

Realistic Initialization of Land Surface States: Impacts on Subseasonal Forecast Skill

RANDAL D. KOSTER,* MAX J. SUAREZ,* PING LIU,*⁺ URSZULA JAMBOR,^{#,@} AARON BERG,&
MICHAEL KISTLER,*⁺ ROLF REICHLER,*[#] MATTHEW RODELL,[@] AND JAY FAMIGLIETTI**

*Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

⁺Science Applications International Corporation, Beltsville, Maryland

[#]Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore, Baltimore, Maryland

[@]Hydrological Sciences Branch, Laboratory for Hydrospheric Physics, NASA Goddard Space Flight Center, Greenbelt, Maryland

&Department of Geography, University of Guelph, Guelph, Ontario, Canada

**Earth System Science, University of California, Irvine, Irvine, California

(Manuscript received 6 February 2004, in final form 21 May 2004)

ABSTRACT

Forcing a land surface model (LSM) offline with realistic global fields of precipitation, radiation, and near-surface meteorology produces realistic fields (within the context of the LSM) of soil moisture, temperature, and other land surface states. These fields can be used as initial conditions for precipitation and temperature forecasts with an atmospheric general circulation model (AGCM). Their usefulness is tested in this regard by performing retrospective 1-month forecasts (for May through September, 1979–93) with the NASA Global Modeling and Assimilation Office (GMAO) seasonal prediction system. The 75 separate forecasts provide an adequate statistical basis for quantifying improvements in forecast skill associated with land initialization.

Evaluation of skill is focused on the Great Plains of North America, a region with both a reliable land initialization and an ability of soil moisture conditions to overwhelm atmospheric chaos in the evolution of the meteorological fields. The land initialization does cause a small but statistically significant improvement in precipitation and air temperature forecasts in this region. For precipitation, the increases in forecast skill appear strongest in May through July, whereas for air temperature, they are largest in August and September. The joint initialization of land and atmospheric variables is considered in a supplemental series of ensemble monthly forecasts. Potential predictability from atmospheric initialization dominates over that from land initialization during the first 2 weeks of the forecast, whereas during the final 2 weeks, the relative contributions from the two sources are of the same order. Both land and atmospheric initialization contribute independently to the actual skill of the monthly temperature forecast, with the greatest skill derived from the initialization of both. Land initialization appears to contribute the most to monthly precipitation forecast skill.

1. Introduction

Numerical weather forecasts rely on atmospheric initialization—the accurate specification of atmospheric pressures, temperatures, winds, and humidities at the beginning of the forecast. Such initialization may contribute to forecast skill at leads of up to 10 days. Forecasts at longer leads, however, require a different strategy. They must take advantage of slower modes of the climate system, modes with states that are not so quickly dissipated by chaos. To this end, operational centers now supply seasonal atmospheric forecasts based largely on forecasts of ocean behavior. The idea is simple—if sea surface temperatures (SSTs) can be predicted months in advance, and if the atmosphere responds in predictable ways to the predicted SSTs, then aspects of the atmosphere's behavior can be predicted months in advance.

Soil moisture, another slow variable of the climate system, is beginning to garner attention in the forecast community (e.g., Dirmeyer et al. 2003). The time scales of soil moisture memory are typically 1 or 2 months (Vinnikov and Yeserkepova 1991; Entin et al. 2000), significantly less than those of the ocean. Nevertheless, soil moisture has a special importance. Some atmospheric general circulation model (AGCM) studies (Kumar and Hoerling 1995; Trenberth et al. 1998; Shukla 1998; Koster et al. 2000) note a strong tropical–extratropical contrast in the ocean's impact on climate. This impact in midlatitudes, where much of the world's population lives, may be significantly modulated by land surface processes, particularly in summer. Soil moisture may play a key role in these areas (Koster et al. 2000).

AGCMs are indeed useful tools for examining the role of soil moisture in the climate system. [See Koster and Suarez (2003) for a literature review of a number of relevant studies.] Of particular relevance to forecasting are those studies that evaluate the improvement of forecast skill associated with the correct initialization

Corresponding author address: Dr. Randal D. Koster, Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Mailstop 974.0, Greenbelt, MD 20771.
E-mail: randal.koster@gsfc.nasa.gov

of soil moisture. Beljaars et al. (1996), Fennessy and Shukla (1999), Douville and Chauvin (2000), and Viterbo and Betts (1999) used reasonably realistic soil moistures (e.g., from reanalyses or offline prescribed forcing analyses) to initialize AGCM simulations, and all found some cause for encouragement—suggestions that the initialization does improve the simulation of precipitation and/or temperature.

For the forecast experiments of Koster and Suarez (2003), a preliminary simulation was performed in which AGCM-generated precipitation rates were replaced with observed rates prior to applying the precipitation to the land surface. The land surface in this simulation thus evolved soil moistures that reflected observed antecedent precipitation. These soil moistures were then used as initial conditions for separate forecast simulations. The study's main contribution was an illustration of how three factors—the size of typical soil moisture anomalies, the sensitivity of evaporation to soil moisture, and the sensitivity of precipitation to evaporation—work together to determine the impact of soil moisture initialization on the forecast. In addition, evaluations of forecasted precipitation and temperature against observations suggested some improvement associated with land initialization. Koster and Suarez (2003), however, argued that the improvement was small enough to require a much larger number of independent forecast periods for its proper quantification—the 5 yr analyzed in the study were insufficient for useful statistics. A similar limitation applies to the other, aforementioned soil moisture initialization studies.

The present study can be considered a substantial broadening of the Koster and Suarez (2003) study. Perhaps most important, we examine here the impact of soil moisture initialization on 1-month forecasts for each of five Northern Hemisphere warm-season months in each of 15 yr, spanning 1979–93. Thus, a total of 75 independent forecasts are evaluated against observations, enough to generate—for the first time—reasonable statistics for the small inherent signal. This study also features several improvements in initialization technique and analysis. For example, we use a more complete set of antecedent forcing data to initialize soil moisture; rather than focusing on observed antecedent precipitation alone, observed antecedent radiation and near-surface air properties from reanalysis are also employed (section 2b). Land initialization effects are examined side by side with atmospheric initialization effects, to demonstrate the relative contribution of land initialization to total skill (section 3d). Impacts of temporal scale (first half versus second half of month; see section 3d) are considered explicitly. Furthermore, the skill levels produced are examined relative to “idealized predictability”—the maximum forecast skill possible in the system—with a more robust diagnostic (section 3a) than that used by Koster and Suarez (2003).

As a result of these and other improvements, we establish in this study much firmer conclusions regarding

the impact of realistic soil moisture initialization on forecast skill.

2. Design of experiment

a. Modeling system

The forecast experiments make use of the seasonal prediction system of the National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO), which is the same as the NASA Seasonal-to-Interannual Prediction Project (NSIPP) system referred to in our earlier studies (e.g., Koster and Suarez 2003; Koster et al. 2000). The AGCM is a state-of-the-art finite-difference model run at a resolution of 2° latitude \times 2.5° longitude. It uses the relaxed Arakawa–Schubert scheme (Moorthi and Suarez 1992) for convection, sophisticated codes for shortwave and longwave radiation (Chou and Suarez 1994), and fourth-order advection of vorticity and all scalars in the modeled dynamics. The land surface model (LSM) is the Mosaic LSM of Koster and Suarez (1996), a soil–vegetation–atmosphere transfer (SVAT) model that uses tiling to account for subgrid vegetation distributions. The behavior of the coupled land–atmosphere system relative to observations is well documented (Becmeister et al. 2000; Koster et al. 2000); the coupled model, while not perfect, successfully reproduces the broad features of precipitation means and variances across the globe.

b. Initialization procedure

In situ soil moisture observations do exist (Robock et al. 2000), but they have a limited spatial distribution and are completely absent in most parts of the world. Satellite-derived values (e.g., Owe et al. 2001) are available across the globe for sparsely vegetated areas but are limited temporally and represent moisture in only the top few millimeters of soil. Thus, for global distributions of soil moisture in the root zone and below, an alternative, indirect approach is required. One promising approach is based on the utilization of antecedent meteorological forcing. Rainfall, for example, is well measured globally, at least relative to soil moisture. If rainfall for the month prior to the forecast start date is known to be anomalously high, the local soil moisture on the forecast start date should be higher than average as well.

A complete and accurate global dataset of observed meteorological forcing used in conjunction with global soil and vegetation properties should, in principle, contain all of the information needed to determine the global field of soil moisture anomalies on any given date, if the data are processed with a good land surface model. This is the basis of the approach used in this paper. Although a perfect observational dataset does not exist, a recent dataset developed by Berg et al. (2003) is a good approximation. This global dataset provides meteorological forcing every 6 h over the period 1979–

93. Values for wind speed, surface pressure, and near-surface air temperature and humidity are extracted from the European Centre for Medium-Range Weather Forecasts (ECMWF) 15-yr Reanalysis (ERA-15; Gibson et al. 1997), as are the submonthly breakdowns of the precipitation and radiation. Surface pressure and air temperature values are elevation corrected. Monthly precipitation amounts agree with values provided by the Global Precipitation Climatology Project (GPCP, version 2; see Adler et al. 2003), and monthly incoming solar and longwave radiation amounts match those estimated by the surface radiation budget (SRB) project (Gupta et al. 1999). (The radiation data were available between July 1983 and June 1991; outside this period, the reanalysis-derived radiation values were scaled to match SRB climatology.) Monthly averaged air temperatures were forced to agree with a merged observational dataset constructed from those of New et al. (2000) and Willmott and Matsuura (2001). Vapor pressures were scaled to those of New et al. (2000).

These data are used to force a 15-yr offline simulation of the Mosaic LSM using the Global Land Data Assimilation System (GLDAS). GLDAS was developed through a collaboration of scientists at the NASA Goddard Space Flight Center and National Oceanic and Atmospheric Administration (NOAA)/National Centers for Environmental Prediction (NCEP), its goal is to produce global, reliable fields of land surface states and fluxes by parameterizing, forcing, and constraining multiple, sophisticated LSMs with data from advanced observing systems (Rodell et al. 2004). For this particular offline simulation, GLDAS vegetation, soil, and elevation parameters were set to match those of the seasonal prediction model. GLDAS/Mosaic was “spun up” by looping over the 1979 forcing 10 times prior to driving the LSM for the full 15-yr period. Output from the beginning of each warm-season month provide the initial conditions used in this study. Recent analyses (Reichle et al. 2004) show that a subset of these model-generated initial conditions, the near-surface soil moisture fields, are reasonably consistent with soil moisture fields inferred from Scanning Multichannel Microwave Radiometer (SMMR) satellite measurements and with in situ observations from the Global Soil Moisture Data Bank (Robock et al. 2000), given the limitations of each dataset.

This offline forcing approach has an important advantage. Koster and Milly (1997) illustrate the model-dependent nature of simulated soil moisture and the danger of blindly inserting soil moisture from one LSM—or even from observations—into another LSM. Because the GLDAS forcing is applied during the initialization procedure to the Mosaic LSM, and because that LSM is also used in the forecast system, this danger is largely avoided.

Despite this guaranteed consistency, however, some adjustment of the initialized fields is still necessary because of climate biases in the forecast system relative to

observations. As explained in Koster and Suarez (2003), use of unmodified fields could lead to suboptimal forecasts—the unmodified fields would lead to a transitional climate “drift” during the forecast period, a drift that could muddle the interpretation of the forecast. To avoid this drift, at least to first order, standard normal deviates are used. Let $\overline{X_{\text{obs}}}$ be the average value of a state variable X (say, soil moisture in the root zone) on a given forecast start date, as determined from GLDAS. Because GLDAS produces data for 15 yr, this mean will be based on 15 values. Similarly, let σ_{obs} be the standard deviation of X on that date, and let $\overline{X_{\text{mod}}}$ and σ_{mod} be the corresponding AGCM statistics. If X_{obs} is the value of the state variable for year n as provided by GLDAS, then the value X_{mod} used to initialize the coupled land–atmosphere system is the one that satisfies

$$\frac{X_{\text{mod}} - \overline{X_{\text{mod}}}}{\sigma_{\text{mod}}} = \frac{X_{\text{obs}} - \overline{X_{\text{obs}}}}{\sigma_{\text{obs}}}. \quad (1)$$

Using (1) ensures, for example, that a relatively wet state from GLDAS translates to a correspondingly wet state in the coupled model. If the coupled system models the statistics of climate perfectly at a given grid cell, then the GLDAS value is effectively used there without modification.

We are implicitly assuming here that the model-generated fields are accurate, to first order. In the future, we expect to increase their accuracy through data assimilation techniques—through the direct combination of satellite-derived soil moisture products, limited as they are, with the model products (e.g., Walker and Houser 2001; Reichle and Koster 2003).

c. Ensembles performed

A total of 75 one-month ensemble forecasts were performed: for each year during 1979–93, we initialized the land surface on the first day of May, June, July, August, and September and integrated the model for one month following each initialization. Each of the 75 ensemble forecasts, which are examined with zero lead, contained nine independent members.

All members of a given ensemble used the same land surface initial conditions, namely, the land states produced by GLDAS, scaled with (1). The members of the GLDAS ensemble differed only in their initial atmospheric conditions, which were taken from nine parallel “Atmospheric Model Intercomparison Project (AMIP) style” simulations, that is, simulations in which the SSTs are prescribed to observed values. Although the nine sets of initial atmospheric conditions in an ensemble are consistent with the SSTs on the forecast start date, they do not necessarily resemble each other or the observed atmospheric conditions on that date; in essence, each set of atmospheric initial conditions represents one realization of what nature could have produced on that start date given chaotic atmospheric dynamics. This is a critical aspect of our simulations. In

our main set of forecasts, we do not make use of the source of skill used by numerical weather prediction (NWP) systems; any skill found reflects only land initialization and SST specification. Supplemental simulations, described in section 3d, examine how the accurate initialization of both the atmosphere and the land affect forecast skill.

To isolate the impact of the land initial conditions from that of SST specification, the above forecasts are compared to corresponding ensemble “forecasts” that do not make use of a specific land surface initialization. These otherwise identical forecasts are simply the appropriate subsets of the nine parallel AMIP-style simulations. By design, the initial atmospheric conditions are equivalent to those used in the first set of ensembles. The initial land surface states for a given forecast in the second set of ensembles are fully consistent with the initial atmospheric states for that forecast, since they are derived from the same long-term AMIP-style simulation. The land initial conditions, however, naturally vary between the forecasts, since the land states in parallel AMIP-style simulations are different. In effect, the land initial conditions for this second set of ensemble forecasts are chosen randomly from the broad distribution of states that are consistent with the concurrent SSTs.

We hereafter refer to the first set of ensembles—the ones using accurate land initialization—as the GLDAS forecasts. The second “control” set of ensembles is referred to as the AMIP forecasts. Note that the use of the term “forecasts” is not precisely correct, since we are prescribing realistic SSTs throughout the 1-month forecast period. In both cases, we are assuming that SSTs can be perfectly predicted for 1 month. A comparison of the GLDAS and AMIP ensembles will nevertheless isolate the impact of the land initialization on forecast skill.

d. Validation data

The temperature data used to evaluate the 1-month temperature forecasts are extracted from the dataset of Berg et al. (2003) for the month in question. The data source for the validation is thus the same as the data source used in the initialization. Of course, the initialization and validation data come from two nonoverlapping segments of this single data source.

For precipitation, the data used across much of the globe is similarly extracted from the Berg et al. (2003) dataset. Over the United States, however, we replace these data with the dataset of Higgins et al. (2000), which is based on a much more complete rain gauge database than was utilized by GPCP. As will be seen below, validation in this area will be key to addressing land initialization impacts.

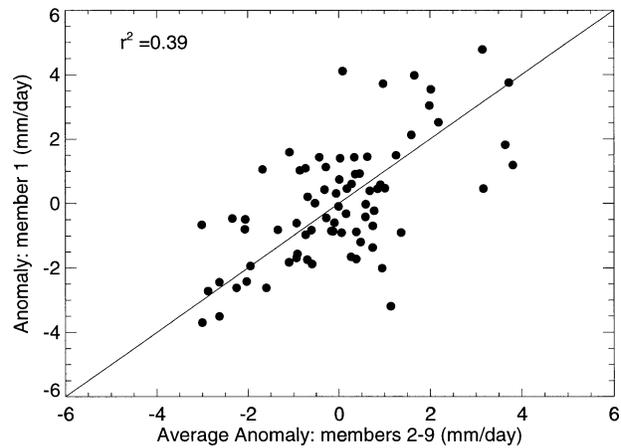


FIG. 1. Scatterplot for the idealized analysis, showing the degree to which the model can “predict itself” at a central U.S. grid cell (40°N , 97.5°W). The x axis represents the forecasted precipitation anomaly averaged over eight members of the GLDAS ensemble, and the y axis represents the precipitation anomaly generated by the ninth member (the “observations” for this idealized analysis). Seventy-five points are plotted, one for each of the five analyzed months in the years 1979–93. The solid line is the 1:1 line.

3. Results

a. Idealized analysis: Maximum skill possible with the modeling system

Before proceeding with a full, global evaluation of forecast skill, we pause to consider the maximum skill possible in the modeling system—the upper bound to what we can hope to achieve. Perfect predictability with a seasonal forecast system is precluded by atmospheric chaos. The vagaries of chaos allow nature to take different evolutionary paths from initial atmospheric conditions that differ only slightly from each other, within measurement error. In effect, nature provides only one realization of seasonal weather from a potentially broad probability density function (PDF). The hope in seasonal forecasting is that this PDF can be reproduced accurately and can be narrowed significantly with the specification of slowly evolving boundary conditions in the ocean and land.

To quantify the maximum possible skill in the system, skill that is not limited by errors in initial conditions, boundary conditions, or validation data, we perform the following idealized analysis. First, we assume that the first member of the GLDAS forecast ensemble represents “nature” and that the remaining eight members of the ensemble represent the actual model forecast. Then, at each grid cell, we perform a scatter analysis, as illustrated in Fig. 1. Each point in the figure represents one of the 15 Mays, Junes, Julys, Augusts, or Septembers analyzed in our experiments. The x coordinate is the “forecasted” precipitation anomaly for the month at a specific grid cell (in the central United States: 40°N , 97.5°W), computed by averaging the precipitation generated by the noted eight ensemble members and then

subtracting from this average the multiyear model mean for that month. The y coordinate is the corresponding “observed” anomaly, that is, the anomaly from the ensemble member representing nature. The square of the correlation coefficient (r^2) is computed through linear regression.

This r^2 value quantifies the ability of the model to predict an assumed “nature,” given that the model and this nature use exactly the same internal physics, the same surface boundary conditions, and the same initial conditions for land. Clearly, if this idealized prediction skill were perfect, the 75 points would lie along the 1:1 line, and r^2 would be exactly 1. If, on the other hand, prediction skill was completely absent, the points would be scattered randomly, and r^2 would be close to zero. In all of our analyses, if the line fitted through the points has negative slope (indicating “negative skill”), r^2 is automatically set to 0. This zeroing is employed to reduce noise in our later evaluations of model forecasts against observations; for the idealized analysis here, the zeroing has almost no impact. In the plotted example, $r^2 = 0.39$, which we interpret as an idealized skill level of 39%. (The equivalent plot constructed from the AMIP ensemble shows much more scatter, with a negligible r^2 .)

Once the global array of r^2 values is established through this procedure, the second ensemble member is chosen to represent “nature,” and the process is repeated. The process is performed a total of nine times, once for each ensemble member, and the resulting nine r^2 arrays are averaged.

The top panel of Fig. 2 shows this average idealized r^2 value for precipitation, as generated from the GLDAS ensemble. These values indicate the maximum potential predictability associated with land initialization combined with SST specification. The middle panel of the figure shows the corresponding r^2 values for the AMIP ensemble, which indicate the maximum potential predictability associated with SST specification alone. The differences between these r^2 values are shown in the bottom panel. These differences reflect the gain in potential predictability associated with land initialization.

Clearly, in this modeling system, land initialization can contribute to predictability over only a few key areas: the central United States, equatorial South America, equatorial Africa, parts of central Asia, and the land skirting the Bay of Bengal. These regions agree, in essence, with those identified by Koster and Suarez (2003) with an alternative, and less robust, statistic. Outside these regions, the chaotic dynamics of the atmosphere overwhelm any control imposed by anomalies at the land surface. Outside these regions, a positive impact of land initialization on precipitation forecast skill cannot be expected. [We caution here that the indicated regions of maximum predictability may be model dependent; indeed, a strong intermodel disparity exists in the calculation of land impacts on atmospheric processes

(Koster et al. 2002). See also the discussion of Fig. 11 in section 4.]

Figure 3 provides the same information for the idealized air temperature forecasts. The inherent predictability of temperature associated with land initialization greatly exceeds that of precipitation throughout most of the globe, in terms of both magnitude and the areal extent of impact. This is not a surprise if one assumes a stronger physical connection between soil moisture and air temperature (through the former’s impact on evaporative cooling) than between soil moisture and precipitation. The areas of impact include some Southern Hemisphere regions. In some areas, however, such as northern Asia and much of either coast of North America, chaotic atmospheric dynamics still prevent any prediction at all.

b. Assessing skill: Main area of focus

Two strong constraints limit our analysis of the impacts of soil moisture initialization on forecast skill: 1) soil moisture initialization must have a statistically significant impact on the forecast, and 2) the applied initial soil moisture must be of acceptable accuracy. The first requirement was addressed in section 3a. In our assessments of skill in this modeling system, we need not look outside the shaded areas in the bottom panels of Figs. 2 and 3. Furthermore, we note again that an idealized r^2 increase associated with land initialization in either figure is indeed an upper bound for the actual r^2 increase for the forecasts with this system—an upper bound that will be difficult to attain, given unavoidable errors in model initialization and parameterized physics. The actual r^2 increase must be accommodated between this upper bound and a value of, say, 0.035, which is significantly different from zero at the 90% confidence level. In this paper, to increase the potential for discernable forecast skill, we focus our analyses on areas for which the idealized r^2 increase in the bottom panel of Figs. 2 or 3 exceeds 0.1. These areas are outlined with black lines in the panels. (Alternative choices for the critical value do not significantly affect the results.)

The second constraint is now addressed. While definitive estimates of soil moisture accuracy are impossible given the paucity of in situ data, an analysis of the factors that determine soil moisture does provide guidance. The initialization system relies on the specification of realistic soil type, vegetation type, and various forcing data: radiation, precipitation, and near-surface meteorological quantities such as specific humidity and temperature. Errors in the specification of any of these quantities could lead to errors in land surface initialization.

For this study we focus on the precipitation forcing, making the assumption that precipitation is the key driver of soil moisture anomalies. Clearly, if the precipitation forcing is poor, soil moisture values cannot be trusted. Oki et al. (1999) demonstrated that a minimum

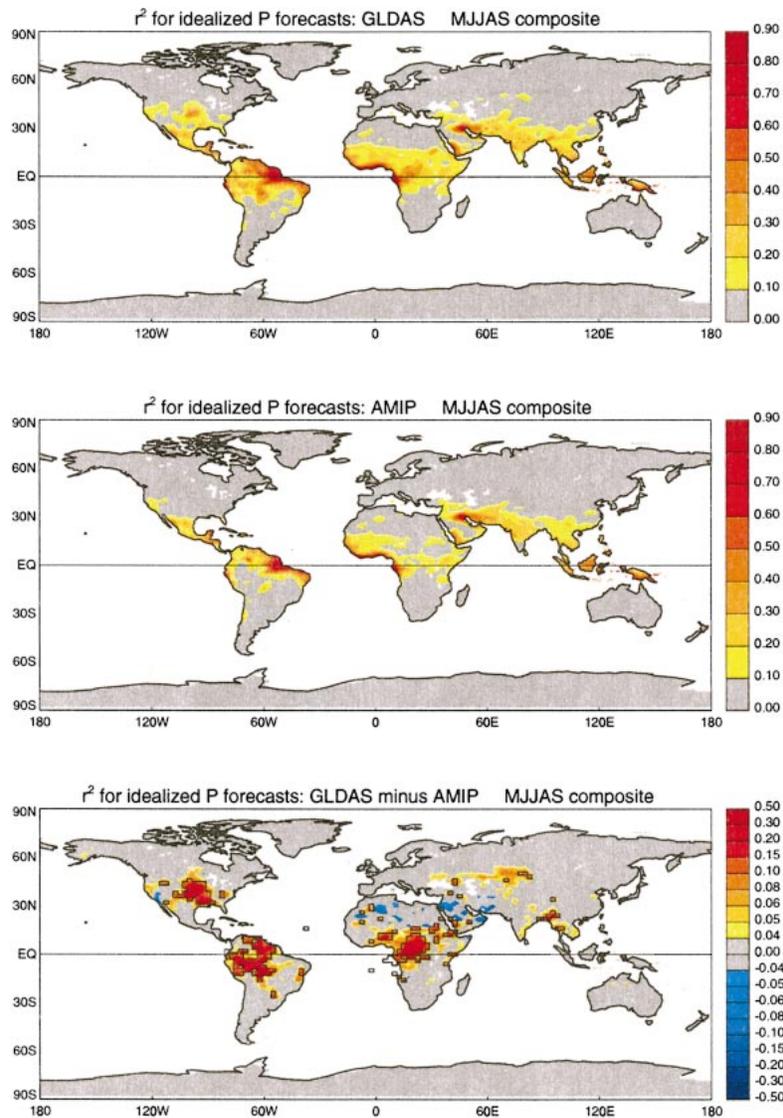


FIG. 2. Idealized analysis that quantifies the upper limit of predictability in the model. (top) Averaged r^2 values for the idealized precipitation forecasts generated by the GLDAS ensemble. (middle) Same, but for the AMIP ensemble. (bottom) Differences. Values of 0.035, 0.05, and 0.08 in the bottom plot are significant at the 90%, 95%, and 99% levels, respectively. The black lines outline the regions for which the differences exceed 0.1, a requirement we impose for our later analyses.

of about 30 precipitation gauges per one million square kilometers, or about 2 gauges per $2.5^\circ \times 2.5^\circ$ GPCP grid cell, are required for accurate streamflow simulation. Because this density would severely limit the areas over which we could evaluate forecast skill, we employ an arbitrarily lower (but still nonzero) critical level for our analysis. We assume here that a rain gauge density of 0.5 gauges per $2.5^\circ \times 2.5^\circ$ GPCP grid cell is required for a reasonably accurate initialization of soil moisture. (The exact value chosen turns out to have little impact on the results.)

Figure 4 shows the density of precipitation gauges used by GPCP to generate monthly precipitation totals.

(Again, monthly GPCP totals underlie the precipitation forcing used to drive the land model in the initialization phase.) In fact, the plot shows the minimum density over the 15 yr of analysis; this was determined by first finding the average density in each GPCP cell for each year (all 12 months) and then identifying, for each cell, the year with the lowest value. Tremendous changes in the yearly coverage of gauges utilized by GPCP—associated with a 1986 switch in gauge network—necessitate the consideration here of minimum density rather than 15-yr average density. As seen from the figure, many parts of the world have unacceptable densities. Furthermore, as seen from a comparison of Fig.

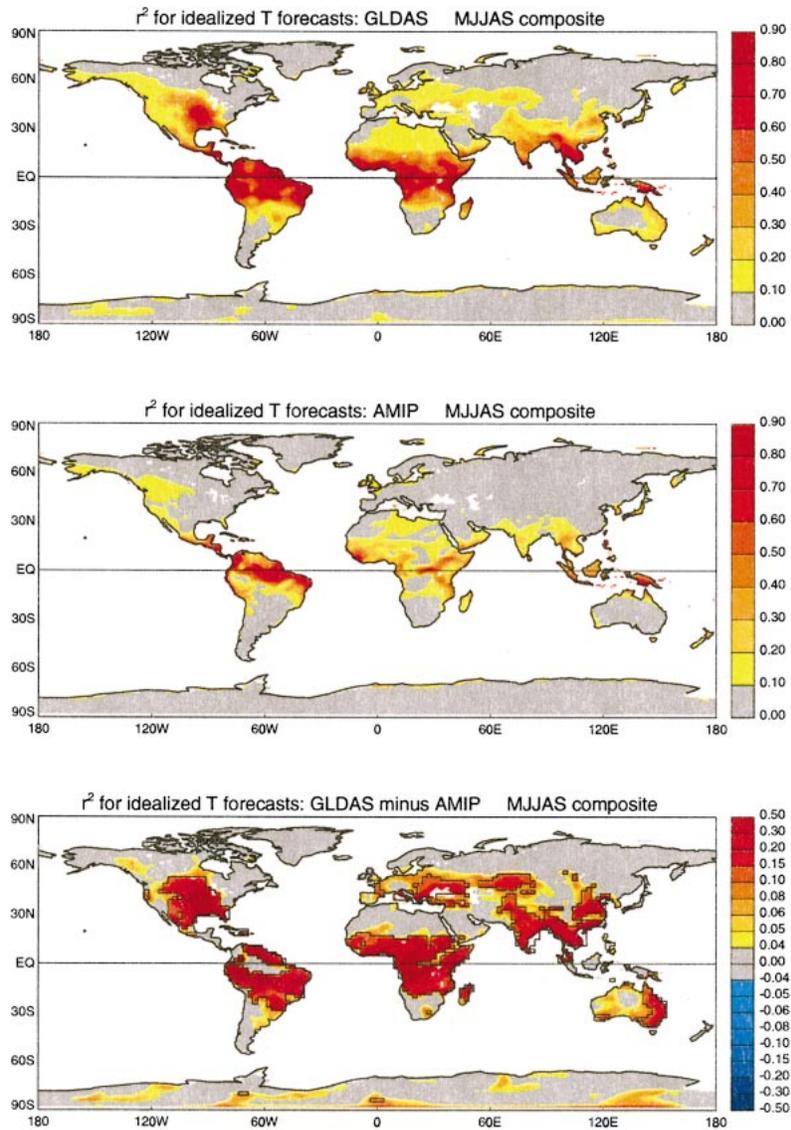


FIG. 3. Same as Fig. 2, but for air temperature.

4 with the bottom panel of Fig. 2, only the central United States has both a significant soil moisture impact on precipitation and an accurate soil moisture initialization, as inferred from gauge density. We will therefore limit to this one region our assessment of soil moisture initialization's impact on precipitation forecast skill. For monthly temperature forecasts, a comparison of Figs. 3 and 4 shows that several isolated regions across the globe satisfy our criteria for forecast assessment.

Note that in applying the rain gauge density criterion, we are implicitly assuming that most soil moisture feedback effects are local. The idealized predictability criterion, on the other hand, carries no such assumption; whether soil moisture effects are local or remote, we

cannot expect, in this modeling system, to see an improvement in skill outside the areas outlined in Figs. 2 and 3. The assumption of purely local impacts will be relaxed in future, more detailed analyses. We note now, however, that for precipitation prediction, the local impacts assumption essentially eliminates from further analysis the regions of high idealized predictability in equatorial South America and Africa. This is not a major problem, since the rain gauge densities in the surrounding areas—in the potential remote controlling areas for these two regions—also tend to be much too small. For temperature prediction, the local impacts assumption is more defensible, since soil moisture has a first-order impact on local temperature through its effect on evaporative cooling.

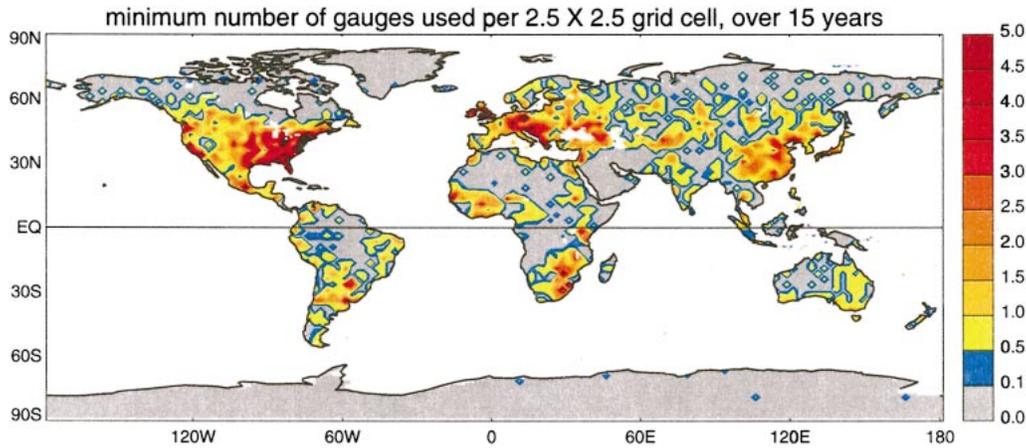


FIG. 4. Density of rain gauges (number of gauges per $2.5^{\circ} \times 2.5^{\circ}$ grid cell) used to generate the GPCP monthly product in the year with the fewest rain gauges at the grid cell, for the period 1979–93.

c. Skill assessment: One-month forecasts

In our assessment of forecast skill, we construct scatterplots (not shown) similar to that in Fig. 1, with the x axis now representing the simulated precipitation anomaly averaged over all nine ensemble members (i.e., the full forecast), and the y axis representing the observed anomaly, relative to the observed mean. We take the resulting r^2 values, one value computed at each grid cell using 75 forecast/observed pairs, as our measure of forecast skill. Note that the experimental design makes the use of certain other skill measures, such as the root-mean-square error (rmse), problematic. Because the experimental forecast is an average of nine ensemble members, the year-to-year variance of the forecast is necessarily smaller than that of the observations, which represents a single “realization” of what nature might produce. This unavoidable variance reduction would inappropriately magnify an rmse diagnostic but does not affect the r^2 diagnostic.

The top-left panel of Fig. 5 shows, for the GLDAS forecasts, the r^2 values in the region of interest in North America—in the set of grid cells having both adequate gauge density (from Fig. 4; see section 3b) and some clear indication of a robust impact of land initialization on precipitation (an r^2 difference value exceeding 0.10 from Fig. 2). The remainder of North America is whited out, to focus the analysis. The top-right panel shows the equivalent field for the AMIP forecasts. The GLDAS forecasts do appear to reproduce observations slightly better, as indicated by the difference map in the lower-left panel. Note that any quantity in the difference map indexed with a color is significant at the 90% confidence level. Differences of 0.05 and 0.08 are significant at the 95% and 99% levels, respectively. The small negative value in the southwest does not imply a negative impact of land initialization, since it would not pass a test of field significance; we can assume that this negative value reflects sampling error.

The maximum increase in r^2 in the difference map is 0.13. While significant, this increase falls far below the idealized increase from Fig. 2, which for comparison is shown in the lower-right panel of Fig. 5. There are many obvious possible reasons for this: the modeling system is presumably deficient; the initial conditions are presumably imperfect, despite the application of a gauge density criterion; and the validation data are themselves imperfect. Nevertheless, the improvement in the designated area does serve as evidence of a positive impact of soil moisture initialization on precipitation forecast skill.

Enclosed by dotted lines in the two lower plots of Fig. 5 are those grid cells that satisfy the rain gauge criterion and for which the idealized r^2 difference value exceeds 0.30 rather than 0.10. Thus, in this smaller area (“area 2”), idealized predictability in the model is much stronger. Notice that the location of the improvement in skill lies largely within area 2—precisely where the model is apt to provide the most skill. This is presumably not a coincidence.

Figure 6 shows the equivalent four plots for air temperature. The success of the forecasts with land initialization is much stronger, covering much more of the region of focus (which is, to begin with, larger than that for precipitation). The maximum r^2 increase is 0.15, which is significantly different from zero at the 99.9% confidence level.

Again, a comparison of Figs. 3 and 4 suggests that for temperature, evaluations can extend beyond the Great Plains. Figure 7 shows a global version of the r^2 increases associated with land initialization. A few spots (e.g., northeast Brazil, eastern equatorial Africa) show a reduction in r^2 , presumably a reflection of sampling error. Far more of the testable regions show increases in r^2 , further supporting the idea that land initialization contributes to temperature forecast skill. Unfortunately, a similar extension of the precipitation analysis to the

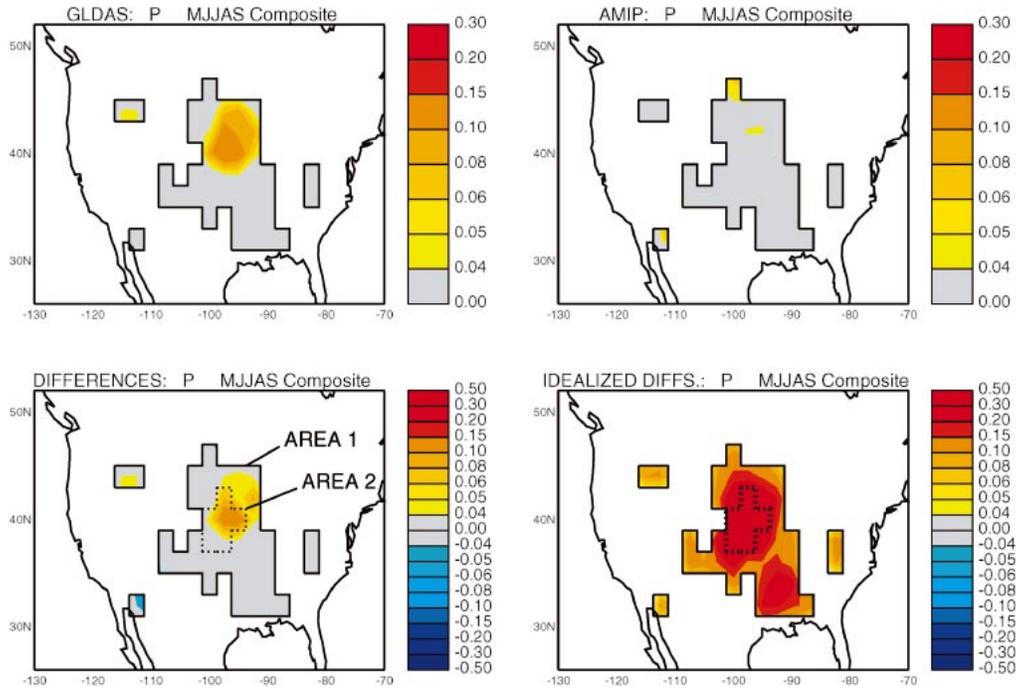


FIG. 5. (top left) Square of the correlation coefficient (r^2) between observed monthly precipitation anomalies and the anomalies predicted by the GLDAS ensemble, computed using 75 data pairs covering May through Sep. (top right) Same, but for the AMIP ensemble. (bottom left) Differences. (bottom right) Differences from the idealized analysis. Areas 1 and 2 are defined by the gauge density criterion and by two levels of potential predictability (0.10 and 0.30) from the idealized analysis.

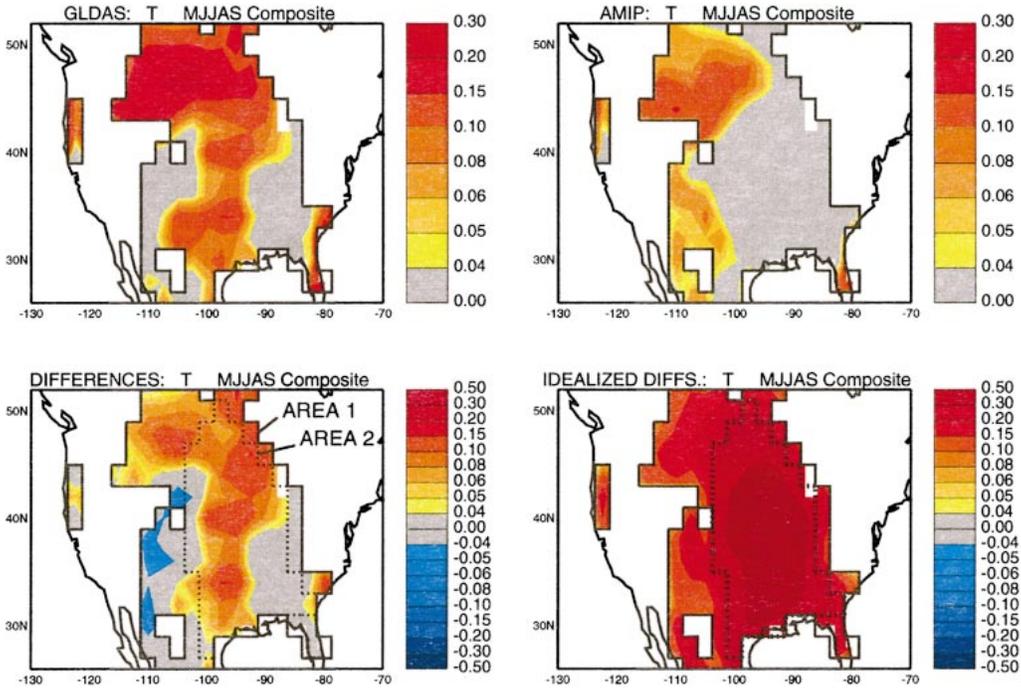


FIG. 6. Same as Fig. 5, but for air temperature.

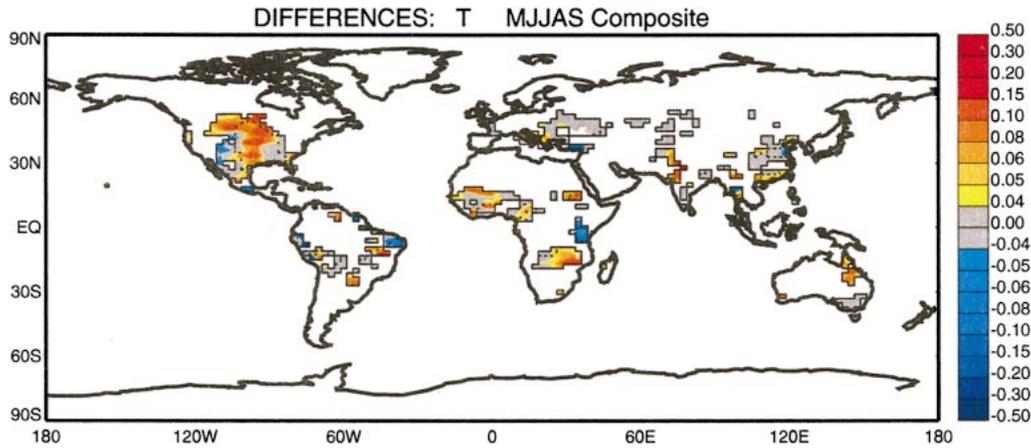


FIG. 7. Global version of the lower-left panel of Fig. 6: differences in the skill levels (the r^2 values, relative to observations) between the GLDAS temperature forecasts and the AMIP temperature forecasts.

globe is not as telling. Only a handful of grid cells outside of North America satisfy the two evaluation criteria, and r^2 differences in these grid cells cannot be distinguished from sampling error. Furthermore, relaxing either the rain gauge density criterion or the potential predictability criterion, in order to allow additional areas for examination of precipitation forecast skill, does not yield useful results; as might be expected, the skill levels in the new, less promising areas are indistinguishable from noise.

Again, all of the skill increases examined in this study are limited by model and data deficiencies. Throughout this analysis, we are, in a sense, determining minimum skill increases. As discussed further in section 4, skill will presumably increase as the models and data get better.

Figure 8 provides a rough indication of how forecast skill varies with month. The r^2 values calculated from a single month's worth of data (15 pairs of values at each grid cell) are relatively unreliable, statistically. If we average these unreliable r^2 values, however, across all grid cells within our areas of focus, we can expect some filtering of the noise and the emergence of an underlying signal. Two areas—area 1 and area 2 from Figs. 5 and 6—are considered here for both precipitation and temperature. The four plots in Fig. 8 show the areal averages of r^2 as a function of month. For both precipitation and temperature forecasts, the average r^2 values in most of the months are indeed generally higher when the soil moisture is properly initialized, especially for area 2, for which predictive skill should indeed be higher. Improvement in precipitation forecast skill is evidently highest in May through July. For temperature forecasts, land initialization seems to provide the most skill during August and September. The reasons for these monthly differences are not currently known. Monthly variations in idealized skill (not shown) are not so large.

d. Impact of atmospheric initialization

Up to now, atmospheric initialization has not been a focus of this paper. The individual ensemble members in the experiments above were not initialized with reanalysis fields but rather with very different atmospheric conditions, taken from the broad range of possible states that are consistent with the imposed SSTs. To examine the relative importance of land and atmospheric initialization, we performed two sets of 1-month nine-member ensemble forecasts for each June during 1979–93. In the first set (hereafter referred to as GLDAS-Atm), the member simulations were initialized with both the GLDAS land surface states and with atmospheric anomalies from the NCEP reanalysis [provided by the NOAA–Cooperative Institute for Research in Environmental Sciences (CIRES) Climate Diagnostics Center, Boulder, Colorado, from their Web site at <http://www.cdc.noaa.gov>]. In effect, the atmospheric anomalies from the NCEP reanalysis (relative to the reanalysis's climatology) were applied to the AGCM's own climatological mean state, and perturbations were imposed in all atmospheric variables to allow the different ensemble members to evolve independently. (These perturbations were, of course, small enough to ensure that each set of imposed anomalies looks very much like the unperturbed set of anomalies.) In the second series of ensembles (hereafter referred to as Atm), the atmosphere in each ensemble member was similarly initialized with reanalysis data, but the land was not initialized with GLDAS states—the same set of land states employed in the AMIP simulations, which represents the full distribution of land states consistent with the imposed SSTs, was used instead.

Because only June simulations were performed, all of the computed statistics for the GLDAS-Atm and Atm simulations are based on 15 values rather than 75 values. Therefore, these statistics, like those for the monthly r^2

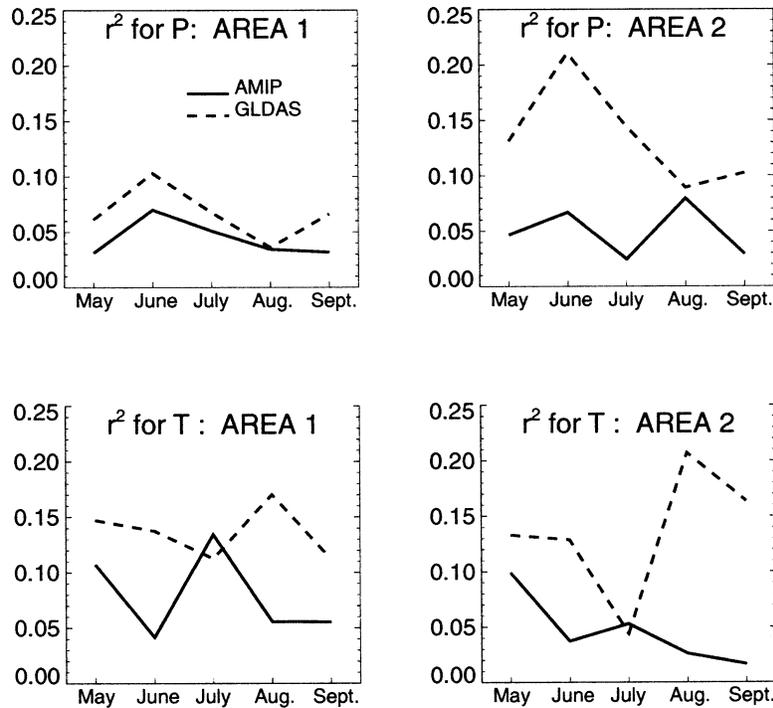


FIG. 8. (top left) Monthly r^2 values averaged over area 1 from Fig. 5, for both the GLDAS and AMIP simulations. (top right) Same, but for averages over area 2 from Fig. 5. (bottom) The corresponding plots for air temperature, with areas 1 and 2 defined in Fig. 6.

values in Fig. 8, are somewhat unreliable. Nevertheless, they can still provide a rough indication of relative skill, especially when averaged over our area of focus (area 1 for each variable in Figs. 5 and 6).

We begin with an idealized analysis, equivalent to that performed in section 3a. The histograms in Fig. 9 show, for both the first and second halves of June, the degree to which the model can predict itself (i.e., the degree to which atmospheric chaos alone would foil the forecast) under the different initialization scenarios. For clarity, the bars are identified according to the aspects of the system that can provide skill. For the AMIP ensemble, this can only be the specification of the SSTs. For the GLDAS ensemble, only the SSTs and the land initialization contribute to skill, and for the Atm ensemble, only the SSTs and the atmospheric initialization do. All three elements contribute to skill in the GLDAS-Atm ensemble. Each bar represents an average of the r^2 values over area 1.

For both precipitation and temperature, the contribution of atmospheric initialization to idealized predictability is quite large during the first half of the month and is much smaller during the second half. This is consistent with current understanding of operational numerical weather prediction. The contribution of land initialization to idealized predictability, on the other hand, is roughly the same in both halves of the month. During the second half of June, the contribution of land

initialization is about the same as that of atmospheric initialization. Notice that the maximum predictability is obtained when both the land and the atmosphere are initialized. The contributions from the land and atmosphere even seem additive, as if the system were linear.

Figure 10 shows the monthly skill levels obtained when the forecasts are compared to observations. For precipitation, land initialization appears to contribute slightly more to monthly skill. For temperature, atmospheric initialization appears more important, though maximum skill is obtained when both land and atmosphere are initialized.

Unfortunately, because only 1 month is considered here and the underlying skill levels are small, sampling error prevents a statistical evaluation of skill for the separate halves of June. Qualitatively, the results look similar to those in Fig. 9, at least in terms of the relative performance of the different initialization procedures. When all 5 months of the GLDAS and AMIP ensembles are considered together (not shown), the land's contribution to temperature forecast skill appears to be weighted heavily to the first half of the month. The land's contribution to precipitation skill, on the other hand, appears roughly the same throughout the month.

4. Summary and discussion

The forecasts examined herein allow a first assessment of the impact of land initialization on 1-month

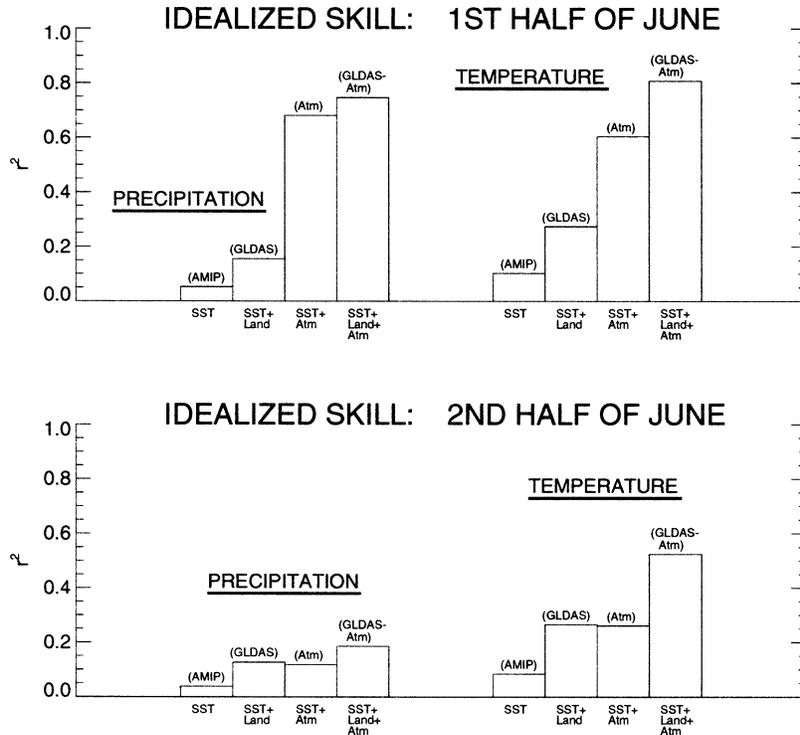


FIG. 9. (top) (left) Average of the r^2 values across area 1 in Fig. 5 for idealized precipitation forecasts (wherein the model “predicts itself”; see section 3a) during the first half of Jun—results are shown for four forecast ensembles (AMIP, GLDAS, Atm, and GLDAS-Atm), each making use of a unique combination of three different elements (SST specification, land initialization, and atmospheric initialization) contributing to forecast skill; (right) same, but for air temperature over area 1 in Fig. 6. (bottom) Same, but for the second half of Jun.

forecast skill. For precipitation, analysis is unfortunately limited to a small area of North America (centered on the Great Plains), for this is the only area that jointly satisfies two criteria during the study period: first, that the model shows some predictability in an idealized analysis (section 3a), and second, that the precipitation

is adequately measured, as determined by a critical rain gauge density (section 3b). In this region, forecast skill—for both precipitation and air temperature—is indeed higher for the ensembles using realistic land initialization than for the ensembles in which the land is not initialized. In places, the improvement is statistically

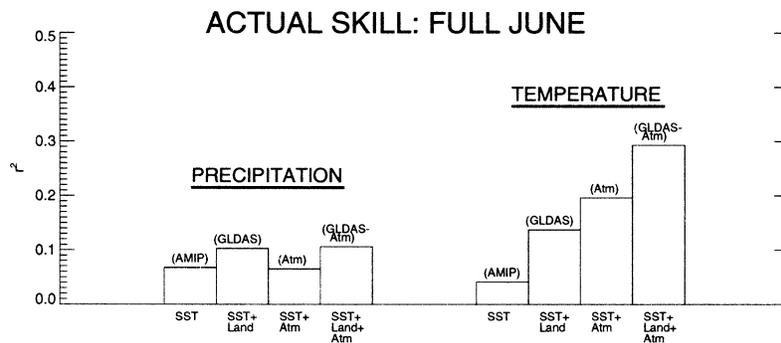


FIG. 10. (left) Average of the r^2 values for Jun precipitation forecasts (computed through regressions against observations) across area 1, as outlined in Fig. 5. Results are shown for four forecast ensembles (AMIP, GLDAS, Atm, and GLDAS-Atm), each making use of a unique combination of three different elements (SST specification, land initialization, and atmospheric initialization) contributing to forecast skill. (right) Same, but for air temperature over area 1 in Fig. 6.

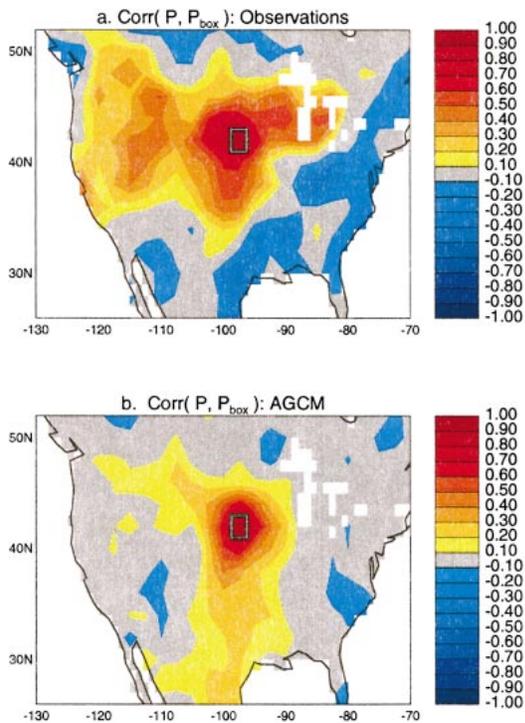


FIG. 11. (a) Correlation between precipitation time series in the outlined box and that in each grid cell of the United States, as determined from an observational dataset (Higgins et al. 2000). (b) Same, but using precipitation from AGCM simulations.

significant at the 99% confidence level. Furthermore, for temperature prediction, significant forecast skill can be seen outside the Great Plains region (Fig. 7).

Further analysis shows that land initialization, coupled with atmospheric initialization, improves over atmospheric initialization alone. The data on submonthly skill levels are noisy, but the idealized analysis suggests that in the last half of the month, land and atmosphere initialization contribute roughly the same amount to the potential predictability of precipitation and temperature.

The skill improvements shown in this paper are significant but are also quite small—perhaps too small to be of practical use. Here we must reiterate that the small improvements are, in a sense, *minimum* improvements. Current skill is limited by inadequacies in both the modeling analysis and the available observational data.

One inadequacy in the modeling is illustrated in Fig. 11. Figure 11a shows the correlation between the time series of monthly precipitation amounts in the outlined box and concurrent time series across the rest of the continental United States, as determined from observations (Higgins et al. 2000). The plotted correlations thus, in a sense, reflect the spatial structure of monthly precipitation anomalies in nature. Figure 11b shows the same diagnostic computed from precipitation rates generated in an AMIP simulation. The spatial structure of the correlation field is much larger in the observations,

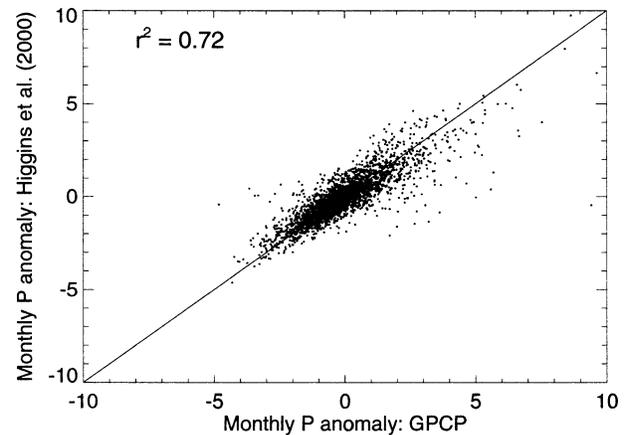


FIG. 12. Scatterplot comparing monthly rainfall anomalies over May through Sep from two data sources: the GPCP version 2 dataset (Adler et al. 2003), as processed by Berg et al. (2003), and the Higgins et al. (2000) dataset. The former was used in the initialization of the land model.

implying that the AGCM underestimates the spatial extent of precipitation anomalies. The box outlined in the figure is representative; other boxes in the area produce similar results.

The misrepresentation of spatial precipitation structures in the AGCM may or may not reflect an inability of the model to capture remote impacts of soil moisture on precipitation. In any case, it suggests that the areas of idealized predictability in the model (Figs. 2 and 3) may be underestimated. Improvements in the model's dynamics may extend significantly the areas of potential skill shown in the figures. Even in the absence of model improvement, statistical procedures that take advantage of the known spatial correlations in Fig. 11a could be used to improve the model forecasts. Preliminary results (not shown) from this line of inquiry are encouraging.

Further limiting prediction skill are imperfections in the land initialization and forecast evaluation procedures. The data we apply in the initialization sequence, for example, have significant uncertainties. Figure 12 shows a scatterplot comparing monthly precipitation totals from the GPCP dataset (which was used in the initialization) and the more measurement-intensive dataset of Higgins et al. (2000) over grid cells within the area of focus (see Fig. 5). The two datasets do agree to first order, but the points are scattered around the 1:1 line ($r^2 = 0.72$), despite the application of the rain gauge criterion in defining the area. The scatter belies an uncertainty that may have compromised the initial soil moistures we used. Additional problems undoubtedly arise from errors in the day-by-day temporal disaggregation of the monthly precipitation totals, which was determined from reanalysis, and from deficiencies in the LSM's ability to convert the forcing into anomalies with the proper magnitude and memory. As datasets and models improve—as we take advantage, for example,

of new satellite measurements of soil moisture analyzed in a data assimilation framework, new estimates of precipitation from global satellite networks, and improvements in model formulation—the skill associated with the initialization should increase.

When evaluating skill in this paper, all years were given equal weight. Worth investigating is the idea that some years—particularly years with extreme initial conditions—may be easier to predict. In the midwestern U.S. drought of 1988, for example, the average observed June precipitation anomaly (area 2 in Fig. 5) was -0.34 mm day $^{-1}$. The GLDAS ensemble predicted an average anomaly of -0.38 mm day $^{-1}$, whereas the AMIP ensemble generated a positive average anomaly of 0.3 mm day $^{-1}$. More research is needed to clarify the relative predictability of extreme versus nonextreme years.

Our long-term goal, of course, is to achieve the potential forecast skill indicated in Figs. 2 and 3—and beyond, if improvements in the model allow. Again, with an improved model and improved observations, the areas over which we can look for skill should expand considerably. For precipitation, we will no longer need to limit our evaluations to the Great Plains of North America.

Acknowledgments. The AGCM runs were funded by the Earth Science Enterprise of NASA Headquarters through the EOS-Interdisciplinary Science Program and the NASA Global Modeling and Assimilation Office (GMAO), with computational resources provided by the NASA Center for Computational Sciences.

REFERENCES

- Adler, R. F., and Coauthors, 2003: The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167.
- Bacmeister, J., P. J. Pegion, S. D. Schubert, and M. J. Suarez, 2000: *Atlas of Seasonal Means Simulated by the NSIPP 1 Atmospheric GCM*. NASA Tech. Memo. 2000-104606, Vol. 17, 191 pp.
- Beljaars, A. C. M., P. Viterbo, M. Miller, and A. K. Betts, 1996: The anomalous rainfall over the United States during July 1993: Sensitivity to land surface parameterization and soil moisture anomalies. *Mon. Wea. Rev.*, **124**, 362–383.
- Berg, A. A., J. S. Famiglietti, J. P. Walker, and P. R. Houser, 2003: Impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes. *J. Geophys. Res.*, **108**, 4490, doi:10.1029/2002JD003334.
- Chou, M.-D., and M. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models. NASA Tech. Memo. 104606, Vol. 3, 84 pp.
- Dirmeyer, P. A., M. J. Fennessy, and L. Marx, 2003: Low skill in dynamical prediction of boreal summer climate: Grounds for looking beyond sea surface temperature. *J. Climate*, **16**, 995–1002.
- Douville, H., and F. Chauvin, 2000: Relevance of soil moisture for seasonal climate predictions: A preliminary study. *Climate Dyn.*, **16**, 719–736.
- Entin, J. K., A. Robock, K. Y. Vinnikov, S. E. Hollinger, S. Liu, and A. Namkhai, 2000: Temporal and spatial scales of observed soil moisture variations in the extratropics. *J. Geophys. Res.*, **105**, 11 865–11 877.
- Fennessy, M. J., and J. Shukla, 1999: Impact of initial soil wetness on seasonal atmospheric prediction. *J. Climate*, **12**, 3167–3180.
- Gibson, J. K., P. Kallberg, S. Uppala, A. Nomura, A. Hernandez, and E. Serrano, 1997: ERA description. ECMWF Reanalysis Project Report Series 1, European Centre, for Medium-Range Weather Forecasts, Reading, United Kingdom, 72 pp.
- Gupta, S. K., N. A. Ritchey, A. C. Wilber, C. H. Whitlock, G. G. Gibson, and P. W. Stackhouse Jr., 1999: A climatology of surface radiation budget derived from satellite data. *J. Climate*, **12**, 2691–2710.
- Higgins, R. W., W. Shi, E. Yarosh, and R. Joyce, cited 2000: *Improved United States Precipitation Quality Control System and Analysis*. NCEP/Climate Prediction Center Atlas 7. [Available online at <http://www.cpc.ncep.noaa.gov/research-papers/ncep-cpc-atlas/7/index.html>.]
- Koster, R. D., and M. Suarez, 1996: Energy and water balance calculations in the Mosaic LSM. NASA Tech. Memo. 104606, Vol. 9, 59 pp.
- , and P. C. D. Milly, 1997: The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models. *J. Climate*, **10**, 1578–1591.
- , and M. J. Suarez, 2003: Impact of land surface initialization on seasonal precipitation and temperature prediction. *J. Hydrometeorol.*, **4**, 408–423.
- , M. J. Suarez, and M. Heiser, 2000: Variance and predictability of precipitation at seasonal-to-interannual timescales. *J. Hydrometeorol.*, **1**, 26–46.
- , P. A. Dirmeyer, A. N. Hahmann, R. Ijpelaar, L. Tyahla, P. Cox, and M. J. Suarez, 2002: Comparing the degree of land–atmosphere interaction in four atmospheric general circulation models. *J. Hydrometeorol.*, **3**, 363–375.
- Kumar, A., and M. P. Hoerling, 1995: Prospects and limitations of seasonal atmospheric GCM predictions. *Bull. Amer. Meteor. Soc.*, **76**, 335–345.
- Moorthi, S., and M. J. Suarez, 1992: Relaxed Arakawa–Schubert, a parameterization of moist convection for general circulation models. *Mon. Wea. Rev.*, **120**, 978–1002.
- New, M., M. Hulme, and P. Jones, 2000: Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate*, **13**, 2217–2238.
- Oki, T., T. Nishimura, and P. Dirmeyer, 1999: Assessment of annual runoff from land surface models using total runoff integrating pathways (TRIP). *J. Meteor. Soc. Japan*, **77**, 235–255.
- Owe, M., R. de Jeu, and J. Walker, 2001: A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Trans. Geosci. Remote Sens.*, **39**, 1643–1654.
- Reichle, R., and R. D. Koster, 2003: Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *J. Hydrometeorol.*, **4**, 1229–1242.
- , —, J. Dong, and A. Berg, 2004: Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation. *J. Hydrometeorol.*, **5**, 430–442.
- Robock, A., K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai, 2000: The global soil moisture data bank. *Bull. Amer. Meteor. Soc.*, **81**, 1281–1299.
- Rodell, M., and Coauthors, 2004: The global land data assimilation system. *Bull. Amer. Meteor. Soc.*, **85**, 381–394.
- Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science*, **282**, 728–731.
- Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N. C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.*, **103**, 14 291–14 324.
- Vinnikov, K. Y., and I. B. Yeserkepova, 1991: Soil moisture, empirical data and model results. *J. Climate*, **4**, 66–79.
- Viterbo, P., and A. K. Betts, 1999: Impact of the ECMWF reanalysis soil water on forecasts of the July 1993 Mississippi flood. *J. Geophys. Res.*, **104**, 19 361–19 366.

Walker, J. P., and P. R. Houser, 2001: A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. *J. Geophys. Res.*, **106**, 11 761–11 774.

Willmott, C. J., and K. Matsuura, cited 2001: Terrestrial air temperature and precipitation: Monthly and annual time series (1950–1999) version 1.02. [Available online at <http://www.siesip.gmu.edu/home.html>.]