NOTES AND CORRESPONDENCE

Low Skill in Dynamical Prediction of Boreal Summer Climate: Grounds for Looking beyond Sea Surface Temperature

PAUL A. DIRMEYER, MICHAEL J. FENNESSY, AND L. MARX

Center for Ocean–Land–Atmosphere Studies, Institute for Global Environment and Society, Calverton, Maryland

16 June 2001 and 18 July 2002

ABSTRACT

Ensemble integrations of three general circulation models (Center for Ocean–Land–Atmosphere Studies, NCAR, and NCEP) have been performed over five different boreal summer seasons (June through September of 1986–88 and 1993–94) with prescribed observed sea surface temperature to assess the predictability of seasonal climate during the boreal summer. Beyond some inconsistent initialization of soil wetness among the models, there is no land surface contribution to predictability that can be assessed. The models show a rapid degradation of skill in global terrestrial surface temperature after the first month, and no skill in precipitation over land. Potential predictability is assessed by examining in tandem the models' skill as measured by their anomaly correlation coefficients, and the models' signal-to-noise ratio (essentially interannual versus intraensemble variance) as a measure of confidence in the results. Collocation of skill in anomaly simulation and a robust signal is a strong indicator of potential predictability. Predictability of interannual climate variations is found to be low outside the deep Tropics, and nil over land. With only SST as a driving boundary condition, the poor performance of these models during summer may indicate that one must turn to the land surface in order to harvest potential predictability.

1. Introduction

Dynamical Seasonal Prediction (DSP) is a multi-institutional research project to examine the predictability of the earth's climate on the seasonal timescale. Dynamical prediction implies the use of physically based numerical models like those used for numerical weather prediction, as opposed to statistical techniques that have historically been used for climate prediction. A special issue of the *Quarterly Journal of the Royal Meteorological Society* (2000, Volume 126, 1989–2350) was recently dedicated to DSP and its sister project in Europe called the Prediction of Climate Variations on Seasonal-Interannual Timescales (PROVOST).

The hypothesis behind DSP is that it is now feasible to extend the capabilities developed for numerical weather prediction to the seasonal scale using state-of-the-art global models of the atmosphere and land surface. Initial atmospheric conditions at the beginning of the season may provide some skill, as there is some "memory" in the atmosphere that persists for several weeks. However, the bulk of the memory of the climate system on seasonal scales comes from the surface boundary—the ocean and land. Sea surface temperature (SST) varies slowly, and may be predictable at a useful level of skill up to a year in advance (Kirtman et al. 1997). Soil moisture, snow cover, and the state of vegetation act as agents of climate memory at the land surface. It is believed that predictability at seasonal timescales will come from the memory inherent in the ocean and land.

To date, the DSP project has focused on climate predictability during the boreal winter season (Shukla et al. 2000). Given the well-documented impact of El Niño–Southern Oscillation (ENSO) during boreal winter, this season was a logical place to begin looking for skill. Whereas the response in the Tropics to anomalous surface boundary conditions is usually quite vigorous and potentially predictable (Charney and Shukla 1981), in the midlatitudes the signal can be engulfed by the noise of synoptic-scale baroclinic systems. The Pacific–North America (PNA) region during winter is an exception, particularly when tropical SST anomalies are large (Shukla 1998).

Why should winter climate be predictable in the PNA region more than elsewhere? There is no definitive answer, but its proximity and orientation relative to the tropical Pacific Ocean, where the El Niño SST fluctuations take place, appear to put it downstream of the

Corresponding author address: Dr. Paul A. Dirmeyer, COLA, 4041 Powder Mill Rd, Ste. 302, Calverton, MD 20705-3106.
E-mail: dirmeyer@cola.iges.org

large-scale wavetrain emanating from that equatorial region of anomalous heating (Shukla and Wallace 1983; Rasmusson and Wallace 1983). The mean state of the general circulation affects the refraction of such wavetrains, determining their paths and strength (Hoskins and Karoly 1981).

Climate during boreal summer appears to be somewhat less determined by tropical ocean forcing, but SST may still play an important role (Trenberth et al. 1988; Palmer and Brankovic 1989; Mo and Paegle 2000). During summer in the subtropics and midlatitudes, land surface conditions may act to modulate the response of the atmospheric circulation to SST anomalies (Fennessy et al. 1994), and errors in land surface state variables such as soil wetness can degrade climate simulations (Dirmeyer 2000). The atmosphere is not decoupled from the land as in winter, so land surface fluxes can provide a local feedback to climate anomalies (Reale et al. 2002).

In this paper, we document the results of a boreal summer rendition of DSP. This experiment is more limited in scope than the winter DSP, using three of the same climate models and covering only 5 yr. Nonetheless, it provides a glimpse into SST-forced predictability during summer. Section 2 describes the models and the summer experiment. A discussion of the simulation of boreal summer climate is given in section 3. In section 4, the ability of the models to reproduce observed interannual variations is assessed. An assessment of potential predictability is made in section 5. Conclusions are given in section 6.

2. Models and experiments

Three global general circulation models (GCMs) were integrated at the Center for Ocean–Land–Atmosphere Studies (COLA) as part of the DSP project. The National Centers for Environmental Prediction (NCEP) GCM is similar to the model used for the NCEP–National Center for Atmospheric Research (NCAR) reanalysis (Kalnay et al. 1996). It was run at a spectral resolution of T62 (1.9°) with 28 vertical levels. The NCAR Community Climate Model version 3 (CCM3; Kiehl et al. 1998) was run at T42 resolution (2.8°) with 19 levels. The COLA GCM (Kinter et al. 1997) was integrated at a resolution of R40 ($1.8° \times 2.8°$) with 18 vertical levels.

While the DSP project focused on the boreal winter season, some summer simulations were also made. Initial atmospheric conditions for all summer integrations were taken from the NCEP–NCAR reanalyses on the last days of May, in the years 1986–88 and 1993–94. These years were chosen because significant regional climate events occurred during these summers. The COLA GCM ensembles contain nine members; the NCEP and NCAR GCM ensembles have five members each. The models are integrated through the end of September. Monthly mean statistics were retained, and are compared in the following sections.

In these integrations, the NCEP GCM land surface is initialized from the NCEP–NCAR reanalysis land surface state (Kalnay et al. 1996). The COLA GCM soil wetness comes from a calculation based on the operational analysis of the European Centre for Medium-Range Weather Forecasts (ECMWF) as described by Fennessy and Shukla (1999). The NCAR CCM3 model uses climatological initial conditions at the land surface. All models initialize soil temperatures and snow cover from climatological values.

All runs use a common dataset for SST, namely the optimally interpolated (OISST) dataset of Reynolds and Smith (1994). Thus, the simulations can be thought of as an exercise in seasonal climate forecasting with perfect knowledge of the evolution of the ocean state. The DSP historical seasonal forecasts represent something of an upper estimate of the skill attainable from SST forcing by current climate models. Encouraging skill has been found during boreal winter (Shukla et al. 2000). In this note, the skill during summer is assessed.

3. Boreal summer climate

The quality of simulation of the mean climate is discussed in detail by Dirmeyer et al. (2001). To briefly summarize, all three models overestimate the global mean precipitation rate for the period of June through September. In the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) dataset, the global mean rate is 2.78 mm day^{-1}. The COLA, NCAR, and NCEP GCMs produce 3.86, 3.35, and 3.65 mm day^{-1}, respectively. In addition, the three models also show curiously similar patterns of systematic error, particularly over the oceans, with global rms errors of 2.0–2.4 mm day^{-1}. Near surface temperature errors are not as similar among models, but rms errors are large, ranging from 3.5–4 K globally among the models. Hereafter, we concentrate on the seasonal evolution of predictability.

The real target for DSP is assessment of predictability of intraseasonal to seasonal climate anomalies. Figure 1 shows how the ensemble mean spatial anomaly correlation coefficient (ACC) for surface temperature evolves from month to month during the boreal summer. ACC is calculated for each month (relative to the 5 yr considered), and then averaged over the 5 yr. The top panel is for all global land points, and the bottom panel is for the North American sector alone. Globally, there is a clear degradation in skill from the first month of the integrations to later months. This degradation is evident in regional assessments over Africa and southern Asia as well (not shown). Over North America, there is a decrease in skill from June to July, followed by an increase in the second half of the integration period. This subsequent increase is weakest for the NCEP GCM. All models show similar skill in the anomaly correlation coefficient for the 4-month seasonal mean. The degradation is likely related to the gradual loss of the infor-

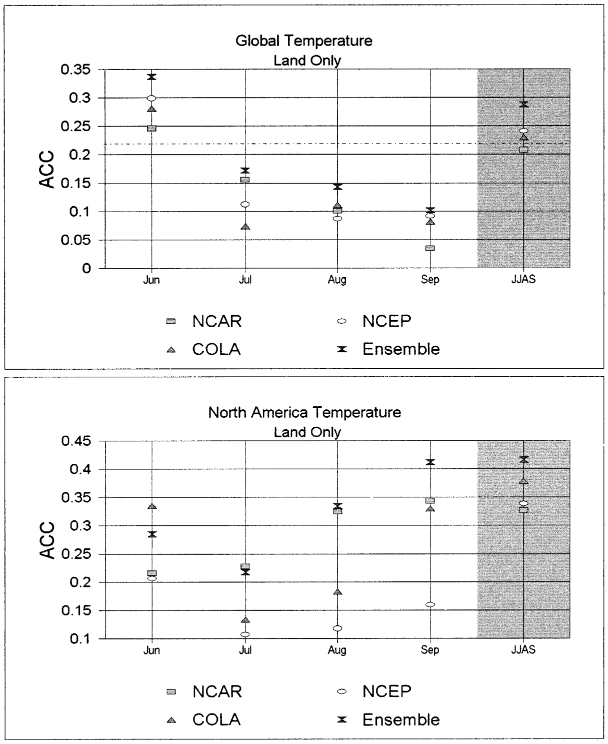

FIG. 1. Mean monthly and mean seasonal spatial anomaly correlation coefficient for surface temperature for each GCM and the intermodel mean over (top) the global (60°S–80°N) and (bottom) North American land areas. Dot–dashed line indicates the 95% significance level (all points lie below this threshold for North America).

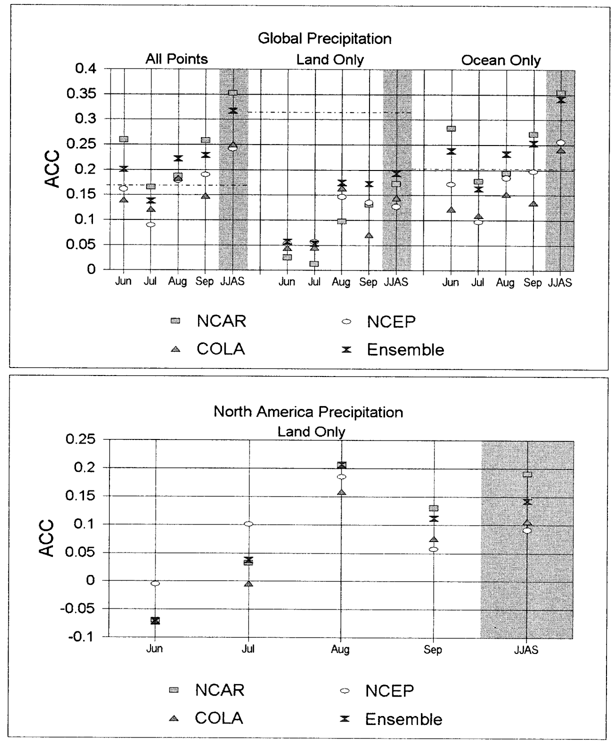

FIG. 2. As in Fig. 1, but for precipitation. The global anomaly correlation coefficients are displayed for all points between 60°S–80°N, land points only, and ocean points only.

mation contained in the initial conditions over the course of the first month as model errors grow and saturate. The reason for the increase in skill over North America going into autumn is less obvious. It may be correlated to the beginning of the transition to the wintertime circulation regime, when tropical SST may exert more control over the midlatitude circulation. On the other hand, it may be a random artifact of the small sample size of 5 yr.

A multimodel ensemble constructed from the mean of the three model ensembles is also plotted for each month and for the season. The multimodel ensemble is nearly always better than any of the members. Even for individual years (not shown), the multimodel ensemble ranks first or second among the four entries. Thus it appears that just as constructing an ensemble mean improves the simulation by a given climate model, ensembling across models may improve skill further, consistent with other similar investigations (e.g., Doblas-Reyes et al. 2000; Krishnamurti et al. 2000).

Figure 2 shows the ensemble mean anomaly correlation coefficients for precipitation, in the same manner as Fig. 1. For global precipitation, three categories are shown: land, ocean, and total. The decay of skill gained from the initial condition is rapid over land. Thus, there is no obvious skill registered during the first month, nor in subsequent months. However, over ocean there is some evidence of inertia from the initial atmospheric conditions. Skill is generally at a minimum in July, and increases in August and September. Again, this may reflect a change in the annual cycle of basic predictability of the climate system, as all three models show a similar behavior. Skill over ocean exceeds skill over land. Over North America, skill is minimum during the first month in all models, and increases to a peak in August before declining somewhat in September. The skill of the multimodel ensemble in precipitation is notable, although not as impressive as for surface temperature.

Overall, there is no significant skill over land beyond the early phase of the integrations in any of the models—any skill that exists in global ACCs is contributed from over the oceans. There is a slow loss of skill with time over land in surface temperature. Skill in precipitation over land is abysmal throughout the season.

4. Interannual variability

There are certain well-known regional climate events that occurred during the seasons simulated. How well did the models reproduce these anomalies? Figure 3 shows the anomaly in surface temperature during the drought and heat wave over North America during the summer of 1988 in the Climate Anomaly Monitoring

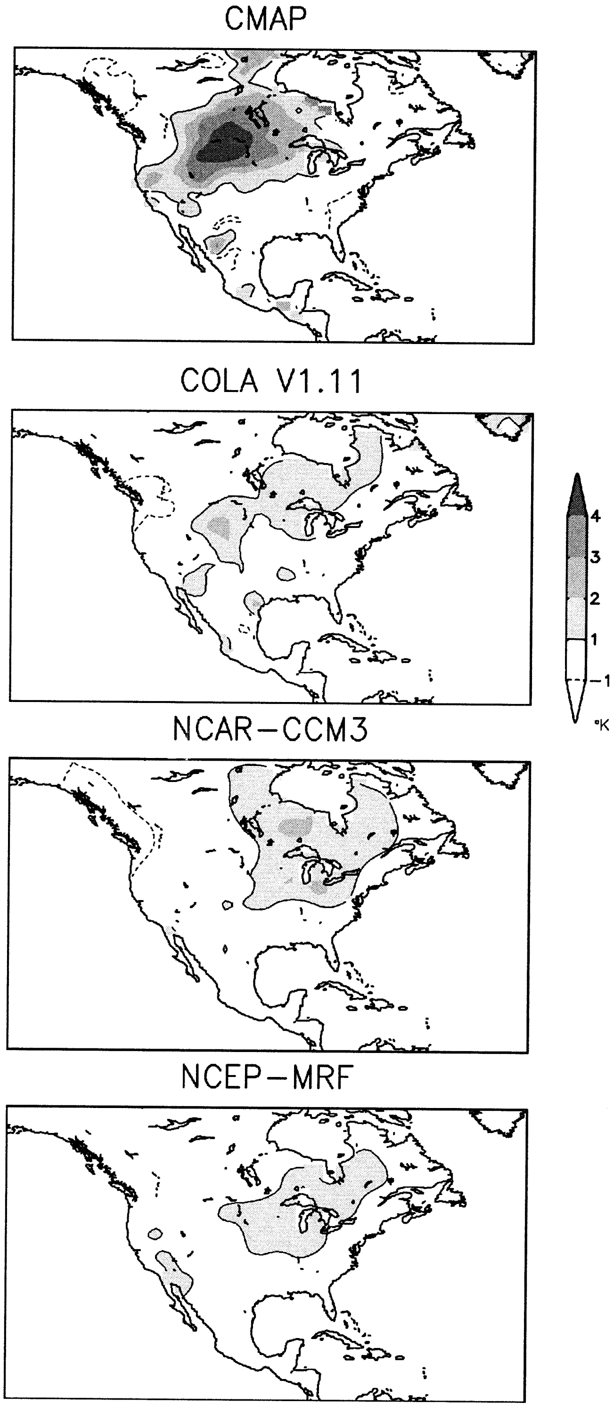

FIG. 3. Mean Jun–Jul surface temperature anomaly (K) for 1988 (compared to the other simulated years) from observations and each of the GCMs over the North American region.

Network (CAMS; Ropelewski et al. 1985) temperature data and as simulated by the three models. Anomalies are calculated for June and July only, relative to the mean of the remaining four years. The warm anomaly is very prominent in the observations over a broad area of southern Canada and the northern United States, stretching from the West Coast to the Great Lakes and Hudson Bay. Each of the models produces a warm anomaly that is weaker by at least a factor of two, and displaced to the east relative to observations. It is remarkable how similar the models are to each other. Two of the models represent the cold anomaly along the northwest coast, with the NCAR model doing the better job.

The year 1988 saw a severe drought over North America, and 1993 was a year of flooding. The difference in rainfall between the flood and drought years during June and July is shown in Fig. 4 in observations from the CMAP (Xie and Arkin 1997) and the three GCMs. It should be noted that in fact the two events did not occur at precisely the same time of year. The drought of 1988 was concentrated mostly during spring, and had begun to abate by July, as rainfall began to return to normal over much of the central United States. The floods of 1993 were largely concentrated during a 4–5-week span from mid-June to mid-July. The observed differences show the heart of the signal to be over the Missouri and upper Mississippi River valleys. There are also widespread differences of the opposite sign over the western Atlantic Ocean, Gulf of Mexico, Greater Antilles, Central America and northern South America. A dipole reflecting a shift in the latitude of the intertropical convergence zone (ITCZ) over the eastern Pacific Ocean between the 2 years is evident as well. Only the NCEP model gives some representation of the midlatitude terrestrial signal with errors in position and reduced magnitude, perhaps as a consequence of its soil moisture initialization. However, the NCEP model also simulates other extrema over the eastern Great Lakes, southeast coastline and the Prairie provinces that are not evident in the CAMS station data. The COLA GCM places the anomaly too far to the south, and the NCAR model produces a very weak and diffuse response. None of the models do well in simulating the patterns over the Atlantic off the coast of North America. However, all models do rather well in simulating the patterns south of about 25°N. Not only is the dipole of the extreme eastern Pacific ITCZ captured, but the dry 1993 versus 1988 over southern Mexico, Mesoamerica, and the Greater Antilles is well captured. Even the small positive anomalies in the Caribbean Sea are represented in each model. It seems that there is more skill at low latitudes than at midlatitudes in these models. This will be examined further below.

The years 1987 and 1988 were also contrasting years for surface hydrology over India and the Sahel: 1988 was unusually wet in both regions (actually over the Sahel, which had been suffering through an extended drought, 1988 rainfall was near normal); 1987 saw a failed monsoon over India and severe drought in the Sahel. These oscillations are evident in the difference maps for these regions (Figs. 5 and 6). Figure 5 also shows similar year-to-year variations in rainfall over Indonesia, and opposite variations over much of Indo-

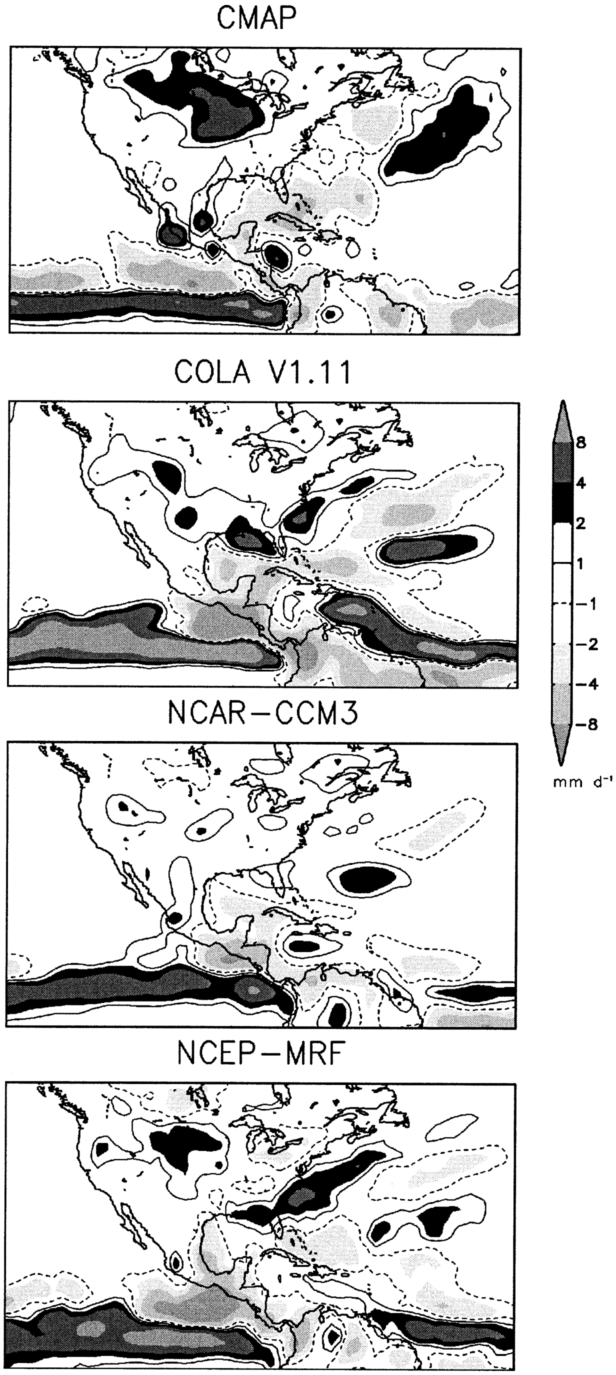

FIG. 4. As in Fig. 3, but for 1993 minus 1998 precipitation (mm day^{-1}).

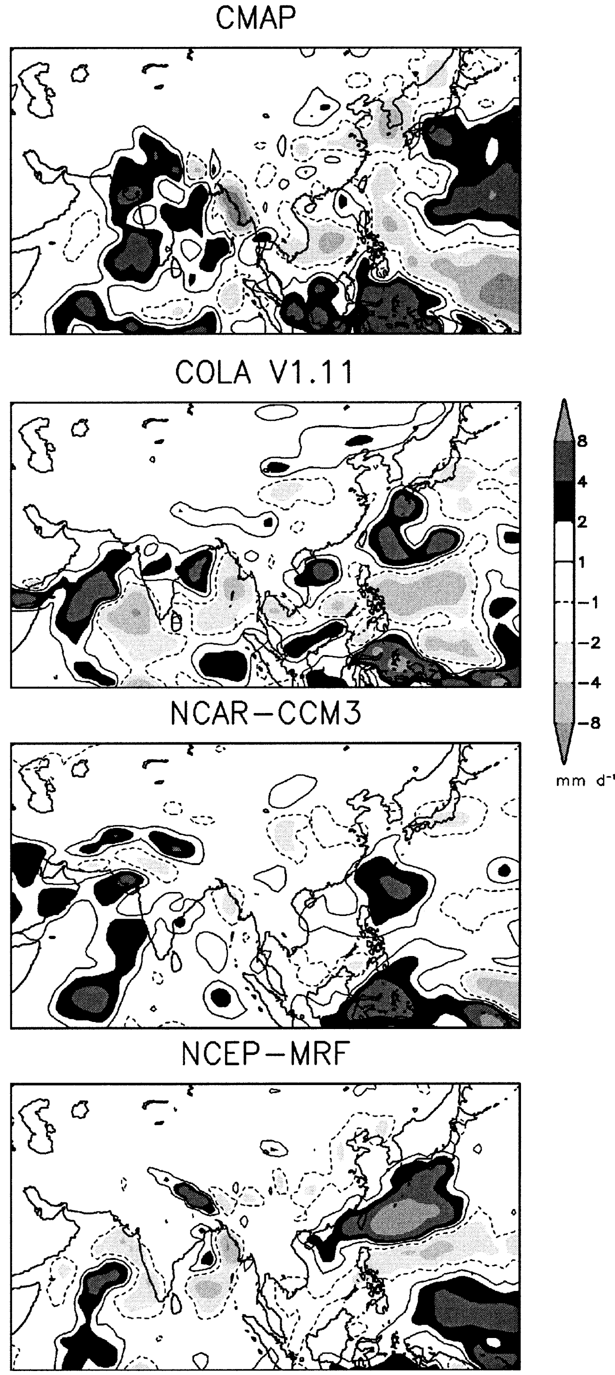

FIG. 5. As in Fig. 4, but for 1988 minus 1987 over southern Asia.

china, especially along the coast of Burma. None of the models do a credible job of simulating the pattern interannual fluctuation of the monsoon rain over the Indian subcontinent (Fig. 5). Yet all three models do appear to resolve the fluctuations over the Burmese coast and Indonesia. But other maritime areas are poorly simulated (e.g., the Philippine Sea). There also appears to be some positive signal over central China. But overall, the performance of the models is poor.

Over sub-Saharan Africa (Fig. 6), the COLA and NCEP models appear to do a fairly good job of capturing the observed fluctuation of rainfall between the two years, including the contrapuntal signal along the Guinea coast. The NCAR model does not appear to capture the signal well, perhaps due to the effect of a rather

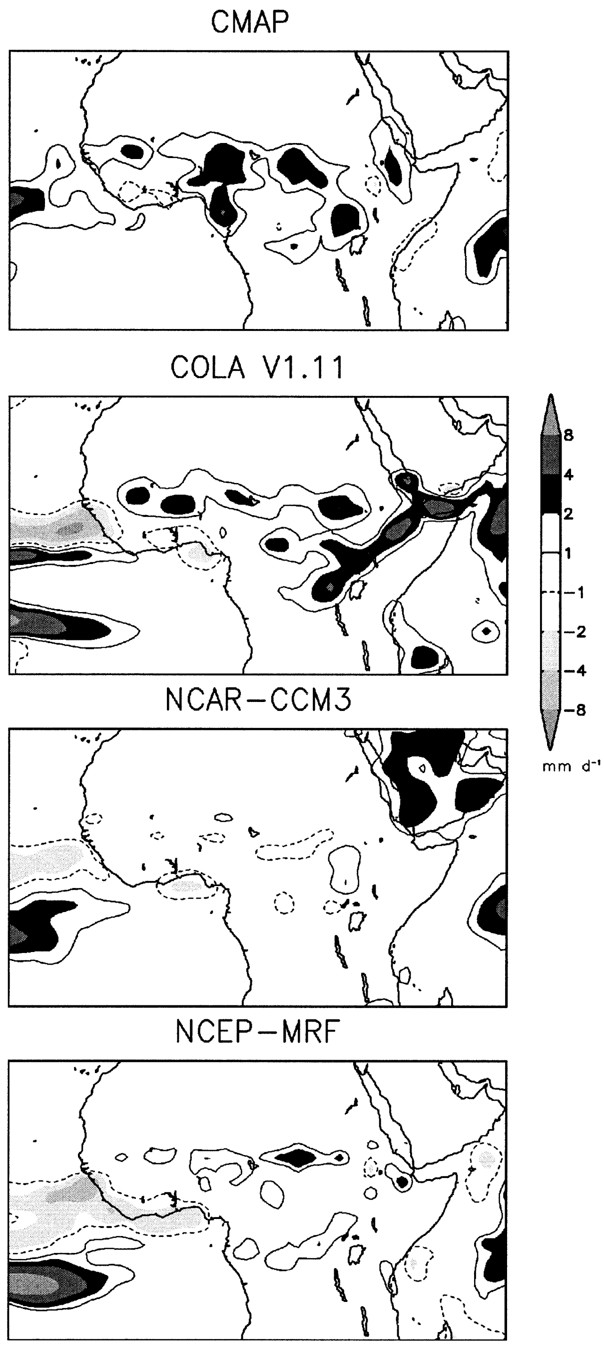

FIG. 6. As in Fig. 5, but over Africa.

broad erroneous precipitation event over Arabia. Again, the areas where skill over land seems to be most apparent are at low latitudes. Yet over the Atlantic the three models produces a dipole pattern much more similar to each other than to the observed pattern in the CMAP dataset.

It should be pointed out that, while using an ensemble mean to predict climate improves a model's ability to capture anomaly patterns, by reducing the contribution of model noise in the solution, the capacity to reproduce extreme anomalies is strongly curtailed. Extreme climate anomalies in the real world may not be purely a function of the boundary conditions, but may be the result of a favorable superposition of boundary-forced response and internal atmospheric variability. The ensemble averaging that serves to enhance the boundary forced signal also suppresses the internal noise. The year-to-year variance of a single realization of a nonlinear system will be reduced systematically as larger and larger ensembles are considered, so the ensemble will not have the same temporal variance statistics as a single integration. This may explain the propensity of the models to underforecast the relative strength of anomalies.

5. Assessment of potential predictability

In order to assess the potential predictability of the ensemble hindcasts in capturing the prominent year-to-year climate anomalies we examine an additional characteristic of the simulations. An assessment of the signal-to-noise ratio (SNR), calculated from the variance statistics of each model for key fields, reveals when and where a model may exhibit potential predictability. However, predictable behavior in a climate model is not of much use if the model does not exhibit skill in simulating the interannual variability at that location. We use temporal ACC as a measure of model skill. If high ACC is collocated with a high SNR for a given variable, we may presume that the model possesses some useful forecast ability at that location. Low SNR does not preclude useful forecast ability, but it may lower our confidence when the model is used in an actual forecast mode.

As described in Shukla et al. (2000), the noise for a given variable is calculated as the intraensemble variance for each season, averaged over the 5 yr. The signal is the variance of the ensemble means from the grand mean, with account taken for the noise component since the ensembles are of finite size. Figure 7 shows the noise and signal separately for the June–September average precipitation for each model. On the panels showing signal, the region where signal exceeds noise is outlined with a bold black contour. Several features are evident among the models. First, the signal in each model is largely confined to parts of the tropical oceans, with little signal in the midlatitudes of either hemisphere or over land. Second, regions of high noise cover a much larger area than the regions of high signal for every model. The pattern of noise is quite similar to the overall pattern of precipitation variance, or the mean precipitation for that matter. Third, each of these models shows a distinct combination of signal and noise consistent with the findings for winter described by Shukla et al. (2000). That is, the NCAR model shows noticeably less intraensemble noise than the other two models, while the COLA GCM shows the strongest signal. None of

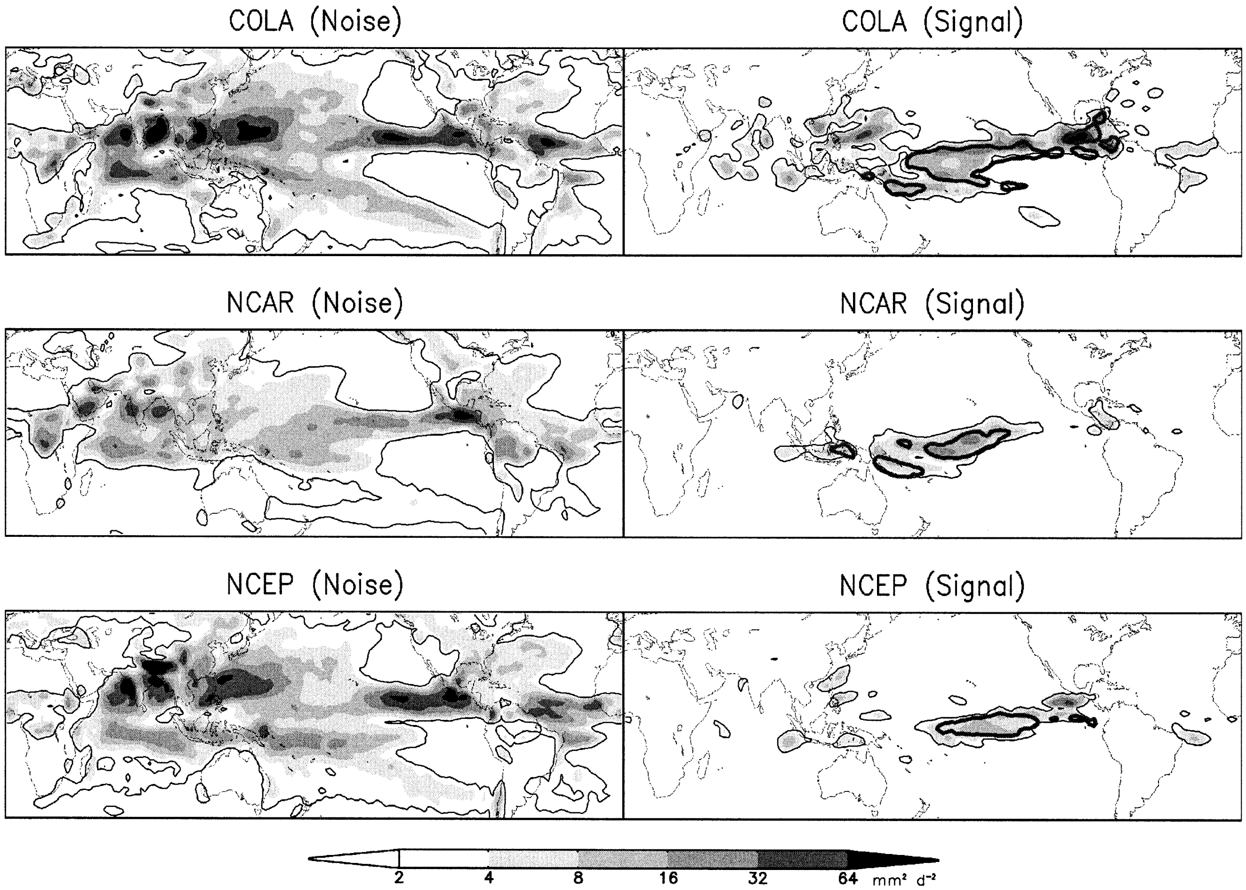

FIG. 7. (left) Noise and (right) signal in mean Jun–Sep precipitation as simulated by the three GCMs, following the definitions of Shukla et al. (2000). In the right-hand panels, areas where signal exceeds noise are outlined by bold contours.

the models show much SNR greater than unity outside the deep Tropics of the Pacific Ocean. Only the COLA model shows signal over the Asian monsoon region, yet none of the models show predictability.

6. Conclusions

An analysis of a set of summer DSP ensemble simulations by three different climate models has been presented. Examination of the interannual and intraensemble statistics verifies the differences between models described for the winter DSP simulations. However, unlike winter, we find that there is very little evidence of useful predictable skill outside the deep Tropics during the 5 yr simulated.

Why are simulations for boreal summer so much poorer than those shown for winter by Shukla et al. (2000)? There are several possibilities. One is that there simply is no exploitable predictability during the summer season. Another is that there is indeed predictability, but the ensemble sizes used here (less than 10 members) is too small to extract useful signal. Finally, it may be that these models, as applied here, are not up to the task.

Certainly there are deficiencies in how these summer DSP integrations were initialized over land. If the general circulation alone determines local anomalies, and SST determines the general circulation, then there is little hope for enhancing prediction during boreal summer by improved land surface representation. If there is warm-season predictability to be harvested from the land surface state, then it may not have been captured by this experiment.

These DSP simulations do not take full advantage of the potential predictability that land surface boundary conditions may contribute. As described in section 2, the COLA GCM initializes soil wetness from values derived from ECMWF analyses, and the NCEP model uses reanalysis values that are heavily damped to climatology. There is no interannual signal in initial soil temperature or snow cover, and none of these models have the capability to represent the interannual variability of vegetation cover or color, aerosols, etc.

Strong systematic errors in precipitation and radiation may quickly overwhelm any useful information contained in the land surface initial conditions. Improvements in the physical parameterizations of the atmospheric models, to reduce systematic errors in precipitation and radiation, may also lead to improved skill.

Also, if the land surface state were realistically prescribed throughout the season, as was SST, there might have been an alteration of surface fluxes that could further improve the simulation of climate variations. The potential impact of complete, realistic land surface initial conditions and boundary conditions also needs to be examined before a definitive statement about the potential predictability of warm season climate can be made.

Acknowledgments. The authors wish to recognize J. Shukla, E. Kalnay, and D. Baumhefner, who provided these models for summer dynamical seasonal prediction experiments. This work was supported by NSF Grant ATM 9814265, NOAA Grant NA96GP0056, and NASA Grant NAG5-8202.

REFERENCES

Charney, J. G., and J. Shukla, 1981: Predictability of monsoons. *Monsoon Dynamics: Proceedings of the Symposium on Monsoon Dynamics*, J. Lighthill and R. P. Pearce Eds., Cambridge University Press, 99–109.

Dirmeyer, P. A., 2000: Using a global soil wetness dataset to improve seasonal climate simulation. *J. Climate,* **13,** 2900–2922.

——, M. J. Fennessy, and L. Marx, 2001: Near surface boreal summer climate as simulated by three general circulation models. COLA Tech. Rep. 100, 36 pp. [Available from the Center for Ocean–Land–Atmosphere Studies, 4041 Powder Mill Road, Suite 302, Calverton, MD 20705; or online at http://www.iges.org/pubs/tech.html.]

Doblas-Reyes, F. J., M. Déqué, and J.-P. Piedelievre, 2000: Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Quart. J. Roy. Meteor. Soc.,* **126,** 2035–2067.

Fennessy, M. J., and J. Shukla, 1999: Impact of initial soil wetness on seasonal atmospheric prediction. *J. Climate,* **12,** 3167–3180.

——, J. L. Kinter III, L. Marx, E. K. Schneider, P. J. Sellers, and J. Shukla, 1994: GCM simulations of the life cycles of the 1988 US drought and heat wave. COLA Rep. 6, Center for Ocean–Land–Atmosphere Studies, 68 pp.

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

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kiehl, J. T., J. J. Hack, G. B. Bonan, B. A. Boville, D. L. Williamson, and P. J. Rasch, 1998: The National Center for Atmospheric Research Community Climate Model: CCM3. *J. Climate,* **11,** 1131–1149.

Kinter, J. L., and Coauthors, 1997: The COLA atmosphere–biosphere general circulation model. Volume 1: Formulation. COLA Rep. 51, Center for Ocean–Land–Atmosphere Studies, 46 pp.

Kirtman, B. P., J. Shukla, B. Huang, Z. Zhu, and E. K. Schneider, 1997: Multiseasonal predictions with a coupled tropical ocean–global atmosphere system. *Mon. Wea. Rev.,* **125,** 789–808.

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate,* **13,** 4196–4216.

Mo, K. C., and J. N. Paegle, 2000: Influence of sea surface temperature anomalies on the precipitation regimes over the southwest United States. *J. Climate,* **13,** 3588–3598.

Palmer, T. N., and C. Brankovic, 1989: The 1988 US drought linked to anomalous sea surface temperature. *Nature,* **338,** 54–57.

Rasmusson, E. M., and J. M. Wallace, 1983: Meteorological aspects of the El Niño/Southern Oscillation. *Science,* **222,** 1195–1202.

Reale, O., P. A. Dirmeyer, and A. Schlosser, 2002: Modeling the effect of land surface evaporation variability on precipitation variability. Part II: Time- and space-scale structure. *J. Hydrometeor.* **3,** 451–466.

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

Ropelewski, C. F., J. E. Janowiak, and M. F. Halpert, 1985: The analysis and display of real-time surface climate data. *Mon. Wea. Rev.,* **113,** 1101–1107.

Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science,* **282,** 728–731.

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

——, and Coauthors, 2000: Dynamical seasonal prediction. *Bull. Amer. Meteor. Soc.,* **81,** 2593–2606.

Trenberth, K. E., G. W. Branstator, and P. A. Arkin, 1988: Origins of the 1988 North American drought. *Science,* **242,** 1640–1645.

Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.,* **78,** 2539–2558.