Simulation Skills of the SST-Forced Global Climate Variability of the NCEP–MRF9 and the Scripps–MPI ECHAM3 Models

PEITAO PENG

Climate Prediction Center, NCEP/NWS/NOAA, Washington, D.C.

ARUN KUMAR

Climate Modeling Branch, EMC, NCEP/NWS/NOAA, Washington, D.C.

ANTHONY G. BARNSTON

Climate Prediction Center, NCEP/NWS/NOAA, Washington, D.C.

LISA GODDARD

Climate Research Division, Scripps Institution of Oceanography, La Jolla, California

(Manuscript received 8 October 1998, in final form 15 December 1999)

ABSTRACT

The global responses of two atmospheric general circulation models (AGCM), the National Centers for Environmental Prediction–Medium Range Forecast (NCEP–MRF9) and the University of Hamburg climate model–3 (ECHAM), to simultaneous global SST forcing are examined on a 3-month timescale. Rotated principal components analysis of the model and observations is also used to identify and compare their leading modes of coherent variability. The scope of the present analyses is largely descriptive and does not attempt to explain the differences in model behavior in terms of their formulations. The authors' main focus is to quantify the simulation skill of the two comprehensive AGCMs on seasonal timescales and compare it to skill obtained using empirical prediction models.

Both models are found to exhibit realistic responses to El Niño–Southern Oscillation (ENSO)-related forcing, with the ECHAM response slightly more accurate in the spatial phasing and structure of the atmospheric anomalies. The ECHAM model exhibits realistic atmospheric responses to tropical Pacific SST forcing as well as patterns associated with extratropical internal atmospheric dynamics [e.g., North Atlantic oscillation (NAO) and a high latitude north–south dipole in the Pacific]. It shows a slightly higher signal-to-noise ratio than that found in the real world, while the NCEP model's signal-to-noise ratio is approximately equal to that in nature. The NCEP model responds with more zonally symmetric atmospheric patterns than observed, although this does not prevent it from forming realistic responses to ENSO over the Pacific–North American region. The NCEP model's NAO variability is only about half as strong as that observed.

In terms of simulation skill with respect to observations, the ECHAM model generally tends to outperform the NCEP model for global 500-hPa geopotential height and surface climate. A decomposition of the observed and model data into rotated principal components indicates that both models reproduce the ENSO-related anomalies in circulation and surface climate of the real atmosphere quite well. The ECHAM model, which handles ENSO variability and impacts slightly better than the NCEP model, shows a larger increment of capability in reproducing other global climate processes. Two linear statistical benchmarks, which are used as skill control measures, sometimes outperform the NCEP model but are more comparable, on average, to the skill of the ECHAM model. Thus as noted in other recent studies, the dynamical models and the statistical models have roughly the same simulation skill and would be expected to have similar forecast skill if the models used forecasted SSTs as their boundary conditions.

To first order, the linear component of the relationships appears to be modeled well by the two dynamical models. It is undetermined whether instances of better performance of the dynamical models than the statistical benchmarks are partly attributable to the models' effective exploitation of nonlinearities in the relationships between tropical SST and global climate. One reason for this inconclusiveness is that evidence for nonlinearities in the present analyses is not compelling. Hence, the question of whether dynamical models have untapped potential to consistently outperform statistical models on the seasonal timescale remains open and may require close examination of each physical formulation in the dynamical models.

Corresponding author address: Dr. Peitao Peng, Climate Prediction Center, W/NP51, 5200 Auth Rd, Room 604, Washington, DC 20233-9910.
Email: ppeng@ncep.noaa.gov

1. Introduction

An important goal in seasonal-to-interannual climate prediction is to be able to rely on an atmospheric general circulation model (AGCM) as a dominant climate forecasting vehicle. Currently, predictions from statistical models are still used in combination with those of dynamical AGCM models when there is a practical need to maximize the skill. We expect that dynamical models eventually will outperform models based on the empirical analysis of historical data because the former explicitly use the physical laws of the ocean–atmosphere system (with empirical tuning) while the latter can only draw on empirical, and often linear, relationships. To date, dynamical models have not consistently been able to make more skillful predictions than statistical models. Possible reasons for this may include a) dynamical model errors that compromise their skill and b) too little nonlinearity in the relationships between SST forcing and atmospheric responses. If these relationships are largely linear, then AGCM-based dynamical seasonal prediction methods will largely duplicate the predictions obtained from empirical methods.

An approach to understanding and quantifying the limits of predictability based on AGCMs is to examine the similarities and differences between their atmospheric responses to imposed SST anomalies and to compare the skill of these dynamical methods against that of empirical approaches. Here we examine the behavior of two AGCMs: the Scripps–Max-Planck Institute University of Hamburg climate model (MPI ECHAM3; Barnett et al. 1994) and the MRF9 version of the National Centers for Environmental Prediction (NCEP) Medium Range Forecast (MRF) model (Kumar et al. 1996). The analyses, based on an ensemble of multiple AGCM realizations for the period of 1958–94, are intended mainly to describe model behavior rather than explain it in terms of the physical formulations and parameterizations.

The now widely known impacts of tropical Pacific SST anomalies on global climate on seasonal timescales have been described statistically (e.g., Horel and Wallace 1981; Van Loon and Madden 1981; Barnett 1981; Lau and Chan 1983; Ropelewski and Halpert 1986, 1987; Graham and Barnett 1995; Barnston 1994). These impacts largely reflect the dynamical effects of the SST anomalies on the planetary waves (Opsteegh and Van den Dool 1980; Hoskins and Karoly 1981). The dynamics of ENSO-related extratropical climate anomalies can now be modeled with moderate accuracy (e.g., with correlation skill of ≥ 0.6 in the core ENSO response regions) for forecast purposes (e.g., Barnett et al. 1994; Ji et al. 1994; Stockdale et al. 1998). Starting with the knowledge of these SST–climate relationships, the linear contemporaneous relationship between the global atmospheric circulation and a tropical Pacific SST index for the two models is first compared with the corre-

sponding relationship in the observational data. This comparison helps reveal the extent to which the AGCMs are able to reproduce the influence of tropical Pacific SST variability on the global atmospheric circulation. Next, the skill of ensemble-averaged simulations¹ of 3-month average 500-hPa heights, precipitation, and surface temperature is evaluated and compared with the skill obtained from two linear empirical models (simple linear regression and canonical correlation analysis). In contrast to the empirical models, the skill of the ensemble-averaged AGCM simulations is not constrained by an assumption of linearity between the SST and the atmospheric response. This comparison, therefore, is useful in assessing the possible role of nonlinearities in atmospheric seasonal prediction.

2. Data and analysis method

Here, three-month mean model responses of two AGCMs to simultaneous prescribed observed SST anomalies are examined for the period of 1958–94. This study bears some similarity to Barnett et al. (1997), which examined these same two models for potential predictability for the Pacific–North American (PNA) mode of climate variability over a recent 14-yr period. However, the comparisons differ in that the present study looks at some descriptive aspects of model behavior globally using a 37-yr period, including an assessment of simulation skill for surface climate as well as for geopotential height.

Because variations in the ENSO state are responsible for a sizeable proportion of the atmospheric interannual variability, we look specifically at model responses to the SST in an ENSO-related region of tropical Pacific SST bounded by 5°N–5°S, 130°W–180°. This region, positioned 10° west of the Niño 3.4 region (Barnston et al. 1997), is highly correlated with the ENSO phenomenon, having climatological SST values near 28°C, which in this region acts as an approximate threshold for convection. This makes the region's SST a sensitive indicator of the remote teleconnections caused by changes in convection during ENSO episodes. This region will be called “the SST box.” We examine model correlations with the SST in this box despite the *global* SST being used as the models' lower boundary condition, and thus we determine to what extent the SST in the box is responsible for the global climate response.

a. Two models

The NCEP–MRF9 (Kumar et al. 1996; NMC Development Division 1988) and the Scripps–MPI ECHAM3

¹ The term “simulation,” used throughout this paper, means prediction with zero lead time (simultaneous with the forcing, in this case averaged over a 3-month time period).

(Barnett et al. 1994; Deutsches Klimarechenzentrum 1992) models, whose responses and performances are compared here, have comparable horizontal and vertical resolution (NCEP has T40 with 18 sigma levels; ECHAM has T42 with 19 sigma levels). However, the two models have a wide range of differences in their parameterization schemes. For example, while the ECHAM model uses a mass flux scheme for convective parameterization (Tiedtke 1989), a Kuo convective parameterization scheme is used in the NCEP model (Kuo 1974). Since our intention is not to diagnose differences in model behavior in terms of differences in model formulation, a detailed comparison of model formulations is not presented here. The reader is referred to the above-cited references for further details on the dynamical formulations and the parameterized physics.

b. Data

The model data come from AGCM simulations for the 1950–94 period, forced by reconstructed observed SST (Smith et al. 1996) for the lower boundary. The ECHAM model uses this SST through 1994, while the NCEP model switches to optimum interpolation (OI) data (Reynolds and Smith 1994) for the 1982–84 period without disruptive effects from the discontinuity. (By 1982, SST observations were dense enough that the two datasets are very similar.) Data for 13 individual runs of the NCEP–MRF9 model and for 10 runs from the ECHAM model are used. For observations, the National Centers for Environmental Prediction–National Center for Atmospheric Research 40-Year (NCEP–NCAR) Reanalysis (Kalnay et al. 1996) data are used for global 500-hPa height, 2-m temperature, and precipitation. While the model used for the reanalysis and the NCEP–MRF9 models both come from versions of the NCEP operational medium-range forecast model separated by several years, they are not closely related.² Hence, there should not be a noticeable verification bias in favor of the NCEP model. The reanalysis data, currently spanning back to 1958, have been demonstrated to be broadly reliable (Chelliah and Ropelewski 1998a,b), despite the presence of biases in specific regions and seasons during the early portion of the period.³ While geopo-

tential height data are available back to 1947 in much of the Northern Hemisphere (NH), this is not the case in other regions, limiting the analysis to 1958–94.

The NCEP model data contain a systematic error caused by a slow leakage of model mass, amounting to approximately 3 mb in surface pressure over the 1958–94 period. This results in a downward trend in geopotential heights but has been found to have little effect on wind, surface temperature, or precipitation. To obtain clearer results in this study, the trends in 500-hPa height are largely removed by linearly correcting that field while retaining the model's interannual variability. The correction is done by making the NCEP model's linear trend equal to that of the ECHAM model by individual grid point and season. The ECHAM model is used as the benchmark because of doubts regarding trends in the reanalysis data, particularly in geopotential height over the regions indicated above. Another reason for calibrating to the ECHAM model is our interest in model behavior differences related to interannual rather than interdecadal SST variability.

c. Analysis design

In diagnosing the models' behavior and skill, anomalies of 3-month mean 500-hPa height, 2-m surface temperature, and total precipitation are considered for the boreal winter and summer periods of January–March (JFM) and July–September (JAS), at which times the extratropical SST is near its extremes in the annual cycle and when the winter hemisphere extratropical atmosphere shows the clearest responses to the SST. Both model and observed anomalies are defined relative to the 1958–94 base period. In computing area average skill, grid points are weighted in proportion to their latitude-dependent areas represented.

Following a look at the models' interannual seasonal variances, responses to tropical Pacific SST are examined. To characterize the relationship between ENSO-related tropical Pacific SST and the atmosphere, global fields of correlation are computed between the SST in the tropical SST box (5°N–5°S, 130°W–180°) and 500-hPa height, surface 2-m temperature, and precipitation. Correlation fields are shown for each AGCM and compared with those of the observed data. Results from a concatenation of all individual model runs rather than ensemble means are used for comparability to the observations, which have only one "member."

The second set of analyses focuses on the skills of the models, comparing them with one another and with statistical models. Skills for 500-hPa height, 2-m temperature, and precipitation are discussed with respect to the globe as a whole and for subareas, including the Tropics, each extratropical hemisphere, and the PNA region. Skills are shown for the entire 1958–94 period and also for the subset of years during which a warm or cold episode of ENSO occurred based on the exceedance of a SST threshold in the tropical SST box.

² The reanalysis model has considerably higher resolution than the NCEP–MRF model and is qualitatively different. As discussed in Kalnay et al. (1996), for example, the reanalysis model used an Arakawa–Schubert convection scheme, radiation is updated every 3 h, spatial resolution is T62 horizontally and 28 levels vertically, and land surface schemes are quite different from those of the MRF. There are still other differences not mentioned here. The two models are therefore not considered to be closely related.

³ Geopotential in the NCEP–NCAR 40-Year Reanalysis appear to have some negative bias in the early part of the period in certain regions (e.g., parts of northern tropical Africa and central Asia near Mongolia), especially in NH summer. The bias gradually dissipates in the 1960s and 1970s.

Skills are presented for the NH winter (JFM) and summer (JAS). The temporal correlation coefficient is used as the primary verification score, and the statistical significance of differences between the skills of the two models and of each model and each of the statistical models is evaluated.

Rotated principal components analysis (RPCA) is used to characterize the dominant modes of model variability in the 500-hPa field based on NH data; these modes are compared to those based on observations. Information about the role of the SST in forcing, or at least associating with, these atmospheric modes is provided by maps of correlation between the SST observations and the amplitude time series of the leading model modes, again in comparison with results for observed modes.

3. Results

a. Interannual variance

The interannual standard deviations of the 500-hPa height, 2-m temperature, and precipitation of the model fields are compared with those of the observed data. Results for 500-hPa height are in Fig. 1 for a) JFM and b) JAS. In JFM, generally good correspondence between model and observed variability is shown for both models, with the ECHAM showing a slightly better pattern match with observations than NCEP (0.94 vs 0.91 pattern correlations, respectively). ECHAM's extratropical variability maxima just south of the Aleutians and in southern Greenland are quite accurately positioned with respect to those observed, as are the relative minima in the NH and Southern Hemisphere (SH) subtropics. The NCEP model shows maxima in the same vicinities, but the high latitude maxima are displaced slightly to the east, especially the southern Greenland maximum, which is also underestimated. Arctic variability is somewhat exaggerated by the NCEP model. Variability in the Aleutian maximum and in northern Alaska is somewhat higher in the ECHAM model than in the observations. Results for JAS (Fig. 2) similarly reveal satisfactory model reproduction of the observed standard deviation field (spatial correlations are 0.95 and 0.94 for ECHAM and NCEP, respectively).

b. Atmospheric response to tropical Pacific SST forcing

Given that much of the predictable portion of the global climate anomalies is linked to ENSO-related SST forcing (Bengtsson et al. 1993; Barnett et al. 1994; Lau and Nath 1994), a straightforward method of assessing the AGCMs' responses is to check the simultaneous SST versus atmosphere correlations. Correlations are examined between the average SST anomaly in the tropical Pacific box and 500-hPa height, continental 2-m surface temperature, and precipitation for the JFM and JAS pe-

riods for 1958–94. Correlations are computed using the concatenated data from the individual model runs: 37×13 (37×10) “yr” for the NCEP (ECHAM) model.

Results for 500-hPa height are shown in Figs. 2 and 3 for JFM and JAS, respectively, for the observations and the ECHAM and NCEP models. For JFM, the two models clearly reproduce the overall correlation pattern found in the observations. The ECHAM model yields somewhat stronger SST–height correlations in the Tropics and NH extratropics than found in the observations, while the NCEP–MRF model shows slightly weaker correlations than observed in the NH extratropics but higher correlations in the Tropics. Using the correlations of the individual model runs (not shown) to estimate the sampling variability of the models' correlations, it is found that all of the ECHAM model runs' total variance explained in the PNA region are well above that of the observed variance explained in JFM, with the latter falling more than 3 ECHAM standard deviations below the ECHAM's mean variance. This implies that the ECHAM model has a statistically significantly higher signal-to-noise ratio than that observed, that is, greater consistency of climate response to the lower boundary forcing. This statistically stable positive difference between the SST–height correlations for model versus observation turns out to be true also for the globe for both models in both JFM and JAS due to high correlations in the Tropics. The correlation strength of the NCEP model in JFM in the PNA region is not significantly lower than that of the observations, as 2 out of 13 NCEP runs produced a higher strength.

The model correlations with 500-hPa height for JAS (Fig. 3) also represent good approximations of those observed. Both models have higher correlations with tropical Pacific SST than observed, particularly the NCEP model in the central tropical Pacific. The ECHAM model responds more consistently than the real atmosphere in the Indian and Atlantic oceans; this is the case mainly in the Atlantic for the NCEP model. In both models, but especially in the NCEP, the meridional extent of the high correlations in the Atlantic and Pacific basins is greater than in nature. The extratropical response in the South Pacific is fairly well reproduced in both models.

While the above correlation approach to the models' responses to SST addresses their spatial distribution of signal-to-noise ratio, it ignores their response amplitudes. The amplitude would be shown by the regression coefficient as opposed to the correlation coefficient or would be shown using composite analysis. Results of regression analyses are shown for the two models (and two additional models) in Kumar et al. (2000; see their Fig. 4 and compare to observations in their Fig. 2b), where it is evident that the ECHAM model's response to ENSO-related SST forcing is stronger than that in nature, while the NCEP model's response is approximately the same as in nature. Thus, the ECHAM model

FIG. 1. Field of the interannual standard deviation of 500-hPa height for (a) JFM and (b) JAS over the 1958–94 period, as reflected in (top) the observations, (middle) the ECHAM model forced with observed SST, and (bottom) the NCEP–MRF (b9x) model forced with same. A concatenation of the individual dynamical model runs is used. Contour interval is 10 m; 30–60 m lightly shaded, 60 m and higher heavily shaded.

appears not only to have a more reliable climate signal than in reality but a higher amplitude response.

The correlation field between the SST in the tropical Pacific box and the 2-m temperature is shown for JFM and JAS in Figs. 4 and 5, respectively. The models' temperature responses closely match those of the observations at both times of the year. Correlations are high over the oceans because the models correctly apply the SST anomalies to the overlying atmosphere at 2 m.

Consistent with 500-hPa height, the ECHAM model's responses over North America are more consistent than those found in nature. The NCEP model's response in JFM is satisfactory and averages slightly weaker than that observed over certain regions, such as Asia.

In JAS the integrated global response is weaker than that in JFM in both models as well as in the observations, and both models do generally well in reproducing the observed pattern.

FIG. 2. Field of the temporal correlation between average SST in the region 5°N – 5°S , 130°W – 180° and 500-hPa height for JFM over the 1958–94 period for (top) the observations, (middle) the ECHAM model, and (bottom) the NCEP–MRF (b9x) model, where the two models are forced with observed SST. A concatenation of the individual dynamical model runs is used. Contour interval is 0.1 for $|r| \geq 0.3$. Areas having correlation magnitude of 0.33 or higher are shaded, representing two-sided local significance at the 0.05 level.

Results of the same examination applied to precipitation are shown for JFM and JAS in Figs. 6 and 7, respectively. Both models capture the broad features of the observed precipitation anomalies in JFM. Evidence of success is found in the tropical Pacific and the north subtropical Pacific, where correlations with ENSO-re-

FIG. 3. As in Fig. 2 (temporal correlation between SST in tropical Pacific box and 500-hPa height), except for JAS.

lated SST anomalies are positive and negative, respectively. The band of negative correlations in the south subtropical Pacific is somewhat underestimated by both models, while too strong a negative correlation is found over Australia. Positive precipitation correlation across the southern United States is reproduced well by ECHAM but underestimated by NCEP. The ECHAM model's response over the U.S. Great Lakes and the United States–Canadian border near the west coast is a good approximation of that observed.

In JAS, the observed region of positive precipitation

FIG. 4. Field of the temporal correlation between average SST in the region 5°N – 5°S , 130°W – 180° and continental 2-m temperature for JFM over the 1958–94 period for (top) the observations, (middle) the ECHAM model, and (bottom) the NCEP–MRF (b9x) model where the two models are forced with observed SST. A concatenation of the individual dynamical model runs is used. Contour interval is 0.1 for $|r| \geq 0.3$. Areas having correlation magnitude of 0.33 or higher are shaded, representing two-sided local significance at the 0.05 level.

correlations in the tropical Pacific is interrupted near the date line, while this is not the case for the two models. The observations indicate a long band of weak negative correlation just south of the equator near Indonesia and a positive correlation in the northern Tropics near the Philippines. Both models reproduce the negative cor-

FIG. 5. As in Fig. 4 (temporal correlation between SST in tropical Pacific box and continental 2-m surface temperature), except for JAS.

relation and the positive correlation with some westward displacement. It should be kept in mind, however, that the reanalysis precipitation data is imperfect, especially over oceans.

c. Skill with respect to observations

In examining the geographical distribution of the temporal correlation between the model simulations and the observations, we keep in mind that the model atmosphere is forced by the global SST field, and the tropical

FIG. 6. Field of the temporal correlation between average SST in the region 5°N – 5°S , 130°W – 180° and precipitation for JFM over the 1958–94 period for (top) the observations, (middle) the ECHAM model, and (bottom) the NCEP–MRF (b9x) model where the two models are forced with observed SST. A concatenation of the individual dynamical model runs is used. Contour interval is 0.1 for $|r| \geq 0.3$. Areas having correlation magnitude of 0.33 or higher are shaded, representing two-sided local significance at the 0.05 level.

Pacific SST box used in subsection B above does not play an exclusive role in the analysis. ENSO still acts as the major factor forcing the atmosphere, but the distribution of the global SST anomalies is also permitted to influence the results. Also, in contrast to the regressions in subsection B, nonlinearity may contribute to

FIG. 7. As in Fig. 6 (temporal correlation between SST in tropical Pacific box and precipitation), except for JAS.

the model skill. In evaluating simulation skill, one of the questions asked is whether the stronger signal of the ECHAM relative to the NCEP–MRF is accompanied by higher correlations between 500-hPa simulations and the corresponding observations. This would be the case if the high signal-to-noise ratio shown in the ECHAM model not only stems from the appropriate SST signals but also from the correct positioning of the related atmospheric climate features.

As a baseline comparison, two linear statistical simulations are also evaluated. The first is simple and is

based only on the ENSO state: a cross-validated version of the linear correlation between the SST in the tropical Pacific SST box used above and the field being simulated (500-hPa height, temperature, or precipitation). In cross validation, the data for the year being simulated are held out of the regression such that only the remaining years are used to define the climatology (mean and standard deviation) and the linear regression equation. The year held out is simulated as an independent case. Each year is held out this way in turn, and a temporal correlation skill is computed from the corresponding simulations and observations. The cross-validated skill is lower than the correlation between the SST and the predicted variable using all years. Skills are adjusted for a degeneracy that occurs in cross validation with regression.⁴

Canonical correlation analysis (CCA) is used as a second skill benchmark for the simulations of 500-hPa height using the global field of 3-month mean SST as the predictor. The CCA uses cross validation (and adjustment for correlation skill degeneracy) and a preorthogonalization filter, as in Barnett and Preisendorfer (1987) and Barnston (1994). CCA skills are examined for cases in which only the leading CCA mode is retained and in which the first two, three, and four modes are retained. The use of only one mode limits the predictor SST field largely to the ENSO phenomenon, residing mainly in the Tropics; this may be the most fair comparison with the dynamical models, which may not benefit materially from the specification of the extratropical SST. The smallness of the extratropical SST's feedback on the atmosphere has been documented for the North Atlantic by Delworth (1996) and more generally by Saravanan (1998). These studies, as well as those of Barsugli and Battisti (1998) and Bladé (1997, 1999) use an interactive ocean and show that extratropical SST anomalies arise mainly through the heat fluxes related to the overlying atmospheric anomalies that develop as a result of internal atmospheric dynamics and/or tropical SST forcing (Lau 1997), and these have little feedback on the atmospheric variability. CCA, however, has indirect access to consistent relationships between the atmosphere and the observed extratropical SST anomalies. The higher order CCA modes, in par-

ticular, often contain extratropical features in addition to portions of the Tropics not exhausted in mode 1. Thus, there is some ambiguity about which number of modes yields the fairest CCA benchmark for the dynamical models. We cautiously estimate that one or two CCA modes provide a reasonable statistical control, and we choose two modes as the benchmark. CCA skill results for up to four modes are shown to illustrate what might be attainable in fully coupled dynamical models of the future and to observe the increase in cross-validated linear statistical skill as a function of the number of modes to estimate the complexity of the ocean-atmosphere system for the season and region in question.

1) 500-hPa HEIGHTS

Figure 8 shows the fields of correlation between the model simulations and observations for the simple regression model and the ECHAM and NCEP models for the JFM period. The correlations of the ECHAM and NCEP models are comparable in the Tropics in that both are higher than those of the simple regression. In the extratropics (particularly over North America), the simple regression is harder to beat. While the ECHAM model appears to have a slight edge over the regression and the NCEP model falls barely short of it, the skill differences among the three forecasters are not statistically significant. It is noteworthy that the skill maps for the dynamical models are similar to the fields of linear correlation between the heights and the SST in the tropical box, shown in Fig. 2. This suggests that the relationship between tropical SST and the global atmosphere may be largely linear.

The above skill computations were repeated except that only the years whose SST anomaly in the tropical Pacific box exceeded one standard deviation in magnitude (i.e., warm or cold ENSO years) were included. Selected years for the JFM season are **1958, 1966, 1969, 1971, 1973, 1974, 1976, 1983, 1985, 1987, 1989, and 1992** (12 years, warm years in bold); for JAS they are 1963, 1964, **1965**, 1970, 1971, **1972**, 1973, 1974, 1975, **1982, 1986, 1987**, 1988, **1991**, and **1994** (15 years). The resulting correlation fields are shown in Fig. 9. Noticeably higher correlations result when only ENSO years are included. These are centered in the same locations as found for all-years skill, indicating that ENSO is the primary skill source. While overall differences in skill among the three models appear smaller for the ENSO year skills, the same relative skill results appear as found for all years in the Tropics and the extratropics.

In conducting tests of the statistical significance between the skills of any two models (dynamical or statistical), it is necessary to estimate the number of degrees of freedom, or independent, samples. Rather than testing for locally significant differences at each grid point, we opt to test certain regions, including the whole globe, in a field significance sense (Livezey and Chen 1983). The method of estimating the total degrees of

⁴ Because of a degeneracy, when cross validation is used in regression, strongly negative skills may occur when the correlation between the SST and the predictand is near zero in the full sample (Barnston and van den Dool 1993). In such cases, the interannual variability of the simulations is tiny due to the low skill and the consequent heavy damping. Because highly negative correlation scores are unwarranted for near-climatological simulations, an adjustment is made for degenerate cases in which negative correlation skills are weakened by multiplying them by the ratio of the standard deviation of the simulations to that of the observations. This adjustment is not applied to the dynamical model skills because cross validation is not used for the models, and negative skills are not related degeneratively to the standard deviation of the simulations.

FIG. 8. Field of temporal correlation between model-simulated and observed 500-hPa height for (top) the one-predictor linear statistical model (middle) the ECHAM-3 model and (bottom) the NCEP model, for the JFM period for 1958–94. Contour interval is 0.1 for $|r| \geq 0.3$. Areas having correlation magnitude of 0.5 or higher are shaded. Model simulations are from the mean of 13 ensemble members for the NCEP model and the mean of 10 members for the ECHAM mode. The local threshold for two-sided statistical significance at the 0.05 level is 0.33.

freedom for each of the regions and seasons examined here and the details of the statistical test itself are discussed in the appendix.

The geographical average skill for 500-hPa height simulation in JFM is summarized in the top portion of

FIG. 9. As in Fig. 8 (temporal correlation skill for simulating 500-hPa height in JFM), except for warm and cold ENSO cases only. The local threshold for two-sided statistical significance at the 0.05 level is 0.53.

Table 1a, in which area-averaged correlation skills are shown for the two models and two statistical controls for several subregions and for the globe. In each cell in Table 1a, skills are shown first for all years, followed by skills for only ENSO years in parentheses. Somewhat better performance of the ECHAM model relative to the NCEP model is found in most regions and for the globe. The two models perform comparably in the Tropics in JFM.

TABLE 1. Simulation skill of the NCEP–MRF model, Scripps–MPI ECHAM-3 model, and benchmark statistical models for (a) JFM and (b) JAS for 1958–94, expressed as a temporal correlation ($\times 100$). Entries in parentheses show skills for ENSO years only. Statistically significant differences between pairs of skills for the same region and variable (500-hPa height, 2-m land temperature, or precipitation) are identified by letter superscripts accompanying the correlation coefficients in the table, and the significance level is indicated as a footnote. An asterisk shows a significant skill difference between the two dynamical models. Numerical superscripts denote significant skill differences between the two statistical control models (regression and CCA). Only the CCA that uses two modes participates in the significance evaluations.

Field-season	Model	PNA region	Northern Hemisphere	Tropics	Southern Hemisphere	Globe
500 hPa	NCEP	20 (33)	18 (24)	80 (88)	23 (27)	43 (49)
	ECHAM-3	30 (40)	27 (35)	81 (88)	32^{RC} (37) ^c	49^R (56)
	Regression	25 (39)	18 (27)	70 (78)	17 (29)	37^{Rn} (47)
	CCA _{1 mode}	20 (39)	14 (25)	80 (86)	12 (13)	43 (49)
	CCA _{2 mode}	29 (46)	28 (38)	80 (86)	14 (18)	47^R (55)
	CCA _{3 mode}	32 (48)	26 (35)	80 (87)	22 (30)	49 (56)
	CCA _{4 mode}	38 (51)	31 (39)	79 (87)	21 (32)	48 (56)
Land temperature	NCEP	12 (23)	9 (12)	41 (51)	6 (2)	18 (22)
	ECHAM-3	19 (26)	20 (26)	43 (57)	12 (12)	26 (33)
	Regression	13 (23)	7 (11)	41 (49)	8 (9)	18 (23)
Precipitation	NCEP	6 (12)	7 (13)	22 (31)	4 (8)	12 (18)
	ECHAM-3	15 (20)	13 (20)	27 (35)	6 (7)	16 (22)
	Regression	13 (22)	10 (16)	15 (21)	8 (13)	11 (17)
500 hPa	NCEP	24 (31)	17^r (24)	60^r (75)	16 (24)	*33^R (44) ^R
	ECHAM-3	25 (33)	23^R (31) ^R	68^R (81) ^r	27 (31)	*41^{nR} (50) ^R
	Regression	11 (14)	7 (9)	30 (35)	16 (21)	18 (23)
	CCA _{1 mode}	10 (13)	19 (22)	67 (72)	22 (27)	40 (44)
	CCA _{2 mode}	13 (17)	22^R (26) ^r	71^R (78) ^r	27 (34)	42^{nR} (47) ^R
	CCA _{3 mode}	24 (35)	30 (36)	76 (84)	27 (35)	47 (56)
	CCA _{4 mode}	23 (36)	29 (38)	75 (84)	26 (33)	47 (56)
Land temperature	NCEP	22 (22)	16 (15)	31 (37)	9 (15)	20^r (22)
	ECHAM-3	16 (19)	13 (19)	37 (46)	16 (19)	21^r (27)
	Regression	6 (9)	6 (7)	16 (20)	6 (12)	9 (12)
Precipitation	NCEP	4 (2)	2 (1)	21 (22)	7 (14)	11 (13)
	ECHAM-3	6 (2)	6 (4)	30 (29)	11 (13)	16^r (16)
	Regression	7 (11)	6 (9)	14 (21)	10 (17)	10 (16)

r(*R*): Higher skill than regression at 95% (99%) significance.

c(*C*): Higher skill than 2-mode CCA at 95% (99%) significance.

n: Higher skill than the NCEP-coupled model at 95% significance.

*: Dynamical model skills differ at 95% significance.

The statistical significance of the difference between the skills of the two dynamical models, or either model and either of the two statistical models (the regression or the CCA using two modes), is indicated in Table 1a by bold-typed correlation coefficients for both the superior and the inferior coefficient. Significant differences between the two statistical models is also indicated, since this has implications about the simplicity of the climate anomaly structure. Significance is evaluated for one variable, region, and season at a time. The higher coefficient of a pair of coefficients that differ significantly is identified by a superscript *r*, *c*, or *n*, depending on whether it is significantly higher than the regression, the CCA (with two modes), or the NCEP-coupled model, respectively. A lowercase (uppercase) superscript denotes a significance level of 95% (99%). For example, for 500-hPa heights in JFM for the globe, although the ECHAM model outperformed the NCEP model, the 0.49 versus 0.43 skill difference is nonsignificant, having only 91% confidence. However, the ECHAM model outperformed the regression (0.49 vs 0.37) at better than 99% confidence, as indicated by the

“*R*” superscript. The regression skill is also significantly lower than the two-mode CCA skill, as indicated by the “*R*” superscript for the CCA score of 0.47. This last result suggests that for the global domain, a univariate tropical SST index does associate with as much variance in the 500-hPa field as two full-pattern CCA predictors. When the skills of the two dynamical models are statistically significantly different, an asterisk appears beside the correlation skills in Table 1.

In the PNA region in JFM, the simple regression model performs approximately as well as the NCEP and ECHAM models for ENSO years only. This outcome may be due to the relative simplicity of the ENSO impacts on this region in northern winter. The skill of the CCA using only one mode is no higher than that of the regression. The leading CCA mode contains a global ENSO pattern in the predictor SST, including the Indian Ocean (with the same polarity as that of the ENSO phase) and an oppositely phased center in the North Pacific. The inclusion of more SST information than used in the simple regression is beneficial when the additional information is relevant but would be detri-

mental if some of it (e.g., in the Indian Ocean) does not matter much to the atmosphere in the PNA region. As the number of modes included in the CCA is increased, the skill increases and eventually exceeds the skills of all other models. Recall, however, that including extratropical SST components in a benchmark for physical models forced from the ocean largely through tropical convection may not be appropriate.

In JFM, the skill of the ECHAM model is significantly greater than that of both statistical controls in the SH for all years and the two-mode CCA for ENSO years. This suggests that the ECHAM may be able to reproduce the weaker and less obvious responses to ENSO in the summer hemisphere and/or some atmospheric processes unrelated to ENSO.

Results of the same analyses applied to the JAS season are shown in Fig. 10 for all years. Skills for 500 hPa tend to be lower in JAS than in JFM. While this is most salient for the low order statistical models (regression and low-mode CCAs), the ECHAM and NCEP models also perform somewhat worse than for JFM. The CCA tends to have highest skill in the winter hemisphere and thus does better in the SH in JAS than in JFM. Overall skills for JAS (top part of Table 1b) indicate that the regression model performs poorly while the dynamical models, especially ECHAM, deliver more usable skill levels. The large skill difference between the regression (see Fig. 10, top) and the one-mode CCA in the Tropics suggests that the Pacific SST box is an inadequate representation of the ENSO state in JAS. While the skill of the regression model is much lower than that of the dynamical models in the Tropics, Fig. 10 shows that the regression is about as successful as the models in the midlatitude South Pacific ENSO-related region east of New Zealand. The two-mode CCA performs about as well as the two dynamical models. For ENSO years, only (see Table 1b), skills are higher for all models as is the case for JFM. In JAS, the ECHAM model slightly outperforms the NCEP model in all regions except for the PNA region where the two perform comparably. While nonsignificant in each region, the skill difference between the two dynamical models is statistically significant at 96% for the globe as a whole.

2) TWO-METER TEMPERATURE

In examining the temporal correlation between model simulations and observations for 2-m temperature, results over the oceans are not included because the dynamical models correctly apply local SST anomalies to the near-surface air temperature heavily (resulting in very high skills), while the regression model and the low-mode CCA models have no direct access to the local SST.

The skill results for JFM as a function of location are shown in Fig. 11, and overall results are shown in the middle portion of Table 1a. Figure 11 indicates that

FIG. 10. As in Fig. 8 (temporal correlation skill for simulating 500-hPa height in JFM for all years within 1958–94), except for JAS.

much of the available continental skill comes from the Tropics of Africa and South America, with some extratropical contributions from North America and Asia. The ECHAM model tends to perform the best of the three tools, with the NCEP and simple regression models slightly lower. The NCEP model performs poorly in eastern Asia and northern Africa, possible reasons for which are discussed in Kumar and Hoerling (1998a). A large portion of ECHAM's superiority comes from the NH rather than the Tropics, where the three models perform more comparably. All skills increase when

FIG. 11. Field of temporal correlation between model-simulated and observed continental 2-m temperature for the one-predictor linear statistical model (top), the ECHAM-3 model (middle), and the NCEP model (bottom) for the JFM period for 1958–94. Contour interval is 0.1 for $|r| \geq 0.3$. Areas having correlation magnitude of 0.5 or higher are shaded. Model simulations are from the mean of 13 ensemble members for the NCEP model and the mean of 10 members for the ECHAM mode. The local threshold for two-sided statistical significance at the 0.05 level is 0.33.

ENSO-only years are considered. It is important to recognize that while suggestive, none of the skill differences for JFM are statistically significant at the 95% confidence level. (The 0.26 vs 0.18 ECHAM–NCEP global skill difference is significant at only 90%.)

In JAS (map not shown; summarized in middle portion of Table 1b), both dynamical models show relatively high performance in simulating near-surface temperature over equatorial Africa, parts of southeast Asia, parts of tropical South America and Australia, and along much of the coastlines (especially windward). The NCEP and ECHAM models perform fairly comparably for this season. The regression model does poorly, statistically significantly falling short of the dynamical models for the globe. Exceptions to this occur in northern Australia and eastern tropical South America.

3) PRECIPITATION

The simulation skill evaluation for precipitation yielded results, as shown for JFM in Fig. 12 and as summarized for both JFM and JAS for all years and ENSO-only years in the bottom portions of Tables 1a and 1b. Skills are lower for precipitation than for 2-m temperature or 500-hPa height, as precipitation patterns are often noisier. As evident in Fig. 12, substantial JFM precipitation simulation skill occurs in the two dynamical models along much of the equatorial tropical Pacific, with additional skill off the southwest United States and northeast Brazilian coasts, in portions of the Indian Ocean, and along Africa's tropical west coast. While the regression has skill in many of these same regions in JFM, its relatively poorer performance in the Tropics makes it the lowest-scoring model for the globe as a whole. For ENSO-only years (Table 1a), all skills are increased, but model-to-model results remain mainly similar in a relative sense. The regression compares favorably against the dynamical models, however, in the PNA region for ENSO years, as was also found for 500-hPa heights. While the ECHAM model appears better than the regression and the NCEP model for the globe, the difference is insignificant (but is 92% confident against the regression).

Results for JAS (not shown geographically) are somewhat similar to those of JFM except that skill in the tropical Pacific is concentrated farther east (100°–170°W), and the tropical Atlantic and Indonesian regions become greater skill sources. The regression model underperforms the dynamical models overall due mainly to the Tropics and the subtropical North Pacific. The ECHAM model is more skillful than the other tools in the Tropics. Counting ENSO years only does not increase skills appreciably. For all years, the ECHAM model has statistically significantly higher skill for the globe than the regression and nearly significantly higher skill (92% confidence) than the NCEP model.

4. Modes of variability in the ECHAM and NCEP models

The model skill assessment given above suggests that the ECHAM model performs slightly better than the NCEP model in many regions (Table 1). For the case

FIG. 12. Field of temporal correlation between model-simulated and observed precipitation for (top) the one-predictor linear statistical model (middle) the ECHAM-3 model and (bottom) the NCEP model, for the JFM period for 1958–94. Contour interval is 0.1 for $|r| \geq 0.3$. Areas having correlation magnitude of 0.5 or higher are shaded. Model simulations are from the mean of 13 ensemble members for the NCEP model and the mean of 10 members for the ECHAM mode. The local threshold for two-sided statistical significance at the 0.05 level is 0.33.

of 500-hPa height in the JAS season, the skill difference between the two models is statistically significant for the globe as a whole. The near significance in several other cases and the large percentage of cases in which the ECHAM skill was insignificantly higher than that

of NCEP motivates a more detailed analysis of the responses of the two models. The ECHAM model's high signal-to-noise ratio, while allowing ECHAM to converge to the ensemble mean without using very many ensemble members, may be a partial explanation for its slightly higher simulation quality. However, an equally possible (or additional) explanation is the nature of the models' responses to the specified SST as well as their modes of extratropical internal atmospheric variability.

a. Rotated principal components of observed and model 500-hPa heights

To help identify the differences in model behavior that might lead to skill differences, a RPCA is applied to the JFM and JAS 500-hPa heights over the 1958–94 study period for the observations and for the NCEP and the ECHAM simulations. Because of some doubt about the accuracy of the SH data, RPCA is applied only to data from 0° – 85° N, and SH data are then projected onto the results to construct a global field. It is anticipated that a more realistic set of principal modes of atmospheric variability might indicate a potential for higher model skill. This would be true whether the variability were associated with internal dynamics, with anomalous SST-related boundary forcing, or both. Of course, having realistic modes of variability does not guarantee skill because the model must also exhibit the responses at the appropriate times. Poor model skill could result from model modes that do not resemble the observed modes (in structure and/or in location) or from poor timing of realistic principal responses. The use of RPCA for diagnostic examination of AGCM behavior is found in Renshaw et al. (1998). Our analyses were done for all of the data from the individual model runs (without ensemble averaging) and for the ensemble average data. The individual run analyses correspond better to the single-realization observations. A correlation matrix is used as input to the RPCA to ensure that the heights at all locations have equal influence on the results regardless of their interannual variances. Varimax rotation is used, and truncation is done at 10 modes, capturing about 85% of the total height variance in JFM for observed as well as model data. Ensemble mean model data with reduced total variance is near 95% for the two dynamical models. A rotation truncation at 10 modes is known to be appropriate for winter observations in the NH (Barnston and Livezey 1987). While the models are not expected to reproduce this many observed modes, satisfactory model reproduction of the leading three modes can be considered favorable, given the current state-of-the-art. In using RPCA to diagnose coherent variability structures in the observed and modeled atmosphere, the differences in the results will be interpreted with respect to what could occur due to natural sampling variability.

Table 2 summarizes the outcome of the RPCA for JFM for the observations and the two models, with mod-

TABLE 2. Percentages of variance explained in leading RPCs of JFM observations and individual and ensemble mean model simulations. Superscripts indicate rotated mode number. Percentages of variance explained by the leading 10 modes are shown below those for individual modes. Some variance statistics are shown at bottom.

Mode	Observed	ECHAM model indiv/ensemble	NCEP model indiv/ensemble
ENSO	27 ¹	30 ¹ /40 ¹	27 ¹ /36 ¹
Arctic–NH high latitude	12 ²	11 ^{3*} /22 ^{2*}	19 ² /24 ²
North Atlantic Oscillation	9 ³	11 ² /11 ³	4 ⁶ /6 ⁵
ENSO–extratropical portion	—	—	7 ³ /7 ³
% Variance of top 10 modes	86	83/95	85/94
Total data variance (m ²)	1189	1294/494	1293/300
Variance ratio: ensemble/indiv	—	0.38	0.23
Expected ratio without SST forcing	—	0.10	0.08

* In the ECHAM model, the arctic pattern undergoes a moderate structure change upon ensemble averaging.

el results shown using both simulation data from concatenated individual runs and data from ensemble mean data. The identities of the leading three observed JFM modes, indicated in the left column, include ENSO, an arctic/high latitude zonally symmetric pattern, and the NAO. The entries in Table 2 indicate the percentages of explained variance, and the superscripts indicate the mode number. The structure of the RPCA patterns using ensemble means are very similar to those for individual member model runs (with one exception to be mentioned below) but explain more variance because some of the noise has been removed in ensemble averaging. The percentages of variance explained by the first 10 modes, collectively, are indicated in the table. In the bottom part of the table, total variances are indicated, and variance ratios of before and after ensemble averaging are shown for the models. The last row shows variance ratios expected for an atmosphere determined exclusively by internal dynamics (equaling the reciprocal of the number of ensemble members), assuming that the ensemble members are independent realizations from the same underlying, approximately Gaussian distributed population. While common modes of internal variability are expected in any of the individual model runs, their random timing across runs would be expected to cause complete cancellation when averaging an infinity of ensemble runs.

By looking at the tabulated differences in explained variance for the “indiv” versus “ensemble” entries on an individual mode basis for each model, some idea can be obtained about the relative influences of SST boundary forcing (signal) versus internal atmospheric dynamics (noise). Large increases in variance explained for ensemble average data indicate that the pattern tends to occur at the same time across all ensemble members, implying a major role of SST forcing. Proportionally smaller increases, or even decreases, are expected when the mode is less consistent across the ensemble members and thus more likely to be associated with the randomly occurring internal dynamics. (This indication is only rough because the changes are also partly determined by the signal-to-noise ratio changes of the other modes

in the RPCA.) The ENSO, known to be largely associated with SST boundary forcing, shows a $\sim 33\%$ increase in explained variance in both NCEP and ECHAM models when ensemble averaging is done. By contrast, the NAO, reproduced among the leading 3 modes only in the ECHAM model, represents a slightly lower percentage of variance in the RPCA of ensemble mean data than individual model run data, and declines from mode 2 to mode 3 for ensemble mean data. The arctic mode appears to behave more like an internal dynamical model in the NCEP and more like an SST-forced mode in the ECHAM model. However, inspection of the patterns (not shown) reveals that ensemble averaging in the ECHAM model leads to an arctic pattern with a marked difference from its individual run counterpart (a strong center in the Bering Sea rather than a weaker one near Mongolia). This structural change makes a signal-to-noise comparison not meaningful, as mixing with higher order modes may have occurred in the ensemble mean data.

The next-to-last row in Table 2 shows that ensemble averaging removes a considerable portion of the variance associated with internal atmospheric activity in both models. It is apparent that much less variance survives the process in the NCEP model than in the ECHAM model, even when accounting for NCEP’s larger ensemble size (note last row in Table 2). Part of the reason for this is that the ECHAM model responds more reliably and more strongly (perhaps *too* much so; see section 3b above) to SST forcing than the NCEP model. A more minor reason, discussed below, is that the ECHAM model may have a somewhat greater variety of responses to SST than the NCEP model, the latter responding more exclusively to ENSO. (Note, however, that ENSO may not necessarily be well described by a one-dimensional index in SST.)

The RPCA spatial loading patterns of the leading three modes for JFM are shown in Fig. 13 for a) observations, b) ECHAM model simulations, and c) NCEP model simulations. The patterns of the dynamical models are based on concatenated single runs. The leading mode, representing ENSO, is satisfactorily represented

FIG. 13. Spatial loading patterns of the first three rotated principal components of 500-hPa heights for JFM over the 1958–94 period for (a) observations and individual run simulations of (b) the ECHAM model and (c) the NCEP model. All individual model run data are concatenated into one long time series. Contour interval is 0.2; the zero contour is omitted and negative contours are dashed. Values represent the correlation between the amplitude time series of the mode and the raw 500-hPa height data. The percentage of original data variance, as reflected in Table 2, is shown above each panel.

across the Tropics in both models for both individual model runs and ensemble averaging (not shown). The ECHAM model's ENSO response is particularly realistic (top panel of Fig. 13b), resembling mode 1 of the observations (top panel of Fig. 13a) in both tropical and extratropical aspects. The percentage of variance explained for the individual-run ECHAM model is slightly more than that for the observations, consistent with its

higher signal-to-noise ratio than found in nature; this is increased further through ensemble averaging, resulting in 40% of model variance explained by ENSO alone (Table 2).

The NCEP model ENSO response, while quite adequate, has slightly ill-defined extratropical features in the individual run result (Fig. 13c). With ensemble averaging, the extratropical clarity is improved (not

FIG. 13. (Continued)

shown). The shape of the PNA pattern in the NCEP's ENSO mode is basically correct. The extratropical portion of the ENSO response also appears in mode 3 of the NCEP model (bottom panel of Fig. 13c). The appearance of a PNA-like pattern in two temporally orthogonal modes in the NCEP model may reflect an SST-forced versus an internal atmospheric dynamical cause. More will be said about this in the next section. In the Tropics, the region of very high (>0.8) loadings in the two models reproduces fairly well that in nature.

Mode 2 in the observed 500-hPa heights (Fig. 13a, middle panel) is a NH arctic pattern, with a center of opposite sign north of Mongolia. This mode appears more strongly and with adequately good placement as mode 2 (mode 3) in the NCEP (ECHAM) model results for individual model run data and in mode 2 for both models using ensemble mean data but with substantial pattern modification in the ECHAM model (not shown). The variance explained by the NCEP model is especially high for both individual and ensemble mean data, and the pattern has more zonal symmetry than that of the observations, as evidenced by loadings of >0.8 in the arctic region.

Mode 3 in the observed 500-hPa heights represents the NAO. This pattern occurs only as mode 6 (not shown) in the NCEP model for the individual run RPCA while more strongly and realistically as mode 2 in the ECHAM. In the observed results the center just north of Mongolia appears as part of the arctic pattern in mode 2 while in the ECHAM model, it appears as part of the NAO. An observational study by Thompson and Wallace (1998) showed that the Mongolian center is part of the Arctic oscillation pattern, which contains the NAO in the middle atmosphere; this gives credibility to the ECHAM's mode 2 pattern and suggests that the NAO and the Arctic-NH high-latitude pattern may have some physical linkage despite their appearance on separate RPCA modes.

The RPCA diagnostics above, repeated also for JAS (not shown) and found to be fairly analogous to those for JFM, show that the ECHAM model's preferred response patterns tend, on average, to be in keeping with those of nature to a greater extent than those of the NCEP model.

To be able to place the above model differences in RPCA patterns in proper perspective, the expected sampling variability of the RPCA results must be taken into consideration. The 37 realizations of observed data result in a noticeable sampling problem when adjacent unrotated modes explain comparable percentages of variance (North

TABLE 3. Sensitivity of RPCA results to sampling variations in the dynamical model ensemble simulations.

Random sample $\rightarrow$ of 5 runs ($N = 185$)	Sample #1 ($N = 185$)	Sample #2 ($N = 185$)	Sample #3 ($N = 185$)	Sample #4 ($N = 185$)	Sample #5 ($N = 185$)	Sample #6 ($N = 185$)	Original 10-run ($N = 370$)
Mode 1 (ENSO)	30.4	30.3	29.5	31.8	30.0	30.7	30.3
Mode 2 (NAO)	9.9*	11.8	10.4*	11.8*	11.0*	10.8	11.2
Mode 3 (polar)	13.1*	10.0	11.6*	11.8*	13.1*	10.8	11.1

* Indicates that modes 2 and 3 switched order from that of the original 10-run RPCA.

et al. 1982), and while not analytically quantifiable for rotated modes, a comparable principle applies. In the case of the RPCAs of the concatenated model data, the sample is much larger and the opportunity for modes to define signals against the background noise of random atmospheric variability is more favorable. With approximately 400 model realizations, adjacent modes can be considered as statistically separated if their explained variances differ by only $(2/400)^{1/2}$ or about 7%. [Here, the North et al. (1982) rule for unrotated modes was used as a rough approximation.] To provide empirical confirmation of this expected robustness, RPCA was performed on each of 6 different randomly selected sets of 5 out of the total of 10 runs of the ECHAM model. Thus, each RPCA used 185 (i.e., 5×37) cases. The resulting patterns (not shown) looked highly similar to the eye, although modes 2 and 3 sometimes reversed order. The spatial correlation among all 15 pairs of resulting patterns averaged 0.997 for mode 1 (and 0.999 against the mode using all 10 runs), 0.97 for the NAO mode (0.98 against the 10-run mode), and 0.96 with the polar mode (0.98 against the 10-run mode). The lowest correlation among the 15 pairs for each of the 3 modes was 0.995, 0.93, and 0.89, respectively. The percentage of variance explained for each of the six RPCAs for each of the leading three modes is shown in Table 3. The curve of variance explained appears fairly stable among RPCAs. In RPCA sets 1, 3, 4, and 5, the NAO and polar modes reversed order from that of the full 10-run RPCA. This is not surprising in view of their explained variances being within 1% of each other in the full RPCA. This explanation of the virtual equality of variance did not cause the two patterns to structurally mix as evidenced by high spatial correlations among differing five-run samples. The degree of robustness seen in these tests may be used in judging whether differences between corresponding modes of the two models, discussed above, is attributable to differing model behaviors as opposed to sampling variability. The authors believe that most of the intermodel difference features are larger than what would be expected from sampling variability. In comparisons with the observed modes, more caution is necessary because a greater degree of sampling error (roughly 20%–25%) is expected in the one-member observations. Nonetheless, identification of model modes with their observed counterparts was not difficult for the leading three modes.

b. Geographical distribution of associated SST anomalies

Above, we roughly estimated the degree to which an RPCA response pattern is associated with the SST by the amount of increase in variance explained for ensemble average model data as compared with individual run data. Now we wish to characterize the SST association geographically by inspecting the distribution of the correlation field between the amplitude of the modes of atmospheric variability and the SST (Renshaw et al. 1998).

For the leading (ENSO) mode, the resulting corre-

lation fields for JFM are shown in Fig. 14 for the observations and the ECHAM and NCEP models. The correlations to the observed mode form a familiar pattern of strong SST anomalies in the tropical Pacific, except for the western portion (whose northern flank has anomalies of opposite sign), and the Indian Ocean. Both oceans are known to be involved (directly or indirectly) in the ENSO phenomenon (Rasmusson and Carpenter 1982; Pan and Oort 1983) and associated with the height pattern described by mode 1 of the observations (Fig. 13a). Figure 14 indicates that ECHAM's ENSO height pattern is forced by SST in approximately the same regions and with about the same strength as those of the real world. The NCEP model's ENSO height pattern is also forced by a nearly identical SST pattern, despite some differences in the spatial structure of its leading mode from that of the ECHAM model. It is reassuring that the two models respond to virtually the same SST, even if their responses differ.

The field of SST correlation of the ENSO RPCA mode of the ensemble average model data (not shown) is highly similar in both structure and strength to that for concatenated individual model run data. The near-equality is a result of the high spatial correlation between the individual- and ensemble-based loading patterns, and thus a high correlation between their amplitude time series, despite a greater amplitude in the latter.

The patterns of SST correlation for the second and third modes for the observations and for both models (not shown) are weak and noisy. The polar mode of the observed data shows no meaningful SST correlation, while for both models it shows a weak (0.3–0.4) west–southwest to east–northeast correlation band in the northwestern tropical Pacific from the Philippines to the north of Hawaii, with the same sign as the polar height anomaly. In both the observations and the ECHAM model, the NAO mode correlates weakly (maximizing slightly over 0.5), with a somewhat congruent SST pattern in the underlying Atlantic. The NCEP model's NAO (not shown) is adequately formed but unrealistically weak, appearing as mode 6 in the individual run RPCA (4% of variance) and as mode 5 in the ensemble average RPCA (6%). The signature of the NAO in the Atlantic SST is thought to be initiated by the atmospheric pattern rather than forcing the atmospheric pattern (e.g., Bladé 1997, 1999). After the atmospheric NAO has persisted for a while, however, it can also be maintained to some extent via local extratropical SST anomalies through hydrostatic considerations: the SST anomaly imparts a like-signed anomaly in the lower atmospheric temperature, changing the thickness and thus the 500-hPa height. This characterization can be applied to the weak SST associations of modes 2 and 3 of both models for atmospheric centers located over ocean.

Recall from Fig. 13c that an extratropical PNA-like pattern appears in isolation as mode 3 for the NCEP model. The associated SST pattern (not shown) is a weak pattern of positive SST anomalies limited to the

FIG. 14. Field of correlation between the amplitude time series of the first rotated principal component of 500-hPa heights for JFM (whose spatial loading patterns are shown in the top panels of Figs. 13a,b,c) and the SST at each grid point over the 1958–94 period for observations, the ECHAM model, and the NCEP model. For the model results, correlations are with modes of concatenated individual model runs. Contour interval is 0.2 and negative contours are dashed. Shaded areas denote correlation magnitudes of >0.325 , which are statistically significant at the 95% level; negative areas are shaded more lightly than positive areas.

eastern portion (80° – 150° W) of the tropical Pacific. This weakness, combined with the lack of increase in variance explained with ensemble averaging (Table 2), suggests that NCEP mode 3 is associated with extratropical internal atmospheric dynamics and is not a distinct component of ENSO-related SST forcing. The amplitude time series of mode 3 (not shown) indicates a weak

relationship to ENSO, with some strong ENSO years participating heavily but others not at all.

To summarize, the SST correlation fields from the RPCA modes of model data suggest that nearly all of the SST forcing occurs in association with ENSO. This is borne out in the SST correlation fields for the observations, supporting similar indications of other recent

studies (e.g., Kumar and Hoerling 1997; Bladé 1999), and indicating that the models are responding to SST boundary conditions realistically while also varying at least moderately as observed due to internal dynamics. Coherent patterns of atmospheric variability other than the ENSO-related pattern (e.g., the NAO and the polar pattern) appear largely unrelated to SST. A PNA-like pattern, devoid of its tropical component (as in NCEP's mode 3), also occurs independently of SST forcing.

5. Conclusions

The atmospheric responses of the NCEP–MRF9 and the ECHAM-3 atmospheric GCMs to simultaneous 3-month mean anomalous SST forcing, and consequent skill in simulating observed midtropospheric and surface climate anomalies, have been examined for the JFM and JAS seasons.

The ECHAM model is found to exhibit realistic atmospheric patterns both in response to SST forcing and as a result of internal atmospheric dynamics. It behaves with a slightly higher signal-to-noise ratio than that estimated for the real world. The NCEP model, whose signal-to-noise ratio is about the same as that in nature, tends to respond with more zonally symmetric atmospheric patterns than those observed. However, this does not prevent it from forming a realistic ENSO response pattern. The NCEP model has NAO variability considerably weaker than that found in nature, while the ECHAM model exhibits strong and realistic NAO variability. In an overall sense, the ECHAM model shows a somewhat higher capability to reproduce the atmospheric variability of the real atmosphere than the NCEP model. This capability difference is not significant, however, in the SST-forced ENSO atmospheric responses, particularly in the Tropics, where both models behave realistically. In many of the ECHAM versus NCEP model skill evaluations in which interannual temporal correlations between model SST-forced simulations and observations are compared, the ECHAM model has slightly higher skill than the NCEP model, but the skill difference is not statistically significant. For the globe as a whole, however, the ECHAM model's simulation superiority (0.41 vs 0.33 correlation) for 500-hPa height for the JAS season—a season whose climate is not typically dominated by ENSO as the JFM season (Kumar and Hoerling 1998b)—is significant with 96% confidence. In JFM, the ECHAM also has the better global performance (0.49 vs 0.43) but with statistically nonsignificant (91%) confidence. Model skill differences in continental temperature and global precipitation simulation, while usually favoring the ECHAM model, do not reach statistical significance. In the numerous cases in which skill differences are visible but nonsignificant, a reasonable conclusion is that the differences—even when not tiny—are not large enough to trust, given the sample sizes. A field significance test that incorporates all the individual results would need to account for their

large overlaps. The results for the globe may be considered an approximation to such a test.

While only one case of significantly different model skills appears, there are several instances in which either model (most often the ECHAM) significantly outperforms at least one of the statistical benchmarks, both for surface climate and 500-hPa height. Of the two statistical benchmarks, the two-mode CCA usually scores higher than the simple regression that uses the tropical Pacific SST box as the sole predictor. In the PNA region during JFM, however, the simple regression is competitive with the CCA as well as both dynamical models, suggesting a more univariate ENSO-driven climate system in that region and time of year.

A decomposition of the observed and model 500-hPa interannual height variability into rotated principal components for JFM shows that the ECHAM model has atmospheric responses to SST boundary conditions that resemble those of the real world to a somewhat greater extent than the NCEP model. This *may* be one explanation for its somewhat better skill performance. In simulating the boreal winter effects of ENSO-related tropical Pacific forcing in the PNA region of the NH, the NCEP model performs slightly (and statistically insignificantly) less well than the ECHAM model both during the years of ENSO extremes and the years having mild or neutral ENSO conditions. This skill difference prevails also in parts of the globe less strongly affected by ENSO or affected by other phenomena such as the Arctic oscillation or the NAO. The two linear statistical models, used as skill benchmarks, tended to perform about the same as or slightly better than the NCEP model and about the same as or slightly lower than the ECHAM model.

The only significant systematic deviation of the ECHAM model from reality found in this study is its slightly higher signal-to-noise ratio than that of the real world. Its atmospheric responses to anomalous SST boundary conditions occur more clearly than in nature. In addition, other work (Kumar et al. 2000) shows that its amplitude of response (as indicated by a regression or a composite analysis) to SST forcing is greater than that in nature. While this may enable climate forecasts to be made from fewer ensemble members, it could also cause responses to boundary forcing that are too strong.

In evaluating the models' exploitation of nonlinear relationships, it is noted that the distributions of the models' skill in simulating 500-hPa height (Fig. 8) greatly resemble those of the linear correlation between the SST in an ENSO-related tropical Pacific box (Fig. 2). This suggests that the relationship between tropical SST and extratropical climate on the seasonal timescale may be largely linear and even near-univariate. This finding was discussed in Kumar et al. (1996) with respect to the NCEP–MRF9 model. However, the skill patterns seen in Fig. 8 may come about in part from nonlinear relations, and their magnitudes could still be similar to those of Fig. 2 due to other model-related problems. The extent to which the models benefit from

using nonlinear relationships is indeterminable from our results. The fact that a successively greater number of CCA modes often increases cross-validated statistical skill suggests that the earth's SST-forced climate is more than univariate, despite that the leading modes account for the majority of the predictable variance.

The attainment of equal or slightly higher (and occasionally statistically significant) skill from the ECHAM (and, more occasionally, the NCEP) model than from statistical models is encouraging, implying that the resources and effort allocated to research in physical modeling may be reaping rewards. However, outperformance of statistical models by dynamical models is not general in this study, indicating that not enough modeling progress has been made. The questions of whether a sizeable portion of the relationship between tropical SST and the global atmosphere is nonlinear on seasonal timescales and how models can best represent nonlinear components need attention. If nonlinear components are important, improved dynamical models of the future should be able to outperform linear statistical models more consistently. If and when such success in the use of models becomes a reality, there is a huge need for better seasonal forecasts during summer and for all seasons controlled substantially by phenomena other than ENSO.

Acknowledgments. The support offered by NOAA's Climate Dynamics and Experimental Prediction Program is gratefully acknowledged. We also thank Drs. Huug van den Dool, Kingtse Mo, and the two anonymous reviewers for their thoughtful comments.

APPENDIX

Statistical Significance Tests of Differences among Model Correlation Skills

Differences among model correlation skills are tested for significance in pairs. The Fisher Z (Draper and Smith

1981) is used. Since the true correlations (i.e., as would be produced by a theoretically infinite sample of cases) are unknown for any of the models, the sampling variability of both sample correlations must be taken into account, making the correlation difference required for significance higher than if one of the correlations could be considered as an absolute (i.e., true) baseline.

To perform the Fisher Z test, it is necessary to estimate the number of independent samples in each correlation comparison. For any location, whether a point or a large region, the number of samples is the product of the number of independent time realizations and the number of independent spatial samples. In this study, the temporal sample size is assumed to equal the number of years included (37 for the "all-years" cases, 12 for "ENSO-only" cases in JFM, and 15 for "ENSO-only" cases in JAS). This is based on the fact that the interannual autocorrelation of the SST in the tropical Pacific is near zero for both the JFM and JAS seasons. The assumption would be violated if knowledge of the climate situation in a given year provided information about the likely climate situation of an adjacent year or if there were significant trends in the data. While trends and the interannual autocorrelation are not zero everywhere, they are generally low enough for the number of years to give a reasonable approximation to the temporal sample.

The number of spatial samples is one when a point in space (e.g., a grid point or station) is considered and more than one when enough additional area is included to add variability unrelated to that of the original location. The lower the radius of the spatial correlation, the less additional area is required to gain more independent spatial samples. The number of independent spatial samples for 1-month mean geopotential height in the extratropical NH has been estimated by Van den Dool and Chervin (1986) as about 15–20 for data in boreal winter and about 40 in summer. This basic in-

TABLE A1. (top) Estimated number of independent spatial samples and (bottom) number of independent total samples after multiplying by the temporal sample size by variable (500-hPa height, land temperature, and precipitation), season, region, and ENSO status used in the Fisher Z significance tests of differences in correlation skill. Temporal sample sizes are 37 for all years and 12 (15) for ENSO years only for JFM (JAS).

	JFM					JAS				
	Globe	NH	SH	PNA	Tropics	Globe	NH	SH	PNA	Tropics
Spatial sample size										
500 hPa–all	27.0	12.0	20.7	3.0	1.6	27.0	22.0	9.3	6.2	1.6
500 hPa–ENSO	24.0	8.5	17.7	1.9	1.3	24.0	20.8	7.7	5.4	1.3
Land temperature–all	20.2	9.0	12.4	2.2	1.8	20.2	16.5	5.6	4.6	1.8
Land temperature–ENSO	18.0	6.4	10.6	1.4	1.4	18.0	15.6	4.6	4.0	1.4
Precipitation–all	62.1	27.6	47.6	6.9	3.7	62.1	50.6	21.4	14.3	3.7
Precipitation–ENSO	55.2	19.6	40.7	4.4	3.0	55.2	47.8	17.7	12.4	3.0
Total sample size										
500 hPa–all	999.0	444.0	766.0	111.0	59.2	999.0	814.0	344.0	229.0	59.2
500 hPa–ENSO	288.0	102.0	212.0	22.8	15.6	360.0	312.0	116.0	81.0	19.5
Land temperature–all	749.0	333.0	460.0	83.2	64.6	749.0	610.0	206.0	172.0	64.6
Land temperature–ENSO	216.0	76.5	127.0	17.1	17.0	270.0	234.0	69.3	60.8	21.3
Precipitation–all	2298.0	1021.0	1762.0	255.0	136.0	2298.0	1872.0	791.0	528.0	136.0
Precipitation–ENSO	662.0	235.0	489.0	52.4	35.9	828.0	718.0	266.0	186.0	44.8

formation is used here to derive estimates of the number of spatial samples for the regions and seasons treated here. Allowance is made for 1) lower spatial sample for 3-month than 1-month means (based on month-to-month autocorrelation), 2) differences between NH and SH (SH generally has a somewhat lower number of independent spatial samples), 3) lower spatial sample for ENSO years alone than all years, 4) greater spatial sample for surface variables than geopotential height, 5) smaller spatial sample for continental temperature than global temperature, and 6) greater spatial sample for precipitation than for temperature. While each model has its own number of independent samples, we approximate here by using that estimated for the observations for all models also. The number of independent spatial samples and the total sample after multiplying by the temporal sample size are shown in Table A1 for the 60 combinations of variable, season, region, and ENSO status used in the significance tests performed in this study.

REFERENCES

- Barnett, T. P., 1981: Statistical prediction of North American air temperatures from Pacific predictors. *Mon. Wea. Rev.*, **109**, 1021–1041.
- , and R. W. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for North American surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.*, **115**, 1825–1850.
- , and Coauthors, 1994: Forecasting global ENSO-related climate anomalies. *Tellus*, **46A**, 381–397.
- , K. Arpe, L. Bengtsson, M. Ji, and A. Kumar, 1997: Potential predictability and AMIP implications of midlatitude climate variability in two general circulation models. *J. Climate*, **10**, 2321–2329.
- Barnston, A. G., 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Climate*, **7**, 1513–1564.
- , and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115**, 1083–1126.
- , and H. M. van den Dool, 1993: A degeneracy in cross-validated skill in regression-based forecasts. *J. Climate*, **6**, 963–977.
- , M. Chelliah, and S. B. Goldenberg, 1997: Documentation of a highly ENSO-related SST region in the equatorial Pacific. *Atmos.–Ocean*, **35**, 367–383.
- Barsugli, J. J., and D. S. Battisti, 1998: The basic effects of atmosphere–ocean thermal coupling on midlatitude variability. *J. Atmos. Sci.*, **55**, 477–493.
- Bengtsson, L., U. Schlese, E. Roeckner, M. Latif, T. P. Barnett, and N. Graham, 1993: A two-tiered approach to long-range climate forecasting. *Science*, **261**, 1026–1029.
- Bladé, I., 1997: The influence of midlatitude ocean–atmosphere coupling on the low-frequency variability of a GCM. Part I: No tropical SST forcing. *J. Climate*, **10**, 2087–2106.
- , 1999: The influence of midlatitude ocean–atmosphere coupling on the low-frequency variability of a GCM. Part II: Interannual variability induced by tropical SST forcing. *J. Climate*, **12**, 21–45.
- Chelliah, M., and C. F. Ropelewski, 1998a: Comparison of Reanalysis tropospheric mean temperatures to satellite observations. *Proc. 22d Annual Climate Diagnostics and Prediction Workshop*, Berkeley, CA, Climate Prediction Center, 206–209.
- , and —, 1998b: Low frequency variability and composite atmospheric circulation over the globe associated with ‘ENSO’ based on 40 years of NCEP/NCAR Reanalysis data. *Ninth Symp. on Global Change Studies, Namias Symp. on Status and Prospects for Climate Prediction*, Phoenix, AZ, Amer. Meteor. Soc., 192–195.
- Delworth, T. L., 1996: North Atlantic interannual variability in a coupled ocean–atmosphere model. *J. Climate*, **9**, 2356–2375.
- Deutsches Klimarechenzentrum, 1992: The ECHAM-3 Atmospheric General Circulation Model. Tech. Rep. 6, ISSN 0940-9327, 189 pp. [Available from the Modellbetreuungsgruppe, Deutsches Klimarechenzentrum, Max Planck Institut für Meteorologie, Bundesstr. 55, D-20146, Hamburg, Germany.]
- Draper, N. R., and H. Smith, 1981: *Applied Regression Analysis*, 2d ed. John Wiley & Sons, 709 pp.
- 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.
- 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. J. Karoly, 1981: The steady state linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179–1196.
- Ji, M., A. Kumar, and A. Leetmaa, 1994: An experimental coupled forecast system at the National Meteorological Center: Some early results. *Tellus*, **46A**, 398–418.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kumar, A., and M. P. Hoerling, 1997: Interpretation and implications of the observed inter-El Niño variability. *J. Climate*, **10**, 83–91.
- , and —, 1998a: Specification of regional sea surface temperatures in atmospheric general circulation model simulations. *J. Geophys. Res.*, **103** (D8), 8901–8907.
- , and —, 1998b: Annual cycle of Pacific–North American seasonal predictability associated with different phases of ENSO. *J. Climate*, **11**, 3295–3308.
- , —, M. Ji, A. Leetmaa, and P. Sardeshmukh, 1996: Assessing a GCM’s suitability for making seasonal predictions. *J. Climate*, **9**, 115–129.
- , 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.
- Kuo, 1974: Further studies of the parameterization of the influence of cumulus convection on large-scale flow. *J. Atmos. Sci.*, **31**, 1232–1240.
- Lau, K. M., and P. H. Chan, 1983: Short-term climate variability and atmospheric teleconnections from satellite-observed outgoing longwave radiation. Part I: Simultaneous relationships. *J. Atmos. Sci.*, **40**, 2735–2750.
- Lau, N. C., 1997: Interactions between global SST anomalies and the midlatitude atmospheric circulation. *Bull. Amer. Meteor. Soc.*, **78**, 21–33.
- , and M. J. Nath, 1994: A modeling study of the relative roles of tropical and extratropical SST anomalies in the variability of the global atmosphere–ocean system. *J. Climate*, **7**, 1184–1207.
- Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59.
- NMC Development Division, 1988: *Documentation of the Research Version of the NMC Medium Range Forecasting Model*, 347 pp. [Available from NCEP Environmental Modeling Division, W/ NP2, 5200 Auth Rd., Camp Springs, MD 20746, USA.]
- North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 699–706.
- 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.
- Pan, Y. H., and A. H. Oort, 1983: Global climate variations connected with sea surface temperature anomalies in the eastern equatorial

- Pacific Ocean for the 1958–73 period. *Mon. Wea. Rev.*, **111**, 1244–1258.
- Rasmusson, E. M., and T. H. Carpenter, 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Wea. Rev.*, **110**, 354–384.
- Renshaw, A. C., D. P. Rowell, and C. K. Folland, 1998: Wintertime low-frequency weather variability in the North Pacific–American sector 1949–93. *J. Climate*, **11**, 1073–1093.
- Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate*, **7**, 929–948.
- Ropelewski, C. F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362.
- , and —, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **115**, 1606–1626.
- Saravanan, R., 1998: Atmospheric low-frequency variability and its relationship to midlatitude SST variability: Studies using the NCAR Climate System Model. *J. Climate*, **11**, 1386–1404.
- Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate*, **9**, 1403–1420.
- Stockdale, T. N., D. L. T. Anderson, J. O. S. Alves, and M. A. Balmaseda, 1998: Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. *Nature*, **392**, 370–373.
- Thompson, D. W. J., and J. M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297–1300.
- Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800.
- Van den Dool, H. M., and R. M. Chervin, 1986: A comparison of month-to-month persistence of anomalies in a general circulation model and in the earth's atmosphere. *J. Atmos. Sci.*, **43**, 1454–1466.
- Van Loon, H., and R. A. Madden, 1981: The Southern Oscillation. Part I: Global associations with pressure and temperature in northern winter. *Mon. Wea. Rev.*, **109**, 1150–1162.