forecast lead: domain I (full line), domain II (long–short dashed line), domain III (dot–dot–dashed line) and domain IV (dotted line). The horizontal axis is labelled in months.

TABLE 1. VALUES OF THE CORRELATION AND THE LINEAR-ERROR IN PROBABILITY SPACE (LEPS) SCORES BETWEEN THE ANNUAL MEANS OF THE ORIGINAL (UNFILTERED) AND THE NINE-MONTH-LEAD FORECAST FIELDS

	Correlation coefficient	LEPS score
Domain I	0.53	0.46
Domain II	0.37	0.31
Domain III	0.21	0.10
Domain IV	−0.17	0.21

In both cases the values of the skill are averaged over the following regions:
Domain I = (75°W–30°W)(20°N–40°N);
Domain II = (30°W–0°W)(35°N–50°N);
Domain III = (25°W–10°W)(60°N–70°N);
Domain IV = (85°W–55°W)(45°N–65°N).

because of its importance for the NAO index. Domain IV (85°W–55°W, 45°N–65°N) corresponds to a region over Canada (near Newfoundland), with the lowest skill values. Forecasts are useful up to nine-months lead in domain I and up to three months ahead in domain II.

Some complementary experiments have also been carried out to investigate the sensitivity of the skill to the training sample size. Altogether, the optimal length is 12 years but there are geographical differences. In domains I and II, the best skill is obtained for small training samples sizes (12 years or even shorter), while for domains III and IV, better skill is obtained from training sizes between 16 and 22 years. The LEPS scores computed, as explained above, are given in Table 1 for the four domains.

(*b*) *Forecasts with other predictor fields*

In an attempt to improve the forecast skill over the north and north-west of the domain, we have sought alternative predictors. Thus, experiment 2 probes into the influence of the northern Pacific anomalous state on the North Atlantic atmospheric evolution. On the other hand, interannual variability in the region might be primarily linked to the internal variability in the northern region of the domain, and be only weakly related to the SST predictor field. To test this hypothesis several atmospheric fields have been used as predictors: 850 hPa air temperature (AT) anomalies over the North Atlantic domain (experiment 3), over the northern Pacific domain (experiment 4), and over North America (experiment 5), and SLP anomalies over the North Atlantic basin (experiment 6). We have also experimented with low-pass filtered versions of two predictor fields, the SST anomaly and the 850 AT anomaly; these are experiments 7 and 8, respectively. Forecasts were also produced using as a predictor the North Atlantic 700 hPa geopotential height anomaly field (experiment 9). The predictand field remained the same through all the experiments. In Table 2 we present a summary of the different experiments at nine months lead and compare performances in terms of the mean forecast skill averaged through the four selected regions described above. Values of the skill of the cross-validation procedure for these regions appear as experiment 1^{+}.

Whereas all of those mentioned above were single-predictor experiments, combinations of two predictor fields were also employed at a further stage. The combined fields and the resulting average skill in each of the domains are also summarized in Table 2.

TABLE 2. VALUES OF THE SKILL FOR THE DIFFERENT FORECAST EXPERIMENTS WITH A NINE-MONTH LEAD TIME

Experiment	Predictor field	Domain I	Domain II	Domain III	Domain IV
1	North Atlantic SSTA	0.62	0.48	0.31	−0.10
1^+	North Atlantic SSTA	0.73	0.74	0.60	0.35
2	North Pacific SSTA	0.56	0.52	0.16	0.11
3	North Atlantic ATA 850HPa	0.58	0.48	0.14	0.06
4	North Pacific ATA 850HPa	0.51	0.40	0.19	0.05
5	North American ATA 850HPa	0.48	0.46	0.14	0.12
6	North Atlantic SLPA	0.54	0.41	0.12	0.05
7	North Atlantic LF ATA 850HPa	0.53	0.41	0.15	0.06
8	North Atlantic LF SSTA 850HPa	0.60	0.42	0.23	0.01
9	North Atlantic HGTA 700HPa	0.58	0.39	0.26	0.03
1+2		0.58	0.47	0.24	0.04
1+3		0.54	0.45	0.16	0.02

Experiments 1 to 9 used a single predictor field and experiment 1^+ used the cross-validation procedure. Two last experiments used a combined predictor field. The predictor variables are: SSTA (sea surface temperature anomalies); ATA (air temperature anomalies); LF (low-frequency air–sea surface temperature anomalies, the band-pass filtered data retaining periods from 10–20 years); SLPA (sea-level pressure anomalies); and HGTA (geopotential height anomalies). The values of the skill (correlation between the predictand field and its forecast) are averaged over the domains defined in the footnote to Table 1.

It is clear from the table that, among all the possible predictors, the SST anomalies in the North Atlantic score the highest skill values in all the regions except domain 4. In this last region, the atmospheric predictors (850 hPa AT anomalies, the same AT anomalies low-passed, and the 700 hPa geopotential height) do slightly better though, in general, the use of these predictor fields does not lead to significant skill improvements.

(c) *Two case studies: The cold winter of 1976 and the warm summer of 1994*

In order to analyse the forecasts supplied by our model, we turn now to two 'case studies': the winter of 1976 and the summer of 1994. To study these periods the predictand's evolution was monitored by four indices, obtained from air temperature anomalies averaged over the four regions mentioned above. In the western subtropics (domain I) and near the Azores region (domain II), air temperature anomalies are characterized by a persistent cooling, from the beginning of the forecast interval up to the start of the 1980s, followed by a prolonged warming thereafter. To study these case studies, seasonal means (e.g. for winter (December, January and February), and so on) of the raw predictand, the filtered predictand, the predictor and the forecast field are computed.

The persistent cooling in the low latitudes of the domain started at the end of the 1960s. A very strong negative anomaly was formed in the central North Atlantic, and persisted there for almost three years (from spring 1968 to autumn 1970). At the end of this period, the negative anomalies were displaced towards the American coast and subsequently weakened. In the period 1972–1973 a positive anomaly appeared, with maximum in the region of the Gulf Stream separation. The anomalous situation still remained in winter 1976. This is the anomalous SST pattern (Fig. 10(d)) that we use to forecast the air temperature anomalies nine months later (Figs. 10(c)). The unfiltered SST anomalies observed during this period are shown in Fig. 10(e). The filtered and unfiltered target observations are also displayed in Figs. 10(a) and (b).

The prolonged warming of the early 1990s in the lowest latitudes of the domain seems to have started early in 1991. There was a diffuse warming along the North American coast with a negative anomaly in the north (Labrador Sea). The positive

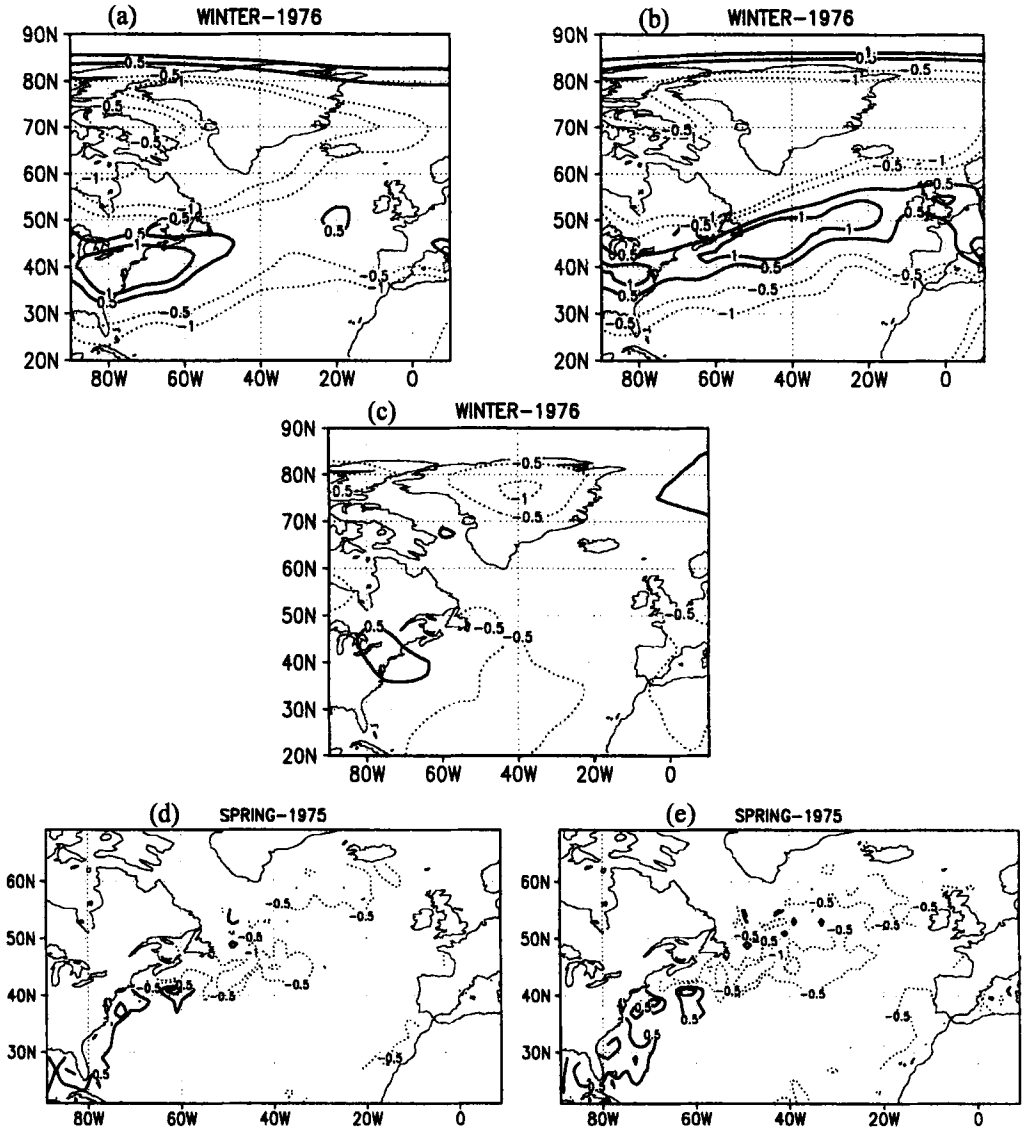

Figure 10. A case study: The forecast in the winter of 1976 against the observations. The observed 850 hPa air temperature anomalies for (a) the filtered and (b) the unfiltered observations, (c) the forecast 850 hPa air temperature anomalies, and the sea surface temperature (SST) anomalies for (d) the filtered and (e) the unfiltered observations used in the forecast. The SST anomalies lead the 850 hPa air temperature anomalies by nine months.

anomaly gave rise to a subtropical warming. At the end of 1992, a negative anomaly appeared south of 40°N making the positive anomaly propagate eastward, up to the coast of the Iberian peninsula and the Mediterranean Sea, where a warming appeared in the summer of 1994. We have illustrated the SST anomalous pattern used in the forecast (Fig. 11(d)), the corresponding unfiltered SST anomalies (Fig. 11(e)), the observed filtered and unfiltered 850 hPa air temperature anomalies (Figs. 11(a) and (b)) and the forecast (Fig. 11(c)). Notice that, although the general evolution of the anomalies is well captured, the forecast values are too weak.

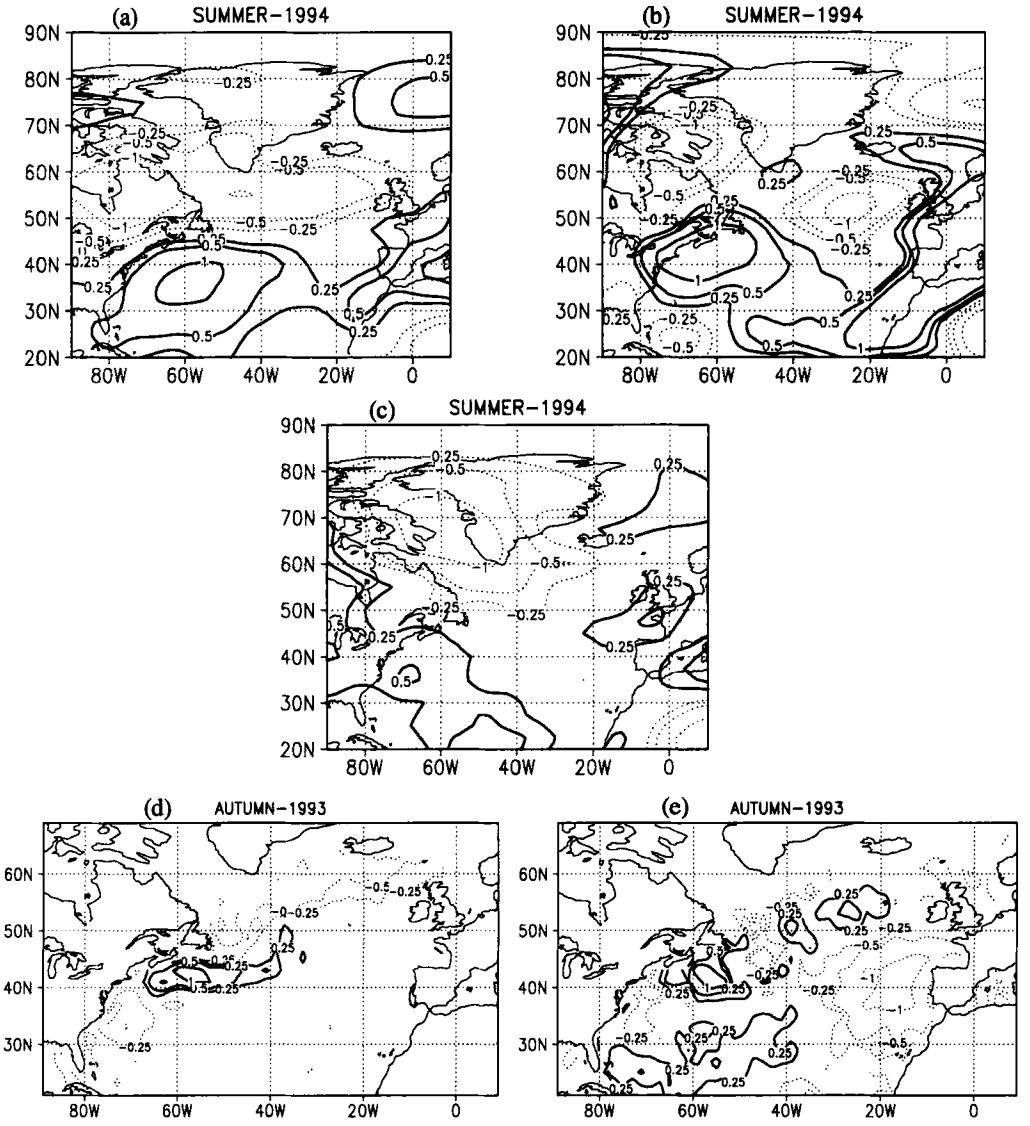

Figure 11. As Fig. 10, but for the forecast in the summer of 1994.

5. RESULTS

We analyse here the values of the skill obtained for three decades of forecasts of the 850 hPa air temperature anomalies in the North Atlantic basin with North Atlantic SST as the predictor field. Spatial, as well as temporal (seasonal, decadal), characteristics of the skill are examined. In addition, by an analysis of all the singular vectors used in the production of the forecasts, we sought physical insight into the atmospheric and oceanic processes leading to skilful forecasts.

(*a*) *Seasonal dependence of the skill*

We have investigated the seasonal dependence of the forecast skill. Results can be summarized as follows: in domain I autumn and winter are the best forecast seasons

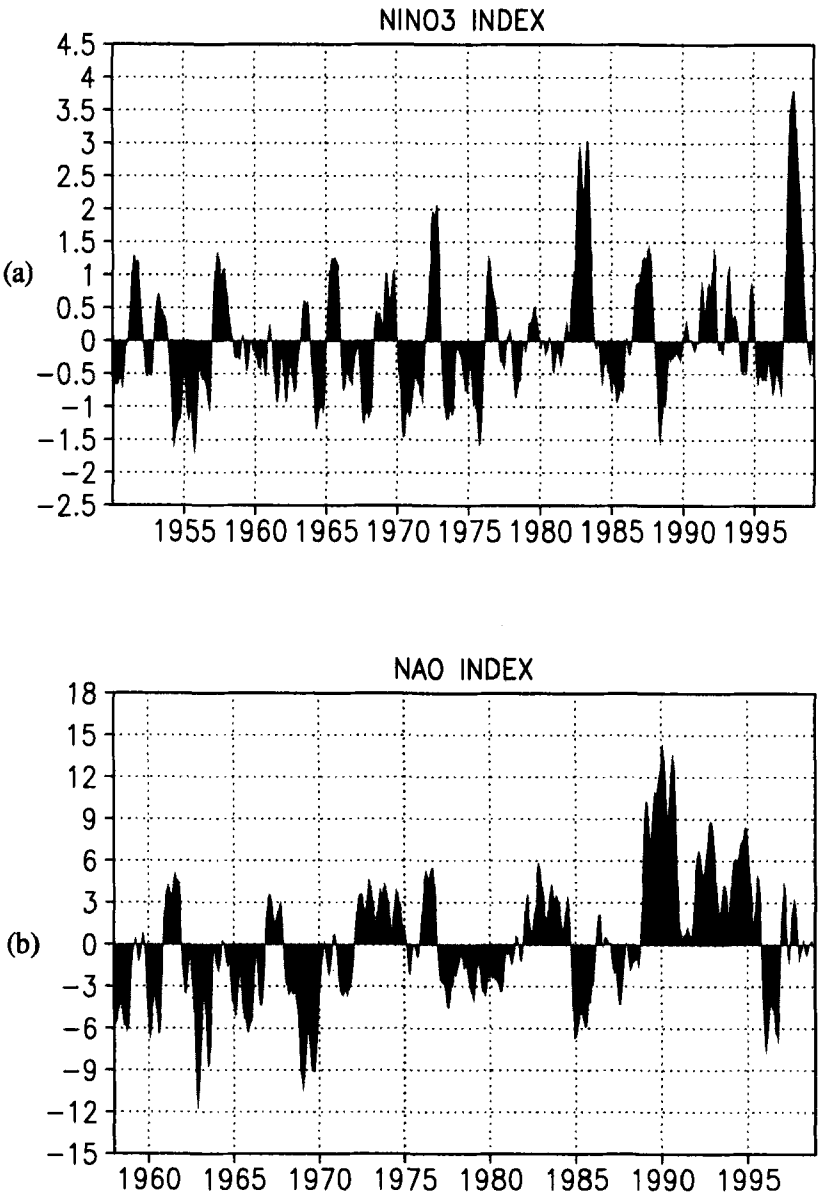

Figure 12. (a) The Niño3 index and (b) the NAO index (see text). The Niño3 index was obtained by averaging the sea surface temperature anomalies in the region (150°W–90°W, 5°S–5°N). The NAO index was calculated from monthly values of sea-level pressure anomalies from the National Centers for Environmental Prediction dataset (filtered to remove the variability below eight months) as the pressure difference between Lisbon (Portugal) and Stykkisholmur (Iceland), according to Hurrell (1996).

and the worst is spring; predictions in region II achieve the best results in summer but, again, spring scores the lowest; in region III the best forecasts are obtained in summer and winter and the worst in spring; in region IV the skill in spring is considerably lower than those in the other seasons. For all of the domains, it is clear that the worst forecasts are for spring, and it can be concluded that the seasons our model is most skilful at forecasting are winter and summer. These results agree with previous similar studies

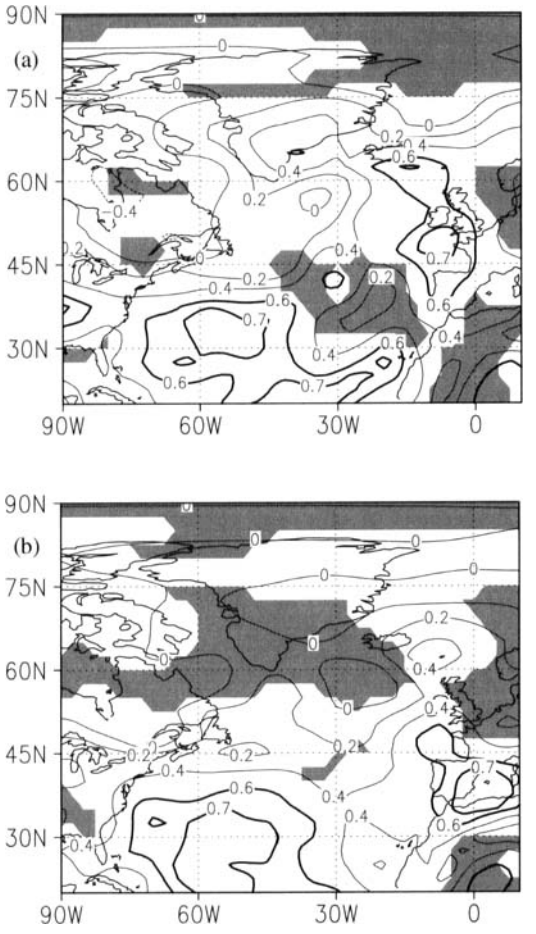

Figure 13. Comparison of the spatial distribution of the forecast skill for the El Niño–Southern Oscillation (ENSO) warm and cold phases. (a) The skill calculated as the correlation between the forecast and observed fields for the warm ENSO conditions, and (b) the skill calculated as the correlation between the forecast and observed fields for the cold ENSO conditions. The shaded areas indicate those points where the skill is below the values estimated for the artificial skill.

(Barnett and Preisendorfer 1987; Shabbar and Barnston 1996; Johansson *et al.* 1998) for the USA, Canada and northern Europe. In these studies, winter turned out to be the best forecast season, followed by summer.

(*b*) *Possible relationship between the origin of the forecast skill and coupling between the ENSO and NAO*

To study the origin of the forecast skill, we also tried to find a connection between the ENSO phase (as identified from the Niño3 index—Fig. 12(a)) and with the NAO phase (as identified from the NAO index—Fig. 12(b)). An $ENSO^+$ skill is estimated by averaging the correlation between the observed and forecast fields for the years of warm ENSO phase. The spatial structure of this skill is shown in Fig. 13(a). A similar $ENSO^-$ skill is computed by averaging over years with cold ENSO conditions. This is shown in Fig. 13(b).

The distribution of the skill for positive values of the NAO index (NAO^+ conditions) and for negative values of the NAO index (NAO^- conditions) is represented in Fig. 14.

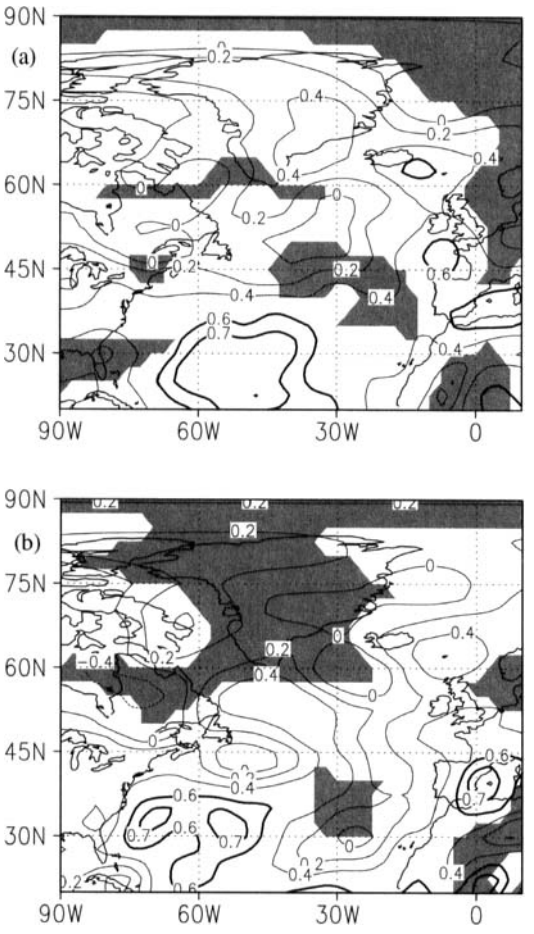

Figure 14. As Fig. 13, but for the North Atlantic Oscillation (NAO) positive and negative phases.

During NAO^+ conditions the best values of the skill are found in the central North Atlantic, whereas during the NAO^- years the predictability is shifted to the north-east.

Figure 15 gives a clue about the decadal dependence of the skill. We have represented a running correlation calculated at every step from a 12-year-long section for the period 1958–1998 between the Niño3 index and the NAO index. These correlations have been computed so that each of the 12-year frames is always coincident with the one of the training samples employed to forecast the 1970–1998 interval. Labels on the x-axis in Fig. 15 indicate the starting years of the segments when the initial month used to calculate the correlation is January. For example, the ordinate for the first point corresponds to the correlation computed from January 1958 to December 1969 (labelled 1958), the second one corresponds to the correlation taking from January 1960 to December 1972 (labelled 1960), and so on. The 12-year segment is progressively displaced so that the ordinate for the last point corresponds to a correlation estimated from 1987 to 1998. Segments beneath the axis indicate the forecast period whose training samples are contained within the time interval in the axis. The correlations summarize the information about the coupling between the NAO and ENSO that is included in the training sample used to make the forecast. We see that, although the values of the correlation are low,

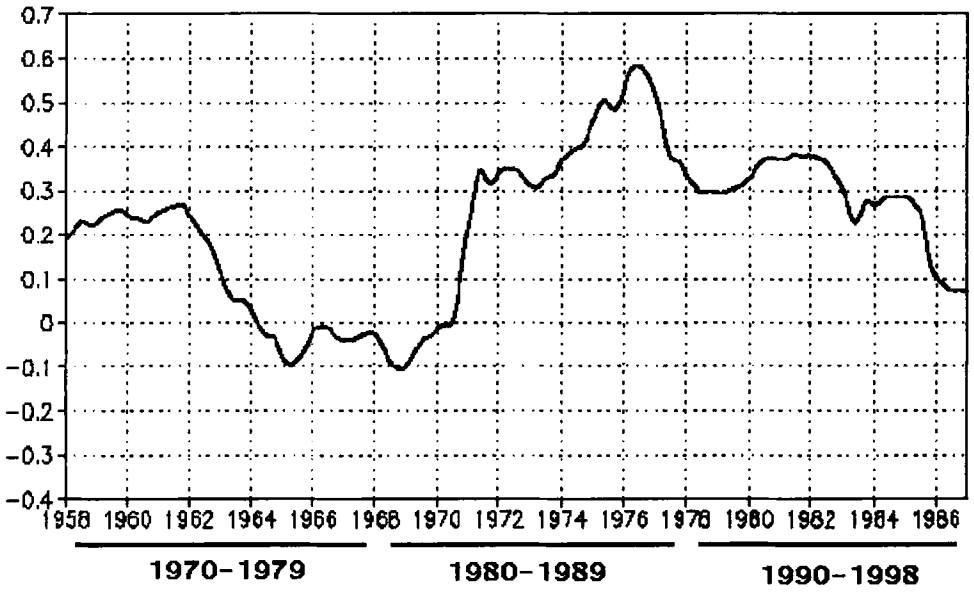

Figure 15. The twelve-year running correlation between the Niño3 index and the NAO index (see text). The calculations were carried out from January 1958 to December 1998 taking 12 years for each of the calculations. The horizontal axis indicates the first temporal point of each 12-year frame. Segments beneath the x-axis indicate the forecast period whose training samples are contained within the time interval.

there is an important peak in the 1976–1988 and 1977–1989 periods when the coupling between ENSO and NAO was stronger than normal. We postulate that this anomalous situation can affect the forecast in the 1980s, due to an abrupt change in the information contained in the training sample. However, for the 1990s, the relationship between the relative phases of the ENSO and NAO signals seems to remain almost constant. This is further supported by visualizing the errors of the forecasts against the running correlation (not shown). In all four domains used throughout this study the largest forecast errors appear when there is a contrast between the relative phases of the ENSO and NAO in the training sample and the relative phases in the forecast period.

(*c*) *Sources of predictability*

The spatial and temporal scales responsible for the forecast skill can also be investigated. All of the first and second singular vectors used as predictors (hereafter denoted by $\mathbf{v}_1$ and $\mathbf{v}_2$) for each forecast in our first experiment were saved. We proceeded likewise with the corresponding singular vectors of the predictand field (hereafter $\mathbf{u}_1$ and $\mathbf{u}_2$). These patterns account for roughly 36% and 20% of the total variance of the predictand field. Then we carried out a cluster analysis on each set of the patterns $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{u}_1$ and $\mathbf{u}_2$ based on principal component analysis. By this means, patterns are grouped according to the principal vectors (classes). Only the first three classes within each of the four sets were considered. The spatial patterns representative of these classes for the first singular vectors are shown in Figs. 16(a), (c) and (e) for the predictor $\mathbf{v}_1$ vector and Figs. 16(b), (d) and (f) for the predictand $\mathbf{u}_1$ vector. By projecting the patterns in Fig. 16 onto the original fields we obtain the time coefficients that give their temporal evolution (not shown). The first of these clusters is important in the forecasts of the mid 1970s, and reverses sign during the 1990s. The second cluster is important through the 1980s and 1990s, while the third is important only during the early 1970s. Notice

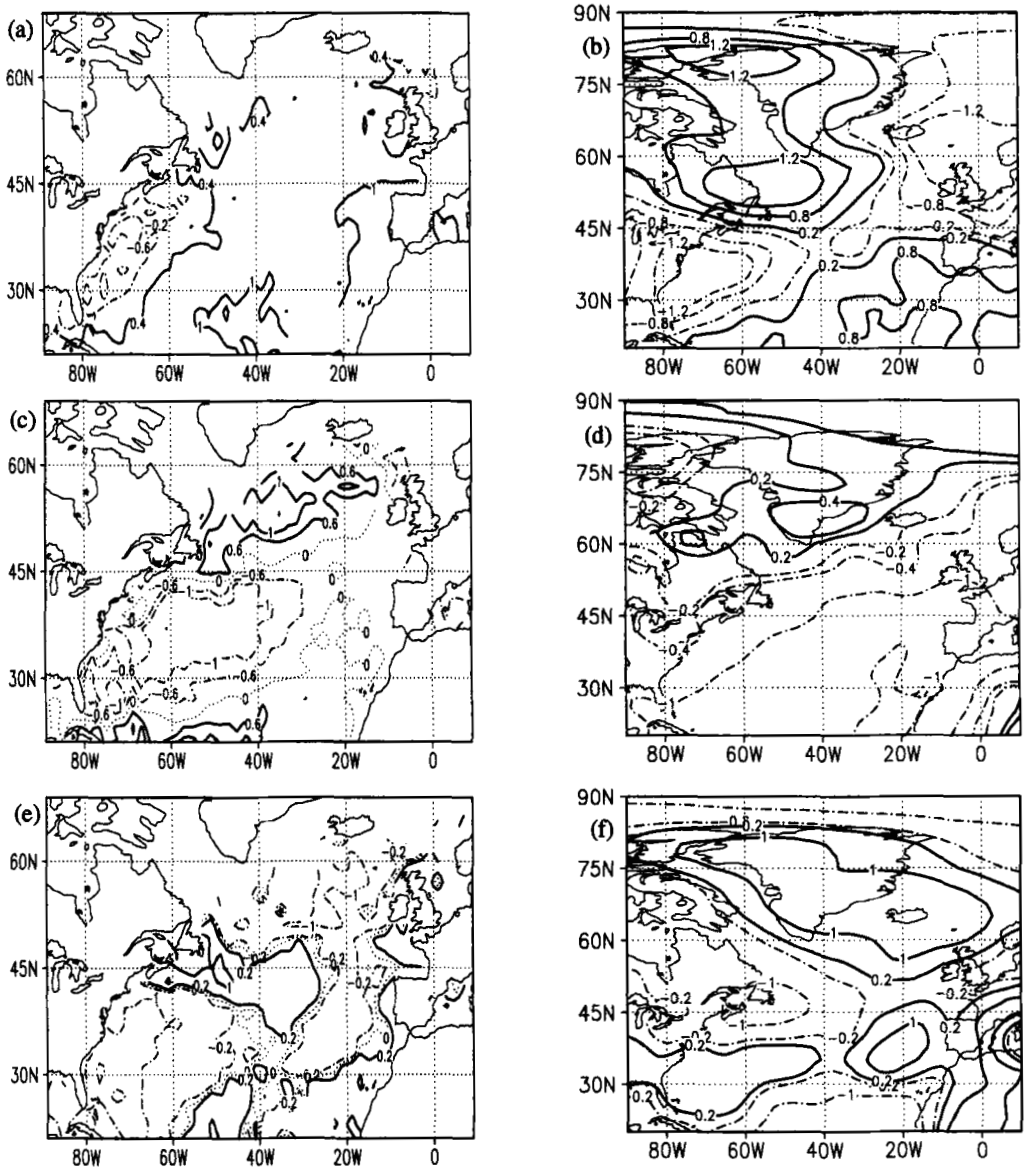

Figure 16. The spatial patterns representing (a) and (b) the first, (c) and (d) the second and (e) and (f) the third cluster of the first singular vectors. The left-hand panels show the sea surface temperature anomaly patterns (predictor vectors). The right-hand panels show the 850 hPa air temperature anomaly patterns (predictand vectors). The clustering analysis is carried out on the 348 first singular modes obtained from each of the singular-value decomposition calculations.

how, on some occasions, the predictor structure is mostly bipolar (cluster 1, Fig. 16(a)), while the predictand response has a tripolar structure (cluster 1, Fig. 16(b)). On other occasions it is just the other way round (cluster 2, Figs. 16(c) and (d)).

The power spectra (not shown) of the time coefficients of the first, second and third cluster of the first singular vectors for the SST anomaly fields present significant peaks at biennial and ENSO related timescales (3–4 years and 7–8 years). Time coefficients of the first, second and third cluster of the first singular vector of the air temperature

anomalies, besides the biennial and ENSO related time scales, show a significant decadal peak (15–20 years). A spectral peak at frequencies corresponding to decadal timescales can also be found in the corresponding oceanic time series, but is not statistically significant. The spectra of the time coefficients of oceanic and atmospheric fields for the clusters of the second singular vector show fully correspondence in their significant peaks.

6. Discussion

We present some forecasts for the interannual variability of 850 hPa air temperature anomalies at timescales greater than eight months. These timescales account for more than 60% of the total interannual variability of this field, and for only 20% less variability than its seasonal average, taken as the target of other mid-latitude empirical forecasts (Johansson *et al.* 1998; Shabbar and Barnston 1996). On the other hand, the skill of forecasts obtained here is more than 20% larger than for these forecasts.

Forecasts were issued for each month from 1970 to 1998. The performance of these forecasts is assessed in terms of the correlation skill between the forecast field and the predictand one (the filtered 850 hPa air temperature anomaly). To be significant, values of the forecast skill are required to be above those of the artificial skill, the latter being estimated from a Monte Carlo experiment. The artificial skill of the forecast is also tested from a cross-validation experiment, where only the year targeted in the forecast is excluded from the training sample. To be useful, the values of the skill are required to reach at least 0.6. The LEPS scores give another measure of the usefulness of the forecasts.

Different predictor datasets, oceanic and atmospheric, have been tested. From these, the SST anomaly over the North Atlantic basin is the field that performs best. Large-scale SST anomalies are generated by atmospheric forcing. This atmospheric high-frequency forcing has no predictability. The large-scale SST anomalies force, in turn, the atmosphere through turbulent surface heat fluxes. This explains why the SST anomalous state is a skilful predictor of the interannual atmospheric anomalies. The atmospheric damping of SST anomalies arises through air–sea interaction that is captured by our SVD scheme, and is the origin of the forecast skill.

The spatial distribution of the skill presents important values in the subtropics (centred on Bermuda) and related regions like the Mediterranean or in the western part of the North Atlantic gyre (near the Iberian peninsula). Only in the last of these regions does the skill beat the persistence one. The forecasts presented here can be useful for those regions. In another region of great interest for the European climate, the region around Iceland, the skill is statistically significant but not high enough to be useful. Finally, in the region around the Labrador peninsula the forecasts show little or no skill. Through complementary experiments it has been shown that the forecast skill of the northern regions cannot be improved appreciably, either by a small displacement of the boundaries or by a change of predictor field. Also, we have established the robustness of our results with respect to changes in the forecast lead.

The fact that our forecasts show almost no ability in regions of the USA and Canada, where other statistical forecasts do not lack skill, could be due to the differences in the treatment of the observations in obtaining the predictor and predictand fields, rather than to differences in the statistical method or to differences in the predictor and predictand domain. The important spectral peak corresponding to timescales of eight months and lower, which have been removed from the predictor field, is a feature of the higher

latitudes that is related to the seasonal signature of the anomalies (Deser and Blackmon 1993). This explains the lack of skill for those regions.

In other empirical predictions for the northern hemisphere (Shabbar and Barnston 1996; Vautard *et al.* 1998; Johansson *et al.* 1998) there is an important seasonal dependence of the skill, winter being the best forecast targeted season. In the present case, spring also seems to be the worst forecast season, whereas winter has higher skill scores. These seasonal variations in skill are connected with the seasonal recurrence of SST anomalies in the mixed layer (Namias and Born 1974) and with the advection of such anomalies. In spring, when the mixed layer is becoming more shallow, SST anomalies can be subducted and persist in the seasonal thermocline until the following winter, when the mixed-layer thickness reaches its maximum value. This mechanism explains the high values of the skill in region 1 that is characterized by a thick thermocline and the absence of currents. The skill in region 2 could be due to the advection of the resurfaced anomalies (Garcia Ortiz and Ruiz de Elvira 1995). The features associated with the above-mentioned seasonal signature present a biennial timescale of variability.

A part of the forecast skill is undoubtedly due to the ability of our model at capturing the ENSO (warm) signal in the subtropics. During the warm phase of the ENSO (Fig. 13(a)), the region of good forecast skill (above 0.5) covers an extended region over western Europe and eastern Greenland. During the cold phase of the ENSO (Fig. 13(b)), this region recedes to the south of 45°N, while the skill over the Mediterranean region improves. There is also an enhancement of the skill in some of these regions when the NAO is in its positive phase in comparison with what happens during the negative NAO phase. For western Europe, for instance, the improvement in the skill for the ENSO occurs during the NAO^+ conditions.

The change in the relative phases of the ENSO and NAO is a possible explanation for the loss of skill during the 1980s. When the two signals are in phase in both the training sample and the period being forecast, as in the 1990s, the predictability over the North Atlantic is considerably enhanced. Of course, there are other explanations for the lack of predictability, for instance the reduced length of the training period. Nevertheless, we have already mentioned that the skill worsened when 20 years were included in the training sample. This is the maximal training-sample length that enables us to forecast the 1980s and the 1990s. To capture the quasi-decadal timescales at least 60 years of training-sample length would be necessary.

The cluster analysis of each of the first two singular vectors used for the forecast shows that the skill is related to the ability of our model to capture the interannual variability at biennial, and also at ENSO, timescales. The biennial timescales, as well as the seasonal dependence of the forecast skill, point to physical mechanisms for the re-emergence of the anomalies. The ENSO timescales, as well as the spatial pattern of the skill in the subtropics, are evidence of the ability of the model at capturing the ENSO signal as the other origin of the forecast skill. The spatial structure of the skill (Fig. 7(a)) is reminiscent of barotropic waves responsible for the tropical–extratropical interactions. On the other hand, the evolution of the anomalies in the two 'case studies' can be related to the advection of the anomalies along the North Atlantic gyre, connected with decadal timescales (Halliwell 1997).

Lastly, the LEPS scores computed between the yearly means of the observations and the forecasts, provide evidence that, at the present state of the art and in terms of real observations, the forecasts are only meaningful for yearly-mean values in a region surrounding Bermuda. Although they are skilful in other regions, an important part of the observed anomalies near Bermuda is linked to the timescales of variability below one year. These are overlooked by the statistical procedure followed here, since it focuses

on longer timescales. We think that a statistical down-scaling would allow us to account for this variability. A probabilistic formulation of this down-scaling is being considered.

ACKNOWLEDGEMENTS

We acknowledge here C. Gutierrez's contribution in the first stages of this work, M. Seisdedos for her help with technical matters, and A. Ruiz de Elvira for his encouragement and support.

This work was financed by the European Union's Environment and Climate Programme under contract CT95-0109. The work of M. Seisdedos was financed by a grant from the Research Directorate of Madrid County. We acknowledge here the support given to the present research by the Spanish Interministerial Commission of Science and Technology and the Research Board of the University of Alcala through a grant (IN96-0071) that gave us access to improved computer facilities.

Thanks are also given to the two anonymous reviewers, for their suggestions.

REFERENCES

Barnett, T. P. and Preisendorfer, R. 1987 Origins and levels of monthly and seasonal skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Weather Rev.*, **9**, 1825–1850

Barnett, T. P., Heinz, H. D. and Hasselmann, K. 1984 Statistical prediction of seasonal air temperature over Eurasia. *Tellus*, **36A**, 132–146

Barnston, A. G. and Ropelewski, C. F. 1992 Prediction of ENSO episodes using canonical correlation analysis. *J. Climate*, **5**, 1316–1345

Bretherton, C. S., Smith, C. and Wallace, J. M. 1992 An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, **5**, 541–560

Davis, R. E. 1976 Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. *Ocean J. Phys.*, **6**, 249–266

Deser, C. and Blackmon, M. 1993 Surface climate variations over the North Atlantic ocean during winter: 1900–1989. *J. Climate*, **6**, 1743–1753

Feddersen, H., Navarra, A. and Ward, M. N. 1999 Towards a method for combined statistical-dynamical seasonal prediction. *J. Climate*, **12**, 1974–1989

Frankignoul, C. and Hasselmann, K. 1977 Stochastic climate models. Part II. Application to sea surface temperature anomalies and thermocline variability. *Tellus*, **29**, 289–305

Garcia Ortiz, J. M. and Ruiz de Elvira, A. 1995 An analysis of North Pacific SST anomalies by means of a linear thermodynamic stochastic two-dimensional model. *Tellus*, **47A**, 118–131

Halliwell, G. R. 1997 Decadal and multidecadal North Atlantic SST anomalies driven by standing and propagating basin-scale atmospheric anomalies. *J. Climate*, **10**, 2405–2411

Hurrell, J. W. 1996 Influence of variations in extratropical wintertime teleconnections on the northern hemisphere temperature variability of the northern hemisphere 500 hPa field. *Geophys. Res. Lett*, **33**, 665–668

Jenkins, G. M. and Watts, D. G. 1968 *Spectral analysis and its applications*, Holden-Day, USA

Johansson, A., Barnston, A., Saha, A. and van der Dool, H. 1998 On the level and origin of seasonal variability forecast skill in northern Europe. *J. Atmos. Sci.*, **55**, 103-127

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetma, A., Reynolds, R., Jenne, R. and Joseph, D. 1996 The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteorol. Soc.*, **77**, 437–471

Namias, J.	1953	Thirty-day forecasting: A review of a ten-year experiment. *Bull. Amer. Meteorol. Soc*, **33**, 279–285
Namias, J. and Born, R. M.	1974	Further studies of temporal coherence in north Pacific sea-surface temperature patterns. *J. Geophys. Res*, **75**, 5952–5955
Navarra, A.	1993	A new set of orthonormal modes for linearized meteorological problems. *J. Atmos. Sci.*, **50**, 2569–2583
Palmer, T. N. and Shukla, J.	2000	Editorial. *Q. J. R. Meteorol. Soc.*, **126**, 1989–1990
Palmer, T. N. and Sun, Z. B.	1985	A modeling and observational study of the relationship of sea surface temperature in the north-western Atlantic and the atmospheric general circulation. *Q. J. R. Meteorol. Soc.*, **111**, 947–925
Penland, C. and Magorian, T.	1993	Prediction of Niño3 sea surface temperature anomalies using linear inverse modeling. *J. Climate*, **6**, 1067–1076
Potts, J. M., Folland C. K., Joliffe, I. T. and Sexton, D.	1996	Revised 'LEPS' scores for assessing climate model simulations and long range forecasts. *J. Climate*, **9**, 34–53
Ratcliffe, R. A. S. and Murray, R.	1970	New lag associations between North Atlantic sea temperature and European pressure applied to long-range weather forecasting. *Q. J. R. Meteorol. Soc.*, **96**, 236–246
Rayner, N. A., Parker, D. E., Frich, F., Horton, E. B., Folland, C. K. and Alexander, L. V.	2000	'SST and sea ice fields for the ERA-40' Pp. 18–22 in Proceedings of the second WRCP international conference on reanalysis'. WCR 109 (WMO/TD 985), World Meteorological Organization, Geneva, Switzerland
Reynolds, R. W. and Smith, T. M.	1994	Improved sea surface temperature analyses. *J. Climate*, **7**, 929–948
Rodwell, M. J., Rowell, D. P. and Folland, C. K.	1999	Oceanic forcing of the wintertime North Atlantic oscillation and European climate. *Nature*, **398**, 320–323
Shabbar, A. and Barnston, A. G.	1996	Skill of seasonal climate forecasts in Canada using canonical correlation analysis. *Mon. Weather Rev.*, **124**, 2370–2385
Van Loon, H. and Rogers, J. C.	1977	The seesaw in winter temperatures between Greenland and northern Europe. Part 1: General description. *Mon. Weather Rev.*, **106**, 296–310
Vautard, R., Plaut, G., Wang, R. and Brunet, G.	1998	Seasonal prediction of North America surface air temperature using space–time principal components. *J. Climate*, **10**, 389–394
Walker, G. T and Bliss, E. W	1932	World Weather V. *Mem. R. Meteorol. Soc.*, **4**, 53–84
Wallace, J. M., Smith, C. and Bretherton, S.	1992	Singular value decomposition of wintertime sea surface temperature and 500 mb height anomalies. *J. Climate*, **5**, 561–576
Webster, H. and Palmer, J. C.	1997	The past and the future of El Niño. *Nature*, **190**, 562–564
Woodruff, S. D., Slutz, R. J., Jenne, R. L. and Steurer, P. M.	1987	A comprehensive atmosphere–ocean data set. *Bull. Amer. Meteorol. Soc.*, **68**, 1239–1250