TECHNICAL MEMORANDUM

263

Seasonal skill and predictability of ECMWF PROVOST ensembles

Č. Branković and T. Palmer

Research Department

November 1998

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Seasonal skill and predictability of ECMWF PROVOST ensembles

Cedo Brankovic and Tim Palmer

Abstract

Variations in seasonal skill and predictability during the 15 years (1979-1993) of the ECMWF reanalysis (ERA) have been studied using 120-day ensemble integrations of the ECMWF model. 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 standardised 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 from 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 itself there are regions where the precipitation exhibits chaotic behaviour and is correspondingly less predictable.

1. Introduction

This is one of a pair of companion papers discussing seasonal predictability estimates from a set of ensemble integrations made with the ECMWF model. These integrations form part of the European Union PROVOST (PRediction Of climate Variations On Seasonal-to-interannual Timescales) project. The main objectives of the PROVOST project are: 1) to quantify the predictability of the atmosphere on seasonal timescales using ensembles of integrations of atmospheric models with specified observed sea surface temperature (SST), 2) to develop comprehensive coupled models of the oceans, atmosphere and land surface together with ocean data assimilation techniques to provide initial conditions for such coupled models, and 3) to test the skill of these coupled systems through coordinated 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 United Kingdom Meteorological Office. The ensembles were made with prescribed observed SSTs, implying that the results give upper bounds on predictive skill (though see the *caveats* in the concluding section of this paper).

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, Brankovic *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 times scales from observed (reanalysis) 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 all four models participating in the PROVOST into account, the maximum ensemble size is 36, a relatively large number by past standards. However, for the purposes of this paper (and in contrast with the companion paper), analysis has been restricted to the sub-ensemble from just one model, the ECMWF model. A pilot study for this PROVOST experimentation was reported in Brankovic and Palmer (1997).

Initial dates of the integrations spanned the 15-year period of the ECMWF reanalysis (1979-1993). 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 summarised 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 a companion paper (Palmer *et al.* 1998).

2. Organization of experiments

2.1 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 reanalysis project (ERA-15; Gibson *et al.* 1997).

The decision to use semi-Lagrangian time-stepping was based on results from seasonal timescale 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 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 that no corrective fixes were applied.

2.2 Seasonal ensembles

Nine-member ensembles were run over all seasons for the ERA-15 period, 1979-1993. The experiment set consists of ensembles for 14 northern winter and 15 northern spring, summer and autumn seasons. The experiments were initiated

from consecutive 12Z ERA-15 analyses, from 1 to 9 days preceding the season of interest (see Table 1). 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 - DJFM, spring - MAMJ, summer - JJAS and autumn - SOND. The model was run with observed prescribed SSTs taken from ERA-15 and updated daily in the integrations.

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

The model output was archived every 24 hours at 12Z, 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. Ten-day mean data 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.

2.3 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). Fig.1 shows monthly mean SST anomalies from 1979 to 1993 averaged over the region 7°N-7°S, 160°E-80°W. The 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 be readily found elsewhere (see, for example, *Climate Diagnostics Bulletin*, published by the Climate Prediction Center, NOAA/NWS/NCEP) and will not be shown here.

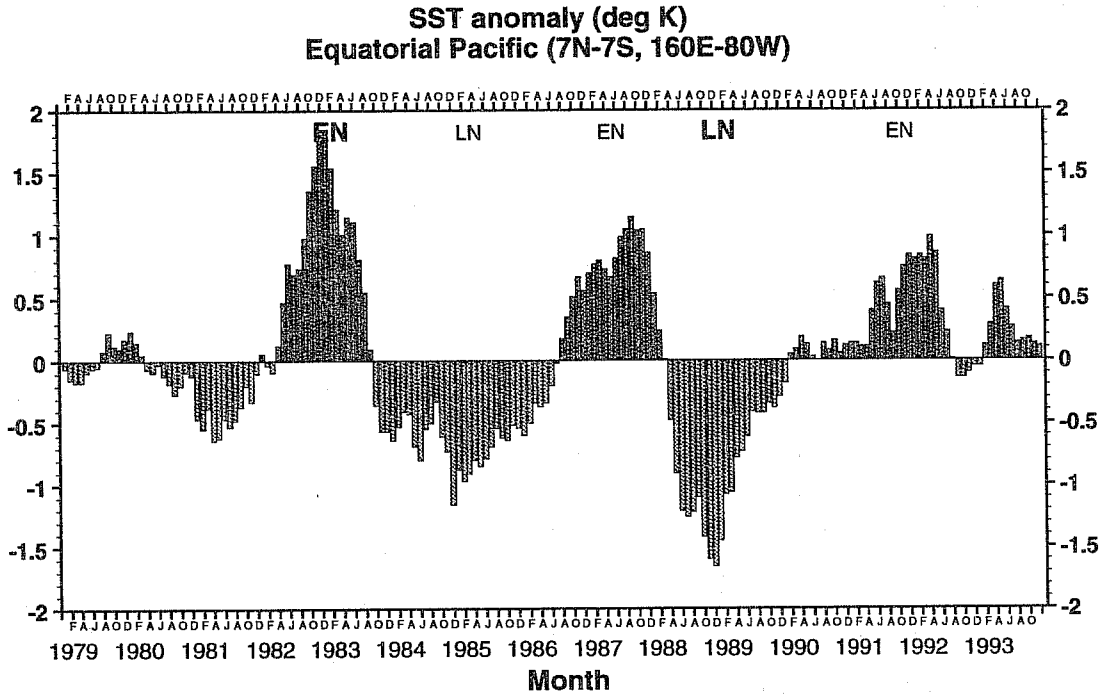


Fig.1 Monthly mean SST anomalies (deg) from 1979 to 1993 averaged over the region 7^0N-7^0S, 160^0E-80^0W. Strong ENSO events are denoted by large heavy EN and LN; moderately strong events are denoted by small EN and LN.

3. Model systematic error

3.1 Errors in the extratropics

For brevity, we focus on seasonal-mean 500 mb height systematic error. Fig.2a shows the systematic error for the last 3 months (months 2-4) of the winter integrations (JFM). The largest errors, of approximately 9 dam, occur over the north Pacific and north Atlantic, and are associated with a midlatitude westerly bias over these oceanic areas. The error pattern, in particular over the Atlantic, is almost unchanged when compared with that of previous ECMWF model cycles (Brankovic 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 JFM is found for the late spring season (AMJ, Fig.2b). The trough-to-ridge amplitude in both Atlantic and Pacific error dipoles is somewhat smaller, whilst the negative error in the Pacific extends over the whole of the northern Asian continent. 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 is 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 (forest areas were not distinguished from clearances; Viterbo and Betts 1998). For the late summer (JAS), the northern hemisphere error is much smaller, apart from very high latitudes, and positive errors are dominant over the continents (Fig.2c). 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.2d), 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.

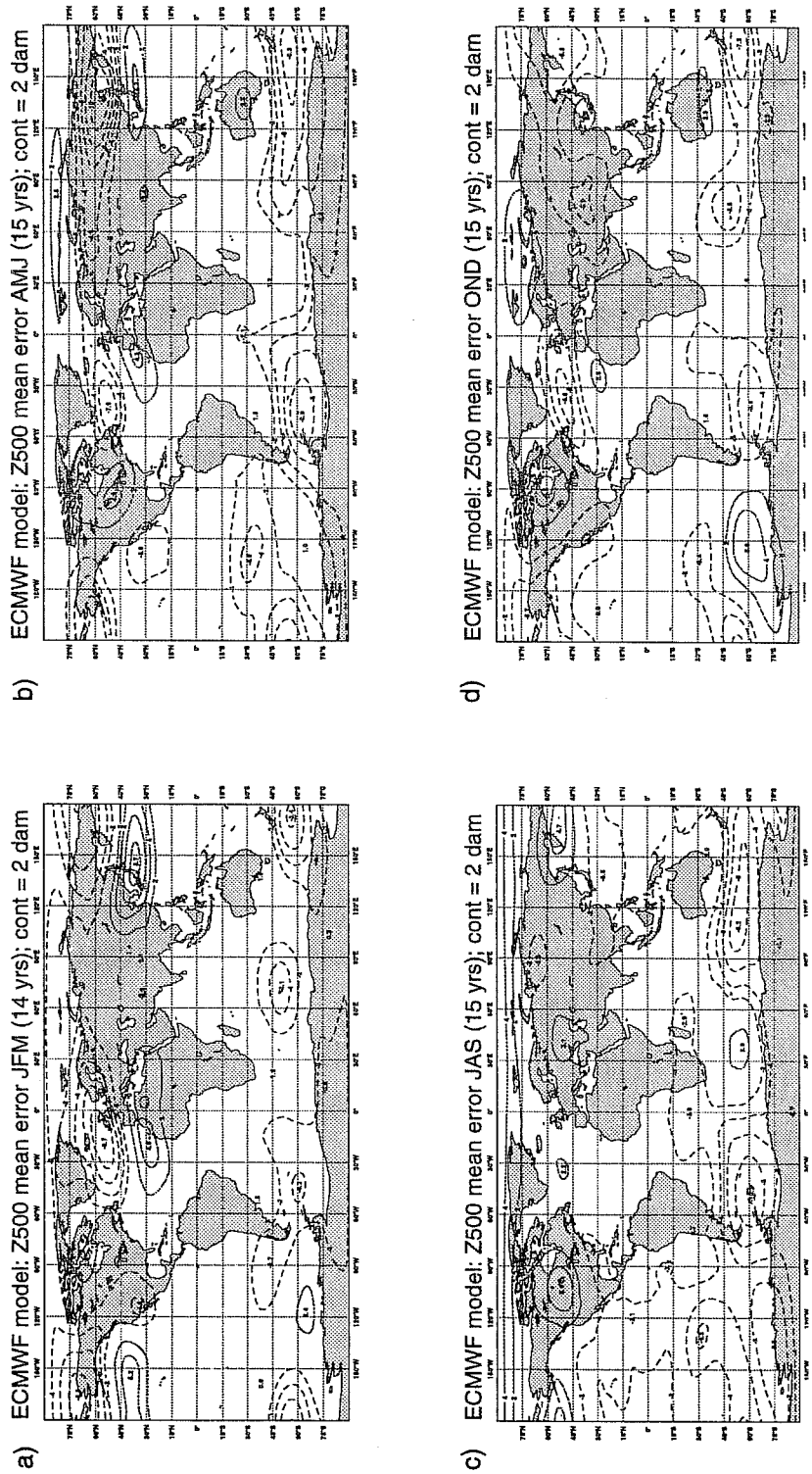


Fig.2 500 mb height systematic error for a) JFM, b) AMJ, c) JAS, and d) OND. Contours every 2 dam, negative errors dashed.

The model error as illustrated in Fig.2 (locally up to 10 dam) is similar in amplitude with a typical magnitude of 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 normalised 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.3b) and smallest in summer (Fig.3c). In the tropics, the relative error is largest in summer, following 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 a relatively small impact on model performance in winter, despite the mean error having large amplitudes.

Given the intrinsic nonlinearity of the atmosphere, it is clear that further work is required to reduce the level of model error on seasonal timescales, at least so that they are small compared with the signal one is attempting to predict. The progress that has been made 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.

3.2 Errors in the tropics

For tropical regions, model errors are assessed in terms of precipitation. As discussed above, the "verification" is based upon the 0-24 hour accumulated precipitation from ERA-15 short-range forecasts, since no precipitation analyses for the whole 1979-1993 period were available. This is certainly a shortcoming in verification; however, such an "error" field gives a large-scale estimate, albeit possibly conservative, of the drift in precipitation over a season.

Fig.4 shows the differences between the simulated and observed climate for the last 3 months (months 2-4) of each season. Over the central and eastern tropical Pacific, the error pattern in all seasons signifies excessive precipitation rates in the simulated inter-tropical convergence zone (ITCZ). In JAS, the excessive ITCZ rainfall is largest of all seasons, reaching 14 mm/day in the tropical Pacific, 20 mm/day in the Atlantic and over 10 mm/day in the tropical Indian Ocean (Fig.4c). The amplitude and the extent of errors in the tropical rainfall in JAS are consistent with relatively large errors in geopotential heights shown in Fig.3c. 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 and equatorial African regions.

4. Assessment of model skill

This section describes some conventional skill scores, appropriate to deterministic ensemble-mean fields. In a companion paper, Palmer *et al.* (1998) discuss the skill of the PROVOST integrations within a probabilistic forecast framework.

4.1 Ensemble mean skill

Figs.5 and 6 show anomaly correlation skill scores for the ensemble-mean 500 mb height, averaged over the months 1-3, and months 2-4 of the integrations. Fig.5 is for the northern hemisphere between 20^{0}N and 80^{0}N, Fig.6 is for Europe only (35^{0}N-75^{0}N, 12.5^{0}W-42.5^{0}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 5 (out of 59) cases. The drop in skill between months 1-3 and months 2-4 is especially large in spring and summer.

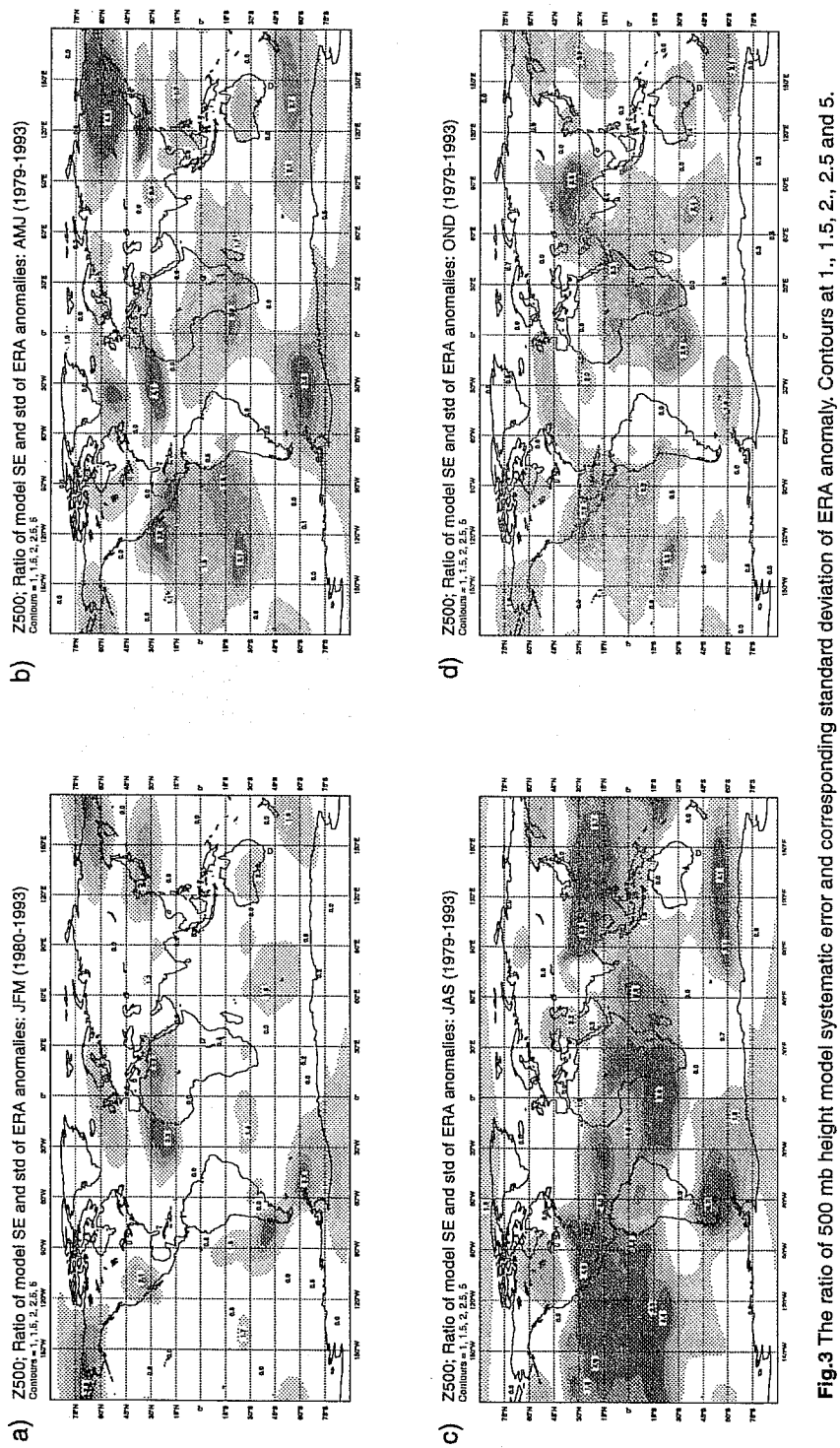


Fig.3 The ratio of 500 mb height model systematic error and corresponding standard deviation of ERA anomaly. Contours at 1., 1.5, 2., 2.5 and 5.

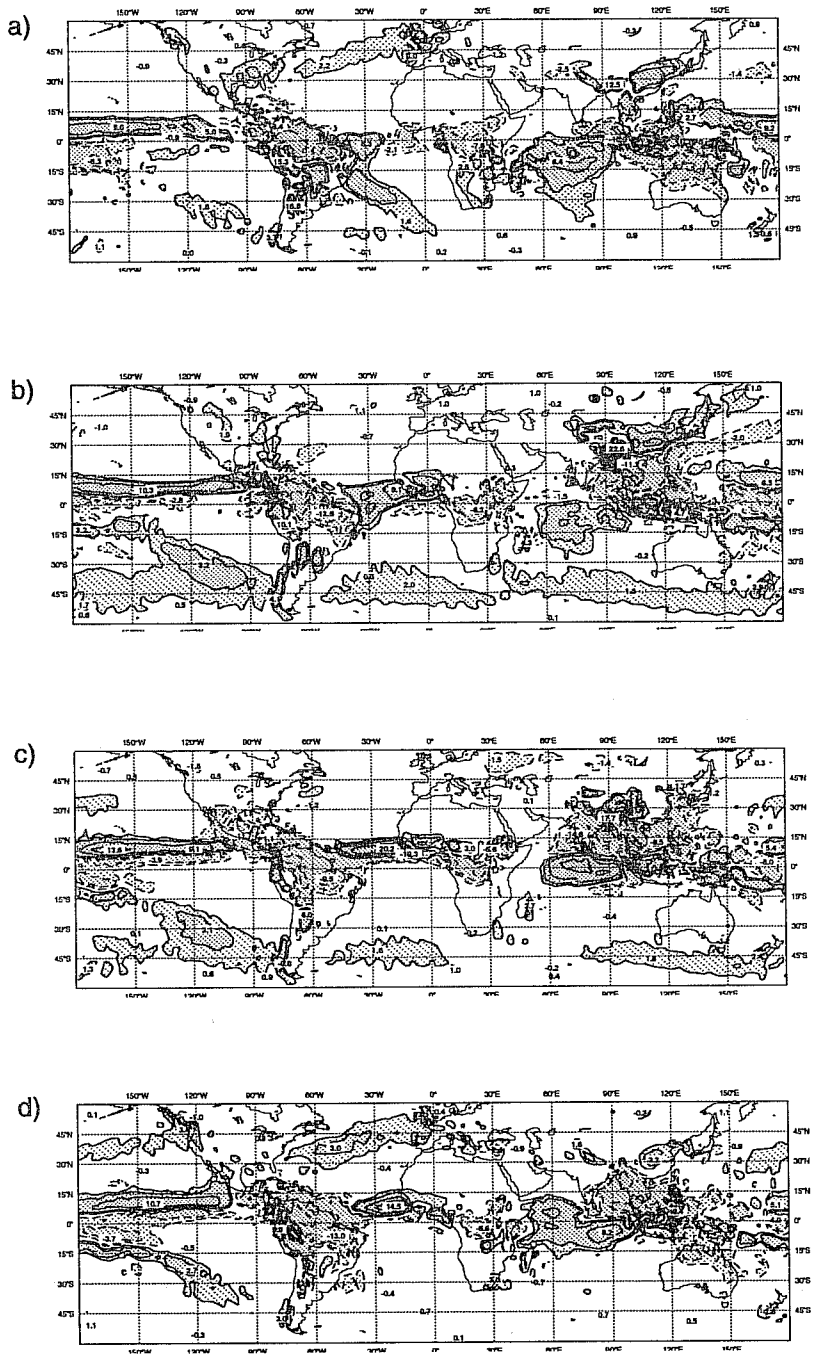


Fig.4 Total precipitation "error" (the difference between model and 24-hour ERA forecasts) for:
a) JFM, b) AMJ, c) JAS, and d) OND. Contours +/- 1, 2, 5, 10, 20 mm/day. Negative errors dashed.

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 the months 2-4 is more often above the 0.5 mark for Europe (11 cases) than for the northern hemisphere (5 cases). In autumn (SOND), skill is rather poor for Europe - in SON skill is negative (worse than climatology) in 12 out of 15 years (Fig.6a).

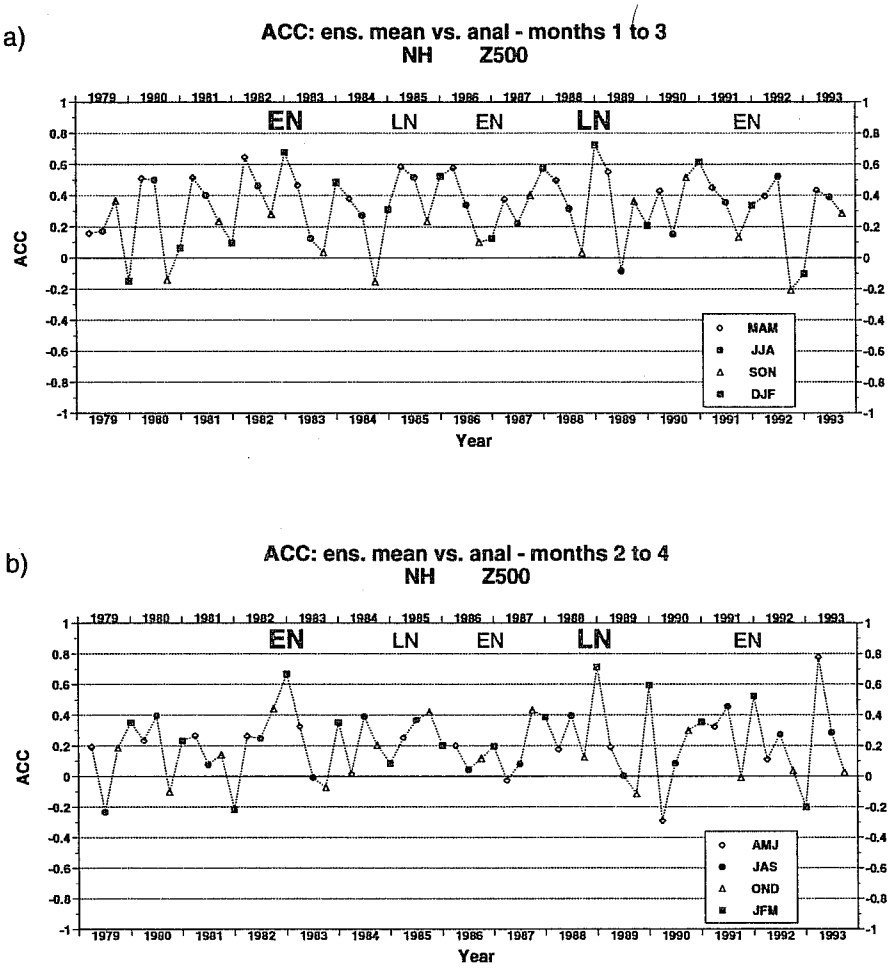


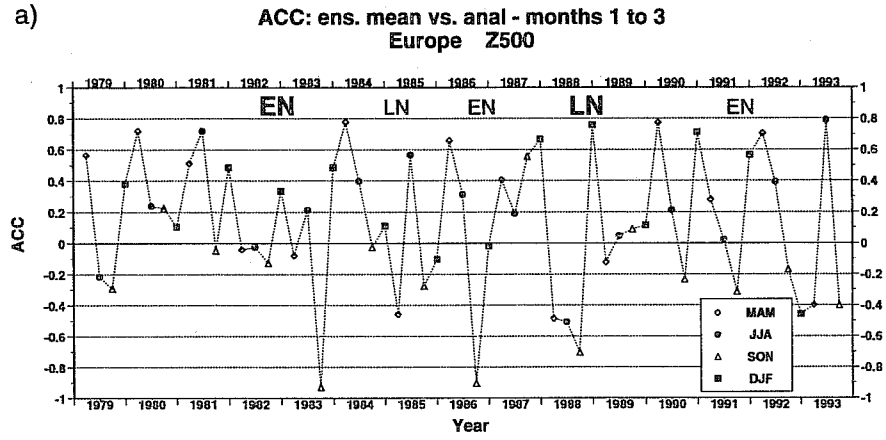
Fig.5 The northern hemisphere (20-80^{0}N) 500 mb height ensemble-mean anomaly correlation coefficients for all four seasons for a) months 1-3 and b) months 2-4. Strong ENSO events are denoted by large heavy EN and LN; moderately strong events are denoted by small EN and LN.

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 the winter 1988/89 has high skill, relatively low skill (below 0.4) is found in the winter 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 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

a)



b)

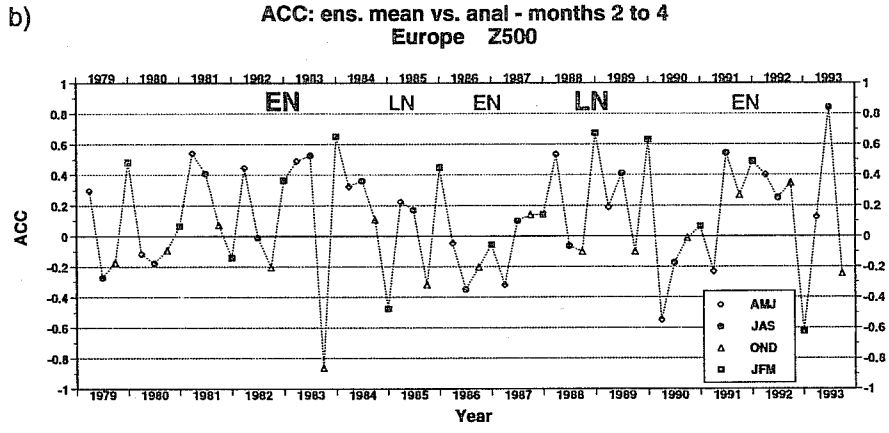


Fig.6 Same as Fig.5 but for Europe (35-75^{0}N, 12.5^{0}W-42.5^{0}E).

hemisphere sectors was found in MAM (Brankovic 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 initial conditions are more predictable in spring than in winter. The impact of initial conditions 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^{0}N, 150-60^{0}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 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, despite being relatively low, skill is consistently enhanced for months 2-4 (AMJ, JAS, OND and JFM) and for DJF as well. This may indicate a delay in time it takes ENSO to influence seasonal predictability over the European region. However, due to the small sample, values in Table 2 (5 cases) may not be

considered definitive. (In contrast to these results, forecasts for the ENSO winter 1997/98 were found to be quite skilful for European region, see Stockdale *et al.* 1998.)

Table 2: 500 mb height ensemble-mean anomaly correlation coefficients.Figures in brackets are the means from strong and moderately strong ENSO years

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)

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.

Fig.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 rather well anomalies over the Pacific/North America (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.7d) 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: broadly, negative anomalies in the north and positive anomalies in the south. This illustrates the difficulties in seasonal prediction over Europe even in the presence of a strong ENSO forcing.

4.2 Skill of individual integrations

In some sense, the ensemble mean skill characterizes the overall 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 ensemble but the skill of the ensemble mean field is low.

Fig.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.8b) there is an overall higher degree of skill consistency within ensembles than in winter (Fig.8a). This is reflected in 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

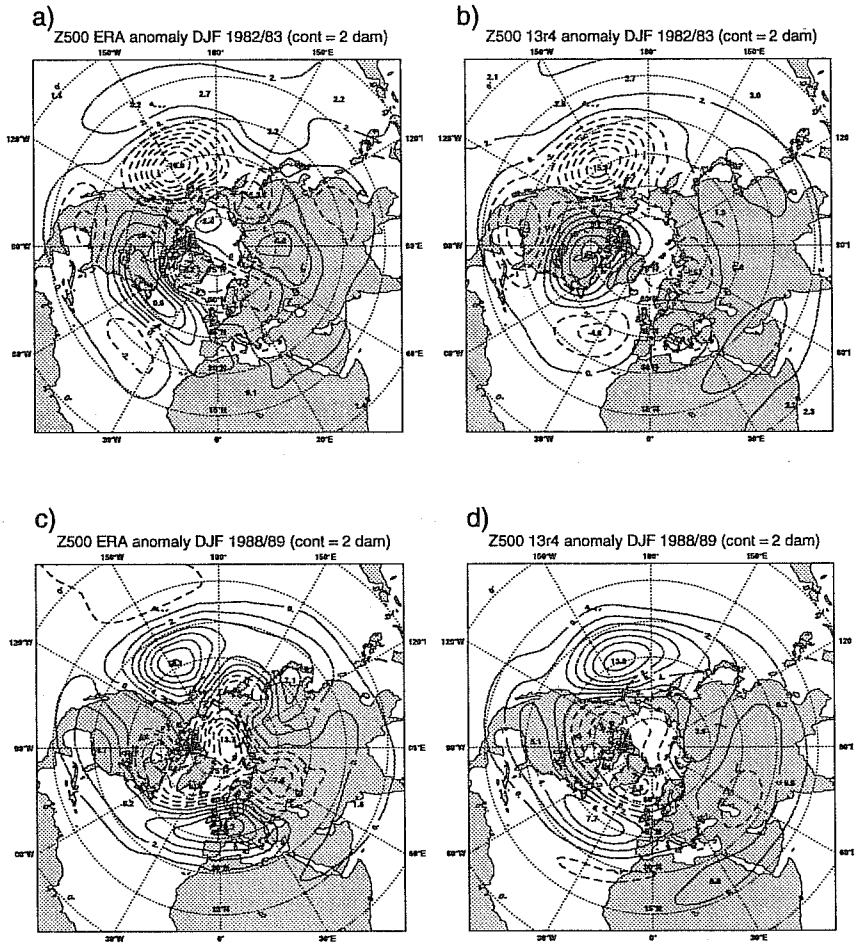


Fig.7 500 mb height anomalies: a) ERA for DJF 1982/83, b) ensemble-mean for DJF 1982/83. c) and d) are as a) and b), but for DJF 1988/89. Contours every 2 dam. Negative anomalies dashed.

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.

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. Figs.9 and 11 show simulation of interannual variations of such anomalies for some prescribed tropical and extratropical regions respectively. The solid line shows the rainfall anomaly based on the ensemble mean field. The dashed line shows the anomaly in terms of the 0-24 hour accumulated field in the ERA-15 archives. The dotted lines show the extreme (wet and dry) members of the ensemble. In all cases the curves have been standardised 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.9a), east Africa (MAM; Fig.9b) and the African Sahel (JJA: Fig.9c), are clearly skilful. For the latter region, it is

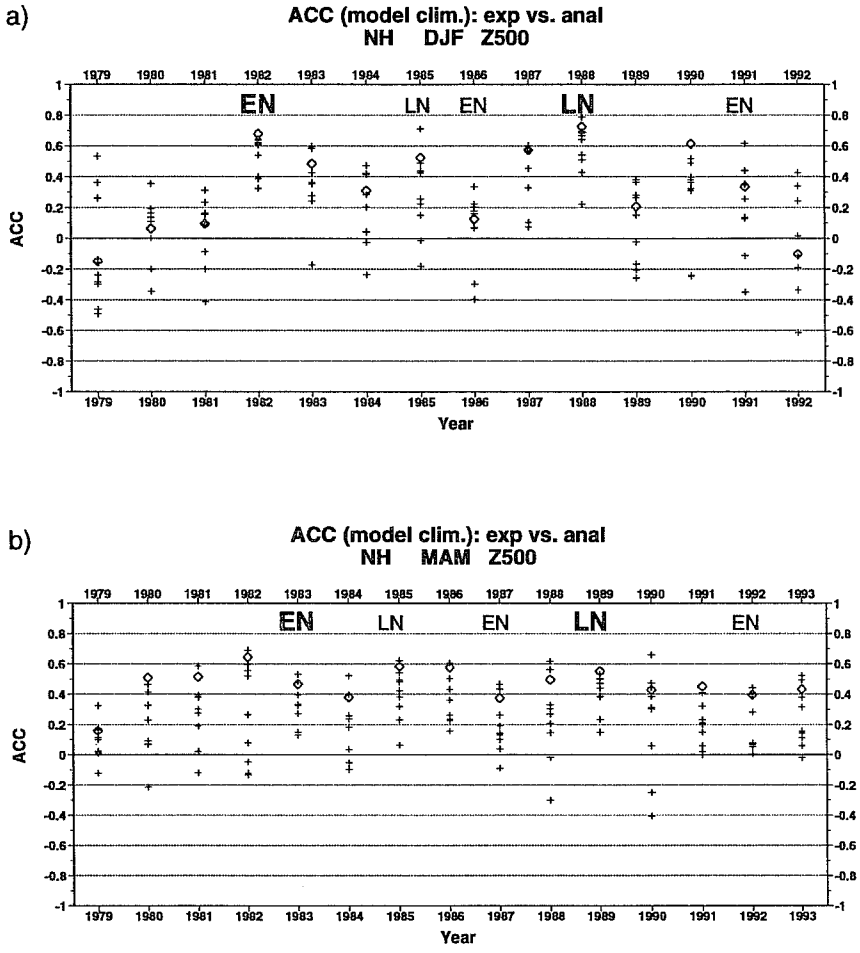


Fig.8 The northern hemisphere (20-80^{0}N) 500 mb height anomaly correlation coefficients for: a) DJF, and b) MAM. Open diamonds - ensemble means, crosses - individual model integrations.

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.9d), 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 models' success is possibly 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.9d, where the distance between the thin dotted lines indicates a range of model realizations. In contrast with 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 the southeast Asia are quite predictable - all nine members of the 1987 ensemble indicate a large positive anomaly over that

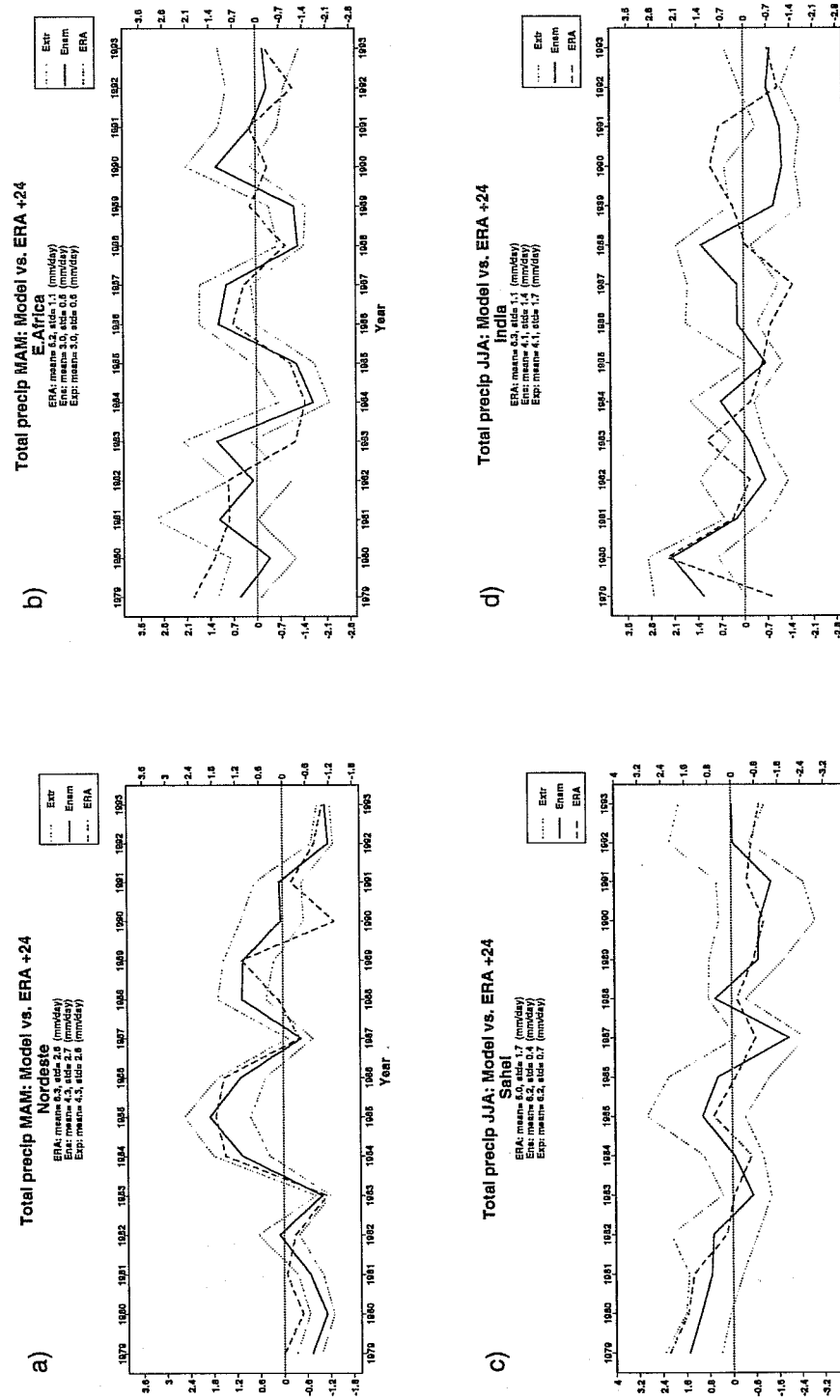


Fig.9 Standardised precipitation rate for ensemble mean (solid), 0-24 hour accumulated from ERA (dashed) and extremes of ensemble (dotted) for: a) Brazilian Nordeste in MAM, b) east Africa in MAM, c) African Sahel in JJA, and d) India in JJA. The standardisation is performed relative to the mean and standard deviation of the model climate or ERA climate, as appropriate. In the headers, "Ens" denotes 15-year mean and standard deviation for ensemble-mean fields. "Exp" and "ERA" denote individual integrations and reanalysis respectively.

region. For JJA 1988, the results are more robust over India in the sense that positive rainfall anomalies are dominant in 6 out of 9 experiments (not shown).

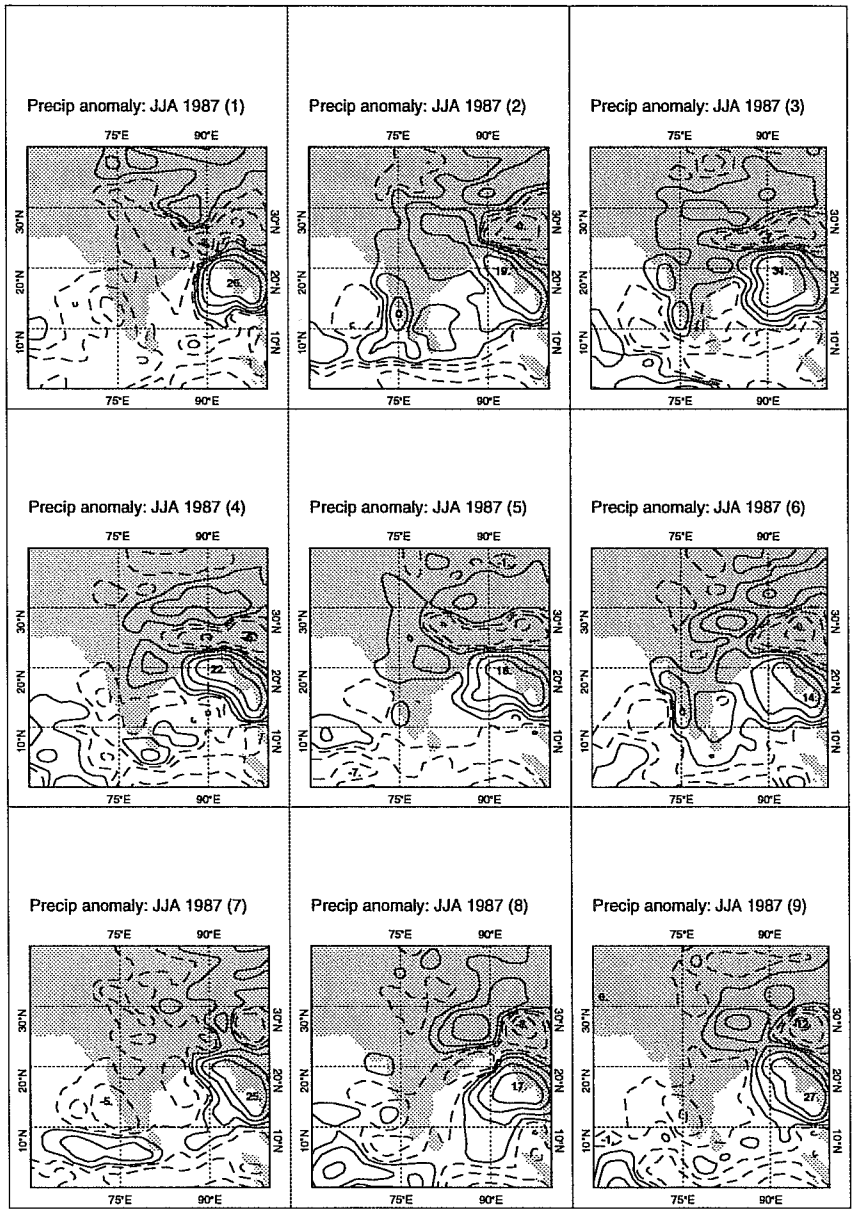


Fig.10 Precipitation anomalies for JJA 1987 over the Indian region for all 9 indvidual model realizations. Contours +/- 0.5, 2, 5, 10, 30 mm/day. Negative anomalies dashed.

In the headers 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 ERA-15. 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 about 2 mm/day drier than the reanalysis values. On the other hand, the ensemble and reanalysis interannual variability are about the same. For the east Africa the ensemble is again drier by about 2 mm/day

and the ensemble standard deviation is smaller than the reanalysis interannual variability. For the Sahel the ensemble is wetter than the reanalysis, but the reanalysis has much more interannual variability than the ensemble. 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, though the model tends to underestimate interannual variability of rainfall rates.

The standardised rainfall estimates have also been studied for a number of extratropical regions. In general the level of skill is much lower when compared with the tropics. However, for some regions the results were encouraging. Fig.11 shows that the interannual variability in JFM rainfall for the northern Europe (50-70°N, 10°W-30°E) and the eastern United States (25-50°N, 100-70°W) was captured reasonably well. For northern Europe the ensemble is wetter than the reanalysis, though the standard deviation is about the same.

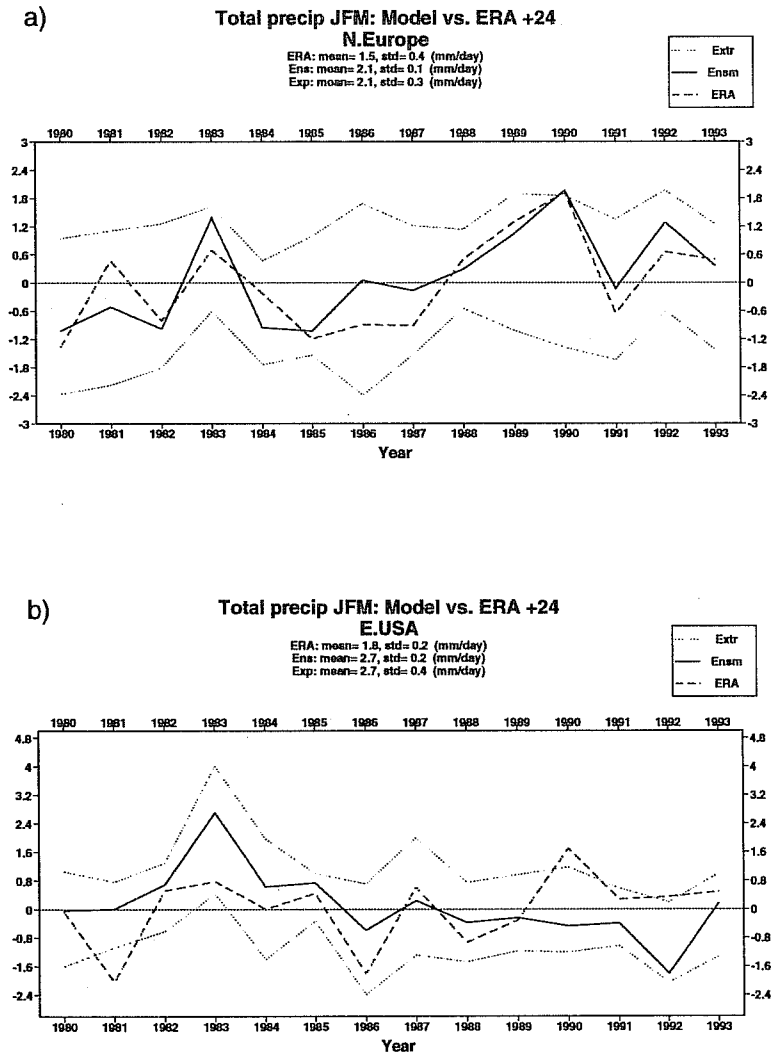


Fig.11 Same as Fig.9 but for: a) northern Europe in JFM, and b) eastern United States in JFM.

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 model's internal dynamics (see, for example, Zwiers 1996).

6.1 Significance of ensemble-mean anomalies

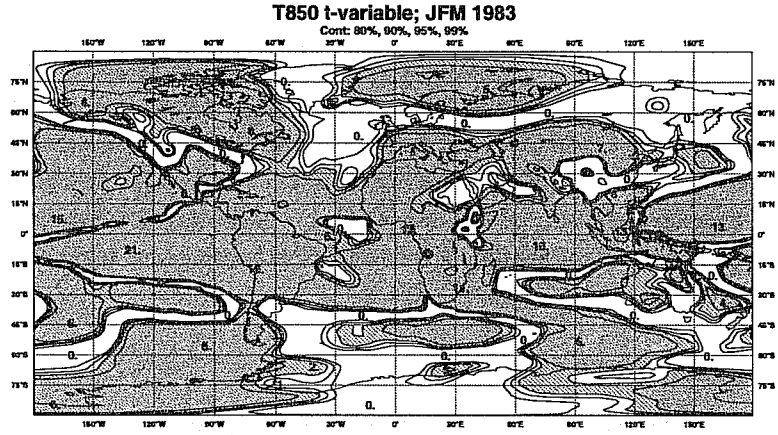
A *t*-test is performed, grid point by grid point, on the 850 mb temperature anomalies, for a given year, where climatology is obtained from all the integrations excluding those from the given year. Fig.12 shows some examples of such spatial distribution of *t*-values for the late winter season (JFM). The years displayed have been chosen as to illustrate winters when the equatorial Pacific SST forcing was strong, moderately strong and weak respectively. 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.12b and 12a 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 the northern Europe in the winter 1984/85 was enhanced, for example, by north Atlantic SST anomalies. For comparison, Fig.12c 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).

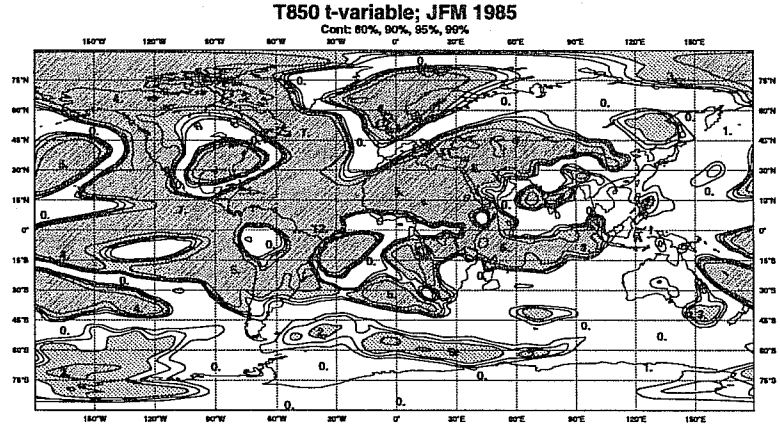
In Table 3 is shown the time-mean *t*-statistic for three selected regions. For the northern hemisphere, *t*-values remain almost unchanged throughout winter, whereas in other seasons there is a decrease in predictability between months 1-3 and months 2-4. For Europe, predictability estimates are generally 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 winter, a reduction in predictability estimates in months 2-4 is seen as well. The enhanced predictability in winter is probably associated with the seasonality of teleconnections between the tropics and extratropics. An 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, 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 line represents 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 subregions 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

a)



b)



c)

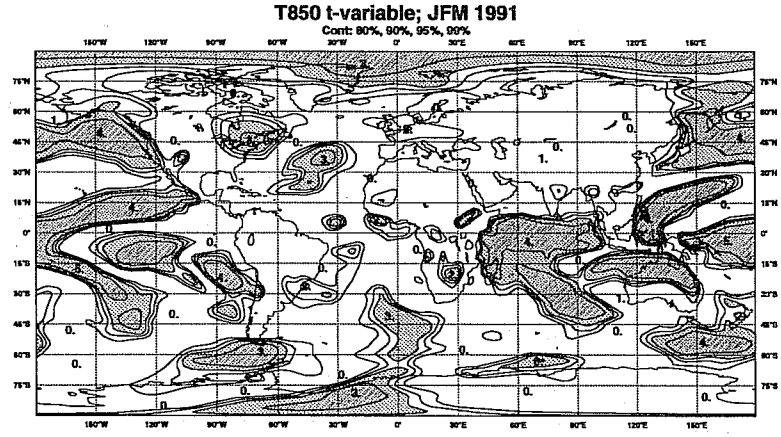


Fig.12 Student t-statistic of 850 mb temperature for the JFM season in: a) 1983, b) 1985, and c) 1991. Contours at 80%, 90%, 95% and >99% confidence levels. Shading for levels greater than or equal to 95%, which corresponds to t-values of 1.980 or greater.

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, northern Europe shows predictability in four years, southern Europe in two years. The opposite is found in summer (Fig.14); there are three years above the 90% threshold for southern Europe, whereas no years are predictable for northern Europe.

Table 3: Mean *t*-statistic for 850 mb temperature. Figures in brackets are the means from strong and moderately strong ENSO years

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)

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 eastern United States and for rainfall over east Africa and India. The eastern US region is in the extratropics, much closer to the centre of action in the equatorial Pacific than Europe; the other two regions are in the tropics.

For the eastern US, 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.15a). A similar result is seen for MAM tropical precipitation *t*-values over eastern Africa (Fig.15b). In addition, predictability of east Africa rainfall is, on average, relatively high - most of the open square symbols are lying above 90% significance level. For India, apart from the year 1980, the *t*-values are always below the 99% confidence limit, and in 8 years (out of 15) are below or equal 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.

6.2 Ratio of variance

As mentioned above, the gridpoint 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.16a), it can be seen that across much of the tropical Pacific, where the ensemble mean is significantly different from climatology (cf. Fig.12a), 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

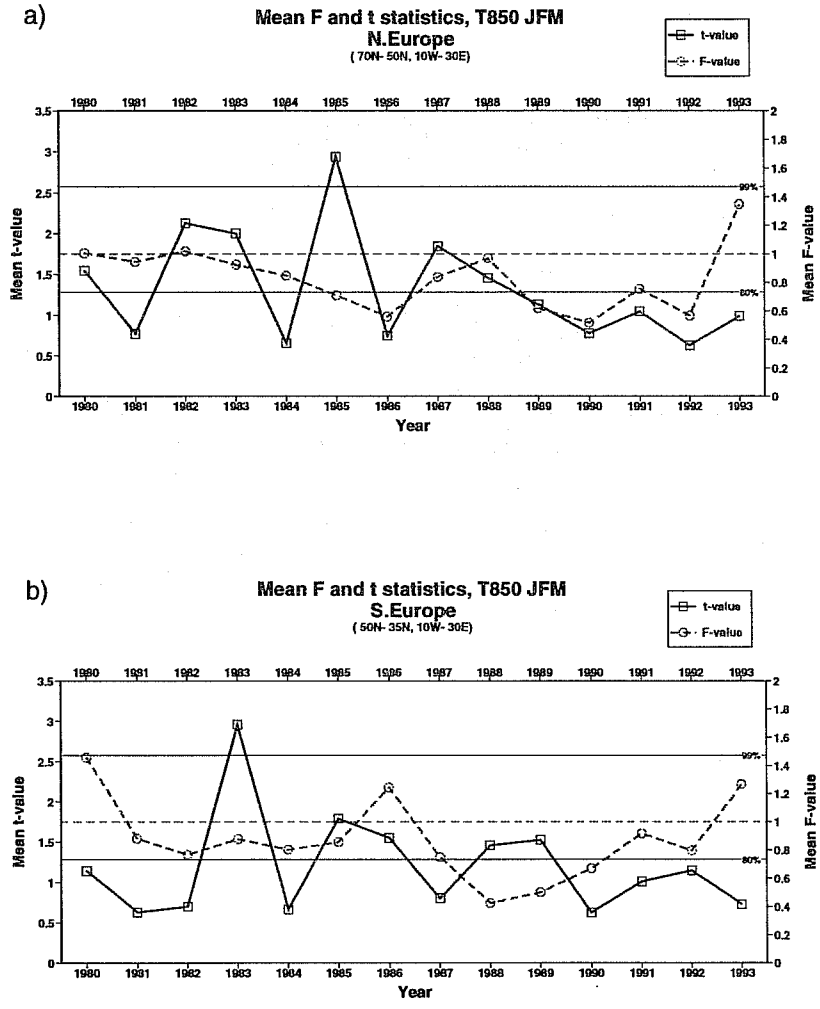


Fig.13 The interannual variation of the JFM area mean 850 mb temperature t-statistic (solid) and ratio of variance (dashed) for a) northern Europe, and b) southern Europe.

1. By contrast, the area over the northern Scandinavia and to the east, which was significant according to the *t*-test, had variance ratio exceeding 1. 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.16a (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 sub-domains are shown by the heavy dashed line 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 thin dashed horizontal line. Years with 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

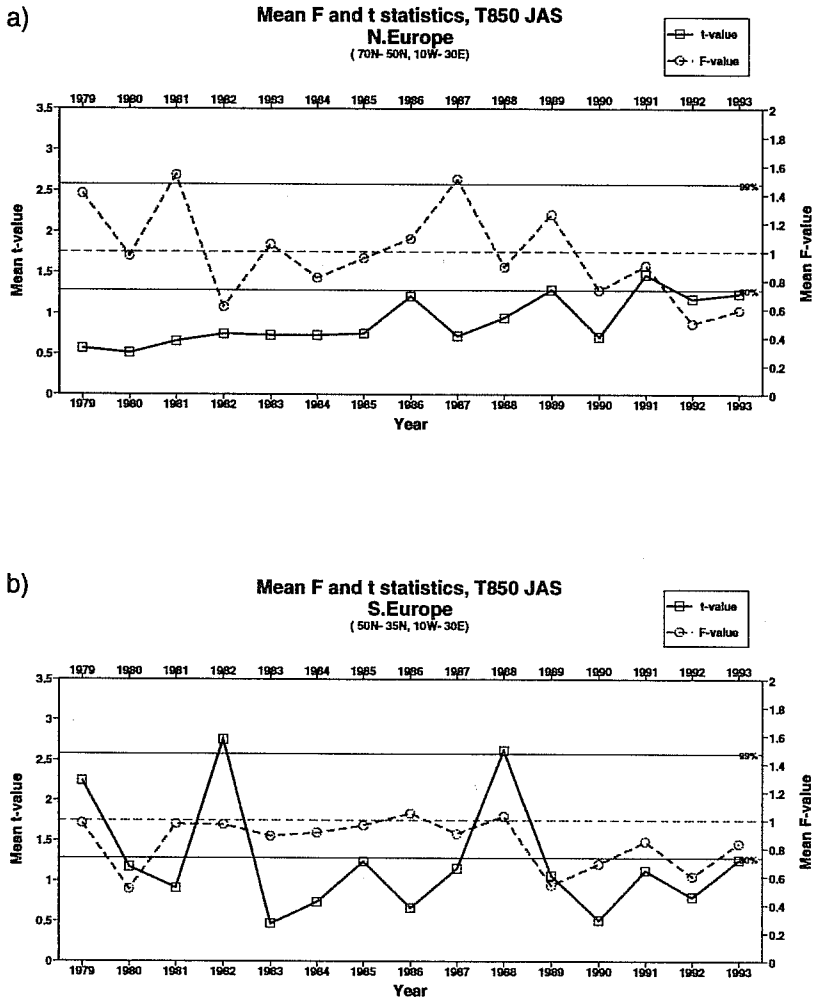


Fig.14 Same as Fig.13 but for the JAS season.

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 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.15b)) are found to be much lower than those over Europe.

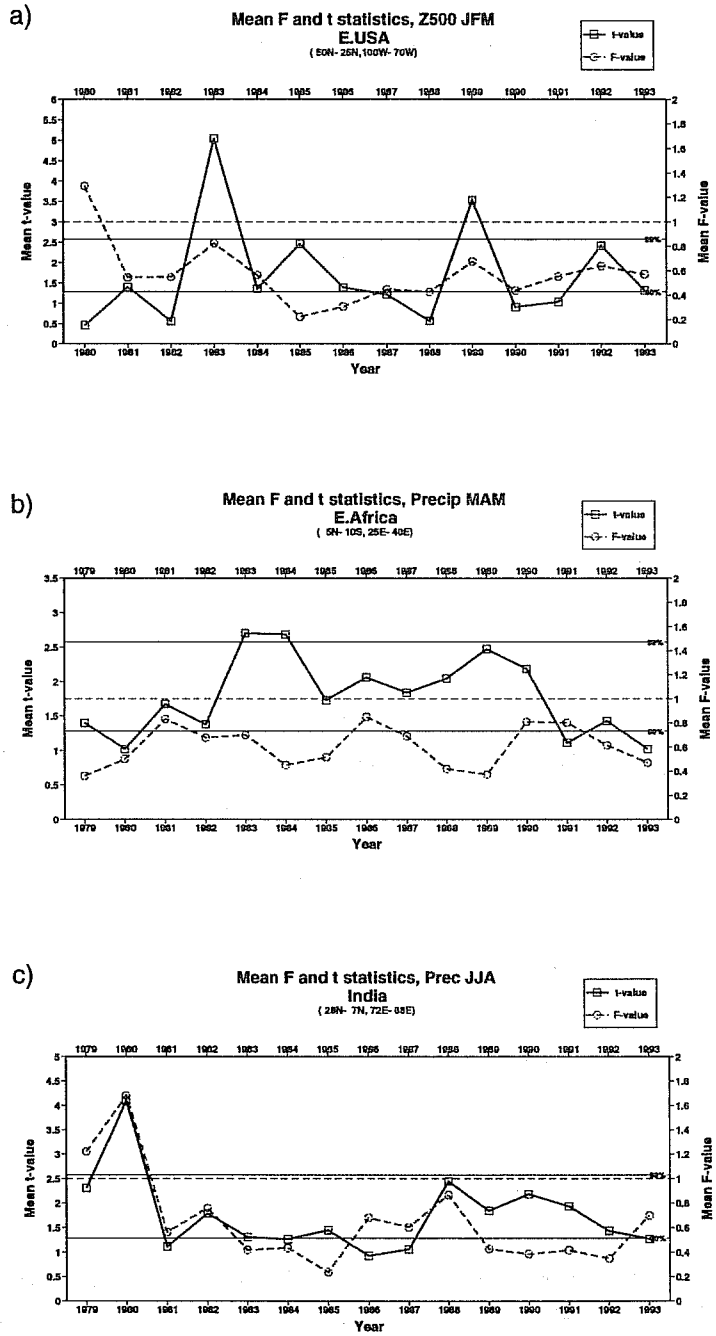


Fig.15 Same as Fig.13 but for a) eastern United States 500 mb height in JFM, b) eastern Africa precipitation in MAM, and c) Indian precipitation in JJA.

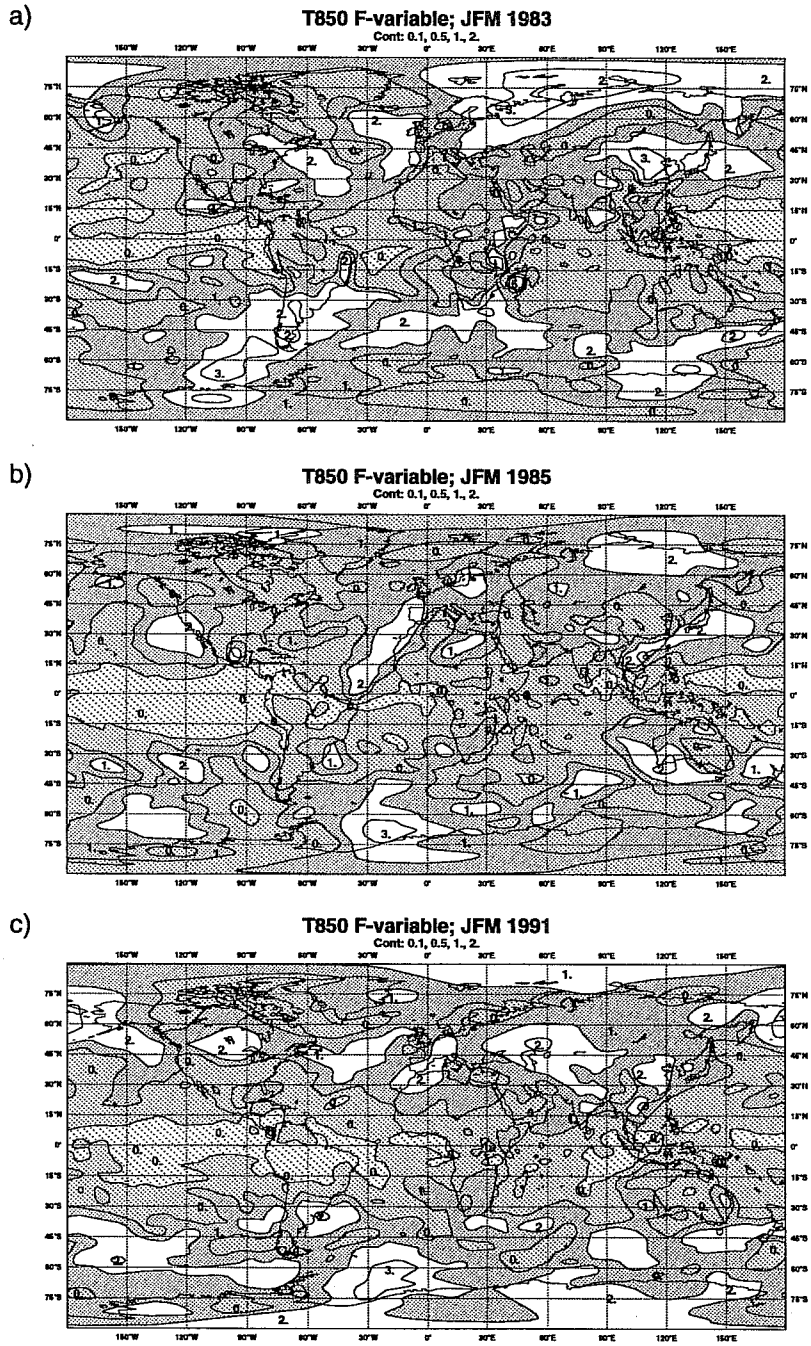


Fig.16 The ratio of variance of 850 mb temperature for the JFM season in: a) 1983, b) 1985, c) 1991. Contours at 0.1, 0.5, 1.0 and 2.0. Shading for values below 1.0 only.

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

Additional ensembles have been run for 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 initial conditions (ICs; which include land surface initial conditions). These additional ensembles consist of 6 members each; the control ensembles are the subsets of the 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.

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}

7.1 Regional impact

Results are shown in Figs.17 and 18 for precipitation in two regions: northern subtropical Africa (Sahel) during JJA and United States during AMJ. For the Sahel, 1987 was a drought year; for the eastern US, 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. For the 1988 US drought, see, for example, Trenberth and Branstator 1992 and Palmer and Brankovic 1989).

For much of the Sahel, the model control ensemble simulated well the observed increase in precipitation in JJA 1988 relative to JJA 1987 (Fig.17a). The associated *t*-values, significant at 95% confidence level, are found to the west of the Greenwich meridian and to the east of the Lake Chad (15°E, Fig.17b). Fig.17c 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 initial conditions in the hybrid ensemble $\{IC_{87}, SST_{88}\}$ were 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 the subtropical northern Africa (Fig.17d).

The impact of ICs is smaller than that of SSTs ($\{IC_{88}, SST_{87}\}$-$\{IC_{87}, SST_{87}\}$, Fig.17e), however, by no means unimportant. The increase in precipitation is localised 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 strongly significant as 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).

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

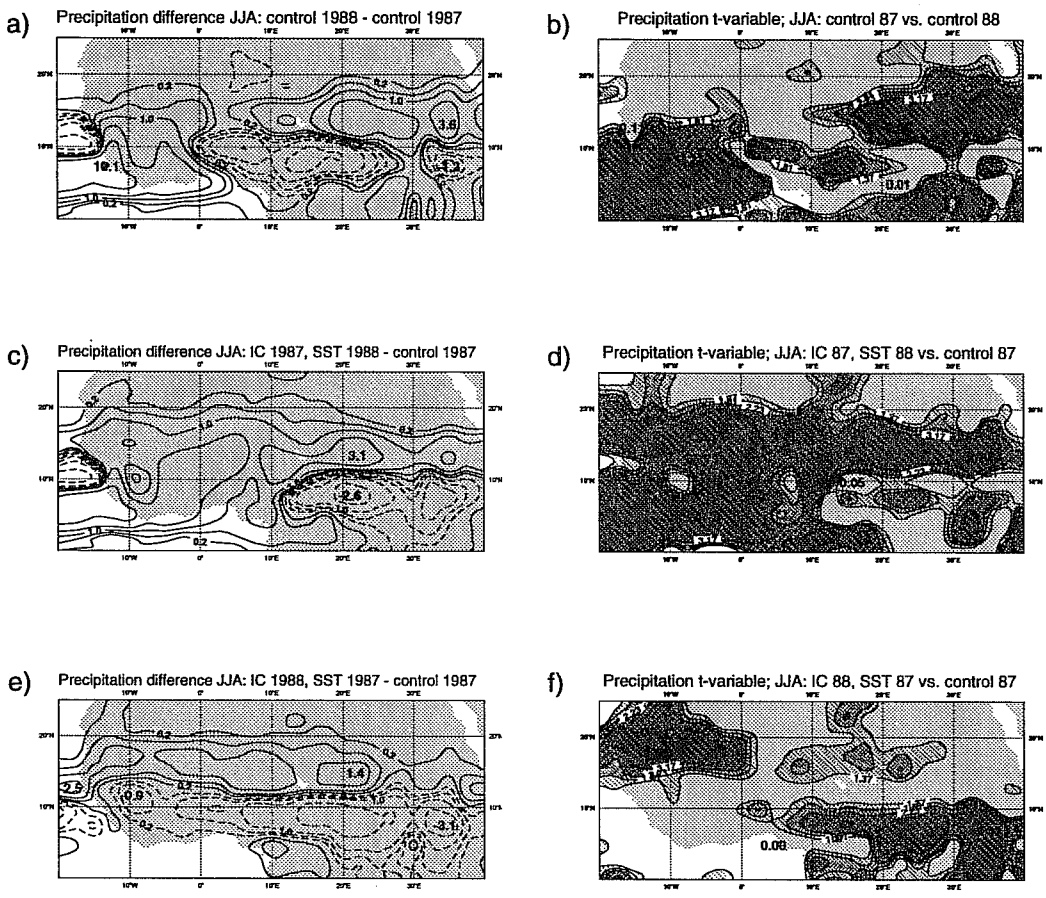


Fig.17 Seasonal mean JJA rainfall over the subtropical Africa (Sahel): a) difference between the 6-member control ensembles for 1988 and 1987, b) t-statistics for a); c) difference showing the impact of SST forcing, d) t-statistics for c); e) difference showing the impact of initial conditions, f) t-statistics for e). Contours in a), c) and e) +/- 0.2, 0.5, 1, 2, 5, 10 mm/day. Contours in b), d) and f) at 80%, 90%, 95% and >99% confidence levels (these correspond to the t-values of 1.37, 1.81, 2.23 and 3.17 respectively).

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

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 southwestern Europe, the influence of SST prevails over that of ICs. The diversity of the above results illustrate the difficulty in attributing rainfall predictability over Europe.

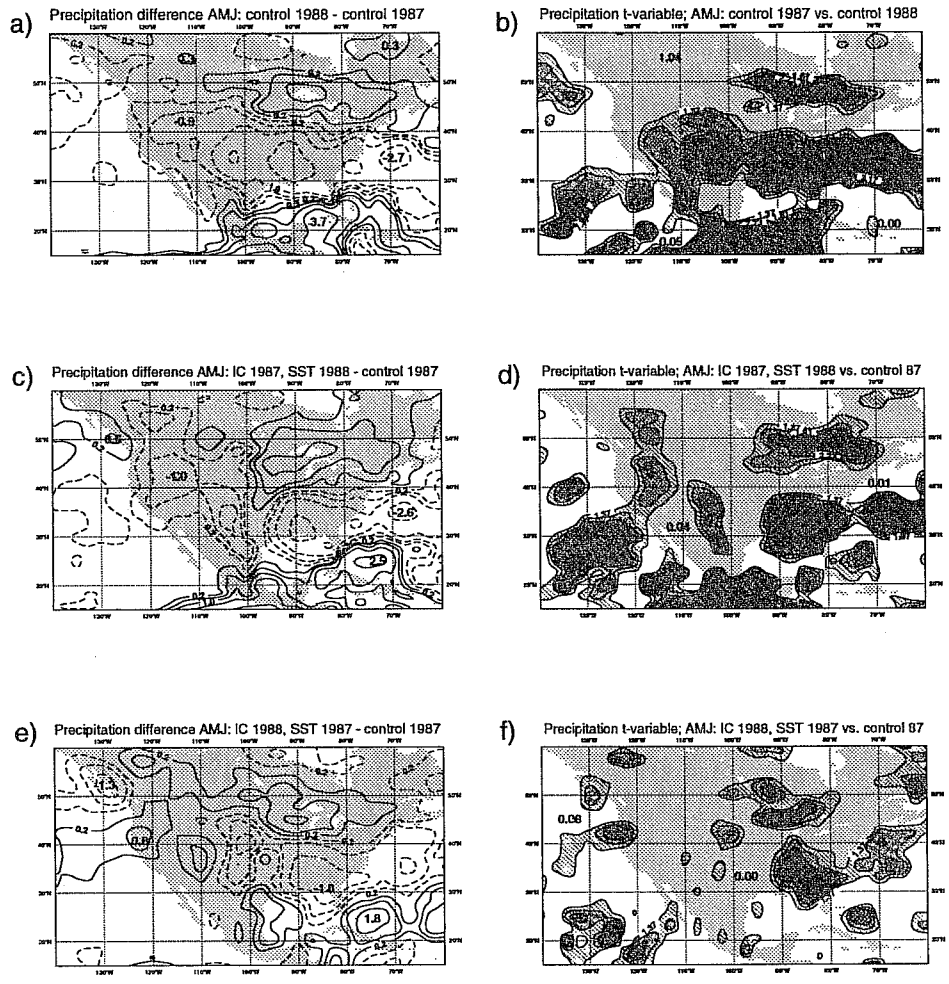


Fig.18 As Fig.17 but for AMJ rainfall over the United States.

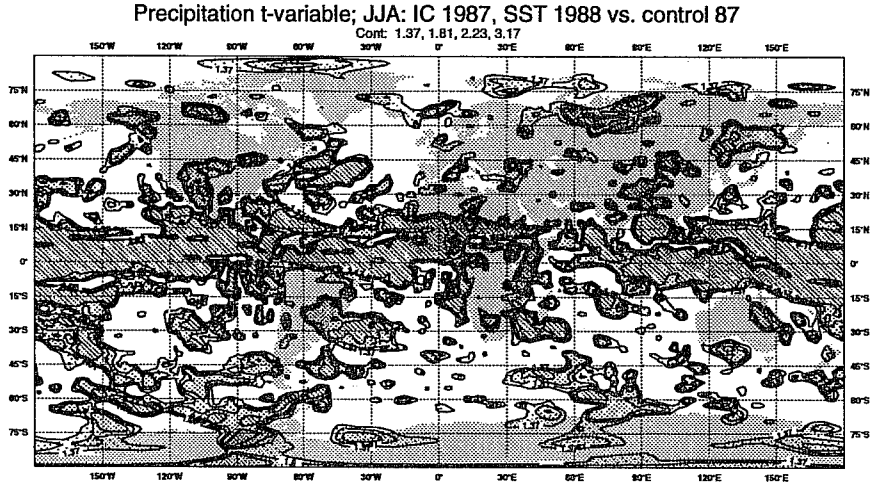
7.2 Global impact

The relative impact of initial conditions and SST forcing on global precipitation rates was also studied; here, we discuss the results for global precipitation difference fields.

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 made 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 of land areas are very similar (in both pattern and amplitude) to the differences between the control ensembles for the two years considered. In Fig.19a, 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" initial conditions 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 a much patchier pattern over both oceans and continents. Over some regions these differences are similar to those between the controls of different years, however, no

a)



b)

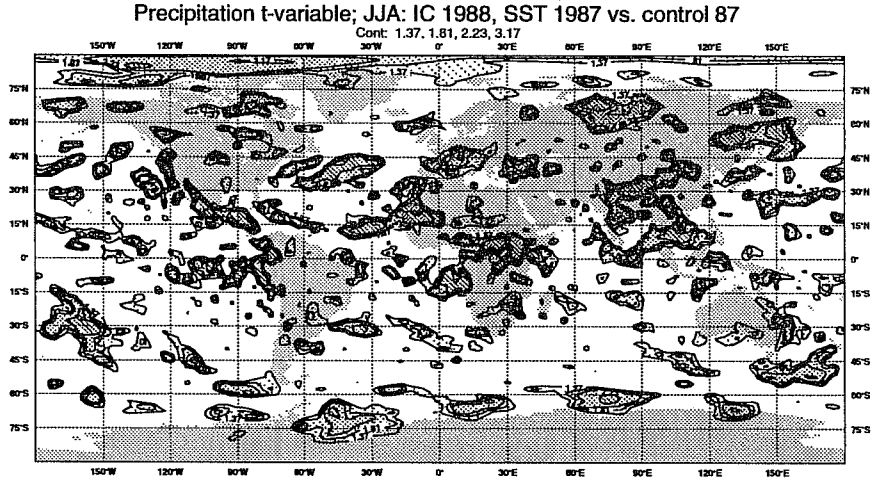


Fig.19 *t*-statistics for seasonal mean JJA rainfall difference between hybrid and control ensembles for: a) "wrong" SSTs in hybrid experiments, and b) "wrong" ICs in hybrid experiments.

truly global pattern is evident (Fig.19b depicts the $\{IC_{88}, SST_{87}\}$-$\{IC_{87}, SST_{87}\}$ case). Based of 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 7.1.

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 SST indicates that the results shown here may be assumed as an upper limit on what can be obtained with a coupled ocean-atmosphere model. However, two *caveat*s should be pointed out. 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 midlatitudes 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, 1998). 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 consistent 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 both 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 initial conditions (which includes land-surface initial conditions) by performing additional ensembles of hybrid integrations with SST from one particular year, and initial conditions from a different year. Overall the results confirm the predominant importance of SST forcing. However, on a regional level, the results show that initial condition effects cannot be ignored for quantitative seasonal prediction. On this basis, seasonal predictability estimates from AMIP integrations (Gates 1992), in which the initial conditions are largely irrelevant, can be expected to be quantitatively different from the results of these PROVOST integrations. The impact of SSTs and initial conditions are not linearly additive, another manifestation of the nonlinearity of climate.

Acknowledgment: This research was supported under the European Union Environment and Climate Programme under contract CT95-0109.

References

Arpe, K., L. Dümenil, and L. Bengtsson, 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 Bulletin*, **47**, 276-280.

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

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

Boer, G.J., K. Arpe, M. Blackburn, M. Déqué, W.L. Gates, T.L. Hart, H. le Treut, E. Röckner, D.A. Sheinin, I. Simmonds, R.N.B. Smith, T. Tokioka, R.T. Wetherland and D. Williamson, 1991: *An intercomparison of the climates simulated by 14 atmospheric general circulation models*. WMO/ICSU WCRP-58, WMO/TD No. 425.

Brankovic, C. and F. Molteni, 1997: Sensitivity of the ECMWF model northern winter climate to model formulation. *Clim. Dyn.*, **13**, 75-101.

Brankovic, C. and T.N. Palmer, 1997: Atmospheric seasonal predictability and estimates of ensemble size. *Mon. Wea. Rev.* **125**, 859-874.

Brankovic, C., T.N. Palmer and L. Ferranti 1994: Predictability of seasonal atmospheric variations. *J. Clim.* **7**, 217-237.

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

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

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

Fennessy, M.J., J.L. Kinter III, B. Kirtman, L. Marx, S. Nigam, E.Schneider, J. Shukla, D. Strauss, A. Vernekar, Y.Xue and J. Zhou, 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. Amer. Meteor. Soc.*, **73**, 1962-1970.

Gibson, J.K., P. Kållberg, S. Uppala, A. Hernandez, A. Nomura and E. Serano, 1997: *ERA description*. ECMWF Re-Analysis project Report Series.

Harzallah, A. and R. Sadourny, 1995: Internal versus SST-forced atmospheric variability as simulated by an atmospheric general circulation model. *J. Climate*, **5**, 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. Meteor. Soc. Japan*, **69**, 271-291.

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

Kumar, A. and M.P. Hoerling, 1998: Annual cycle of Pacific/North American seasonal predictability associated with different phases of ENSO. *J. Climate* (to appear).

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

Miller, M., M. Hortal and C. Jakob, 1995: A major operational forecast model change. *ECMWF Newsletter*, No. **70**, 2-5.

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

Palmer, T.N., 1994: Chaos and predictability in forecasting the monsoons. *Proc. Indian Nat. Sci. Acad.*, **60**, 57-66.

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

Palmer, T.N., C. Brankovic, P. Viterbo and M.J. Miller, 1992: Modeling interannual variations of summer monsoons. *J. Climate*, **5**, 399-417.

Palmer, T.N., C. Brankovic and D. Richardson, 1998: A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. Submitted.

Peng, S. and J.S. Whitaker, 1998: Mechanisms determining the atmospheric response to midlatitude SST anomalies. *J. Climate* (to appear).

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

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

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

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

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

Trenberth, K.E., G.W. Branstator, D. Karoly, A. Kumar, N.-C. Lau and C. Ropelewski, 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 A.K. Betts, 1998: The forecast impact of changes to the albedo of the boreal forests in the presence of snow. Submitted to *J. Geophys. Res.* (ECMWF Tech. Memo. No. 256)

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.