

# Dynamical Seasonal Prediction



J. Shukla,\* J. Anderson,+ D. Baumhefner,# C. Brankovic,@ Y. Chang,&  
E. Kalnay,\*\* L. Marx,\* T. Palmer,@ D. Paolino,\* J. Ploshay,+  
S. Schubert,& D. Straus,\* M. Suarez,& and J. Tribbia#

## ABSTRACT

Dynamical Seasonal Prediction (DSP) is an informally coordinated multi-institution research project to investigate the predictability of seasonal mean atmospheric circulation and rainfall. The basic idea is to test the feasibility of extending the technology of routine numerical weather prediction beyond the inherent limit of deterministic predictability of weather to produce numerical climate predictions using state-of-the-art global atmospheric models. Atmospheric general circulation models (AGCMs) either forced by predicted sea surface temperature (SST) or as part of a coupled forecast system have shown in the past that certain regions of the extratropics, in particular, the Pacific–North America (PNA) region during Northern Hemisphere winter, can be predicted with significant skill especially during years of large tropical SST anomalies. However, there is still a great deal of uncertainty about how much the details of various AGCMs impact conclusions about extratropical seasonal prediction and predictability.

DSP is designed to compare seasonal simulation and prediction results from five state-of-the-art U.S. modeling groups (NCAR, COLA, GSFC, GFDL, NCEP) in order to assess which aspects of the results are robust and which are model dependent. The initial emphasis is on the predictability of seasonal anomalies over the PNA region. This paper also includes results from the ECMWF model, and historical forecast skill over both the PNA region and the European region is presented for all six models.

It is found that with specified SST boundary conditions, all models show that the winter season mean circulation anomalies over the Pacific–North American region are highly predictable during years of large tropical sea surface temperature anomalies. The influence of large anomalous boundary conditions is so strong and so reproducible that the seasonal mean forecasts can be given with a high degree of confidence. However, the degree of reproducibility is highly variable from one model to the other, and quantities such as the PNA region signal to noise ratio are found to vary significantly between the different AGCMs. It would not be possible to make reliable estimates of predictability of the seasonal mean atmosphere circulation unless causes for such large differences among models are understood.

## 1. Introduction

Prediction of seasonal mean rainfall and circulation in the atmosphere for different regions of the globe is of great scientific and societal interest. For example, if the major floods of 1993 and 1995 and the drought of 1988 in the United States had been predicted even one season in advance, there would have been substantial savings for the U.S. economy. Likewise, the agrarian societies in the tropical belt in general, and in the monsoonal regions in particular, could benefit enormously from reliable predictions of rainfall for the coming rainy season. This paper describes a multi-institution joint study project, to be referred to as

---

\*Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland.

+Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey.

#National Center for Atmospheric Research, Boulder, Colorado.

@European Centre for Medium-Range Weather Forecasting, Reading, United Kingdom.

&Goddard Space Flight Center, Greenbelt, Maryland.

\*\*National Centers for Environmental Prediction, Camp Springs, Maryland.

*Corresponding author adress:* Prof. J. Shukla, George Mason University, Center for Ocean–Land–Atmosphere Studies, 4041 Powder Mill Rd., Suite 302, Calverton, MD 20705-3106.

E-mail: shukla@cola.iges.org

In final form 3 March 2000.

©2000 American Meteorological Society

Dynamical Seasonal Prediction (DSP), to investigate the predictability of seasonal mean climate anomalies. The long-term purpose of this project is to carry out a comprehensive study to determine the predictability of seasonal anomalies, especially in the extratropics, and more specifically in the United States. This project initially will be focused on the study of predictability of seasonal mean anomalies only about 15 days in advance. In other words, with the knowledge of the state of the global climate system today, to what extent can we predict the seasonal mean atmospheric circulation and rainfall beyond 15 days? Since the theoretical limit of prediction of day-to-day weather is about two weeks, close to the actual upper limit of skill for current numerical weather prediction (NWP) models, we investigate the predictability of seasonal mean climate anomalies beyond the theoretical upper limit of deterministic predictability of weather, and study the predicted average for days 16 through 105.

The scientific underpinnings of this project are based on a large number of GCM sensitivity studies, which have shown that anomalous boundary conditions of sea surface temperature (SST), sea ice, snow cover, soil wetness, and other land surface conditions have significant influence on seasonal mean circulation and rainfall anomalies (Charney and Shukla 1981; Shukla and Wallace 1983; Livezey et al. 1996; Barnett et al. 1997; Kumar and Hoerling 1998; Anderson et al. 1999; Fennessy and Shukla 1999). A number of operational centers are now making predictions of seasonal mean circulation anomalies a season or more in advance using a combination of dynamical and statistical models. It has been shown (Anderson et al. 1999) that statistical models are currently competitive with the dynamical models indicating serious deficiency of the dynamical models. However, the existence of operational prediction systems and the higher skill of statistical predictions does not mean that the scientific basis for such predictions is well understood. The results of atmospheric general circulation model (AGCM) sensitivity to prescribed SST anomalies during the past have shown varieties of responses and there is still a great deal of uncertainty about how much the details of various AGCMs impact conclusions about seasonal predictability. One of the main motivations of the DSP project is to reduce this uncertainty by using several state-of-the-art dynamical models with identical initial and boundary conditions. Past research has already shown that the nature of boundary forced predictability is remarkably different for the tropical and the extratropical circulation. While the

tropical response is relatively robust and predictable, the extratropical response is sensitively dependent upon the initial conditions (Shukla 1981). Parts of the tropical seasonal mean response to anomalous SST forcing are highly reproducible and characterized by large signal-to-noise ratios. However, the extratropical response has generally a much reduced ratio, due in part to much larger amounts of noise related to mid-latitude synoptic scale systems. However, in cases of extremely strong anomalous forcing, certain parts of the extratropical circulation, in particular the Pacific–North American (PNA) region during Northern Hemisphere winter, is also highly reproducible and has high signal to noise ratio (Shukla 1998). This is in contrast to the calculations with the previous generation of GCMs for which SST forced variability was indistinguishable from the variability produced by internal dynamics (Chervin 1986; Lau 1997), which indicated poor prospects for dynamical seasonal prediction in midlatitudes.

In keeping with the primary goal of exploring extratropical seasonal predictability, this project utilizes several state-of-the-art models and large ensembles of seasonal integrations to make a definitive statement on the problems and promises of routine dynamical seasonal predictions. Since the large-scale tropical flow is strongly determined by the boundary conditions, it should be possible to make routine dynamical seasonal predictions in the Tropics if the boundary conditions can be predicted. However, since the internal dynamics of extratropical flows are chaotic, the seasonal mean has little, if any, predictability in the absence of anomalous boundary conditions. The important question addressed by this study is: with perfect predictions and specification of anomalous SST boundary conditions, how well and with what degree of confidence can the extratropical seasonal anomalies be predicted? In the future, we would like to explore whether there is information in the initial conditions of either the atmosphere or the land surface that can also contribute to seasonal predictability.

## 2. Scientific background

It is well known from the history of NWP that estimates of predictability can be quite different for different models (Charney et al. 1966). It is only after 20–30 years of improvements in weather prediction models that the estimates of weather predictability have nearly converged to about 1.5 days of doubling time for small errors in the initial conditions (Simmons

et al. 1995). Recognizing similar model dependence of the estimates of seasonal predictability, the present work utilizes several state-of-the-art models to investigate seasonal predictability. An intercomparison of results among several models is helpful in understanding which aspects of results are robust and which are model dependent.

The ultimate objective of this project is to determine the predictability of the seasonal mean circulation using coupled ocean–land–atmosphere models. However, as a first step toward that ultimate goal, observed global SST has been used as a lower boundary condition for a number of atmospheric models. This should help determine the *upper limit* of SST forced seasonal predictability. The next step will either be to perturb the global SST field to mimic the possible errors in predicting the global SST using coupled ocean–atmosphere models, or to make predictions with the coupled models. This project is being carried out in two phases. In phase 1, global SST and sea ice are prescribed from observations for the entire duration of GCM integration. In phase 2, SST will be produced by a combination of coupled ocean–atmosphere models and statistical prediction techniques, or seasonal hindcasts will be done with coupled ocean–land–atmosphere models.

In this study (phase 1), global SST produced by Reynolds and Smith (1994) using optimal interpolation (denoted OI SST) was used. Since different GCMs have different schemes for land–atmosphere interactions, the future evolution of land conditions was determined by the GCMs. The initial conditions of land-surface conditions were also different for different GCMs, to be consistent with the respective models.

This research project is different from the Atmospheric Model Intercomparison Project (Gates et al. 1999) because all the models have used the same boundary conditions of SST, as well as the same initial conditions of global atmosphere, and model simulations have been verified against actual global observation for the particular season and the particular year. This project is focused solely on seasonal predictability issues and it utilizes ensembles that are essential to calculate signal to noise ratio, and as a continuation of this project, the next step is to go on to real predictions with coupled models. Because of the chaotic nature of the extratropical flow, predictions must be probabilistic and therefore dynamical prediction will require a large ensemble. Large ensembles will be produced by aggregating somewhat smaller ensembles from several different GCMs with different initial conditions.

### 3. Experiments

All models were integrated to produce 5- to 10-member ensembles with slightly perturbed initial conditions of the atmosphere. The initial conditions were perturbed either by adding a suitable scaled random field of model variables, or by taking observed analyses 12 or 24 h apart. Since the first 15 days of model integrations were not to be included for defining the seasonal means, the manner in which the initial conditions were perturbed was not considered to be significant. Global SST and sea ice were prescribed from observations (the OI SST) produced by Reynolds and Smith (1994) except for the Goddard Space Flight Center (GSFC) model for which runs were carried out with an older SST dataset: the so-called blended SST data (Reynolds and Marsico 1993). Compared with the OI SST, the blended SST has weaker anomalies due to excessive trimming of the observations during the SST analysis cycle (Hoerling and Kumar 1997). The impact of the differences in the SST was examined with the GSFC model for selected years with strong ENSO forcing. The results show only a modest increase in the response over the North Pacific, and do not alter the basic conclusions.

Although the goal of this project was to carry out historical seasonal forecasts for all the seasons, in this paper we present results for winter season only. The initial conditions for all models except the European Centre for Medium-Range Forecasts (ECMWF) are the observed analyses centered around 15 December, and integrations extend through the end of March. ECMWF integrations started at the end of November. Winter season mean was defined to include the months of January, February, and March. This choice of winter initial conditions is based on the observation that the extratropical circulation anomalies over North America in association with tropical SST anomalies are better organized in January–February–March (JFM) means rather than the conventional winter season defined as December–January–February (DJF) means (Anderson et al. 1999). The last two weeks of December were discarded so that the estimate of seasonal predictability of the winter season was not biased by the potential skill in the high-frequency component during the first 15 days.

### 4. Models

At present five U.S. groups have carried out ensemble seasonal forecasts for winter season using ob-

TABLE 1. Model resolutions and ensemble size for six models.

	Institution	Resolution	Years	Ensemble size	Reference
1	NCAR	T42, L19	1982–97	10	Kiehl et al. (1998)
2	COLA	R40, L18	1982–98	9	DeWitt (1996)
3	GSFC	2°lat × 2.5°long, L43	1981–95	9	Suarez and Takacs (1995)
4	GFDL	T42, L18	1980–96	10	Anderson and Stern (1996)
5	NCEP	T62, L28	1983–96	5	Kalnay et al. (1996)
6	ECMWF	T63, L31	1980–93	9	Gibson et al. (1997)

served initial and boundary conditions. Results from the ECMWF model, which is one of the models participating in a similar European project called PROVOST (Predictability of Climate Variations on Seasonal-to-Interannual Timescales) are also included in this paper. The model resolutions used for these integrations and references where descriptions of these models can be found are given in Table 1.

The National Centers for Environmental Prediction (NCEP) model is the same model that was used for the NCEP/National Center for Atmospheric Research (NCAR) Reanalysis, and the NCAR model is the CCM3 atmospheric component of the climate system model. The GSFC model is an early implementation of version 2 of the Goddard Earth Observing System (GEOS-2) model of the Data Assimilation Office. The final version of GEOS-2 includes an interactive land surface model, and a moist turbulence scheme. Table 1 also gives the years for which winter season (JFM) hindcasts have been produced, and the ensemble size (number of hindcasts per calendar winter) for each model.

### 5. Results

All modeling groups are preparing separate comprehensive reports on the results of their respective models (Anderson and Ploshay 2000; Chang et al. 2000; Shukla et al. 2000; Bankovic and Palmer 2000), and these papers are being published in a special volume of the *Quarterly Journal of the Royal Meteorological Society*. We present here only a few results to highlight the fact that all models show significant skill

in predicting seasonal mean extratropical circulation anomalies in the presence of large scale tropical SST anomalies associated with ENSO. A remarkable aspect of these model intercomparisons is the large model-to-model difference in the partitioning of the total variance between the variance of the ensemble means, and the intra-ensemble variance. Since this partitioning is the main factor that determines the magnitudes of SST forced “signal” and internal “noise,” and hence the estimates of the potential predictability for any model, it is necessary to understand the reasons for large model-to-model differences before a definitive conclusion can be drawn about the predictability of the seasonal mean extratropical circulation. To reduce ambiguity in the interpretation of results, the same 11 years (1983–93) were used for comparison of model results.

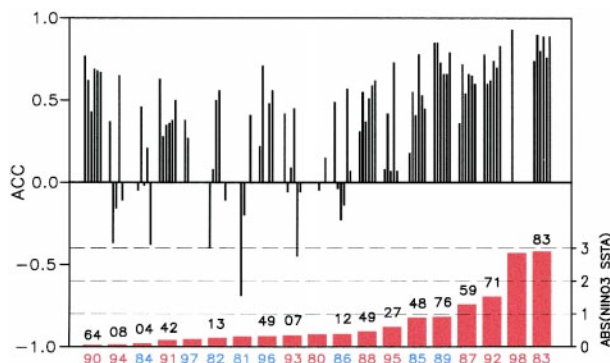


FIG. 1. ACC for ensemble mean JFM 500-hPa height forecasts over PNA region for six models in the following order: NCAR, COLA, GSFC, GFDL, NCEP, ECMWF. The red bars give the absolute value of Niño-3 SST anomalies.

a. Verification over the Pacific North-American (PNA) region

Figure 1 shows the anomaly correlation coefficient (ACC) over the PNA region ( $15^{\circ}$ – $70^{\circ}$ N,  $180^{\circ}$ – $60^{\circ}$ W) for ensemble mean JFM average geopotential height anomalies at 500 hPa. This figure presents a summary of all available integrations from each participating group. Along the abscissa, the years are arranged in the ascending order of the absolute value of SST anomaly over the Niño-3 region. Years in blue and red indicate negative and positive SST anomalies, respectively. The six black bars represent the ACC for the six models in the same order [NCAR, Center for Ocean–Land–Atmosphere Studies (COLA), GSFC, Geophysical Fluid Dynamics Laboratory (GFDL), NCEP, ECMWF] for each year. The numbers on the top of each red bar for each year show the average value of ACC for the six models. For some years, ACC is not available for all six models. The procedure for calculating ACC is given in the appendix.

It is remarkable that for the 8 yr in the plot with absolute value of Niño-3 greater than or equal to  $0.5^{\circ}\text{C}$ , all models show positive ACC and for Niño-3 greater than or equal to  $1.0^{\circ}\text{C}$ , ACC for all models is 0.5 or higher. It is also remarkable that there is a large model-to-model variability especially when SSTA is not large. This is more clearly seen in Fig. 2 and Fig. 3, which show ACC and root-mean-square-error (rmse), respectively, for 500-hPa height over the PNA region, for six models, for all the common years, for all the members of the ensemble. The order of the years is based on decreasing value of average ACC for all six models. The solid bars denote the values of ACC and rmse for ensemble mean hindcasts. An important feature of these figures is a large variability among different members of ensemble for the same SST boundary condition. However, even for models

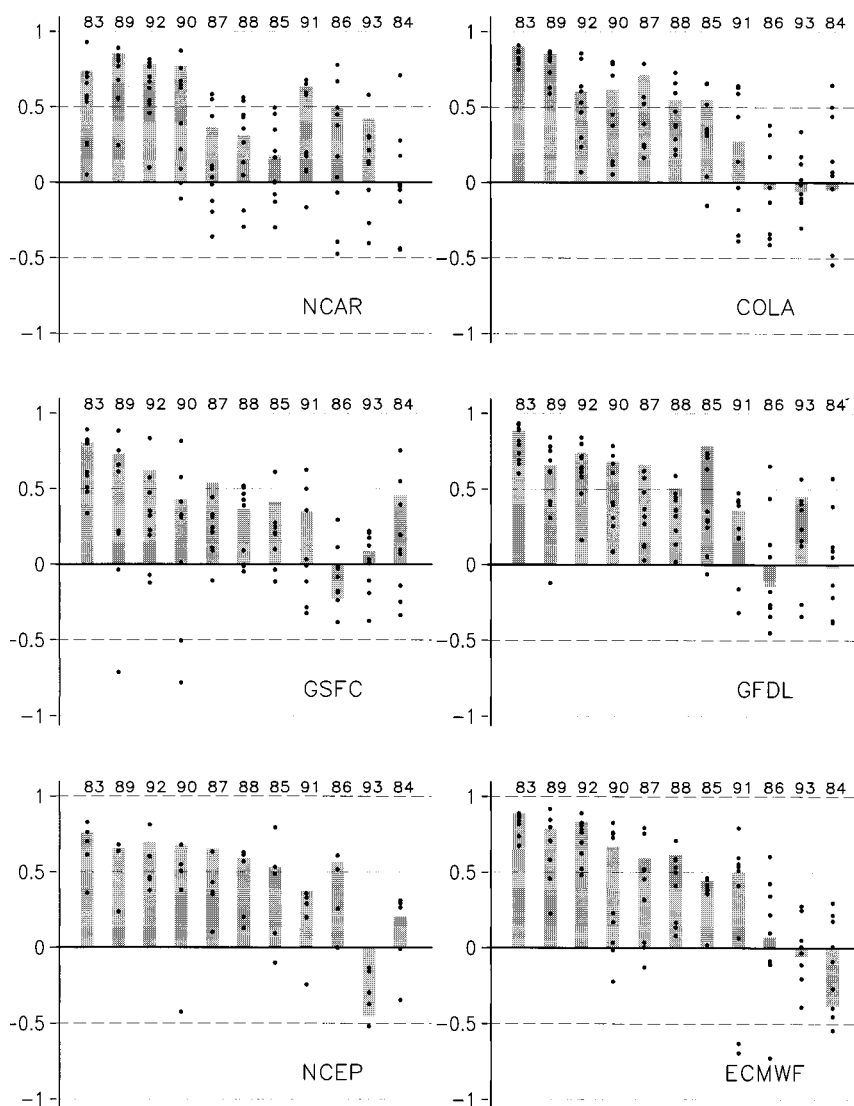


FIG. 2. ACC over PNA region for JFM mean 500-hPa height forecasts. Solid bars are the ACC for the ensemble mean, and the dots represent ACC for each member of the ensemble.

with lower ACC for ensemble mean forecasts, there are a few members of the ensemble for which ACC is quite high.

The years 1990 and 1991 do not have strong forcing by our measure (SST anomaly amplitude), however, they have consistently higher anomaly correlations. We do not know if it happened by chance that the observed response was consistent with some common systematic response in GCMs (which is highly unlikely), or there may be other external forcing (other than Niño-3 SST) for the atmosphere that we have not examined.

Table 2 gives a summary of ensemble mean forecast verification for all the six models for the PNA region and the European region. Average values of

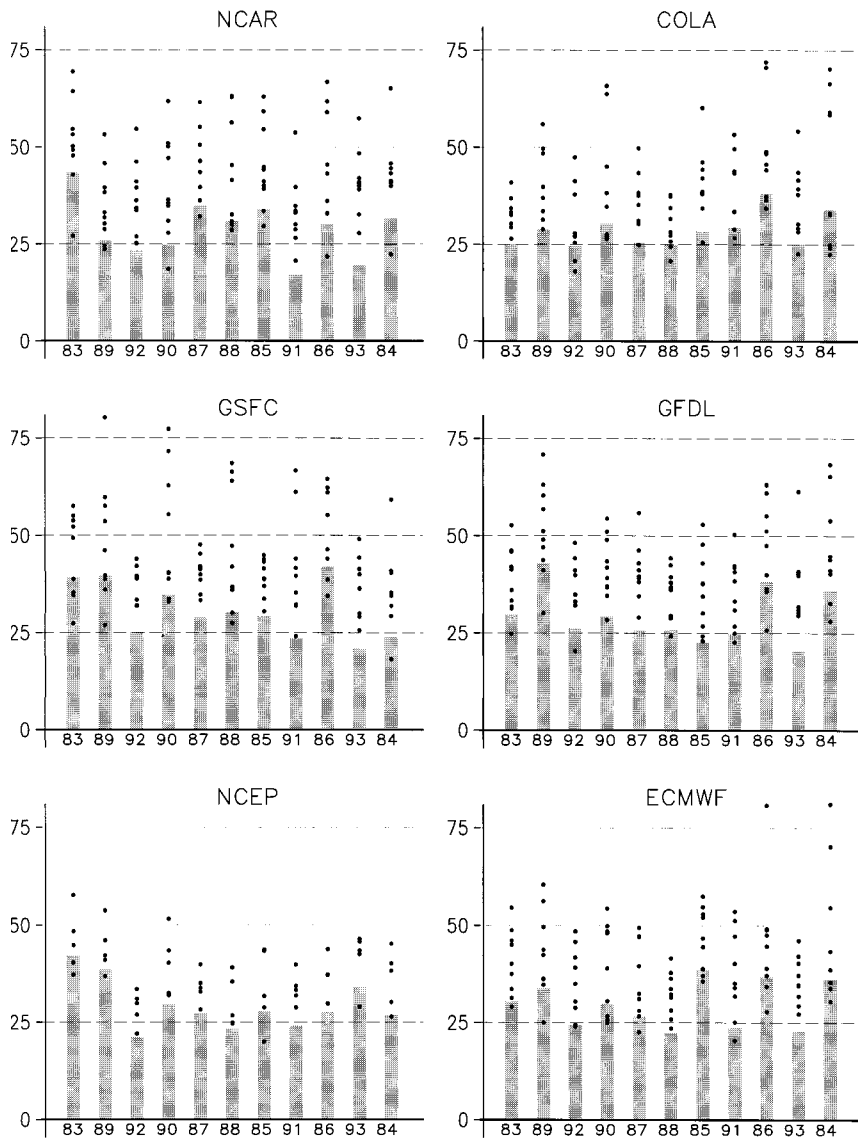


FIG. 3. Same as Fig. 2 for rmse (m).

ACC and rmse for ENSO years and non-ENSO years are presented separately. A remarkable aspect of this table is that the average ACC and rmse for all cases have a small range of model-to-model variability (0.42–0.51 for ACC, and 29–31 m for rmse) for the PNA region. The intramodel differences are much larger for the European region, and the differences become even larger when ACC and rmse are averaged separately for ENSO and non-ENSO years. The years 1983, 1985, 1987, 1989, 1992 for which absolute value of Niño-3 is greater than  $0.5^{\circ}\text{C}$  are defined as ENSO years. It will be shown later that the ensemble mean ACC for any model depends on the manner in which the model partitions the total variance between the SST forced signal, and the internal noise.

when  $\sigma_{\text{signal}}^2$  for tropical precipitation is relatively large, the variance of extratropical height especially at 200 hPa is also large. However, this relationship breaks down for noise in tropical precipitation and extratropical height. For example, the NCAR model has one of the lowest values for  $\sigma_{\text{noise}}^2$  for tropical precipitation ( $2 \text{ mm}^2 \text{ day}^{-2}$ ), but the highest value for  $\sigma_{\text{noise}}^2$  for 200-hPa extratropical height. Differences among the models become even larger when only the five ENSO years (1983, 1985, 1987, 1989, 1992) are used to calculate the variances. For these calculations, the GSFC model results using OI SST are shown.

Figure 4 shows maps of the total variance for five ENSO years for geopotential height at 500 hPa over the PNA region for six models and corresponding

*b. Total variance, SST forced variance (signal), and internal dynamics variance (noise)*

The total variance is the sum of the signal plus noise variances. The variance of the ensemble average seasonal means among all the years, modified appropriately for sampling, is referred to as SST forced variance or signal. The variance within each ensemble averaged for all the years is referred to as the internal dynamics variance, or noise. The procedure for calculating variances is given in the appendix.

Table 3 gives the total variance ( $\sigma_{\text{total}}^2$ ), SST forced variance ( $\sigma_{\text{signal}}^2$ ) and internal dynamics noise ( $\sigma_{\text{noise}}^2$ ) for the common years (1983–93) for all six models for precipitation and geopotential height at 500 and 200 hPa. The model generated total (area averaged) variance for tropical precipitation ranges from  $7 \text{ mm}^2 \text{ day}^{-2}$  for the GSFC model to  $46 \text{ mm}^2 \text{ day}^{-2}$  for the ECMWF model. While there is some uncertainty in estimating changes in precipitation over the oceanic regions, the total variance calculated from observed precipitation is  $14 \text{ mm}^2 \text{ day}^{-2}$ .

From Table 3, it is clear that

TABLE 2. Mean Anomaly Correlation Coefficient (ACC) and mean root-mean-square error (rmse) for JFM mean 500-hPa geopotential height hindcasts compared to observations (NCEP reanalysis) for (A) PNA region (15°–70°N, 180°–60°W), and (B) Europe (35°–75°N, 12.5°W–42.5°E). ENSO = mean for ENSO years 1983, 1985, 1987, 1989, 1992; Other = mean for non-ENSO years 1984, 1986, 1988, 1990, 1991, 1993; All = mean for all years 1983–93. NINO3 is mean of absolute value of Niño-3 SST anomaly.

<b>(A) PNA region (15°–70°N, 180°–60°W)</b>															
	<b>ACC × 100</b>								<b>Rmse (m)</b>						
	NCAR	COLA	GSFC	GFDL	NCEP	ECMWF	MEAN	NINO3	NCAR	COLA	GSFC	GFDL	NCEP	ECMWF	MEAN
ENSO	58	72	62	75	66	71	67	1.50	32	27	32	29	31	31	30
Other	44	22	24	31	33	24	30	0.26	26	30	29	29	28	29	29
All	50	45	42	51	48	45	47	0.83	29	29	31	29	29	30	30

<b>(B) Europe (35°–75°N, 12.5°W–42.5°E)</b>															
	<b>ACC × 100</b>								<b>Rmse (m)</b>						
	NCAR	COLA	GSFC	GFDL	NCEP	ECMWF	MEAN	NINO3	NCAR	COLA	GSFC	GFDL	NCEP	ECMWF	MEAN
ENSO	23	46	–05	57	17	22	27	1.50	44	39	50	43	47	48	45
Other	46	54	21	57	17	30	38	0.26	44	42	48	39	45	45	44
All	36	50	09	57	17	26	33	0.83	44	41	49	41	46	46	44

TABLE 3. Interannual (1983–93) Variance of JFM mean tropical precipitation and extratropical geopotential height for 200 and 500 hPa. Grid point values of variance are averaged over 10°S–10°N, 140°E–130°W for precipitation, and 25°–70°N, 60°W–180° for geopotential height.

	<b>Precipitation (mm<sup>2</sup> day<sup>-2</sup>)</b>			<b>200-hPa height (10<sup>2</sup>m<sup>2</sup>)</b>			<b>500-hPa height (10<sup>2</sup>m<sup>2</sup>)</b>		
	$\sigma^2_{\text{total}}$	$\sigma^2_{\text{signal}}$	$\sigma^2_{\text{noise}}$	$\sigma^2_{\text{total}}$	$\sigma^2_{\text{signal}}$	$\sigma^2_{\text{noise}}$	$\sigma^2_{\text{total}}$	$\sigma^2_{\text{signal}}$	$\sigma^2_{\text{noise}}$
NCAR	20	18	2	45	11	34	21	6	15
COLA	31	27	4	42	22	20	24	13	11
GSFC	7	5	2	31	7	24	19	4	15
GFDL	22	15	7	41	18	23	24	11	13
NCEP	16	12	4	24	10	14	14	6	8
ECMWF	46	39	7	46	26	20	29	17	12
Observed	14			30			18		

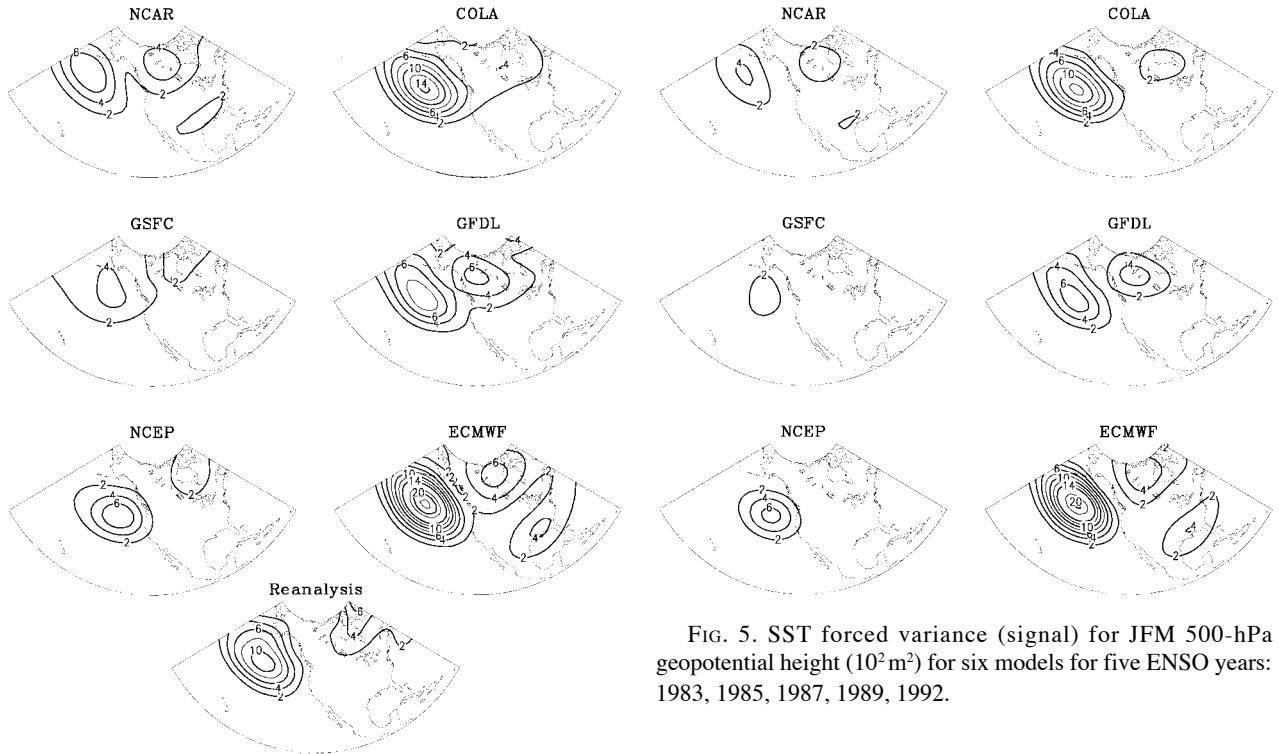


FIG. 4. Total variance for JFM 500-hPa geopotential height ( $10^2 \text{ m}^2$ ) for six models and observations from NCEP Reanalysis for five ENSO years: 1983, 1985, 1987, 1989, 1992.

observations. The PNA region has the highest values of interannual variance of seasonal (JFM) mean height over the globe. There is considerable model-to-model variability, especially over the north Pacific region, the region of highest interannual variability. The differences among the models become even larger when the SST forced variance (signal) and internal dynamics variance (noise) are calculated separately. Figures 5 and 6 (note different contour intervals) show the corresponding values of  $\sigma^2_{\text{signal}}$  and  $\sigma^2_{\text{noise}}$ . Values of noise in geopotential height for four models (COLA, ECMWF, GFDL, GSFC) are similar to each other, but the noise is considerably larger for CCM3 and smaller for NCEP. The SST forced signal is clearly the lowest for the GSFC model and relatively high for COLA and ECMWF.

Figures 7, 8, 9 show the total variance, signal and noise, for JFM mean rainfall over the tropical Pacific region for the ENSO years. The corresponding observed variance from Xie and Arkin (1994) is shown in Fig. 10. The model-to-model variability is again found to be quite large. In particular, it is seen that the ECMWF model has the largest signal for rainfall and also the largest signal for geopotential height, whereas

FIG. 5. SST forced variance (signal) for JFM 500-hPa geopotential height ( $10^2 \text{ m}^2$ ) for six models for five ENSO years: 1983, 1985, 1987, 1989, 1992.

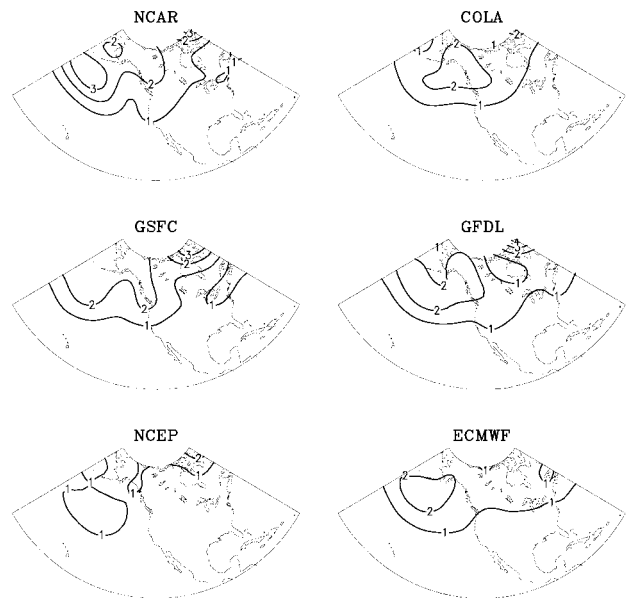


FIG. 6. Internal dynamics variance (noise) for JFM 500-hPa geopotential height ( $10^2 \text{ m}^2$ ) for six models for five ENSO years: 1983, 1985, 1987, 1989, 1992.

the GSFC model has the smallest signal for tropical rainfall and also the smallest signal for geopotential height.

Figure 11 shows the 1983 minus 1989 difference for ensemble mean JFM height anomalies (meters) for 500 hPa for six models and for observations (NCEP



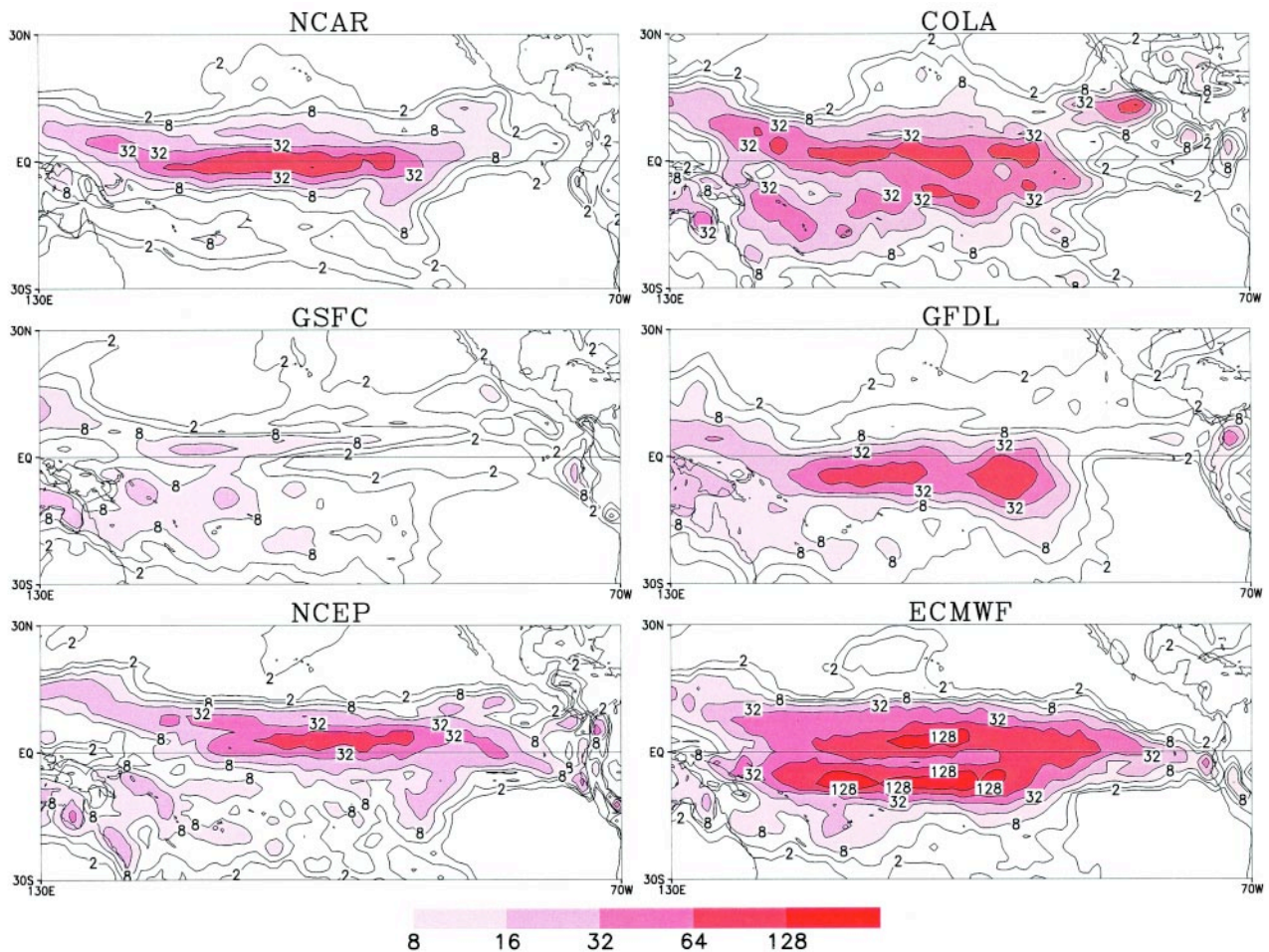


FIG. 7. Total variance for rainfall ( $\text{mm}^2 \text{day}^{-2}$ ) for six models for five ENSO years: 1983, 1985, 1987, 1989, 1992.

Reanalysis). Consistent with the earlier figures, it is seen that the models with higher SST forced variance have larger simulated height anomalies, whereas the models with the lower SST forced variance have smaller simulated height anomalies. The simulated height anomaly over the North Pacific region is higher than observed for the ECMWF model and only about half of the observed values for the GSFC and NCAR models.

### c. Probability distribution for variance explained by SST

The variability of winter season (JFM) means of 500-hPa height related to tropical SST anomalies has been estimated separately for six models for the 11 common winters of 1983–93. The idea is to compile a large number (100) of samples of GCM integrations, where for each GCM a sample is obtained by randomly drawing one ensemble member for each calendar winter. (Each sample is thus a series of 11 seasonal means, comparable to observations.) The midlatitude height

variance explained by tropical SST anomalies in the Pacific is calculated for each sample by linear regression on a base series of SST anomalies.

The base time series of tropical SST anomalies is obtained from singular value decomposition analysis of 30 winter means (1968–97) of 500-hPa height ( $120^\circ\text{E}$ – $60^\circ\text{W}$ ,  $20^\circ$ – $75^\circ\text{N}$ ) and SST ( $120^\circ\text{E}$ – $80^\circ\text{W}$ ,  $20^\circ\text{S}$ – $20^\circ\text{N}$ ). The grid used for the height field was an equal area grid, while that used for the SST field was a regular  $2.5^\circ$  grid. The first mode explains 88% of the squared covariance, and the SST pattern associated with it (not shown) has a broad center at  $130^\circ\text{W}$  on the equator. The associated time series for the 11 winters (1983–93) is used as the basis for the regression.

The regression of height on the base time series is carried out for each of 100 samples for each GCM. The geographical distribution of the percentage of explained variance averaged over all 100 samples (Fig. 12) is quite different for different models, especially in the north Pacific region. The probability dis-

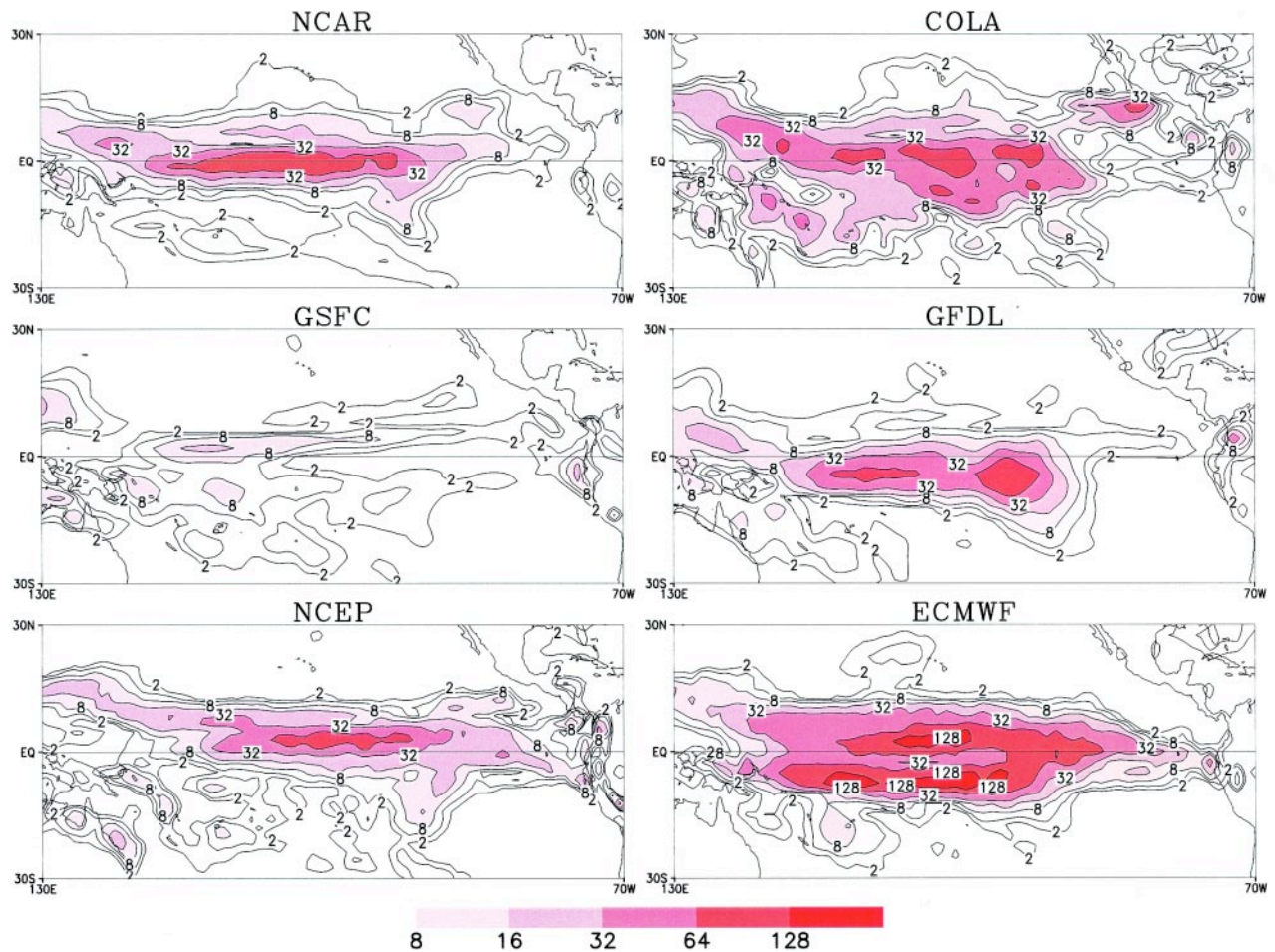


FIG. 8. SST forced variance (signal) for rainfall ( $\text{mm}^2 \text{day}^{-2}$ ) for six models for five ENSO years: 1983, 1985, 1987, 1989, 1992.

tribution among the samples of the area average of the explained variance (over the area  $20^\circ\text{--}75^\circ\text{N}$ ,  $180^\circ\text{--}30^\circ\text{E}$ ) is shown in Fig. 13 for the GCMs, and can be compared to the single number (27.5%) for the observations. It is seen that the peak of the explained variance ranges from 14% to 34% among different models. Although the sample sizes were different for different models, this should not be a problem because it only affects our confidence in the estimates, and not the estimates themselves.

## 6. Summary and discussion

Comparison of ACC and rmse with respect to observations, a common metric for verification of weather prediction models, shows that ensemble mean predictions of JFM mean-height over the PNA region for the 11-yr period have comparable skill for all the six models. All models show higher forecast skill for

the ENSO years compared to the non-ENSO years. While some of these conclusions might be a consequence of small sample size (11), it is incontrovertible that the winter season mean circulation over the PNA region is highly predictable during years of large SST anomalies in the tropical Pacific Ocean. The reproductibility of the predicted circulation anomalies among different members of an ensemble for a given SST anomaly is quite different for different models.

Recently, there has been rapidly expanding interest in the use of multimodel ensembles to improve the skill of prediction (Krishnamurti et al. 1999). This topic is beyond the scope of this paper which is to understand differences in GCM simulation behavior. The model outputs of this project provide a good dataset to test ideas on multimodel ensembling. It could also be possible to use the ensembles to make probabilistic forecast evaluation and make forecasts of more than just the mean response. Such an evaluation will be more appropriate for subsequent investigations.

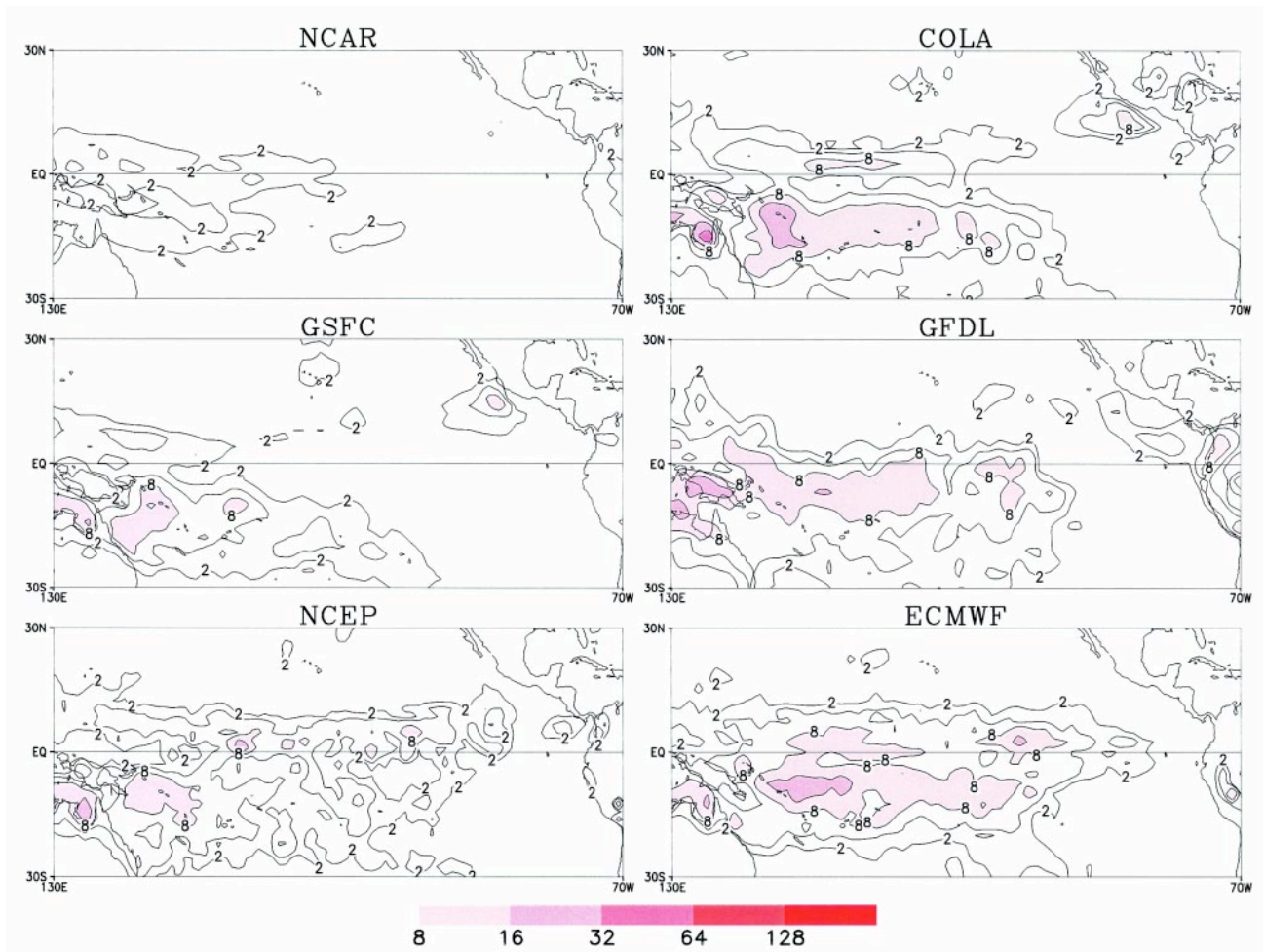


FIG. 9. Internal dynamics variance (noise) for rainfall ( $\text{mm}^2 \text{day}^{-2}$ ) for six models for five ENSO years: 1983, 1985, 1987, 1989, 1992.

It is found that in spite of utilizing identical global SST boundary conditions and global atmospheric initial conditions, different models show quite different levels of interannual variability. It is assumed that an ensemble average of five to ten integrations with identical SST for any given year is a reasonably good measure of the SST forced response. Therefore, the differences in the simulation of signal may be attributed to the differences in the model's ability to generate different diabatic heat sources for the same SST anomaly, and to produce different local and remote responses for the same diabatic heat sources. However, to the extent that the seasonal mean extratropical response of diabatic heating is strongly affected by the transient variability in each model, and by the accuracy of the model climatology, it is difficult to estimate the relative roles of diabatic heating and internal dynamics in producing different SST forced signals in different models.

Likewise, it is reasonable to assume that the departures from the ensemble mean for each year are a good

measure of the internal dynamics variability, since the SST forced signal is the ensemble mean that is taken out. Therefore, the differences in the simulation of noise may be attributed primarily to the differences in the dynamics of each model. Since all the models used in this study are state-of-the-art models, and since it

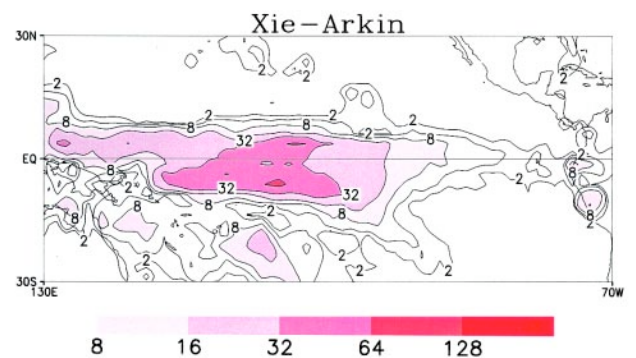


FIG. 10. Total variance for rainfall ( $\text{mm}^2 \text{day}^{-2}$ ) for observations from Xie and Arkin (1996) for five ENSO years: 1983, 1985, 1987, 1989, 1992.

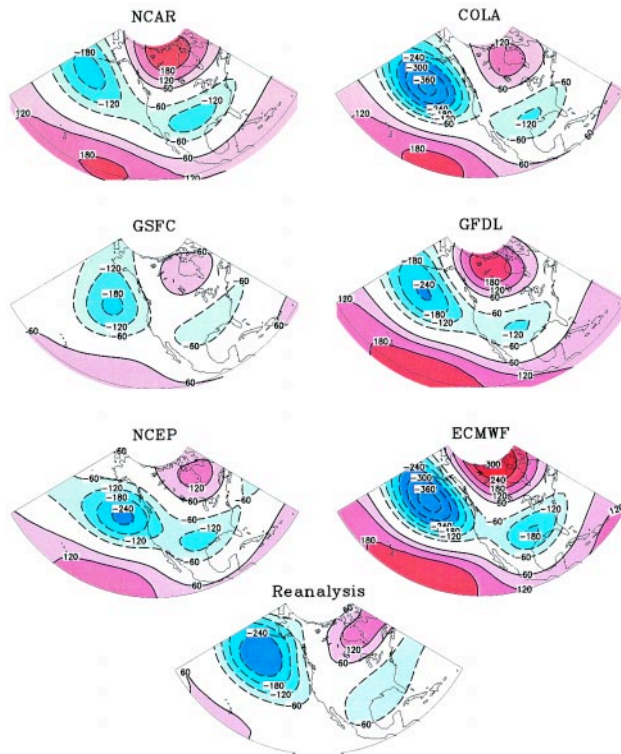


FIG. 11. Difference (1983 minus 1989) of ensemble mean JFM height at 500 hPa for six models and observations from NCEP Reanalysis.

is generally accepted that the accuracy of the dynamics in these models is quite comparable, we cannot explain why the internal dynamics variability (noise) varies so much between models. In particular, it should be noted that although the NCEP model and the ECMWF model have the same horizontal resolution, and both are derived from their respective operational weather prediction models with comparable skills in medium range weather forecasts, the simulations of seasonal variability by the two models is quite different. Likewise, the NCEP and the COLA model have an identical dynamical core, but different physics, yet the internal dynamics variability is quite different. These qualitative intercomparisons lead us to suggest that the treatment of physical processes is perhaps the most important factor in determining the internal variability in these models.

The model simulated interannual variability of tropical rainfall is generally higher than the observed. It is difficult to resolve whether this is a model deficiency or it is due to inadequacy of observations in estimating interannual variability of rainfall over the tropical oceanic regions. Since the simulated height anomalies over the PNA region are remarkably simi-

lar to the observed, especially during the ENSO years, it is likely that the observations underestimated the interannual variability of rainfall, unless, of course, some other (unknown) model deficiency compensates for unrealistic large changes in tropical rainfall. We would like to address these questions later when seasonal forecasts for other seasons are completed. The present study does not allow us to estimate the limits of predictability of seasonal averages for operational climate prediction because we have not considered the uncertainty in predicting SST. This will be addressed in phase 2, when coupled ocean-atmosphere models will be used for seasonal forecasts.

Another factor that we have not taken into account is the initial conditions of soil wetness and other land surface properties. It is not well understood how the variability of the winter season mean circulation over the PNA region is influenced by the land surface initial conditions. However, more controlled model integrations are needed to get an insight into the reasons for large model-to-model variability. A more detailed diagnosis of the structure and variability of the diabatic heating field and the transient variability in the models will be needed to understand the mechanisms of interannual variability.

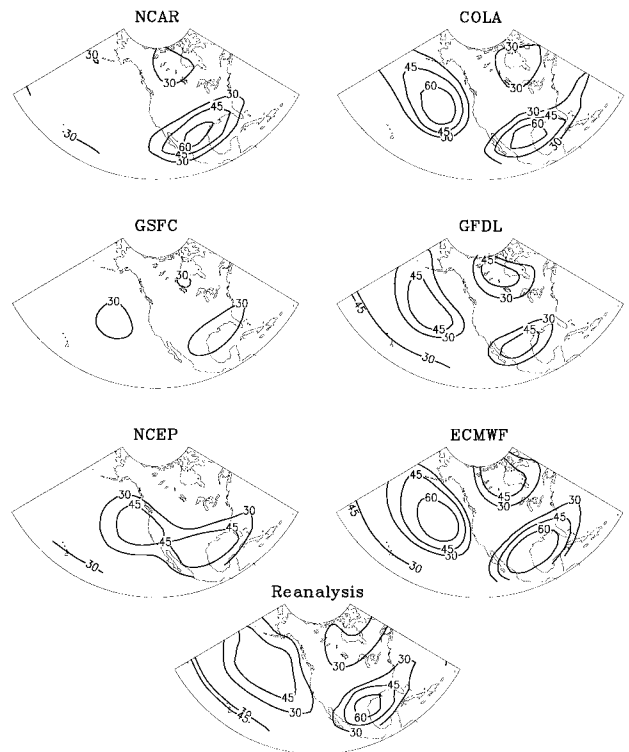


FIG. 12. Percentage of JFM 500-hPa height variance explained by tropical SST.

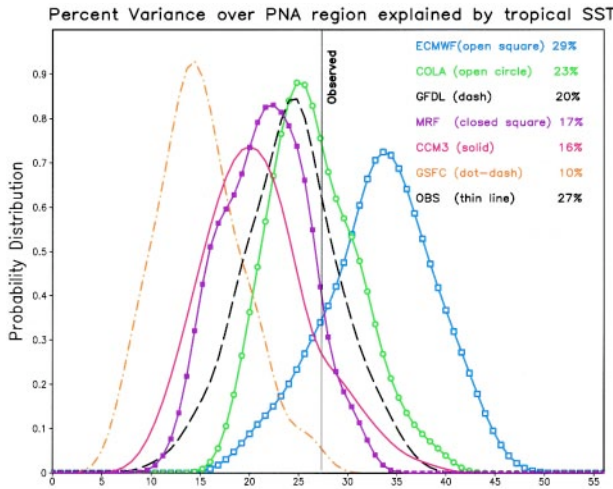


FIG. 13. Probability distribution of percent variance averaged over PNA region explained by tropical SST.

Such an understanding is essential to define the limits of predictability of seasonal averages. Generally, the signal-to-noise ratio is considered to give a measure of predictability. It can be seen from the preceding results that even for the identical SST anomalies and the same eleven year period, the signal-to-noise ratio over the PNA region can be different by a factor of 8 among different models. Such large divergence in the model-dependent estimates of predictability of seasonal means is not dissimilar to the situation about 35 years ago when different models gave quite different estimates of weather predictability.

*Acknowledgments.* The COLA staff was supported by grants from NOAA (NA76-GP0258), NSF (ATM-9321354), and NASA (NAGW-5213). The computer time for integrations of NCAR, COLA, and NCEP models was provided by the climate system laboratory at NCAR.

## Appendix: Measures of error and variance

### a. Calculation of ACC and rmse

For a seasonal mean climate variable  $x_{ij}$  (height, rainfall, etc.) for  $N$  years ( $i = 1, 2, \dots, N$ ), and  $n$  ensemble members ( $j = 1, 2, \dots, n$ ), ensemble mean  $\bar{x}_p$ , and climatological (ensemble) mean  $\bar{x}$  are defined as

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}; \quad \bar{x} = \frac{1}{nN} \sum_{i=1}^N \sum_{j=1}^n x_{ij}.$$

Rmse is calculated between the model anomaly and

observed anomaly. The model and observed seasonal anomalies for a given year are calculated by subtracting the model and observed seasonal climatologies, respectively. This is nearly equivalent to removing monthly mean climatology and calculating the seasonal mean of monthly anomalies, except for the slight-difference in the numbers of days in each month. The model climatology for each model is calculated using all available model integrations shown in Table 1. Before calculating the ACC for a given region, the regional mean anomaly is subtracted from anomaly values at each grid point, both for models and observations. All means and anomalies are area weighted. The JFM seasonal mean is defined as the average of 1200 UTC 1 January–0000 UTC 1 April, twice daily for both models and observations. The algebraic expressions for calculating ACC and rmse are same as given in Anderson et al. (1999).

### b. Calculation of total variance, SST forced variance (signal), and internal dynamics variance (noise)

$$\sigma_{\text{noise}}^2 = \frac{1}{N(n-1)} \sum_{i=1}^N \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$$

The SST forced variance ( $\sigma_{\text{signal}}^2$ ) is given as

$$\sigma_{\text{signal}}^2 = \sigma_{\text{EM}}^2 - \frac{1}{n} \sigma_{\text{noise}}^2.$$

Where the variance of the ensemble mean ( $\sigma_{\text{EM}}^2$ ) is

$$\sigma_{\text{EM}}^2 = \frac{1}{N-1} \sum_{i=1}^N (\bar{x}_i - \bar{x})^2.$$

Following Rowell et al. (1995), the total variance  $\sigma_{\text{total}}^2$  is given as

$$\sigma_{\text{total}}^2 = \sigma_{\text{noise}}^2 + \sigma_{\text{signal}}^2.$$

## References

- Anderson, J. L., and W. F. Stern, 1996: Evaluating the potential predictive utility of ensemble forecasts. *J. Climate*, **9**, 260–269.  
 —, and J. Ploshay, 2000: Impact of initial conditions on seasonal simulations with an atmospheric general circulation model. *Quart. J. Roy. Meteor. Soc.*, **126**, 2241–2264.

- , H. van den Dool, A. Barnston, W. Chen, W. Stern, and J. Ploshay, 1999: Present-day capabilities of numerical and statistical models for atmospheric extratropical seasonal simulation and prediction. *Bull. Amer. Meteor. Soc.*, **80**, 1349–1361.
- Barnett, T. P., K. Arpe, L. Bengtsson, M. Ji, and A. Kumar, 1997: Potential predictability and AMIP implications of midlatitude climate variability in two general circulation models. *J. Climate*, **10**, 2321–2329.
- Brankovic, C., and T. N. Palmer, 2000: Seasonal skill and predictability of ECMWF PROVOST ensembles. *Quart. J. Roy. Meteor. Soc.*, **126**, 2035–2067.
- Chang, Y., S. D. Schubert, and M. Suarez, 2000: Boreal winter predictions with the GEOS-2: The role of boundary forcing initial conditions. *Quart. J. Roy. Meteor. Soc.*, **126**, 2293–2321.
- Charney, J. G., and J. Shukla, 1981: Predictability of monsoons. *Monsoon Dynamics*, J. Lighthill and R. P. Pearce, Eds., Cambridge University Press, 99–109.
- , R. Fleagle, V. E. Lally, H. Riehl, and D. Q. Wark, 1966: The feasibility of a global observation and analysis experiment. *Bull. Amer. Meteor. Soc.*, **47**, 200–220.
- Chervin, R. M., 1986: Interannual variability and seasonal climate variability. *J. Atmos. Sci.*, **43**, 233–251.
- DeWitt, D. G., 1996: The effect of cumulus convection on the climate of COLA general circulation model. COLA Tech. Rep. 27, 58 pp. [Available from COLA, 4041 Powder Mill Rd., Calverton, MD 20705.]
- Fennessy, M. J., and J. Shukla, 1999: Impact of initial soil wetness on seasonal atmospheric prediction. *J. Climate*, **12**, 3167–3180.
- Gates, W. L., and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIPI). *Bull. Amer. Meteor. Soc.*, **80**, 29–55.
- Gibson, J. K., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serano, 1997: ERA description. ECMWF Re-analysis project report series 1, 72 pp. [Available from ECMWF, Shinfield Park, Reading RG29AX, United Kingdom.]
- Hoerling, M. P., and A. Kumar, 1997: Origins of extreme climate states during the 1982–93 ENSO winter. *J. Climate*, **10**, 2859–2870.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kiehl, J. T., J. J. Hack, G. B. Bonen, B. A. Boville, D. A. Williamson, and P. J. Rasch, 1998: The National Center for Atmospheric Research Community Climate Model: CCM3. *J. Climate*, **11**, 1131–1178.
- Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Willford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548–1550.
- Kumar, A., and M. Hoeling, 1998: Annual cycle of Pacific–North American predictability associated with different phases of ENSO. *J. Climate*, **11**, 3295–3308.
- Lau, N. C., 1997: Interactions between global SST anomalies and the midlatitude atmospheric circulation. *Bull. Amer. Meteor. Soc.*, **78**, 21–33.
- Livezey, R. E., M. Masutani, and M. Ji, 1996: SST-forced seasonal simulation and prediction skill for versions of the NCEP/MRF model. *Bull. Amer. Meteor. Soc.*, **77**, 507–517.
- Reynolds, R. W., and D. S. Marsico, 1993: An improved real-time global sea surface temperature analysis. *J. Climate*, **6**, 114–119.
- , T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate*, **7**, 929–948.
- 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 modeling. *Quart. J. Roy. Meteor. Soc.*, **121**, 699–704.
- Shukla, J., 1981: Dynamical predictability of monthly means. *J. Atmos. Sci.*, **38**, 2547–2572.
- , 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science*, **282**, 728–731.
- , and M. Wallace, 1983: Numerical simulation of the atmospheric response to equatorial Pacific sea surface temperature anomalies. *J. Atmos. Sci.*, **40**, 1613–1630.
- , D. A. Paolino, D. M. Straus, D. DeWitt, M. Fennessy, J. L. Kinter, L. Marx, and R. Mo, 2000: Dynamical seasonal predictions with the COLA atmospheric model. *Quart. J. Roy. Meteor. Soc.*, **126**, 2265–2291.
- Simmons, A. J., R. Mureau, and T. Petroliaigis, 1995: Error growth and estimates of predictability from the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **121**, 1739–1771.
- Suarez, M. J., and L. L. Takacs, 1995: Documentation of the Aries-GEOS Dynamical Core: Version 2. NASA Tech. Memo. 104606, Vol. 5, 45 pp. [Available from Goddard Space Flight Center, Greenbelt, MD 20771.]
- Xie, P., and P. A. Arkin, 1996: Analyses of global monthly precipitation using gauge observation, satellite estimates and numerical model predictions. *J. Climate*, **9**, 840–858.

