Atmospheric predictability in seasonal ensemble integrations

Čedo Branković

ECMWF
Shinfield Park, Reading, RG2 9AX, England

Abstract

In the first part of the paper, some general ideas on seasonal predictability based on statistical properties of atmospheric circulation are discussed. The distinction between seasonal predictability in the tropics and extratropics is signified. The importance of the ensemble approach for seasonal predictions is emphasised.

In the second part, results from a set of 9-member ensemble seasonal integrations with a T63L19 version of the ECMWF model are presented. The integrations are made using observed specified sea surface temperature (SST) from the 5-year period, 1986 to 1990, which included both warm and cold El Niño-Southern Oscillation (ENSO) events. The distributions of ensemble skill scores and internal ensemble consistency are studied. For years in which ENSO was strong, the model generally exhibits a relative high skill and high consistency in the tropics. In the northern extratropics, the highest skill and consistency are found for the northern Pacific/North American region in winter, whereas for the northern Atlantic/European region the spring season appears to be both skilful and consistent.

Applying a t-test to interannual fluctuations, estimates of a minimum useful ensemble size are made. Explicit calculations are made with ensemble size varying between 3 and 9; estimates for larger sizes are made by extrapolating the t-values. Based on an analysis of 2m temperature and precipitation, the use of relatively large ensembles for extratropical predictions is likely to be required; in the tropics, smaller-sized ensembles may be adequate during years in which ENSO is strong.

The role of the SST forcing in a seasonal time-scale ensemble is to bias the probability distribution function (PDF) of atmospheric states. Such PDFs can, in addition, be a convenient way of condensing a vast amount of data usually obtained from ensemble predictions. Interannual variability in PDFs of monsoon rainfall and regional geopotential height probabilities is discussed.

Finally, the effect of the changes to the model formulation on seasonal predictability estimates is discussed.

1. Introduction

The scientific basis for extended-range atmospheric prediction derives principally from the predictability of the atmosphere's lower boundary conditions, particularly sea surface temperature (SST). However, even if SST could be predicted without error, the associated atmospheric evolution would not be uniquely determined, essentially because of the chaotic nature of atmospheric dynamics. As a result, the SST anomaly should be thought of as having a well-defined impact, not on a specific phase-space trajectory corresponding, say, to the atmosphere's evolution over one season, but on the phase-space geometry of the whole atmospheric climate attractor. This impact can be specified in terms of changes to the atmospheric probability distribution (or density) function (PDF) over atmospheric states (Palmer 1993, Kumar and Hoerling 1995).

In practice, estimating the impact of prescribed SST anomalies on such probability distributions can only be determined from ensembles of integrations of a dynamical atmospheric model. Examples of such seasonal ensemble integrations have been discussed recently by Branković *et al.* (1994), Palmer and Anderson (1994) and by Barnett (1995). In all these studies, the issue of what constitutes a reasonable lower bound on ensemble size was raised; this paper in part addresses the same issue. Stern and Miyakoda (1995) have also explored the feasibility of seasonal prediction from ensembles of 10-year long integrations.

In sections 2 and 3 some general notions on atmospheric predictability on seasonal time-scales are introduced and discussed, with emphasis on the ensemble approach to seasonal predictability. More specific results, based on the experimental experience at ECMWF, are discussed in the second part of the paper. In section 4, the organization of 9-member ensembles is described. The distribution of ensemble skill scores and ensemble consistency values are calculated on a regional basis (section 5). Based on statistical tests, we determine the extent to which the increase in ensemble size has increased our confidence in being able to estimate reliably the impact of the imposed SSTs on regional variables of practical interest, specifically precipitation and near-surface temperature (section 6). By extrapolation of the results of the statistical analyses, the likely impact of further increases of ensemble size is assessed. The impact of model formulation on some of the seasonal time-scale predictability estimates presented in this paper is discussed in section 7. An assessment of probability forecasts using the 9-member ensembles is made in section 8. Summary and conclusions are given in section 9.

2. Evidence of atmospheric seasonal predictability

Only a brief account on the evidence of seasonal time-scale predictability is given here, focusing primarily on the results of atmospheric general circulation models (AGCMs). For more details, the reader is referred to a review paper by Palmer and Anderson (1994).

Predictions beyond the average limit of deterministic predictability of synoptic-scale systems are possible because of the following two reasons. First, some components of the large-scale flow (or circulation types) are inherently more predictable than individual synoptic systems. In the medium-range, for example, the relatively successful prediction of blocking circulation patterns is nowadays well established fact (Tibaldi *et al.* 1995). Beyond the medium-range, there is some evidence that GCMs can reproduce relatively faithfully the statistics of the blocking occurrence. For the ECMWF model, Branković and Molteni (1995) have shown that in their seasonal integrations the frequency of occurrence of blocking events improved dramatically with the newest model version and has become comparable with observed frequency. The results from a multi-year integrations of various models with observed SSTs (the AMIP dataset) indicate some potential in simulating the frequency of blocking events (Fabio D'Andrea, University of Bologna, *personal communication*).

The second reason which makes predictions on longer time-scales possible is that slowly varying (and often predictable) lower boundary forcing can communicate a significant predictability on atmospheric development and influence the statistics of atmospheric circulation. This will be discussed in some details in the next two subsections. In addition to SSTs, boundary forcing may also be due to other variables, like soil moisture, snow cover, sea ice cover, land surface temperature and albedo.

For the short time-scales, in addition to internal dynamics, inadequacies in the definition of initial data significantly influence the error growth rates and affect the limit of predictability. On the longer time-scales, however, the effect of initial uncertainties is presumably less important. We shall assume that for seasonal time-scales the importance of lower boundary conditions outweighs the importance of uncertainties in initial conditions. Thus, using Lorenz's (1975) convention, predictions on seasonal time-scale are, as far as atmosphere is concerned, of the *second kind*, in contrast to those that are almost completely dependent on definition of initial conditions (predictions of the *first kind*).

2.1 Seasonal predictability and statistical properties of the atmospheric circulation

Molteni *et al.* (1993) and Palmer (1993) have demonstrated how the lower boundary forcing can transfer a significant predictability 'signal' to the atmosphere and influence statistical properties of atmospheric circulation. They presented a simple paradigm for interactions between an extratropical system and a tropical system. (The former is considered to be chaotic and the latter less chaotic or predictable system.) In particular, the well known Lorenz model (Lorenz 1963) was coupled to the so-called linear oscillator which provides an external forcing to the conventional Lorenz model. This oscillator is also called the tropical linear oscillator, because, in general terms, it can be assumed to represent or mimic the tropical ocean SST forcing in a coupled atmosphere-ocean system. The phase space estimates of probability density function (PDF) of the Lorenz coupled system are shown in Fig.1 for the three different cases. We present only a qualitative discussion of the Lorenz coupled model here.

When no forcing is applied in the coupled system (Fig.1a), the PDF shows a bimodal structure with almost equal amplitudes (given the sampling error). The regimes shown in Fig.1 can be thought of as equivalent to weather regimes or circulation types in the real atmosphere. The two most prominent circulation types in the middle latitudes are the blocked flow and the zonal flow. Fig.1a implies that in the absence of boundary forcing the two mutually exclusive regimes have equal probability of occurrence.

With a weak (tropical) forcing, the positioning of the two maxima in the phase space remains unchanged, but the magnitude of one maximum has decreased and the magnitude of the other maximum has increased (Fig.1b). It should be noted that a weak external forcing causes a weak relative amplitude change.

When a strong forcing is applied, the structure of the PDF is still bimodal, and the positioning of the two maxima has remained unchanged (Fig.1c). In other words, the spatial pattern of extratropical "weather regimes" is almost unaffected by the tropical forcing. However, the relative amplitude difference between the two maxima has become very large; in statistical sense, one of the two regimes is more likely to occur than the other.

The real atmosphere equivalent of the Lorenz model regimes are, for example, the positive and negative Pacific/North American (PNA) regimes. The PNA pattern can occur during almost any

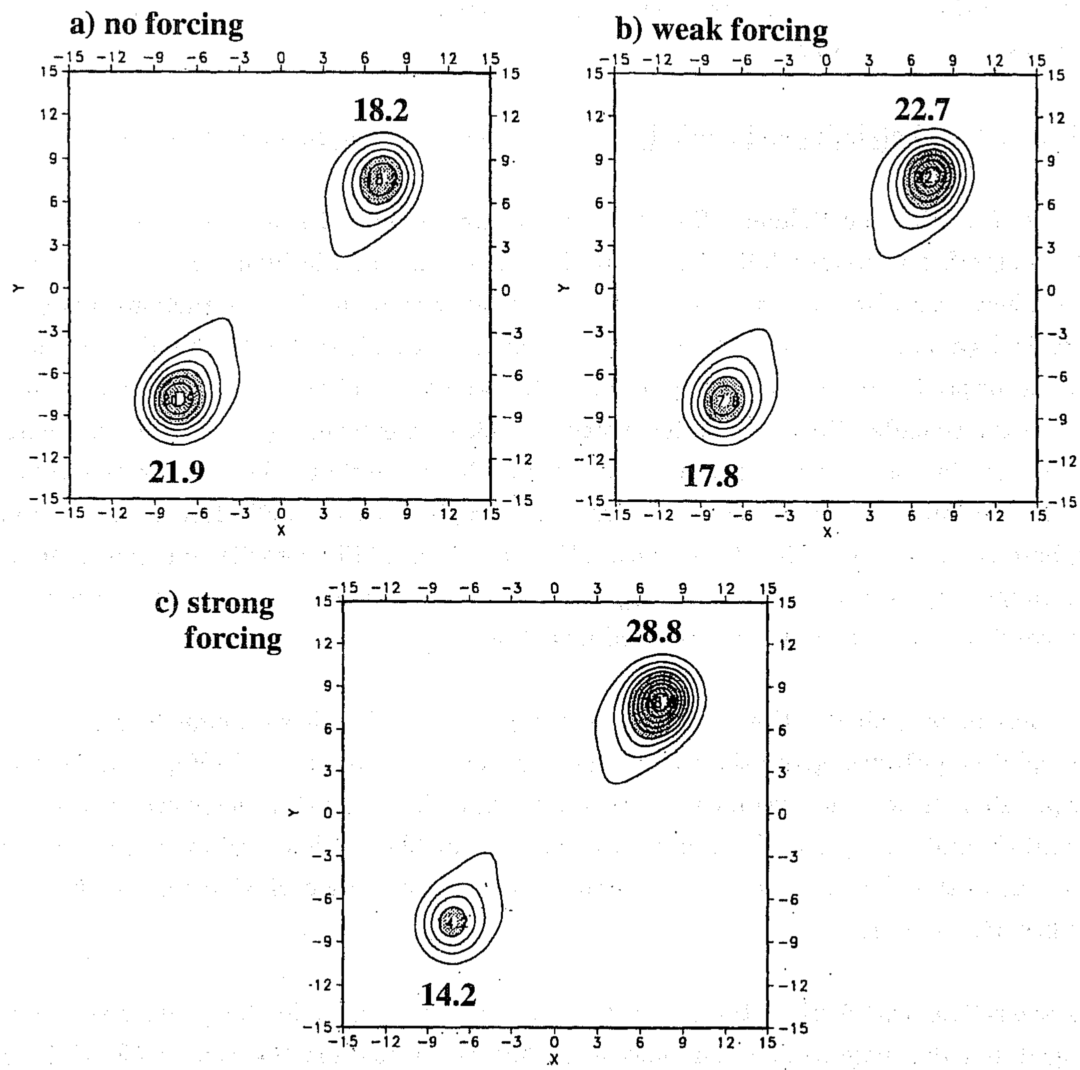

Fig.1 Probability density function (PDF) for the Lorenz model coupled to a tropical oscillatory system with: a) no forcing, b) weak tropical forcing and c) strong tropical forcing. (From Molteni *et al.* 1993)

winter, regardless of the Pacific SST forcing. However its frequency of occurrence is increased during the El Niño winters when a positive SST anomaly in the equatorial Pacific is observed (Yarnal and Diaz 1986).

2.2 Tropical versus extratropical seasonal predictability

(a) Tropics

Based on chaotic properties of the atmosphere we are able to clearly delineate the predictability problem between the tropics and extratropics. In the tropics, the internal variability of the large-scale flow is relatively weak and can explain only a fraction of seasonal time-scale fluctuations. Externally forced variability exceeds the level of internal variability and fluctuations in the large-scale flow are mainly due to variations in boundary conditions. The tropical oceans are known to be the main source of forcing for the tropical atmosphere. Luckily, the dominant mode of variability of the tropical oceans is predictable several seasons in advance (see Anderson 1996, this Volume, for the discussion on predictability of the tropical oceans).

Of course, day-to-day variations do exist in the tropics and they are associated with synoptic scale disturbances. Apart from the tropical cyclones, the tropical synoptic scale disturbances are generally weak and cannot give rise to the tropical seasonal variability. The tropical cyclones, though generally stronger than their extratropical counterparts, do not significantly contribute to the tropical seasonal variability because they are weakly non-linearly coupled to the tropical large-scale flow. They "interfere" with and alter the tropical large-scale circulation less strongly, and their prediction, in statistical sense, is relatively successful (see, for example, Gray *et al.* 1993).

The dependence of the tropical atmosphere variability on the land-surface forcing (soil moisture, snow cover, surface albedo) has also been studied by various authors. For example, Shukla *et al.* (1990) found that the Amazon deforestation has a significant impact on both regional and global climate. Rowell *et al.* (1995) found that the impact of soil moisture on summer rainfall amounts over the Sahel region in Africa is important, but in general the SST forcing is still dominant. The possible impact of extratropical boundary forcing on the tropical circulation is studied in numerous papers. Some of them, for example, focus on how the interannual variation in the Eurasian snow cover affects the south Asian monsoon circulation (see Zwiers, 1993 as an example of a relatively recent modelling study).

(b) Extratropics

In the extratropics, on the other hand, fluctuations due to the flow internal (inherent, nonlinear) instabilities are dominant and they will exist even under prescribed boundary forcing. However, these fluctuations are not completely independent, because of the coupling between flow instabilities and boundary anomalies.

The main characteristic of the extratropical flow is constant shuffle between various weather regimes and a relatively short transition periods between regimes. On seasonal time-scales, the timing of transitions is not predictable, however, estimates of prevailing regime(s) *are* predictable. As discussed in section 2.1, the rôle of SST anomalies is to influence (or bias) the statistics of weather regimes. This bias may not be large for some regions and the predictable signal may be indistinguishable from climatological variance; this undoubtedly presents the major obstacle to predictability in the extratropics. In terms of a GCM prediction, a single model realization will certainly be insufficient to capture the signal from the boundary forcing.

3. Ensemble approach to seasonal prediction

3.1 Probability density function and the concept of ensembles

In order to introduce the concept of a GCM ensemble for predictions on seasonal time-scales, we make use of the properties of PDFs. By definition, a PDF describes the probability that the random data will assume a value within some defined range at any instant of time (see, for example, Bendat and Piersol 1971). Following the discussion by Kumar and Hoerling (1995) we extend the notion of PDF to seasonal prediction.

The impact of interannual variation in boundary forcing (say SST) on PDF in the tropics is shown schematically in Fig.2a. The effect of the two different SST forcing (for example, El Niño and La Niña in the equatorial Pacific) is to make a clear separation between mean atmospheric states, i.e. the distance between the two corresponding PDFs is large. The internal variability, measured as the horizontal distance from the vertical, is small when compared with the distance between the two PDFs.

Therefore, for the tropics, a single GCM realization would normally be sufficient to capture the signal from interannual variation in boundary forcing. Of course, the "sufficient number" of model realizations very much depends on the amplitude of SST forcing - the weaker the SST forcing, the larger the number of model realizations may be required to capture such a signal.

In the extratropics (Fig.2b), on the other hand, internal variability is large due to chaotic properties of extratropical flow, and may dominate over the external forcing. This would imply that the horizontal distance within the PDF curve may be larger than the distance between the two PDFs. In view of our discussion in section 2.1, Fig.2c describes the extratropical PDFs more accurately than Fig.2b if the extratropical PDF is characterized by the bimodal structure.

In the case of Fig.2b (or Fig.2c), a single model realization may fall, for example, between medians of the two PDFs (depicted by vertical lines). In this case, it is not possible to affirm to which half the given realization belongs to. In other words, for the extratropics, we need more than one model integration to extract the effect of the SST-forced interannual variation on regime statistics. Multiple model realizations are normally referred to as an *ensemble* of model integrations. In addition, an ensemble of forecasts gives a *probabilistic* estimate of the likely occurrence of specific

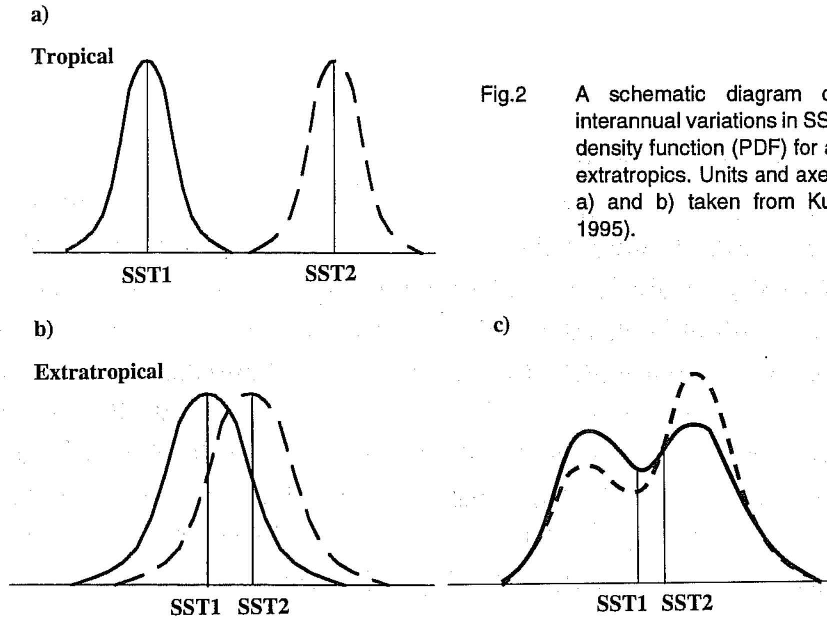

Fig.2 A schematic diagram of the impact of interannual variations in SST on the probability density function (PDF) for a) tropics, b) and c) extratropics. Units and axes arbitrary. (Panels a) and b) taken from Kumar and Hoerling 1995).

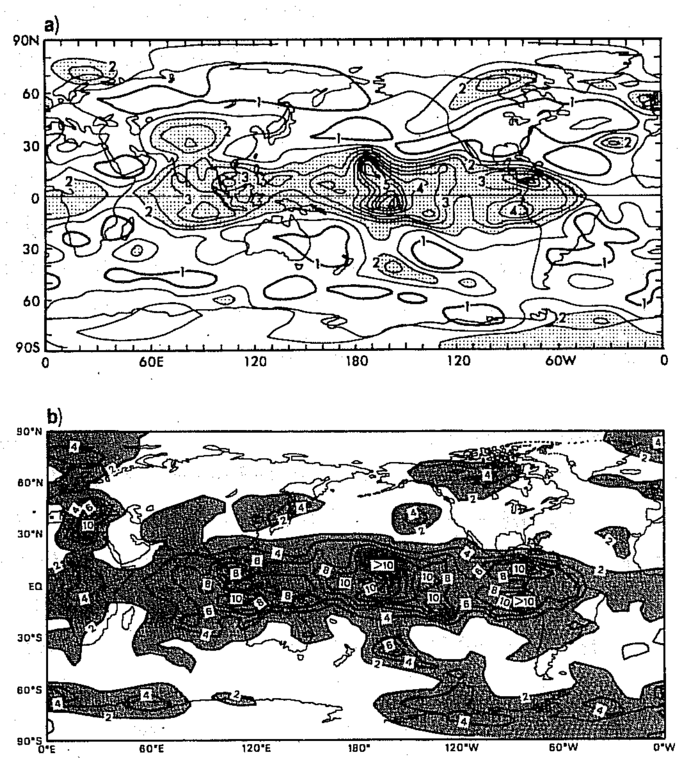

Fig.3 Distributions of the ratio of variances of 200 mb geopotential height between the perturbed and control runs. a) monthly mean values for December-February, b) seasonal mean values for december to February. Shading indicates values of the ratio larger than two. (From Palmer 1987)

weather regime(s) over a season. The second part of this paper partly addresses important question on how many integrations or how large ensembles are required to obtain reliable signal from the boundary forcing.

3.2 Time averaging

Time averaging is an additional important factor that also contributes to seasonal predictability and is unrelated to the physical origins of predictability mentioned above. Shukla (1981) has argued that the limit of predictability is closely related to the variance of the quantity we want to predict. For example, daily variations, which are mainly due to internal atmospheric dynamics, set a relatively low upper limit of predictability than variations of monthly or seasonal averages. In other words, for a longer time-scales, time averaging of atmospheric variables filters out contributions from a smaller temporal (and spatial) scales and extends the limit of predictability. The time averaging essentially increases the proportion of variability due to boundary forcing. Based on variance analysis from a very long integrations with the GFDL model, Palmer (1987) has convincingly demonstrated that predictability of seasonal averages increased substantially over predictability of monthly averages, both in the tropics and extratropics (Fig.3).

4. Seasonal predictability experimentation at ECMWF

Most of the discussion below is identical to that in Branković and Palmer (1996). As in Branković *et al.* (1994; hereafter referred as BPF), all integrations described in the second part of this paper were made with the ECMWF model at the reduced horizontal resolution of T63L19, using the so-called cycle 36 physics package (Simmons *et al.* 1988, Miller *et al.* 1992). The integrations were about 120 days long, depending on season and initial date. They cover all seasons over the 5-year period, from spring (MAM) 1986 to winter (DJF) 1990/91. The nine initial dates for each calendar season are shown in Table 1. They were chosen around the first day of the month preceding the season of interest. Thus, the range of the BPF initial dates is extended to include four days before and two days after the dates of the original 3-member ensembles.

Within a given season, the same SSTs (based on the U.S. National Meteorological Center analyses), were used for each ensemble member and were updated every 5 days throughout the integration. For every calendar season, the interannual variation of the SST anomalies for the period spring 1986 to winter 1990/91 includes both warm and cold El Niño-Southern Oscillation (ENSO) events. As discussed in BPF, we categorise seasons according to whether anomalies in an equatorial Pacific SST index, computed over the tropical Pacific strip (7°N-7°S, 160°E-80°W), were either strong and positive, strong and negative, weak and positive or weak and negative (see Figs 1 and 2 of BPF). In the rest of this paper we shall be discussing differences between pairs of ensemble integrations, each ensemble having been made using SSTs for a specific year. Where results are described as associated with "strong ENSO-index years" one ensemble was made with SSTs where the Pacific index had strong positive anomalies; the other ensemble was made with SSTs where the index had strong negative anomalies. Similarly, for "weak ENSO-index years" one ensemble was made using

Table 1 Initial dates for ECMWF seasonal simulations.

Ensemble member	DJF	MAM	JJA	SON
1	28 October	27 January	27 April	28 July
2	29 October	28 January	28 April	29 July
3	30 October	29 January	29 April	30 July
4	31 October	30 January	30 April	31 July
5	1 November	31 January	1 May	1 August
6	2 November	1 February	2 May	2 August
7	3 November	2 February	3 May	3 august
8	4 November	3 February	4 May	4 August
9	5 November	4 February	5 May	5 August

Table 2 Division of the experimental years/seasons according to the index based on the equatorial Pacific SST anomalies.

Amplitude of ENSO	Season	Positive	Negative
Strong	DJF	1986/87	1988/89
	MAM	1987	1989
	JJA	1987	1988
	SON	1987	1988
Weak	DJF	1990/91	1989/90
	MAM	1988	1986
	JJA	1990	1989
	SON	1990	1989

SSTs where the index had weak positive anomalies; the other ensemble was associated with an index having weak negative anomalies. In practice, the term "weak ENSO-index years" refers to years in which significant El Niño SST anomalies were absent. Table 2 shows how each season can be characterized with this index.

Throughout the paper we focus our discussion on seasonal averages only. For each experiment, the last three months, corresponding to conventional calendar seasons, were averaged. The seasons are denoted conventionally as: spring - MAM, summer - JJA, autumn - SON and winter - DJF. For verification purposes, seasonal averages were also computed from ECMWF analysis data.

5. Objective verification of ensembles

5.1 Skill scores of anomaly fields

In this section, model skill scores are given in terms of anomaly correlation coefficients (ACCs) between seasonally averaged observed anomaly fields and model anomaly fields. For a given season, the observed anomalies have been computed with respect to the 5-year average (1986 to 1990); we refer to this average as the 'observed climate'. For the model, anomalies have been computed with respect to the two different mean fields. First, we used the 'observed climate' as above. This methodology is consistent with the ECMWF operational practice for determining the skill of medium-range forecasts. (Of course, in the ECMWF operations the climate is derived from a much longer period than used here.) In addition to this, model anomalies were also computed with respect to the mean from all integrations in the same 5-year period, i.e. from the total of 45 runs. We refer to this average as the 'model climate'. For reason of space, the discussion in this sub-section will be restricted to ACCs of the 500 mb heights in the three northern hemisphere regions (NH1, NH2, NH3; see Fig.4) for the DJF season only.

The top row in Fig.5 shows ACCs when the 'observed climate' is used to calculate model anomalies; the bottom row of Fig.5 shows ACCs when the 'model climate' is used. The ACCs are shown for all individual integrations within an ensemble (depicted for each year as 9 small crosses) and for ensemble averages (larger diagonal crosses). The distribution of small crosses in the vertical is indicative of the intra-ensemble range of scores, while the small shift between the crosses in the horizontal represents the different initial dates for individual model integrations.

When the 'observed climate' is used, the highest scores are found for the NH2 region (covering the northern Pacific and much of North America) during DJF 1988/89. The range of scores for this winter appears to be the smallest of all winters considered. According to Table 2, the 1988/89 winter was classified as a strong negative ENSO-index season. Skill scores for DJF 1986/87, the strong positive ENSO-index winter, do not differ very much from, for example, skill scores for DJF 1987/88.

On the other hand when the 'model climate' is used (Fig.5, bottom row), the highest scores for the

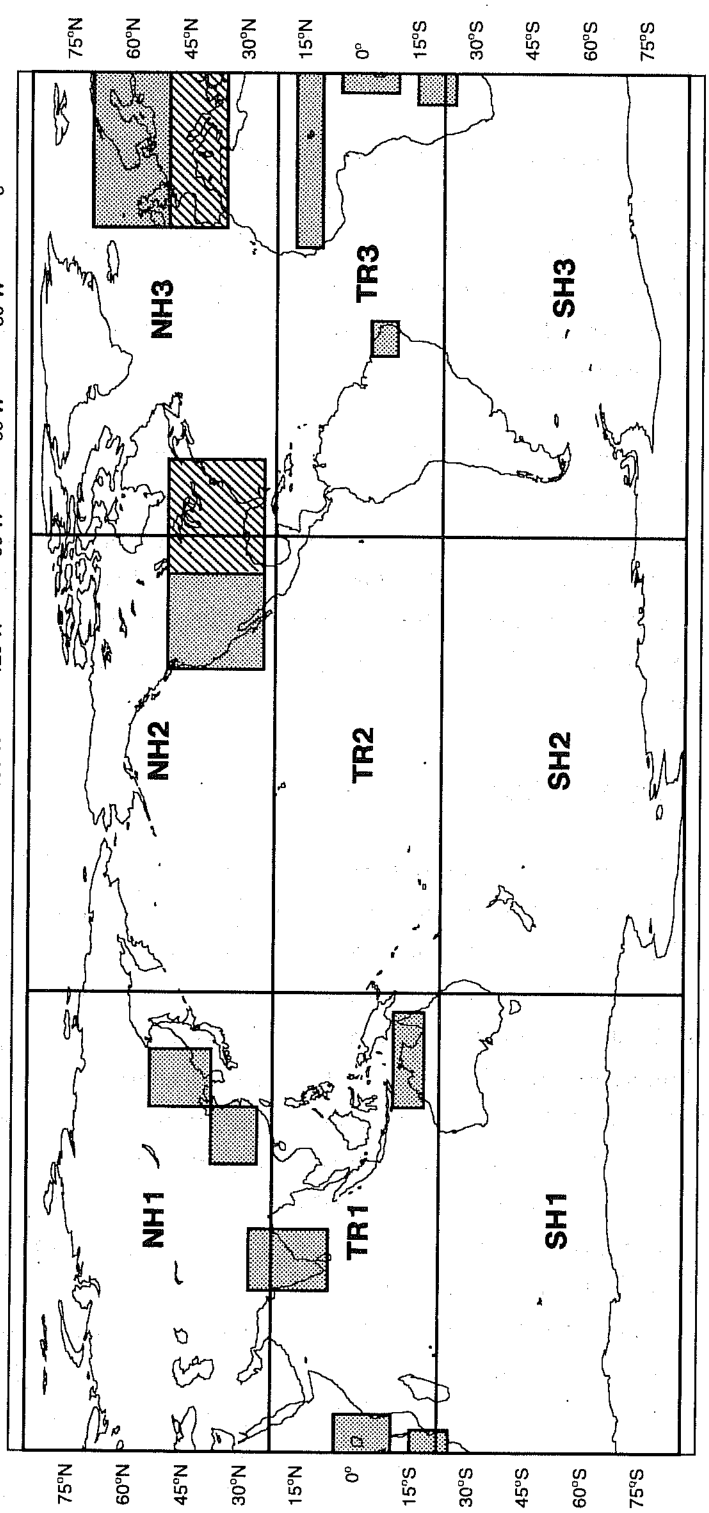

Fig.4 Nine regions (NH1, NH2, ...) for which distribution of skill scores and ensemble consistency were computed. Regions for which t-test statistic was calculated are shaded. (The eastern Africa and northern Kalahari regions are broken along 30°E.)

NH2 region are found during DJF 1986/87. For all but one integration, the ACCs for that season fall between 0.6 and 0.9, and the range of scores is small when compared with the other winters.

The difference between the two sets of skill scores for NH2 during DJF 1986/87 is striking. The differences for DJF 1988/89 and for the weak-ENSO index winters are somewhat smaller, though by no means negligible. Similar differences in skill scores are also seen for the NH3 (Asian) and for NH3 (north Atlantic/European) regions.

Clearly, these inconsistencies are associated with the different reference fields used to calculated model anomalies. Apparent improvement in the model skill in DJF 1986/87, when the 'model climate' is used, may be because of a larger covariance between model and observed anomalies than that obtained when the 'observed climate' is used. Such an increase occurs because the 1986-1990 'model climate' projects more strongly on the La Niña than on the El Niño flow pattern over the northern Pacific/North American region.

These results highlight the dependence of the skill scores, and in particular the interannual variability of the skill scores, on the choice of a reference climate. Even using an observed climate, the scores will be arbitrary to some degree, being dependent on which years are chosen to form the climate fields. Because of this, results are shown for most of the body of this paper, not in terms of the skill of anomalies, but rather in terms of the skill of differences between pairs of chosen years (e.g. between an El Niño year and a La Niña year). For such measures, we do not need to refer to any sample climatology.

5.2 Skill-score distributions of difference fields

The ensemble distribution of skill scores was estimated by comparing simulated and observed seasonal-mean difference fields for the pairs of years shown in Table 2. Specifically, for a given season let $E^1=\{e_i^1\}$ and $E^2=\{e_i^2\}$ denote two ensembles, the first for a year taken from the third column of Table 2 (denoted "positive"), the second corresponding to the year shown on the same row in the fourth column of Table 2 (denoted "negative"). The differences $(e_i^1-e_j^2)$ are then correlated with the corresponding observed difference field, O^1-O^2, for all combinations of subscripts i,j, where i and j run from 1 to N, and N denotes the ensemble size. For two 9-member ensembles, there are 81 such difference fields $(e_i^1-e_j^2)$ with 81 corresponding correlation coefficients. The distribution of relative frequencies of these correlation coefficients is then computed by binning them into categories of equal correlation intervals of 0.2. The distributions have been computed corresponding to all pairs of years from the rows of Table 2, and for 9 regions which together cover the globe (see Fig.4). For the six extratropical regions (three in the northern hemisphere, three in the southern hemisphere), the skill scores are shown for the 500 mb height field. For the three tropical regions, they are shown for the 200 mb zonal wind field (geopotential height being rather featureless in the tropics).

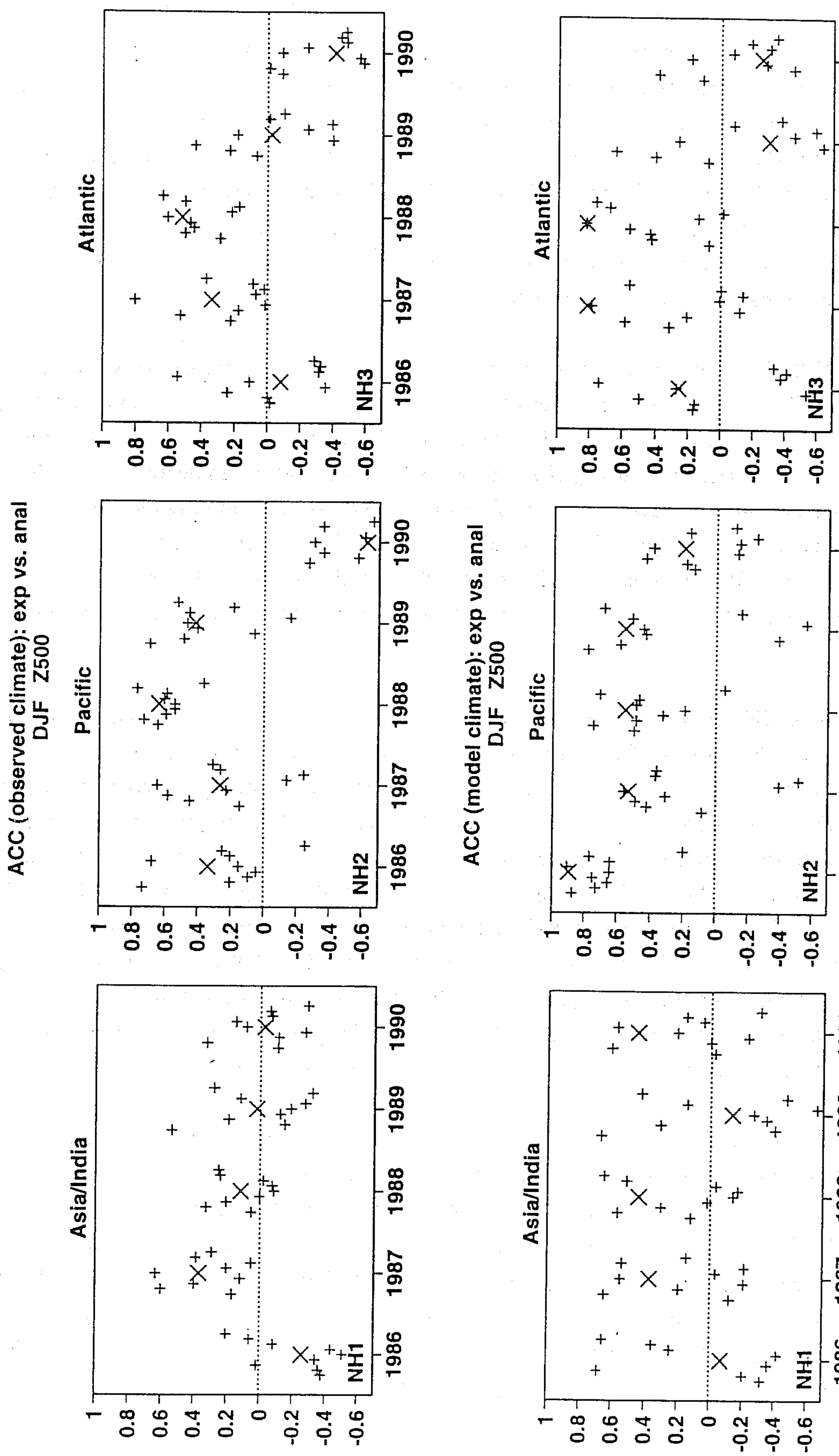

Fig.5 The DJF 500 mb anomaly correlation coefficients for the three northern hemisphere regions (NH1, NH2, NH3): when the 'observed climate' is used to calculate model anomalies (top), and when the 'model climate' is used (bottom).

(a) Strong ENSO-index years

Fig.6 shows the distribution of skill scores for strong ENSO-index years for DJF over the 9 regions. Categories between -1 and +1 are depicted on the x-axis; on the y-axis the relative frequency of skill scores (in %) in each category is shown. Compared with the extratropics, the distributions are strongly peaked in the tropical regions. This is consistent with the relatively chaotic nature of the extratropics (e.g. Charney and Shukla, 1981; Palmer *et al.*, 1995). In the region TR2, covering much of the tropical Pacific, the distribution is peaked towards the most skilful category, whilst in the other two tropical regions (TR1 and TR3), the distributions are strongly peaked towards the second most skilful category. Such a well-defined shift in the skill scores in the regions TR1 and TR3 is presumably associated with model error. Overall, these results are consistent with the fact that the dominant signal is associated with ENSO itself.

For the region NH2, the distribution is clearly peaked towards the two categories with largest correlation values. However, for other northern hemisphere regions, further away from the ENSO signal, the distribution of skill scores is both broader and shifted towards less skilful categories. Nevertheless, the distributions are clearly skewed towards positive values, indicative of an overall level of skill. The distributions are also broad in the southern hemisphere, and, as for the northern hemisphere, the most skilful region (SH2) lies closest to the ENSO region. The shift towards negative skill values in SH3 may again be indicative of the influence of model error, but it also may be related to the quality of ECMWF verifying analysis for the years in question in these data sparse areas.

The ensemble-mean difference skill scores have also been categorised in the same way as individual members' differences. They are shown as small vertical arrows in Fig.6 (and in subsequent skill score figures), pointing to the category they fall in. Generally, if the distribution is skewed strongly towards positive values, there is a tendency for the ensemble-mean score to either fall within the same category as distribution's peak value, or to reside in the adjacent higher category of the peak value. This reflects the fact that the spatial variance of individual runs is larger than the spatial variance of the ensemble mean field and, consequently, ensemble averaging will increase the value of the ACC (Branković *et al.* 1990). If the distribution is skewed strongly towards negative values (as found for some weak-ENSO cases, see discussion and figures below), the ensemble-mean score may again coincide with distribution's peak value or fall in the adjacent lower category (here, ensemble averaging can make poor scores even worse!).

Fig.7 shows the skill score distributions in the three northern extratropical regions for the strong ENSO-index years, for the three remaining seasons, MAM, JJA and SON. It is interesting to note that whilst in NH2 (the northern Pacific), the distributions are skewed to the most skilful categories in DJF (Fig.6), in NH1 and NH3, the distributions are most skilful in MAM. It is not clear at present whether this is associated with the additional influence of more local (extratropical) lower boundary forcing anomalies in this season. The relatively high levels of skill around the northern hemisphere in spring may have some important practical consequences in the application of seasonal forecasts to agricultural production.

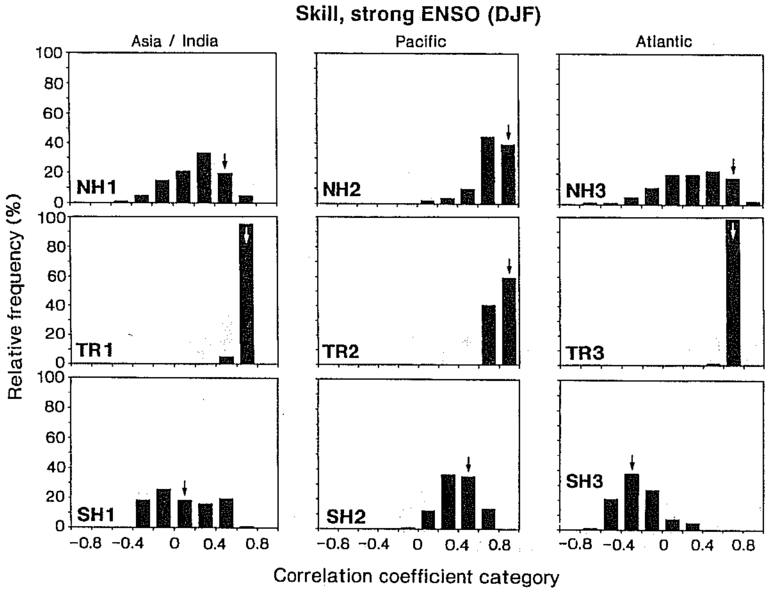

Fig.6 Distributions of relative frequencies (%) of the difference skill scores over the nine regions of the globe (Fig.1) for the strong ENSO-index DJF season. In the extratropics (top and bottom panels) correlation coefficients are computed for 500 mb height differences, in the tropics (centre row panels) for 200 mb zonal wind differences. Vertical arrows point to category of ensemble-mean difference skill score.

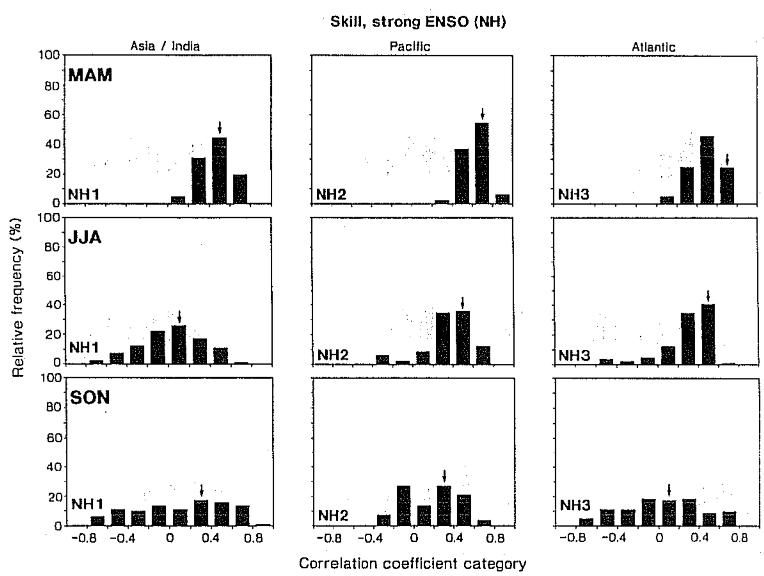

Fig.7 Same as Fig.6 but for the strong ENSO-index MAM (top row), JJA (middle) and SON (bottom) seasons over the northern hemisphere regions (NH1, NH2, NH3) only.

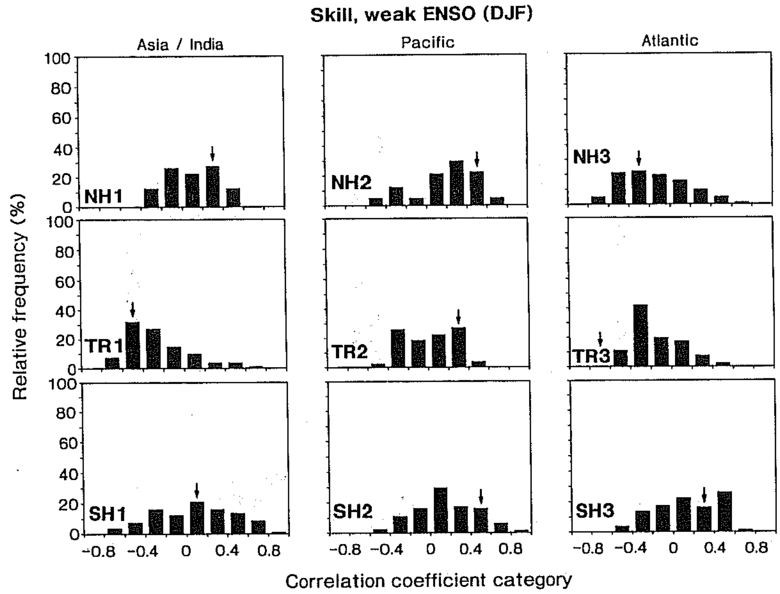

Fig.8 Same as Fig.6 but for the weak ENSO-index DJF season.

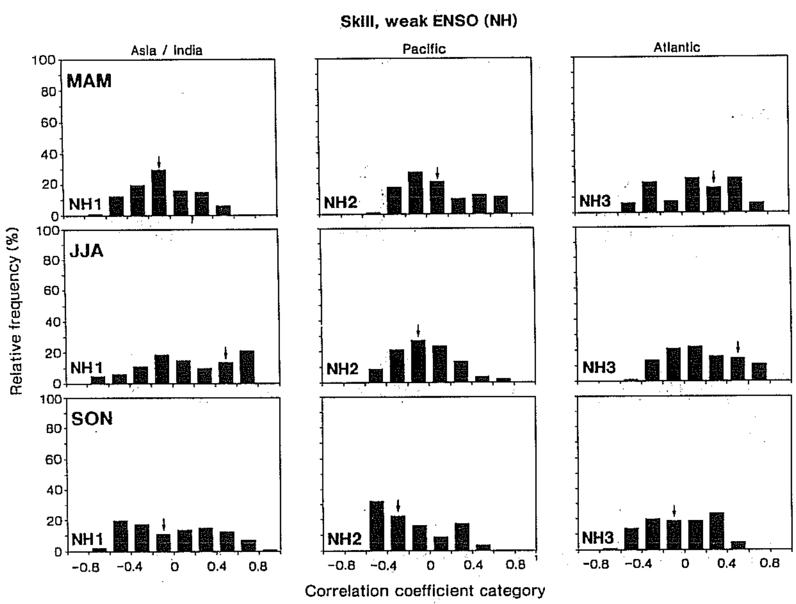

Fig.9 Same as Fig.7 but for the weak ENSO-index MAM, JJA and SON seasons.

In general, it can be seen that the broadest distributions of skill scores occurs in SON. The distributions for NH1 and NH3 in SON are a particularly striking illustration of the nature of internal chaotic variability in the atmosphere. Within the ensemble difference fields, there are (pairs of) members with skill scores exceeding 0.8, and yet others with skill scores between -0.6 and -0.8.

(b) Weak ENSO-index years

Fig.8 shows the DJF skill score distributions for the weak ENSO-index years for all 9 regions. It can be seen that for these years, comparing with Fig.6, there is no clear-cut difference in the distributions in the tropics and extratropics. In general it is clear that these weak-ENSO years are associated with much weaker levels of skill. In the tropics, both TR1 and TR3 are strongly shifted towards negative skill scores. Once more, this would appear to be associated with the impact of model error. It is interesting to note that despite a relatively broad distribution, the NH2 region still shows a shift towards positive skill values.

Fig.9 shows the northern hemisphere distributions for the other three seasons for the weak ENSO-index years (these can be compared with Fig.7). Although the distributions are broad, there is some evidence of a shift towards positive skill values, especially in the Atlantic (NH3) region in MAM and JJA. It is possible (see subsection *(a)* above), that lower boundary forcing, local to NH3, may be having an influence on the predictability in that region.

5.3 Consistency of ensembles

As mentioned a number of times above, distributions of skill scores are influenced by model error. In order to study the impact of SST variations on distributions of "internal" ensemble differences, we again consider two ensembles $\{e_i^1\}$ and $\{e_i^2\}$ from the third and fourth columns (respectively) of Table 2. For a given difference field, e_i^1-e_j^2 (i-th element from E^1 and j-th element from E^2), we calculate N×N-1 correlation coefficients between this field and all possible pairs of difference fields, e_k^1-e_l^2, where k,l=1,N. This calculation is performed N×N times for all different correlations $C(e_i^1-e_j^2, e_k^1-e_l^2)$. For 9-member ensembles, we obtain 80×81 correlation coefficients. As for the skill scores, correlation coefficients measuring consistency have been calculated for the 500 mb height in the extratropical regions and for the 200 mb zonal wind in the tropics. The distribution of relative frequencies of correlation coefficients is estimated, as with skill scores, by binning into equal categories of 0.2.

Although not shown, in strong ENSO-index years, the consistency between ensembles is very high in the tropics, regardless of season. Generally speaking, values are strongly peaked in the highest correlation category. In the extratropics, the distributions are much broader. Fig.10 shows the distributions for the northern hemisphere regions for all four seasons. (The meaning of the graph axes is the same as in Figs. 6 to 9.) Interestingly, it can be seen that for all regions, MAM has the highest consistency. We noted above that MAM was the most skilful for NH1 and NH3. For the Atlantic sector (NH3), both MAM and JJA are more consistent than DJF.

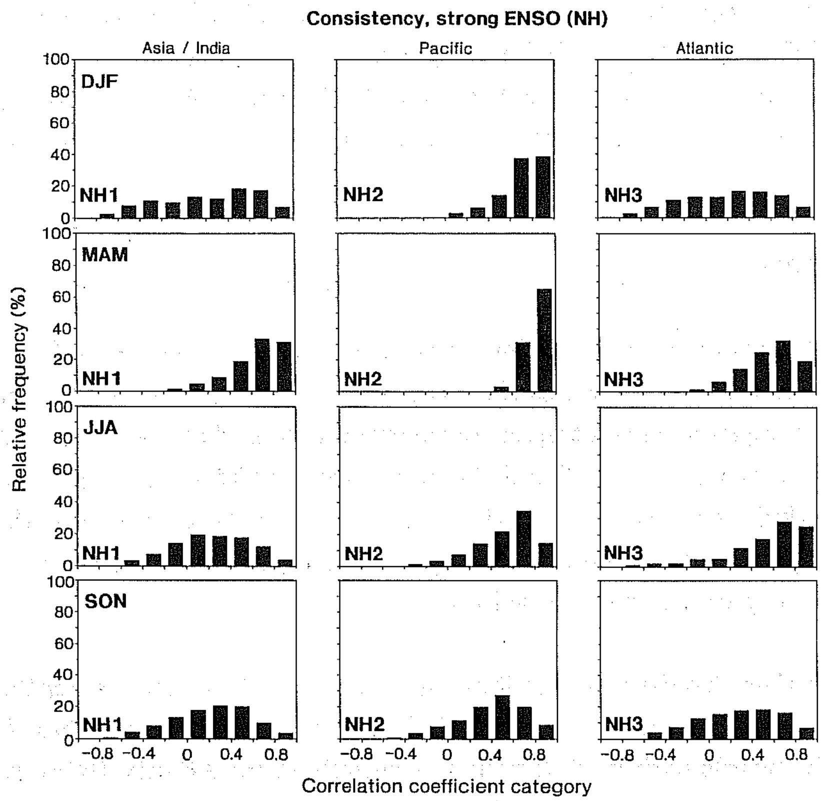

Fig.10 Distributions of relative frequencies (%) of the 500 mb height consistency correlation coefficients over the northern hemisphere regions (NH1, NH2, NH3) for the strong ENSO-index DJF (top row panels), MAM (second row), JJA (third row) and SON (bottom row) seasons.

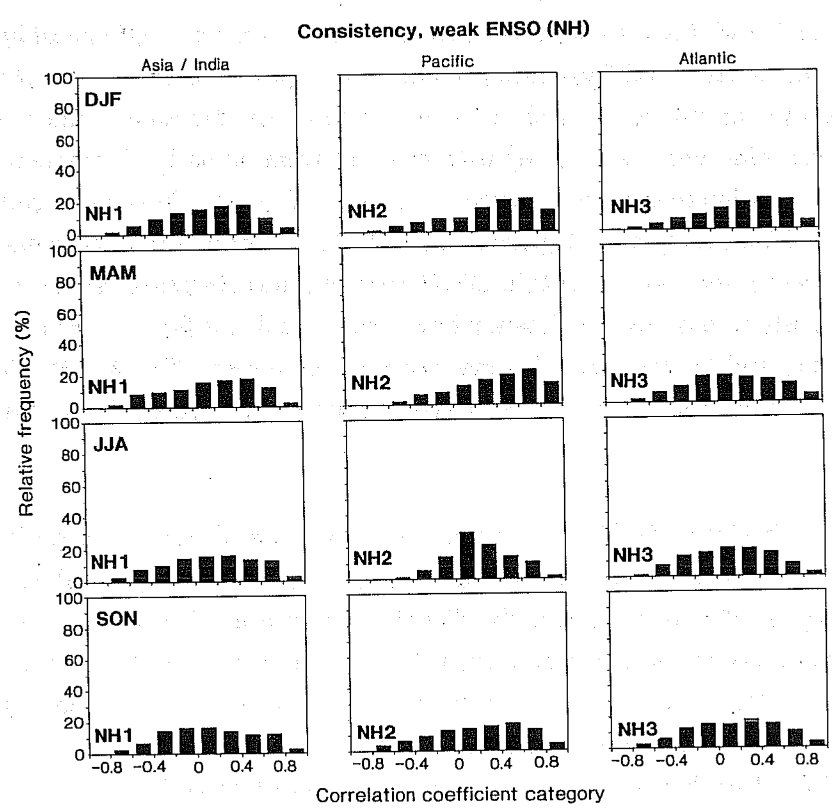

Fig.11 Same as Fig.10 but for the weak ENSO-index DJF, MAM, JJA and SON seasons.

For the weak ENSO-index years (Fig.11), there is a clear shift towards positive correlation values, however, the overall distributions are much broader than for the strong ENSO-index years. The DJF seasons shows a stronger signal than JJA and SON, and in the NH2 area the shift towards positive values is stronger than in the other northern hemisphere regions.

If we compare consistency distributions with skill score distributions, it can be seen that there are some clear similarities. For example, both sets of distributions are broader for the weak ENSO-index years. Moreover, the NH2 has the tightest distributions for both consistency and skill. In general, the MAM season is seen as the most consistent and most skilful season at least for the strong ENSO-index years. In NH3, both MAM and JJA are consistent and skilful for strong ENSO-index years.

6. Potential predictability and ensemble size

In this section we focus on confidence values associated with the SST forcing of precipitation and near-surface temperature (2m post-processed temperature) as a function of ensemble size. These estimates are made for a number of pre-specified land subregions, shown in Fig.4 as shaded squares. (These have been chosen as representative examples from a much larger set of regions for which calculations have been made.) The chosen regions include some tropical areas where predictability may be expected to be high, together with extratropical areas where internal atmospheric variability may be expected to obscure, at least partially, the influence of lower boundary forcing.

The method of analysis is as follows. For each ensemble, there is a unique 9-member ensemble-mean value for the regionally-averaged precipitation or 2m temperature. On the other hand, there are 84 possible values for a 3-member subensemble-mean of the same precipitation or 2m temperature from the original 9-member ensemble (corresponding to the number of ways of choosing a 3-member subset from a 9-member set). The number of possible values for a 4-, 5-, 6-, 7- or 8-member subensemble is 126, 126, 84, 36 and 9 respectively. For each n-member subensemble ($3 \leq n \leq 9$), we calculate the subensemble-mean difference between years in the third and fourth columns in Table 2, and the corresponding t-statistic based on the null hypothesis H_0 that the subensembles are not significantly different. We then calculate a mean t-statistic by averaging over all possible subensemble t-values. In performing this averaging, the sign of the t-variable is ignored, because the sense of the interannual variation has no relevance to the discussion here.

For reference, in Figs. 12 to 14 the t-value corresponding to the rejection of H_0 at the 90% and 99% confidence levels is shown. We comment on the minimum ensemble sizes required to exceed these t-values, for different seasons, regions and magnitude of the ENSO index. To aid the discussion below, we have also remarked on likely t-values for hypothetical larger ensembles based on an extrapolation of the computed values. In all diagrams we note an almost linear relationship between ensemble size and the t-value. This is due to a linear increase in the number of degrees

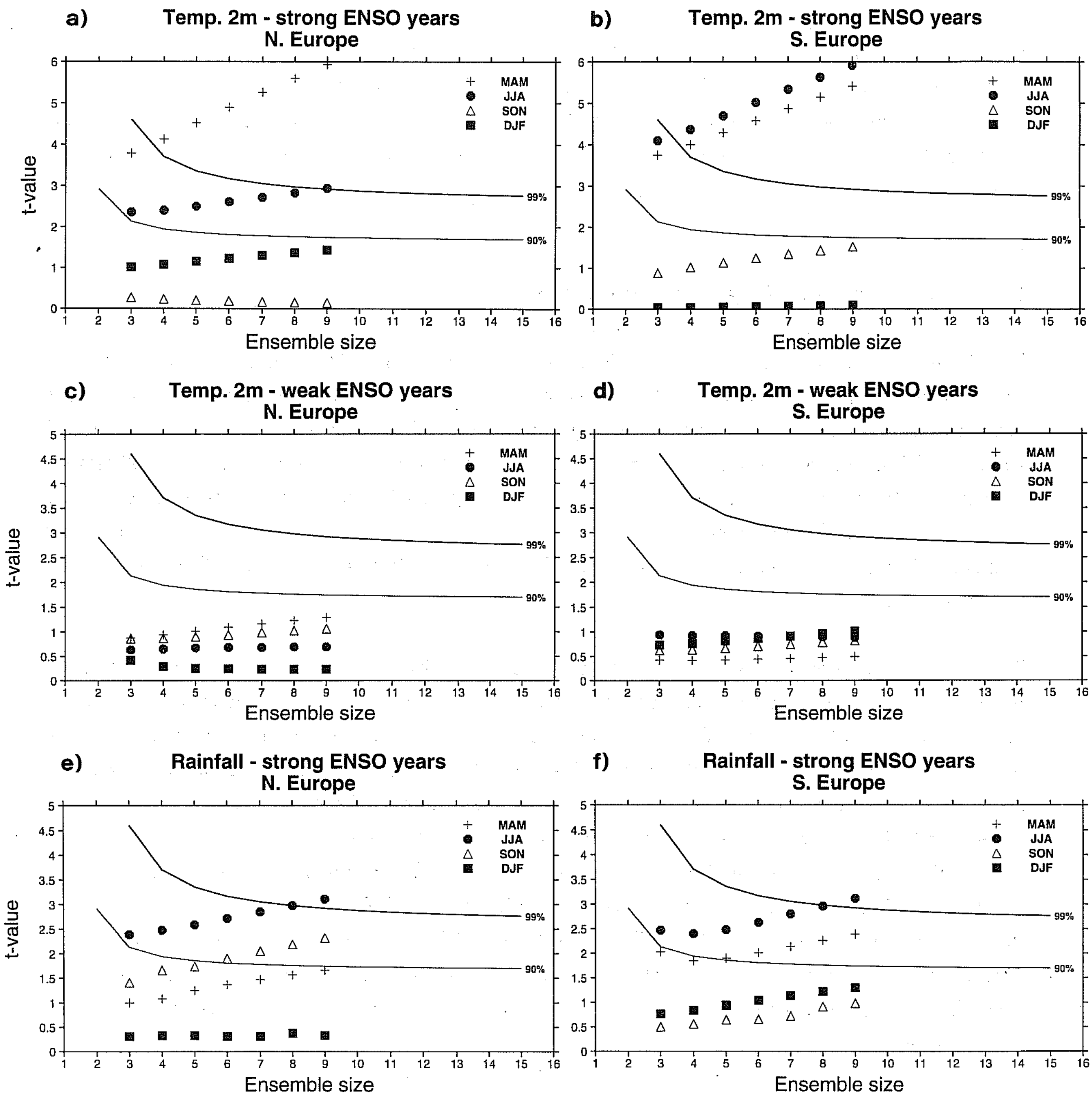

Fig.12 The dependence of t-values on ensemble size over the northern Europe (left) and southern Europe (right). a) and b) 2m temperature, strong ENSO-index years, c) and d) 2m temperature, weak ENSO-index years, e) and f) rainfall, strong ENSO-index years.

of freedom in calculating t-values as the size of the ensemble increases.

6.1 Extratropics

(a) Europe

Fig.12 a-d, shows t-values for 2m temperature in the northern and southern European regions. For strong ENSO-index years (Fig.12 a,b) it can be seen that MAM and JJA are the most predictable seasons. These results are broadly in agreement with the NH3 500 mb height consistency distributions discussed in the previous section. For MAM, it appears that only 4-member ensembles are required to reject H_0 with 99% confidence level. For SON and DJF, large ensemble sizes appear necessary to distinguish the two years.

In the weak ENSO-index years (Fig.12 c,d), only the MAM season in northern Europe seems to become predictable when the ensemble size increases significantly over 9 members (again in agreement with NH3 height consistency diagrams). For this season and region, H_0 would appear to be rejected with 90% confidence, with approximately 20-member ensemble.

Fig.12 e,f shows the t-values for rainfall for the northern and southern European regions for strong ENSO-index years. Consistent with the 2m temperature results, MAM and JJA show evidence of potential predictability. However, for this variable, larger ensembles are required to reject H_0 with 99% confidence (in excess of 16 members for the northern Europe spring rainfall). The t-values for weak ENSO-index rainfall are not shown, but are generally smaller than those found in the strong ENSO-index years.

(b) USA

For reason of space, predictability estimates for the USA are shown in Fig.13 for the strong ENSO-index years only. The near-surface temperature appears to be fairly predictable in general (Fig.13 a,b). Apart from JJA in the western USA, all seasons reach 90% confidence level of predictability with 9-member ensembles.

By contrast, during the weak ENSO-index years (not shown), significant predictability of 2m temperature is found for the western USA only in summer, and in the eastern USA for winter.

For both the western and eastern USA rainfall (Fig.13 c,d), there is a dramatic seasonal cycle effect for the strong ENSO-index years, with spring (MAM) showing much more significant values than for other seasons. For other seasons, there is relatively little difference between strong and weak ENSO-index years (not shown).

6.2 Tropics

To assess the relationship between predictability and ensemble size in the tropics we discuss the

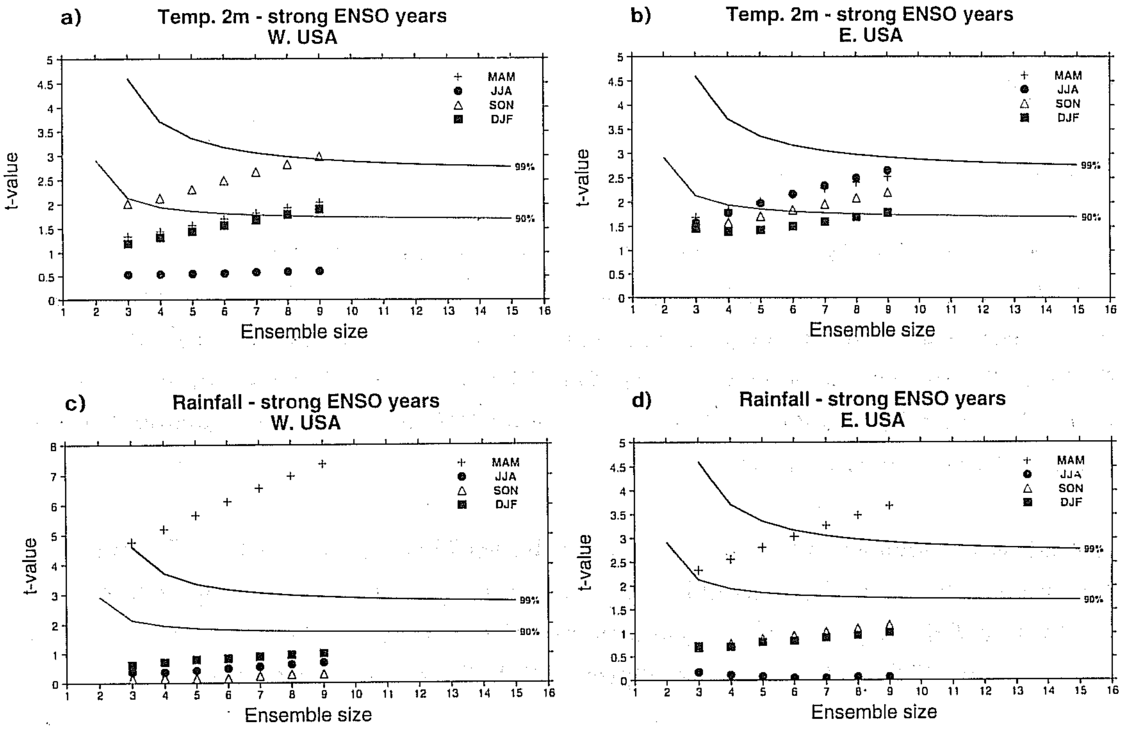

Fig.13 Same as Fig.12 but for the western USA (left) and eastern USA (right) and for the strong ENSO-index years only. a) and b) 2m temperature, c) and d) rainfall.

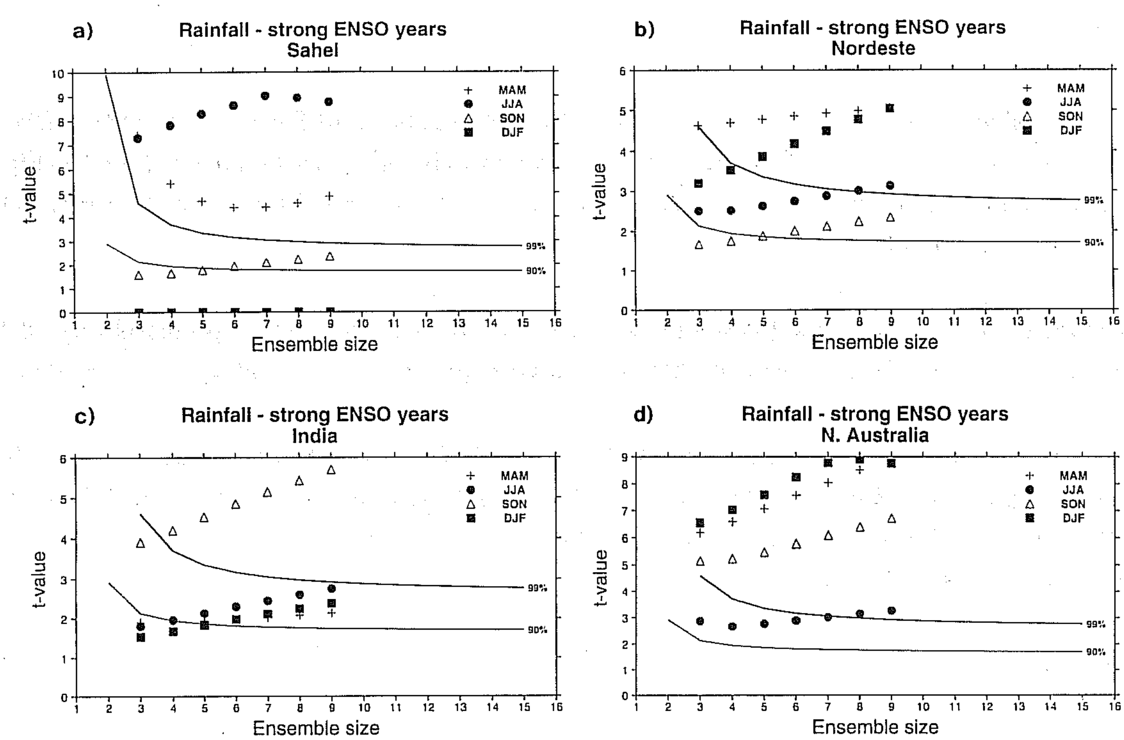

Fig.14 Same as Fig.12 but for selected tropical regions, rainfall and strong ENSO-index years only.

t-value for the rainfall in four different regions (Sahel, India, Brazilian Nordeste and the northern Australia; see Fig.4 for the regions' boundaries) during the strong ENSO-index years (Fig.14). As discussed by Charney and Shukla (1981), tropical rainfall should be intrinsically more predictable than extratropical rainfall. Clearly, as inferred from Fig.14, in the tropics relatively small-size ensembles are required to reach significant levels of predictability. Note the variable range of t-values on the ordinate axes in Fig.14. (In the Sahel during DJF there was no rainfall, hence all t-values are zero.) The high level of skill over the Sahel for JJA and over the Nordeste for MAM is consistent with results from earlier studies (e.g. Rowell *et al.* 1995, Ward and Folland 1991). The Indian region has the highest proportion of t-values below the 99% (and 90%) confidence level, indicating relatively lower predictability than in the other tropical regions shown. This could be associated with a model tendency to underestimate the overall level of the Indian rainfall, as shown in Branković and Palmer (1994), or with the influence of quasi-chaotic intraseasonal monsoon variations (Palmer 1994).

For the weak ENSO-index years (not shown), it can be noted that rather high levels of predictability were found for Nordeste and north Australian rainfall, while for the Sahel and India a noticeable deterioration in predictability for most seasons was found.

6.3 Summary for all regions

We have summarised the results for all seasons in Table 3. This shows the number of regions in the tropics or the extratropics for which the rainfall or the near-surface temperature differences are statistically significant (i.e. H_0 is rejected) at the 90% confidence level, as a function of ensemble size. The regions in question are shown as shaded areas in Fig.4. There are 6 regions in the extratropics (N. Europe, S. Europe, W. USA, E. USA, NE. China and central China) and 6 in the tropics (the Sahel, E. Africa, N. Kalahari, India, N.Australia and the Brazilian Nordeste). Table 3a is for the strong ENSO-index years, Table 3b is for the weak ENSO-index years.

In the strong ENSO-index years (Table 3a), only 9 out of 48 possible regionally-averaged extratropical rainfall or temperature values are significant with 3-member ensembles[1]. This increases to 22 with a 9-member ensemble and a projected 32 (out of 48) with a 16-member ensemble. Hence for extratropical prediction, the use of relatively large ensembles can be justified. For the tropical regions, the number of significant values increases from 31 to 40 to a projected 42 (out of 48) as the ensemble size increases from 3 to 9 to 16.

On the other hand, for the weak ENSO index years (Table 3b), even tropical predictions appear to benefit from large ensembles, where the number of significant values increases from 15 to 24 to a projected 32 as the ensemble size increases from 3 to 9 to 16. For these weak ENSO-index years, a large ensemble benefits detection of extratropical predictability as well.

---

[1] The number 48 is obtained when the total number of extratropical regions considered (6) is multiplied by the number of seasons (4) and by the number of variables (2). The same is valid for the tropics.

Table 3a Number of regions in the tropics and extratropics (depicted as shaded areas in Fig.4) for which interannual differences for either rainfall or 2m temperature are statistically significant at 90% confidence level. Strong ENSO-index years.

Regions located at	Ensemble size		
	3	**9**	**16**
Extratropics	9	22	32
Tropics	31	40	42

Table 3b Same as Table 3a but for the weak ENSO-index years.

Regions located at	Ensemble size		
	3	**9**	**16**
Extratropics	2	10	15
Tropics	15	24	32

Although this information is not shown in Table 3, the greatest benefit from having a 16-member ensemble comes from detecting rainfall predictability in extratropical regions (for strong ENSO-index years from 5 to 12 to a projected 22 (out of 24) significant values with 3-, 9- and 16-member ensembles respectively).

7. Impact of model formulation on seasonal predictability estimates

We briefly discuss the impact of a more recent version of the ECMWF model, cycle 12r1, on seasonal potential predictability estimates. This model cycle differs from cycle 36 in a number of components related to the physical parametrizations. Branković and Molteni (1995) found that, overall, cycle 12r1 has an improved climate when compared with cycle 36.

In Fig.15, we compare the rainfall t-values for model cycles 36 and 12r1 in the strong ENSO-index JJA seasons over selected extratropical and tropical regions. For northern Europe, the 99% confidence level is reached with a smaller ensemble size. In contrast, for southern Europe, a somewhat increased 12r1 ensemble size is needed to achieve the same level of predictability. In the Sahel, the t-values are lower with cycle 12r1 than with cycle 36, but this has little impact on the high predictability over this region. By contrast, t-values are higher over the Indian region with the newer model.

Overall, 6 out of 12 regions shown in Fig.4 have higher rainfall t-values with cycle 12r1. For 2 regions the impact of this cycle is neutral, and for 4 regions some (slight) reduction of predictability estimates is found. It is obviously not possible to make any general conclusions about the impact of the newer model on predictability estimates, other than to caution that such estimates will be model dependent.

8 Probability forecasts

As discussed in the introduction, the underlying SST fields can be thought of as influencing the geometry of the atmospheric attractor. We can represent this influence through changes in the PDF of atmospheric states. By focusing on changes to the PDF of specific weather variables, such as rainfall, the discussion is relevant to the practice of operational seasonal weather prediction.

Fig.16 depicts the distribution of rainfall amounts over southern Europe and the Sahel for all individual integrations of the (strong ENSO-index) JJA 1987 (left-hand bars) and JJA 1988 (right-hand bars) ensembles. The experiment number on the x-axis depicts individual members of ensembles with respect to their (ascending) initial dates shown in Table 1. For example, experiments denoted by the number 6 were initiated on 2 May of 1987 and 1988 respectively. The rainfall is averaged over land grid points only (the regions' boundaries are shown in Fig.4). The variation of regional rainfall amounts within ensembles is seen in both diagrams, though the response of the model to the same SST forcing is much more stable in the Sahel. From such

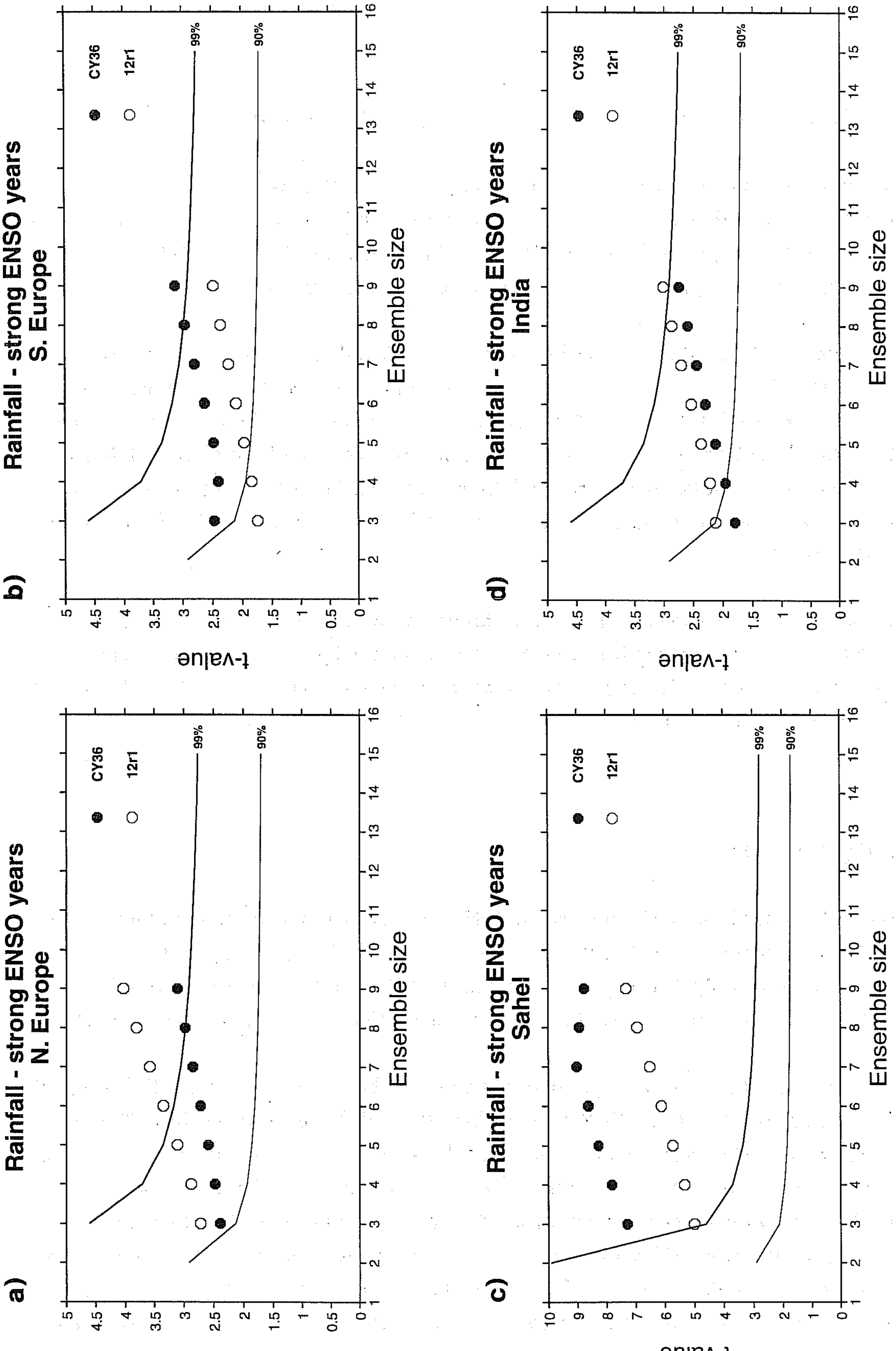

Fig.15 Same as Fig.12 but for rainfall in the strong ENSO-index JJA seasons only and model cycles 36 (solid circles) and 12r1 (open circles) over: a) northern Europe, b) southern Europe, c) Sahel and d) India.

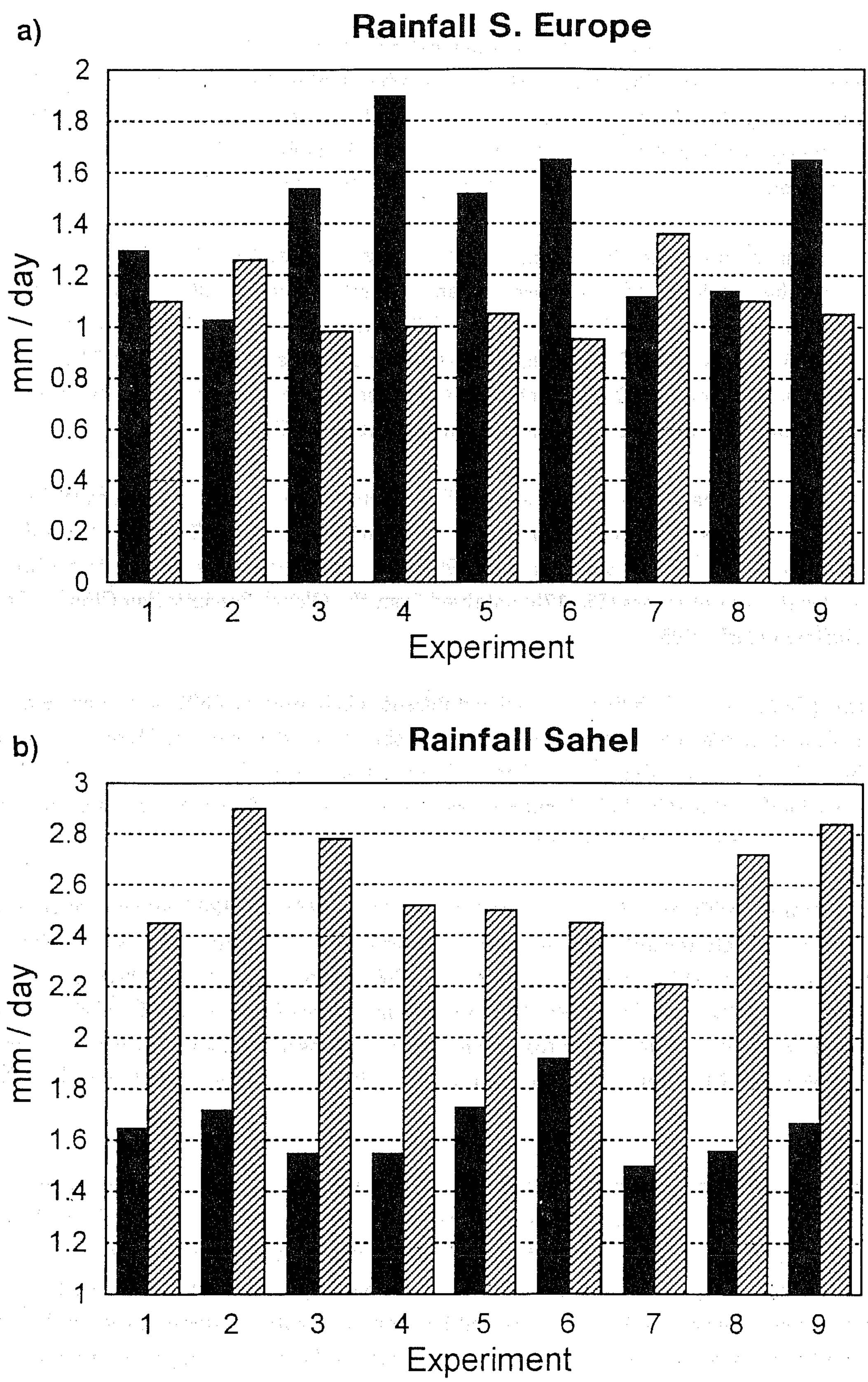

Fig.16 Distribution of seasonally averaged rainfall (in mmday^{-1}) in the JJA 1987 (left-hand, solid bars) and JJA 1988 (right-hand, hatched bars) ensembles for: a) southern Europe and b) Sahel.

distributions we can estimate changes to rainfall PDFs.

It is interesting to note from Fig.16a that a 3-member subensemble {2,7,8} would actually give the opposite sign than other combinations of a 3-member subensemble. However, for most of 3-member combinations in Fig.16a, the average rainfall difference between JJA 1987 and JJA 1988 would be of the same sign and opposite to that for the {2,7,8} sub-ensemble. This is why the value for the JJA 3-member ensemble shown in Fig.12f reached a relatively high confidence level.

Fig.17a shows an example for the strong ENSO-index years over the African and south Asian region. From the incidence of both positive and negative rainfall differences, a grid-point probability (or proportion of positive and negative differences) is assigned. For each grid point, the value obtained is an estimate of the probability that, for JJA, the El Niño year 1987 was wetter or drier than the La Niña year 1988. Probabilities for both positive and negative differences are then combined into one map with the discriminating contour of 60%.

Over the Sahel, the probability of negative rainfall difference (light stipple) is high, exceeding 90% over much of the region. This result implies that for most of the 81 differences, the ensemble simulations of JJA 1987 were drier than JJA 1988. This is in good agreement with verification differences for the two summers (Fig.17b) obtained from the Global Precipitation Climate Centre (GPCC; Huffman *et al.* 1995).

Over much of India, the probability of negative differences is in excess of 60%, and more than 90% in the north and in the south. Generally smaller probability estimates over India imply lower predictability for that region than for the Sahel. This is associated with a relatively larger internal ensemble variability over the Indian subcontinent than over the Sahel and is consistent with estimates of the t-variable shown in Fig.14.

The GPCC rainfall verification data were available for years 1987 and 1988 only and at this stage it is difficult to validate precipitation forecasts in general; the production of global precipitation datasets is an ongoing effort. Hence, in order to validate such probability forecasts, we have created fields giving the probability that the local 500 mb geopotential height difference is either positive or negative. We would expect that for gridpoints enclosed by a given probability contour (here we choose the 60% probability contour), at least 60% of these gridpoints would validate correctly.

Table 4 shows a set of validations for the 9 regions that span the globe (Fig.4). For each region, and each season from the strong ENSO-index years, we create a 2×2 table. The two diagonal elements of each table show the percentage of points where: a) the probability of a positive difference exceeded 60%, and a positive difference occurred (top left element of the table), and b) the probability of a negative difference exceeded 60%, and a negative difference occurred (bottom right element). These two diagonal elements essentially show the degree of agreement between the probability forecasts and verifying analyses. The off-diagonal elements correspond to forecasts which did not verify.

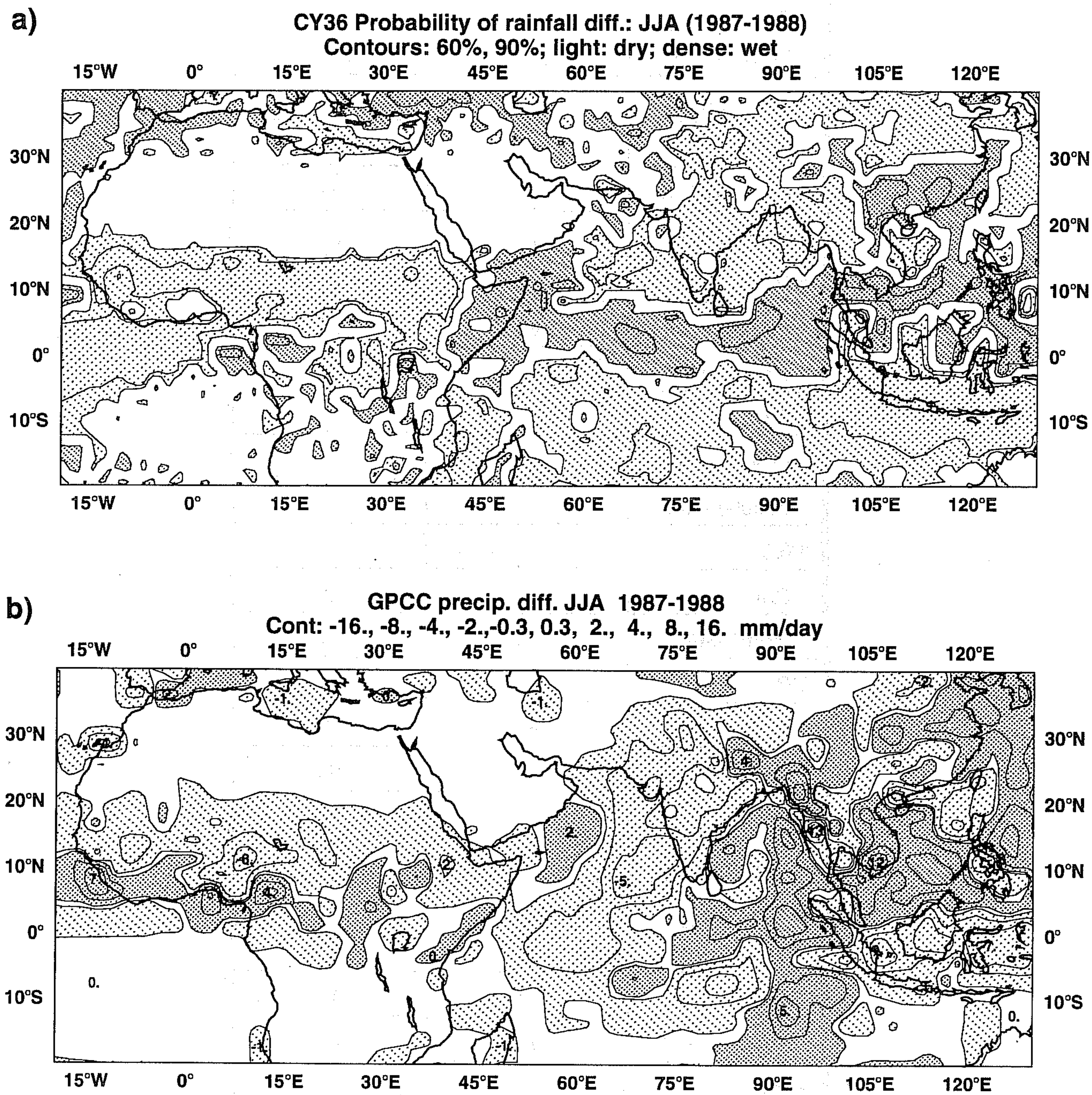

Fig.17 a) Probability of rainfall differences, JJA 1987 minus JJA 1988. Light stipple probability of negative differences, dense stipple probability of positive differences. Contours 60% and 90% for both probabilities. b) The GPCC rainfall differences JJA 1987 minus JJA 1988. Contours ±0.3, ±2, ±4, ±8, ±16 mmday[-1].

Table 4 Percentage of grid points for 9 "global" regions (Fig.4) for which probability of positive (top row) and negative (bottom row) 500 mb height difference exceeds 60%. Strong ENSO-index, MAM season.

Region	Anal diff > 0	Anal diff < 0
NH1		
Prob(+)>60%	66%	34%
Prob(-)>60%	21	79
NH2		
Prob(+)>60%	59	41
Prob(-)>60%	17	83
NH3		
Prob(+)>60%	89	11
Prob(-)>60%	25	75
TR1		
Prob(+)>60%	97	3
Prob(-)>60%	-	-
TR2		
Prob(+)>60%	97	3
Prob(-)>60%	33	67
TR3		
Prob(+)>60%	95	5
Prob(-)>60%	0	100
SH1		
Prob(+)>60%	45	55
Prob(-)>60%	88	12
SH2		
Prob(+)>60%	84	16
Prob(-)>60%	17	83
SH3		
Prob(+)>60%	57	43
Prob(-)>60%	37	63

As an example, in Table 4 we show results for the strong ENSO-index MAM season. The best agreement between the model probabilistic fields and observed differences is found in the tropical Atlantic (TR3). In the northern hemisphere, this verification method gives satisfactory results for all three regions, whereas in the southern hemisphere, results are poor for SH1.

In JJA (and also in SON), the agreement in the three northern hemisphere regions is generally lower than in the boreal cold seasons, DJF and MAM (not shown). In the tropics, negative differences are entirely accurately predicted by the model, and in the southern hemisphere they are better predicted than positive differences.

9. Summary and conclusions

Results from a large set of 120-day, 9-member ensemble integrations of a T63L19 version of the ECMWF model have been presented. Individual ensemble members were initiated from consecutive operational ECMWF analyses, separated by 24 hours. Integrations were made using specified observed SST, updated in the model every 5 days. The last three months of each individual integration, corresponding to conventional calendar seasons, were analysed. This set of ensembles is an extension of the 3-member ensembles reported by Branković *et al.* (1994; BPF).

We focus on the ability of the model to simulate interannual atmospheric variations on seasonal time-scales over the 5-year period, 1986-1990. This period was characterized by significant variability in the El Niño-Southern Oscillation (ENSO). Based on ENSO, an index was defined that varied from large positive values in the first part of the 5-year period to large negative values in the middle of the period with weak negative and weak positive values towards the end of the period. As in BPF, difference fields were computed from seasons when the ENSO index was large and opposite, and weak and opposite.

The skill of the model was first discussed in terms of anomaly correlation coefficients (ACCs). The ACCs have been derived with respect to the two different reference fields (used in the computation of model anomalies). The first reference field was the 'observed climate', the second reference field was the 'model climate'. It was shown that the model skill scores depend strongly on the choice of a reference climate. Therefore, in our analysis of skill scores we focus on difference fields between pairs of years with the opposite index of ENSO.

Distributions of ensemble skill scores for difference fields were calculated for 9 regions over the globe. During strong ENSO-index years, the highest and most sharply peaked distribution of skill in the northern extratropics was found for the northern Pacific/North American region for the winter (DJF) season. Over the northern Atlantic/European region, the ensemble skill is highest and most sharply peaked in spring (MAM). This may have some important implications for the application of seasonal predictions in the growing season in Europe. In the tropical regions, skill is generally high with very sharply peaked distributions. In DJF and MAM, the highest skill is found in the tropical Pacific region, and in JJA and SON in the tropical Atlantic and tropical

Indian Ocean regions.

For weak ENSO-index years, the distributions of skill are found to be generally broader than for the strong ENSO-index years. However, for many regions there is a tendency of skill distributions to be skewed towards positive values. A shift towards negative correlation coefficients in the southern hemisphere may be associated with non-negligible model systematic errors.

For a given season, estimates of consistency for all possible differences between members of two ensembles were also made. These distributions can be thought of as giving some measure of intra-ensemble ("internal") variability. For strong ENSO-index years, consistency distributions in the tropical regions are peaked at high positive values. In the northern hemisphere, spring has the highest consistency between ensembles, similar to distributions of ensemble skill. For weak ENSO-index years the consistency distributions are generally broad, though clearly shifted towards positive correlations.

The above conclusions are generally consistent with those from BPF for 3-member ensembles. However, some differences that may exist between the two papers could be attributed to a much poorer sampling in our earlier work. In BPF, for example, for the northern Atlantic/European area the model skill for both winter and spring seasons was found to be relatively high and almost identical. The increase in ensemble size as well as the analysis of skill distribution performed in this paper help to better discriminate between those two seasons.

Based on the above ensemble skill and consistency estimates, we can distinguish between the regions of the globe and seasons with relatively good prospects for seasonal prediction. Apart from the tropical regions in general, the northern Pacific/North American region in winter and the northern Atlantic/European region in spring appear to have such a potential during ENSO years. These results are not inconsistent with observational and empirical seasonal predictability studies (e.g. Halpert and Ropelewski 1992, Livezey 1990, Barnston 1994).

In this paper, an estimate, based on the t-statistic, is given of the minimum size of an ensemble required to simulate with confidence the impact of the underlying SST anomalies on the probability distribution of atmospheric states. These t-values were derived as a function of ensemble size, when the latter increases from 3 to 9 members. In addition, an extrapolation of the t-statistic for larger ensembles is discussed. This evaluation is performed for 2m temperature and precipitation over a number of predefined regions in both extratropics and tropics.

In general, relative large ($\geq$20 member) ensembles may be needed for extratropical seasonal prediction of regional weather, even in the presence of a relative strong tropical signal. On the other hand, in the tropics during strong ENSO events, the same level of confidence can be attained with much smaller ensembles. However, these estimates may vary widely, depending on the region and season considered. For example, whereas the JJA rainfall prediction for the Sahel may require only a 2-3 member ensemble, for India at least 9 or 10 member ensembles would be desirable. In weak ENSO-years, a relatively large ensemble would be needed even for tropical regions.

Some examples of probability fields were shown, focusing on monsoon rainfall probability. An objective verification of such probabilities was given indicating good agreement between observed differences and probabilities inferred from model difference fields.

Results were shown indicating some model dependence of seasonal predictability estimates. However, overall, these results were not definitive. An extensive set of multi-model ensembles, based on a research project made jointly with the United Kingdom Meteorological Office, Météo France and the French Electricity Board, currently in progress, will allow more conclusive investigation on the rôle of model formulation on seasonal predictability estimates.

References

Anderson. D.L.T., 1996: The legacy of TOGA: Seasonal climate prediction. (this Volume).

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

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

Bendat, J.S. and A.G. Piersol, 1971: *Random data: analysis and measurement procedures*. Wiley-Interscience, New York, 407 pp.

Branković Č. and F. Molteni, 1995: Sensitivity of ECMWF model wintertime climate to model formulation. ECMWF Tech. Memo. No. 222.

Branković, Č., and T.N. Palmer, 1994: Predictability of summer monsoons. *Proceedings of the International Conference on Monsoon Variability and Prediction*, 9-13 May 1994, Trieste, Italy, Vol.2, 629-636. WMO/TD-No. 619.

Branković, Č., and T.N. Palmer, 1996: Atmospheric seasonal predictability and estimates of ensemble size. *Mon. Wea. Rev.* (to appear)

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

Branković, Č., T.N. Palmer, F. Molteni, S. Tibaldi and U. Cubasch, 1990: Extended-range predictions with ECMWF models: Time-lagged ensemble forecasting. *Q. J. R. Meteorol. Soc.*, **116**, 867-912.

Charney, J.G., and J. Shukla, 1981: Predictability of monsoons. *Monsoon Dynamics*, J. Lighthill and R. Pearce, Eds. Cambridge University Press, 735 pp.

Gray, W.M., C.W. Landsea, P.W. Mielke and K.J. Berry, 1993: Predicting Atlantic seasonal hurricane activity 6-11 months in advance. *Weather and Forecasting*, **3**, 440-455.

Halpert, M.S., and C.F. Ropelewski, 1992: Surface temperature patterns associated with the Southern Oscillation. *J. Climate*, **5**, 577-593.

Huffman, G.J., R.F. Adler, B. Rudolf, U. Schneider and P.R. Keehn, 1995: Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and

NWP model precipitation information. *J. Clim.* **8**, 1284-1295.

Kumar, A., and M.P. Hoerling, 1995: Prospects and limitations of seasonal atmospheric GCM predictions. *Bull. Amer. Meteo. Soc.*, **76**, 335-345.

Livezey, R.E., 1990: Variability of skill of long-range forecasts and implications for their use and value. *Bull. Amer. Meteo. Soc.*, **71**, 300-309.

Lorenz, E.N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130-141.

Lorenz, E.N., 1975: Climatic predictability. In *The physical basis of climate and climate modelling*, WMO-GARP publications series No. 16.

Miller, M.J., A.C.M. Beljaars, and T.N. Palmer, 1992: The sensitivity of the ECMWF model to parametrization of evaporation from the tropical oceans. *J. Climate*, **5**, 418-434.

Molteni, F., L. Ferranti, T.N. Palmer and P. Viterbo, 1993: A dynamical interpretation of the global response to equatorial Pacific SST anomalies. *J. Climate*, **6**, 777-795.

Palmer, T.N., 1987: Modelling low frequency variability of the atmosphere. In *Atmospheric and oceanic variability*, Ed. H. Cattle, Royal Meteorological Society, Bracknell, 182 pp.

Palmer, T.N., 1993: Extended-range atmospheric prediction and the Lorenz model. *Bull. Amer. Meteo. Soc.*, **74**, 49-65.

Palmer, T.N., 1994: Chaos and predictability in forecasting the monsoons. *Proc. Indian Nat. Sci. Acad.* **60**, No.1, 57-66.

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

Palmer, T.N., D. Battisti, Č. Branković, R. Buizza, Y.-Q. Chen, and K. Sperber, 1995: Ensemble techniques for seasonal prediction. Pp 620-632, *Proceedings of the International Scientific Conference on TOGA Programme*, 2-7 April 1995, Melbourne, Australia. WCRP-91, WMO/TD No. 717.

Rowell, D.P., C.K. Folland, K. Maskell, and M.N. Ward, 1995: Variability of summer rainfall over tropical north Africa (1906-92): Observations and modelling. *Q. J. R. Meteorol. Soc.*, **121**, 669-704.

Shukla, J., 1981: Dynamical predictability of monthly means. *J. Atmos. Sci.*, **38**, 2547-2572.

Shukla, J., C. Nobre and P. Sellers, 1990: Amazon deforestation and climate change. *Science*, **247**, 1322-1325.

Simmons, A., D.M. Burridge, M. Jarraud, C. Girard, and W. Wergen, 1988: The ECMWF medium-range prediction models: Development of the numerical formulations and the impact of increased resolution. *Meteor. Atmos. Phys.*, **40**, 28-60.

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

Tibaldi, S., P. Ruti, E. Tosi and M. Maruca, 1995: Operational predictability of winter blocking at ECMWF: an update. *Ann. Geophysicae*, **13**, 305-317.

Ward, M.N, and C.K. Folland, 1991: Prediction of seasonal rainfall in the north Nordeste of Brazil using eigenvectors of sea surface temperature. *Int. J. Climatol.*, **11**, 711-743.

Yarnal, B. and H.F. Diaz, 1986: Relationship between extremes of the Southern Oscillation and the winter climate of the Anglo-American Pacific Coast. *J. Climatol.*, **6**, 197-219.

Zwiers, F.W., 1993: Simulation of the Asian summer monsoon with the CCC GCM-1. *J. Climate*, **6**, 470-486.