*Q. J. R. Meteorol. Soc.* (2000), **126**, pp. 2035–2067

# Seasonal skill and predictability of ECMWF PROVOST ensembles

By ČEDO BRANKOVIĆ* and T. N. PALMER
*European Centre for Medium-Range Weather Forecasts, UK*

(Received 20 November 1998; revised 20 October 1999)

SUMMARY

Variations in seasonal-forecasting skill and predictability during the 15 years (1979–93) of the European Centre for Medium-Range Weather Forecasts (ECMWF) re-analysis (ERA), have been studied using 120-day ensemble integrations of the ECMWF model. These integrations form part of the European Union PROVOST (PRediction Of climate Variations On Seasonal to interannual Time-scales) project. Observed sea surface temperatures (SSTs) were updated daily at the model's lower boundary. Two major and three moderate El Niño Southern Oscillation (ENSO) events occurred during the ERA period. Results are interpreted as giving an upper bound on the predictive skill of a coupled ocean–atmosphere system as a function of season, location, and state of ENSO.

The model systematic error was found to be comparable with a typical amplitude of interannual variation. When standardized by the corresponding ERA anomaly variance, systematic error appears to be largest in boreal spring in the northern extratropics, and in boreal summer in the tropics.

Ensemble-mean skill scores were found to be positive overall. Apart from the northern winter season, the ensemble-mean skill for months 2–4 drops significantly when compared with months 1–3. The interannual variation of skill scores is much larger for the European region than for the hemispheric domain. Over the northern hemisphere, skill is much higher when only ENSO years are considered; for Europe, the enhancement in skill for ENSO years is much weaker.

Estimates of intrinsic predictability were made for each year of the dataset. These estimates, defined both by a $t$-test and variance ratio, indicate generally high predictability in years when ENSO was strong. Apart from northern winter, the predictability estimates also showed a systematic drop between months 1–3 and months 2–4. It is therefore concluded that the fall in skill scores between months 1–3 and 2–4 indicates more a weakening of the impact of initial conditions (ICs) than, say, an increase in the effects of model error. In order to study this further, the relative impacts of SSTs and ICs, including land surface ICs, on interannual variation of precipitation have been examined in an additional set of experiments. Overall, SSTs have a dominant role, though the impact of ICs is not negligible.

The predictability of tropical and extratropical precipitation is also discussed. The level of skill for precipitation in the extratropics is generally lower than in the tropics. However, within the tropics there are regions where the precipitation exhibits chaotic behaviour and is correspondingly less predictable.

KEYWORDS: El Niño Ensemble prediction Seasonal predictability

## 1. INTRODUCTION

This is one of two companion papers discussing seasonal-predictability estimates from a set of ensemble integrations made with the European Centre for Medium-Range Weather Forecasts (ECMWF) model. These integrations form part of the European Union PROVOST (PRediction Of climate Variations On Seasonal to interannual Time-scales) project. The main objectives of the PROVOST project were: (1) to quantify the predictability of the atmosphere on seasonal time-scales, using ensembles of integrations of atmospheric models with specified observed sea surface temperatures (SSTs); (2) to develop comprehensive coupled models of the oceans, atmosphere and land surface together with ocean data assimilation techniques to provide initial conditions (ICs) for such coupled models; and (3) to test the skill of these coupled systems through co-ordinated experimentation.

This paper addresses the first of these objectives. Within the terms of the PROVOST uncoupled experimentation, 9-member ensembles of 120-day integrations were made by four different modelling groups: ECMWF, Electricité de France, Météo France and The Met. Office. The ensembles were made with prescribed observed SSTs, implying that the results give upper bounds on predictive skill (though see the caveats in section 8).

* Corresponding author: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK

Seasonal-simulation studies using specified observed SSTs have an extensive history (e.g. Blackmon *et al.* 1983, Shukla and Wallace 1983; Lau 1985; Owen and Palmer 1987; Kitoh 1991; Branković *et al.* 1994; Deque *et al.* 1994; Barnett 1995; Dix and Hunt 1995; Stern and Miyakoda 1995; Harzallah and Sadourny 1995; Bengtsson *et al.* 1996; Davies *et al.* 1997; Kumar and Hoerling 1998). However, the integrations discussed in this paper differ in one or more ways from these earlier studies. First, the integrations are made over seasonal time-scales from observed (re-analysis) initial conditions. Second, the integrations are made over all four seasons. Third, the integrations have been made over a relatively large number of years. As a result, variations of skill can be assessed as a function of season, region, and state of El Niño Southern Oscillation (ENSO). Finally, the integrations have been made in ensemble mode. Taking into account all four models participating in PROVOST, the maximum size of such a multi-model ensemble is 36, a relatively large number by past standards. However for the purposes of this paper, analysis has been restricted to the 'sub-ensemble' from just one model, the ECMWF model (the 36-member multi-model ensemble is discussed in Palmer *et al.* 2000). A pilot study for this PROVOST experimentation was reported in Branković and Palmer (1997).

Initial dates of the integrations spanned the 15-year period of the ECMWF re-analysis (1979–93). More details concerning the model and the ensemble construction are given in section 2. The systematic error of the ECMWF model is analysed in section 3, and is compared with the standard deviation of observed interannual variability. The main analysis is performed in sections 4–6. Conventional estimates of skill are described in section 4. In section 5 attention is focussed on the simulation of interannual variation of precipitation anomalies in a number of specified tropical and extratropical regions. In view of the deleterious effect of model error, measures of intrinsic predictability (which are independent of the verifying analysis) are discussed in section 6. These measures include both a $t$-statistic and an analysis of variance. In section 7 we discuss results from an additional set of experiments, designed to assess the relative impact of initial conditions and SST forcing on the seasonal predictability of precipitation. Conclusions are summarized in section 8.

The use of ensemble forecasts for seasonal prediction derives from the fact that such forecasts are essentially probabilistic. An analysis of the skill of these PROVOST integrations from a probabilistic perspective is made in Palmer *et al.* (2000).

## 2. Organization of experiments

### (*a*) *The model*

The model version used for the integrations described in this paper was ECMWF cycle 13R4 with semi-Lagrangian dynamics at T63L31* resolution. This model cycle was introduced into operations (at T213L31) on 4 April 1995 (see Miller *et al.* 1995). One of the main reasons for choosing cycle 13R4 was that the same model was used in the ECMWF 15-year re-analysis project (ERA-15; Gibson *et al.* 1997).

The decision to use semi-Lagrangian time stepping was based on results from seasonal-time-scale trial integrations. Consistent with Chen and Bates (1996), it was found, for example, that a strong cooling of the polar stratosphere, common to many general circulation models with Eulerian schemes (see, for example, Boer *et al.* 1991), was dramatically reduced with the semi-Lagrangian scheme. It is well known that semi-Lagrangian dynamical schemes are not necessarily mass conserving, and this may pose

* Spectral triangular truncation T63 and 31 levels in the vertical.

TABLE 1. INITIAL DATES FOR THE PROVOST EXPERIMENTS

| Member | Spring* (MAMJ) | Summer (JJAS) | Autumn (SOND) | Winter (DJFM) |
|---|---|---|---|---|
| 1 | 20 February | 23 May | 23 August | 22 November |
| 2 | 21 February | 24 May | 24 August | 23 November |
| 3 | 22 February | 25 May | 25 August | 24 November |
| 4 | 23 February | 26 May | 26 August | 25 November |
| 5 | 24 February | 27 May | 27 August | 26 November |
| 6 | 25 February | 28 May | 28 August | 27 November |
| 7 | 26 February | 29 May | 29 August | 28 November |
| 8 | 27 February | 30 May | 30 August | 29 November |
| 9 | 28 February | 31 May | 31 August | 30 November |

*Add one day for leap years

a problem for long model integrations. However, in the ECMWF model it was found that the globally averaged change in surface pressure was at most 3 mb at the end of the four-month integration period. This was considered sufficiently small, so that no corrective fixes were applied.

### (*b*) *Seasonal ensembles*

Nine-member ensembles were run over all seasons for the ERA-15 period, 1979–93. The experimental set consists of ensembles for 14 northern winter and 15 northern spring, summer and autumn seasons. The experiments were initiated from consecutive 1200 UTC ERA-15 analyses, from 1 to 9 days preceding the season of interest (see Table 1), i.e. initial conditions were separated by 24 hours. The length of integration was 4 months plus 1 to 9 days depending on the initial date, so as to include the full calendar month at the end of each season. The seasons are defined as: winter, December–March (DJFM); spring, March–June (MAMJ); summer, June–September (JJAS); and autumn, September–December (SOND). The model was run with observed prescribed SSTs taken from ERA-15 and updated daily in the integrations. The choice of *atmospheric* initial conditions should not strongly affect seasonal predictability, since on these time-scales the dominant forcing is provided by SSTs. However, initial conditions in our experiments also include land surface initial conditions which may influence seasonal predictability (see discussion in section 7).

The model output was archived every 24 hours at 1200 UTC, and monthly and seasonal averages were derived from such daily data. For verification purposes, the same mean fields were computed from ERA-15. For precipitation, the ERA-15 'verification' was based on 24-hour accumulated rainfall from 24-hour forecasts run from the ERA-15 analyses. We have also used an independent set of precipitation analyses based mainly on observation data (Xie and Arkin 1997; hereafter referred to as Xie–Arkin data). Ten-day mean data from the ECMWF PROVOST experiments have also been archived on CD-ROMs (Becker 1998).

The model seasonal climatology is defined as the mean of all ensemble members over the ERA-15 period (14 years for winter and 15 years for other seasons). The observed climate was computed in the same way from the ERA-15 data. In all cases below, the simulated anomalies are defined relative to the model climate, the ERA-15 anomalies are defined relative to the observed climate. In this way, comparison of simulated and observed anomalies implies that an *a posteriori* linear correction for model systematic error has been applied.

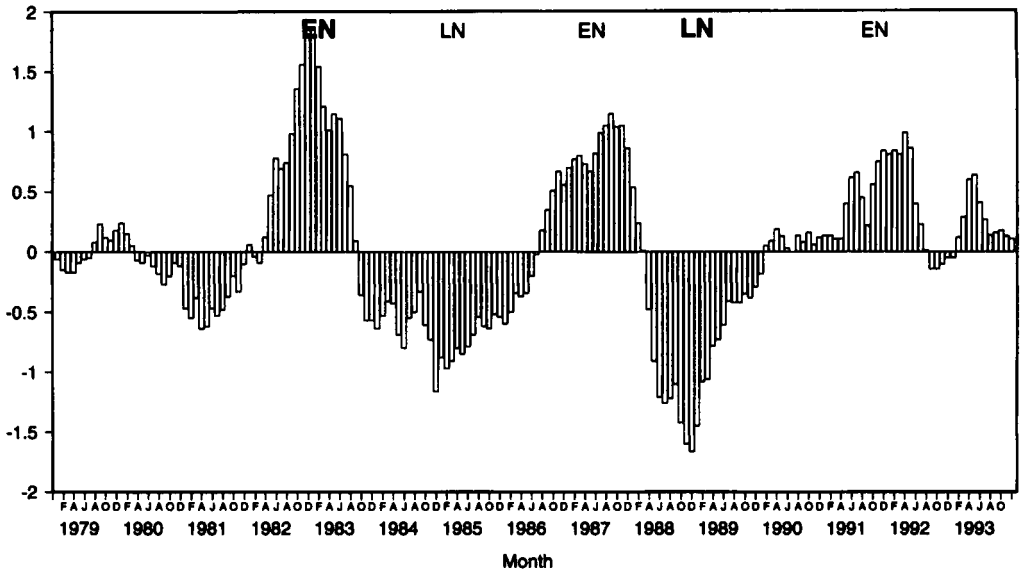


Figure 1. Monthly mean sea surface temperature anomalies (K) from 1979 to 1993 averaged over the equatorial Pacific region, 7°N–7°S, 160°E–80°W. Strong El Niño/La Niña Southern Oscillation events are denoted by large bold EN and LN; moderately strong events are denoted by smaller EN and LN.

### (*c*) *SST anomalies during ERA-15*

It is well known that SST anomalies in the equatorial Pacific have a significant influence on the global circulation (e.g. Trenberth *et al.* 1998). Figure 1 shows monthly mean SST anomalies from 1979 to 1993 averaged over the region 7°N–7°S, 160°E–80°W. Two periods of large anomalies, with magnitude over 1.5 K, are seen: one warm, El Niño, event in 1982/83; and one cold, La Niña, event in 1988/89. Other moderately strong events were the El Niño in 1986/87 and 1991/92, and a prolonged La Niña from 1984 to 1986. The cold event in 1981 and the warm event in 1993 were weaker than these major events. The spatial distribution of SST anomalies for the period considered can readily be found elsewhere (see, for example, *Climate Diagnostics Bulletin*, published by the Climate Prediction Center, NOAA/NWS/NCEP*) and will not be shown here.

## 3. MODEL SYSTEMATIC ERROR

### (*a*) *Errors in the extratropics*

For brevity, we focus on seasonal-mean 500 mb height systematic error. Figure 2(a) shows the systematic error for JFM, the last 3 months (months 2–4) of the winter integrations. The largest errors, of approximately 9 dam, occur over the north Pacific and north Atlantic, and are associated with a mid-latitude westerly bias over these oceanic areas. The error pattern, in particular over the Atlantic, is almost unchanged compared with that of previous ECMWF model cycles (Branković and Molteni 1997). The error amplitude in JFM is, on average, only marginally larger than in DJF (not shown). Over the northern hemisphere, an error pattern similar to that of JFM is found for the late spring season (AMJ, Fig. 2(b)). The trough-to-ridge amplitude in both Atlantic and

* National Oceanic and Atmospheric Administration/National Weather Service/National Centers for Environmental Prediction.

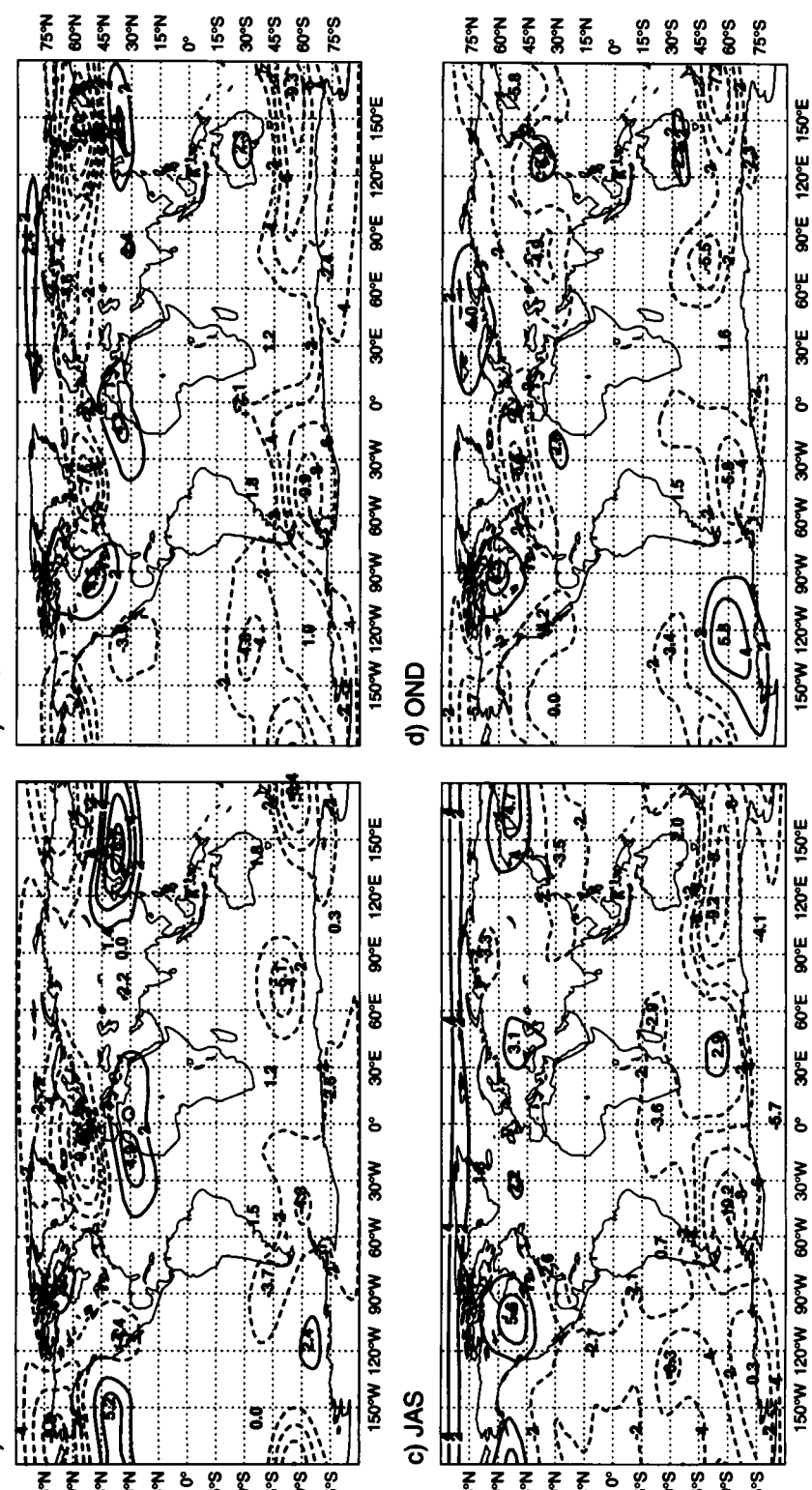


Figure 2. ECMWF model 500 mb height systematic errors for three-month periods: (a) JFM; (b) AMJ; (c) JAS; and (d) OND. Contours are every 2 dam; negative errors are dashed. See text for details.

Pacific error dipoles is somewhat smaller, whilst the negative error in the Pacific extends over the whole of northern Asia. This error is associated with a relatively large cooling over the same region extending almost over the whole depth of the model troposphere (not shown). (The erroneous cooling was found to be the consequence of an inadequate treatment of the snow albedo over the northern hemisphere's large forest areas in the model version used for the PROVOST integrations (Viterbo and Betts 1999).) For the late summer (JAS), the northern hemisphere error is much smaller, apart from at very high latitudes, and positive errors are dominant over the continents (Fig. 2(c)). In the southern hemisphere, the amplitude of the JAS 500 mb height error reaches its maximum of 10 dam in the southern Atlantic. Between JAS and OND (Fig. 2(d)) negative errors in the southern hemisphere are reduced, whereas a positive error in the south Pacific, near Antarctica, has increased substantially. There is an indication that the error in the southern Indian Ocean, reaching about 5 dam in JFM and OND, might be partly caused by isolated and possibly biased radiosonde soundings at Kerguelen Island (Per Kållberg, the ERA project, personal communication), and therefore should not be ascribed solely to the model.

The magnitude of model error, as illustrated in Fig. 2 (locally up to 10 dam), is similar to the magnitude of observed atmospheric interannual variability. Indeed, the standard deviation for ERA-15 anomalies reaches 9.1 dam in JFM (not shown), and is smaller in other seasons (up to 7.8 dam in MAM, 6.9 dam in JAS and 6.0 dam in OND). Since the model error depicted in Fig. 2 is clearly influenced by the annual cycle, the magnitude of model systematic error for each season has been normalized by the corresponding seasonal standard deviation of analysis anomaly (Fig. 3). In shaded areas the model systematic error is greater than the amplitude of observed interannual variation. In the northern extratropics, this *relative* systematic error appears to be largest in spring (Fig. 3(b)) and smallest in summer (Fig. 3(c)). In the tropics, the relative error is largest in summer, followed by spring. (However, the 500 mb height field is rather featureless in the tropics, and its prediction is not as important as in the extratropics). In contrast with the other seasons, relative errors in winter are confined mainly to extratropical regions. The results from Figs. 2 and 3 suggest that model errors may have less impact on model performance in winter, despite the mean error having larger absolute amplitude.

Given the intrinsic nonlinearity of the atmosphere, it is clear that further work is required to reduce the level of model error on seasonal time-scales, at least so that it is small compared with the signal one is attempting to predict. The progress that has been made in reducing model error in medium-range forecasts, suggests that the level of systematic error shown in Fig. 2 can be significantly decreased over the coming years.

### (*b*) *Errors in the tropics*

For tropical regions, model errors are assessed in terms of precipitation. As discussed above, the verification is based upon both Xie–Arkin data and the 0–24 hour accumulated precipitation from ERA-15 short-range forecasts. The latter may be viewed as a shortcoming in verification; however, an 'error' field with respect to ERA-15 gives a large-scale estimate, albeit possibly conservative, of the drift in precipitation over a season.

Figure 4 shows the differences between the simulated climate and verification for the last 3 months (months 2–4) of extreme seasons (JFM in the upper two panels and JAS in the lower two panels). Over the central and eastern tropical Pacific, the error pattern in both seasons signifies excessive precipitation rates in the simulated inter-tropical convergence zone (ITCZ). A similar error pattern is found in the other seasons,

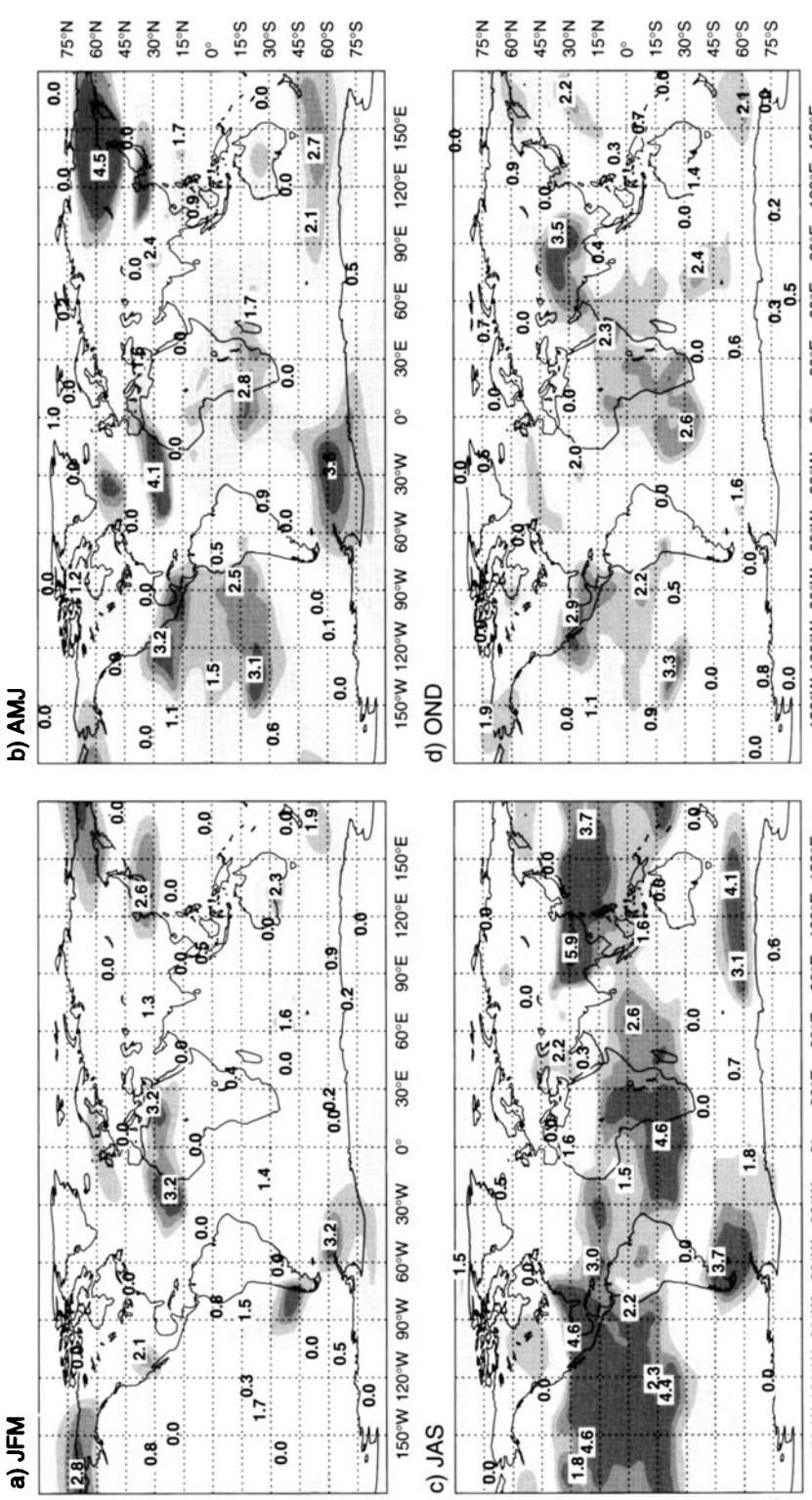


Figure 3. The ratio of ECMWF model 500 mb height systematic errors and corresponding standard deviation of ERA anomaly in three month periods: (a) JFM, (b) AMJ, (c) JAS, (d) OND. Contours are at 1., 1.5, 2., 2.5 and 5. See text for details.

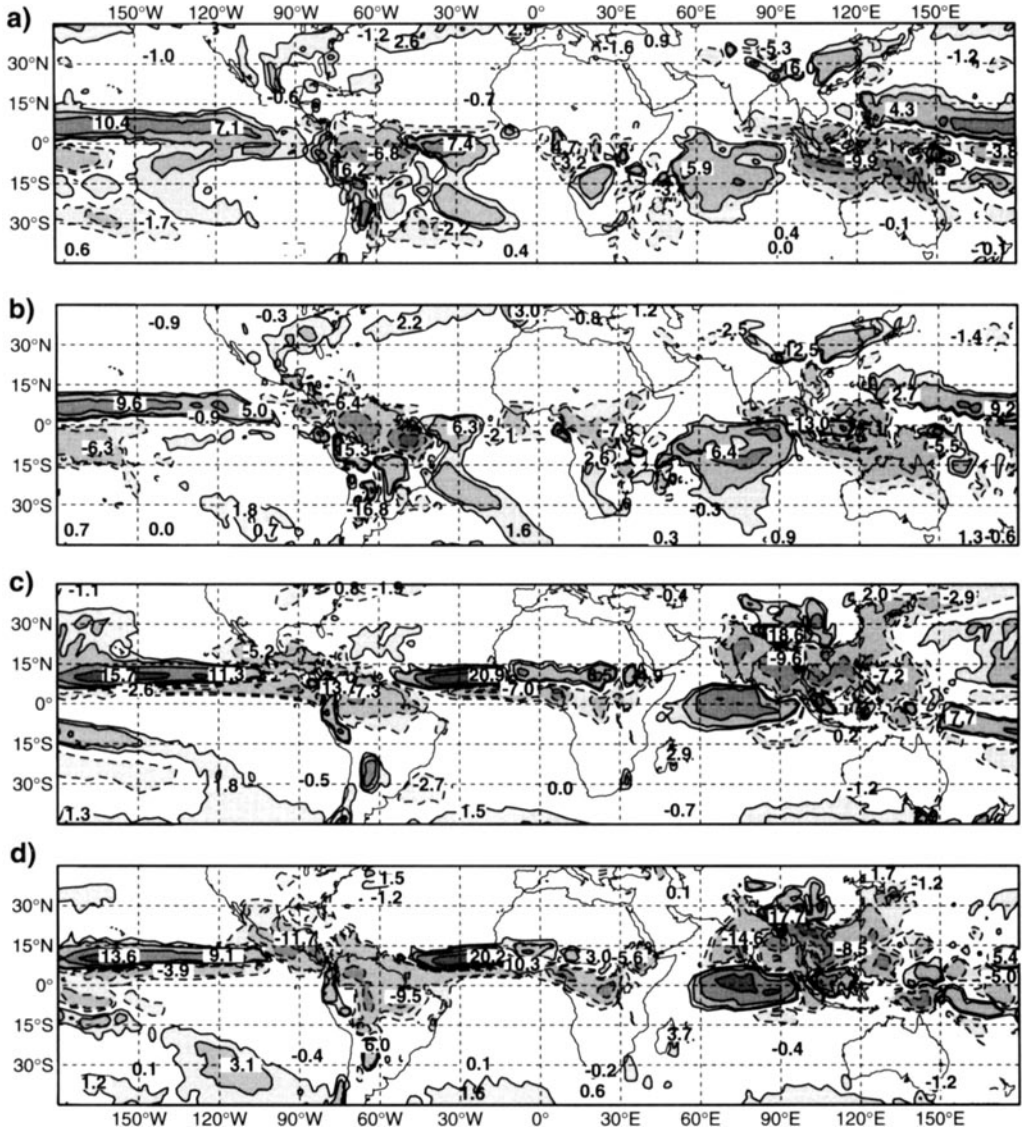


Figure 4. Total precipitation error for: (a) JFM with respect to Xie–Arkin; (b) JFM with respect to ERA; (c) same as (a) but for JAS; (d) same as (b) but for JAS. Contours are $\pm 1, 2, 5, 10, 20$ mm day$^{-1}$; negative errors are dashed. See text for details.

AMJ and OND (not shown). In JAS (Fig. 4(c) and (d)), the excessive ITCZ rainfall is the largest of all seasons, reaching between 14 and 16 mm day$^{-1}$ in the tropical Pacific, around 20 mm day$^{-1}$ in the Atlantic and over 10 mm day$^{-1}$ in the tropical Indian Ocean. The amplitude and the extent of errors in the tropical rainfall in JAS are consistent with the relatively large errors in geopotential heights shown in Fig. 3(c). They are also consistent with the model's overestimation of the near-surface convergence in the ITCZ (not shown). In contrast with the overestimation of the tropical ocean rainfall, the model underestimates rainfall rates in the Amazon region and equatorial Africa. The model differences with respect to both Xie–Arkin data and to ERA are very similar.

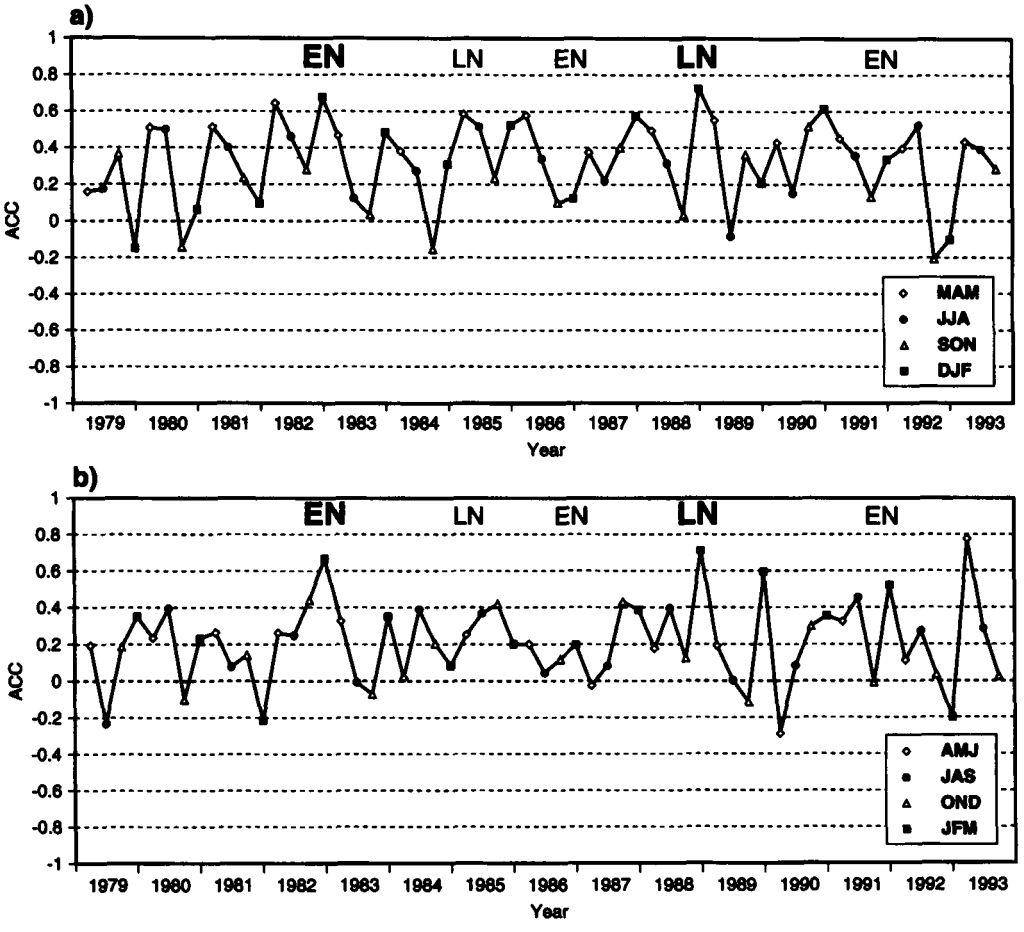


Figure 5. Northern hemisphere (20–80°N) 500 mb height ensemble-mean anomaly correlation coefficients for all four seasons for: (a) months 1–3; and (b) months 2–4. Strong El Niño/La Niña Southern Oscillation events are denoted by large bold EN and LN; moderately strong events are denoted by small EN and LN. Symbols correspond to three-month periods, see key.

## 4. Assessment of model skill

This section describes some conventional skill scores, appropriate to deterministic ensemble-mean fields. Palmer *et al.* (2000) discuss the skill of the PROVOST integrations within a probabilistic-forecast framework.

### (*a*) *Ensemble-mean skill*

Figures 5 and 6 show anomaly correlation skill scores for the ensemble-mean 500 mb height, averaged over months 1–3, and months 2–4 of the integrations. Figure 5 is for the northern hemisphere between 20 and 80°N; Fig. 6 is for Europe only (35–75°N, 12.5°W–42.5°E). Results for the four seasons are shown separately (see legend on Figs. 5 and 6). Clearly, the results show an overall positive level of skill, both for the northern hemisphere and for Europe. Over the northern hemisphere, skill scores for months 1–3 are at least equal to 0.5 in 17 (out of 59) seasons, nearly 30% of cases. However, for months 2–4 skill scores are at least equal to 0.5 in only five (out of 59)

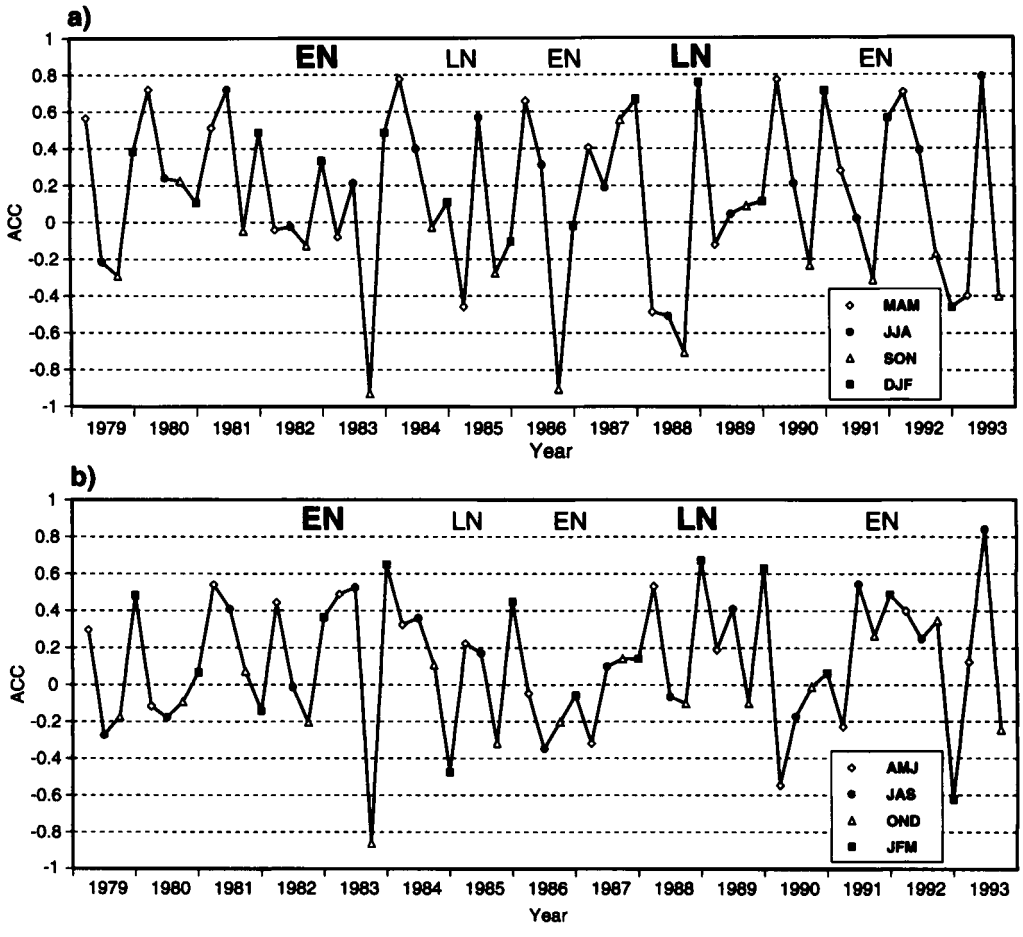


Figure 6. As Fig. 5 but for Europe (35–75°N, 12.5°W–42.5°E).

cases. The drop in skill between months 1–3 and months 2–4 is especially large in spring and summer.

Interannual variability in skill is much more marked for Europe than for the whole hemisphere, not least because of its smaller area. However, regardless of this, the skill for months 2–4 is more often above the 0.5 mark for Europe (11 cases) than for the northern hemisphere (five cases). In autumn, skill is rather poor for Europe—in SON skill is negative (worse than climatology) in 12 out of 15 years (Fig. 6(a)).

The level of skill is enhanced for the northern hemisphere when only years in which ENSO was active are considered. This can be seen readily in Fig. 5; for example, the El Niño winter of 1982/83, and the La Niña winter of 1988/89 stand out as having particularly high skill. On the other hand, skill over Europe is not well correlated with ENSO events (Fig. 6). For example, whilst winter 1988/89 has high skill, relatively low skill (below 0.4) is found in the winter of 1982/83.

In contrast, European skill is found to be relatively high in some years that are not related to ENSO. For example, high skill is found in DJF 1987/88, which was a transition season from a relatively strong El Niño to a strong La Niña. However, in the late winter, JFM 1988, the skill score drops substantially below that for DJF. Similar scores are

TABLE 2. 500 mb HEIGHT ENSEMBLE-MEAN ANOMALY CORRELATION COEFFICIENTS

| Season | N hemisphere | Europe | N America |
|---|---|---|---|
| MAM | 0.47 (0.48) | 0.31 (0.12) | 0.51 (0.59) |
| AMJ | 0.22 (0.17) | 0.16 (0.21) | 0.22 (0.24) |
| JJA | 0.35 (0.40) | 0.28 (0.13) | 0.34 (0.44) |
| JAS | 0.21 (0.30) | 0.22 (0.31) | 0.15 (0.32) |
| SON | 0.20 (0.30) | −0.31 (−0.17) | 0.37 (0.50) |
| OND | 0.16 (0.35) | −0.11 (0.02) | 0.25 (0.47) |
| DJF | 0.36 (0.52) | 0.31 (0.36) | 0.55 (0.75) |
| JFM | 0.34 (0.53) | 0.21 (0.23) | 0.44 (0.76) |

Figures in brackets are the means from strong and moderately strong ENSO years

found in the winter of 1990/91 which was characterized by a weak warm SST anomaly in the equatorial Pacific (see Fig. 1). These two examples indicate that factors other than SST forcing from the equatorial Pacific may influence European skill in early winter.

Table 2 shows the average anomaly correlation skill scores of 500 mb height for three chosen regions over all years, for the months 1–3 and months 2–4. The skill scores are averaged using the Fisher $z$-transform technique. For the northern hemisphere the highest skill is in spring (MAM), though there is a sharp drop in the late spring (AMJ, months 2–4). This is consistent with the results from an earlier ECMWF model version, where the highest skill over the three northern hemisphere sectors was found in MAM (Branković and Palmer 1997). Other studies have also found that the northern spring season has relatively high predictability (Davies *et al.* 1997; Kumar and Hoerling 1998). The second-highest skill score is in winter (DJF), and is maintained into the late season (JFM). The contrasting behaviour of skill scores in the late spring and late winter seasons may indicate the weakening of dynamical teleconnections between the tropics and the northern hemisphere extratropics as summer approaches. On the other hand, it may indicate that the effects of ICs are more predictable in spring than in winter. The impact of ICs on seasonal predictability is further discussed in sections 6 and 7.

For Europe, the mean skill is lower than for the northern hemisphere. There is no clear winner between the winter, spring and summer seasons. For autumn, the ECMWF model performs very poorly, having a negative anomaly correlation. For North America (25–70°N, 150–60°W), on the other hand, the highest skill is in winter (DJF) and is maintained at a relatively high level into the late winter (JFM). Skill in summer and autumn clearly lags behind that in the colder part of the year.

When only strong and moderately strong ENSO events are considered (figures in brackets in Table 2), skill is consistently enhanced for North America. For Europe the skill, despite being relatively low, is consistently enhanced for months 2–4 (AMJ, JAS, OND and JFM) and also for DJF. This may indicate a delay in the time it takes for ENSO to influence seasonal predictability over the European region. However, due to the small sample (five cases), values in Table 2 may not be considered definitive. (In contrast to these results, forecasts for the ENSO winter 1997/98 were found to be quite skilful for the European region, see Stockdale *et al.* (1998).)

Many of the scores shown above were also computed in the cross-validation mode, i.e. using a definition for which the anomaly for year $x$ was computed by using a climate in which year $x$ was omitted. This made no noticeable difference to the skill score results. Over the northern hemisphere, the difference between any of the 'variable' climate fields and the 'full' climate does not exceed 2 dam.

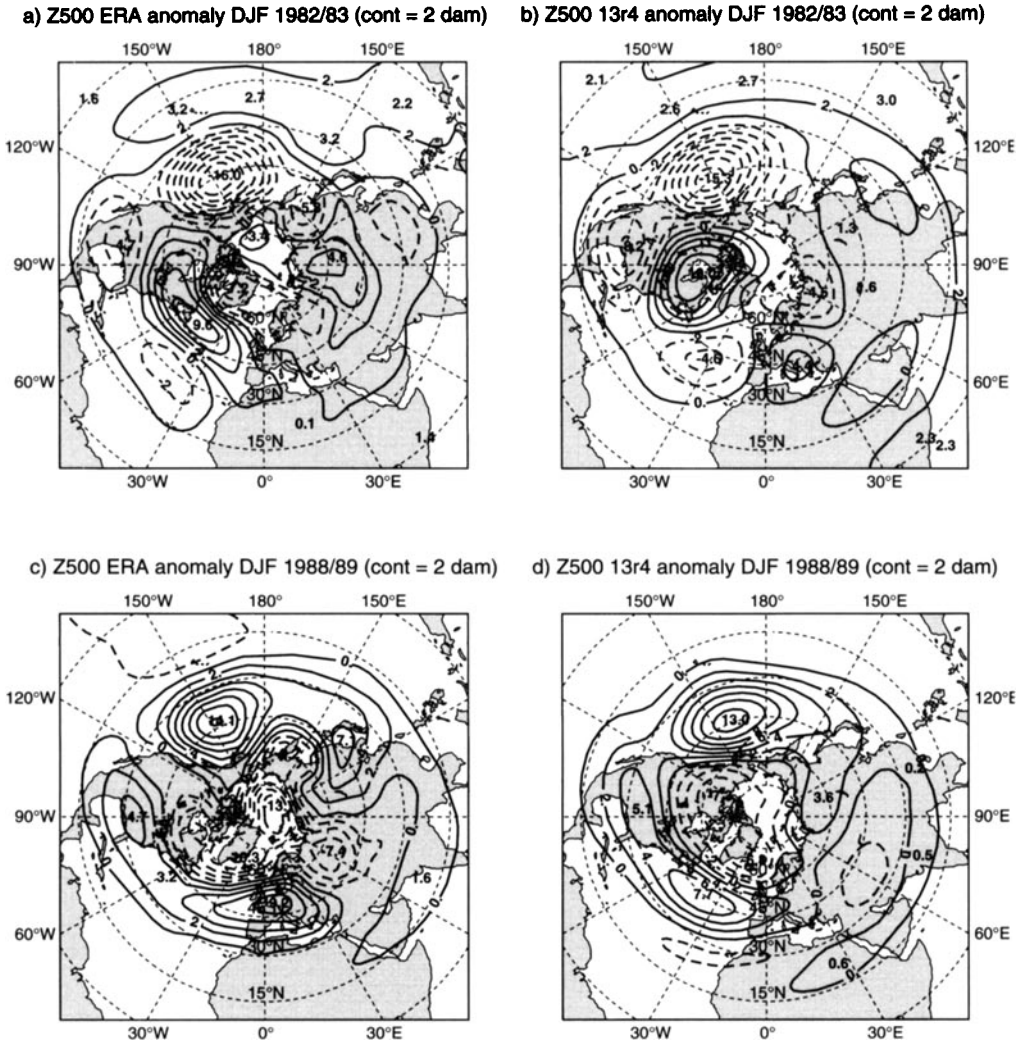


Figure 7. 500 mb height anomalies: (a) ERA for DJF 1982/83; (b) ensemble-mean for DJF 1982/83. (c) and (d) are as (a) and (b), respectively, but for DJF 1988/89. Contours are every 2 dam; negative anomalies are dashed. See text for details.

Figure 7 shows the 500 mb height anomaly over the northern hemisphere from ERA-15 and from the ensemble mean for the two winters when the northern hemisphere skill was high. In DJF 1982/83, the model simulated anomalies rather well over the Pacific/North American (PNA) quadrant of the northern hemisphere. Over Europe and the north Atlantic, the simulated anomalies bear much less correspondence with the observed anomalies. In DJF 1988/89, the simulation of a positive anomaly over the north Atlantic and Europe (Fig. 7(d)) gave a much higher skill score for Europe and overall for the whole of the northern hemisphere. It is interesting to note that, whilst the 1988/89 observed anomaly pattern over the PNA region is almost exactly the opposite to that in 1982/83, over the European region it is somewhat similar in both winters with, broadly, negative anomalies in the north and positive anomalies in the south. This illustrates the

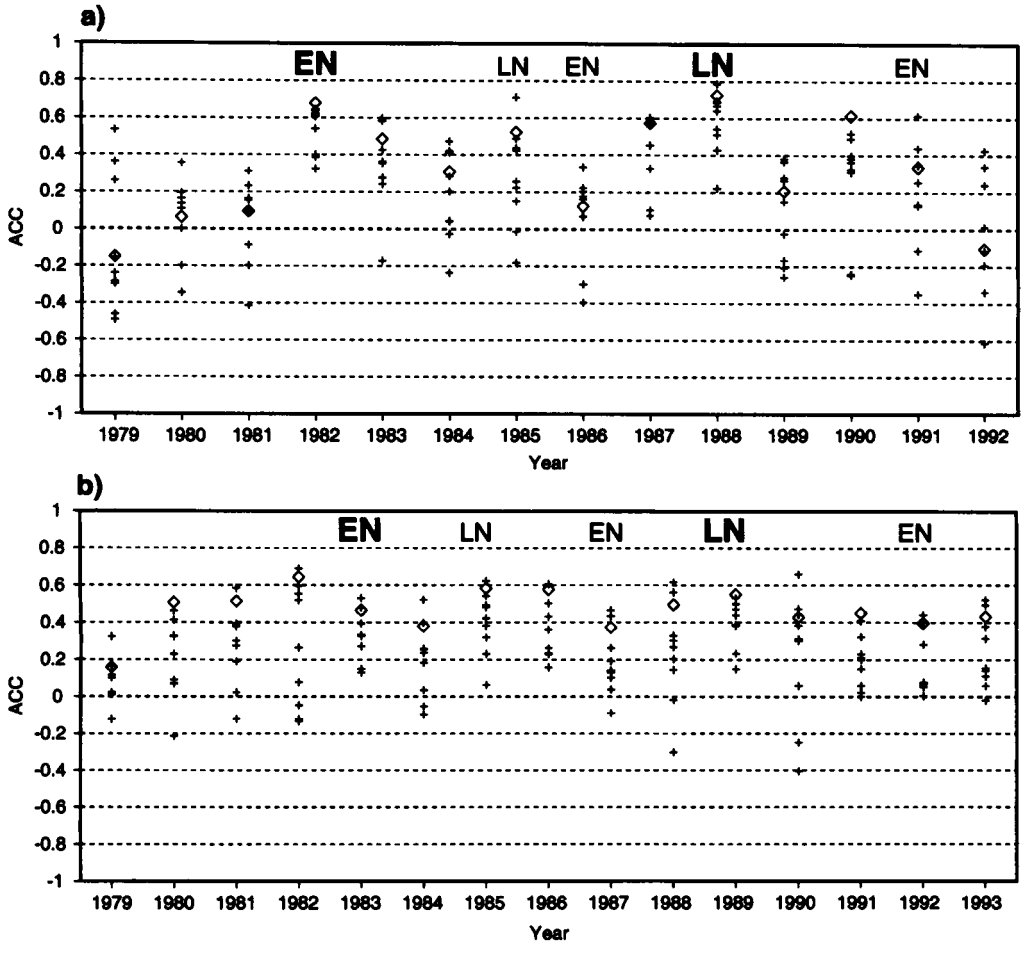


Figure 8. Northern hemisphere (20–80°N) 500 mb height anomaly correlation coefficients for: (a) DJF, and (b) MAM. Open diamonds correspond to ensemble means, and crosses to individual model integrations. See text for details.

difficulties in seasonal prediction over Europe even in the presence of a strong ENSO forcing.

### (*b*) *Skill of individual integrations*

The ensemble-mean skill characterizes one aspect of the forecast performance of the ensemble. However, in addition, it is important to assess the spread or dispersion of the skill of individual forecasts within an ensemble. For example, it is possible that the verification field lies within the range of the ensemble, but that the skill of the ensemble-mean field is low.

Figure 8 shows 500 mb height anomaly correlations of individual model realizations for the northern hemisphere in DJF and MAM. Open diamonds depict ensemble-mean skill, crosses are related to individual forecasts. Clearly, in spring (Fig. 8(b)) there is an overall higher degree of skill consistency within ensembles than in winter (Fig. 8(a)). This is reflected in a consistent level of skill for this season over the 15-year period, without much interannual variability. The smallest intra-ensemble dispersion of

skill scores is found for the two strong ENSO winters 1982/83 and 1988/89; in both years more than half of the ensemble integrations have anomaly correlation coefficients greater than 0.6. In other years a much lower degree of skill consistency is found.

A discussion of the relationship between such skill scores and population (sample) statistics is given in section 6(c).

## 5. SIMULATION OF OBSERVED PRECIPITATION ANOMALIES

From a practical point of view, prediction of seasonal-mean precipitation anomalies is of prime interest, particularly in the tropics. Figures 9 and 11 show simulations of interannual variations of such anomalies for some prescribed tropical and extratropical regions, respectively. The solid lines show the rainfall anomalies based on the ensemble-mean fields; the dashed lines show the anomalies in terms of the 0–24 hour accumulated field in the ERA-15 archives. Open diamonds indicate Xie–Arkin precipitation anomalies. The dotted lines show the extreme (wet and dry) members of the ensemble. In all cases the curves have been standardized with respect to their own climatologies (i.e. the mean is subtracted, and the anomaly divided by the standard deviation, based on values from the appropriate 15-year dataset).

From Fig. 9 it can be seen that seasonal simulations of precipitation for some regions, like the Brazilian Nordeste (MAM; Fig. 9(a)), east Africa (MAM; Fig. 9(b)) and the African Sahel (JJA; Fig. 9(c)) are clearly skilful. For the latter region, it is interesting to note that the model has successfully captured the decadal trend in precipitation. These results highlight the fact that considerable skill can be expected from seasonal forecasts in many parts of the tropics.

For India (JJA; Fig. 9(d)) the model was not as successful, particularly in the second half of the ERA-15 period, though some variations (e.g. between the El Niño year 1987 and La Niña year 1988) were simulated. The difficulty in modelling Indian rainfall has also been reported by other authors (e.g. Palmer *et al.* 1992; Fennessy *et al.* 1994; Krishnamurti *et al.* 1995; Arpe *et al.* 1998). This lack of model success is in part associated with the chaotic nature of monsoon variations (Palmer 1994). This chaotic variability is illustrated in Fig. 10 which shows precipitation anomalies for individual integrations for the JJA season in 1987 over the Indian region. Negative precipitation anomalies dominate over the Indian subcontinent in ensemble members 1, 7 and 9; positive anomalies dominate in ensemble members 2, 3 and 6. In ensemble members 4, 5 and 8 neither anomaly prevails. All the ensemble members have been initiated from late May 1987 initial data separated by 24 hours (see Table 1). The relatively large spread for 1987 is also seen in Fig. 9(d), where the distance between the thin dotted lines indicates a range of model realizations. In contrast to the Indian subcontinent, it can be seen in Fig. 10 that precipitation anomalies over the western part of the Bay of Bengal and the adjacent coast of south-east Asia are quite predictable—all nine members of the 1987 ensemble indicate a large positive anomaly over that region. For JJA 1988, the results are more robust over India in the sense that positive rainfall anomalies are dominant in six out of nine experiments (not shown).

In the headings of Fig. 9, the rainfall rates associated with the 15-year average ensemble-mean fields and the 15-year standard deviation of ensemble-mean rainfall are given along with the equivalent values from Xie–Arkin and ERA-15 datasets. Note that the ensemble-mean field necessarily has a smaller standard deviation than that from the individual integrations. For the Nordeste region, the model is slightly drier than Xie–Arkin data, and more than 2 mm day$^{-1}$ drier than the re-analysis values. The ensemble and re-analysis interannual variability are about the same, and larger than that observed.

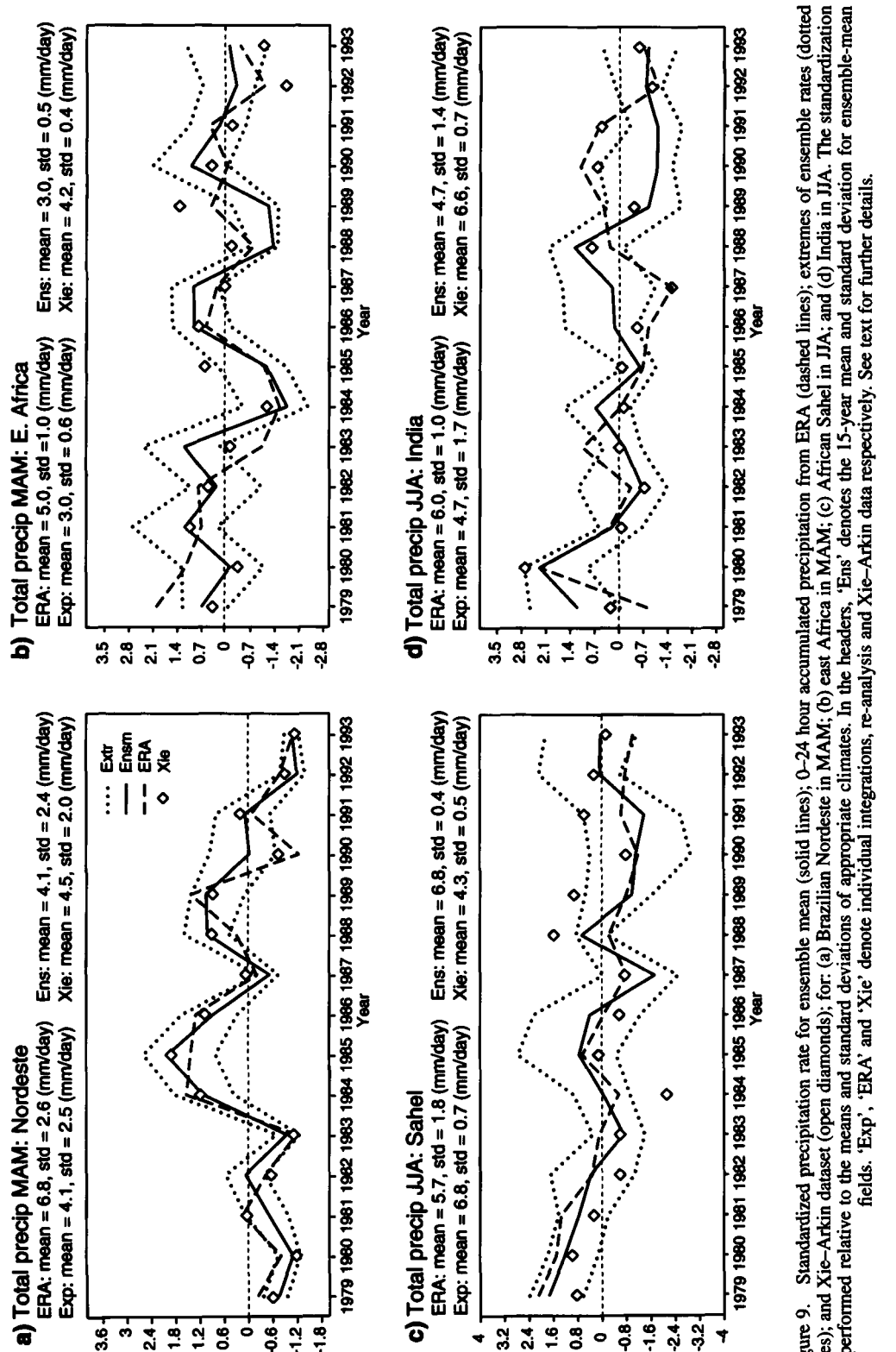


Figure 9. Standardized precipitation rate for ensemble mean (solid lines); 0–24 hour accumulated precipitation from ERA (dashed lines); extremes of ensemble rates (dotted lines); and Xie–Arkin dataset (open diamonds); for: (a) Brazilian Nordeste in MAM; (b) east Africa in MAM; (c) African Sahel in JJA; and (d) India in JJA. The standardization is performed relative to the means and standard deviations of appropriate climates. In the headers, 'Ens' denotes the 15-year mean and standard deviation for ensemble-mean fields. 'Exp', 'ERA' and 'Xie' denote individual integrations, re-analysis and Xie–Arkin data respectively. See text for further details.

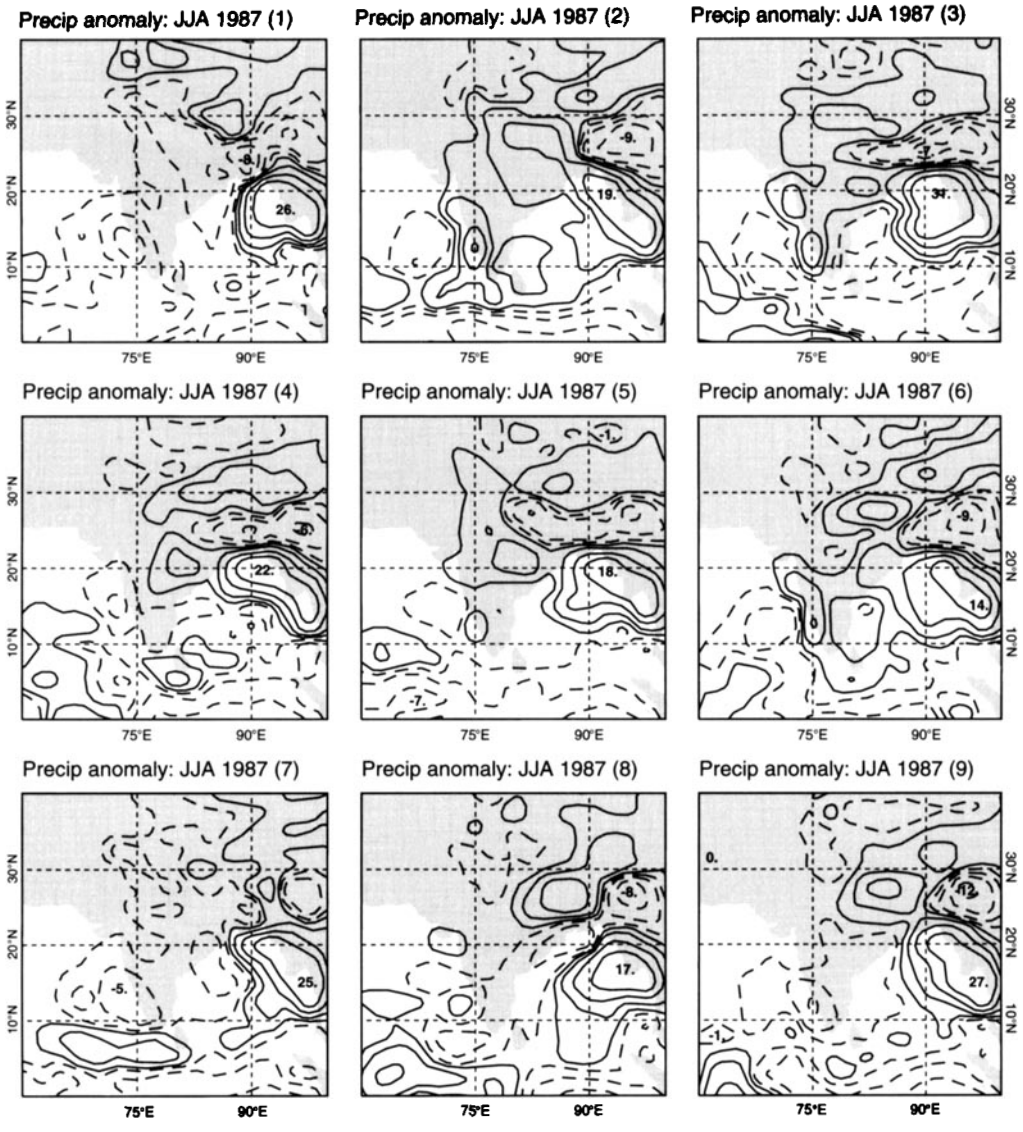


Figure 10. Precipitation anomalies for June, July and August (JJA) 1987 over the Indian region for all nine individual model realizations. Contours are at ±0.5, 2, 5, 10, 30 mm day$^{-1}$. Negative anomalies are dashed.

For east Africa the ensemble is again drier than both Xie–Arkin and ERA, and the ensemble standard deviation is about the same as that for Xie–Arkin but smaller than the re-analysis interannual variability. For the Sahel the ensemble is wetter than both Xie–Arkin and the re-analysis, but the re-analysis has larger interannual variability. In contrast to the Sahel, the model underestimates the precipitation rate over India, but has more variability. Overall, there is no clear over- or underestimation of the 15-year mean rainfall rates. The model tends to underestimate interannual variability of rainfall rates when compared with ERA, but is often about right when compared with Xie–Arkin variations.

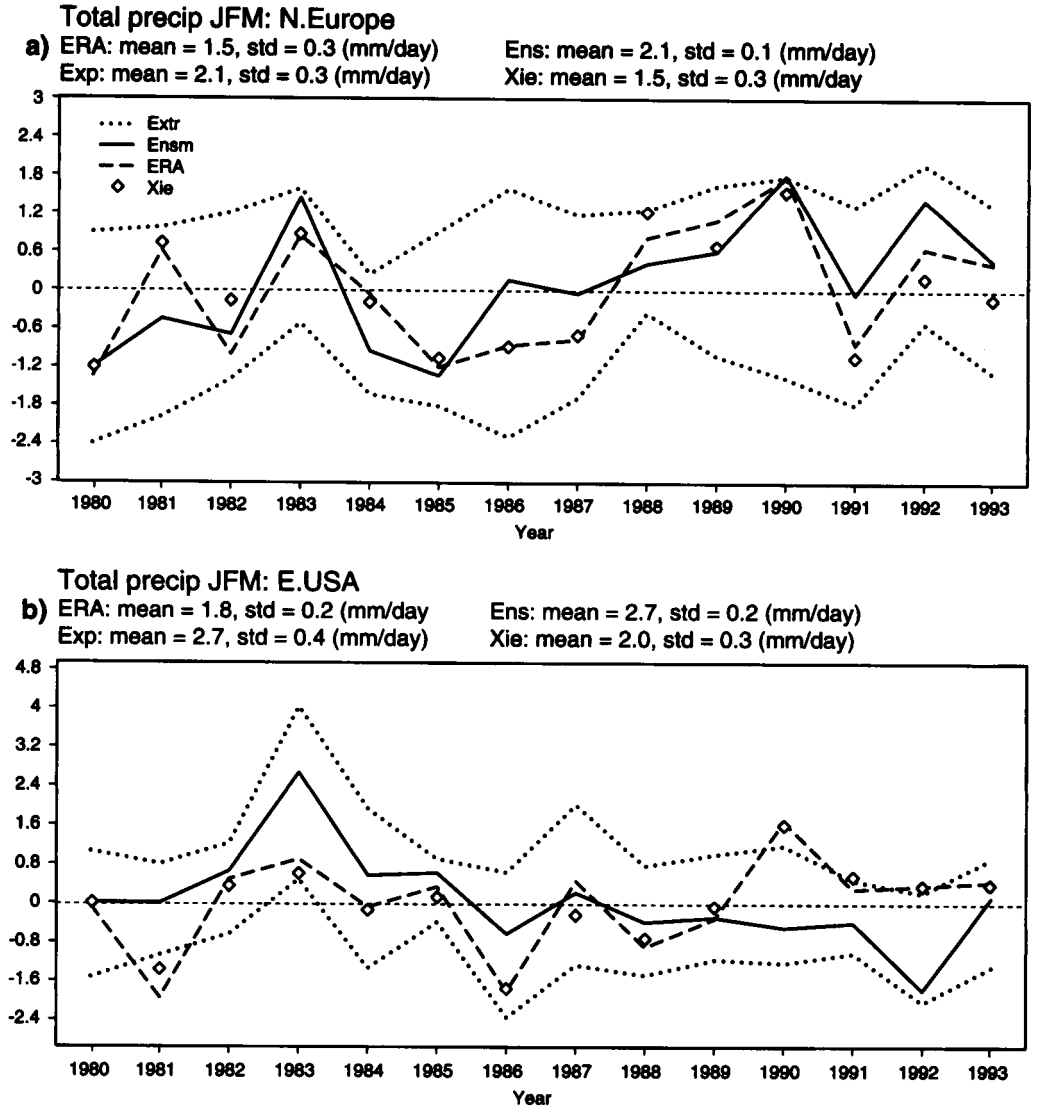


Figure 11. As Fig. 9 but for: (a) northern Europe in January, February and March (JFM); and (b) eastern USA in JFM.

The standardized rainfall estimates have also been studied for a number of extra-tropical regions. In general the level of skill is much lower when compared with the tropics. However, for some regions the results were encouraging. Figure 11 shows that the interannual variability in JFM rainfall for northern Europe (50–70°N, 10°W–30°E) and for the eastern USA (25–50°N, 100–70°W) were captured reasonably well. For both regions, the ensemble is wetter than Xie–Arkin precipitation and re-analysis, though the standard deviation is about the same.

## 6. Estimates of predictability

The skill scores shown in section 4 are influenced by model error. In order to assess the potential for seasonal prediction without the direct impact of model error,

we discuss variations of intrinsic seasonal predictability in this section. Since the ability to discriminate between years with high and low predictability is of great practical importance, estimates of seasonal predictability are made for each year. This approach contrasts with those where attempts to derive an overall measure of seasonal predictability were made (e.g. Chervin 1986; Rowell 1998).

In the following two subsections, two different measures of predictability are evaluated for seasonal-mean 850 mb temperature. The first is the $t$-test, measuring whether the ensemble mean for a given year was significantly different from the climatological state. The second is a ratio of variance, measuring the size of ensemble variance for a particular year relative to the overall ensemble variance using all ensemble members for all years. High predictability corresponds to large $t$-values and/or small ratio of variance. These essentially correspond to measuring the signal imposed from the boundary forcing in relation to the 'climatological' noise from unpredictable nonlinearities due to a model's internal dynamics (see, for example, Zwiers 1996).

### (*a*) *Significance of ensemble-mean anomalies*

A $t$-test has been performed, grid point by grid point, on the 850 mb temperature anomalies for a given year, where climatology is obtained from all the integrations excluding those from the given year. Figure 12 shows some examples of such a spatial distribution of $t$-values for the late winter season (JFM). The years displayed have been chosen so as to illustrate winters when the equatorial Pacific SST forcing was (a) strong, (b) moderately strong and (c) weak. For reference, the null hypothesis that the ensemble mean is not different from climate can be rejected with at least 80%, 90% or 99% confidence levels, if the $t$-variable exceeds 1.29, 1.66 or 2.62, respectively. The sign of the $t$-variable is irrelevant for this analysis.

For example, 1984/85 was a predictable winter for much of northern Europe, whilst 1982/83 was a predictable winter for much of southern Europe (see Figs. 12(b) and 12(a), respectively). Interestingly, whilst there was a strong El Niño event in 1982/83 (giving rise to extremely large $t$-values over most of the tropics), there was only a moderately strong La Niña event in 1984/85. It is possible that the predictability over northern Europe in the winter 1984/85 was enhanced, for example by north Atlantic SST anomalies. For comparison, Fig. 12(c) shows the global distribution of $t$-values for a winter when the ensemble distribution was not significantly different from the 15-year climatology over most of Europe (JFM 1991).

In Table 3 is shown the time-mean $t$-statistic for three selected regions. For the northern hemisphere, $t$-values remain almost unchanged throughout winter, whereas in other seasons there is a decrease in predictability between months 1 and 3 and months 2 and 4. For Europe, predictability estimates are generally lower than those for the northern hemisphere and are reduced in months 2–4 relative to months 1–3 irrespective of the season. The highest $t$-values are found for North America where, apart from in winter, a reduction in predictability estimates in months 2–4 is also seen. The enhanced predictability in winter is probably associated with the seasonality of teleconnections between the tropics and extratropics. The overall higher predictability in months 1–3 than in months 2–4 may be related to a possible stronger influence of initial conditions. When the ENSO years are considered (figures in brackets in Table 3), the predictability estimates are generally increased in all regions and in all seasons. The only exceptions are the late spring (AMJ) and summer (JJA) over Europe.

To emphasize the interannual and seasonal variations in predictability, the grid point $t$-values of the seasonal mean 850 mb temperature have been averaged over two European subdomains: northern Europe, 50–70°N, 10°W–30°E; and southern Europe,

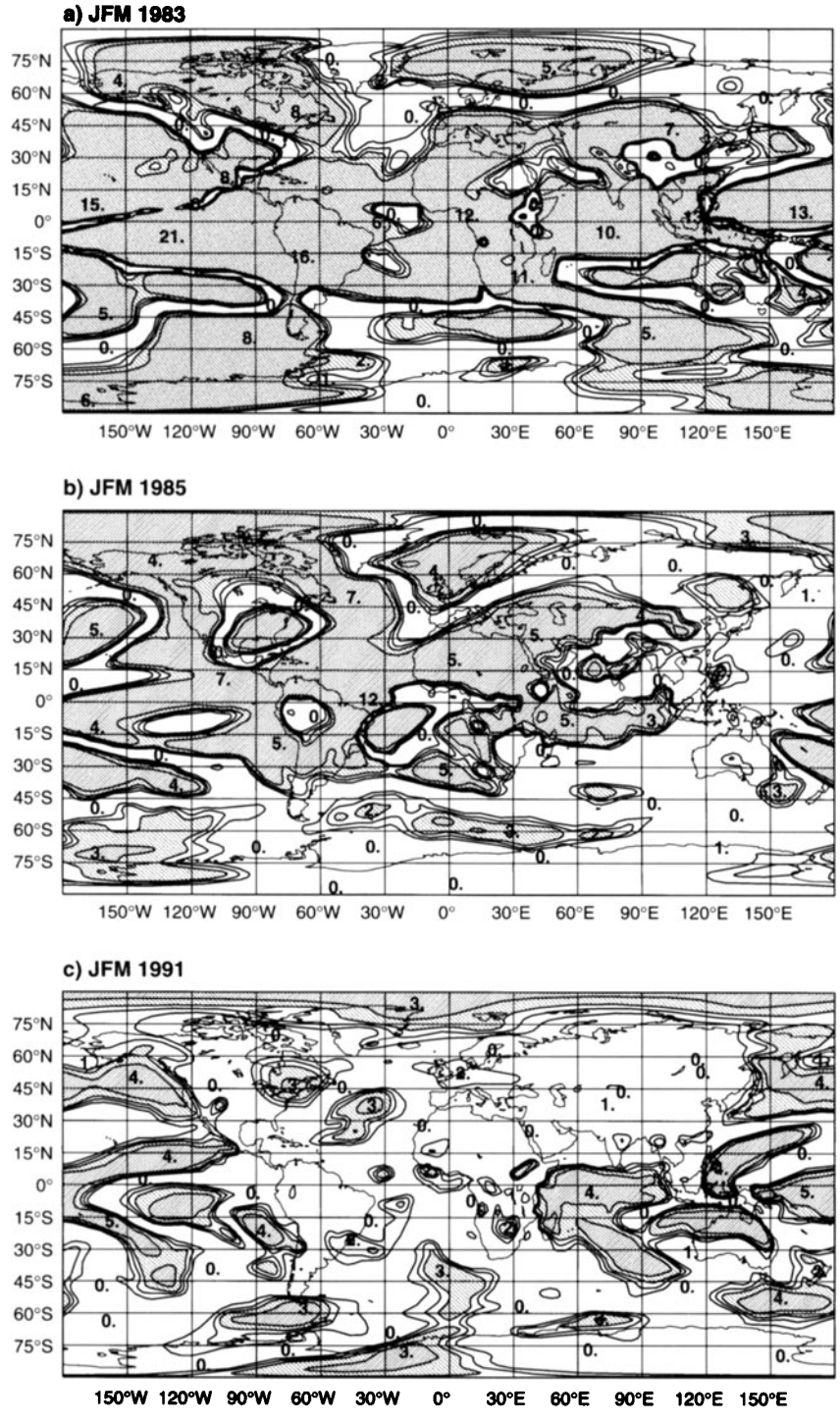


Figure 12. Student's $t$-statistic of 850 mb temperature for January–March (JFM) in: (a) 1983; (b) 1985; and (c) 1991. Contours are at 80%, 90%, 95% and 99% confidence levels, with shading where greater than 95% which corresponds to $t$-values of 1.980 or greater.

TABLE 3. MEAN $t$-STATISTICS FOR 850 mb TEMPERATURE

| Season | N hemisphere | Europe | N America |
|---|---|---|---|
| MAM | 1.42 (1.67) | 1.11 (1.24) | 1.59 (2.04) |
| AMJ | 1.21 (1.28) | 0.94 (0.94) | 1.26 (1.38) |
| JJA | 1.49 (1.69) | 1.14 (0.92) | 1.48 (1.74) |
| JAS | 1.32 (1.61) | 1.07 (1.10) | 1.28 (1.58) |
| SON | 1.21 (1.31) | 1.17 (1.46) | 1.25 (1.31) |
| OND | 1.05 (1.21) | 0.86 (0.99) | 1.10 (1.21) |
| DJF | 1.48 (1.83) | 1.36 (1.74) | 1.58 (2.22) |
| JFM | 1.47 (2.02) | 1.28 (1.77) | 1.64 (2.60) |

Figures in brackets are the means from strong and moderately strong ENSO years.

35–50°N, 10°W–30°E. The results are shown in Fig. 13 for JFM and in Fig. 14 for JAS. The heavy solid lines represent the interannual variation of the $t$-value with the $y$-axis reference on the left-hand side of each panel. For reference, the 80% and 99% confidence limits are shown as horizontal lines. There is certainly evidence of seasonal predictability for these two subdomains of Europe, at least in winter, but the level of predictability varies considerably from year to year. Moreover, a year which shows predictability over northern Europe does not necessarily show predictability for southern Europe. In winter there are more predictable years for northern Europe than for southern Europe (Fig. 13). For example, based on the 90% confidence threshold ($t$-value equals 1.66), northern Europe shows predictability in four years, and southern Europe in two years. The opposite is found in summer (Fig. 14), when there are three years above the 90% threshold for southern Europe, whereas no years are predictable for northern Europe.

The $t$-values for other fields, such as geopotential height and precipitation, have also been calculated. For brevity, in Fig. 15 we show $t$-values for 500 mb height for the eastern USA and for rainfall over east Africa and India. The eastern US region is in the extratropics, much closer than Europe to the centre of action in the equatorial Pacific; the other two regions are in the tropics.

For the eastern USA, the late-winter predictability estimates for 500 mb heights for the strong ENSO years (1983 and 1989) are high, and differ significantly from the other years (Fig. 15(a)). A similar result is seen for MAM tropical precipitation $t$-values over eastern Africa (Fig. 15(b)). In addition, predictability of east Africa rainfall is, on average, relatively high—most of the open square symbols are lying above the 90% significance level. For India, apart from the year 1980, the $t$-values are always below the 99% confidence limit, and in eight years (out of 15) are below or equal to the 90% threshold. In this sense, Indian rainfall is not generally as predictable as, for example, east African rainfall. This is consistent with the results in Fig. 10.

### (*b*) *Ratio of variance*

As mentioned above, the grid point value of the ratio of ensemble variance for a given year to the total variance from the set of all ensemble fields taken over all the years is calculated. Using this measure, for regions where there is high intrinsic predictability, the ensemble spread should be much less than the interannual variability of the ensemble mean, and the ratio of ensemble variance to the total variance will be small. Note that, in contrast to the $t$-test, high predictability is associated with small values of the diagnostic.

The ratio of variance of seasonal-mean 850 mb temperature is shown in Fig. 16 for the same three years as in Fig. 12. In JFM 1983 (Fig. 16(a)), it can be seen that

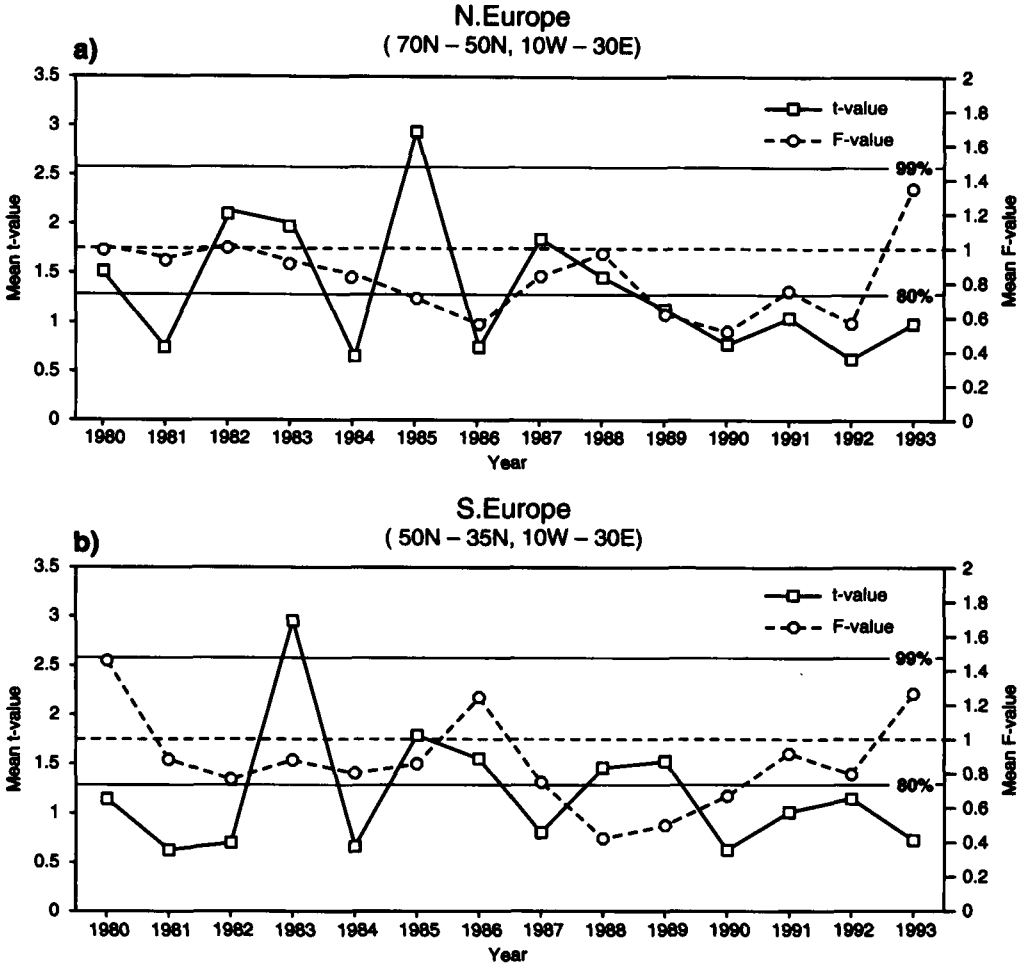


Figure 13. The interannual variation of the January–March (JFM) area mean 850 mb temperature $t$-statistic (solid lines) and ratio of variance ($F$-value, dashed lines) for: (a) northern Europe; and (b) southern Europe.

across much of the tropical Pacific, where the ensemble mean is significantly different from climatology (cf. Fig. 12(a)), the variance ratio is particularly small. Over much of the Mediterranean area, where the ensemble mean was also significantly different from climate, the variance ratio was also substantially less than unity. By contrast, the area over northern Scandinavia and to the east, which was significant according to the $t$-test, had a variance ratio exceeding unity. Hence for this region the ensemble for JFM 1983 had substantial spread of temperature estimates, though with many members consistently different from climatology. It is interesting to note that there are regions in Fig. 16(a) (e.g. over parts of the north Atlantic) where the variance ratio was small but where the $t$-value was not significant. For such regions, one would forecast a climatological mean temperature with confidence!

The area-averaged variance ratio of seasonal mean 850 mb temperature for the two European subdomains are shown by the heavy dashed lines in Figs. 13 and 14 for the winter and summer seasons, respectively. The reference y-axis is given on the right-hand side, and the reference level of unity is shown by the dashed horizontal line. Years with

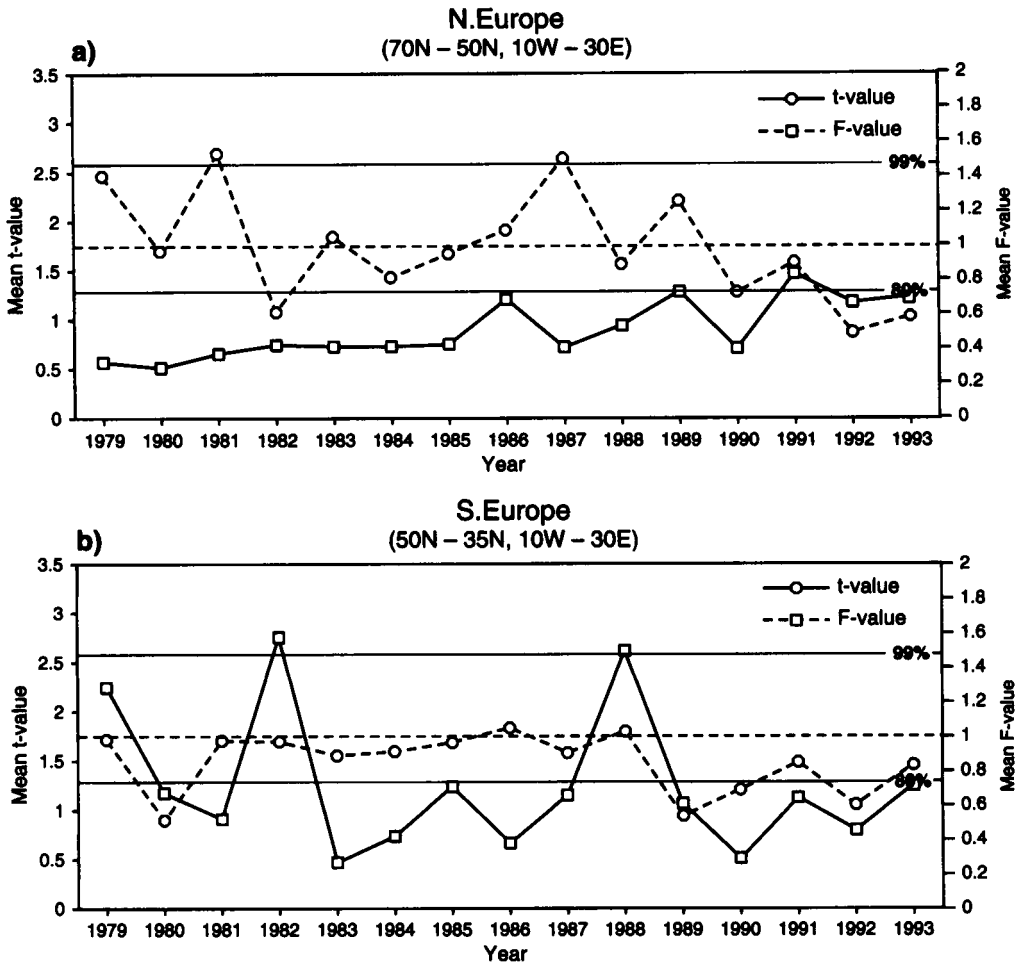


Figure 14. As Fig. 13 but for the July–September (JAS) season.

relatively large $t$-values and small values of the variance ratio, such as winter 1984/85 for northern Europe and winter 1982/83 for southern Europe, can be singled out. In general, the ratio of variance is smaller in winter than in summer, thus indicating a higher level of predictability in winter by this measure.

The variance ratio was also calculated for precipitation. The values for Europe were found to be fairly consistent in spring, falling just below unity (not shown). In winter, however, there is much more variability, with some individual ensembles showing more consistency than any of the spring ensembles. In summer, rainfall does not appear to be very predictable over Europe, when compared with, for example, the United States. The variance ratio for some tropical regions (e.g. Brazilian Nordeste, east Africa (shown in Fig. 15(b))) are found to be much lower than those over Europe.

### (c) *Comparison with conventional skill scores*

We have shown in section 4(b) that in the sample of PROVOST integrations there is a correlation between ensemble-mean skill and the intra-ensemble dispersion of skill (see Fig. 8). On the other hand, this correlation does not prove the existence of

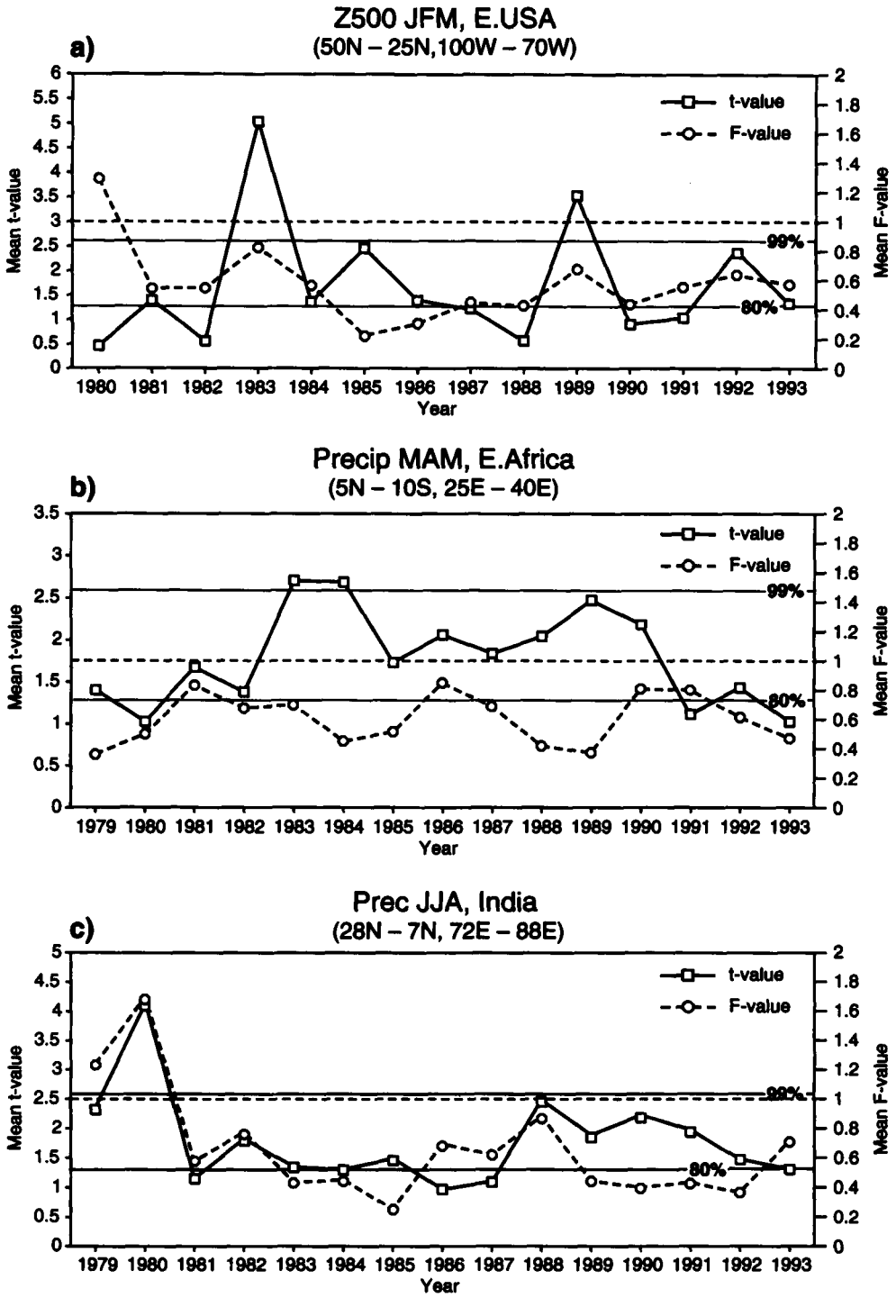


Figure 15. As Fig. 13 but for: (a) eastern USA 500 mb height in JFM; (b) eastern Africa precipitation in March–May (MAM); and (c) India precipitation in June–August (JJA).

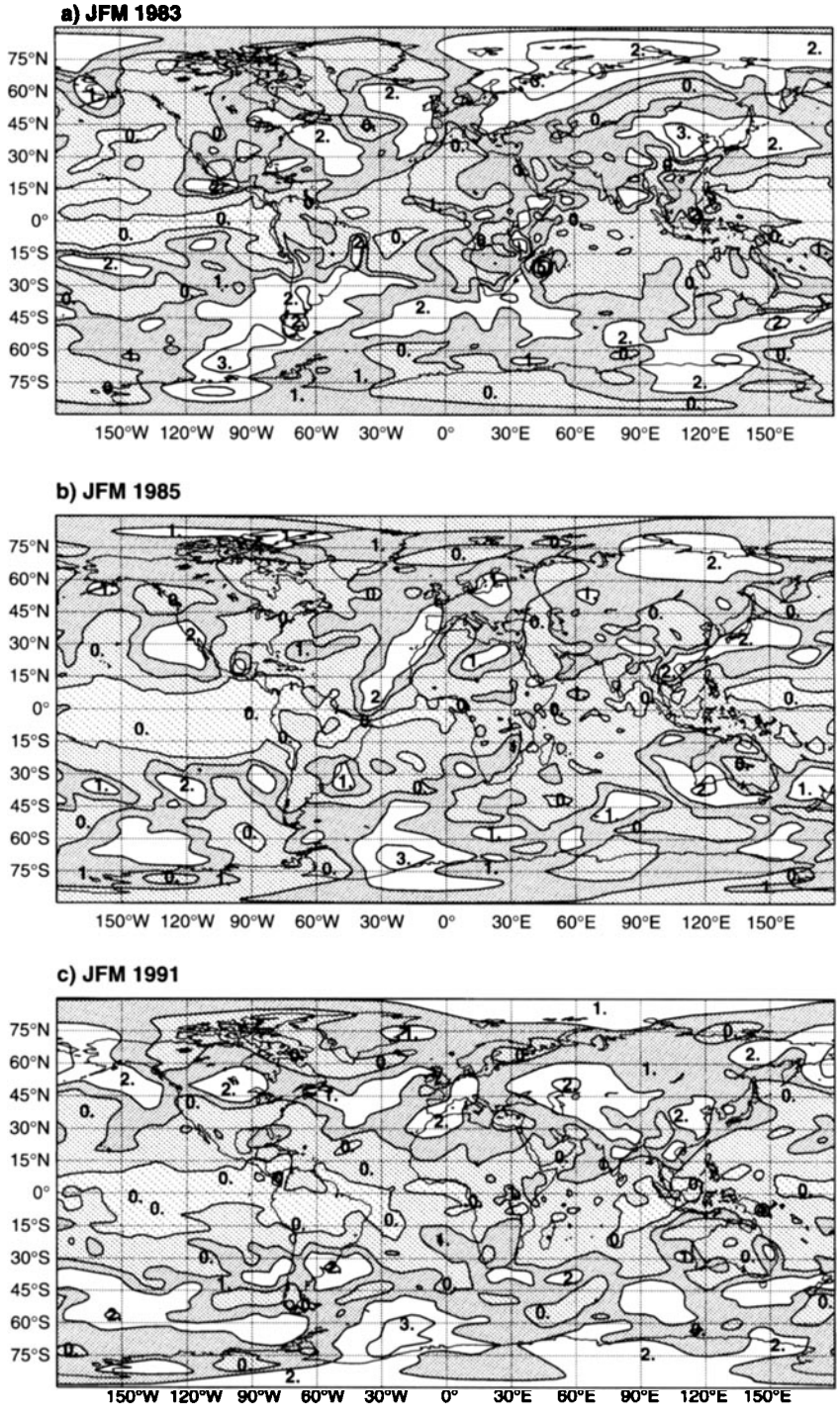


Figure 16. The ratio of variance of the 850 mb temperature for the January–March (JFM) season in: (a) 1983, (b) 1985, (c) 1991. Contours are at 0.1, 0.5, 1.0 and 2.0. Shading for values below 1.0 only. See text for further details.

a relationship between ensemble spread and ensemble-mean skill. To investigate this further, we compare ensemble-mean skill with the ensemble $t$-statistic and the ratio of variance introduced and discussed in the previous two subsections. For brevity, we discuss some results without showing figures.

For the northern hemisphere 500 mb height in DJF of 1982/83, 1983/84, 1987/88, 1988/89 and 1990/91, there is a reasonable correspondence between relatively high $t$-values (above 90% confidence level) and relatively low ratio of variance (below 0.8). These winters have ensemble-mean anomaly correlations greater than or equal to 0.5, i.e. they are among the most skilful winter forecasts (see Fig. 5(a)), and the consistency of skill scores within these ensembles is relatively high (Fig. 8(a)). The lowest northern hemisphere $t$-values in 500 mb height are found for SON, where for almost all years it stays below 80% confidence level. This is consistent with the poor skill scores for SON discussed earlier.

Based on these considerations, one may infer that these sample statistics are reasonable predictors of the ensemble mean. However, this is not always true. For example, the ensemble-mean forecast for the 1985/86 winter has a skill score of around 0.5, but based on sample statistics it cannot be distinguished as being skilful. However, it can be said that a high $t$-value or low ratio of variance is a necessary (but not sufficient) condition for skill. No cases of substantial skill with a low $t$-statistic and a large variance ratio were found.

Despite these somewhat mixed results, we can use the $t$-statistic in an attempt to shed more light on the rather high hemispheric-forecast skill seen during the strong La Niña winter of 1988/89 (see Fig. 5). This result may seem somewhat surprising, especially in view of the evidence offered in scientific literature. For example, Kumar and Hoerling (1998) found that potential predictability over the PNA region for the strongest warm ENSO events consistently exceeds that for the strongest cold events. In addition, Palmer (1988) found that when in a negative PNA phase, possibly associated with a La Niña SST anomaly, the growth of barotropic perturbations in the northern extratropical atmosphere is much stronger (giving rise to lower predictability) than when in a positive PNA phase.

The 500 mb height $t$-statistic for the 1988/89 winter has lower values over the PNA region when compared with, for example, that for a strong El Niño winter 1982/83 (not shown). However, the spatial extent of relatively high $t$-values is much larger than that for 1982/83. For example, the regions with a very high confidence level (above 99%) extend from North America over much of the northern and central Atlantic into Europe in 1988/89. This is not the case in the winter of 1982/83 where this high $t$-value is mainly confined to the PNA region. Moreover, in 1988/89 high $t$-values extend westward from the northern central Pacific into eastern Asia. This spatial extension of large $t$-values in the 1988/89 winter coincides with the regions of significant height anomalies (cf. Fig. 7). Clearly, Fig. 7 indicates that the northern hemisphere skill is highly influenced by anomalies over the whole of the western hemisphere, including the northern Pacific region.

## 7. Relative impact of initial conditions and SST forcing on rainfall predictability

Additional ensembles have been run for the 1987 and 1988 spring and summer seasons, to address the issue of the extent to which seasonal predictability is due to the underlying SST anomalies or the ICs (which include land surface ICs). These additional ensembles consist of six members each; the control ensembles are the subsets of the

TABLE 4. ORGANIZATION OF HYBRID EXPERIMENTS

| | Year A: 1987 | Year B: 1988 |
|---|---|---|
| Control | $IC_{87}$, $SST_{87}$ | $IC_{88}$, $SST_{88}$ |
| Impact of SST | $IC_{87}$, $SST_{88}$ | $IC_{88}$, $SST_{87}$ |
| Impact of IC | $IC_{88}$, $SST_{87}$ | $IC_{87}$, $SST_{88}$ |

original 9-member PROVOST ensembles. The additional ensembles are referred to as 'hybrid' experiments, because of the blending of ICs for one year with SST data from the different year.

Let $\{IC_{87}, SST_{87}\}$ denote an integration made with 1987 ICs and 1987 SSTs. The hybrid integrations can be denoted as $\{IC_{87}, SST_{88}\}$, and so on; see Table 4 for details. In order to establish the statistical significance of these differences we also show related $t$-statistics.

### (*a*) *Regional impact*

Results are shown in Figs. 17 and 18 for precipitation in two regions: northern subtropical Africa (Sahel) during JJA, and the USA during AMJ. For the Sahel, 1987 was a drought year; for the eastern USA, 1988 was a drought year. The droughts in these two regions were discussed extensively elsewhere, and will not be discussed here; see, for example, Palmer *et al.* (1992) for the discussion on the Sahel drought, and Trenberth and Branstator (1992) and Palmer and Branković (1989) for the 1988 US drought.

For much of the Sahel, the model control ensemble simulated well the observed increase in precipitation in JJA 1988 relative to JJA 1987 (Fig. 17(a)). The associated $t$-values, significant at the 95% confidence level, are found to the west of the Greenwich meridian and to the east of the Lake Chad (15°E, Fig. 17(b)). Figure 17(c) shows that this increase in JJA 1988 precipitation comes primarily from the impact of the 1988 SST data, $\{IC_{87}, SST_{88}\} - \{IC_{87}, SST_{87}\}$, since the ICs in the hybrid ensemble $\{IC_{87}, SST_{88}\}$ are identical to those in the control ensemble $\{IC_{87}, SST_{87}\}$, and the SSTs were taken from 1988. For the $\{IC_{87}, SST_{88}\} - \{IC_{87}, SST_{87}\}$ difference, the confidence level is more than 99% for much of subtropical northern Africa (Fig. 17(d)).

The impact of ICs is smaller than that of SSTs ($\{IC_{88}, SST_{87}\} - \{IC_{87}, SST_{87}\}$, Fig. 17(e)), though by no means unimportant. The increase in precipitation is restricted to the sub-Saharan region, whereas to the south of 12°N there is a reduction in precipitation. The signal from the ICs is not as highly significant as that from the SSTs, though it is large in certain sub-regions. Similar results to those in Fig. 17 apply for $\{IC_{88}SST_{87}\} - \{IC_{88}SST_{88}\}$ and $\{IC_{87}, SST_{88}\} - \{IC_{88}, SST_{88}\}$ differences, however with a smaller statistical significance (not shown).

Figure 18(a) and (b) shows that the late spring (AMJ) control ensemble precipitation difference over the eastern USA agrees well with the observed interannual variation (wetter in 1987, drier in 1988). In terms of precipitation differences and $t$-statistics, it can be concluded that both SST forcing ($\{IC_{87}, SST_{88}\} - \{IC_{87}, SST_{88}\}$, Fig. 18(c) and (d)) and initial conditions ($\{IC_{88}, SST_{87}\} - \{IC_{87}, SST_{87}\}$, Fig. 18(e) and (f)), seem to play important roles in explaining interannual variability. Bearing in mind that the initial conditions were defined in late February 1988, their impact on the AMJ precipitation is striking.

When the differences with respect to the 1988 control ensemble are considered ($\{IC_{88}, SST_{87}\} - \{IC_{88}, SST_{88}\}$ and $\{IC_{87}, SST_{88}\} - \{IC_{88}, SST_{88}\}$), the model

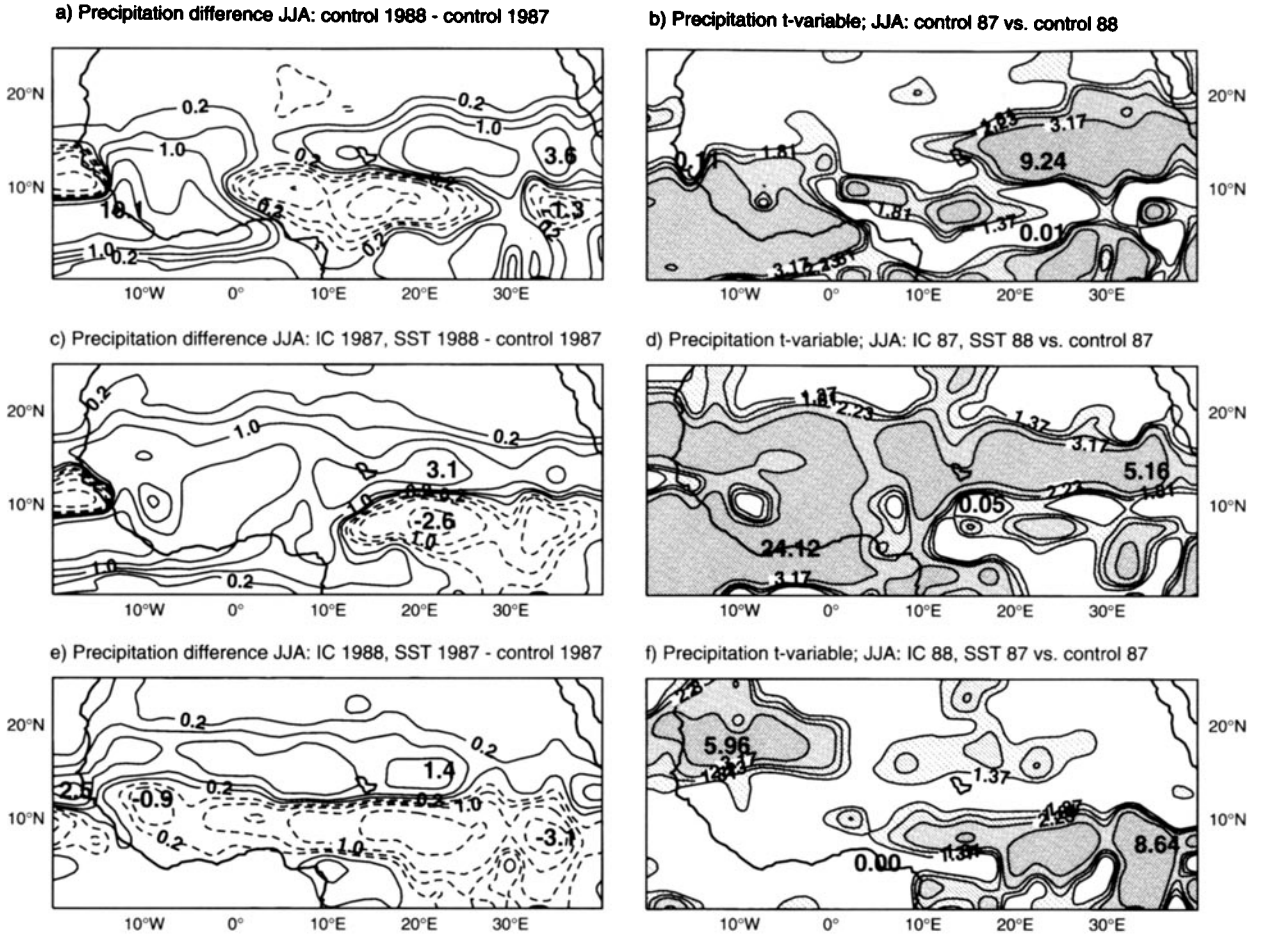


Figure 17. Seasonal mean June–August (JJA) rainfall over subtropical Africa (Sahel): (a) difference between the 6-member control ensembles for 1988 and 1987; (b) $t$-statistics for (a); (c) difference showing the impact of sea surface temperature forcing; (d) $t$-statistics for (c); (e) difference showing the impact of initial conditions; (f) $t$-statistics for (e). Contours in (a), (c) and (e) are $\pm 0.2$, 0.5, 1, 2, 5, 10 mm day$^{-1}$. Contours in (b), (d) and (f) are at 80%, 90%, 95% and 99% confidence levels (these correspond to $t$-values of 1.37, 1.81, 2.23 and 3.17 respectively).

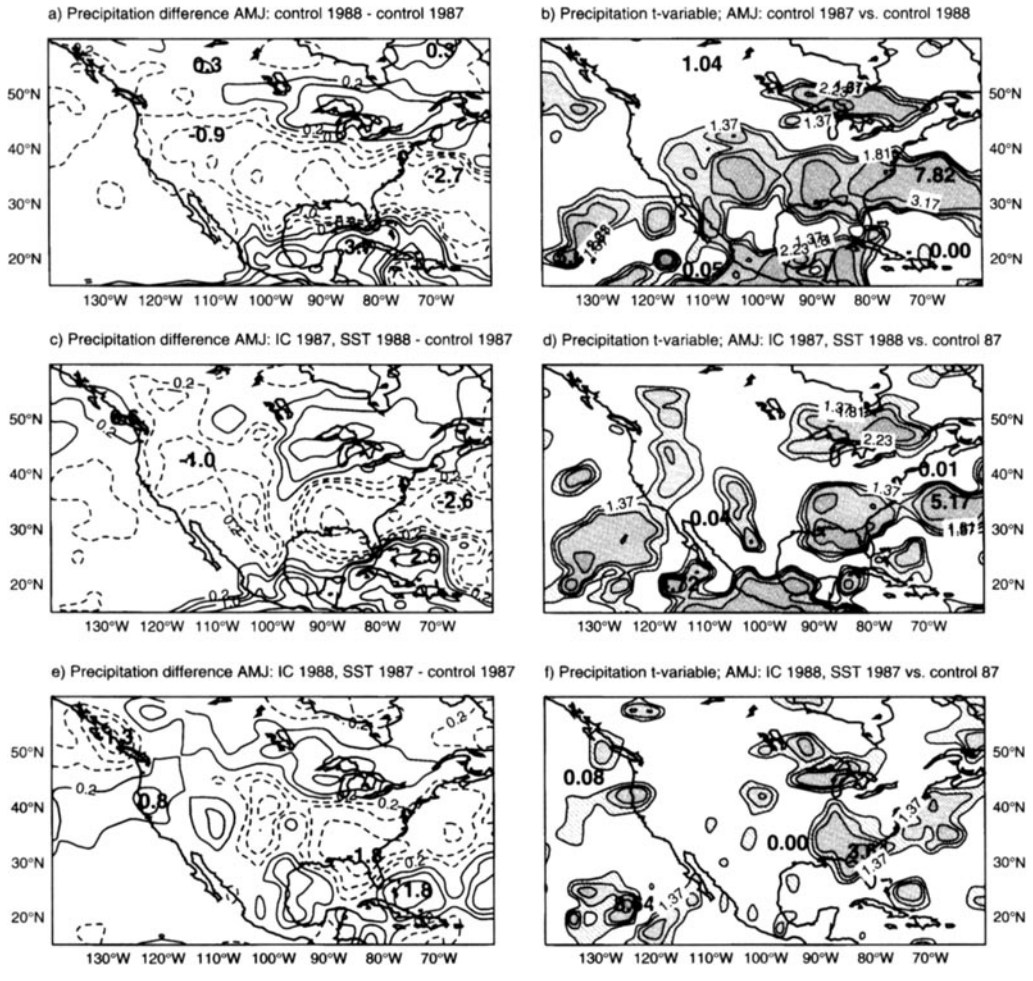


Figure 18. As Fig. 17 but for April–June (AMJ) rainfall over the USA.

response is essentially qualitatively similar to that shown in Fig. 18, but the statistical significance over the eastern part of the USA is weaker.

For brevity, the results for Europe in late spring (AMJ) are summarized as follows. For northern Europe, the precipitation difference (with respect to the control ensemble) due to SSTs is very similar to the precipitation difference due to ICs. However, the *t*-statistics indicate that the impact of the SSTs is more significant than the impact of the ICs. For eastern Europe, the precipitation differences due to SSTs and ICs are again very similar, and both factors seem to play an equally significant role in the interannual variation. For south-western Europe, the influence of SSTs prevails over that of ICs. The diversity of the above results illustrates the difficulty in attributing rainfall predictability over Europe.

### (*b*) *Global impact*

The relative impacts of ICs and SST forcing on global precipitation rates were also studied; here we discuss the results for global precipitation difference fields.

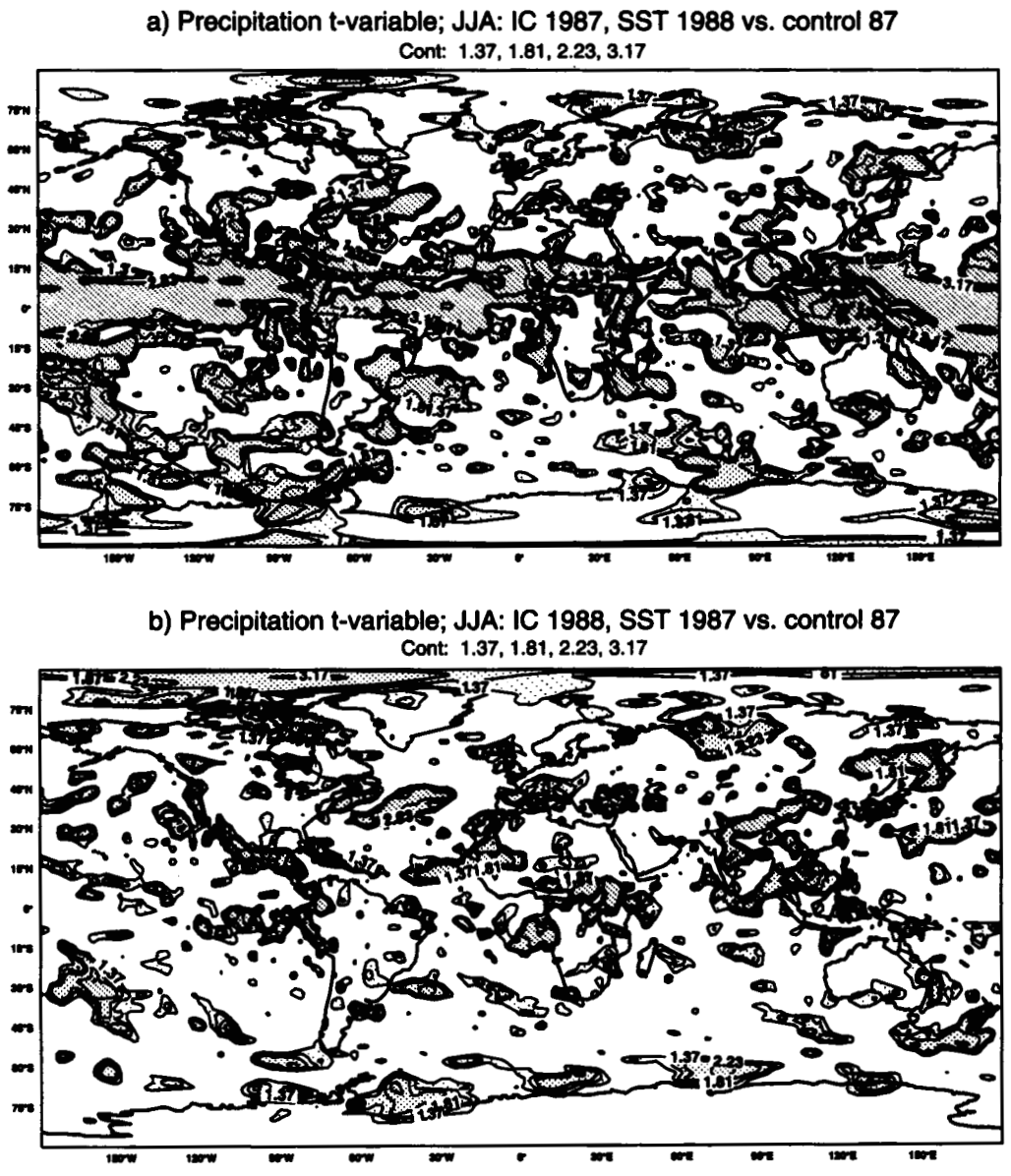


Figure 19. $t$-statistics for seasonal-mean June–August (JJA) rainfall differences between hybrid and control ensembles for: (a) 'wrong' SSTs in hybrid experiments; and (b) 'wrong' initial conditions (ICs) in hybrid experiments. See text for details and discussion.

Irrespective of the season considered, the following pattern in the difference fields emerges. When the 'wrong' SSTs in the hybrid experiments are used and differences are calculated with respect to corresponding control ensembles (cases ($\{IC_{87}, SST_{88}\} - \{IC_{87}, SST_{87}\}$ and $\{IC_{88}, SST_{87}\} - \{IC_{88}, SST_{88}\}$ in Table 4), the precipitation differences over oceanic areas and most land areas are very similar (in both pattern and amplitude) to the differences between the control ensembles for the two years considered. In Fig. 19(a) the JJA $t$-values for the case $\{IC_{87}, SST_{88}\} - \{IC_{87}, SST_{87}\}$ are illustrated.

This shows that the oceanic boundary forcing is the major contributor to interannual variations of global precipitation.

When the 'wrong' ICs are used, differences between hybrid and control ensembles (cases $\{IC_{88}, SST_{87}\} - \{IC_{87}, SST_{87}\}$ and $\{IC_{87}, SST_{88}\} - \{IC_{88}, SST_{88}\}$ in Table 4) show much patchier patterns over both oceans and continents. Over some regions these differences are similar to those between the controls of different years, however, no truly global pattern is evident (Fig. 19(b) depicts the $\{IC_{88}, SST_{87}\} - \{IC_{87}, SST_{87}\}$ case). Based on these results, it could be argued that the impact of ICs is confined to relatively small spatial scales, and is predominantly local in character as discussed in section 7(a).

In addition, the 'sum' of the individual contributions from SSTs and ICs does not correspond to (the differences between) the control ensembles. This would imply that nonlinear processes are important in determining the combined effect of SSTs and ICs.

## 8. Conclusions

Objective measures of seasonal skill and predictability have been estimated from ensembles of 120-day integrations of the ECMWF model, run with prescribed observed SSTs. The use of observed rather than predicted SSTs indicates that the results shown here may be assumed to provide an upper limit on what can be obtained with a coupled ocean–atmosphere model (assuming the adequate accuracy of observed SSTs). However, two *caveats* should be made. In section 3, we noted that the systematic error in the model was, in many regions, at least as large as the signal (the observed interannual variability) one is trying to predict. In the northern winter, for example, the ratio of systematic error to interannual variability was large over the extratropical Atlantic and Pacific Oceans. Obviously these systematic errors can have a significant impact on the level of model skill, e.g. as determined by anomaly correlation coefficients. However, there is a second, more subtle way in which model systematic error can influence even predictability estimates. It is known that SSTs in mid latitudes are both forced by, and in turn force, the overlying atmospheric flow. The storm tracks appear to play a key role in this two-way process (e.g. Peng and Whitaker 1999). If the model's storm tracks are systematically mishandled, then the model will not respond correctly to the observed SSTs. In this sense, the model's response to the observed SSTs, and hence extratropical predictability, may increase when model error has been reduced.

The results using conventional anomaly correlation scores showed a fairly uniformly positive level of skill over the whole northern hemisphere. Over Europe the level of skill was positive overall, but not consistently so. This variable and inconsistent level of skill was also reflected in estimates of predictability based on $t$-tests and ratio-of-variance tests. Other areas were studied: for example, for tropical regions the level of predictability is much higher, as is well known. However, some parts of the tropics suffer from chaotic atmospheric processes, Indian monsoon rainfall being a prime example. The impact of ENSO on estimates of skill and predictability was also studied. Whilst most regions show a clear impact, it is less apparent overall for the European region.

The impact of SST forcing was compared with the influence of ICs (which include land surface ICs) by performing additional ensembles of hybrid integrations with SSTs from one particular year, and ICs from a different year. Overall the results confirm the predominant importance of SST forcing. However, on a regional level the results show that IC effects cannot be ignored for quantitative seasonal prediction. On this basis, seasonal-predictability estimates from AMIP integrations (Gates 1992), in which ICs are largely irrelevant, can be expected to be quantitatively different from the results of

these PROVOST integrations. The impact of SSTs and ICs are not linearly additive, another manifestation of the nonlinearity of climate.

These results, together with the positive impact of using multi-model ensemble techniques (Palmer *et al.* 2000), suggest that the next phase of seasonal-ensemble evaluation should be performed in coupled ocean–atmosphere mode. The next European Union project DEMETER (Development of European Multi-model Ensemble systems for seasonal to inTERannual prediction) is concerned with the development of a database of multi-model coupled-model ensemble integrations over the period of the ECMWF 40-year re-analysis. Results from DEMETER will be reported in due course.

## ACKNOWLEDGEMENT

This research was supported under the European Union Environment and Climate Programme contract CT95-0109. We thank Rob Hine and Jocelyn Williams for help with figures.

## REFERENCES


Arpe, K., Dümenil, L. and Giorgetta, M. A. 1998 Variability of the Indian monsoon in the ECHAM3 model: Sensitivity to sea surface temperature, soil moisture, and the stratospheric quasi-biennial oscillation. *J. Climate*, **11**, 1837–1858

Barnett, T. P. 1995 Monte Carlo climate forecasting. *J. Climate*, **8**, 1005–1022

Becker, B. D. 1998 ECMWF seasonal ensemble simulations on CD-ROM. *WMO Bull.*, **47**, 276–280

Bengtsson, L., Arpe, K., Roeckner, E. and Schulzweida, U. 1996 Climate predictability experiments with a general circulation model. *Clim. Dyn.*, **12**, 261–278

Blackmon, M. L., Geisler, J. E. and Pitcher, E. J. 1983 A general circulation model study of January climate anomaly patterns associated with interannual variation of equatorial Pacific sea surface temperature. *J. Atmos. Sci.*, **40**, 1410–1425

Boer, G. J., Arpe, K., Blackburn, M., Déqué, M., Gates, W. L., Hart, T. L., le Treut, H., Röckner, E., Sheinin, D. A., Simmonds, I., Smith, R. N. B., Tokioka, T., Wetherland, R. T. and Williamson, D. 1991 'An intercomparison of the climates simulated by 14 atmospheric general circulation models'. WMO/ICSU WCRP-58, WMO/TD No. 425. World Meteorological Organization, Geneva, Switzerland

Branković, Č. and Molteni, F. 1997 Sensitivity of the ECMWF model northern winter climate to model formulation. *Clim. Dyn.*, **13**, 75–101

Branković, Č. and Palmer, T. N. 1997 Atmospheric seasonal predictability and estimates of ensemble size. *Mon. Weather Rev.*, **125**, 859–874

Branković, Č., Palmer, T. N. and Ferranti, L. 1994 Predictability of seasonal atmospheric variations. *J. Climate*, **7**, 217–237

Chen, M. and Bates, J. R. 1996 A comparison of climate simulations from a semi-Lagrangian and Eulerian GCM. *J. Climate*, **9**, 1126–1149

Chervin, R. M. 1986 Interannual variability and seasonal climate variability. *J. Atmos. Sci.*, **43**, 233–251

Davies, J. R., Rowell, D. P. and Folland, C. K. 1997 North Atlantic and European seasonal predictability using an ensemble of multidecadal atmospheric GCM simulations. *Int. J. Climatol.*, **17**, 1263–1284

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

Dix, M. R. and Hunt, B. G. 1995 Chaotic influences and the problem of deterministic seasonal predictions. *Int. J. Climatol.*, **15**, 729–752

Fennessy, M. J., Kinter III, J. L., Kirtman, B., Marx, L., Nigam, S., Schneider, E., Shukla, J., Strauss, D., Vernekar, A., Xue, Y. and Zhou, J. 1994 The simulated Indian monsoon: a GCM sensitivity study. *J. Climate*, **7**, 33–43

Gates, W. L. 1992 AMIP: The atmospheric model intercomparison project. *Bull. Am. Meteorol. Soc.*, **73**, 1962–1970

Gibson, J. K., Kållberg, P., Uppala, S., Hernandez, A., Nomura, A. and Serano, E. 1997 'ERA description'. ECMWF Re-Analysis project report series. Available from the European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK

Harzallah, A. and Sadourny, R. 1995 Internal versus SST-forced atmospheric variability as simulated by an atmospheric general circulation model. *J. Climate*, **8**, 474–495

Kitoh, A. 1991 Interannual variations in an atmospheric GCM forced by the 1970–1980 SST. Part II: Low frequency variability of the wintertime northern hemisphere extratropics. *J. Meteorol. Soc. Japan*, **69**, 271–291

Krishnamurti, T. N., Bedi, H. S., Rohaly, G., Fulakeza, M., Oosterhof, D. and Ingles, K. 1995 Seasonal monsoon forecast for the years 1987 and 1988, *Global and Planetary Change*, **10**, 79–95

Kumar, A. and Hoerling, M. P. 1998 Annual cycle of Pacific/North American seasonal predictability associated with different phases of ENSO. *J. Climate*, **11**, 3295–3308

Lau, N.-C. 1985 Modeling the seasonal dependence of the atmospheric response to observed El Niños in 1962–76. *Mon. Weather Rev.*, **113**, 1970–1996

Miller, M., Hortal, M. and Jakob, C. 1995 'A major operational forecast model change'. Pp. 2–5 in ECMWF Newsletter, No. 70. Available from the European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK

Owen, J. A. and Palmer, T. N. 1987 The impact of El Niño on an ensemble of extended-range forecasts. *Mon. Weather Rev.*, **115**, 2103–2117

Palmer, T. N. 1988 Medium and extended range predictability and stability of the Pacific/North American mode. *Q. J. R. Meteorol. Soc.*, **114**, 691–713

1994 Chaos and predictability in forecasting the monsoons. *Proc. Indian Nat. Sci. Acad.*, **60**, 57–66

Palmer, T. N. and Branković, Č. 1989 The 1988 US drought linked to anomalous sea surface temperature. *Nature*, **338**, 54–57

Palmer, T. N., Branković, Č., Viterbo, P. and Miller, M. J. 1992 Modeling interannual variations of summer monsoons. *J. Climate*, **5**, 399–417

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

Peng, S. and Whitaker, J. S. 1999 Mechanisms determining the atmospheric response to midlatitude SST anomalies. *J. Climate*, **12**, 1393–1408

Rowell, D. P. 1998 Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. *J. Climate*, **11**, 109–120

Shukla, J. and Wallace, J. M. 1983 Numerical simulation of the atmospheric response to equatorial Pacific sea surface temperature anomalies. *J. Atmos. Sci.*, **40**, 1613–1630

Stern, W. and Miyakoda, K. 1995 Feasibility of seasonal forecasts inferred from multiple GCM simulations. *J. Climate*, **8**, 1071–1085

Stockdale, T. N., Anderson, D. L. T., Alves, J. O. S. and Balmaseda, M. A. 1998 Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. *Nature*, **392**, 370–373

Trenberth, K. E. and Branstator, G. W. 1992 Issues in establishing causes of the 1988 drought over North America. *J. Climate*, **5**, 159–172

Trenberth, K. E., Branstator, G. W., Karoly, D., Kumar, A., Lau, N.-C. and Ropelewski, C. 1998 Progress during TOGA in understanding and modelling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.*, **103 (C7)**, 14291–14324

Viterbo, P. and Betts, A. K. 1999 Impact on ECMWF forecasts of changes to the albedo of the boreal forests in the presence of snow. *J. Geophys. Res.*, **104D**, 27803–27810

Xie, P. and Arkin, P. A. 1997 Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates and numerical model outputs. *Bull. Am. Meteorol. Soc.*, **78**, 2539–2558

Zwiers, F. W. 1996 Interannual variability and predictability in an ensemble of AMIP climate simulations conducted with the CCC GCM2. *Clim. Dyn.*, **12**, 825–847