Propects and Limitations of Seasonal Atmospheric GCM Predictions

Arun Kumar* and Martin P. Hoerling*

Abstract

Climate simulations and hindcast experiments of increasingly large ensemble size are being performed to determine the predictive capability of atmospheric general circulation models (AGCMs) on seasonal or longer timescales. These have exhibited large sensitivity to anomalous boundary forcing associated with global sea surface temperatures (SSTs). Large-scale patterns of climate anomalies are at times generated in the extratropics when the AGCM is forced by the SSTs associated with El Niño events. It remains to be determined whether on average such results imply useful predictive skill for seasonal means in the extratropics. Indeed, given the prospects for small, if not negligible, skill in the extratropics as revealed in variance tests of boundary-forced potential predictability, one is forced to question and examine the limits of AGCM methods.

These issues are addressed within the context of a large ensemble of climate simulations using an AGCM forced with observed SSTs for the 1982–93 period. From the analysis of the model data it is argued that the impact of interannual changes in SSTs is to create a shift in the extratropical-mean state, although this shift is small and resides within the envelope of atmospheric states attained with climatological SSTs. This effect does not have any appreciable impact on the total variance of seasonal-mean atmospheric states and confirms the conclusions drawn from earlier studies.

A reliable detection of the boundary-forced shift in the mean state, however, is shown to be feasible when a sufficiently large ensemble of model runs is considered. The shift in the mean state has a certain probability of being in phase with the observed seasonal anomalies. Indeed, the benefit of generating the ensemble prediction lies in the fact that it is the ensemble-mean response that nature has the greatest probability of selecting. Nonetheless, to the extent that the observed anomalies are at least partly the result of natural variability, AGCM-based seasonal predictions will be inherently probabilistic. Implications for AGCM simulations of the extratropical response to the boundary forcing, and for seasonal-mean predictions in general, are discussed.

1. Introduction

Whether dynamical prediction of seasonal means is feasible, especially in the extratropics, remains a question. It is only recently that quantitative assessments of the level of skill achieved in hindcasts of extratropical circulation anomalies have been reported (e.g., Brankovic et al. 1994). The hindcast experiments incorporate both observed initial conditions and the observed evolution of global sea surface temperatures (SSTs). A complementary approach, referred to as a sensitivity study, employs only boundary condition information, and Barnett (1995) and Kumar et al. (1994) report the level of skill achieved using such an approach with the prescribed SSTs for the period since 1970. These atmospheric general circulation models (AGCMs) have generated seasonal-mean extratropical climate anomalies during El Niño–Southern Oscillation (ENSO) events that at times are in phase with observed patterns.

However, the tantalizing suggestion of predictive skill here must be weighed against the evidence of small boundary-forced predictability in the middle latitudes as indicated by observational and model studies (e.g., Madden 1976; Lau 1985; Chervin 1986). One is forced to examine not only the possible limitations of AGCM methods for seasonal and longer timescales, but also the methodology for further GCM improvements. Specifically, with continued model developments, can prediction derived from such models reach a stage where they demonstrate a consistently high degree of accuracy? This paper focuses on these and related issues of dynamical seasonal forecasting. A comprehensive review of the prospects for seasonal forecasting, which includes the discussion of empirical methods and predictions of boundary conditions themselves, can be found in Palmer and Anderson (1994).

Potential predictability of time-averaged atmospheric states has been historically assessed by comparing the internally generated natural variability to the total variability of the system (Madden 1976; Lau 1985; Chervin 1986; Shea and Madden 1990). The latter may be viewed as resulting from the linear combination of the natural and the boundary-forced variability (e.g., due to anomalous SSTs, soil moisture, snow cover, or sea ice). The extent to which the boundary-forced signal exceeds the inherently unpredictable background climate noise determines the
potential for predictability. This predictability is entirely due to the boundary forcing and should be distinguished from the dynamic predictability of the initial value problem that at times can lead to skillful forecasts beyond the accepted range of deterministic prediction (Shukla 1981; Miyakoda et al. 1983).

The underlying assumption within the variance tests of predictability is that the extreme climate states are associated with the extremes in the boundary forcings. The validity of this assumption is irrefutable for the tropical latitudes where interannual SST variability drives the atmospheric interannual climate variability. The situation is less clear in the extratropical latitudes, however. Although stationary wave anomalies of appreciable amplitude are frequently observed in the extratropics during El Niño (e.g., Horel and Wallace 1981; Livezey and Mo 1987), large-amplitude climate perturbations also occur during non–El Niño years. Similarly, AGCMs subjected to fixed climatological boundary forcings exhibit appreciable interannual variability in the midlatitudes comparable to that observed (Lau 1985; Chervin 1986).

The AGCM results combined with Madden’s (1976) observational study of sea level pressure demonstrate that in the midlatitudes between 40° and 60°N, the ratio of total to natural variability does not greatly exceed unity, and as such the potential predictability is judged to be low. Results can differ with season and variable examined—for example, the indications of potential predictability of January-mean surface temperature along coastal western North America (Shea and Madden 1990). There are difficulties, of course, in estimating the observed natural variability, and different estimates can lead to different conclusions. For example, Shukla and Gutzler (1983) used a slightly different method to estimate natural variability of 500-mb heights than Madden (1976) had used for sea level pressure, and they found more optimistic results for the potential predictability in midlatitudes.

Recently, Palmer (1993) has also noted that the impact of SST anomalies is small on the variance of the atmospheric extratropical flows but argued that the probability distribution of preferred weather regimes would be altered. For example, during El Niño, the probability of finding the atmospheric state in the Pacific/North American (PNA)-like pattern would be higher than during non–El Niño years. Results of his arguments, based upon a Lorenz model, are consistent with the observational results of Yamal and Diaz (1986). They find that the frequency of occurrence of the PNA pattern is greatly altered during the El Niño winters, although as noted in other studies, the PNA pattern can occur during any winter.

Using a comprehensive AGCM and multiple integrations, our analysis of variance of model data also confirms the results of earlier studies on the lack of potential predictability of extratropical monthly means. As described in the section 2, our approach makes an unambiguous separation of total variance into the internal and boundary-forced components. Specifically, in section 3 it is shown that the impact of interannual changes in SST is to create a shift in the extratropical-mean state, and that this shift is within the envelope of the internal spread. This effect does not have any appreciable impact on the total variance. Yet, a consistent SST-forced signal on the extratropical flow can be obtained when ensemble AGCM runs are conducted. Implications for ensemble AGCM integrations as tools for seasonal climate prediction are discussed in sections 3 and 4.

2. Data and analysis procedure

a. Data

A total of nine long-term climate simulations in which the atmosphere is forced with the monthly mean observed SSTs have been performed to assess the predictive capability of a seasonal prediction system developed at the National Meteorological Center (NMC). The GCM used for this purpose is referred to as MRF9 in Kumar et al. (1994), and a detailed description of this model including its climatology can be found therein. Using perturbed atmospheric initial conditions derived from balanced AGCM states, all the integrations start from 1 February 1982 and continue until 31 December 1993. The experiments are performed at spectral T40 horizontal resolution and 18 unequally spaced sigma levels. The analysis reported here is done on the 200-mb monthly means of zonally asymmetric (or eddy) height anomalies for January. This month is chosen due to the large observed and modeled SST-forced signal in the Northern Hemisphere extratropics [e.g., as implied in the observational study by Shea and Madden (1990)]. Schematically, all the available Januarys (A_w) can be arranged as in Fig. 1. The index i represents the sample within an ensemble for a particular January. The index α corresponds to individual years during the 1982–93 period with each year representing a different SST forcing. The model-simulated anomalies are compared with the observed anomalies obtained from the NMC analysis. For consistency, the observed anomalies are also computed from the mean of 11 Januarys from 1983 to 1993.

b. Analysis procedure

We first define the internal variability that occurs in the absence of interannual SST variability. For a
FIG. 1. A matrix representation of AGCM climate simulations used in this study. The index $i$ represents the sample within an ensemble that has been subjected to identical SST forcing. The index $\alpha$ represents the individual years within the period 1983–93, with each possessing different SSTs. Dots represent a particular realization of monthly mean Januaries simulated by the AGCM.

particular SST, an ensemble mean from a sample of $N = 9$,

$$\bar{A}_\alpha = \frac{1}{N} \sum_{i=1}^{N} A_{\alpha i},$$

(1)

can be defined. Recall that although each sample has been subjected to the same SST forcing, individual January realizations for a fixed SST can differ due to the internal variability, the spread of which is defined as

$$\sigma^2_{\alpha} = \frac{1}{N} \sum_{i=1}^{N} (A_{\alpha i} - \bar{A}_\alpha)^2.$$

(2)

Since the spread defined by (2) may also depend on the particular SSTs, a mean internal spread $\sigma^2_t$ is defined as the average of the internal variabilities over all the SSTs,

$$\sigma^2_t = \frac{1}{M} \sum_{\alpha=1}^{M} \sigma^2_{\alpha},$$

(3)

where $M$ denotes the number of years spanning our model integrations. For this study we employ 11 Januaries from 1983 to 1993.

We next define the atmospheric variability due to interannual changes in SSTs. Analogous to (2) the spread $\sigma^2_E$ due to the interannual changes in SST can be defined by

$$\sigma^2_E = \frac{1}{M} \sum_{\alpha=1}^{M} (\bar{A}_\alpha - \bar{A})^2,$$

(4)

where $\bar{A}$ is the mean of the entire population. This spread is an estimate of the external variance caused by the interannual changes in the SSTs. It can be shown that the total variance $\sigma^2_T$ of the population is the sum of mean internal and external variances ($\sigma^2_t + \sigma^2_E$).

We also perform a similar analysis where $\bar{A}_\alpha$, instead of being a particular model realization of January, is the result of averaging an ensemble of $k$ realizations of AGCM simulations subjected to the same SST forcing. From nine January realizations for a particular SST, such ensembles can be constructed by randomly selecting $k$ Januarys from them. Thus, for each $\alpha$, we can construct a sample of $N = (9 + k - 1)!/ (k!8!)$ distinct possible ensembles with $k$ members each and obtain ensemble averages $\bar{A}_\alpha$. We can then arrange ensemble averages $\bar{A}_\alpha$ in a manner similar to Fig. 1. Figure 1 is a trivial case of this procedure for $k = 1$.

Thus defined, the internal and external variance, but now for the ensemble averages, can again be calculated. The internal variance now represents the spread among AGCM ensemble averages, while the external variance remains unchanged. This analysis is performed to demonstrate the effect of AGCM ensemble averaging on the internal spread.

3. AGCM results

The period 1982–93 contains four major ENSO events with mature warm phases during the winters of 1982/83, 1986/87, and 1991/92 and a mature cold phase during 1988/89. The observed and the AGCM average 200-mb eddy height anomalies for these Januaries are shown in Figs. 2 and 3, respectively. Over the Pacific/North American region, the AGCM demonstrates some skill in simulating the spatial pattern of the observed atmospheric anomalies. For the domain bounded by 180° and 60°W and extending from 20° to 70°N, the spatial anomaly correlations are 0.81, 0.69, 0.59, and 0.30 for January 1983, 1987, 1989, and 1992, respectively. Of particular importance, note that the spatial anomaly correlations are not consistently high. A hypothesis can be put forth that this variability in the skill is due to natural variability in the observations, at least to the extent that...
FIG. 2. Observed 200-mb eddy height anomalies for (a) January 1983, (b) January 1987, (c) January 1989, and (d) January 1992. Units are in meters and contour interval is 20 m. Dashed contours are negative.

FIG. 3. Nine-member ensemble-mean AGCM simulation of 200-mb eddy height anomalies for (a) January 1983, (b) January 1987, (c) January 1989, and (d) January 1992. Units are in meters and contour interval is 20 m. Dashed contours are negative.
the ensemble model response is a proxy for the SST-forced component of the individual observed monthly means.

To clarify the impact of SST forcing in general, we calculate the mean model internal variance and the external variance defined by (3) and (4), the results from which are given in Figs. 4a and 4b. For the extratropical latitudes, the external variance is much smaller than the internal variance, whereas the reverse tends to be true for the Tropics. Note that the centers of maximum external variability (Fig. 4b) coincide with the centers of action of the PNA-like pattern, and one may claim that the PNA region offers the best hope for climate prediction in extratropics during winter, a conclusion also reached by empirical studies (e.g., Horel and Wallace 1981; Kiladis and Diaz 1989). Indeed, the results in Fig. 4b offer a more rigorous confirmation of this fact, since the empirical studies are based on linear correlations only, whereas the AGCM incorporates the full nonlinear interactions occurring within the climate system.

Due to the high internal and low external variance in the extratropics, the ratio of the total to the internal variance is close to unity (Fig. 4c). As such, an assessment of the potential predictability based upon the variance approach yields a rather pessimistic statement in the extratropical latitudes. This is in agreement with the earlier observational and modeling studies.

To illustrate this point further, Fig. 5 presents a schematic of the impact of the interannual SST forcing on the tropical and extratropical climate.

---

**Fig. 4.** Model-simulated mean internal (a) and SST-forced external (b) variance for monthly mean 200-mb eddy height anomaly. Units are in meters squared and contour interval is 500. Values greater than 2000 m$^2$ are shaded light, and values greater then 4000 m$^2$ are shaded dark. The ratio of the total and the internal variance in dimensionless units is shown in (c). Successive contours in (c) are drawn at 1.2, 1.8, 3.0, 6.0, and 12.0. Values greater than 1.8 are shaded light, and values greater than 6.0 are shaded dark.
the Tropics the effect of interannual SST variations is to create a detectable shift in the atmosphere’s mean state. In the extratropics, the changes in the mean state are smaller than the spread due to the internal variability. As a consequence, the SST impact on the total variance of the tropical atmosphere may be quite large, whereas a much smaller impact on the total variance of the extratropical atmosphere exists. Since for the latter region $\sigma_{T}^2/\sigma_{T}^2 = 1$, one is led to conclude that little, if any, boundary-forced predictability exists.

This ratio of variance results can be interpreted as signifying the reliability with which the impact of the boundary forcing can be detected (or simulated) if only a single case (or AGCM simulation) is considered. In the Tropics, for example, a single model realization is generally sufficient to establish the SST-forced interannual change. This fact has been well established by the highly reproducible results for the SST-forced tropical interannual variations in most AGCMs (e.g., Blackmon et al. 1983; Lau 1985; Hoerling and Ting 1994). However, for the extratropics, a single realization of an AGCM integration typically is insufficient for capturing the boundary-forced signal since the externally forced (and thus potentially predictable) component of the shift in the mean state is within the envelope of the mean internal spread.

Stated otherwise, any single observed or AGCM-simulated anomaly in the extratropics has an appreciable probability that its outcome is a consequence of the internal spread alone. Here $\sigma_{E}^2/\sigma_{E}^2 = 1$, and to extract reliably the SST-induced interannual signal requires analysis of ensemble averages of multiple realizations subjected to identical boundary forcing. Regions where $\sigma_{E}^2/\sigma_{E}^2 = 1$, although possessing little or no potential predictability on a single AGCM integration basis, can nonetheless have a detectable boundary-forced signal if an ensemble approach is taken. It should be recognized that even if the boundary-forced signal obtained from ensemble averaging is taken as a prediction for the observed anomaly, the AGCM-based prediction will still be probabilistic in nature. This inherent property of an AGCM’s seasonal prediction for individual observed events is a fundamental limitation of ensemble methods applied to climate timescales.

There is a practical benefit of ensemble methods for climate prediction that needs to be recognized. Specifically, the ensemble-mean GCM response is the maximum likelihood region for the observed anomalies for particular events. This can be seen in Fig. 5 by the vertical lines that coincide with both the GCM ensemble-mean response and the highest probability region selectable by nature.

One frequently hears the comment that failure to accurately simulate an observed anomaly in the extratropics is due to model errors. While there undoubtedly is some validity to such claims, consider for the moment the best-case AGCM scenario, when in Fig. 5b the model-simulated probability density function (PDF) for extratropical latitudes is a perfect analog of the observed PDF. The ensemble-averaged AGCM anomaly for SST is then given by the centroid (denoted by the vertical line in Fig. 5b), whereas the observed anomaly for any single case will have a nonzero probability of residing anywhere on the curve. If the distance between the centroid and the observed anomaly is some measure of the skill, then the skill itself will exhibit randomness. This is entirely consistent with, and a possible explanation for, the case-to-case variability in the skill for four ENSO Januarys noted earlier.

It follows that even for a perfect AGCM, the skill of ensemble averages need not be consistently high. Table 1 provides further evidence in support of these
The correlations are computed for the Pacific/North American domain bounded by 180° to 60°W and 20° to 70°N. Correlations among the GCM patterns are 0.7 or greater, indicative of the canonical ensemble-mean model response. Much larger spread exists among the individual observed anomaly patterns as indicated by greatly reduced spatial anomaly correlations.


table 1. spatial anomaly correlations for 200-mb eddy height among observed (upper right triangle) and 9-member ensemble-mean GCM responses (lower left triangle) for the indicated ENSO Januarys. the correlations are computed for the Pacific/North American domain bounded by 180° to 60°W and 20° to 70°N. Correlations among the GCM patterns are 0.7 or greater, indicative of the canonical ensemble-mean model response. Much larger spread exists among the individual observed anomaly patterns as indicated by greatly reduced spatial anomaly correlations.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>1.00</td>
<td>0.52</td>
<td>-0.58</td>
<td>0.71</td>
</tr>
<tr>
<td>1987</td>
<td>0.73</td>
<td>1.00</td>
<td>-0.10</td>
<td>0.44</td>
</tr>
<tr>
<td>1989</td>
<td>-0.70</td>
<td>-0.82</td>
<td>1.00</td>
<td>-0.14</td>
</tr>
<tr>
<td>1992</td>
<td>0.71</td>
<td>0.82</td>
<td>-0.77</td>
<td>1.00</td>
</tr>
</tbody>
</table>

It could be argued that due to model errors, the AGCM ensemble-averaged responses are too canonical, and the implied lack of sensitivity to the details in the spatial structure of ENSO SST anomalies is unrealistic. There are two avenues available for addressing this issue. First, the lower left triangle in Table 1 has been reconstructed using individual AGCM integrations instead of ensemble means, and the correlations between these model responses for the four ENSOs show behavior similar to that observed (not shown). This is indicative of the fact that the atmospheric response on average, and different modes of ENSO atmospheric responses exist. However, the observational data are not sufficient for this approach to be viable. As an alternative to this, the results from the AGCM may already suggest the fact that the atmospheric mean response is not very sensitive to the details in the ENSO SST anomalies.

What minimal ensemble size is then required to reliably detect the boundary-forced signal? This question is best addressed by examining the relationship between ensemble size and the ratio of variance using multiple integrations of the AGCM. Recall that the internal variance shown in Fig. 4 is a trivial case of internal variance between ensembles with member size $k = 1$. If the model integrations are independent (i.e., are beyond the influence of initial conditions), the expected mean internal spread for the ensemble averages with $k$ members each will be $\sigma_f^2/k$ (Bendat and Piersol 1986). However, the external variance is expected to remain unchanged. To illustrate, for each $\alpha$ as in Fig. 1 ensembles are constructed by randomly sampling $k$ members out of nine possible Januarys. For $k > 1$, $N = 20$ ensembles and their corresponding means have been obtained, although as indicated in section 2, many more distinct ensembles could be constructed. The resulting globally averaged internal and external variances for different $k$ are shown in Table 2. As expected, the mean internal variance shows an approximate linear decrease with increasing ensemble size, while the
external spread remains approximately constant (the external variance would have been identical if all possible ensembles were used).

The spatial distribution of the mean internal and the external variance for \( k = 5 \) is shown in Figs. 6a and 6b, respectively. The reduction in the internal variance is quite uniform globally (compare Figs. 4a and 6a), whereas the distribution of the external variance shows no appreciable change (compare Figs. 4b and 6b). Now repeating the ratio of variance test reveals extratropical regions with values near 2 (Fig. 6c) versus near unity in Fig. 4c. This in fact is the fundamental consequence of ensemble averaging since it acts to reduce the influence of the random (and hence unpredictable) internal variations. One may judge from Table 2 that an ensemble size of at least six is required to reliably detect the extratropical signal in the GCM.

To the extent that ensemble averaging is equivalent to time averaging, these results imply that variance estimates of potential predictability for seasonal means may be more optimistic (see also Palmer 1987). This is consistent with the skillful seasonal simulations and hindcasts reported recently by Hoerling et al. (1992), Brankovic et al. (1994), and Kumar et al. (1994). The effect of longer time-averaging intervals is to reduce the contribution of white-noise high-frequency effects in favor of red-noise low-frequency phenomena that may have a significant externally forced component (see also Shea and Madden 1990). Thus, for the seasonal or longer time averages, the reduction in the internal variance will improve the probability of the AGCM prediction being successful, although the fundamental limitation that the prediction is probabilistic still remains.

The purpose of the ensemble approach in climate prediction is to have...
a large enough sample to obtain a representative measure of the probability distribution of the internal spread. This requirement is essential; otherwise, the statistical properties associated with the ensemble—for example, the ensemble mean—will be biased estimates. It should be further noted that for climate predictions, ensembles with smaller than typical internal spread (i.e., $\sigma^2$), instead of indicating high confidence in the ensemble mean prediction, may point to poor sampling. This is in contrast to initial value medium-range prediction, where the spread among the perturbed initial condition integrations is considered to be a measure of the probability of the ensemble-averaged forecast being skillful. For example, if the majority of climate integrations yield consistent seasonal anomalies (i.e., small $A_{\text{int}} - \bar{A}_{\text{int}}$), then such an occurrence can be attributed to the chance (albeit small) associated with the probability distribution of the internal spread. This is particularly true for extratropical regions. Indeed, consistent climate anomalies simulated by individual members of an AGCM ensemble can be argued to contradict the basic notion of chaotic atmospheric behavior. A further discussion on the breakdown of the correlation between the spread among ensemble members and confidence in the predictive skill for extended integrations is given by Barker (1991).

4. Conclusions and discussions

This study has attempted to explore the basic science issues relevant to dynamical seasonal prediction, particularly in the middle latitudes. The potential for seasonal predictions is derived largely from the influence of anomalous conditions at the earth’s surface, and the most significant of these is associated with the El Niño–Southern Oscillation phenomenon. Outside the Tropics, classic ratio of variance tests of potential predictability demonstrate that for the monthly means the signal associated with even this major boundary perturbation is easily masked by the internal variability of the midlatitude atmosphere.

Our own results based on large ensembles of climate simulations using observed SSTs for 1982–93 when the ENSO cycle was especially pronounced confirm the conclusions drawn in the earlier studies. For the extratropics, we showed that the low potential predictability stems from the fact that the shift in the extratropical-mean state due to interannually varying SSTs resides within the envelope of climate states attainable using fixed SSTs.

Confronted with these facts, one is forced to inquire into the levels of skill that are achievable on monthly, seasonal, or longer timescales using atmospheric GCMs forced with boundary conditions only. There is some evidence on seasonal timescales that ensemble AGCM hindcasts and simulations for El Niño events can be quite successful in the extratropics (e.g., Brankovic et al. 1994; Molteni et al. 1993; Kumar et al. 1994), although this skill appears quite case dependent. This can be understood from the fact that a reliable detection of the boundary-forced shift in the mean state can be attained if an ensemble of multiple AGCM integrations is conducted. Yet, the ensemble prediction for single events will necessarily remain probabilistic. In the presence of extratropical natural variability, the constraint on the observed extratropical state to select the boundary-forced component is weak, although it depends on the time-averaging interval. There are other important scientific questions related to the development and implementation of atmospheric GCM for seasonal prediction problem, and they are discussed next.

An implication of the current analysis is the minimum ensemble size for climate simulations of the atmospheric response to SST forcing. Analysis of variance for the ensemble averages summarized in Table 2 gives some guidance about the choice of the ensemble size needed to reliably detect the boundary-forced interannual signal. The magnitude of the global internal variance for an ensemble size of 5 is nearly equal to the external variance and decreases progressively for larger ensembles. This suggests that an ensemble size of 6–10 may be sufficient to reliably detect the boundary-forced signal. Of course, this is likely to be a model-dependent result, and for AGCMs with greater internal variance (or smaller external variance), larger ensemble sizes will be required. Nonetheless, once the typical magnitude of an AGCM’s internal and external variance is known, an a priori estimate of the ensemble size can be made.

The effect of ensemble averaging may also be viewed to be equivalent to time averaging. The difference, of course, is that the boundary-forced signal is not forfeited under ensemble averaging, whereas the seasonal dependency of the boundary-forced signal (e.g., Lau 1985; Kiladis and Diaz 1989; Barnston and Livezey 1987) argues for care in choosing the time-averaging interval. The focus of our study has been on monthly mean ensembles. It is evident, however, that for longer time averages (i.e., seasonal to annual) the influence of internal variability will diminish, and to the extent that the boundary-forced signal is temporally coherent, the external signal may actually dominate. Possible impacts of time averaging on signal-to-noise ratio can be clearly seen in Table 2. This fact may explain the recent reports of skillful midlatitude simulations on seasonal timescales. However, AGCM pre-

Bulletin of the American Meteorological Society

343
dictions for the seasonal means may not be the best averaging period. The optimal time-averaging interval will ultimately be determined by the timescale of the external forcing and the seasonality of the boundary-forced atmospheric anomalies. The determination of such intervals is clearly a central problem in climate prediction since it is for the optimal time mean that the ensemble AGCM response to boundary forcing will have the largest probability of capturing the observed atmospheric anomalies. Of course, from a practical viewpoint, it can be argued that AGCM predictions for shorter time averages will be more beneficial. However, for such periods the AGCM predictions will also have a greater probability of deviating from the observed anomalies, thus compromising their optimal utility.

This study has focused on the potential predictability of 200-mb eddy height anomalies, and it is important to inquire into the potential predictability of other meteorological variables. For example, the low internal variance for the 200-mb anomalous eddy heights in the Tropics suggests that a single model realization may be sufficient to detect the boundary-forced signal, but the prospects may not be so promising for other variables, such as precipitation (Dix and Hunt 1995). The predictability estimates for different variables in the Tropics are likely to differ and may still warrant ensemble averaging. A similar dependence of ensemble size on atmospheric variable is likely to exist for the extratropics.

An additional issue concerns the upper limits on the ensemble-mean AGCM predictions for a particular case. Ensemble averaging acts to reduce the influence of random internal variability, and furthermore the ensemble mean points to the region where the observed anomalies have the largest probability of occurrence. Yet the observed anomalies are equivalent to a single realization of the AGCM and can have an appreciable random, and hence unpredictable, component. These considerations alone may explain the low correlation between the observed and AGCM anomaly patterns for January 1992. We do not argue that the present results denote upper limits on the capabilities of the model-based simulations of extratropical boundary-forced variability, since model errors undoubtedly exist. Indeed, given the ongoing effort in model development, the extent to which an AGCM's ensemble-averaged boundary-forced interannual variability can reproduce the observed variability remains an open question. Insights are likely to come from a careful analysis of observed and simulated anomalies, and by identifying that variability unrelated to interannual SST variations (e.g., Ting and Hoerling 1993; Hoerling and Ting 1994; Lau et al. 1994).

An important related question is, To what extent are the spatial patterns of the ensemble-averaged simulated anomalies themselves dependent upon the details of the spatial structure in the SST forcing? Some observational evidence appears to suggest that the atmospheric anomalies over the PNA region during ENSO cases are not perfectly correlated may suggest small sensitivity to different SST forcing, although analysis of more diverse cases is required.

Given the above problems, and the implied limitations of the ensemble AGCM approach, one needs to be careful in assessing apparent AGCM deficiencies. Instances when, due to chance alone, the observed extratropical atmospheric anomalies (even in the presence of large SST forcing) get overwhelmed by the random internal variability cannot be discounted. For such cases, one can claim a priori that the observed anomalies and the SST-forced AGCM response will compare unfavorably. As a case in point, the winter of 1972/73 is widely considered to be among the strongest El Niño episodes of the century (Quinn et al. 1978), yet the extratropical circulation anomalies were not particularly noteworthy and deviated significantly from the composite analysis of teleconnection patterns during the ENSO. The assessment of an AGCM's simulated interannual variability and the judgments for model developments may be better based upon the comparison of simulations of the tropical interannual variability, since here the influence of internal variability is comparatively small. If, however, the ensemble-averaged simulations are to be compared to observed anomalies in the extratropics, conclusions about the model performance should be based upon averages over many ENSO events (Kumar et al. 1994).

To summarize, while recognizing and confirming the limitations imposed by the extratropical natural variability, the prospects for atmospheric GCM-based seasonal to longer timescale boundary-forced predictions may still be bright. The level of optimism has to be evaluated ultimately within the perspective of (i) the ability to predict the boundary conditions themselves and (ii) accuracy requirements of such time-averaged predictions as they pertain to economic benefits (see also Palmer and Anderson 1994).

Acknowledgments. Support provided by NOAA's Office of Global Programs through the Climate and Global Change Program is gratefully acknowledged. Original versions of this manuscript greatly benefited from careful reviews by Drs. R. Madden, Huug van den Dool, and two anonymous reviewers. We are also thankful for insightful discussions with Drs. A. Leetmaa and P. Sardeshmukh.
References


