Tellus (2005), 57A, 424–434
Printed in UK. All rights reserved

TELLUS

Downscaling of DEMETER winter seasonal hindcasts over Northern Italy

By V. PAVAN[1]*, S. MARCHESI[1], A. MORGILLO[1], C. CACCIAMANI[1] and F. J. DOBLAS-REYES[2], [1]*ARPA-SIM, Bologna, Italy;* [2]*ECMWF, Shinfield Park, Reading RG2 9AX, UK*

(Manuscript received 31 March 2004; in final form 25 October 2004)

ABSTRACT

A novel method is applied in order to obtain winter predictions over Northern Italy using state-of-the-art multi-model seasonal ensemble hindcasts. The method consists of several stages. In the first stage, the best predictions are computed for a group of eight indices of large-scale circulation variability using the multi-model ensemble data set. The predictions are multiple linear regressions of single-model ensemble mean hindcasts produced within the European project DEMETER using six different coupled models. The regression is obtained using the method of the best linear unbiased estimate (BLUE). In the second stage, a standard statistical downscaling technique of the 'perfect prog' kind is applied in order to predict a group of 12 surface predictands starting from a group of predictors selected between the large-scale indices identified during the first stage. The selection of the predictands is carried out empirically, using those which lead to the best final prediction, while the regression coefficients are defined using observational data only, as in a 'perfect prog' downscaling technique. All steps of the prediction computation up to this point are performed in cross-validation mode. Finally, the full high-resolution surface winter predictions are reconstructed using an adequate selection of the forecasted predictands.

The predictions obtained have a much higher detail than the DEMETER direct model output predictions and, in parts of the domain, they are characterized by substantially significant skill. The improvement of the skill with respect to single-model ensembles is due to the use of the BLUE technique, while the statistical downscaling allows us to increase significantly the detail of the prediction. The study includes a discussion on the sensitivity of the results to both the period in years and the number of models used to produce the forecasts, and a comparison with the results obtained using a simple multi-model forecast in which all models are given the same weight.

1. Introduction

Interannual variability is a very important component of regional climate. Within variability particular attention has always been given to the frequency of occurrence of weather extremes. The occurence of an extreme or even just anomalously intense season can impact several branches of local economy (agriculture, tourism, energy production and consumption, etc.) and also translate in large humanitarian losses (Kunkel et al., 1999; Hartmann et al., 2002; Pielke and Carbone, 2002). Particularly large costs have been recognized to be implied by the occurrence of temperature extremes.

One possible approach to the reduction of losses and the improvement of risk management strategies is to use seasonal forecasts. Although the value of this approach has been largely debated (Katz and Murphy, 1997; Murnane et al., 2002; Pielke and Carbone, 2002), it remains one of the great challenges faced by meteorology and climatology.

In view of this application, great efforts have been made in the improvement of seasonal and climate forecast scores. A good level of reliability has been reached for a selected group of regions and for particular times of the year, such as for the North American continent wintertime seasonal forecasts. All the same, issuing skilful forecasts for other regions, such as Europe, has always proven not to be a straightforward task (Palmer and Anderson, 1994). Furthermore, in regions characterized by a strong spatial variability of climate, in order to be valuable, forecasts must be issued at relatively high resolution, higher than that at which all seasonal forecasts are produced at the main weather centres. This implies that global seasonal forecasts must be downscaled over each region of interest.

All these considerations apply also to a small and geographically complex region such as Northern Italy, where surface fields are characterized by large spatial variability and, at the same time, the weather is influenced by different and sometimes independent patterns of large-scale flow in each subregion

*Corresponding author.
e-mail: vpavan@smr.arpa.emr.it

(Cacciamani et al., 1994; Tomozeiu et al., 2000, 2002b; Quadrelli et al., 2001). It must be kept in mind that given the low level of success of previous attempts, it is possible that the final scores of the forecasts are not significantly different from zero and so of no real value.

The choice made within the present work is to apply a standard statistical downscaling method of the 'perfect prog' type in cross-validation mode (Wilks, 1995) to the winter seasonal hindcasts produced within the European project DEMETER (Development of a European Multi-model Ensemble system for seasonal to inTERannual prediction). This choice is of affordable cost for a local weather service, as it requires the availability of large-scale circulation predictions at monthly time-scales, together with relatively long time series of surface data at high spatial resolution.

The crucial requirement for this choice to lead to acceptable results is that the scores of the large-scale forecasts used as input are significantly good. As mentioned earlier, this is a strong requirement for the Euro-Atlantic region, which has long been known to be characterized by a weakly predictable atmospheric flow at medium to long time-scales. One problem of great relevance is the weakness of large-scale teleconnection patterns over this region. The existence of correlations between atmospheric large-scale circulation anomalies and Atlantic sea surface temperature (SST) anomalies was first addressed in the early studies of Bjerknes (1962, 1964) where the author noticed the existence of a relation between SST anomalies over the North Atlantic and the strength and eastward extension of the Atlantic storm-track. These results were confirmed in later studies (Ratcliffe and Murray, 1970; Palmer and Sun, 1985), often mentioned in support of the existence of some predictability at seasonal time-scales over the Euro-Atlantic (Palmer and Anderson, 1994). Although these results describe the presence of significant correlations between Atlantic SST anomalies and European atmospheric circulation anomalies, several later studies have shown that the intensity of this correlation is weak or not sufficient to support the existence of a significant level of atmospheric predictability at these time-scales (Lin and Derome, 2003).

Another possibility is that planetary scale teleconnection patterns link the occurrence of SST anomalies over the Equatorial Pacific Ocean with circulation anomalies over Europe. The existence of planetary scale teleconnection patterns has been largely documented by many authors (Hoskins and Karoly, 1981; Wallace and Gutzler, 1981; Branstator, 1985; Ferranti et al., 1994). Both modelling and observational studies have shown that although these patterns imply a strong connection between the Pacific SST anomalies and circulation anomalies over the North Pacific and North America, this is not true when these same SST anomalies are related to Euro-Atlantic circulation anomalies (Hoskins and Karoly, 1981; Pavan et al., 2000a).

These problems connected with the intrinsic characteristics of the observed atmospheric flow over Europe are aggravated by problems of all state-of-the-art atmospheric and coupled general circulation models (AGCMs and CGCMs) in representing the mean state and the low-frequency variability of the large-scale flow over Europe (d'Andrea et al., 1998; Doblas-Reyes et al., 2000; Pavan et al., 2000a,b; Pavan and Doblas-Reyes, 2000). The model errors lead to a further reduction in predictability of the large-scale flow to non-significant levels.

All these issues have made the production of valuable operational seasonal forecasts over Europe by single institutes extremely difficult up to present times.

The negative impact of model errors on the final forecast can be reduced by using multi-model ensembles. This method has been successfully applied to both medium-range and climate predictions (Harrison et al., 1996; Krishnamurti et al., 1999; Kharin and Zwiers, 2002). In the present paper, it is planned to apply it to the multi-model ensemble large-scale winter DEMETER hindcasts. In particular, the large-scale multi-model forecast is obtained using the method described in Pavan and Doblas-Reyes (2000), which has already given promising results when applied to the PROVOST experiments. Large-scale predictions are obtained by identifying the main patterns contributing to interannual variability over the Euro-Atlantic region and combining several model data, using different weights for each model and for each pattern considered. Weights are determined using the method of the best linear unbiased estimate (BLUE) suggested by Thompson (1977) and Sarda et al. (1996), using all data available in cross-validation mode. This method allows us to reduce the model bias and error in the representation of the interannual or interseasonal variability of each of the large-scale patterns considered, a possibility which is closed when using either the dynamical or the statistical-dynamical downscaling, or single-model large-scale seasonal hindcasts. Because different models are characterized by different skills in reproducing the variability of each pattern (Hagedorn et al., 2005), this method builds on the possibility of extracting the maximum possible information from the multi-model ensemble for each pattern.

Once the best possible seasonal forecast for each of the large-scale circulation indices is available, this is used within the multiple-regression linear method in order to obtain a prediction for a group of surface local predictands. Finally, the local predictands are used to reconstruct the full high-resolution wintertime seasonal prediction over Northern Italy.

In Section 2, the data sets and the techniques used are briefly described. In Sections 3 and 6, large-scale circulation indices and local predictands characteristics are described with particular attention to their interannual variability and their link with the main teleconnection patterns of the Euro-Atlantic region. The production of the downscaled forecasts consists of two stages. In the first stage, the best seasonal predictions for all large-scale predictors are computed using DEMETER data (Section 4). Section 4 also includes a short discussion of the results with particular attention to their dependence on the number of models in the ensemble and on the period considered. In Section 5 we present a comparison between DEMETER and PROVOST results. The latter stage of

this work consists of the realization of the multiple-regression scheme between a selection of the large-scale circulation indices used as predictors in order to obtain the local predictand forecasts. A description of this last part of the method and of the downscaled prediction scores is given in Section 7. Conclusions and a brief discussion of the results are included in the final section.

2. Data and methods

The identification and description of both large-scale circulation indices and local predictands are carried out using the following data sets: the new European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (hereafter ERA40), the precipitation analysis over Northern Italy and the Alps produced within the Mesoscale Alpine Project (hereafter MAP) and the minimum and maximum surface temperature analyses made available by the Ufficio Centrale di Ecologia Agraria (hereafter UCEA).

ERA40 is the new reanalysis produced at the ECMWF. A description of the project and a preliminary analysis of the characteristics of this data set can be found in Simmons and Gibson (2000) and in Uppala (2002). This data set includes all surface and upper air fields originally produced using the T159L60 version of the Integrated Forecasting System. Analyses are produced every six hours. Within this data set it is possible to find for all fields monthly and daily means and 6-h time-resolution snapshots. The analysis has been run for the period from 1958 to 2002. From this data set we extract only December–January–February–March (DJFM) monthly mean 500-hPa geopotential height (Z500) and 850-hPa temperature (T850) interpolated from the original grid on a $2.5^\circ \times 2.5^\circ$ regular grid, going from 87.5°S to 87.5°N and covering the period from 1959 to 2002.

The characteristics of the MAP precipitation analysis are described in Frei and Schär (1998). This analysis was produced using surface observations recorded by local weather services in the different regions covering the whole Alpine area and Northern Italy. This data set includes daily precipitation analysis over a regular $0.3^\circ \times 0.22^\circ$ grid, for the period 1966–1999. In the present work, monthly mean (DJFM) precipitation fields are used for the regions from 6.6°E to 14.1°E and from 43.72°N to 47.02°N, covering Northern Italy, for the period 1966–1993. All data from the other available years have to be discarded, because in the region of interest more than 20% of the data are missing for each grid point.

The analysis produced by the UCEA includes both minimum and maximum daily temperatures and covers the whole Italian national territory. It has been produced starting from data taken at several groups of observational stations run by the UCEA itself, by the ex-Servizio Idrografico e Mareografico Nazionale, by the Rete Agrometeorologica Nazionale and by the Aereonautica Militare. The data cover the period 1961–1999 and are given on a regular $0.37^\circ \times 0.27^\circ$ grid. Details about the techniques used to produce it can be found in Girolamo and Libertà (1990) and in UCEA (1990).

The seasonal predictions for the large-scale circulation indices are obtained using the data produced within the DEMETER project. A description of this project and of its preliminary results can be found in Palmer et al. (2004). The main deliverable of this project is a data set of seasonal predictions obtained using seven different models. Each model has been run so as to produce ensembles of nine seasonal predictions for each of the four seasons. Within the same ensemble, the nine experiments are initialized using the burst method: the ensemble is initialized on the first day of the month preceding the beginning of the season at 00 GMT, with a spread created using wind stress and SST perturbations. This implies that the seasonal predictions obtained for December for the Euro-Atlantic region are already sufficiently uncorrelated with the initial conditions and, if used, they are unlikely to produce artificially high forecast scores.

In this study, only the following six models are used to produce the forecasts.

(i) ECMWF coupled model (SCWC hereafter), consisting of the T95, 40 levels 23r4 version of the ECMWF AGCM and of HOPE as ocean GCM. This model has been run so as to obtain a set of ensemble seasonal hindcasts from 1958 to 2002.

(ii) Centre National de Recherches Météorologiques coupled model (CNRM hereafter), consisting of ARPEGE-Climat version 3 as an AGCM with T63 truncation and 31 vertical levels, and OPA 8.1 as an oceanic GCM coupled using OASIS.2.2. This model has been also run so as to produce seasonal hindcasts from 1958 to 2002.

(iii) UK Meteorological Office coupled model (UKMO hereafter), which identifies with the HadCM3 climate model. The atmospheric component (HadAM3) has a horizontal resolution of 2.5° latitude by 3.75° longitude, with 19 levels in the vertical. This model has been run so as to produce seasonal hindcasts from 1959 to 2002.

(iv) Max-Planck-Institut für Meteorologie (MPI) coupled model (SMPI hereafter), consisting of the ECHAM5 atmospheric general circulation model at T42 resolution with 19 vertical levels and the MPI-OM1 ocean GCM, coupled using the OASIS coupler. This model has been run so as to produce seasonal hindcasts from 1969 to 2002.

(v) Laboratoire d'Océanographie Dynamique et de Climatologie (LODYC) coupled model (LODY hereafter), consisting of the same atmospheric component of the SCWC model and of the OPA ocean model in its global configuration ORCA2. This model has been run so as to produce seasonal hindcasts from 1974 to 2002.

(vi) Istituto Nazionale di Geofisica e Vulcanologia (INGV) coupled model (SCNR hereafter), consisting of ECHAM4 atmospheric model coupled with the OPA 8.1 ocean model. The atmospheric model is run with 19 hybrid vertical levels and T42 horizontal truncation, while the ocean model has space

resolution roughly equivalent to a geographical mesh of 2° × 1.5° (with a meridional resolution of 0.5° near the equator) and 31 vertical levels. This model has been run so as to produce seasonal hindcasts from 1973 to 2002.

One more model, the European Centre for Research and Advanced Training in Scientific Computation (CERFACS) model, was run within this project, for the period 1980–2001. Given that this period is considerably shorter than all the others, implying the availability of a considerably smaller number of seasons for the downscaling computation, the data from this model are not employed within this work and the results here presented are obtained using only the data of the six other models.

Given that three models have produced seasonal forecasts for a long period and all six for a shorter period, the scores obtained for large-scale patterns for each model and for the multi-model ensemble are analysed by grouping the experiments together in the following way:

(i) long period (DJF 1959–2002) for three models (SCWC, UKMO and CNRM), hereafter LON-3;
(ii) short period (DJF 1974–2002) for three models (SCWC, UKMO and CNRM), hereafter SHO-3;
(iii) short period (DJF 1974–2002) for six models (SCWC, UKMO, CNRM, SMPI, LODY and SCNR), hereafter SHO-6.

This allows us to check the dependence of results on both the number of models included in the multi-model ensemble and the length of the period covered by the predictions.

The patterns of interannual variability of both large-scale upper-air and local surface fields are identified using a standard principal component (PC) analysis applied on monthly mean anomalies for the months from December to March. In order to obtain the seasonal forecasts for each large-scale pattern, the monthly anomalies of each experiment are projected on the EOF patterns associated with the observed large-scale PCs following the same technique as in Pavan and Doblas-Reyes (2000). Only predictions averaged over the whole winter season (DJF) and over each single-model ensemble are considered. The method used to compute the weights to be used to combine the models so as to produce the BLUE multi-model ensemble predictions is the same as in Pavan and Doblas-Reyes (2000). These weights are computed using all data available in cross-validation mode.

Forecast quality is also evaluated for the simple multi-model ensemble, obtained giving to all models the same weight, in order to compare its performance with respect to that of the BLUE prediction.

Once the time series of BLUE predictions for all large-scale circulation indices are available, it is necessary to choose between the indices in order to optimize the downscaled forecasts. The predictors used during this stage must be mutually orthogonal in time and their combination must lead to the best of all available predictions. If all possible independent circulation indices were used, the score of the final forecast would not be the optimal one, because a large number of predictors are used for which a good prediction is not available. All possible subsets of independent predictors are considered and only the combination of predictors leading to the best prediction is chosen. The quality of the prediction has been evaluated by computing the linear correlation between the predicted and observed time series of each predictor. The correlation between predictors is estimated on the period covered by observational data. The regression coefficients between predictors and predictands are evaluated each time using only observational data in cross-validation mode, so as to avoid the correlation between observed and forecasted predictors leading to an overestimation of the final scores. The prediction is finally obtained by substituting the BLUE forecasts to the observational data for the predictors chosen, as in a standard 'perfect prog' downscaling technique (Wilks, 1995).

3. Large-scale circulation indices

A group of large-scale circulation indices is identified starting from the time series of monthly mean anomalies of Z500 and T850 from ERA40 for the months from December to March: the first four PCs of Z500 over the Euro-Atlantic region (from 90°W to 60°E and from 20°N to 90°N) and the first four PCs of T850 over the European region (from 10°W to 60°E and from 30°N to 70°N). Together, the first four PCs of Z500 explain 75.0% of the total Z500 variance, while the first four T850 PCs explain 80.3% of the total T850 variance.

The first four PCs of Z500 have the following characteristics.

(i) PC1, which explains 36.7% of the total variance, is strongly correlated with the North Atlantic Oscillation (NAO) pattern (correlation value 0.86).
(ii) PC2, which explains 15.5% of the total variance, is strongly correlated with the Scandinavian (SCA) pattern (correlation value 0.80) and has a weaker but significant correlation (−0.46) with the Eastern Atlantic (EA) pattern.
(iii) PC3, which explains 13.2% of the total variance, is strongly correlated with the European Blocking (EB; correlation value 0.85).
(iv) PC4, which explains 9.6% of the total variance, is strongly correlated with the EA pattern (correlation value 0.74).

These patterns have been widely described in the literature (Wallace and Gutzler, 1981; Barnston and Livezey, 1987; Rogers, 1990; Tibaldi and Molteni, 1990; Deser and Blackmon, 1993; Tibaldi et al., 1994; Hurrell and van Loon, 1997; Pavan et al., 2000a,b). The correlation values are computed using data for the months from December to February, and averaging the anomalies over each DJF season for the period 1958–2002.

The cross-correlation matrix between the T850 and the Z500 PCs (not shown), computed using only mean DJF data, shows that these two sets of predictors are strongly dependent on each

other, so that care must be taken when choosing the final set of predictors in order to avoid overfitting.

4. DEMETER large-scale winter seasonal predictions

Having defined the circulation indices as the PCs of Z500 and T850 monthly DJFM anomalies, predictions are computed by projecting the monthly mean, ensemble mean anomalies on the associated EOF patterns. This process leads to the production of a time series of forecasts for each model and each index. Forecast quality is evaluated for each model, for the simple multi-model (MM) and for the BLUE multi-model ensemble forecasts for each of the groups identified in Section 2 (LON-3, SHO-3 and SHO-6). Scores are obtained by computing the correlation between ERA40 and predicted PCs. Those obtained for Z500 PCs and for all three groups of forecasts are shown in Table 1, while those for T850 PCs are shown in Table 2.

Table 1. Correlation between mean DJF Z500 forecasted anomaly projection on ERA40 EOFs and observed PCs for the LON-3 group (a), the SHO-3, SHO-6 and SEL-6 groups (b), and for the PROVOST period for DJF and JFM winter definitions for 3 models (c). In parts a and b, values are shown for both single-model, simple multi-model and multi-model BLUE ensembles. Part c includes only BLUE scores

		PC1	PC2	PC3	PC4
	CNRM	0.35[a]	0.36[a]	0.03	0.23
(a)	SCWC	−0.001	−0.02	−0.05	0.08
	UKMO	−0.1	0.13	−0.06	−0.20
	MM	0.13	0.27[a]	−0.05	0.05
	BLUE	**0.40**[b]	**0.38**[a]	**0.08**	**0.33**[a]

		PC1	PC2	PC3	PC4
	CNRM	0.42[a]	0.47[a]	−0.16	0.23
	SCWC	−0.03	−0.05	−0.07	0.04
	UKMO	0.06	0.15	−0.14	−0.25
	SMPI	0.002	0.01	0.23	0.24
(b)	LODY	0.21	−0.05	−0.23	−0.04
	SCNR	0.19	0.19	−0.07	0.41[a]
	MM-3	0.22	0.31	−0.20	0.01
	MM-6	0.26	0.22	−0.15	0.17
	BLUE-3	**0.44**[a]	**0.49**[b]	**0.20**	**0.37**
	BLUE-6	**0.48**[b]	**0.59**[b]	**0.39**[a]	**0.61**[b]
	SEL-6	**0.44**[a]	**0.53**[b]	**0.39**[a]	**0.61**[b]

		PC1	PC2	PC3	PC4
(c)	**BLUE DJF**	0.78[b]	0.67[b]	0.41	0.73[b]
	BLUE JFM	0.58[a]	0.33	0.37	0.36

[a]0.95 confidence level.
[b]0.99 confidence level.

Table 2. Same as Table 1 but for T850

		PC1	PC2	PC3	PC4
	CNRM	0.18	−0.06	0.28	0.27
(a)	SCWC	−0.04	0.18	0.05	−0.04
	UKMO	−0.06	0.19	0.05	−0.22
	MM	0.05	0.16	0.18	−0.001
	BLUE	**0.21**	**0.25**	**0.28**	**0.34**[a]

		PC1	PC2	PC3	PC4
	CNRM	0.002	−0.06	0.31	0.43[a]
	SCWC	−0.14	0.07	0.02	0.17
	UKMO	−0.05	0.11	0.002	−0.29
	SMPI	0.01	−0.06	0.39[a]	−0.25
(b)	LODY	0.13	0.17	0.003	0.11
	SCNR	0.03	0.08	0.11	−0.09
	MM-3	−0.11	0.06	0.16	0.17
	MM-6	−0.005	0.13	0.21	0.04
	BLUE-3	**0.15**	**0.12**	**0.35**	**0.50**[b]
	BLUE-6	**0.25**	**0.23**	**0.52**[b]	**0.51**[b]
	SEL-6	**0.24**	**0.23**	**0.47**[a]	**0.51**[b]

		PC1	PC2	PC3	PC4
(c)	**BLUE DJF**	0.37	0.17	0.64[a]	0.12
	BLUE JFM	0.74[b]	0.26	0.59[a]	0.34

[a]0.95 confidence level.
[b]0.99 confidence level.

Given that several of the models used are built in a similar way, or even share parts of their structure, it is possible that the predictions obtained with each of them are significantly correlated with the others. This could be a source of overfitting within the method. A check on the single-model large-scale indices time series shows that neither the LON-3 nor the SHO-3 experiment series are significantly affected by this problem, apart from PC3-T850. Results are obtained by evaluating the correlation between time series and assuming two series are linearly dependent when they have a value of correlation different from 0.0 at 95% confidence level. In this last case, the model characterized by a lower correlation with ERA40 is discarded from computations and the BLUE prediction is recomputed using only two of the three models. The new prediction has a correlation with ERA40 equal to 0.32, which is not significantly different from the BLUE-3 value 0.35 in Table 2b. As for the SHO-6 experiments, several single-model time series have significant correlation with each other. For this reason, a new BLUE prediction is computed and its correlation with ERA40 is presented on the last line of Tables 1b and 2b (SEL-6). This prediction is obtained by using for each index a selection of the six single-model predictions available, including only those time series which can be considered independent at

95% confidence level. In none of the cases, apart from PC2-Z500 and PC3T850, are the final SEL-6 results substantially different from the BLUE-6 results.

Tables 1 and 2 show that sometimes the correlation between BLUE predictions and observations is statistically significant at the 99% level, while in most cases no significance is associated with the correlation between single-model results and observations – a result that by itself indicates the great superiority of this with respect to the single-model approach. Relatively low skill is associated with the prediction of PC3 of Z500, strongly correlated with a pattern, the EB, characterized by low predictability. Tables 1 and 2 also show that, for each predictor, the BLUE score is always as good as the best model if not better. This confirms the result, shown in Pavan and Doblas-Reyes (2000), that the use of the BLUE technique allows us to extract all useful information from the numerical forecasts and can help increase the skill of seasonal predictions with respect to single-model ensembles, by removing part of the model bias and by reducing misrepresentation of the interannual variability. BLUE scores are also always significantly better than those obtained for the simple multi-model ensemble.

A comparison between SHO-3 and LON-3 results indicates that the performance of the method is strongly dependent on the length of the time period considered. This dependence may be linked with problems of the models in representing the strong decadal variability of Z500 large-scale patterns. In order to check this hypothesis, the seasonal prediction scores are compared with those of the forecasts obtained after filtering the data, so as to remove the decadal time-scale variability. This is done by simply subtracting from the original data a centred 9-yr running mean. This operation has slightly shortened the long time series, reducing it to the period 1963–1998, a length which is still substantially longer than the short period. The scores obtained for the filtered forecasts of both Z500 and T850 predictors are shown in Table 3. Single-model skill for the filtered time series is better than that of the full unfiltered time series only in some cases. All the same, this has a positive impact on the BLUE predictions, which, for Z500, are always substantially better than those obtained for the unfiltered series. In the Z500 case, values are also comparable with those obtained for the SHO-3 BLUE predictions apart for PC3. This same result does not hold for all T850 PCs, possibly due to the minor relevance of decadal variability for this field. These results suggest, on the one hand, that DEMETER experiments have low skill in reproducing large-scale decadal variability, but that, on the other hand, this component is of relatively low importance on the final. These results are encouraging in view of an application of the method here described to the production of operational seasonal forecasts. In fact, they suggest that in order to reproduce large-scale forecast scores comparable to those here presented, only relatively short time series of seasonal forecasts obtained using the same model are needed.

Comparing SHO-3 results with SHO-6 results, it is always possible to notice an improvement in the forecast skill when increasing the number of models used within the ensemble, suggesting the importance of using the greatest possible number of models when computing multi-model seasonal forecasts.

5. Comparison between DEMETER and PROVOST results

In general, DEMETER single-model scores are low, if not very low. In particular, both Z500 and T850 single-model scores for the ECMWF and UKMO models are much lower than those obtained in the PROVOST project (Pavan and Doblas-Reyes, 2000). In order to make the comparison more objective, Tables 4a and b show the correlations between the predicted and observed anomalies with respect to the same EOF patterns, but using PROVOST data experiments, for both Z500

Table 3. Same as Tables 1a and 2a but for filtered data

		PC1	PC2	PC3	PC4
(a)	CNRM	0.43[a]	0.47[b]	−0.09	0.36[a]
	SCWC	0.08	−0.06	0.06	0.05
	UKMO	−0.06	0.27	0.005	−0.13
	BLUE	**0.45**[b]	**0.54**[b]	**0.10**	**0.39**[a]

		PC1	PC2	PC3	PC4
(b)	CNRM	−0.03	−0.32	0.20	0.24
	SCWC	−0.15	0.20	0.17	0.07
	UKMO	0.02	0.17	−0.004	−0.15
	BLUE	**0.15**	**0.40**[a]	**0.22**	**0.31**

[a]0.95 confidence level.
[b]0.99 confidence level.

Table 4. Correlations between predicted and observed PC anomalies for Z500 (a) and T850 (b) from PROVOST experiments. Values refer to time series of JFM averages

		PC1	PC2	PC3	PC4
(a)	EDF	0.28	0.29	0.36	−0.08
	ECMWF	0.22	0.60[a]	0.36	−0.21
	UKMO	0.57[a]	0.10	0.18	0.34
	BLUE	0.59[a]	0.62[a]	0.41	0.47

		PC1	PC2	PC3	PC4
(b)	EDF	0.03	0.10	−0.10	0.09
	ECMWF	0.09	0.06	−0.01	0.20
	UKMO	0.30	−0.30	0.36	0.20
	BLUE	0.37	0.28	0.43	0.17

[a]0.95 confidence level.

and T850. In Table 4, EDF indicates the winter experiments run by Electricité de France using the ARPEGE model at T63 horizontal resolution, ECMWF denotes the European Centre model, and UKMO denotes the UK Met Office model. Notice that, in the case of PROVOST data, the JFM hindcasts refer to months 2–4 of each integration, while in the case of DEMETER data, JFM hindcasts refer to months 3–5 of each integration. This implies that the two sets of forecasts are characterized by different lead times. Given that the minimum lead is one month, all forecasts are anyway equally decorrelated with the initial conditions. Comparing Table 1a with Table 4a, DEMETER Z500 ECMWF and UKMO values are much lower than those of PROVOST, while the CNRM model skill seems to be comparable. This produces a net worsening of Z500 BLUE scores in DEMETER with respect to PROVOST. As for T850, the UKMO model skill is worse in DEMETER, while the CNRM skill is slightly better and ECMWF has comparable skill in the two groups of experiments. The worsening of UKMO skill produces a decrease in BLUE scores in DEMETER with respect to PROVOST.

These changes in forecast quality can be a result of PROVOST experiments being run for a shorter and more predictable time period, or a result of the difference between DJF and JFM as a winter definition, or, finally, as a result of a change in the models used to produce the experiments. In fact, the models used to run PROVOST experiments were atmosphere-only GCMs, forced with observed (ERA15) SST data, while the models used in DEMETER were obtained by coupling a new and possibly improved version of the PROVOST AGCMs with ocean GCMs.

Tables 1c and 2c report the BLUE scores for the DEMETER experiments over the 1979–1993 period obtained using only SCWC, UKMO and CNRM models, for both DJF and JFM winter definitions. Single-model scores are not presented, as a detailed discussion of the performance of single models is beyond the scope of this paper.

The comparison between Tables 1a and 2a and the BLUE scores of the DEMETER experiments over the PROVOST period (DJF 1979–93), in Tables 1c and 2c, shows that within this period both Z500 and T850 DJF BLUE scores are better or comparable with respect to those for the full LONG period (this is connected to the general better performance of UKMO and CNRM models, for Z500, and CNRM and SCWC models, for T850, over this period).

Changing the winter definition from DJF to JFM produces a slight improvement of T850 scores but a slight worsening of Z500 scores. The general worsening of Z500 scores as spring starts is consistent with the results presented in the literature and linked to a worsening of single-model scores for March with respect to December (not shown) (Palmer and Anderson, 1994; Branković and Palmer, 2000).

The analysis carried out suggests that, only in some cases, the changes in score due to a change in the winter definition and, especially, to the time period considered are able to explain the differences between PROVOST and DEMETER results.

6. Local predictands

The objective of our work is to produce a high-resolution seasonal forecast for wintertime precipitation and minimum and maximum temperatures over Northern Italy, which takes into account the local main characteristics of surface climate. The skill of a very detailed forecast taking into account the full spatial and temporal variability might be low, because the details are mostly related to unpredictable and, in some cases, random processes. This part of the variability can be shown to explain a relatively small fraction of the total variance. As a consequence, we decided to choose as local predictands the first four PCs of each of the three surface fields of interest. This should allow us to produce a full forecast for each field, taking into account the most important local climate anomalies, without covering the predictable part of the signal with the unpredictable noise.

This choice leads to the identification of the following 12 predictands: the first four PCs of precipitation (PREC) over Northern Italy, explaining 53.1%, 15.3%, 8.9% and 4.2% of the total variance, respectively; the first four PCs of minimum temperature ($T_{\min}$) over the same area, explaining 76.9%, 10.5%, 5.4% and 1.9% of the total variance respectively; the first four PCs of maximum temperature($T_{\max}$), explaining 86.0%, 5.9%, 2.6% and 1.3% of the total variance. In all cases, the first PC explains most of the total variance, this being particularly true for temperature, a fact already observed also in other regions (Tomozeiu et al., 2002a).

Retaining only these PCs in the computation, the final downscaled forecast can at most reproduce 81.5% of the total variance in the case of precipitation, 94.7% of the total variance in the case of minimum temperature and 95.8% of the total variance in case of maximum temperature.

From the correlation matrix between the 12 predictands and the eight predictors (not shown), it can be noticed that all predictands have values of correlation significant at the 95% level with at least one predictor. In particular, the first PC of each field presents significant values of correlation with two or more large-scale circulation indices. This result is encouraging in the prospect of building a statistical downscaling method, because, for each local predictand, it opens the possibility of choosing the predictors of the downscaling between several large-scale indices, all significantly correlated with the predictand considered, depending on the skill of their forecasts.

7. Downscaled forecasts

Once the best possible forecasts are available for the large-scale indices, the predictands are forecast using a multiple-regression technique. The values of the coefficients used within this computation are determined using observational data for both predictors

Table 5. Values of correlation between observed and predicted predictands for the three groups of experiments (LON-3, SHO-3 and SHO-6) and for the experiments SEL-6 for precipitation (a), T_{min} (b) and T_{max} (c).

		LON-3	SHO-3	SHO-6	SEL-6
(a)	PC1 PREC	0.34	0.37	0.50[a]	0.44
	PC2 PREC	0.56[b]	0.62[b]	0.47[a]	0.49[a]
	PC3 PREC	0.39[a]	0.36	0.42	0.36
	PC4 PREC	0.26	0.32	0.43	0.42

		LON-3	SHO-3	SHO-6	SEL-6
(b)	PC1 T_{min}	0.30	0.38	0.61[b]	0.60[b]
	PC2 T_{min}	0.30	0.33	0.37	0.35
	PC3 T_{min}	0.19	0.14	0.12	0.12
	PC4 T_{min}	0.22	0.30	0.32	0.32

		LON-3	SHO-3	SHO-6	SEL-6
(c)	PC1 T_{max}	0.40[a]	0.57[b]	0.73[b]	0.67[b]
	PC2 T_{max}	0.20	0.33	0.34	0.35
	PC3 T_{max}	0.33[a]	0.38	0.53[b]	0.45[a]
	PC4 T_{max}	0.27	0.35	0.53[b]	0.51[a]

[a]0.95 confidence level.
[b]0.99 confidence level.

and predictands, from the full period over which both data sets are available (1966–1993 for precipitation and 1961–1999 for temperature) in cross-validation mode. The exact choice of the set of predictors to be used is made year-by-year by computing the coefficients for all possible subsets of independent predictors and choosing the set leading to the best forecast once observed predictors are substituted by forecasted predictors. The independence of the predictors is checked for the period over which the observations are available. Predictors are considered linearly dependent if their correlation is different from 0.0 at the 0.99 confidence level. The forecast score is obtained computing the correlation between observed and predicted seasonal values for each predictand. Tables 5a–c report the scores obtained for the three groups of four predictands and for the SEL-6 group.

In several cases, the scores are significantly high. In particular, very good scores are always obtained for PC1 -T_{max} and PC2-PREC, while poor scores are obtained for all groups of forecasts only for PC4-PREC, for PC2 -T_{max} and for all PCs of T_{min} but PC1. Remembering that in this case PC1 explains most of the total variance of all these fields, we can consider this result quite satisfactory.

In general, some improvements can be noticed when going from the LON-3 group to the SHO-3 group of predictions, and also when increasing the number of models used to produce the forecasts.

Starting from the predictions of the first four PCs of each field and the associated observational EOFs, the predicted full field anomalies are reconstructed at each grid point of Northern Italy and for each of the three variables considered. Figure 1 shows the anomaly correlation coefficient (ACC) between the reconstructed full field anomaly time series and observations for the SHO-3 experiments. Results are shown only for these experiments for brevity and because it seems the most relevant to possible future operational applications.

Grid point values of correlation are always relatively high, especially for T_{max}. Comparing Figs. 1b and c, we notice that T_{max} seems much more predictable than T_{min}over the south-western part of the domain. A preliminary investigation has shown that these results may be influenced by the poor quality of the data used. In particular, it has been found that T_{min} presents a particularly high number of missing values over part of the domain and that in some cases it presents higher values than T_{max}. In these last cases, T_{max} has still a reasonable value, while the T_{min} high value may be linked again to the great number of missing within the grid box considered.

Although at present the UCEA analysis is the only high-resolution temperature data set over our region, a new and improved data set is going to be produced by this same institute. As soon as this new data set is available, it will be possible to check the validity of these results.

We suspect that a similar problem also affects precipitation, which presents a region of relatively low scores over the south-western part of the domain. In this case, the final forecast has been upscaled to the three grid points at which ERA40 reanalysis data are available, falling over the Po Valley (grid point 1 being the westernmost of the three). The final upscaled–downscaled (UD) prediction can be compared with the simple multi-model prediction obtained using the direct model output (DMO) from DEMETER large-scale forecasts. The upscaling is carried out retaining all non-missing values so that the UD prediction is less sensitive to the presence of missing data in the validation data set.

Table 6 shows both the ACC and the root mean square error (RMSE) for the UD and for the DMO predictions over the three grid points for the SHO-3 group. From this table, it is clear that the proposed method of downscaling has produced a substantial improvement of the forecast, in terms of both ACC and RMSE. A comparison of Fig. 1 with Table 6 shows that the value of ACC in each grid point is substantially greater than the mean of the high-resolution ACC over the associated grid box, confirming the bad influence of missing data on the final high-resolution downscaled forecast. All the same, the upscaling cannot reduce the impact of the artificial variability introduced by missing data in the computation of the downscaled forecast. We conclude that an improvement of the quality of the original observation data is expected to produce a net improvement in the quality of the downscaled prediction.

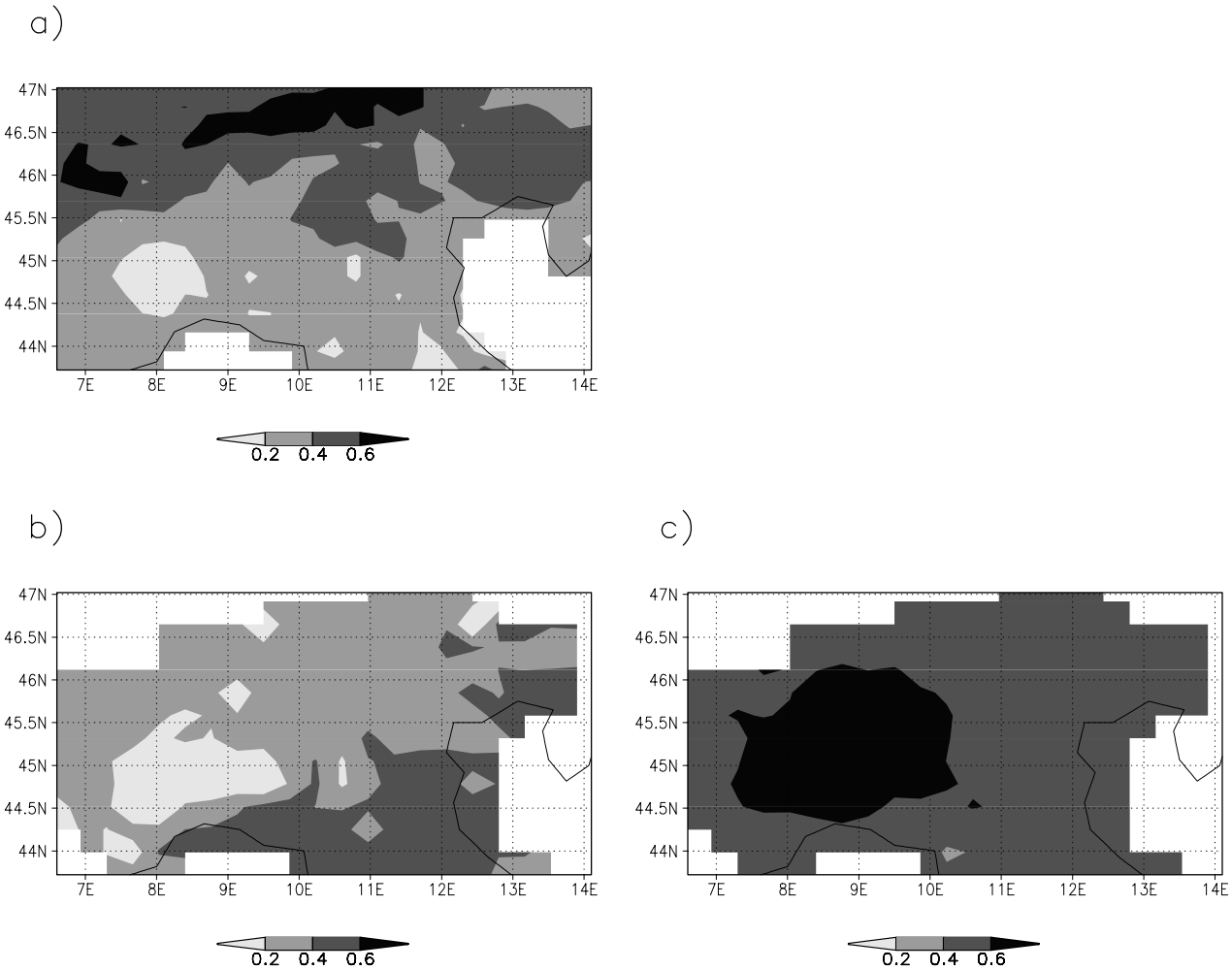

Fig 1. ACC for mean DJF downscaled forecasts of (a) precipitation (b) $T_{\min}$ and (c) $T_{\max}$ obtained using the SHO-3 group of experiments. Contour interval 0.2.

Table 6. ACC and RMSE for the UD predictions and for the DMO over the three Po Valley grid points for PREC

	ACC		RMSE	
	UD	DMO	UD	DMO
Grid point 1	0.51[a]	0.14	0.80	0.87
Grid point 2	0.42	0.10	0.48	0.54
Grid point 3	0.29	0.02	0.63	0.65

[a]0.95 confidence level.

A final check has been made on the possible effects on the final downscaled results of the presence of overfitting due to linear dependence between the models. In none of the three surface fields considered does the use of a selection of models lead to substantial differences in terms of ACC, although small differences can be found in the values of SHO-6 and SEL-6 correlations for some of the predictands, shown in the last column of Tables 5a–c.

8. Conclusions

State-of-the-art seasonal forecasts for Europe, as they are issued by the main weather centres, are of limited value, because of both their coarse resolution and the presence of model errors affecting the mean state and the interannual variability. In this paper, we propose a method which is meant both to reduce model errors and to increase the resolution of the final product. The method is used to produce downscaled seasonal forecasts over Northern Italy starting from DEMETER hindcasts. As for now, the results are limited to the winter (DJF) season and to mean seasonal anomalies of precipitation, $T_{\min}$ and $T_{\max}$.

The method used within this work consists of two stages. First, the best seasonal mean DJF forecasts are obtained from DEMETER multi-model ensemble experiments using all available data in cross-validation mode, for a group of possible large-scale circulation indices. Then, these forecasts are used within a multiple regression scheme of the 'perfect prog' type, also used in

cross-validation mode, in order to obtain detailed seasonal forecasts for a group of local predictands describing the interannual variability of precipitation, $T_{\min}$ and $T_{\max}$.

The large-scale indices seasonal forecast time series are produced by combining several model results using different weights for each model in order to obtain a forecast for each index (BLUE). This technique produces better results with respect to both single-model and simple multi-model ensemble forecasts. The multi-model approach shows with this example its superiority versus a single-model strategy.

The dependence of the results on both the time period and the number of models considered is analysed. Increasing the number of models within the ensemble always has positive impacts on the final prediction skill. As for the dependence on the time period considered, it is shown that in most cases multi-model forecast skills are comparable for long and short periods, once the decadal variability is removed from the data. This is related to the fact that models have some problems in reproducing this component of the variability. The robustness of the multi-model BLUE forecast skill with respect to the period considered is encouraging on the prospect of producing operational downscaled seasonal forecasts using the method described here. In fact, this method requires the existence of time series of seasonal forecasts run using always the same model. Although, in general, short series of this type are made available by most climate prediction centres, it is very unlikely that their time length extends beyond 20 yr. The independence of the final results from the period considered suggests that this length is sufficient in order to determine the values of the coefficients.

In the second part of this work, the downscaled forecast for Northern Italy were produced for three fields: precipitation, $T_{\min}$ and $T_{\max}$. First, the full surface fields are filtered using a standard PC analysis and only the first four PCs are retained. These are identified as the local predictands in the multiple linear regression scheme. The final forecast is then obtained recombining the associated EOFs using the predicted PCs.

The prediction displayed a satisfactory forecast skill for each of the fields considered, although some problems are found in a small region in the south-western part of the domain for precipitation and over the whole domain for $T_{\min}$. A preliminary analysis has shown that these problems may be a result of the poor quality of the observational data used to describe the interannual variability of the surface fields and to verify the forecasts.

9. Acknowledgments

AM has been funded under the Italian national project CLIMAGRI. FJDR has been funded by the European Union under the project DEMETER (EVK2–1999-00024). This work has also been partially funded by the Italian national project SINAPSI. The authors are grateful to two anonymous reviewers who have helped to improve the quality of this paper.

References

Barnston, A. G. and Livezey, R. E. 1987. Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.* **115**, 1083–1126.

Bjerknes, J. 1962. Synoptic surveyof the interaction of sea and atmosphere in the North Atlantic. *Geophys. Norvegica* **24**(3), 115–145.

Bjerknes, J. 1964. Atlantic air–sea interaction. *Adv. Geophys.* **10**, 1–82.

Brancković, Č. and Palmer, T. N. 2000. Seasonal skill amd predictability of ECMWF PROVOST ensembles. *Q. J. R. Meteorol. Soc.* **126**, 2035–2068.

Branstator, G. 1985. Analysis of general circulation model sea surface temperature anomaly simulations using a linear model. Part I: forced solutions. *J. Atmos. Sci.* **42**, 2225–2241.

Cacciamani, C., Nanni, S. and Tibaldi, S. 1994. Mesoclimatology of winter temperature and precipitation in the Po Valley of Northern Italy. *Int. J. Climatol.* **14**, 777–814.

d'Andrea, F., Tibaldi, S., Blackburn, M., Boer, G., Déqué, M. and co-authors. 1998. Northern Hemisphere atmospheric blocking as simulated by 15 atmospheric general circulation models in the period 1979–1988. *Clim. Dyn.* **14**, 385–407.

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

Doblas-Reyes, F. J., Déqué, M. and Piedeliévre, J.-Ph. 2000. Multi-model spread and probabilistic seasonal forecasts of the North Atlantic Oscillation. *Q. J. R. Meteorol. Soc.* **126**, 2069–2088.

Ferranti, L., Molteni, F. and Palmer, T. N. 1994. Impact of localized and extratropical SST anomalies in ensembles of seasonal GCM integrations. *Q. J. R. Meteorol. Soc.* **120**, 1613–1645.

Frei, C. and Schär, C. 1998. A precipitation climatology of the Alps from high-resolution rain-gauge observations. *Int. J. Climatol.* **18**, 873–900.

Girolamo, A. and Libertà, A. 1990. A national climatic database: the Italian experience. Internal Technical Report, AGRISIEL SpA.

Hagedorn, R., Doblas-Reyes, F. J. and Palmer, T. N. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. *Tellus* **57 A**, 219–233.

Harrison, M. S. J., Palmer, T. N., Richardson, D. S., Buizza, R. and Petroliagis, T. 1996. Joint ensembles from the UKMO and ECMWF models. In: *Proceedings of ECMWF seminar on Predictability* (4–8 September 1995), Vol. II, ECMWF, Reading, UK, 61–120.

Hartmann, H. C., Pagano, T. C., Sorooshian, S. and Bales, R. 2002. Confidence builder: evaluating seasonal climate forecasts for user perspectives. *Bull. Am. Meteorol. Soc.* **83**, 683–698.

Hoskins, B. J. and Karoly, D. J. 1981. The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.* **38**, 1179–1196.

Hurrell, J. W. and van Loon, H. 1997. Decadal variations in climate associated with the North Atlantic oscillation. *Clim. Change* **36**, 301–326.

Katz, R. and Murphy, A. 1997. *Economic Value of Weather and Climate Forecasts*. Cambridge University Press, Cambridge, 1–222.

Kharin, V. V. and Zwiers, F. W. 2002. Climate prediction with multimodel ensembles. *J. Climate* **15**, 793–799.

Krishnamurti, T. N., Kishtawal, C. M., Timoty, E. L., Bachiochi, D. R., Zhang, Z. and co-authors. 1999. Improved weather and seasonal climate forecasts from multi-model superensemble. *Science* **285**, 1548–1550.

Kunkel, K., Pielke, R. A. Jr. and Changnon, S. A. 1999. Temporal fluctuations in weather and climate extremes that cause economic and human health impacts: a review. *Bull. Am. Meteorol. Soc.* **80**, 1077–1098.

Lin, H. and Derome, J. 2003. The atmospheric response to North Atlantic SST anomalies in unseasonal prediction experiments. *Tellus* **55A**, 193–207.

Murnane, R. J., Crowe, M., Eustis, A., Howard, S., Koepsell, J. and co-authors. 2002. The weather risk manegement industry's climate forecast and data needs. *Bull. Am. Meteorol. Soc.* **83**, 1193–1198.

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

Palmer, T. N. and Sun, Z. 1985. A modelling and observational study of the relationship between sea surface temperature anomalies in the north-west Atlantic and the atmospheric general circulation. *Q. J. R. Meteorol. Soc.* **111**, 947–975.

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M. and co-authors. 2004. DEMETER: Development of a European multi-model ensemble system for seasonal to interannual prediction. *Bull. Am. Meteorol. Soc.* **85**, 853–872.

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

Pavan, V., Molteni, F. and Branković, Č. 2000a. Wintertime variability in the Euro-Atlantic region in observations and in ECMWF seasonal ensemble experiments. *Q. J. R. Meteorol. Soc.* **126**, 2143–2173.

Pavan, V., Tibaldi, S. and Branković, Č. 2000b. Seasonal prediction of blocking frequency: results from winter ensemble experiments. *Q. J. R. Meteorol. Soc.* **126**, 2125–2142.

Pielke, R. Jr. and Carbone, R. E. 2002. Weather impacts, forecasts, and policy. *Bull. Am. Meteorol. Soc.* **83**, 393–403.

Quadrelli, R., Lazzeri, M., Cacciamani, C. and Tibaldi, S. 2001. Observed winter alpine precipitation variability and links with large-scale circulation patterns. *Climate Res.* **17**, 275–284.

Ratcliffe, R. A. S. and Murray, R. 1970. New lag assosiation between North Atlantic sea temperature and European pressure applied to long-range weather forecasting. *Q. J. R. Meteorol. Soc.* **96**, 226–246.

Rogers, J. C. 1990. Patterns of low-frequency monthly sea-level pressure variability (1899–1989) and associated wave cyclones frequencies. *J. Climate* **3**, 1364–1379.

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

Simmons, A. J. and Gibson, J. K. 2000. The ERA-40 Project Plan. ERA-40 Project Report Series No. 1. ECMWF, Reading, UK.

Thompson, P. D. 1977. How to improve accuracy by combining independent forecasts. *Mon. Wea. Rev.* **105**, 228–229.

Tibaldi, S. and Molteni, F. 1990. On the operational predictability of blocking. *Tellus* **42A**, 343–365.

Tibaldi, S., Tosi, E., Navarra, A. and Pedulli, L. 1994. Northern and Southern Hemisphere seasonal variability of blocking frequency and predictability. *Mon. Wea. Rev.* **122**, 1971–2003.

Tomozeiu, R., Busuioc, A., Marletto, V., Zinoni, F. and Cacciamani, C. 2000. Detection of changes in the summer precipitation time series of the region Emilia-Romagna, Italy. *Theor. Appl. Climatol.* **67**, 193–200.

Tomozeiu, R., Busuioc, A. and Stefan, S. 2002a. Changes in seasonal mean maximum air temperature in Romania and their connection with large-scale circulation. *Int. J. Climatol.* **22**, 181–196.

Tomozeiu, R., Lazzeri, M. and Cacciamani, C. 2002b. Precipitation fluctuations during winter season from 1960 to 1995 over Emilia-Romagna, Italy. *Theor. Appl. Climatol.* **72**, 221–229.

UCEA 1990. Analisi climatologica e progettazione della rete agrometeorologica nazionale. Technical report for the Ministero dell'Agricoltura e delle foreste, Direzione generale della produzione agricola. UCEA, Rome, Italy.

Uppala, S. 2002. ECMWF Reanalysis 1957–2001, ERA-40. In: *ERA-40 Project Report Series No. 3*, ECMWF, Reading, 1–10, (also available from http://www.ecmwf.int/publications/library/ecpublications/proceedings/ERA40-reanalysis_workshop/index.html).

Wallace, J. M. and Gutzler, D. S. 1981. Teleconnection in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.* **109**, 784–812.

Wilks, S. D. 1995. *Statistical Methods in the Atmospheric Sciences*, International Geophysics Series (eds R. Dmowska and J. R. Holton), Academic, New York, 1–467.