Improving Seasonal Prediction Practices Through Attribution of Climate Variability

Anthony G. Barnston[1], Arun Kumar[2], Lisa Goddard[1], and Martin P. Hoerling[3]

[1]International Research Institute for Climate Prediction, Columbia Univ., Palisades, NY

[2]NOAA/NCEP, Climate Prediction Center, Washington, D.C

[3]NOAA, Climate Diagnostics Center, Boulder, Colorado

Summary

Identification of the part of seasonal climate variability that is predictable from SST is discussed, for benefits in forecasting and in establishing attribution for observed climate states.

Abstract

The Seasonal Diagnostics Consortium of the Applied Research Centers is engaging in a real-time activity to detect and understand the role of sea surface temperature (SST) anomalies in observed climate anomalies. The activity is aimed to improve practices in seasonal climate forecasting by fully harvesting the accumulated research evidence of the climate's sensitivity to ocean forcing. The approach, in the first phase of the activity, involves performing ensembles of atmospheric general circulation models (AGCMs) at several institutions, using the most recently observed global SST anomalies as prescribed forcings. The runs are routinely updated each month as the latest SST observations become available, adding to the archive of historical simulations spanning the last half-century.

The SST-forced signal in the seasonal mean climate is detected through the agreement among ensemble mean anomalies drawn from the simulations of the various AGCMs. The Consortium activity also compares the dynamically forced signals with those estimated empirically, based on the observational archive. A comparison of the coordinated simulations with the observed climate anomalies is then made for two principle reasons: (1) to offer an attribution for the ocean's role in the origin of the observed seasonal climate anomalies, and (2) to determine the causes for success or failure of operational seasonal climate predictions, whose tools may be either mainly dynamically or empirically derived. It is expected that routine climate diagnostics and attribution efforts for climate anomalies will help further develop the knowledge base for improving the practice of seasonal climate predictions, and advance understanding of global climate on seasonal to decadal time-scales.

1. Motivation

Operational seasonal climate prediction is an emerging practice with far reaching societal applications. An ability to anticipate climate fluctuations one or more seasons in advance would have measurable benefits for decision-making in hydrology, agriculture, health, energy, and other sectors of society. It would allow for proactive reservoir management (Cunha 2003), crop planting adjustments (Dilley 1997; Hammer et al. 2001), vector control for epidemic disease prevention (Thomson et al. 2000), fuel storage, and other mitigative measures. It also influences other risk management activities implicit in the weather derivatives and reinsurance industries (Murnane et al. 2002; Murnane 2004). Until the most recent decades, climate prediction had been viewed as a speculative and largely unproven venture. In the 1980s, the U.S. National Weather Service's seasonal outlooks were developed using mainly lag correlations of observed upper atmospheric pressure anomalies (Wagner 1989) and analogues (Livezey and Barnston 1988), with some consideration of newly discovered teleconnections from tropical Pacific sea surface temperatures (SSTs)(Horel and Wallace 1981). More recently, and largely as a result of better quantification of the climate effects of the El Niño/Southern Oscillation (ENSO) phenomenon, seasonal forecasts of 3-month average surface temperature or precipitation have been clearly demonstrated to have skill in particular seasons, regions and circumstances (Livezey 1990; Kumar et al. 1996; Shukla et al. 2000; Graham et al. 2000). Nonetheless, users often remain cautious and reluctant to make use of officially issued climate forecasts in their decision-making process. Underlying this hesitancy is the complication of accommodating the inherent probabilistic nature of climate forecasts. As such, the value of the forecasts becomes clear only through their consistent application over a set of cases rather than in any single case. A related issue is the fact that it is not always clear why seasonal predictions succeed in some instances but fail in others.

In the emerging practice of seasonal climate prediction, the following questions can be posed: Are the failures of individual seasonal predictions largely reflections of an inherent natural limit of seasonal climate predictability? How much do inaccuracies and biases in seasonal prediction methodologies contribute to errors in seasonal climate predictions? Is there an optimal way to practice seasonal climate prediction?

In this paper, we focus on some issues related to the practice of making seasonal climate predictions. Starting from a discussion of why climate predictability exists, we propose what the best practice for seasonal climate prediction may be. The best practice is viewed as a procedure that would lead to the most accurate seasonal climate predictions, resulting in highest possible skill when evaluated over an extended validation period[1]. We then describe a Consortium activity aimed at realizing some aspects of making the best possible predictions. Other practical applications for the Consortium activity are also discussed.

2. Climate Predictability and Climate Predictions

Seasonal climate predictability has its scientific basis in the fact that slow variations in the earth's boundary conditions, e.g., SST, snow cover, soil wetness or vegetation, can

---

[1] We apply the term *accuracy* to forecast quality for an individual case (e.g. a single location for one time), and the term *skill* for quality over a collection of cases (either many locations for a single time, many times for a single location, or, more typically, for many cases over both time and space).

influence global atmospheric circulation and thus global surface climate[2]. If future evolution in the boundary conditions can be anticipated, then from the knowledge of their influences on global atmospheric circulation, skillful seasonal predictions for atmospheric climate anomalies are possible. A key requirement in making successful seasonal climate forecasts, therefore, is *documenting and understanding atmospheric responses to a broad range of anomalous boundary forcings*. Because SST forcing is principle among the boundary conditions influencing atmospheric seasonal variability, we focus mainly on it in this paper.

It can be shown that the most accurate prediction of seasonal mean climate, i.e., a seasonal climate prediction that *on average* would have the highest skill, is the atmospheric response associated with the attending anomalous boundary forcing. To the extent that there is a "forcing-to-response" relationship between the boundary conditions and the atmosphere, the predictable signal can be estimated from dynamical experiments in which the boundary conditions are specified. Indeed, this is the presumption of the so-called two-tiered practice of climate forecasting that has been in wide use (Bengtsson et al. 1993). In this initial phase of the Consortium activities, the focus is on the sensitivity of such atmospheric models to specified SST boundary conditions; subsequent efforts will address the suitability of the two-tiered design itself.

In theory, over increasingly large numbers of cases of the same anomalous boundary forcing, the statistical distribution of the atmospheric anomalies would converge to that which is most likely for that forcing. In the absence of large samples of repeated observations of any specific SST anomaly pattern, this distribution is estimated from large ensembles of AGCM integrations for the same SST anomaly pattern but for differing atmospheric initial conditions—the latter giving rise to variability in the climate simulations across ensemble members. The summary information obtained from the statistical distribution (hereafter called the probability density function, or PDF) would then be considered the most accurate and hence the best forecast. For example, the mean anomaly across the ensemble members can be considered as a possible deterministic prediction, and the uncertainty of this prediction would be reflected in the spread (dispersion) of the PDF about the mean anomaly.

Deterministic predictions of seasonal mean climate based on the mean anomaly, however, are not always accurate for individual cases, even for a perfect AGCM and perfectly represented boundary conditions. This is the case because a single observed seasonal climate anomaly convolves the boundary forced signal with the noise—the atmosphere's internal variability that is not boundary forced, and hence largely unpredictable on a seasonal time-scale[3]. This variability consists of individual weather events over the course of the season whose timing, intensities, and trajectories are not predictable beyond about two weeks into the future. Included in these weather events are certain longer-lived atmospheric anomalies such as high latitude blocking patterns and

---

[2] Seasonal forecast skill can also be derived to a lesser extent from within the atmosphere, from phenomena having long lifetimes such as the Madden-Julian Oscillation (MJO; Madden and Julian 1971) or the stratospheric Quasi-biennial Oscillation (QBO; Baldwin et al. 2003). Here, however, we limit our attention to external boundary conditions.

[3] Within the paradigm of deterministic prediction, signal refers to the location of the mean of the PDF (and its deviation from the climatological mean), whereas noise is represented by the PDF's spread. For probabilistic predictions, signal is represented as the entire PDF itself (and its difference from the climatological PDF), and noise as the uncertainty of the prediction (i.e. the forecast probabilities).

stationary troughs. Consequently, even the best possible seasonal predictions have notable uncertainty. It is for this reason that seasonal climate predictions are cast in probabilistic terms, to reliably reflect the uncertainty related to intrinsic "weather noise". Our current knowledge about the relationships between boundary conditions and climate is incomplete, consisting primarily of associations related to ENSO and a few other, mainly more localized, tropical phenomena. Essential for improving seasonal climate predictions is obtaining a wider knowledge of associations between boundary conditions and the statistical distribution (PDF) of the atmospheric climate states, together with a better understanding of the predictability of those boundary conditions.

Within the paradigm that the predictable part of climate variability emerges from boundary forcing, and acknowledging that the best possible seasonal climate prediction is the PDF of atmospheric states consistent with that forcing (in the "forcing-to-response" paradigm mentioned above), one pathway for improving the practice of climate predictions is to: (1) obtain a best prediction for the future boundary condition anomalies, and (2) specify the true PDF corresponding to this boundary condition prediction as the seasonal prediction having maximum possible skill. Such an approach minimizes error in the estimate of the expected atmospheric state caused by the imposed forcing. Returning to our example of the mean anomaly as a deterministic prediction, the mean anomaly has the maximum likelihood of occurrence (assuming the true PDF is not highly irregular), and is therefore the best possible prediction. It is stressed that this "best" prediction may not be accurate with respect to the observed result in individual cases, or even skillful on average over many such cases, since the signal may be disturbed by the climate noise whose amplitude is represented by the spread of the PDF.

Such a 2-tiered prediction system is already in use at many operational centers. One may argue that the 2-tier approach is a transitory phase that is necessitated by biases in the current generation of coupled ocean-atmosphere prediction systems. With future improvements in dynamical models, seasonal climate predictions may eventually be done using a one-tier system, in which the boundary conditions and the atmospheric responses evolve together. Such an evolution notwithstanding, understanding seasonal atmospheric climate variability, and its connection to the earth's slowly evolving boundary conditions, could still rely on the 2-tier approach where boundary conditions can be considered an external forcing.

3. Two Approaches to Inferring Atmospheric Responses to Boundary Forcings[4]
The atmospheric responses to the slowly changing boundary conditions can be inferred using either of two broad approaches. An empirical approach involves analysis of historically observed boundary conditions and the accompanying global circulation and surface climate. That is, using observations, patterns of correspondence between SST anomalies and the climate can be identified. This approach is illustrated in studies such as Barnett (1981), Ward and Folland (1991), Barnston and Smith (1996), Yang et al. (1996), Ward (1998), Landman and Mason (1999), and Mason and Goddard (2001), to name only a few. While often identifying an approximately correct atmospheric signal forced by the ENSO-related SST anomaly pattern and to a lesser extent by one or two more localized tropical SST patterns (Hastenrath, 1995; Anderson et al. 1999), the period of globally

---

[4]A review of the history of the development and current state of climate prediction science is provided in Goddard et al. (2001).

adequate observational analyses is not long enough to resolve differences in the relationships between different "flavors" of ENSO SST forcing and climate, or between presently unrecognized non-ENSO-related SST forcings and climate. This is the case because inadequate samples of such SST variations exist upon which stable empirically-derived climate models can be built (see also Nicholls 1984).

It is this latter gap in particular that can be filled by a second, and complementary, approach to defining the climate signals forced by boundary conditions using dynamical approaches. A multitude of atmospheric general circulation model (AGCM) simulations forced by the historical SSTs can be produced to yield a large sample of climate states for each any every boundary state in the record. Permitted thereby is a statistically robust estimate of the climate sensitivity to the various, known SST conditions. In addition to detecting the mean climate signal due to the influence of the boundary condition, the unpredictable portion of the atmosphere's behavior for that forcing is also estimated by running the AGCM repeatedly using the same boundary conditions but differing atmospheric initial conditions, thereby yielding the PDF of climate states consistent with the forcing. For the typical 15-day lead times of seasonal climate predictions, details of the initial atmospheric conditions are not important because the forecast is, on average, beyond the range of deterministic predictability. The result is an ensemble of simulations whose mean is representative of the signal coming from the boundary conditions and whose variations (which are often substantial) express the uncertainty related to the different possible sets of unpredictable weather events. The larger the ensemble of AGCM runs, the smaller the role of sampling error and the more representative the resulting PDF is expected to be, aside from biases of the AGCM. Relationships among boundary condition signal strength, AGCM ensemble size and atmospheric climate simulation skill are discussed in Kumar et al. (2001a), Kumar and Hoerling (2000), and Wehner (2000), among others.

The AGCM simulations are computer intensive, which partly explains why the dynamical approach emerged most rapidly during the most recent decade. Yet, this approach to determining SST-atmosphere relationships circumvents problems associated with the limitations of observed data, and makes possible simulations of atmospheric responses to boundary forcings that have hitherto been unsampled in the instrumental record—an attribute that is particularly germane within a non-stationary climate.

It remains an open question to what extent dynamical methods may improve upon the information gleaned from empirical methods alone for purposes of climate attribution and prediction. The answer depends largely on the sensitivity of the atmosphere to changes in boundary forcings. If there is mainly just a single preferred atmospheric pattern, the one that is forced by ENSO-related tropical SSTs, then empirical approaches have adequate historical archives upon which to define that mode of variability. The question is whether inter-ENSO variations of tropical SST anomalies matter to the atmosphere, and furthermore whether there exist significant response patterns to non-ENSO sources of SST variations—including ones that might be unprecedented in the observed archive. Notwithstanding debates on the merits of empirical vs. dynamical methods in inferring atmospheric responses to boundary forcings, there is little doubt that a synergistic use of both empirical and dynamical approaches would be useful in the advancement of our understanding of seasonal climate variability.

The complexity of the atmospheric processes and of the sequence of physical events leading from SST forcings to atmospheric responses makes individual AGCMs prone to specific biases and raises well justified questions about any conclusions based on individual AGCMs. This situation leads to potential problems in determining the true atmospheric signal associated with a boundary forcing based on any single model. The observations pose a no less difficult problem, namely whether a single realization of nature can be relied upon to accurately reflect that signal.

Keeping the influence of AGCM biases in mind, the modeling challenge is illustrated in a comparison of the observed and a single-model ensemble mean simulated 200-hPa height anomalies during December through February (DJF) 2001-2002 (Fig. 1). We focus on the AGCM's mean anomaly because it represents the most basic characteristic—the first moment—of the PDF across the AGCM's ensemble members. What can be inferred from the agreement between observations and simulations for this single case, and what is the implication for attribution efforts to discern the role of SST boundary forcing?

Comparison with the observed height anomalies (Fig. 1, bottom) reveals a roughly out of phase relationship over North America. Does this imply a poor performance by the AGCM that could be due to biases? On the contrary, from this comparison between the atmospheric signal and the observed seasonal mean anomaly one cannot reach any definitive conclusion, as the observed anomaly may have been influenced by atmospheric internal variability, while still being consistent with the PDF of seasonal mean atmospheric states associated with the SST forcing. Consistency with the SST-forced PDF means that the observed result is within the range of possibilities implied by the PDF, including occasionally residing on a tail of the PDF.

The same uncertainty would exist if the AGCM ensemble mean had closely resembled the observations for a single case, insofar as agreement may have been coincidental given that internal atmospheric variability can not only mask the boundary forced signal, but also constructively interfere with that signal. It is by comparing a large sample of cases of observed and simulated climate anomaly patterns for winter in North America that the suitability of a particular model can be assessed (Kumar et al. 1996). On the other hand, a principal purpose of the Seasonal Diagnostics Consortium is to detect the boundary forced atmospheric signals on an individual case basis, as for example DJF 2001-2002. A step toward significantly reducing the uncertainty surrounding AGCM behavior for individual cases is proposed in the next section.

4. The Seasonal Diagnostics Consortium

An enterprise that is beyond the resources of any single institute is to perform real-time climate simulations using numerous, different AGCMs—and this leads to a key diagnostic tool of the Consortium for detection and attribution efforts. The premise of this activity is that *detection of the boundary forced signal on an individual case basis is feasible in the context of a multi-AGCM approach,* as this would reduce the effect of AGCM-specific biases. There is reason to suspect that the different models have differing biases regarding their SST forced sensitivity, as suggested by the improvement in dynamical forecast skill when using multi-model (Rajagopalan et al. 2002) and "super-ensemble" (Krishnamurti et al. 1999) approaches. It is nonetheless also recognized that despite the fact that several AGCMs used in the Consortium were developed independently at different institutions, they have some common biases in their mean

climates, and these may lead to common biases in their sensitivities. Those notwithstanding, a dynamical approach based on use of several AGCMs is necessary to advance detection of boundary forced seasonal climate signals, and to provide meaningful attribution for the origin of observed seasonal anomalies. The degree of agreement in the atmospheric responses among AGCMs serves as a first-order indicator of the prominence of the boundary-forced signal, as compared with the internally generated atmospheric variability.

Presently, the Consortium ensures that several AGCMs are run in near-real time using the most recently observed global SSTs and sea ice (Reynolds and Smith 1994) as prescribed boundary states. This suite of AGCM runs is continually updated monthly, and appended to a historical archive of such simulations spanning, for most models, the post-1950 era. The boundary states of soil moisture and snow cover are not specified according to observations, though these are fully interactive in the AGCMs. Trace gases, including ozone, methane and carbon dioxide, are set to modern climatological values. The activity produces simulations designed in the same way as those already archived from the AMIP experiments (Gates et al. 1999). Distinguishing the Consortium's simulations is that they are produced and examined as soon as is physically possible, are compared with corresponding results using SSTs predicted 3.5 months earlier, and in some cases are associated with an ongoing climate state that may affect a seasonal forecast task presently at hand. Adding to the real-time context is that some of the AGCMs used in the Consortium are used routinely at the operational forecast centers (e.g. NOAA/CPC and IRI), whose recent errors and long-term biases are of importance to forecasters in consolidating the indications of the AGCMS and other tools.

The institutions participating in this ARCs Consortium determine the set of AGCMs used in the Consortium activity. However, at present not all institutions provide their model simulations, due to practical considerations. The list of institutions is shown in the sidebar. Some basic features of the participating AGCMs are given in Table 1.

To illustrate the diagnostic activity of the Consortium, we consider surface temperature over North America during the recent El Niño of DJF 2002-03. The moderate intensity of that event (Fig. 2) led to expectations for a canonical winter climate resembling the previously documented teleconnection pattern during El Niños (Opsteegh and van den Dool 1980; Hoskins and Karoly 1981; Horel and Wallace 1981; Halpert and Ropelewski 1992; Barnston 1994; Graham et al. 1994; Graham and Barnett 1995; Hoerling et al. 1997; Higgins et al. 2000).

Figure 3 shows the global surface temperature simulations of five AGCMs as forced by their observed global SST and sea ice analyses for DJF 2002-03. As an assessment of the boundary forced signal for this case, we diagnose the occurrences of unanimous sign agreement among the ensemble mean anomalies of all five AGCM simulations. The regions of unanimous agreement are indicated by color-shading in Fig. 4. Within the regions of agreement there is fairly high confidence of having detected the sign of the boundary forced signal, whereas outside of that region the confidence of detection is judged to be inadequate. Over North America, all five AGCMs simulated above normal surface temperature in the western U.S. and southwestern Canada, and below normal temperature along the east U.S. coast from the Gulf of Mexico northeastward to southeastern Canada (Fig. 4). There is also unanimous agreement that the specified SSTs yielded positive temperature anomalies over nearly all tropical landmasses.

In the analysis above, the mean of the ensemble simulations is interpreted as providing the most likely boundary-forced climate signal for each AGCM. Such analysis focuses only on the polarity of the signal, with no regard to amplitude, and unanimity of sign across all AGCMs is required to claim detection of the atmospheric signal.[5] The probability that five AGCMs having independently designed dynamical cores and parameterization schemes will produce the same atmospheric anomaly sign in the absence of any signal is only $2/2^5$, or 0.06. (This is the same binomial probability as that of flipping a fair coin five times and getting the same outcome each time: 2^{-5} for all heads, plus 2^{-5} for all tails.) The actual probability is higher than 0.06 to the extent that model biases are shared across the models. Nevertheless, agreement among a set of models, each contributing as large an ensemble of runs as possible to its ensemble mean, is one viable approach to detection of the boundary forced signal on an individual case basis.

The approach of the Consortium activity to infer atmospheric signals for SST forcings circumvents the problems of inferring similar relationships using a single AGCM. Agreement in the atmospheric responses among different AGCMs forced with common SSTs enhances our confidence in the "fidelity" of atmospheric responses, even if such responses, for unprecedented SST forcings, cannot be verified from the limited observational data. On the other hand, a first order "sanity check" on the AGCM-simulated atmospheric response to the ENSO SSTs can always be made by comparing it with the observed ENSO composites that are fairly well defined and understood.

5. Broader Applications of Seasonal Diagnostics Consortium Activity

The central goal of the Consortium activity is to increase knowledge of the atmospheric impacts of SST boundary conditions, including relatively subtle impacts that at present cannot be gleaned from the observed data alone. In recent decades enormous focus has been placed on understanding the effects from the tropical Pacific ENSO canonical mode boundary forcings (e.g., see a review by Trenberth et al. 1998). However, forcing related to variations in ENSO's "flavor", from the Indian Ocean in the absence of an ENSO signal, from the tropical Atlantic, or from the western Pacific warm pool, is less well understood. These effects need better documentation not only through isolated ocean forcing experiments (e.g. Goddard and Graham 1999; Barsugli and Sardeshmukh 2002) but also for actual SST forcings for all tropical oceans—a focus of the Consortium activity.

Apart from the straightforward task of documenting atmospheric responses to a wide range of SST forcings, the Consortium activity and its datasets also have other applications as discussed below.

a. Climate Attribution
Current research on seasonal climate variability can be viewed as having two distinct aspects: *monitoring* and *prediction*. Monitoring involves documenting global seasonal observed climate anomalies, without necessarily seeking to understand their causes.

---

[5] Ideally, the multi-model signal would be defined probabilistically by the number of ensemble members falling into specified ranges. In the absence of refined model recalibrations and the large ensemble sizes needed to make such probability assignments meaningful, here we consider only the ensemble mean (which has a lower ensemble size requirement) and cases of unanimous sign agreement, as a first step.

While seasonal prediction activities involve evaluations of the robustness of the skill of the predictions, they often do so without seeking reasons for why predictions are accurate in some cases and inaccurate in others. The Consortium activity can add to both aspects by examining the causality of observed seasonal climate anomalies.

As discussed in section 4, an outcome of the Consortium activity will be to better understand atmospheric signals consistent with global SST forcings. This exercise could be done independently of the observed atmospheric anomalies. However, if the atmospheric signals inferred from the AGCM simulations are also compared with the observed atmospheric anomalies, the Consortium activity can examine the causality of observed seasonal climate. In other words, the same set of AGCM runs can be used to explore attribution (identification of causality) and predictability (robustness of relationship between external forcing and climate) for the observed anomalies.

As an example, in the top panel of Fig. 5 we show the observed surface temperature anomalies for DJF 2002-03. A comparison between the observed anomalies and the model simulations (Fig. 3) indicates that there is good agreement between the two over North America—in particular the negative temperature anomalies along the immediate east coast of the U.S. and positive anomalies in portions of western North America. The mutual agreement in the atmospheric signal among five AGCMs and the observations gives some confidence that the observed anomalies are related to the global SSTs. Shown in the bottom panel of Fig. 5 is an agreement plot similar to that of Fig. 4, but also incorporating the observations with the model-simulated signals. Here it is seen that a few regions of agreement among the AGCMs are lost when the observations are included, as for example a portion of northeastern India and northeastern China. However, over the geographical regions where the atmospheric signal among the models agrees with the observed anomalies, the sign of observed anomalies are found to be consistent with all of the AGCM responses, and one may be tempted to attribute the observation to the SST-forced signal. However, the question remains open as to whether the latter agreement could be coincidental.

In seeking an attribution for the observed anomalies, it becomes important to examine their location within the PDF formed by all simulated anomalies associated with the boundary forcing, rather than just the PDF's mean value. It is necessary therefore to express the attribution of individual observed seasonal mean anomalies in probabilistic terms. Under the current Consortium activity there is an ensemble of approximately 80 simulations for each season. Similar to the agreement plot in Fig. 4, maps showing the spatial distribution of probability for AGCM simulated anomalies to be above or below given thresholds can be analyzed. This probability would be approximated as the number of ensemble members in the given interval divided by the total number of ensembles. It is emphasized that in doing a probabilistic evaluation using the full PDF of simulations, a large (order ~100) ensemble size is required (Sardeshmukh et al. 2000), and correction of systematic errors on an AGCM-specific basis is desirable.

b. Postmortem of Routine Operational Climate Predictions
When coordinated across many models and executed in real time, AGCM simulations can improve the forecast process by clarifying the possible causes for the success or failure of operational seasonal climate predictions.

Consider a forecast for the Northern Hemisphere winter season of DJF, made at a one-half month lead time. Atmospheric and oceanic observations for October would become available during the first 10 days of November, enabling SST predictions for the coming several months to be made. Using these SST predictions for DJF, dynamical (or empirical) predictions of the atmosphere for this season would be made in mid-November. In early March, observations of both the SST and the climate for DJF would become available. As a part of Consortium activity, the ensembles of AGCM simulations are then run using the observed SSTs. Thus, three sets of climate data are present for the finished season: (1) atmospheric predictions by AGCMs using predicted SSTs, (2) simulations by AGCMs using observed SSTs, and (3) observations.

Three possible reasons for a particular forecast's inaccuracy may be identified:

•   The SST predictions have errors that can lead to errors in the subsequent atmospheric forecasts.
•   The prediction models have errors in their atmospheric responses, such that even when the SSTs are predicted correctly, their seasonal predictions have errors. In the case of AGCMs, these would be due to model biases, and may also include the possibility that the 2-tiered design itself is flawed as a prediction practice. In the case of empirical models, these would be due to sampling limitations of the observed data used to train the model, as well as lack of satisfaction of the empirical models' assumptions (e.g. linearity, normality).
•   The observed climate anomalies may have little or no signal associated with the SSTs (or other boundary-forcings), despite well predicted SSTs and AGCMs having realistic atmospheric responses to the SSTs. This would imply a large random, unpredictable component in the observed seasonal mean climate state, and would indicate a large uncertainty in forecasts.

It would not be possible to identify which factors among the three listed above was responsible for the accuracy of predictions if only the predictions and verifications were available; it is important also to generate the simulations based on the true, observed boundary forcings. It is also important to provide such assessments in a timely manner so that the next seasonal forecast cycle can benefit from the thorough evaluation, not only of the causes for prior forecast success or failure, but also the role that the most recent ocean states have played in controlling the trajectory of seasonal climate.[6]
In the approach used in the Seasonal Diagnostics Consortium, the first reason listed above is examined by forcing the AGCMs with observed SSTs to assess the contribution to the error caused by the imperfectly predicted SSTs. (It should be noted that empirical methods can also be used in this attribution effort, though with the various caveats regarding the limitations of that approach in cases of weak, or historically unsampled forcings.) If that reason can be eliminated, we then find where there is anomaly sign unanimity among the climate simulations of the AGCMs. Where there is unanimity, the second reason is substantially reduced because inferences about the atmospheric signals are made using several AGCMs, effectively filtering many of the individual model

---

[6] However, extrapolation of aspects of the most recent season's outcome to a forecast for the coming season requires a careful consideration of the seasonal dependence of the SST-forced climate signal.

biases. By eliminating these two factors, it is possible to diagnose some poor predictions as being likely due to the observations containing a large influence from internal variability. Drawing this conclusion requires believing that the probability that the AGCMs suffer from common biases is minimal—i.e. that their biases are largely independent. The less realistic this belief, the larger the required number of individual AGCMs to attain any given probability that unanimous sign agreement occurs by chance.

c. Improving Climate Predictions
Recent evidence suggests that consolidated output from several of today's most recently developed AGCMs forced by observed SSTs provides valuable simulation guidance that may not be possible using empirical methods. Three examples of this are (1) prolonged high temperatures over North America following the 1997-98 El Niño (Kumar et al. 2001b), (2) mid-latitude drought in association with positive SST anomalies in the western Pacific in 1998-2002 (Hoerling and Kumar 2003), and (3) influence of Indian Ocean SST anomalies over the Pacific North American sector during the 1997-1998 El Niño (Farrara et al. 2000), although here only a single AGCM was used. Identifying the sources of these climate effects using empirical methods would be more difficult due to a combination of two factors: First, historical data needed to train empirical models are scant over parts of the affected regions—e.g. southwestern Asia. The second factor is that two of these occurrences (post-1997-98 El Niño warmth and 3-year southwestern Asia drought) are unprecedented in the most recent several decades of historical data.
The above examples suggest that a routine climate attribution effort using AGCMs will build up our knowledge base of relationships between SST forcings and their atmospheric responses and complement efforts based on empirical analysis alone. By doing so, it may lead to improvements in the seasonal predictions at later times when SST anomalies similar to those documented recur.
However, prescribed observed SSTs are required in obtaining the above dynamical diagnostic results—SSTs that are not perfectly known in advance for real-time forecasting. The impact of imperfectly predicted SST on climate prediction skill can be estimated from the skills of AGCMs forced with predicted versus observed SST (e.g. Goddard and Mason 2002). This should help forecasters be aware of the possible range of influences of errors in the predicted SST fields on the seasonal forecasts.

d. Nowcasting Climate Change and Implications for Seasonal Predictions
A prediction tool used as input to climate forecasts for the U.S. at the NOAA CPC, called optimum climate normals (OCN; Huang et al. 1996), computes the average atmospheric anomaly over the most recent 10 (15) years for surface temperature (precipitation) for the season in question, with the anomaly defined in terms of the current 30-year climatological base period. This anomaly is specified as a prediction for the subsequent season, and can be broadly interpreted as a forward extrapolation of recent climate trends. It is noteworthy that the historical skill of hindcasts of OCN is similar to that of any of NOAA's other prediction tools, including their statistical and dynamical models that focus on the effects of forcings having shorter interannual time scales. However, it is not clear what proportion of the skill of OCN is an artifact of statistical sampling (Wunsch 1999) vs. being physically based, e.g. as influenced by gradual changes in SSTs.

Recent studies have shown that some climate anomalies may be related to decadal variability or trends in SST anomaly. A warmer Indian Ocean (Lau and Weng 1999; Giannini et al. 2003), an expanded warm pool in the tropical Pacific (Barlow et al. 2002), and other warmed portions of the world's oceans are likely causes for certain recent climate trends (Hoerling et al 2001). The example of recent multi-year drought in part of southwestern Asia and North America discussed in Hoerling and Kumar (2003), related to a positive SST anomaly pattern in the western tropical Pacific more commonly observed in recent years, implies a greater likelihood for future drought in these regions than the observed relative frequency in the past. If physical connections in the low-frequency trends in the boundary conditions and atmospheric climate anomalies exist, the Seasonal Diagnostics Consortium activity can "nowcast" the evolution of the longer-term climate system by recognizing and understanding such associations. This low frequency climate behavior may be related to interdecadal variations in the boundary conditions, or may involve changes in anthropogenic forcings such as greenhouse gas concentrations (Timmerman et al. 1999; IPCC 2001), or both. In any case, a coordinated set of AGCM simulations can provide a "nowcast" of an atmospheric response to aspects of SST identified as being low-frequency.

e. Evaluation of Model Updates for Seasonal Predictions

An aspect of seasonal prediction of key importance is the signal-to-noise (S/N) ratio, which represents the relative proportion of the climate variability that is potentially predictable for a given location and time of year, given perfect knowledge of the external forcings. The predictable portion (the signal) is dependent on SST or other boundary condition, external to the atmosphere. The remainder of the climate variability is related to fluctuations internal to the atmosphere (the noise), which is generally unpredictable beyond the first two weeks except in those cases in which the internal variability may depend to some extent on the signal (Renwick and Wallace 1996; Sardeshmukh et al. 2000; Compo et al. 2001; Schubert et al. 2001)[7] or to some degree from long-lived intra-atmospheric phenomena (e.g. MJO, stratospheric QBO). In locations and seasons having negligible contribution from boundary condition forcing, predictability is lacking and climate variations are related mainly to noise. As dynamical models are continually updated, assessments of their seasonal predictability are an important aspect of their evaluation process. The reliance of the coordinated Consortium activity on the output of AGCMs makes possible evaluations of individual AGCMs—whether for S/N, base period correlations with observations, deviation from the majority response of the AGCM set, or other attributes.

6. Conclusion: Future Vision

It is envisioned that in the future, climate prediction will have greater input from dynamical methods either based on a one-tier or two-tier approach. One can also envision intermediate ("1.5-tier") approaches in which SSTs are specified over a portion of the oceans, such as the ENSO domain, but a coupled (1-tier) model is employed over the remainder of the world oceans. Implementation of dynamical approaches, however, requires confirmation that such tools deliver skill and probabilistic reliability similar to or

---

[7] It has been shown, however, that a large majority of the impact of boundary forcing is on the mean of the atmospheric response, with a more subtle impact on the dispersion about the mean (Kumar et al. 2000).

greater than that of empirical approaches—either in general or under specific conditions that are identifiable at the time of the prediction. It is hoped that the Seasonal Diagnostics Consortium activities will lead to greater acceptance in the use of dynamical methods to better understand and predict climate variations on seasonal to decadal time-scales.

In the first stage of the Consortium activity, the coordinated set of AGCM simulations only includes the forcing of AGCMs using global observed SST. In the future, it is planned to perform "targeted" experiments that would include a set of simulations forced using SST anomalies restricted to single ocean basins. The results may enhance the attribution aspect of the Consortium activity, and help increase our understanding of the relative importance of SST, and its prediction, in different ocean basins. More detailed analyses of the typical limitations imposed by using predicted SST in real-time forecasts would naturally follow. Finally, the set of AGCM simulations can be extended to include evolution of anthropogenic forcings to help climate nowcasting efforts, thereby further improving seasonal climate predictions.

Institutions participating in the ARCs Seasonal Diagnostics Consortium (sidebar; goes with bottom of page13 or top of page 14).

The participating institutions are the NOAA Climate Diagnostics Center (CDC; Boulder, Colorado), the NASA Seasonal-to Interannual Prediction Project (NSIPP[8]; Greenbelt, Maryland), the International Research Institute for Climate Prediction (IRI; Palisades, New York), the National Centers for Environmental Prediction (NCEP; Camp Springs, Maryland), the Center for Ocean-Land-Atmosphere Studies (COLA; Calverton, Maryland), the National Center for Atmospheric Research (NCAR; Boulder, Colorado), Scripps Institution of Oceanography (SIO; La Jolla, California), and the Geophysical Fluid Dynamics Laboratory (GFDL; Princeton, New Jersey).

---

[8] The NSIPP organizational name was changed to GMAO during 2003.

References

Anderson, J., 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. Am. Meteor. Soc.*, **80**, 1349-1359.

Anderson, J. A., and Coauthors, 2003: The new GFDL global atmosphere and land model AM2/LM2: Evaluation with prescribed SST simulations. *J. Climate*, **16**, accepted.

Bacmeister, J., P.J. Pegion, S.D. Schubert and M.J. Suarez, 2000: Atlas of Seasonal Means Simulated by the NSIPP 1 Atmospheric GCM. NASA/TM-2000-104505, Vol. 17.

Baldwin, M. P., D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, 2003: Stratospheric memory and extended-range weather forecasts. *Science*, **301**, 636-640.

Barlow, M., H. Cullen, and B. Lyon, 2002: Drought in central and southwest Asia: La Niña, the warm pool, and Indian Ocean. *J. Climate*, **15**, 697-700.

Barnett, T. P., 1981: Statistical prediction of North American air temperatures from Pacific predictors. *Mon. Wea. Rev.*, **109**, 1021-1041.

Barnston, 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Climate*, **7**, 1513-1564.

Barnston, A.G. and T.M. Smith, 1996: Specification and prediction of global surface temperature and precipitation from global SST using CCA. *J. Climate*, **9**, 2660-2697.

Barsugli, J. J., and P. D. Sardeshmukh, 2002: Global atmospheric sensitivity to tropical SST anomalies throughout the Indo-Pacific basin. *J. Climate*, **15**, 3427-3442.

Bengtsson L, U. Schlese, E. Roeckner, M. Latif, T.P. Barnett and N.E. Graham, 1993: A two tiered approach to long-range climate forecasting. *Science*, **261**, 1026-1029.

Compo, G. P., P. D.Sardeshmukh, and C. Penland, 2001: Changes of subseasonal variability associated with El Niño. *J. Climate*, **14**, 3356-3374.

Cunha, M. D., 2003: Water systems planning: The optimization perspective. *Engineering Optimization,* **35**, 255-266.

Dilley, M. 1997: Climatic factors affecting annual maize yields in the Valley of Oaxaca, Mexico. Int. J. *Climatol.,* **17**, 1549-1557.

Farrara, J. D., C. H. Mechoso, and A. W. Robertson, 2000: Ensembles of AGCM two-tier predictions and simulations of the circulation anomalies during winter 1997-98. *Mon. Wea. Rev.*, **128**, 3589-3604.

Gates, W. L., and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Am. Meteor. Soc*., **80**, 29-55.

Giannini, A., R. Saravanan, and P. Chang, 2003: Oceanic forcing of Sahel rainfall on interannual to interdecadal time scales. *Science*, **302**, 1027-1030.

Goddard, L., and N. E. Graham, 1999: The importance of the Indian Ocean for simulating rainfall anomalies over eastern and southern Africa. *J. Geophys. Res*., **104**, 19099-19116.

Goddard, L. and S. J. Mason, 2002: Sensitivity of seasonal climate forecasts to persisted SST anomalies. Climate Dynamics, 19:619-632, DOI 10.1007/s00382-002-0251-y.

Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane, 2001: Current approaches to seasonal to interannual climate predictions. *Int. J. Climatol*., **21,** 1111-1152.

Graham, N. E., T. P. Barnett, R. Wilde, M. Ponater, and S. Schubert, 1994: On the roles of tropical and midlatitude SSTs in forcing interannual to interdecadal variability in the winter Northern Hemisphere circulation. *J. Climate*, **7**, 1416-1441.

Graham, N. E. and T. P. Barnett, 1995: ENSO and ENSO-related predictability. Part II: Northern Hemisphere 700-mb height predictions based on a hybrid coupled ENSO model. *J. Climate*, **8**, 544-549.

Graham, R. J., A. D. L. Evans, K. R. Mylne, M. S. J. Harrison, and K. B. Robertson, 2000: An assessment of seasonal predictability using atmospheric general circulation models. *Quart. J. Roy. Met. Soc*., **126**, Part B, 2211-2240.

Hack, J.J., J.T. Kiehl and J.W. Hurrell, 1998: The hydrologic and thermodynamic characteristics of the NCAR CCM3. *J. Climate*, **11**, 1179-1206.

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

Hammer, G., J. W. Hansen, J. G. Phillips, J. W. Mjelde, H. S. J. Hill, and A. Potgieter, 2001: Advances in application of climate prediction in agriculture. *Agricult. Systems*, **70,** 515-553.

Hastenrath, S., 1995: Recent advances in tropical climate prediction. *J. Climate*, **8**, 1519-1532.

Higgins, R. W., A. Leetmaa, Y. Xue, and A.. Barnston, 2000: Dominant factors influencing the seasonal predictability of U.S. precipitation and surface air temperature. *J. Climate*, **13**, 3994-4017.

Hoerling, M.P., A. Kumar, and M. Zhong, 1997: El Niño, La Niña, and the nonlinearity of their teleconnections. *J. Climate*, **10**, 1769-1786.

Hoerling, M.P., J.S. Whitaker, A. Kumar, and W. Wang, 2001: The mid-latitude warming during 1998-2000. *Geo. Res. Lett.*. **28**, 755-758 .

Hoerling, M.P. and A. Kumar, 2003: The perfect ocean for drought. *Science*, **299**, 691-694.

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

Hoskins, B.J., and D. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.,* **38**, 1179-1196.

Huang, J., H.M. Van den Dool and A.G. Barnston, 1996: Long-lead seasonal temperature prediction using optimal climate normals. *J. Climate*, **9**, 809-817.

Hurrell, J.W., J.J. Hack, B.A. Boville, D.L. Williamson and J.T. Kiehl, 1998: The dynamical simulation of the NCAR Community Climate Model version 3 (CCM3). *J. Climate*, **11**, 1207-1236.

IPCC (Intergovernmental Panel on Climate Change), 2001: *Climate Change 2001. The scientific basis*. Eds. J. T. Houghton, et al. Cambridge University Press, Cambridge, U.K., 881 pp.

Kanamitsu, M., A Kumar, H.M.H. Juang, J.K. Schemm, W.Q. Wang, F.L. Yang, S.Y. Hong, P.T. Peng, W. Chen, S. Moorthi, and M. Ji, 2002: NCEP dynamical seasonal forecast system 2002. *Bull. Am. Meteor. Soc.*, **83**, 1019-1037.

Kiehl, J.T., J.J. Hack, G.B. Bonan, B.A. Boville, D.L. Williamson and P.J. Rasch, 1998: The National Center for Atmospheric Research Community Climate Model*, J. Climate*, **11**, 1131-1149.

Krishnamurti T. N. C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran , 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548-1550.

Kumar, A., and M. P. Hoerling, 2000: Analysis of a conceptual model of seasonal climate variability and implications for seasonal prediction. *Bull. Am. Meteor. Soc.*, **81**, 255-264.

Kumar, A., A. G. Barnston and M. P. Hoerling, 2001a: Seasonal predictions, probabilistic verifications, and ensemble size. *J. Climate*, **14**, 1671-1676.

Kumar, A., M. Hoerling, M. Ji, A. Leetmaa, and P. Sardeshmukh, 1996: Assessing a GCM's suitability for making seasonal predictions. *J. Climate*, **9**, 115-129.

Kumar, A., A. G. Barnston, P. Peng, M. P. Hoerling, and L. Goddard, 2000: Changes in the spread of the variability of the seasonal mean atmospheric states associated with ENSO. *J. Climate*, **13**, 3139-3151.

Kumar, A., W.Q. Wang, M.P. Hoerling, A. Leetmaa and M. Ji, 2001b: The sustained North American warming of 1997 and 1998. *J. Climate*, **14**, 345-353.

Landman, W. A., and S. J. Mason, 1999: Operational long-lead prediction of South African rainfall using canonical correlation analysis. *Int. J. Climatol.*, **19**, 1073-1090.

Lau, K. M., and H. Y. Weng, 1999: Interannual, decadal-interdecadal, and global warming signals in sea surface temperature during 1955-97. *J. Climate*, **12**, 1257-1267.

Livezey, R. E., and A. G. Barnston, 1988: An operational multifield analog antianalog prediction system for United States seasonal temperatures. 1. System design and winter experiments. *J. Geophys. Res* (A), **93**, 10953-10974.

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

Madden, R. A., and P. R. Julian, 1971: Detection of a 40-50 Day Oscillation in the Zonal Wind in the Tropical Pacific. *J. Atmos. Sci.* **28**, 702-708.

Mason, S. J., and L. Goddard, 2001: Probabilistic precipitation anomalies associated with ENSO. *Bull. Amer. Meteor. Soc.*, **82**, 619-638.

Murnane, R. J., 2004: Climate research and reinsurance. *Bull. Amer. Meteor. Soc.*, **85**, in press.

Murnane, R. J., M. Crowe4, A. Eustit, S. Howard, J. Koepsell, R. Leffler, and R. Livezey, 2002: The weather risk management industry's climate forecast and data needs: A Workshop Report. *Bull. Amer. Meteor. Soc.*, **83**, 1193-1198.

Nicholls, N., 1984: The stability of empirical long-range forecast techniques – a case study. *J. Clim. Appl. Meteor.*, **23,** 143-147.

Opsteegh, J. D., and H. M. van den Dool, 1980: Seasonal differences in the stationary response of a linearized primitive equation model – prospects for long-range weather forecasting. *J. Atmos Sci.*, **37**, 2169-2185.

Pegion, P. J., S. D. Schubert, and M. J. Suarez, 2000: An Assessment of the Predictability of Northern Winter Seasonal Means with the NSIPP 1 AGCM. NASA/TM-2000-104505, Vol. 18.

Rajagopalan, B., U. Lall, and S. E. Zebiak, 2002: Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. *Mon. Wea. Rev.*, **130**, 1792-1811.

Renwick, J. A., and J. M. Wallace, 1996: Relationships between North Pacific wintertime blocking, El Niño, and the PNA pattern. *Mon. Wea. Rev.*, **124**, 2071-2076.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analysis using optimum interpolation. *J. Climate*, **7**, 929-948.

Roeckner, E., K. Arpe, L. Bengtsson, M. Christoph, M. Claussen, L. Dumenil, M. Esch, M. Giorgetta, U. Schlese and U. Schulzweida, 1996: The atmospheric general circulation model ECHAM4: Model description and simulation of present-day climate. Max Planck Institut fur Meteorologie, Report no. 218, Hamburg, Germany, 90 pp.

Sardeshmukh, P. D., G. P. Compo, and C. Penland, 2000: Changes of probability associated with El Niño. *J. Climate*, **13**, 4268-4286.

Schubert, S. D., M. J. Suarez, Y. H. Chang, and G. Branstator, 2001: The impact of ENSO on extratropical low-frequency noise in seasonal forecasts. *J. Climate*, **14**, 2351-2365.

Schubert, S. D., M. J. Suarez, P. J. Pegion, M. A. Kistler and A. Kumar, 2002: Predictability of zonal means during boreal summer. *J. Climate*, **15**, 420-434.

Shukla, J, 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, 2000: Dynamical seasonal prediction. *Bull Am. Meteor. Soc.*, **81**, 2593-2606.

Thomson, M. C., T. Palmer, A. P. Morse, M. Cresswell, and S. J. Connor, 2000: Forecasting disease risk using seasonal climate predictions. *The Lancet,* **355**, 1559-1560.

Timmermann, A, J. Oberhuber, A. Bacher, M. Esch, M. Latif, and E. Roeckner 1999: Increased El Niño frequency in a climate model forced by future greenhouse warming. *Nature*, **398**, 694-697.

Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res. (O),* **103**, 14291-14324.

Wagner, A. J., 1989: Medium- and long-range forecasting. *Weather Forecasting*: **4**, 413-426.

Ward, M.N., 1998: Diagnosis and short-lead time prediction of summer rainfall in tropical North Africa at interannual and multidecadal timescales. *J. Climate*, **11**, 3167-3191.

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

Wehner, M. F., 2000: A method to aid in the determination of the sampling size of AGCM ensemble simulations. *Clim. Dyn.*, **16**, 321-331.

Wunsch, C., 1999: The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations. *Bull. Am. Meteor. Soc.*, **80**, 245-255.

Yang, S., K.-M. Lau, and R. M. Sankar, 1996: Precursory signals associated with the interannual variability of the Asian summer monsoon. *J. Climate*, **10**, 1787-1820.

Table 1. Some basic characteristics of the AGCMs used in the discussion. Model Type refers to whether quantities are stored and expressed on the basis of spectral wave numbers or by fixed grid point locations. Resolution refers to the spatial scale of the smallest details able to be captured by the model. For spectral models, T stands for "triangular" truncation, and the following number refers to the fineness of horizontal resolution. (T42, for example, implies 2.81 degrees latitude and longitude resolution). For grid point models, resolution indicates how far adjacent grid points are apart from one another. L refers to vertical resolution—specifically, the number of vertical levels in the model. The number of simulations refers to the ensemble size.

Model	CCM3	NCEP/SFM	NSIPP	ECHAM4.5	GFDL
Model Type	Spectral	Spectral	Grid Point	Spectral	Grid Point
Horizontal Resolution	T40	T42	2 degr	T40	2 degr lat, 2.5 degr lon
Vertical Resolution	L18	L18	L34	L18	L18
Highest level	2.9 mb	2 mb	2.5 mb	10 mb	3 mb
Number of Simulations	20	18	9	24	10
References	Hack et al. (1998); Hurrell et al. (1998); Kiehl et al. (1998); Bacmeister et al. (2000)	Kanamitsu et al. (2002)	Pegion et al. (2000); Schubert et al. (2002)	Roeckner et al. (1996)	Anderson et al. (2003)

Figure Captions

1. Top: Ensemble mean simulation of 200 hPa height anomaly for DJF 2001-02 using a single AGCM forced by the observed global SST for the 3-month period.   Bottom: Observed 200 hPa height anomalies for DJF 2001-02. Contours are drawn at zero, 60 and −60 m; dashed lines denote negative anomalies. Orange (blue) shading denotes positive (negative) anomalies. Shading changes occur at magnitudes of 15, 30, 45, 80 and 120 m.

2. Observed 3-month mean SST anomaly for DJF 2002-03. Contours are drawn for anomaly magnitudes of 1 and 2°C. Shading changes occur at magnitudes of 0.25, 0.5, 1.0, 1.5, and 2°C.

3. Plots of surface temperature anomaly as simulated by five AGCMs for DJF 2002-03, all forced by prescribed observed SSTs for those three months. Red shading denotes positive anomalies, blue shading negative anomalies. Contours are drawn for anomaly magnitudes of 1 and 2°C. Shading changes occur at magnitudes of 0.25, 0.5, 1.0, 1.5, and 2°C.

4. Agreement plot for the observed SST-forced simulations of five AGCMs for surface air temperature over land for DJF 2002-03. Grid points where all five AGCMs simulated positive (negative) temperature anomaly are shown in yellow (blue).

5. Top: Observed surface temperature anomaly for DJF 2002-03. Contours are drawn at zero, 60 and −60 m; dashed lines denote negative anomalies. Orange (blue) shading denotes positive (negative) anomalies. Contours are drawn at each 1°C. Shading changes occur at magnitudes of 15, 30, 45, 80 and 120 m. Bottom: Six-way agreement plot, including the observations as well as the simulations of five AGCMs whose agreement over land was shown in Fig. 4.
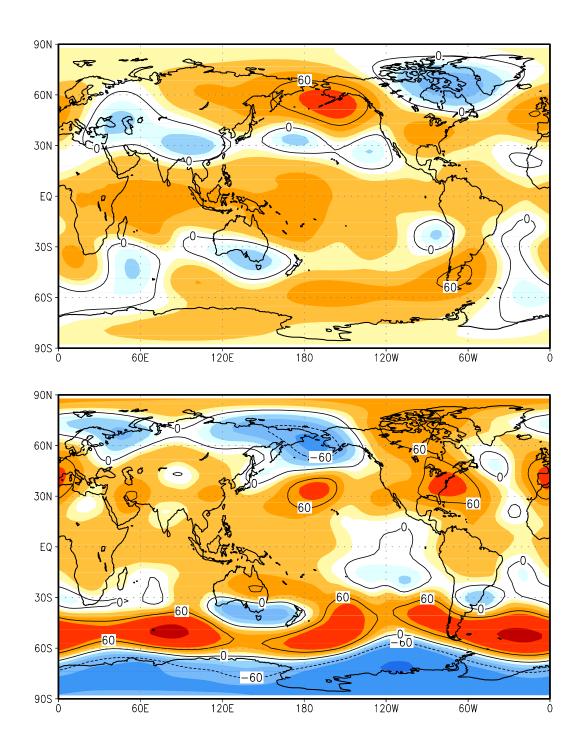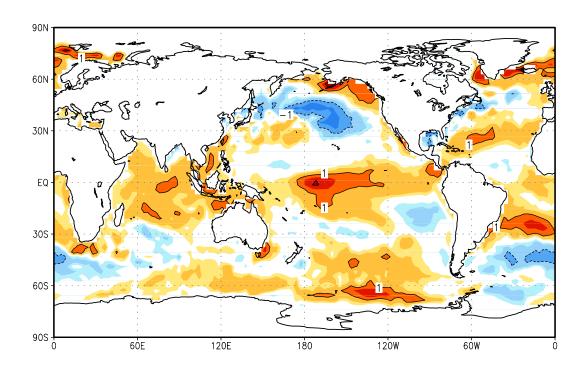
Figure 1. Top: Ensemble mean simulation of 200 hPa height for DJF 2001-02 using a single AGCM forced by the observed global SST for the 3-month period.   Bottom: Observed 200 hPa height anomalies for DJF 2001-02. Contours are drawn at zero, 60 and −60 m; dashed lines denote negative anomalies. Orange (blue) shading denotes positive (negative) anomalies. Shading changes occur at magnitudes of 15, 30, 45, 80 and 120 m.

Figure 2. Observed 3-month mean SST anomaly for DJF 2002-03. Red shading denotes positive anomalies, blue shading negative anomalies. Contours are drawn for anomaly magnitudes of 1 and 2°C. Shading changes occur at magnitudes of 0.25, 0.5, 1.0, 1.5, and 2°C.
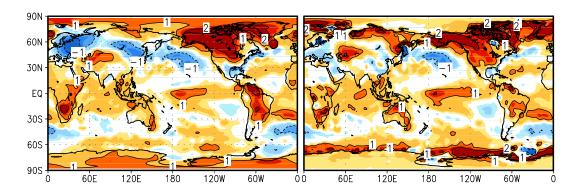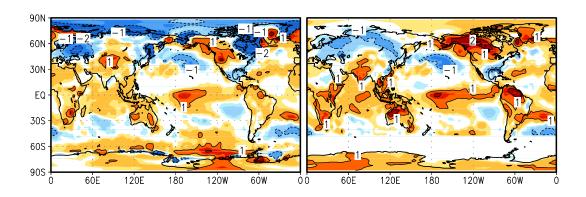
Figure 3. Plots of surface temperature anomaly as simulated by five AGCMs for DJF 2002-03, all forced by prescribed observed SSTs for those three months. Red shading denotes positive anomalies, blue shading negative anomalies. Contours are drawn for anomaly magnitudes of 1 and 2°C. Shading changes occur at magnitudes of 0.25, 0.5, 1.0, 1.5, and 2°C.
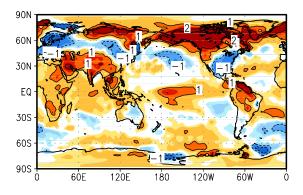
Figure 4. Agreement plot for the observed SST-forced simulations of five AGCMs for surface air temperature over land for DJF 2002-03. Grid points where all five AGCMs simulated positive (negative) temperature anomaly are shown in yellow (blue).
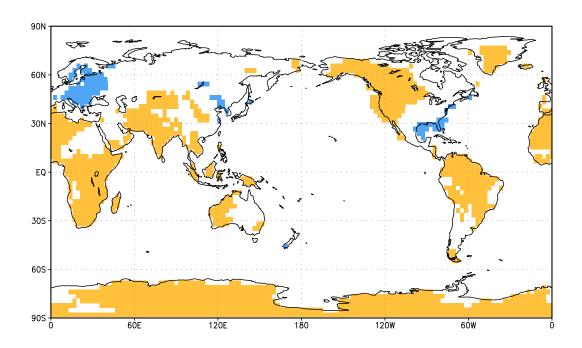
Figure 5. Top: Observed surface temperature anomaly for DJF 2002-03. Contours are drawn at zero, 60 and –60 m; dashed lines denote negative anomalies. Orange (blue) shading denotes positive (negative) anomalies. Contours are drawn at each 1°C. Shading changes occur at magnitudes of 0.25, 0.5, 1, 1.5 and 2°C. Bottom: Six-way agreement plot, including the observations as well as the simulations of five AGCMs whose agreement over land was shown in Fig. 4.