Forecasting global ENSO-related climate anomalies

T.P. Barnett, L. Bengtsson, K. Arpe, M. Flügel, N. Graham, M. Latif, J. Ritchie, E. Roeckner, U. Schlese, U. Schulzweida & M. Tyree

Forecasting global ENSO-related climate anomalies

By T. P. BARNETT[1]*, L. BENGTSSON[2], K. ARPE[2], M. FLÜGEL[2], N. GRAHAM[1], M. LATIF[2], J. RITCHIE[1], E. ROECKNER[2], U. SCHLESE[3], U. SCHULZWEIDA[3] and M. TYREE[1], *[1]Climate Research Division, Scripps Institution of Oceanography, 0224, La Jolla, California, 92093 USA; [2]Max Planck Institute for Meteorology, Bundestrasse 55, D-20146 Hamburg, Germany; [3]Deutsches Klimarechenzentrum, Bundestrasse 55, D-20146 Hamburg, Germany*

ABSTRACT

Long-range global climate forecasts have been made by use of a model for predicting a tropical Pacific sea surface temperature (SST) in tandem with an atmospheric general circulation model. The SST is predicted first at long lead times into the future. These ocean forecasts are then used to force the atmospheric model and so produce climate forecasts at lead times of the SST forecasts. Prediction of the wintertime 500 mb height, surface air temperature and precipitation for seven large climatic events of the 1970 to 1990s by this two-tiered technique agree well in general with observations over many regions of the globe. The levels of agreement are high enough in some regions to have practical utility.

1. Introduction

The capability to forecast some atmospheric climate changes at time scales of a season or more in advance has been demonstrated by a variety of statistical forecasting techniques (e.g., Barnett, 1981a, b; Barnett and Preisendorfer, 1978, 1987; Harnack et al., 1986; Broccoli and Harnack, 1981). The source of this predictive skill has been traced to low frequency changes in the sea surface temperature of the oceans, particularly the tropical Pacific Ocean. Sophisticated numerical models, atmospheric general circulation models or AGCMs, reproduce moderately well contemporaneous atmospheric climate variability provided they are forced with observed SST, i.e., they have skill at "nowcasting" (cf. Shukla and Wallace, 1983; Latif et al., 1990; Barnett et al., 1991; Graham et al., 1994). But their skill at long range climate forecasting is virtually nil if used in a stand alone mode. This is so because the AGCMs by themselves lack knowledge of the ocean memory that the statistical studies have identified

as the source of predictive skill. In an attempt to remedy this deficiency, the AGCMs have been coupled to ocean models producing coupled GCMs (CGCMs) in an attempt to provide the ocean memory and evolution required for successful long-range forecasts. These CGCMs have, to date, showed limited forecast skill due to a variety of complex problems that attend their construction and initialization. (See Neelin et al. (1992) for excellent summaries of coupled model performance and Miyakoda et al. (1988, 1989, 1993) for examples of coupled model forecasts). However, recent forecasts with a CGCM by the Max Planck Institute (Hamburg) are encouraging and suggest some of the major problems are close to solution so that fully coupled GCMs may one day be legitimate competitors for extended climate prediction (Latif et al., 1993). Whether the CGCMs will be more skillful than the method described in this paper remains to be seen. But one thing is clear: The CGCMs will require far more computer time than the 2-tiered forecast technique to be described below.

The above situation for atmospheric forecasting is in marked contrast to the current state of our ability to forecast tropical Pacific Ocean tem-

---

* Corresponding author.

perature changes associated with ENSO, the
major source of ocean memory responsible for the
predictive ability of the statistical models noted
above. Skillful forecasts of tropical SST out to
lead times of 12–18 months have been clearly
demonstrated by 3 widely different techniques
(Barnett, 1981b; Graham et al., 1987; Barnett
et al., 1988; Cane and Zebiak, 1985; Zebiak and
Cane, 1987; Barnett et al., 1993). Unfortunately,
these ocean forecast tools by themselves are of
little use for predicting atmospheric changes out-
side of a narrow Pacific equatorial strip. Yet it is
well-known from empirical studies that there is a
close, simultaneous relation between changes in
tropical SST and various atmospheric variables
over many regions of the world (e.g., Ropelewski
and Halpert, 1986 and 1989; Horel and Wallace,
1981; and many others). We have taken advantage
of this observed fact to improve our long-range
atmospheric prediction capability.

In this paper, we combine our ability to make
long range SST forecasts with the "nowcasting"
abilities of a sophisticated AGCM, in order to
make atmospheric forecasts with an AGCM at
here-to-fore unattainable lead times. Specifically,
we use an ENSO forecast model to predict tropical
Pacific SST anomalies at lead times between 6–14
months in advance. These forecast SST fields are
then used to force an AGCM in order to predict
both the local and remote atmospheric climate
variations that will accompany the predicted
evolution of the tropical ocean's surface tempera-
ture field. The results from this 2-tiered forecast
scheme are highly encouraging.

The subsequent sections of this paper discuss
briefly the models and data used in the study and
the strategy of the forecasts experiments. Examples
of the long range atmospheric forecasts based on
the SST forecasts are shown next and the skill of an
ensemble of forecasts is presented. We conclude
with a discussion of some of the more interesting
findings and a summary of the work and its short
comings.

2. Models and data

2.1. Models

The results of this paper were obtained by using
a sequence of models. These are described briefly
below.

Tropical Pacific coupled model. Long-range
forecasts of the tropical Pacific SST were obtained
by the hybrid coupled model (HCM) described by
Barnett et al. (1993). This section summarizes
briefly the results of that paper. The ocean compo-
nent of this coupled model is the fully nonlinear
ocean GCM described by Latif (1987). The
OGCM is bounded by latitudes 30° and zonally by
Asia and South America. It has 13 levels in the
vertical with 10 of them in the upper 300 m. The
seasonal cycle in the model is driven by a New-
tonian heat flux and observed wind stress
(Goldenberg and O'Brien, 1981). A Richardson
number dependent vertical mixing scheme is
employed (Pacanowski and Philander, 1981). The
performance of this model has been ascertained in
a number of studies, perhaps the most thorough
being that of Miller et al. (1993).

The atmospheric component of the HCM is
statistical in nature. It was devised as a regression
model that relates observed, anomalous wind
stress and observed, anomalous SST. In operation,
the OGCM provides the SST anomaly to this
atmospheric model and it, in turn, provides the
wind stress anomaly that forces the ocean. The
interface between the 2 models includes an MOS
corrector scheme that compensates for errors in
the OGCM-produced SST fields. Tests of the
HCM in both the hindcast and forecast mode
show it produces significant skill in projecting SST
out to lead times of 12–18 months.

The model is most skillful for prediction of
northern wintertime SST in the central equatorial
Pacific (see Figs. 8–11, Barnett et al., 1993). It
has modest to poor performance in the far eastern
and western Pacific and for latitudes polarward
of approximately 8°. Like most current ENSO
forecast models, it performs poorly on forecasts for
the late Spring–early Summer period (see Barnett
et al., 1993 for additional details).

Atmospheric model. The atmospheric model
used in this study is ECHAM3 from the Max
Planck Institute for Meteorology, Hamburg
(cf. Roeckner et al., 1992). This state-of-the-art
AGCM has 19 levels in the vertical and was run at
T42 resolution (about $2.8 \times 2.8°$ grid) in the
experiments described below. Cloud moisture is a
prognostic variable in the model and this affects
cloud formation at all levels. The model also
employs a comprehensive mass flux for cumulus
convection scheme, as well as accounting for the

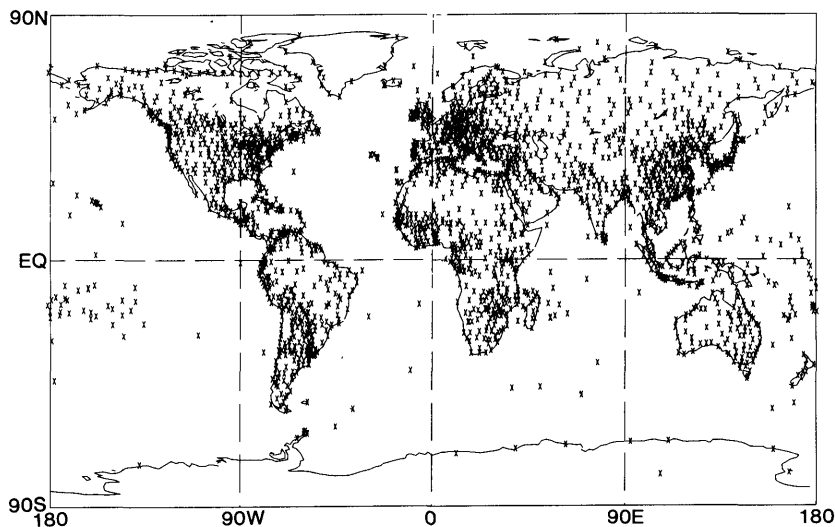LOCATIONS OF 2437 WMO STA. FOR OBSERVED PRECIP 1971-1993

Fig. 1. Distribution of precipitation stations.

effects of clouds on turbulent mixing in the boundary layer. In the current set of experiments, the seasonal cycle of solar radiation and sea ice are prescribed. Numerous tests of the model suggest it reproduces observed climate variability with enough fidelity to be quite useful in the forecast experiments to be described below (Roeckner et al., 1992; Sherwood et al., 1993; Gaffen et al., 1991; and others).

2.2. Observations

The forecasts made in this study were verified against three different global sets of observations: 500 mb height, near-surface air temperature, and precipitation. Each data set and any special treatment it received are described below.

*500 mb height (h) data.*The analyzed data product from both the National Meteorological Center (NMC) and European Center for Medium Range Weather Forecasting (ECMWF) were available from 1979-present. Over most of the span, the 2 products were indistinguishable from each other, so we used them interchangeably. Anomalies were computed using the 1979–92 base period and then regridded to the ECHAM3 T42 grid. The model forecasts are compared against the resulting data set.

Surface air temperature (T). The air temperature data were provided courtesy of C. Ropelewski

(Climate Analysis Center). The data were provided on a $2 \times 2°$ grid and came originally from station observations. The data span the period 1970–1992. Details of the data set construction can be found in Halpert and Ropelewski (1992) and Ropelewski et al. (1992).

Precipitation (p). The precipitation data were again provided by C. Ropelewski (see above reference) and came to us as raw station data for 2437 WMO stations for the period 1971–1992 (Fig. 1). Using stations that had less than 5% missing data, we computed the standardized, anomaly precipitation time series for each station. These irregularly spaced data were projected onto a $2 \times 2°$ grid by simply averaging all the series that fell within individual 2×2 boxes. The data for individual continents were then filtered by an EOF analysis to exclude variability at spatial scales less than about two ECHAM3 grid intervals, 5.6° of latitude or roughly 500 kms. The filtered data sets were used as the verification product.

3. Experimental methods

3.1. Forecast procedure

The basic idea was to forecast the Pacific SST at long lead times with the HCM and then use the predicted SST to force the AGCM. In this scheme,

the application of the AGCM appears essentially as a specification or "nowcast" where the evolution of the SST is prescribed according to the HCM forecast. In fact, the result is an atmospheric forecast at the same lead times as the SST forecast from the HCM. We shall discuss below mainly two season in advance climate forecasting with this method. However, some examples of one year in advance forecasts will be given in Section 5.

In practice, the HCM was spun up to 1 June of a candidate year (hereafter t_0). The spinup was accomplished by providing the observed SST anomaly field to the atmospheric component of the HCM. The resulting wind stress anomalies were used to force the ocean in uncoupled mode for several years prior to t_0. The fully coupled HCM used the initial conditions at t_0 and integrated ahead in time through February of the subsequent year. Thus, HCM forecast SST fields for June through February were available to force the AGCM.

The atmospheric forecasts were made from 1 October initial conditions provided by the extended AMIP run (defined below). Starting the forecast integration in October, instead of June, meant the global atmosphere had 2 model months (October and November) to come into balance with the tropical SST predicted by the HCM. Tests, not reported here, showed the 2-month atmospheric spinup time to be adequate for our purposes and so the 1 October start date saved needless integration over the June–September time frame. The HCM SSTs were only provided to the AGCM in the narrow, equatorial Pacific band described in Subsection 2.1. Over the rest of the global ocean, the climatological SST was specified for the AGCM.

Each atmospheric forecast was repeated three times. These realizations used exactly the same predicted SST, but were begun from three different 1 October initial conditions obtained from the extended AMIP run. The final forecasts were averages over the three realizations and over the period December–February. They represent 2 season in advance forecasts of wintertime climate variability by an AGCM. The variables we chose to concentrate on in this paper were 500 mb height (h), 2-m air temperature (T), and total precipitation (p).

We chose as forecast targets the largest ENSO warm/cold events of the last 22 years. The warm event winter conditions we tried to forecast were for the years 1972–73, 1982–83, 1987–88 and 1991–92, while the cold event years were 1970–71, 1974–75, and 1988–89. We felt that if the forecast procedure did not work for these large events then it would likely not work at all.

3.2. Significance tests

Developing stringent significance tests for the forecasts to be described below turned out to be a difficult task. The approach we finally settled on was based on what we term the "extended AMIP" experiment. The original atmospheric model inter-comparison project (AMIP, Gates et al., 1992) used observed global SST and sea ice distribution to force a variety of AGCMs, including ECHAM3, over the period 1979–1988. We extended this original ECHAM3 AMIP run to cover 1970–1992. Thus, for each winter of this 23-year period we had the fields of 500 mb height, 2 m air temperature, and total precipitation produced by ECHAM3 when it was forced by observed SST. We refer to these integrations as "specifications", essentially "nowcasts" with observed SST.

Two immediate questions come to mind.

(1) Are the forecasts or specifications made for the extremes of the ENSO cycle significantly different from specifications made during non-ENSO situations? Contrasting the skills of ECHAM3 between non-ENSO conditions (when we expect little or no skill) and ENSO conditions represents a significance test of the model's ability to successfully specify/forecast. A corollary is that if ECHAM3, forced by observed SST, cannot successfully specify a given variable in a given locale then we would not expect it to be able to forecast the variable/locale even given a perfect SST forecast.

(2) Is there an appreciable difference between the winter specifications made with observed SST and the winter forecasts made with HCM-predicted SST? Certainly the SST forecasts are flawed (see above) and we wish to know how much this degraded the atmospheric forecasts. In Subsection 4.3, it will be shown the forecasts and specifications made with observed SST are basically the same.

The 1st question is the most important one and it was addressed as follows: The 16 non-ENSO winter fields of T, h, and p from the extended

500mb HT EXTREME WINTERS SPECIFICATION vs. OBS

AIR TEMP EXTREME WINTERS SPECIFICATION vs. OBS

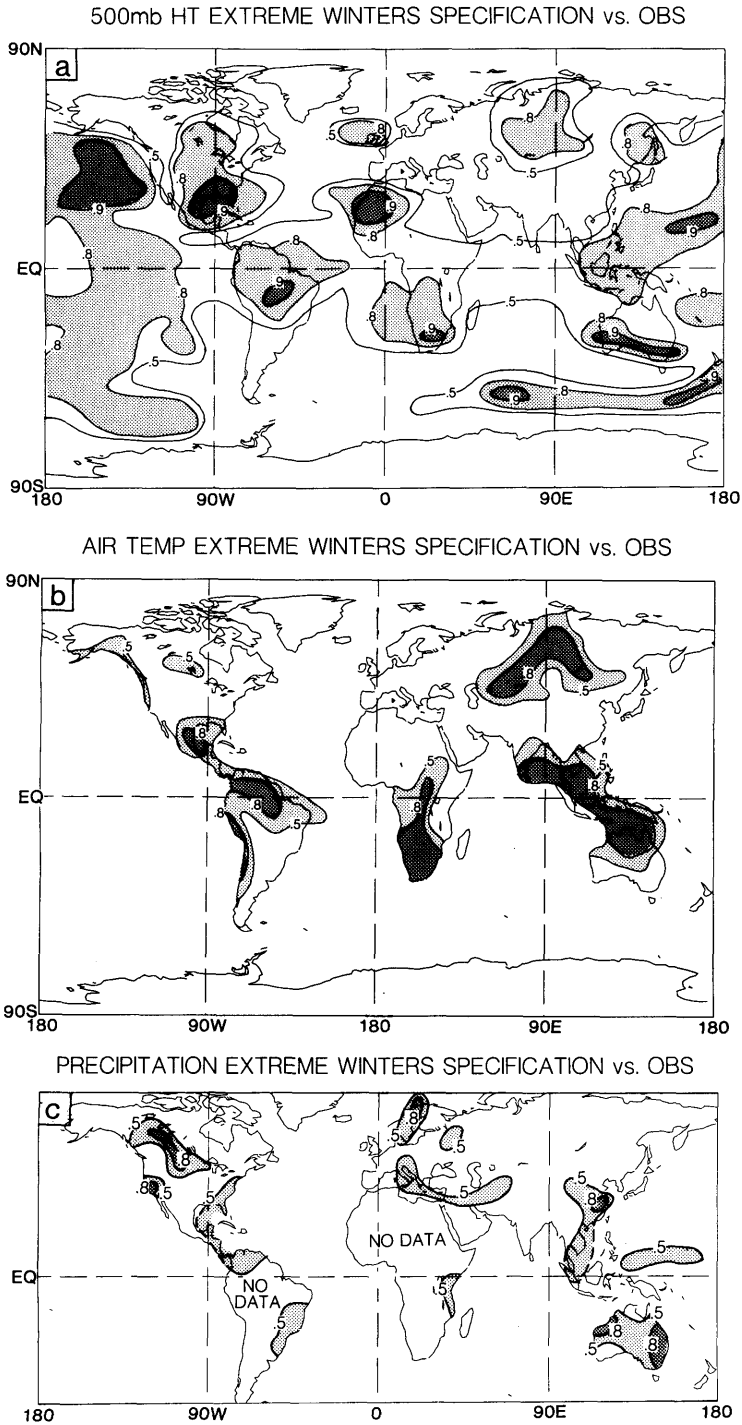PRECIPITATION EXTREME WINTERS SPECIFICATION vs. OBS

Fig. 2. Specification skill (correlation) for (a) winter 500 mb height anomalies (b) for near surface air temperature, and (c) precipitation for 7 large ENSO events. Light (heavy) stipple for 0.10 (0.05) confidence level.

AMIP run were considered a "no skill" sampled population*. 3 winters were randomly selected (with replacement) from this ensemble and averaged, as was done with the 3 realizations of the forecasts, to form one realization of a "no skill" forecast. This procedure was repeated 6 more times to match the 7 real forecasts. The 7 no-skill forecasts were correlated grid point by grid point with the observed variables for the 7 candidate forecast years. The correlation coefficient $r^v(x)$ represented a measure of the model success one might expect by chance for one set of 7 forecast experiments when no real skill exists. Repeating this procedure over 1000 realizations $(v_i, \ i = 1, 2, ..., 1000)$ of sets of 7 forecast experiments provided a population of $r^v(x)$, from which it was easy to construct a probability distribution function (pdf) for expected values of $r(x)$ under the condition of no expected skill. The 0.01, 0.05, and 0.10 values of $r(x)$ from this pdf were used to determine locally the significance of the correlation $(r_f(x))$ obtained by correlating the 7 actual forecasts with the appropriate set of observations. Note the procedure handles indirectly the problems of field significance raised by Livezey and Chen (1983).

4. Results

4.1. Specification skill

In this section, we estimate the ability of ECHAM3 to reproduce observed variations for the seven large ENSO events given "perfect" (observed) information on SST. This will serve as a background against which to measure model forecast skill and to answer indirectly in Subsection 4.3 the 2nd of the 2 questions posed above. In keeping with the forecast strategy, a group of 3 realizations for each event was computed. The 3 realizations were averaged, as we did with the forecasts, to form the specification. The SST (observed) were the same for each realization but the initial conditions were taken to be the same as those used in the actual forecast experiments. Thus, the only difference between specification and

---

* This is a conservative assumption since weak events may have some modest level of skill which is ignored here.

forecast experiments is in the use of observed versus predicted SST.

500 mb height. The restricted time span of the verification set resulted in only 10 "non skill" cases to use in evaluating the 4 ENSO specifications in the AMIP run. We feel this makes our statistical estimates marginal. Bearing this in mind, the correlation between the 4 specifications and the associated observed winter 500 mb height anomaly fields is given in Fig. 2a.

Correlations exceeding 0.5 are found throughout the tropical strip with many of the values exceeding 0.8, the 0.10 confidence level. Impressive correlations at midlatitudes were found in the eastern North Pacific, eastern North America, the Caribbean, northern and eastern Asia. Large values are found over many areas of the southern oceans but most of these just miss significance at the 0.10 level. Given the small sample we have been forced to use and the fundamental lack of data in these regions leads us to conclude that our results in the Southern Hemisphere oceans are encouraging, but are not definitive one way or the other.

It is clear that there are large regions of the world, e.g., Europe, Asia and northern Africa where the ECHAM3 forced by Pacific SST alone has no ability to specify the observed variations in 500 mb height. We expect no forecast skill in this variable over these regions.

Surface air temperature. The discussion is restricted to air temperatures over the land masses since the comparison over most of the ocean, where SST is specified to be climatology, is not meaningful. Regions of correlation, significant at 0.10 or above, were found over the northern half of South America, the west coast of North America as well as the Gulf Coast (Fig. 2b). Much of the rest of the United States had correlations significant at 0.15. South Africa and the eastern half of Australia had correlation value just significant at the 0.10 level.

Virtually no skill was found for Europe, most of Asia, and the majority of Africa north of the equator. These findings are much like those obtained for the 500 mb height field.

Precipitation. The precipitation skill is shown in Fig. 2c and is located in many of the areas expected from earlier empirical studies (e.g., Ropelewski and Halpert, 1989). These areas include northwest and southeastern sections of North America, eastern

Asia, and Australia. The signal over Europe is open to question since it has not, to our knowledge, appeared in prior empirical studies. Some highly significant correlations were found in the region but the spatial pattern was uneven and difficult to contour. Over many areas there appears to be little or no skill. However, there are large regions of sparse or missing data. In fact, the skill map shown in Fig. 2c needs to be recomputed once a global, gridded precipitation field is available (a task well beyond the scope of the present effort.)

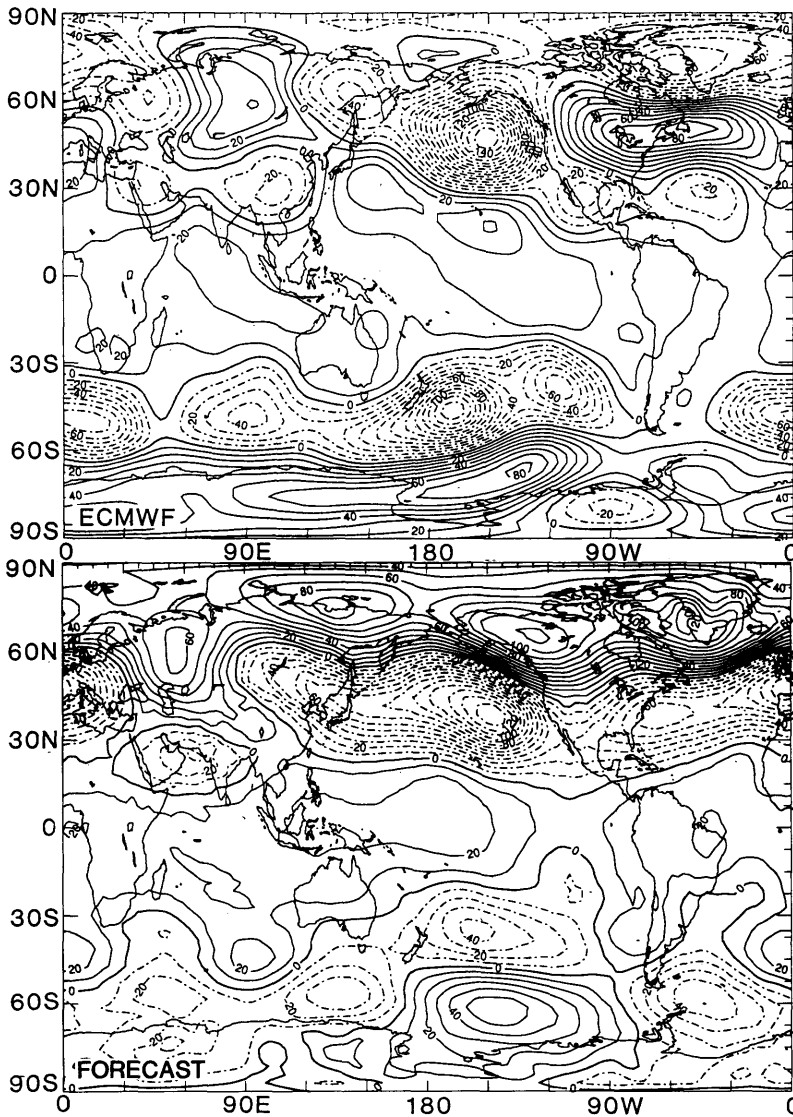## FORECAST vs. ECMWF 500mb FIELD
WINTER 1982 – 83

Fig. 3. Example 2 season lead forecast (lower) and observation (upper) for 1982–83 winter 500 mb height anomaly field. Most warm event forecasts had much the same structure.

4.2. Forecast skill

The distribution of forecast skill at a two season in advance lead time for 3 variables discussed above is presented here. In many regions, the skills reported are high enough to be of practical use.

500 mb height. Examples of forecasts and accompanying verifications for both a warm and cold event are shown in Figs. 3, 4, respectively. The forecasts are seen to capture many gross features of the anomalous height field, especially over the central North Pacific (as expected). The warm/cold forecasts are not simple mirror images of each other, although individual warm (cold)

Fig. 4. Same as Fig. 3 but for the cold event winter of 1988–89.

(a) 500mb HT EXTREME WINTERS FCST vs. OBS

(b) AIR TEMP EXTREME WINTERS FCST vs. OBS

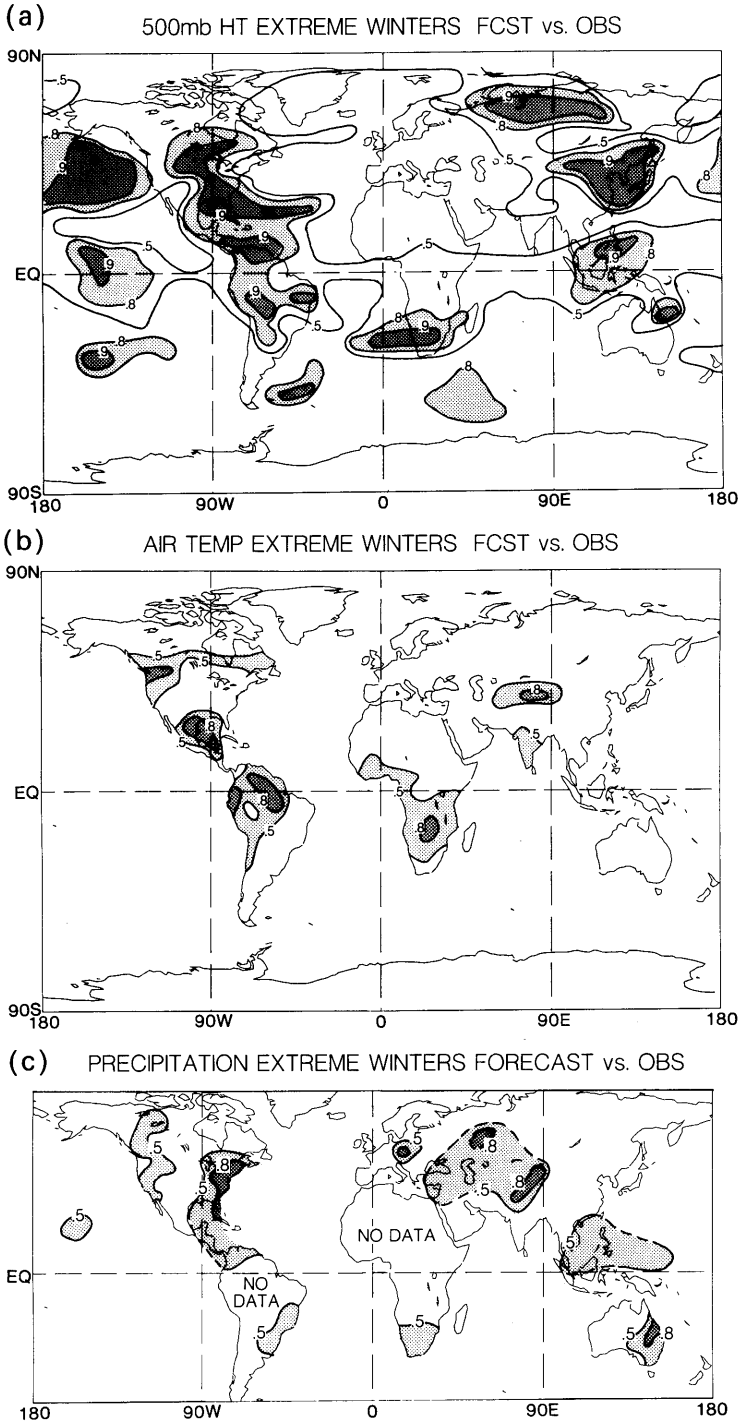**(c)** PRECIPITATION EXTREME WINTERS FORECAST vs. OBS

Fig. 5. Same as 2 but for the 2 season in advance forecasts of (a) 500 mb height (b) air temperature, and (c) precipitation.

events bear a fair similarity over the Pacific and parts of North America. Inspection of the 7 individual forecasts show they are at their worse when the North Atlantic Oscillation, which is not reproduced well in the model forecasts, is highly energetic.

An overview of the skill of the model predictions is provided by the forecast-observation correlation, $r_f(x)$, described above. The values are illustrated in Fig. 5a. The forecast skills demonstrate significant values in many of the same regions found in the specifications, most notably the eastern Pacific, North America, South America, and South Africa. The largest differences occur over North Africa where the forecasts now have no skill and eastern Asia where the forecasts are better than the specification. Note, however, in the latter area the forecasts are not significant at the 0.10 level.

The forecast results also differ from the specifications over much of the southern oceans and the tropical Atlantic and Indian Oceans. Supplying the forecast SST for these latter 2 tropical regions would, one suspects, remedy this latter problem. The situation over Europe is more of a problem, since both forecast and specification demonstrate no skill.

Air temperature. The forecasts of air temperature 2 seasons in advance (Fig. 5b) are essentially identical over the land masses to those obtained from the specification (the "perfect" prediction). The numerical values show some sampling fluctuations but the only spatially coherent variations occur along the West Coast of North America where the forecasts are actually a bit better and over eastern Australia where the specifications are slightly superior. These differences, like some of those noted above, may well be real and are discussed below. We conclude, for now, that the two season in advance surface air temperature forecasts are not appreciably different from the specifications.

Precipitation. The 2 season in advance precipitation skill (Fig. 5c) is quite similar to the specification (Fig. 2c). In some areas, it is appreciably better than the specification (see below). This is especially true for the east coast of the United States and Australia. The improved skill in central Asia is based on widely scattered data and, we feel, apt to be unreliable. In summary, we conclude, as we did with surface air temperature,

that the forecasts are about as good as the specifications.

4.3. Comparison of forecasts and specifications

Even a casual comparison of Figs. 2, 5 shows the forecasts to be as good as the specifications for all intents and purposes. There are several regional differences that may be real. For instance, forecasts for the tropical Atlantic and Indian Oceans are inferior to the specifications. This is presumably due to the use of climatological SST in the areas instead of the actual SST or its predicted value (see Subsection 5.1 and Graham et al., 1994). Work to remedy this shortcoming is in progress.

More interesting are cases where the predictions are clearly better than the specifications. Inspection of these cases shows the forecast SST anomaly to be larger than the observed values used in the specifications. Thus, the atmospheric response has been amplified, giving a better signal/noise ratio and skill. Had we used a skill measure based on variance, instead of correlation, the stronger SST forecast may have been a liability rather than an asset.

Finally, we noted in Subsection 2.1 that the HCM had good SST forecast skill only in the central equatorial Pacific, but not at the ocean margins or areas outside about $8-10°$ latitude. These flaws in the HCM are apparently of no consequence to the midlatitude forecasting skill of our 2-tiered prediction method. This is obvious from the close correspondence between specification and forecast (Figs. 2, 5). The reason for this situation is that the main ENSO teleconnection patterns are based on central Pacific SST; not SST variation off South America or in the far western Pacific. This has been demonstrated in a variety of empirical studies (e.g., Barnett et al., 1981a; Graham and Barnett, 1993; and others) plus GCM sensitivity experiments (e.g., Geisler et al., 1985). Thus, we were fortunate that the HCM had its best performance in just the region of the tropical Pacific that was required to reproduce the observed midlatitude response patterns.

5. Discussion

In this section, we address briefly a number of issues raised by the results presented above.

5.1. The rôle of the midlatitude SST

It is natural to wonder if inclusion of actual midlatitude SST would have enhanced the forecast skill. The answer appears to be "not appreciably". There are several reasons for such an answer. The specifications (AMIP) which used observed *global* SST were not appreciably better than the forecast which used climatological SST outside the equatorial Pacific strip. Apparently the additional observed midlatitude SST information was of little use in helping the AGCM reproduce the observed climate variability.

More to the point, there has been accumulating over the last 15 years a body of evidence that, on the seasonal time scale, the midlatitude ocean is driven by the atmosphere and not vice versa. This answer has been consistently provided by empirical studies since those of Davis (1976), Barnett (1981a), and many others up to modern times where improved data sets and analysis techniques were employed, e.g., Graham and Barnett (1993). Similar results have been obtained via GCM sensitivity experiments. The earlier results in this area are suspect due to model shortcomings. However, recent extensive studies using a good AGCM by Graham et al. (1994) clearly show that information on midlatitude SST does little if anything to improve the performance of the AGCMs in specifying observed change in global and regional
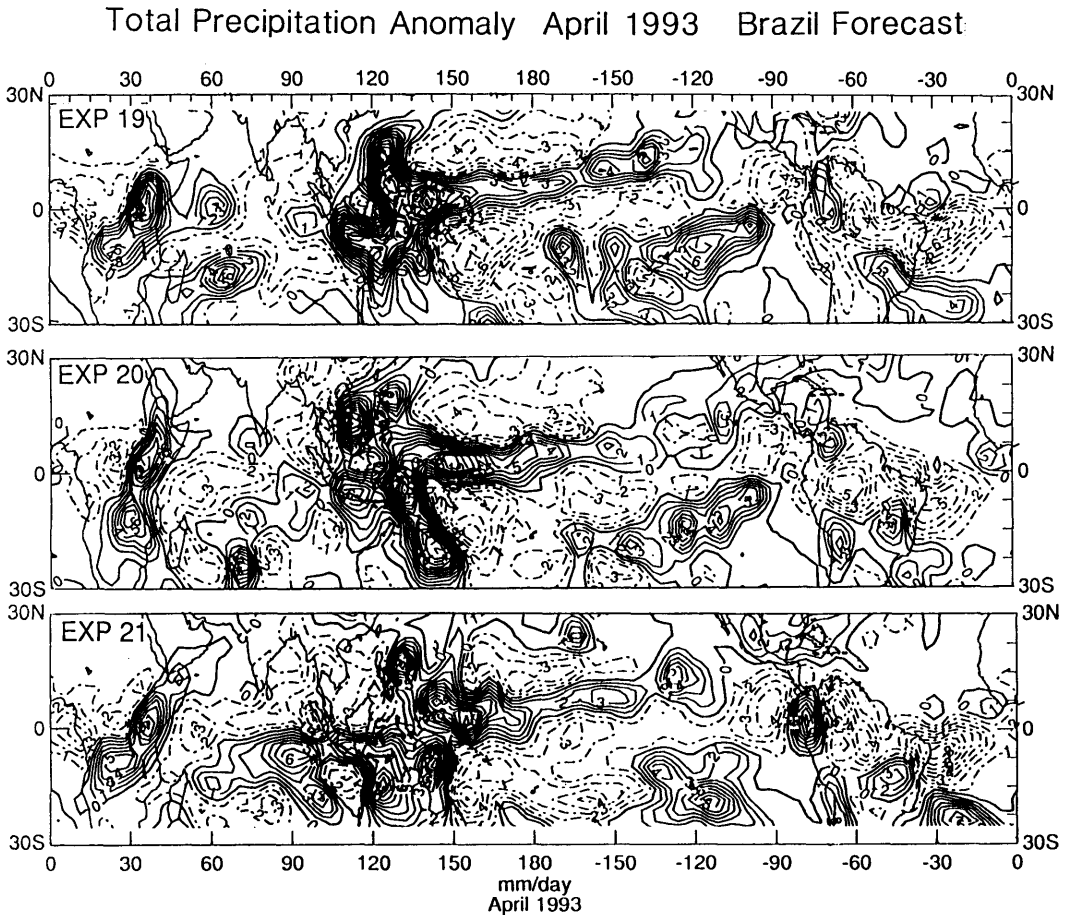

Fig. 6. Example of 3 different predicted realizations of tropical precipitation for April, 1993, courtesy N. Graham. Forecast lead time was 4 months.

climate. The important SST information for successfully modeling observed climate change comes almost entirely from the tropics.

5.2. The forecast ensemble: how many is enough?

The forecast results reported here are based on the simple average of 3 separate forecasts of the same event, each made with different initial conditions but identical tropical SST. Selection of the number 3 was largely intuitive and based on available computer time. We inquire here if that is an adequate number of realizations for an ensemble forecast.

Tropics. Consider first the tropics, say, within 20° of the equator. Inspection of the individual members of forecast ensembles suggests that 3 realizations are likely adequate to obtain a stable estimate of expected conditions. This is illustrated in Fig. 6 (from Graham, personal communication) where the separate forecasts of expected April 1993 precipitation over South America were made from a 1 January initial condition at the request of Brazil which was undergoing a serious drought and wished to know if it would persist. The details by which the forecasts were made are omitted. Rather we show the three different precipitation forecasts made with the same SST but different initial conditions (Fig. 6). Each forecast bears a good similarity to the others, not only with respect to spatial distribution of precipitation but also magnitude of the predicted anomalies. Given the numerous, well-known problems associated with even analyzing precipitation fields, let alone predicting them, we conclude the three forecasts are, for practical purposes, identical.

Midlatitudes. Estimating the required ensemble size to make meaningful forecasts in midlatitudes is an open question but one that is beginning to draw considerable interest (Palmer, 1988; Palmer et al., 1992; Brankovic et al., 1993). In order to obtain a preliminary answer on this subject for this paper, we repeated the 1982–83 forecast from 10 different initial conditions provided by the extended AMIP run in the manner described in Subsection 3.1. The same SST field forecast by the HCM was used in each experiment.

The wintertime 500 mb height field over the region 25–70° N, 120° W–60° W was isolated in both the extended AMIP and the 10 forecast realizations. The first two EOFs of the extended AMIP (60% of the variance) were estimated and

a scatter plot of the principal components was constructed (Fig. 7). The 10 forecast realizations were projected onto the AMIP EOFs and the associated principal components entered onto the same scatter plot. A number of observations may be made from the results.

(i) The 10 realizations do not appear adequate to constitute a stable forecast of the 1982–83 event. The spread about mean (not shown) is nearly as large as the displacement of the mean forecast from the origin. *A posteriori* elimination of the two outlier forecasts would help remedy this problem. In contrast, if one averages the 3 realizations in the upper left end of the group and the three in the lower right end of the group, then the forecast fields would be as shown in Fig. 8. The anomaly patterns are relatively stable since most of the forecasts fall into the same quadrant of the EOF phase space. However, the differences are regionally large and practically significant in terms of their magnitude. This is not good news for those who wish to make extended range forecasts with fully coupled CGCMs for it means, if our results are general, that forecasts must be ensemble averages of a large number of realizations, perhaps 20–30, if one is to get a stable estimate
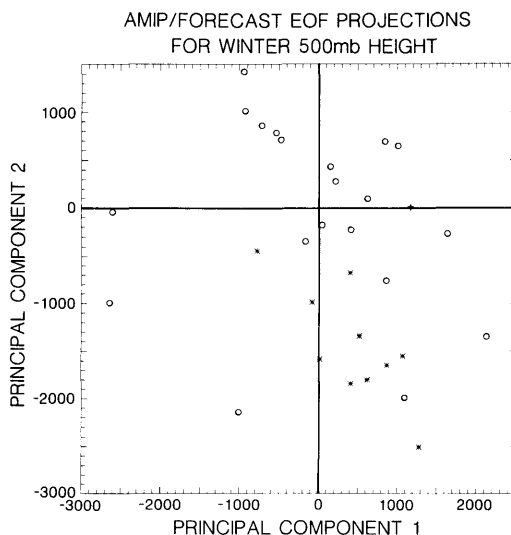

AMIP/FORECAST EOF PROJECTIONS
FOR WINTER 500mb HEIGHT

Fig. 7. EOF projection of winter 500 mb height field over the North Pacific and North America from (i) the extended AMIP run (0), (ii) the 1982–83 AMIP winter (+) and (iii) 10 different realizations of the 2 season lead forecast for the 1982–83 winter (∗).

ENSEMBLE FORECASTS OF 1982-83 WINTER
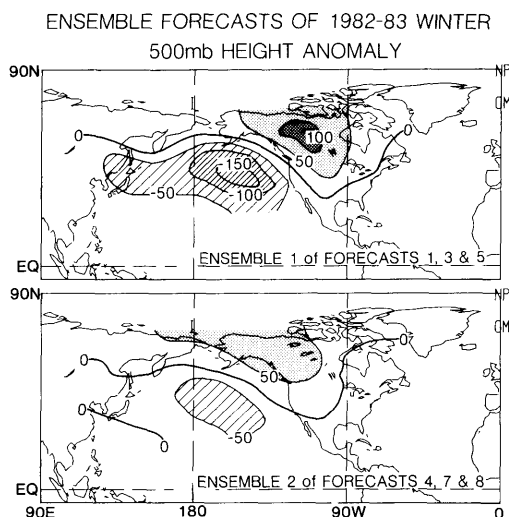500mb HEIGHT ANOMALY

Fig. 8. Two ensemble forecasts made from the three extreme realizations at each end of the 10 forecast realization cluster shown in Fig. 7 (units are geopotential meters).

of the ensemble average. The computer resources required for such an ensemble of forecasts would be immense.

(ii) The difference between forecasts and specifications for the 1982–83 event appear considerably larger here than one would have guessed from comparing Figs. 2a, 5a. This suggests that a more detailed analysis of the forecast error structure is in order.

5.3. Competitors

It is fair to ask if another method could use the SST forecasts and produce atmospheric predictions comparable to those of the AGCM, but with considerably less effort. Specifically, suppose we used a statistical specification method with the HCM SST forecasts. Graham and Barnett (1993) did just this and demonstrated that in some regions and for some variables, the statistical specification yields better results than the AGCM. However, for many regions of the globe and for many variables, the fundamental data sets do not exist that would permit one to construct the statistical specification models. In these cases, the AGCM offers the only possibility for useful forecasts.

Some important climate variables are associated with highly nonlinear processes and have

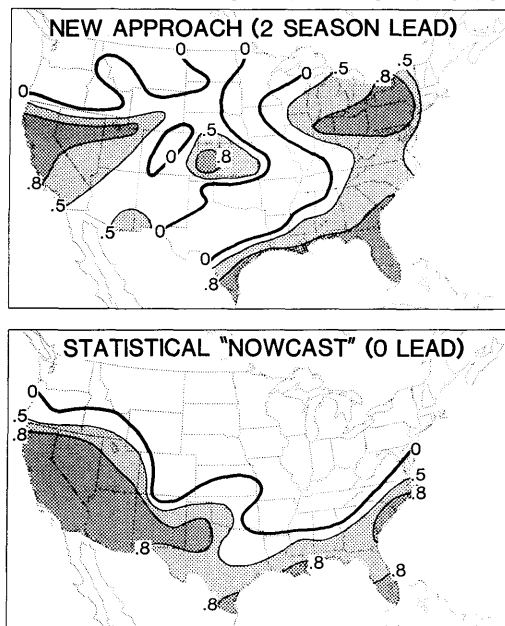JANUARY/FEBRUARY PRECIPITATION

Fig. 9. Upper. Correlation between two (2) season forecasts of winter precipitation for seven large ENSO events and observations. Lower. Correlation between observed central Pacific (NINO3) winter SST anomalies for the same seven events and contemporaneously observed precipitation. The forecast equals or beats the "nowcast" over much of the United States.

decidedly non-Gaussian distributions. Both of these properties make them difficult to model statistically and so, again, the AGCM offers the best chance for useful forecasts. This point is illustrated in Fig. 9 where the 2 season in advance precipitation forecast from our 2-tiered scheme is put against a statistical competitor. The latter is a simple regression model that relates *contemporaneous* wintertime equatorial Pacific SST with observed, wintertime divisional precipitation data for the United States. This is an unfair competitor for the 2-tiered scheme (specification against a 2-season forecast) but nonetheless the new approach more than holds its own. The statistical specification is better in the southwestern US, but slightly inferior to the AGCM forecast precipitation in the southeastern US. However, the AGCM forecast has good skill over most of the eastern third of the US where the statistical method cannot
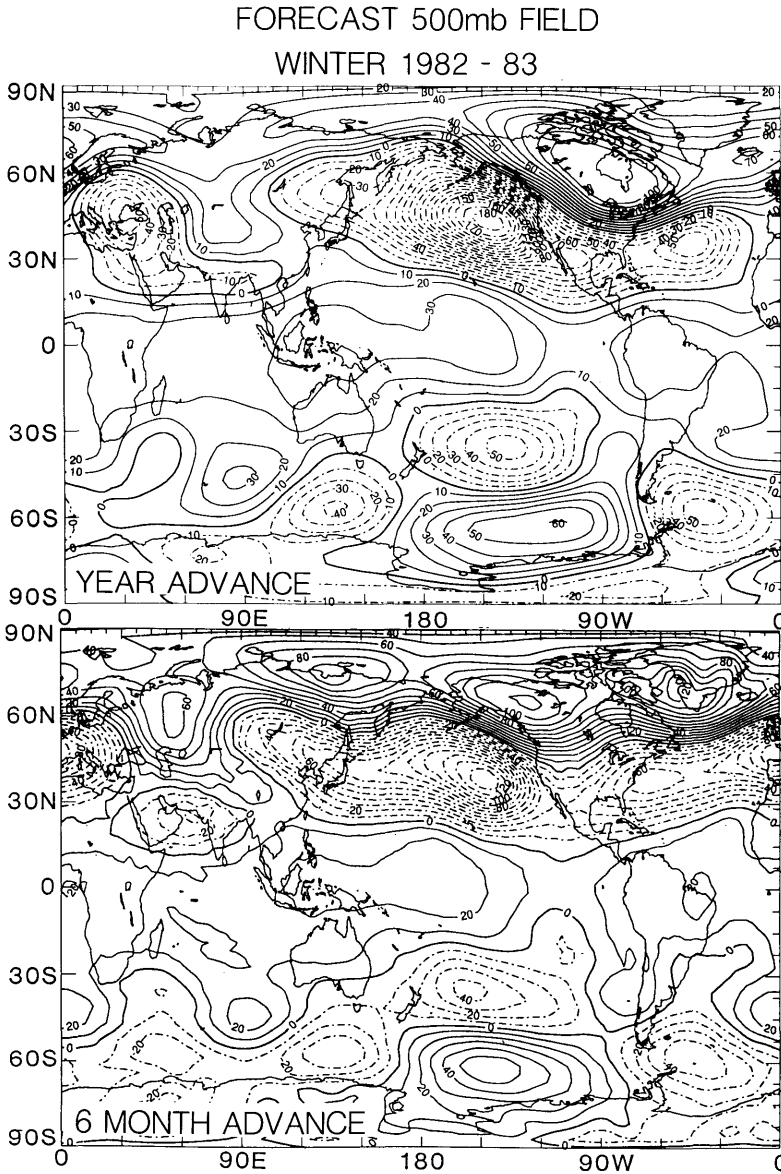
FORECAST 500mb FIELD
WINTER 1982 - 83

Fig. 10. Comparison of 6-month in advance and 12-month in advance forecasts of 500 mb height field for the winter of 1982–83 (units are geopotential meters).

even specify the precipitation given the observed winter SST. We conclude that using a full AGCM as the second leg of the 2-tiered forecast scheme is justified*.

---

* Where possible, we also recommend simultaneous use of statistical specification technique. They are easy, inexpensive and provide an independent forecast that can be compared with the AGCM product.

5.4. Forecast range

The forecast procedures of Sections 3, 4 were repeated on a limited set of events using tropical SST fields predicted one year in advance by the HCM. A typical result (Fig. 10) suggests that the one year lead forecast is essentially as good as the two season lead forecasts. Thus, the key to extended climate forecasts rests on 2 factors:

(i) making a good long-range SST forecast. The basic unanswered question is, "How far in advance can we predict the tropical Pacific SST?" (ii) The second factor is obviously an AGCM that reproduces well the observed climatic variability accompanying observed changes in tropical SST.

Offsetting the highly encouraging result and conclusion presented immediately above is the fact that the scheme we have demonstrated here will have little, if any, success in the absence of a substantial ENSO signal. In the latter case then, the forecast range is effectively zero. This raises a 3rd key point: The long-range ENSO forecasts must be accompanied by some measure of expected forecast success, i.e., we need to forecast the expected forecast skill or else there will be no way to place a confidence factor on the subsequent atmospheric forecast.

6. Conclusions

A 2-tiered scheme for long-range climate forecasting has been demonstrated. The method consists of first making a long range forecast of tropical Pacific sea surface temperatures (SST). The predicted evolution of the SST is used to force an atmospheric GCM thereby producing a climate forecast with the AGCM at the same lead times as the SST forecasts. The approach was demonstrated on the seven large warm/cold ENSO events of the last 23 years. The conclusions were as follow.

(1) The method has highly significant skill in many regions of the world for the 3 predictor fields investigated: 500 mb height, surface air temperature, and precipitation during the Northern Hemisphere winter. The highest skills appear in just the areas one would expect from numerous prior empirical studies: North Pacific, parts of North America, South America, Australia and Africa. Little skill was found over Europe and large regions of Asia. Skill also appears over the southern oceans but lack of reliable verification sets makes this conclusion marginal.

(2) The ability of the AGCM to make 2-season forecasts was, by the correlational skill measures used here, essentially the same as that obtained by the model using observed SST in a specification mode.

(3) The absence of midlatitude SST data seems to have had little, if any, impact on the forecast/specification performance. However, the use of climatological SST in the tropical Atlantic and Indian ocean, instead of forecast SSTs, seem to have degraded forecast performance in localized tropical environs.

(4) Groups of 3 forecasts, made from different initial conditions, were averaged to form a single ensemble forecast and that is the forecast that we evaluated in (1) and (2) above. In the tropics, the three member ensemble appears adequate for a stable forecast. In midlatitudes, even a 10-member ensemble appears inadequate to produce a highly stable forecast without some *a posteriori* intervention.

(5) The skill and forecast range of this approach depends on the occurrence of a substantial warm/cold event in the tropical Pacific. In 2 cases studied, year in advance forecasts were as good as specifications. However, the method will not work in the absence of an ENSO event.

Additional studies are required to investigate a number of shortcomings of the present study, e.g.,

(a) How much did errors in the SST forecasts versus shortcomings of the AGCM contribute to forecast errors?

(b) Is the preliminary conclusion regarding the large number of forecast realizations required for a stable midlatitude ensemble forecast really true, or was it a result of the event year and model chosen?*

(c) How much more skill can be realized by inclusion or prediction of other low frequency forcing fields, e.g., snow cover, sea ice, the North Atlantic oscillation? Will these additional forcing fields help in the "no skill" areas of Europe and much of Asia?

Other such questions can be posed. Fortunately, the results of the current study appear encouraging enough to justify additional pursuit of these problems using the methods we have described here.

---

* The 7 forecasts shown in this study were re-evaluated using 10 member ensembles. The forecasts are slightly better than those shown here (Barnett, 1994, *J. Climate*, submitted).

7. Acknowledgments

REFERENCES

Barnett, T. P. 1981a. Statistical prediction of North American air temperature from Pacific predictors. *Mon. Wea. Rev.* **109**, 1021–1041.

Barnett, T. P. 1981b. Statistical relations between ocean/atmosphere fluctuations in the tropical Pacific. *J. Phys. Oceanogr.* **11**, 1043–1058.

Barnett, T. P. 1985. Variations in near-global sea level pressure. *J. Atmos. Sci.* **42**, 478–501.

Barnett, T. P., Graham, N., Cane, M., Zebiak, S., Dolan, S., O'Brien, J. and Legler, D. 1988. Prediction of the El Niño of 1986–1987. *Science* **241**, 192–196.

Barnett, T. P., Graham, N., Latif, M., Pazan, S. and White, W. 1993. ENSO and ENSO-related predictability: Part I: Prediction of equatorial Pacific sea surface temperature with a hybrid coupled ocean-atmosphere model. *J. Clim.* **6**, 1545–1566.

Barnett, T. P., Kirk, E., Latif, M. and Roeckner, E. 1991. On ENSO physics. *J. Climate* **4**, 487–515.

Barnett, T. P. and Preisendorfer, R. W. 1978. Multifield analog prediction of short-term climate fluctuations using a climate state vector. *J. Atmos. Sci.* **35**, 1771–1787.

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

Brankovic, CC., Palmer, T. N. and Ferranti, L. 993. Predictability of seasonal atmospheric variations. *J. Clim.*, in press.

Broccoli, A. and Harnack, R. 1981. Predictability of monthly north Pacific sea level pressure from monthly sea surface temperature for the period 1922–1976. *Mon. Wea. Rev.* **109**, 2107–2117.

Cane, M. A. and Zebiak, S. E. 1985. A theory for el Niño and the Southern oscillation. *Science* **228**, 1085–1087.

Davis, R. E. 1976. Predictability of sea-surface temperature and sea-level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.* **6**, 249–266.

Gaffen, D., Barnett, T. P. and Elliott, W. P. 1991. Space and time scales of global tropospheric moisture. *J. of Climate* **4**, 989–1008.

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

Geisler, J. E., Blackmon, M. E., Gates, G. T. and Munoz, S. 1985. Sensitivity of January climate response to the magnitude and position of equatorial Pacific sea surface temperature anomalies. *J. Atmos. Sci.* **42**, 1037–1049.

Goldenberg, S. and O'Brien, J. 1981. Time and space variability of the tropical Pacific wind stress. *Mon. Wea. Rev.* **109**, 1190–1207.

Graham, N. E. and Barnett, T. P. 1993. ENSO and ENSO related predictability, Part II: Northern hemisphere 700 mb height predictions based on a hybrid coupled ENSO model. *J. Clim.*, in press.

Graham, N. E., Barnett, T. P., Wilde, R., Ponater, M. and Schubert, S. 1994. On the roles of tropical and mid-latitude SSTs in forcing interannual to interdecadal variability in the winter Northern Hemisphere circulation. *J. Clim.*, in press.

Graham, N. E., Michaelsen, J. and Barnett, T. P. 1987. An investigation of the ENSO cycle with statistical models, Part II: Model results. *J. Geophys. Res.* **92**, 14271–14290.

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

Harnack, R., Harnack, J. and Lanzante, J. 1986. Seasonal temperature predictions using jackknife approach with an intraseasonal variability index. *Mon. Wea. Rev.* **114**, 1950–1954.

Horel, J. and Wallace, J. 1981. Planetary scale atmospheric phenomena associated with the southern oscillation. *Mon. Wea. Rev.* **109**, 813–829.

Latif, M. 1987. Tropical ocean circulation experiments. *J. Phys. Oceanogr.* **17**, 246–263.

Latif, M., Biercamp, J., von Storch, H., McPhaden, M. J. and Kirk, E. 1990. Simulation of ENSO-related surface wind anomalies with an atmospheric GCM forced by observed SST. *J. Clim.* **3**, 509–521.

Latif, M., Maier-Reimer, E. and Olbers, E. 1985. Climate variability studies with a primitive equation model of the equatorial Pacific. In: *Coupled ocean-atmosphere models*, ed. J. C. J. Nihoul. Elsevier Oceanographic Series, vol. 40, Amsterdam, The Netherlands.

Latif, M., Sterl, A., Maier-Reimer, E. and Junge, M. M. 1993. Structure and predictability of the El Niño/Southern Oscillation phenomenon in a coupled ocean-atmosphere general circulation model. *J. Clim.* **6**, 700–708.

Livezey, R. and Chen, W. 1983. Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.* **111**, 46–59.

Miller, A. J., Barnett, T. P. and Graham, N. E. 1993.

A comparison of some tropical ocean models: Hindcast skill and el Niño evolution. *J. Phys. Oceanogr.* **23**, 1567–1591.

Miyakoda, K., Rosati, A. and Gudgel, R. 1993. Towards the GCM el Niño simulation. NATO ASI Series: *Global climate change*, in press. Springer-Verlag, Berlin.

Miyakoda, K., Sirutis, J., Rosati, A. and Derber, J. 1988. Experimental seasonal forecasts with an air-sea model: preliminary results. *Proc. of US-PRC Internat. TOGA Symp. at Beijing*, ed. J. P. Chao and J. A. Young, pp. 417–432.

Miyakoda, K., Sirutis, J., Rosati, A. and Gudgel, R. 1989. Some activities at GFDL relevant to COARE. *Proc. Workshop on Japanese-coupled ocean atmosphere response experiments*, 23–24 October 1989. Meteorol. Res. Rep., Geophys. Inst., Univ. of Tokyo, ed. A. Sumi, pp. 93–108.

Neelin, J. D., Latif, M., Alaart, M. A. F. et al. 1992. Tropical air-sea interaction in general circulation models. *Climate Dynamics* **7**, 73–104.

Pacanowski, R. C. and Philander, S. G. H. 1981. Parameterization of vertical mixing in numerical models of tropical oceans. *J. Phys. Oceanogr.* **11**, 1443–1451.

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

Palmer, T. N., Molteni, F., Mureau, R., Buizza, R., Chapelet, P. and Tribbia, J. 1992. Ensemble prediction. *1992 ECMWF Technical Memorandum*, no. 188.

Roeckner, E., Arpe, K., Bengtsson, L., Brinkop, S., Dumenil, L., Esch, M., Kirk, E., Lunkeit, F., Ponater, M., Rockel, B., Sausen, R., Schlese, U., Schubert, S. and Windelband, M. 1992. Simulation of the present-day climate with the ECHAM model: Impact of model physics and resolution. *Max Planck Institut fur Meteorologie, Rep. no. 93*, October, 1992.

Ropelewski, C., Janowiak, J. E. and Halpert, M. S. 1992. *The climate anomaly monitoring system, "CAMS"*. NOAA/National Weather Service, National Meteorological Center, Climate Analysis Center, Washington, DC 20233.

Ropelewski, C. and Halpert, M. 1986. North American precipitation and temperature patterns associated with the El Niño-Southern Oscillation (ENSO). *Mon. Wea. Rev.* **114**, 2352 ff.

Ropelewski, C. and Halpert, M. 1989. Precipitation patterns associated with the high index phase of the Southern Oscillation. *J. Clim.* **2**, 268–284.

Sherwood, S., Ramanathan, V., Barnett, T. P., Tyree, M. and Roeckner, E. 1993. Response of a GCM to radiative forcing of tropical clouds. *J. Geophys. Res.-Atmospheres*, submitted.

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

Zebiak, S. and Cane, M. 1987. A model el Niño/Southern oscillation. *Mon. Wea. Rev.* **115**, 2262–2278.