Q. J. R. Meteorol. Soc. (2003), **129**, pp. 2299–2313 doi: 10.1256/qj.02.103

Predictability study of the observed and simulated European climate using linear regression

By RICHARD BLENDER[1]* UTE LUKSCH[1], KLAUS FRAEDRICH[1] and CHRISTOPH C. RAIBLE[2]

[1] *Universität Hamburg, Germany*

[2] *University of Bern, Switzerland*

(Received 15 April 2002; revised 19 December 2002)

SUMMARY

Monthly mean temperature anomalies in the regions England, Germany and Scandinavia are predicted by linear regression. Two predictors are selected from monthly mean teleconnection indices, North Atlantic sea surface temperatures (SSTs) projected on the first three empirical orthogonal functions (EOFs), and European climate variables (temperature, sea level pressure, and precipitation) averaged in the three predictand regions. The predictors are chosen separately for each month according to their correlation with the predictand. Observations from 1870–1999 and data from a 600-year integration with the coupled atmosphere–ocean general-circulation model ECHAM/HOPE are used to assess and compare the forecast skill. The skill is measured by the anomaly correlation coefficient (ACC) and the explained variance (EV).

For a one-month lead time the ACC for observations is up to 0.6 (EV $\approx$ 35%) for February–March and August–September in the three regions. The skill for the simulated data is lower (maximum values at ACC $\approx$ 0.5, EV $\approx$ 25%) and its seasonal dependence differs from that of the observations. Main predictors are the preceding temperatures in the predictand region. Using segments of the simulated data the spread of skill is estimated as 0.1 in ACC (10% in EV).

For lead times up to one year there is a small ACC (0.3–0.4) in the observations for England (spring and late summer), and Scandinavia (August–September), but none in Germany. The observed two-month mean England temperature in spring and late summer can be predicted with six months' lead time for 1971–96 with 1870–1969 as a training set, selecting the first two North Atlantic SST EOF coefficients as predictors. A leave-two-out cross-validation in 1870–1999 shows a distinct reduction of skill. In simulated data, the skill beyond one month is negligible compared with the observations.

KEYWORDS: GCM evaluation Monthly forecast

1. INTRODUCTION

Seasonal forecasting in Europe has received renewed interest in recent years (Carson 1998; Goddard *et al.* 2001). Whereas the tropical Pacific climate has useful predictive skill at forecast lead times of up to one year (Latif *et al.* 1998) and the seasonal rainfall in Africa can be successfully predicted using historical sea surface temperature (SST) anomalies (Mutai *et al.* 1998), the skills in Europe are only moderate.

There are two basic approaches for long-range climate prediction: dynamical comprehensive atmosphere–ocean general-circulation models (AOGCMs) for ensemble prediction (Anderson *et al.* 1999), and empirical statistical models. Although it is anticipated that dynamical models may become superior to empirical methods in future, empirical forecasts are still able to compete. A major advantage is that they require orders of magnitude less computations than coupled models. Moreover, these simple models can help to identify causes and relations within the climate system, and reveal deficiencies of complex dynamical models. Furthermore, since combinations of independent forecasts of numerical weather prediction and statistical models improve the short-range predictions considerably (Fraedrich and Leslie 1987; Raible *et al.* 1999), it is expected that this may also be applicable on longer time-scales (Fraedrich and Smith 1989; Sarda *et al.* 1996).

* Corresponding author: Meteorologisches Institut, Universität Hamburg, Bundesstr. 55, D-20146 Hamburg, Germany. e-mail: blender@dkrz.de

Dynamical multi-model ensemble integrations forced by observed SSTs in 1979–93 have been analysed within PROVOST (PRediction Of climate Variations On Seasonal to interannual Time-scales). The skill in the midlatitudes shows strong seasonality with maximum values in late winter and early spring (Doblas-Reyes *et al.* 2000), and nearly no annual cycle over the tropics (Pavan and Doblas-Reyes 2000). It has been suggested that the European climate is difficult to predict due to the intense synoptic variability over the North Atlantic (Palmer and Anderson 1994). Over northern Europe, at the exit region of the Atlantic storm-track, several GCMs have still substantial systematic errors (Pavan and Doblas-Reyes 2000).

Long-range forecasting with empirical techniques has a long history (Rossby 1941); it is mostly based on linear regression (Barnett and Preisendorfer 1987; Barnston and Smith 1996; Colman 1997; Johansson *et al.* 1998; Colman and Davey 1999) or on analogue techniques (Bergen and Harnack 1982; Livezey *et al.* 1994; van den Dool 1994). Recently, neural network models (Hsieh and Tang 1998; Tang *et al.* 2000), space–time principal components obtained from the multichannel singular-spectrum analysis technique (Vautard *et al.* 1999), and principal prediction patterns (Dorn and von Storch 1999) have been used for prediction.

The main potential predictors for the European temperature are given by the large-scale circulation patterns (Rossby 1941; Hurrell 1995), the advection of North Atlantic SST anomalies (Sutton and Allen 1997) and the memory of the European climate itself. The circulation anomalies can be characterized by teleconnection indices (Wallace and Gutzler 1981). The North Atlantic Oscillation (NAO, Defant 1924), defined as the pressure difference between the Icelandic low and the Azores high, is the most relevant index for Europe. The North Pacific pattern (NP, Trenberth and Hurrell 1994) describes the intensity of the Aleutian trough. Finally, the Southern Oscillation Index (SOI) impacts the extratropical circulation (Fraedrich 1994; Sickmöller *et al.* 2000; Raible *et al.* 2001). These teleconnection indices (or the occurrence of planetary interannual oscillations) are often linked to anomalous values of SST and land surface properties (Pavan and Doblas-Reyes 2000).

The predictability of the seasonal temperature in northern Europe during 1955–93 has been analysed by Johansson *et al.* (1998). In the regression based on canonical correlation analysis, the 700 hPa geopotential height north of 20°N, the quasi-global surface temperature, and the predictand itself lead to skill mainly in winter and a weaker maximum in summer. The geopotential height with an NAO-like pattern is the most important predictor in winter, superior to global and local SST and surface air temperature. The area-averaged cross-validated anomaly correlation coefficient (ACC) between observations and forecasts has maximum values above 0.45 in an eastward oriented belt stretching from south-east Sweden towards Moscow. Predictability of the central England temperature in summer has been found by Colman (1997) and Colman and Davey (1999). The forecasts during 1946–95 and 1971–95 use the North Atlantic SST as predictor with a lead time of six months.

The aim of this paper is to study the predictability for observed (1870–1999) and simulated (600 years) monthly mean European temperature anomalies for lead times up to one year. A major advantage of the simulation is the length of the time series which allows one to compare different sampling approaches, and, furthermore, yields an estimate for the spread of the skill. The forecast within the simulated data is completely independent from the observed data. The method is a multivariate linear regression with an adaptive choice of predictors. The available set of predictors are the monthly mean teleconnection indices, North Atlantic SST anomalies, and European climate variables. Out of this predictor set, the two time series showing the highest correlations with the predictand are selected. Although mainly local temperatures are selected for the one-month lead time, this changes for longer lead times. Long lead-time forecasts are performed to detect the limits of the predictability and to compare the model simulation with the observations.

The paper is organized as follows. The observational data, the simulations and the data processing are described in section 2. The forecast method and the verification measures are presented in section 3. The results for the predictability of observed and simulated data with different lead times and for different validation approaches are presented in section 4. Finally, in section 5, the results are summarized and discussed.

2. Data

The empirical prediction scheme is applied to observed as well as simulated climatological data. To permit a comparison of both datasets, the simulated time series are defined in similar geographical regions as the observations. In the sections below, the predictands and the predictors, the data sources and the data processing are described.

(*a*) *Predictands and predictors*

About 130 years of monthly mean observed and 600 years of simulated data are available for model building and verification of the empirical prediction scheme. Predictands and predictors are monthly mean anomalies which are linearly detrended and standardized separately for each month.

The *predictands* are the monthly temperature anomalies (deviations from the long-term monthly mean) averaged in the regions England, Germany, and Scandinavia. The *predictor* data consist of three groups: teleconnection indices, North Atlantic SST anomalies, and European climate variables. The teleconnection indices provide information about the global circulation. These comprise the NAO, the NP index, and the SOI, determined either at stations or as averages of four model grid points. The North Atlantic SST anomalies are included as projections (principal components) of the anomalies on the first three empirical orthogonal functions (EOFs) in the North Atlantic region (20–65°N). The European climate variables are monthly surface data for temperature, sea level pressure (SLP) and precipitation, averaged in the three predictand regions.

(*b*) *Observations*

The observed monthly mean data are available for 1870–1999 with a few missing values for NP, NAO, SOI, and precipitation. The climate indices are available from the Climate and Global Dynamical Division*. The NAO index is based on the difference of normalized pressure between Lisbon/Portugal and Stykkisholmur/Iceland (Hurrell 1995). The NP index is the area-weighted SLP over the region 30–65°N, 160°E–140°W (Trenberth and Hurrell 1994). The SOI is derived from the normalized SLP in Tahiti and Darwin (Trenberth 1984).

The North Atlantic SST, European temperature, SLP, and precipitation are available from the Climate Research Unit† on regular grids. The monthly SST anomalies (Parker *et al.* 1995) are projected on the first three EOFs in the North Atlantic region (20–65°N). The temperature (Jones *et al.* 1997), the SLP (Jones 1987) and the precipitation (Hulme 1992) are averaged over the three regions England, Germany, and Scandinavia.

* http://www.cgd.ucar.edu

† http://www.cru.uea.ac.uk/cru/data/

(*c*) *Simulation*

The simulated data are from a 600-year present-day climate integration (Legutke and Voss 1999; Raible *et al.* 2001) with an AOGCM. The atmosphere is simulated by the European Centre model of Hamburg (ECHAM-4) using triangular truncation at wave number 30 (T30), corresponding to $3.75^\circ \times 3.75^\circ$ resolution, and 19 hybrid sigma-pressure levels. The ocean is simulated by the Hamburg Ocean model in Primitive Equations (HOPE) simplified by the Boussinesq approximation and formulated on a Gaussian T42 Arakawa-E grid. The horizontal resolution is approximately $2.8^\circ \times 2.8^\circ$ which is meridionally increased in the tropics up to 0.5°. In the vertical the model consists of 20 irregularly distributed levels with 10 levels in the first 300 m.

The atmospheric component with a T30 truncation captures the observed storm-track variability (Stendel and Roeckner 1998). The dominant modes of the northern hemisphere, the NAO and the Pacific North America pattern, are simulated adequately and represent the observed variability (Raible *et al.* 2001).

The teleconnection indices for the model simulation are derived using the normalized 1000 hPa geopotential height averaged over four model grid points near the locations of the station data (see subsection (b)). The SST anomalies are projected on the first three EOFs in the North Atlantic region (20–65°N). The EOFs are determined separately for observations and simulations but the patterns are quite similar. The temperature, the SLP and the precipitation are grid-point averages in the regions England, Germany, and Scandinavia.

3. FORECAST AND EVALUATION

For the empirical prediction scheme a linear regression method is employed. This technique assumes linear relationships and stationarity of the time series. Nonlinear relationships, which are observed for the El Niño Southern Oscillation, require other, less parsimonious methods (note that the SOI time series is included as a possible linear predictor). Nonstationary time series with different phases and different interrelations would require much longer time series than the available observations to inhibit a lack of significance. A short description of the forecast technique as well as the measures for model verification are presented below.

(*a*) *Forecasting by linear regression*

The linear regression is the basis of the forecast scheme. For each calendar month and predictand, the prediction is performed separately. The predictors are denoted as $X_{i\beta}$, with the annual time step $i = 1, \ldots, N$, where N denotes the number of years, and the variable $\beta = 1, \ldots, M$; M being the number of predictors. The forecast of a single predictand is F_i, and the signal S_i is the given time series to be predicted, either given by observations or model data. The forecasts are linear combinations obtained by

$$F_i = \sum_{\beta=1}^{M} X_{i\beta} A_\beta, \quad i = 1, \ldots, N \tag{1}$$

with coefficients A_β; in matrix notation this is $\mathbf{F} = \mathbf{XA}$. The differences between the signals and the forecasts are the errors $\mathbf{E} = \mathbf{S} - \mathbf{XA}$. A minimum of the squared error $\mathbf{E}^\top\mathbf{E}$ (where $\top$ is transpose) is obtained for the coefficients $\mathbf{A} = \mathbf{\Sigma}^{-1}\mathbf{X}^\top\mathbf{S}$ with $\mathbf{\Sigma} = \mathbf{X}^\top\mathbf{X}$, which is proportional to the predictor covariance matrix. For example, for one-month lead time, the predictors $X_{i\beta}$ are past signals shifted by one month to the forecast time i. Thus, the last predictor and the predictands are adjacent. The number N of years depends on the dataset and the sampling. The predictors are selected according to their correlations with the predictand in the training (learn) set. Due to correlations between the predictor time series, the over-fitting problem occurs, and, after an increase for small predictor numbers, the skill decreases for large numbers. The number M of predictors with an optimal skill is about two for the observed data and two or three for the simulated data. For simplicity, this number is fixed to $M = 2$ throughout this investigation.

(*b*) *Accuracy, skill and sampling*

The relative accuracy of the forecasts is measured by two skill scores, the temporal ACC, and the explained variance (EV, Wilks 1995).

$$\mathrm{ACC} = \langle FS\rangle(\langle F^2\rangle\langle S^2\rangle)^{-1/2}, \quad \mathrm{EV} = 1 - \mathrm{MSE}/\langle S^2\rangle \tag{2}$$

where S are the signals, F the forecasts, and $\langle \ldots \rangle$ denotes the time mean of the time series. MSE is the squared error of the forecast, $\mathrm{MSE} = \langle (F - S)^2\rangle$. The data are centred, $\langle S\rangle = \langle F\rangle = 0$. EV vanishes if the method is equivalent to the climate forecast ($F = 0$) and becomes unity for a perfect forecast. EV is positive if the regression model is better than the climate reference model.

The skill is estimated using different sampling approaches. For small samples, the so-called leave-one-out cross-validation (Michaelsen 1987) is appropriate and has been applied in a large number of statistical forecast studies (Barnston 1994). The skill is determined by forecasts for every year using all remaining years as the training set. For the observations this method is used and extended to leave-two-out cross-validation (training without the forecast and the preceding year), thus $N = 127$ in Eq. (1).

The skill for the simulated dataset (600 years) is evaluated using three different samplings:

- For a direct comparison with the observations, the 600 years are partitioned in six 100-year segments which are evaluated separately by leave-two-out cross-validation.
- The total 600-year time interval is analysed with leave-two-out cross-validation.
- All six 100-year segments are used as training sets with the remaining five 100-year segments as forecast (test) sets. This yields 30 individual results which are used to determine the mean and the standard deviation of the skill.

In a case-study, the skill for the observed England two-month mean temperature forecast for six months' lead time is estimated by two different samplings, a leave-two-out cross-validation and a two sub-sample validation by separation into learn (1870–1969) and test set (1971–96).

4. Results

The regression forecast is applied to observed and simulated temperature anomalies in the three regions England, Germany, and Scandinavia. For every month, the first two predictors with maximal correlation with the predictand are selected. The annual cycles of accuracy and skill of the temperature-anomaly forecasts are presented using the ACC and the EV. Since the number N of observations for a fixed month is rather small, the forecast skill is determined by a leave-two-out cross-validation. The simulated data skill is evaluated using three different samplings (details in section 3(b)).

First, the regression model is applied to data with one-month lead time. Observations and simulations are compared and the distribution of the skill in the simulations is determined. Second, forecasts with lead times up to one year are performed for observations and simulations. In a case-study, the skill for the two-month mean England temperature with six months' lead time is considered.

(*a*) *Observations, one-month lead time*

In the observed dataset the temperature predictability for one-month lead time is determined in the regions England, Germany and Scandinavia. The skill shows a seasonal cycle simultaneously in England, Germany, and Scandinavia (Figs. 1(a)–(c) respectively) with maximum values in February–March and August–September (ACC $\approx$ 0.5–0.6, and EV $\approx$ 30–35%). The decay during April–June and November is a well-known property and has been observed in several studies (Carson 1998; Johansson *et al.* 1998); it can be interpreted as a reorganization of the atmospheric circulation in spring and autumn. The first, dominant predictors for this forecast are the preceding temperatures in the corresponding regions. The second predictors are mostly other European temperatures; however, in some cases replaced by the SST (England winter and summer), SLP (England autumn), and NAO (Scandinavia winter).

(*b*) *Simulations, one-month lead time*

For a comparison with the observations, the 600 years simulated data are first split in six 100-year segments which are analysed separately by leave-two-out cross-validation. This gives a first hint to the skill–variability. Then, the total 600-year time range is subjected to a leave-two-out cross-validation to obtain the most significant predictability estimate. Finally, an intercomparison of all 100-year segments used as independent training and test sets is performed to estimate the spread of the skill.

The skill, estimated by leave-two-out cross-validation in each segment, is shown in Figs. 1(d)–(f) for the regions England, Germany and Scandinavia respectively. Since the cross-validation and the choice of predictors are performed independently for all six segments, the six curves build an ensemble which can be compared with the observational skill in Figs. 1(a)–(c). The simulated ACC is 0.25–0.5, and EV is 5–25%. Since the spread within the six curves is about 0.1 in the ACC (10% in EV) a clear annual cycle can be attributed to the mean skill in all regions with maximum values in April–May (Germany and Scandinavia) and August (England and Scandinavia). The first predictor is always the regional temperature. The second predictor is the temperature in a different European region, besides SLP (England) and NAO (Scandinavia in spring).

To use all available data in the prediction, a leave-two-out cross-validation is applied to the whole 600 years. Figures 2(a)–(c) show the ACC and the EV in this forecast for the three regions England, Germany and Scandinavia respectively. This result is similar to the mean of the six independent ensembles (100-year forecasts) in Figs. 1(d)–(f)).

To estimate the variability of the skill within the simulations, a sub-sample cross-validation is performed which uses all six 100-year segments as training and all the remaining five 100-year segments as independent forecast (test) sets. This leads to an ensemble of 30 forecasts whose mean and standard deviation are shown in Figs. 2(d)–(f) for the regions England, Germany and Scandinavia respectively. The total height of the error bars, 4σ, where σ is the standard deviation, includes approximately 95% (a Gaussian distribution is roughly provided). The ensemble means show the same behaviour as the 600-year cross-validation skills in the left panel and the 100-year sub-sample cross-validation in Figs. 1(d)–(f). The error-bar widths exclude a constant skill and indicate a distinct annual cycle.

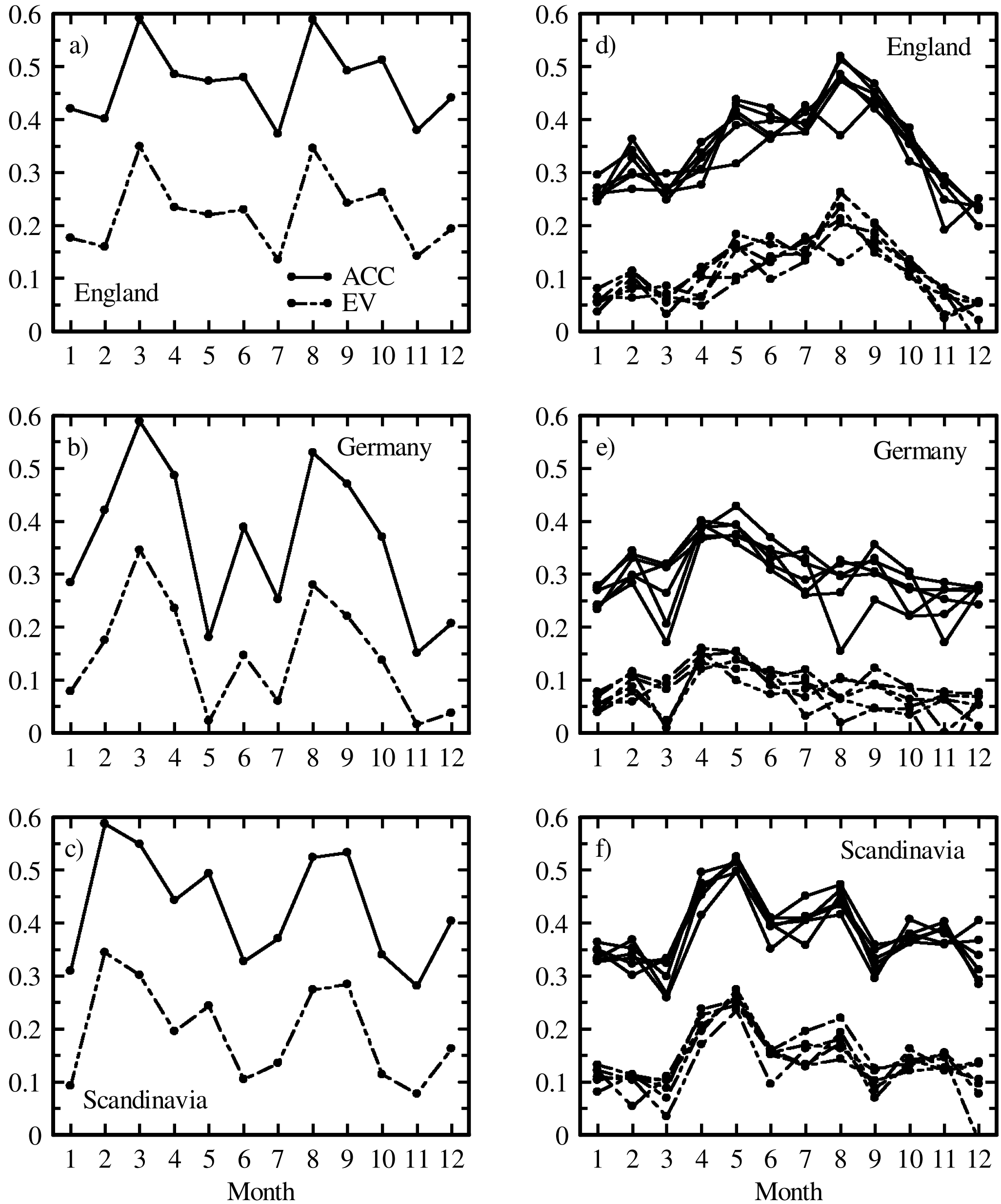

Figure 1. Forecast skill for observations and simulations using leave-two-out cross-validation in three regions. Left panel ((a)–(c)): observations; right panel ((d)–(f)): simulations, six 100-year segments (six curves). Anomaly correlation coefficients (ACCs) and explained variances (EVs) are given for England ((a) and (d)), Germany ((b) and (e)), and Scandinavia ((c) and (f)).

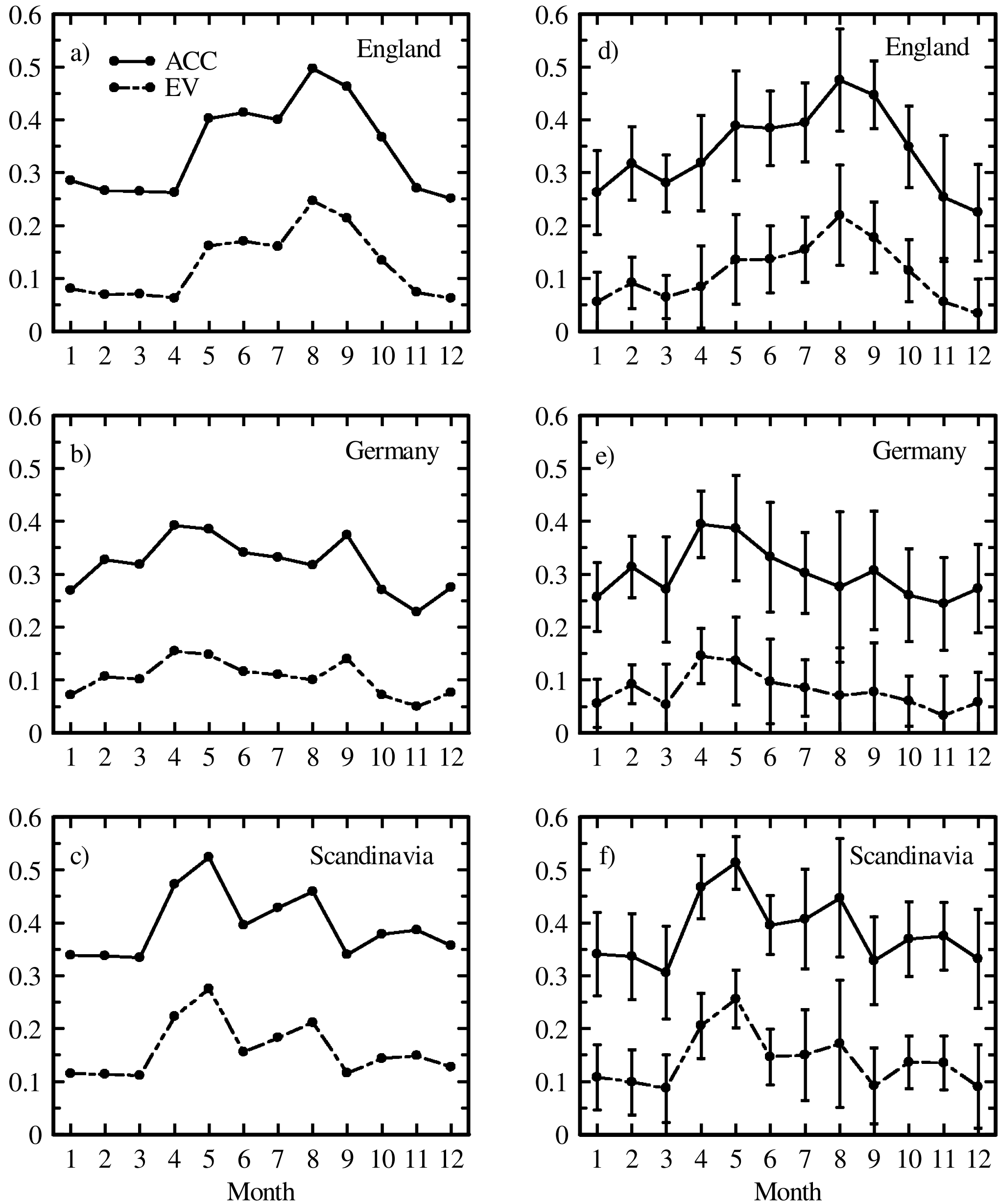

Figure 2. Forecast skill for the 600-year simulation with two sampling methods in three regions. Left panel ((a)–(c)): leave-two-out cross-validation for the total range. Right panel ((d)–(f)): subsample cross-validation using all 100-year learn and all 100-year test periods leading to 30 combinations. The curves show means and error bars with total height 4σ (σ = standard deviation) including 95%. Anomaly correlation coefficients (ACCs) and explained variances (EVs) are given for England ((a) and (d)), Germany ((b) and (e)), and Scandinavia ((c) and (f)).

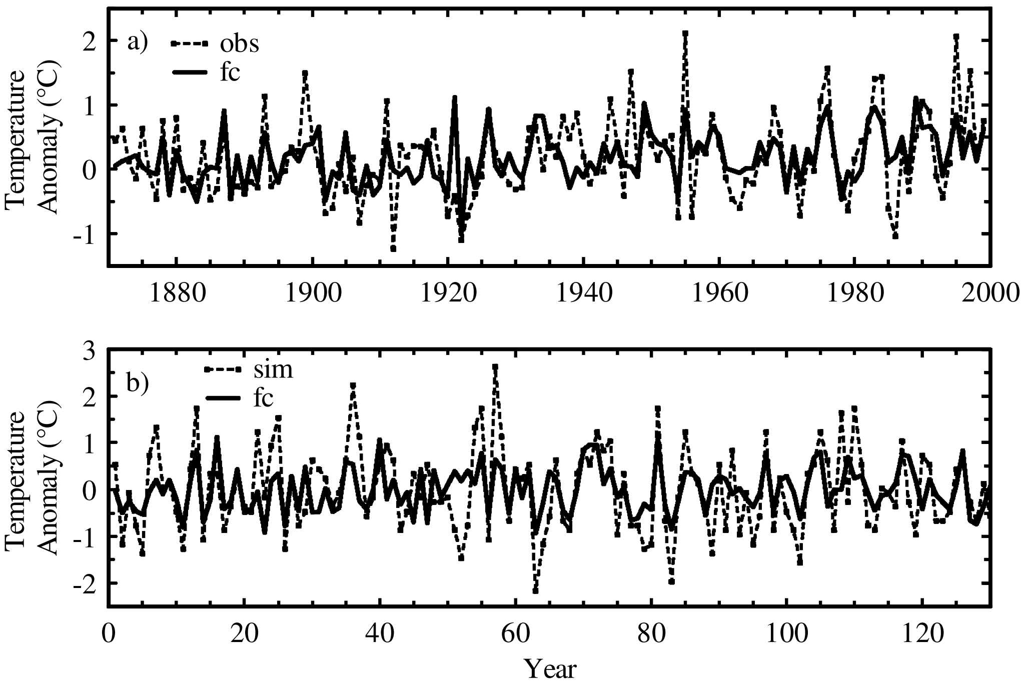

Figure 3. Temperature anomalies in England during August: (a) observation and forecast; and (b) simulation and forecast, only the first 130 years are shown.

For a comparison of the skills for observed and simulated data it is assumed that the observed and simulated spread is similar. The selected predictors are nearly the same in observations and model simulations but the annual cycle of skill for the GCM data (Figs. 2(d)–(f)) deviates from the observed one with its maximum in February–March and August–September (Figs. 2(a)–(c)). Main differences are the low skill in late winter and early spring in all three regions. Furthermore, the late summer skill maximum in Germany is not reproduced in the model simulation.

As an example, the forecast of the observed and simulated England temperature in August with one-month lead time is presented (Figs. 3(a) and (b)). The observed time series correspond to ACC ≈ 0.6 (see Fig. 1(a)). The forecast of the simulated data is shown in Fig. 3(b) with ACC ≈ 0.5 (see Fig. 2(a)). The displayed range is restricted to the first 130 years. Note that the GCM simulation uses an interactive ocean model and is not forced by the observed SST; therefore, there is no year-to-year correspondence of the observed and simulated anomalies.

(*c*) *Lead time 1–12 months*

In this section, the forecast for the monthly mean temperature is extended to lead times up to one year. Figure 4 displays the ACC for the observations (Figs. 4(a)–(c)) and simulations (Figs. 4(d)–(f)). The patterns for one-month lead time correspond to the left panels of Figs. 1 and 2. The thresholds for a 99% significance of correlation coefficients with N degrees of freedom are $r = 0.2$ ($N = 130$) and $r = 0.1$ ($N = 600$). Note that the significance test for correlations and the level for a useful prediction represent different aspects, whereas $r = 0.2$ might be a significant correlation, it is below a useful prediction skill which is considered to be above 0.4.

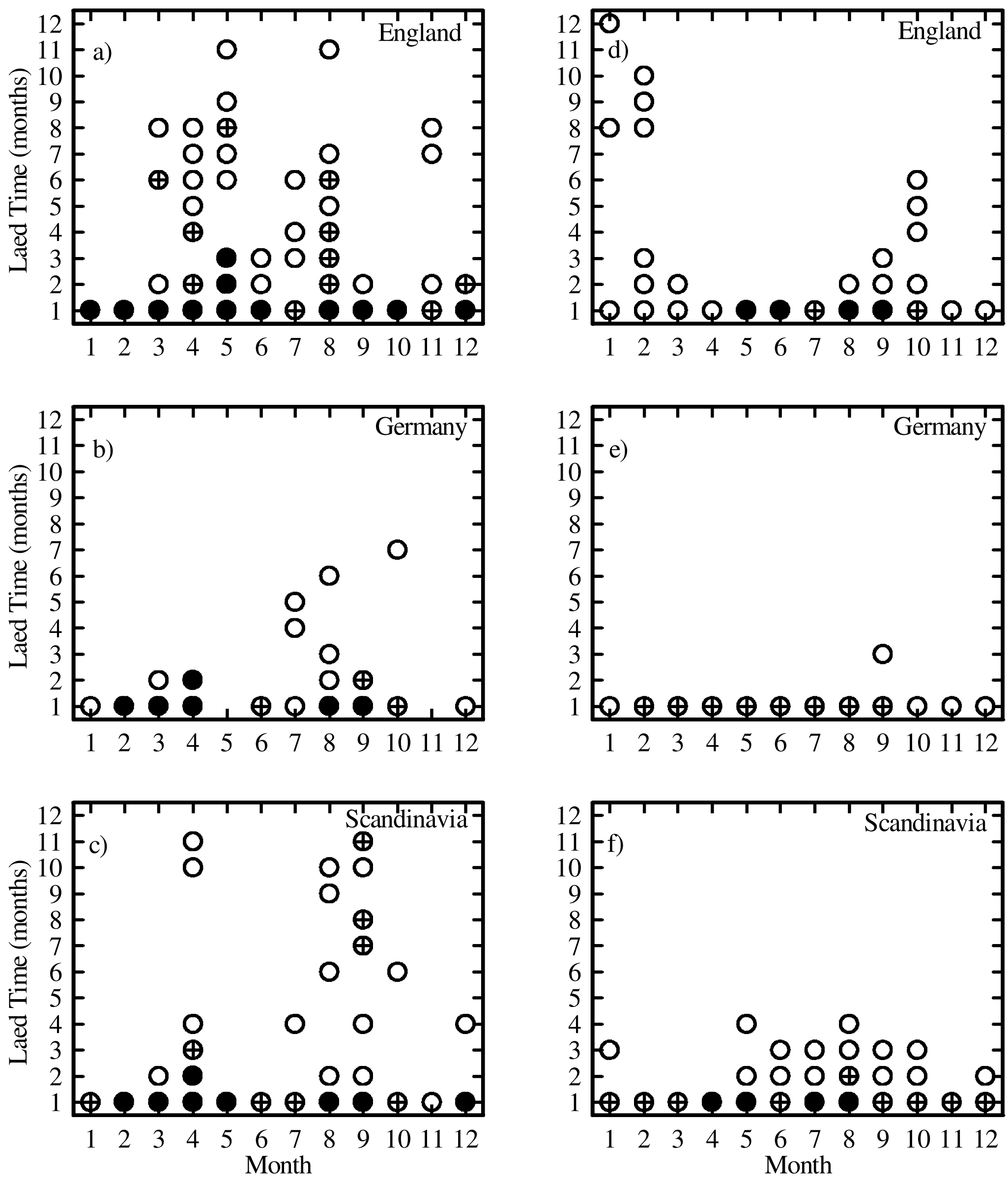

Figure 4. Anomaly correlation coefficient (ACC) for the forecast of monthly temperatures for 1 to 12 months lead time obtained by leave-two-out cross-validation. Left panel ((a)–(c)): observations, 1870–1999; right panel ((d)–(f)): simulations, 600-year; in England ((a) and (d)), Germany ((b) and (e)), and Scandinavia ((c) and (f)). ACC levels are 0.2 (○), 0.3 (⊕), and 0.4 (●).

Observed England temperature shows small skill (ACC > 0.3) up to six months' lead time in March–May and July–August where SST (second or third EOF) and England or Scandinavian temperature are selected as predictors. Germany reveals very weak ACC. Scandinavia hints to long lead-time predictability in late summer, as in England, however, mainly forecasted with the local temperature and different secondary predictors with weak contributions (for example the SOI and the SST EOFs). The ACC for the April forecast is based on the local temperature; NAO contributes for 3–4 months

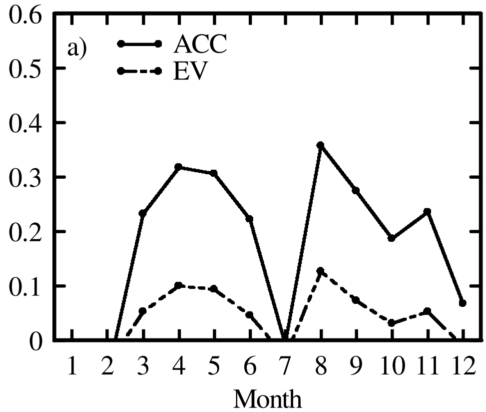

Figure 5. Skill for the observed England two-monthly mean temperature forecast using two-monthly mean predictors six months before. Anomaly correlation coefficients (ACCs) and explained variances (EVs) are obtained for (a) leave-two-out cross-validation in 1870–1999, and (b) forecast range 1971–96 (test) using the training range 1870–1969 (learn). The two-monthly mean is shown at the 2nd month, e.g. 8 is the July–August mean.

lead time. There is no winter skill in Scandinavia for longer lead times. The skill for the forecasts of the two-monthly mean temperature follows a similar behaviour (not shown).

The maximum in July–August for the England temperature has been found by Colman (1997) with higher ACC values; this is investigated further later. The skill during March–May for lead times up to 9 months appears to be new.

For simulated data the forecast skill for longer lead times (Figs. 4(d)–(f)) is considerably lower than in the observations (Figs. 4(a)–(c)). Furthermore, the seasonal dependence of significant skill (ACC > 0.2) in simulations does not coincide with the observations. Obviously, the AOGCM is not able to simulate the observed intra-annual memory of the climate system.

(*d*) *Two-monthly mean England temperature, six months' lead time*

The observed two-monthly mean England temperature is predicted with six months' lead time. To validate the forecast skill two different samplings are applied. A leave-two-out forecast during the whole observed dataset, 1870–1999, shows weak maximum values (ACC around 0.3) in spring and late summer (Fig. 5(a)). In a second approach, the forecast for 1971–96 is trained in 1870–1969. The skill in Fig. 5(b) shows two distinct maxima, a first during winter and spring (ACC ≈ 0.4), and a second for August–September (ACC ≈ 0.5). It is noteworthy that throughout the year the dominant predictor is the coefficient of the first EOF of the SST anomaly. The second predictor varies during the beginning of the year but remains fixed to the second EOF coefficient from June to December. A comparable prediction study by Colman (1997) and Colman and Davey (1999) shows a high skill for the forecast of the England summer temperature (July–August mean) using the preceding winter North Atlantic SST as predictor. The comparison of the two sub-sample validations with the leave-two-out cross-validation demonstrates that the skill strongly depends on the forecast period. Further inspection shows that the skill for the latter sampling is mainly obtained by forecasts during the period 1960–99. This prediction study is also applied to simulated data for two-month mean England temperature and predictors with six months' lead time.

The leave-two-out cross-validation reveals no skill (ACC < 0.2) during the whole year (not shown).

5. Summary and Discussion

This study applies a linear regression to forecast monthly mean temperature in the three regions England, Germany, and Scandinavia with lead times up to one year. In a case-study, the England temperature is predicted with a lead time of six months. The forecast uses observations in 1870–1999 and model data from a 600-year control simulation of a coupled atmosphere–ocean GCM. Note that the AOGCM is solely used to produce a long independent dataset and that the forecasts are performed by the empirical linear regression, separately within both datasets. Predictors are teleconnection indices (NAO, NP, SOI), North Atlantic SST anomalies projected on the first three EOFs, and the climate variables (temperature, sea level pressure, and precipitation) in the three predictand regions. The empirical method selects the two predictors with the largest predictand correlation.

A new aspect of the present study is the use of data simulated by an AOGCM in an empirical forecast. Due to the long interval of 600 years, a more reliable forecast and evaluation of the regression model should be possible. The datasets in observations and simulations are prepared to overlap geographically as much as possible and to contain similar climatological information.

For one-month lead time, observations show skill maxima during spring and late summer in all three regions with ACC ≈ 0.6; the most relevant predictor is the preceding temperature in the respective region. Due to the restricted length of these data, the skill is obtained by leave-two-out cross-validation. The skill within the simulated data is estimated using three different sampling and validation procedures: (i) by leave-two-out cross-validation in six 100-year segments, (ii) by leave-two-out cross-validation in the total 600 years' range, and (iii) by a partition in six 100-year training sets using all remaining 100 years as forecast sets. The latter partitioning yields estimates for the standard deviation of the skill. This result supports a distinct annual cycle which differs from that found in the observations. The skill in the simulated data is up to ACC ≈ 0.5 and in general lower than in the observations. In particular, the observed skill for England during spring, the spring and late summer skill in Germany, and the late winter skill in Scandinavia are not found in the simulation.

Regarding the predictor sets, there is little difference between observations and simulations, mainly in England where the SST is replaced by various other predictors (SLP, NAO, and precipitation) in the simulated data. NP and SOI are of minor importance in observations and simulation. Therefore, discrepancies between the skill in observations and simulation do not originate in the different selection of predictor time series but may be a hint on model deficiencies.

For longer lead times there is small skill (ACC $= 0.3$–0.4) for the observed temperature in England and Scandinavia during spring and late summer, but no skill in Germany. The forecast skill for longer lead times is general smaller in the simulation than in the observation. The result precludes attempts to use simulated data as a training set for the forecast of observational data. A forecast of the observed two-month mean England temperature with six months' lead time using two-month mean predictors yields skill for the forecast in 1971–96 using the training set 1879–1969. A leave-two-out cross-validation in the total observed timed interval demonstrates a clear decrease of skill. This skill is mainly based on successful forecasts after 1960.

A major shortcoming of the model simulation is the negligible skill for lead times larger than one month. On the other hand, there is sufficient skill for one-month lead time in the simulation, for example during autumn in England and spring in Scandinavia (ACC $\approx$ 0.5), to exclude a general low predictability within the AOGCM data. A possible reason is the unsatisfactory representation of the ocean and the land surface, including the surface fluxes, since for long-term prediction the realistic simulation of the slow components of the climate system is crucial. Improvements would not only be necessary for the empirical approach used here, but also even more for the use of AOGCMs in dynamical seasonal prediction, which is frequently performed using simplified versions of weather forecast models.

ACKNOWLEDGEMENTS

We thank Axel Andersson, Konrad Balzer, Thomas Frisius, and Edilbert Kirk for hints and discussions. This work was supported by the Bundesministerium für Bildung und Forschung (BMBF) under contract 01 LA 9820. We acknowledge the Deutsche Forschungsgesellschaft in the Sonderforschungsbereich 512, 'Tiefdruckgebiete und Klimasystem des Nordatlantiks'. The data processing at the Climate Research Unit and the Climate and Global Dynamical Division is appreciated.

REFERENCES

Anderson, J., van den Dool, H., Barnston, A., Chen, W., Stern, W. and Ploshay, J. 1999 Present-day capabilities of numerical and statistical models for atmospheric extratropical seasonal simulation and predication. *Bull. Am. Meteorol. Soc.*, **80,** 1349–1361

Barnett, T. P. and Preisendorfer, R. 1987 Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Weather Rev.*, **115,** 1825–1850

Barnston, A. G. 1994 Linear statistical short-term climate predictive skill in the northern hemisphere. *J. Climate*, **7,** 1513–1564

Barnston, A. G. and Smith, T. M. 1996 Specification and prediction of global surface temperature and precipitation from global SST using CCA. *J. Climate*, **9,** 2660–2697

Bergen, R. E. and Harnack, R. P. 1982 Long-range temperature prediction using a simple analog approach. *Mon. Weather Rev.*, **110,** 1038–1099

Carson, D. J. 1998 Seasonal forecasting. *Q. J. R. Meteorol. Soc.*, **124,** 1–26

Colman, A. 1997 Prediction of summer central England temperature from preceding North Atlantic winter sea surface temperature. *Int. J. Climatol.*, **17,** 1285–1300

Colman, A. and Davey, M. 1999 Prediction of summer temperature, rainfall and pressure in Europe from preceding winter North Atlantic Ocean temperature. *Int. J. Climatol.*, **19,** 513–536

Defant, A. 1924 Die Schwankungen der atmosphärischen Zirkulation über dem nordatlantischen Ozean im 25-jährigen Zeitraum 1881–1905. *Geogr. Ann.*, **6,** 13–41

Doblas-Reyes, F. J., Deque, M. and Piedelievre, J.-P. 2000 Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Q. J. R. Meteorol. Soc.*, **126,** 2069–2088

Dorn, M. and von Storch, H. 1999 Identification of regional persistent patterns through principal prediction patterns. *Contrib. Atmos. Phys.*, **72,** 105–111

Fraedrich, K. 1994 An ENSO impact on Europe? *Tellus*, **46,** 541–552

Fraedrich, K. and Leslie, L. M. 1987 Combining predictive schemes in short-term forecasting. *Mon. Weather Rev.*, **115,** 1640–1654

Fraedrich, K. and Smith, N. R. 1989 Combining predictive schemes in long-range forecasting. *J. Climate*, **2,** 291–294

Goddard, L., Mason, S. J., Zebiak, S. E., Ropelewski, C. F., Basher, R. and Cane, M. A. 2001 Current approaches to seasonal-to-interannual climate predictions. *Int. J. Climatol.*, **21,** 1111–1152

Hsieh, W. W. and Tang, B. 1998 Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull. Am. Meteorol. Soc.*, **79,** 1855–1870

Hulme, M. 1992 A 1951–80 global land precipitation climatology for the evaluation of general circulation models. *Clim. Dyn.*, **7,** 57–72

Hurrell, J. W. 1995 Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, **269,** 676–679

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

Jones, P. D. 1987 The early twentieth century arctic high—fact or fiction? *Clim. Dyn.*, **1,** 63–75

Jones, P. D., Osborn, T. J., Briffa, K. R., Folland, C. K., Horton, E. B., Alexander, L. V., Parker, D. E. and Rayner, N. A. 1997 Adjusting for sampling density in grid box land and ocean surface temperature time series. *J. Geophys. Res.*, **106,** 3371–3380

Latif, M., Anderson, D., Barnett, T., Cane, M., Kleeman, R., Leetmaa, A., O'Brien, J., Rosati, A. and Schneider, E. 1998 A review of the predictability and prediction of ENSO. *J. Geophys. Res.*, **103,** 14375–14393

Legutke, S. and Voss, R. 1999 'The Hamburg atmosphere–ocean coupled circulation model ECHO-G'. Technical Report 18, Deutsches Klimarechenzentrum, Hamburg, Germany

Livezey, R. E., Barnston, A. G., Gruza, G. V. and Ran′Kova, E. Y. 1994 Comparative skill of two analog seasonal temperature prediction systems: Objective selection of predicators. *J. Climate*, **7,** 608–615

Michaelsen, J. 1987 Cross-validation in statistical climate forecast models. *J. Climate Appl. Meteorol.*, **26,** 1589–1600

Mutai, C. C., Ward, M. N. and Colman, A. W. 1998 Towards the prediction of the east Africa short rains based on sea-surface temperature–atmosphere coupling. *Int. J. Climatol.*, **18,** 975–997

Palmer, T. N. and Anderson, D. L. T. 1994 The prospects for seasonal forecasting—A review paper. *Q. J. R. Meteorol. Soc.*, **121,** 317–342

Parker, D. E., Folland, C. K. and Jackson, M. 1995 Marine surface temperature: Observed variations and data requirements. *Clim. Change*, **31,** 559–600

Pavan, V. and Doblas-Reyes, F. J. 2000 Multi-model seasonal hindcasts over the Euro-Atlantic: Skill scores and dynamic features. *Clim. Dyn.*, **16,** 611–625

Raible, C. C., Bischof, G., Fraedrich, K. and Kirk, E. 1999 Statistical single station short-term forecasting of temperature and probability of precipitation: Area interpolation and NWP combination. *Weather and Forecasting*, **14,** 203–214

Raible, C. C., Luksch, U., Fraedrich, K. and Voss, R. 2001 North Atlantic decadal regimes in a coupled GCM simulation. *Clim. Dyn.*, **17,** 321–330

Rossby, C. G. 1941 The scientific basis of modern meteorology. Climate and Man. Pp. 599–655 in *Yearbook of agriculture*. US Government Printing Office

Sarda, J., Plaut, G., Pires, C. and Vautard, R. 1996 Statistical and dynamical long-range atmospheric forecasts: Experimental comparison and hybridization. *Tellus*, **48,** 518–537

Sickmöller, M., Blender, R. and Fraedrich, K. 2000 Observed winter cyclone tracks on the northern hemisphere in re-analysed ECMWF data. *Q. J. R. Meteorol. Soc.*, **126,** 591–620

Stendel, M. and Roeckner, E. 1998 'Impacts of the horizontal resolution on simulated climate statistics in ECHAM 4'. Technical Report 253, Max-Planck-Institut, Germany

Sutton, R. T. and Allen, M. R. 1997 Decadal predictability of North Atlantic sea surface temperature and climate. *Nature*, **388,** 563–567

Tang, B., Hsieh, W. W., Monahan, A. H. and Tangang, F. T. 2000 Skill comparison between neural networks and canonical correlation analysis in predicting the equatorial Pacific sea surface temperature. *Weather and Forecasting*, **13,** 287–293

Trenberth, K. E. 1984 Signal versus noise in the Southern Oscillation. *Mon. Weather Rev.*, **112,** 326–332

Trenberth, K. E. and Hurrell, J. W. 1994 Decadal atmosphere–ocean variations in the Pacific. *Clim. Dyn.*, **9,** 303–319

van den Dool, H. M. 1994 Searching for analogues, how long must we wait? *Tellus*, **46,** 314–324

Vautard, R., Plaut, G., Wang, R. and Brunet, G. 1999 Seasonal prediction of North American surface air temperatures using space–time principal components. *J. Climate*, **12,** 380–394

Wallace, J. M. and Gutzler, D. S. 1981 Teleconnections in the geopotential height field during the northern hemisphere winter. *Mon. Weather Rev.*, **109,** 782–812

Wilks, D. S. 1995 *Statistical methods in the atmospheric sciences: An introduction*. Academic Press