

The Second Phase of the Global Land–Atmosphere Coupling Experiment: Soil Moisture Contributions to Subseasonal Forecast Skill

R. D. KOSTER,^a S. P. P. MAHANAMA,^{a,b,c} T. J. YAMADA,^{a,b,d} GIANPAOLO BALSAMO,^e A. A. BERG,^f
M. BOISSERIE,^{g,h} P. A. DIRMEYER,ⁱ F. J. DOBLAS-REYES,^{j,k} G. DREWITT,^f C. T. GORDON,^l Z. GUO,ⁱ
J.-H. JEONG,^m W.-S. LEE,ⁿ Z. LI,^{a,c} L. LUO,^{o,p} S. MALYSHEV,^p W. J. MERRYFIELD,ⁿ S. I. SENEVIRATNE,^q
T. STANELLE,^q B. J. J. M. VAN DEN HURK,^r F. VITART,^c AND E. F. WOOD^p

^a GMAO, NASA Goddard Space Flight Center, Greenbelt, Maryland

^b UMBC/GEST, Baltimore, Maryland

^c SAIC, Beltsville, Maryland

^d Division of Field Engineering for Environment, Hokkaido University, Sapporo, Japan

^e ECMWF, Reading, United Kingdom

^f Department of Geography, University of Guelph, Guelph, Canada

^g Center for Ocean–Atmospheric Prediction Studies, The Florida State University, Tallahassee, Florida

^h Météo-France, Toulouse, France

ⁱ Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland

^j Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

^k Institut Català de Ciències del Clima (IC3), Barcelona, Spain

^l NOAA/GFDL, Princeton, New Jersey

^m Department of Earth Sciences, University of Gothenburg, Gothenburg, Sweden

ⁿ CCCMA, Environment Canada, Victoria, Canada

^o Department of Geography, Michigan State University, East Lansing, Michigan

^p Princeton University, Princeton, New Jersey

^q Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

^r KNMI, De Bilt, Netherlands

(Manuscript received 3 September 2010, in final form 18 February 2011)

ABSTRACT

The second phase of the Global Land–Atmosphere Coupling Experiment (GLACE-2) is a multi-institutional numerical modeling experiment focused on quantifying, for boreal summer, the subseasonal (out to two months) forecast skill for precipitation and air temperature that can be derived from the realistic initialization of land surface states, notably soil moisture. An overview of the experiment and model behavior at the global scale is described here, along with a determination and characterization of multimodel “consensus” skill. The models show modest but significant skill in predicting air temperatures, especially where the rain gauge network is dense. Given that precipitation is the chief driver of soil moisture, and thereby assuming that rain gauge density is a reasonable proxy for the adequacy of the observational network contributing to soil moisture initialization, this result indeed highlights the potential contribution of enhanced observations to prediction. Land-derived precipitation forecast skill is much weaker than that for air temperature. The skill for predicting air temperature, and to some extent precipitation, increases with the magnitude of the initial soil moisture anomaly. GLACE-2 results are examined further to provide insight into the asymmetric impacts of wet and dry soil moisture initialization on skill.

1. Introduction

The idea that soil moisture can influence the variability of precipitation and air temperature has been explored

extensively in the literature, using both numerical climate models (e.g., Shukla and Mintz 1982, Delworth and Manabe 1989, Dirmeyer 2000, Douville et al. 2001, and Hong and Kalnay 2000, among many others) and observational analysis (e.g., Betts and Ball 1995; Findell and Eltahir 1997). The impact of soil moisture on near-surface air temperature is straightforward and well documented in models and observations (e.g., Koster et al. 2009b):

Corresponding author address: Randal Koster, Global Modeling and Assimilation Office, Code 610.1, NASA GSFC, Greenbelt, MD 20771.

E-mail: randal.d.koster@nasa.gov

DOI: 10.1175/2011JHM1365.1

higher soil moisture can induce higher evaporation and thus greater evaporative cooling of the surface and the overlying air. The impact of soil moisture on precipitation is more complex. If higher soil moisture induces higher evaporation, the correspondingly lower sensible heat flux can lead to shallower boundary layers and thus an easier buildup of the conditions that trigger convective rainfall (Betts et al. 1994). In addition, the higher evaporation can serve as a moisture source. However, under certain conditions, higher evaporation rates may have the opposite effect (Findell and Eltahir 2003; Cook et al. 2006; Van den Hurk and Van Meijgaard 2010)—they may act to inhibit precipitation when the growth of the boundary layer is too small to reach the lifting condensation level. Note that evaporation impacts need not be strictly local; an evaporation anomaly that affects the large-scale circulation may affect precipitation rates in remote regions (Douville 2002).

The numerous modeling studies addressing atmospheric response to soil moisture variations were formalized recently in the first phase of the Global Land–Atmosphere Coupling Experiment (GLACE), a coordinated international numerical experiment involving a dozen independent modeling systems (Koster et al. 2004b, 2006; Guo et al. 2006). In essence, the experiment quantified “land–atmosphere coupling strength” in the participating models (i.e., the degree to which the evolution of simulated precipitation and air temperature can be guided by prescribed, model-specific time variations in soil moisture content). The original GLACE study produced two key results: (i) models differ significantly in their estimates of how soil moisture variations affect precipitation and air temperature, and (ii) they nevertheless tend to agree that these meteorological variables are particularly affected by soil moisture variations in specific regions: the transition zones between arid and humid regions.

One key motivation for studying and quantifying land–atmosphere coupling strength is the idea that soil moisture may play a role in meteorological forecasting (Seneviratne et al. 2010). This coupling strength is indeed one of the two critical elements underlying soil moisture’s ability to influence forecasts. The other critical element is the forecasting of soil moisture itself—to take advantage of the coupling strength, a soil moisture anomaly must be initialized realistically (not always straightforward given limitations in the availability of accurate real-time data) and then be “remembered” into the forecast period. The persistence time scales of soil moisture in nature can extend out to about two months (Vinnikov and Yeserkepova 1991; Entin et al. 2000; Seneviratne et al. 2006). The hope is that if soil moisture is modeled correctly and a forecast model

adequately captures the land–atmosphere coupling present in nature, an initialized soil moisture anomaly in a forecast system may persist long enough to improve forecasts at subseasonal time scales.

A handful of studies have examined the impact of initializing a forecast system with observation-based estimates of soil moisture (e.g., Fennessy and Shukla 1999; Viterbo and Betts 1999; Douville and Chauvin 2000; Koster et al. 2004a; Douville 2010) with generally optimistic results. These studies, however, are uncoordinated, model specific, and very difficult to compare given their different forecast periods and sample sizes and their different verification metrics. Given the potential importance of soil moisture in forecasting, the Global Energy and Water Cycle Experiment (GEWEX) and Climate Variability (CLIVAR) activities of the World Climate Research Programme (WCRP) have recently sponsored the second phase of the Global Land–Atmosphere Coupling Experiment (GLACE-2). In GLACE-2, the same numerical experiment, one specifically designed to isolate soil moisture initialization impacts on subseasonal forecast skill (measured against real observations, with a large number of independent start dates for robust statistics), is performed with 11 independent forecast systems. The result is a coordinated “consensus” view of the degree to which realistic land initialization improves forecast quality in today’s models.

Some first results from GLACE-2 focusing on the United States have already been published (Koster et al. 2010) as has a companion paper focusing on Europe (Van den Hurk et al. 2011). The present paper provides a broad overview of the experiment and a much-expanded analysis of the results.

2. Experimental design

a. Overview

Participants in GLACE-2 performed two series of forecasts covering a wide range of boreal warm-season start dates. Each series was characterized as follows:

Length of each forecast: 2 months (more precisely, 60 days).

Start dates: 1 April, 15 April, 1 May, 15 May, 1 June, 15 June, 1 July, 15 July, 1 August, and 15 August in each of the years 1986–95.

Total number of start dates: 100.

Number of ensemble members per forecast: 10.

The two series differ from each other only in the nature of their land surface initialization. In Series 1, land conditions are initialized to realistic values based on an integration of historical meteorological forcing (see below).

In Series 2, realistic land surface initialization is not utilized. By comparing the forecast skill quantified for Series 1 and 2, we isolate the impacts of land initialization on the skill. Note that much of the wording in this text reflects our presumption that of all the land variables initialized, soil moisture affects forecast skill by far the most; nevertheless, land variables such as subsurface temperature are also initialized in Series 1 and may have some impact in certain areas (Mahanama et al. 2008).

A fundamental pragmatism underlying GLACE-2 is that the modeling groups participate not only to contribute to a multimodel-consensus view of land impacts on skill but also to reveal the potential of their own particular (sometimes operational) systems to make use of soil moisture data. Thus, while certain aspects of the experimental design are enforced across models, all groups are given some freedom in tailoring the experiments to address their own particular needs. An example lies in model resolution: each group was allowed to choose the model resolution used for both the land and atmosphere components of the prediction system. Groups were also allowed to use their own system-specific sets of vegetation boundary conditions.

The design of the experiment involved model-specific technical choices regarding setup and the processing of the simulations. Many of the technical guidelines are provided in an online experimental plan (<http://gmao.gsfc.nasa.gov/research/GLACE-2/docs/GLACE2c.pdf>). The remainder of this section describes some of the key aspects of the experimental design.

b. Land surface variable initialization

In effect, initial land surface states for Series 1 were established through participation in the Global Soil Wetness Project Phase 2 (GSWP-2). GSWP is an environmental modeling research activity of the Global Land–Atmosphere System Study (GLASS) and the International Satellite Land Surface Climatology Project (ISLSCP), both of which are contributing projects of GEWEX. Through GSWP-2, modelers produced global fields of land surface fluxes, state variables, and related hydrologic quantities by driving their land surface models offline with global arrays of observation-based meteorological forcing (Dirmeyer et al. 2002, 2006). Much of this forcing is based on reanalysis, but key aspects of the forcing were scaled so that their long-term means agree with independent observational datasets. In particular, precipitation was adjusted to agree at the monthly time scale with a gauge-based GPCC dataset.

The forcing used for GSWP-2 spans the period 1986–95 at a resolution of 1°; GLACE-2 participants were given some freedom in regard to the configuration of the forcing to their own model grid (see the aforementioned

Web site). Note that the Florida State University (FSU) and Canadian Centre for Climate Modelling and Analysis (CCCma) models (see section 3) used analogous datasets produced by Sheffield et al. (2006) and Berg et al. (2005), respectively, for the offline forcing exercise; because all of these datasets rely, for the most part, on the same observations, we do not expect the use of different datasets to have a material effect on the results.

It is worth noting that the real-time forcing data available for use in operational forecasting is probably somewhat less accurate than the forcing data used here. Thus, while the skill levels achieved in Series 1 will illustrate the potential for skill improvement connected with land initialization, these skill levels can, in at least one sense, be interpreted as an “upper bound” for what might be achieved with current real-time measurement networks. Upcoming versions of operational systems, however, will use soil moisture estimates based on the assimilation of satellite-derived soil moisture data, which may give them a relative advantage.

The model soil moisture states at the forecast start times produced through the offline forcing exercise were used to initialize the 2-month Series 1 forecasts in GLACE-2. First, though, the initial states were “scaled” to the forecast model’s climatology to account for possible biases between the model’s climate and the values derived offline with GSWP-2 forcing: a relatively dry (wet) state obtained through the offline exercise for a given region may be a relatively wet (dry) state in the forecast system because the forecast system may be biased dry (wet) in the region. By scaling, a relatively dry state generated offline with GSWP-2 forcing can be converted to a correspondingly dry state for the coupled model system. We note, however, the philosophical counterargument that soil moisture generated offline, while biased relative to a model’s climatology, is still more “realistic,” and thus some may consider the use of non-scaled variables preferable for operational prediction.

Most groups scaled a given land state X (e.g., soil moisture content in a given layer) at a given location and time through the use of standard normal deviates, or Z scores, based on the first two statistical moments of X for the location and initialization day of year considered. In essence, the Z score for X obtained in the offline forcing system (using the offline system’s moments) was assumed to apply in the coupled system and was converted to an absolute value of X using the coupled system’s moments (see Koster et al. 2004a for details). Suitable constraints, of course, must be applied to the scaled fields, ensuring, for example, that soil moistures don’t fall below the wilting point or become supersaturated. Other groups used a more rigorous scaling approach, such as the matching of cumulative distribution functions,

and FSU–Center for Ocean–Atmospheric Prediction Studies (COAPS) avoided scaling altogether by using a land surface assimilation technique in coupled mode.

For the Series 2 forecasts, the initial land states for the members of a given forecast ensemble are not identical, and they are not set to realistic values; rather, they are drawn from a distribution of potential states, the distribution determined from long-term simulations with the model. Most groups had archived restart files spanning decades or more for their modeling system; the land states from these restart files on a given start date (for 10 different years, spaced as far apart as possible) were used to initialize the different ensemble members. Some other groups took advantage of existing parallel Atmospheric Model Intercomparison Project (AMIP) simulations, for which multiple restart files for a given start date and year could be used to provide the same number of sets of independent land initial conditions that were each consistent with that year’s sea surface temperature (SST) distribution. [Note that with this approach, if the Series 2 initializations extract realism from antecedent SST-induced springtime forcing, then the models using this approach will be penalized when skill differences (Series 1 minus Series 2) are computed; this effect, however, is expected to be small. The Series 2 SST-induced skill scores are, in any case, small for boreal summer.] Still others “shuffled” the initial conditions generated for the Series 1 forecasts for use in the Series 2 forecasts.

c. Atmospheric initialization

If possible, the atmosphere was initialized realistically (i.e., with fields representing the actual state of the atmosphere on the forecast start date). Appropriate atmospheric conditions were extracted from existing reanalyses. Participants then aggregated, disaggregated, or interpolated these atmospheric conditions onto their own model grids and generated 10 different sets of atmospheric initial conditions for each ensemble, using their choice of ensemble generation technique (e.g., sampling every few hours from a one-day simulation initialized with the reanalysis fields, or perturbing the initial states with singular vectors). For logistical reasons, some modeling groups [e.g., the groups using the National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO) models and ECHAM–Jena Scheme for Biosphere–Atmosphere Coupling in Hamburg (JSBACH)] did not utilize reanalysis-based initialization, using instead, for example, atmospheric conditions produced in a free-running atmospheric model simulation with prescribed SSTs. Analyzing the impact of the different atmospheric initialization approaches is beyond the scope of this study; we implicitly assume here that the impact is small at the subseasonal

[beyond numerical weather prediction (NWP)] time scale, particularly given that the same approach is used within an individual model for both Series 1 and 2, with the land impacts isolated by differencing the results from the two series.

d. The ocean boundary condition

GLACE-2 is designed to isolate the impacts of land initialization on subseasonal predictability (e.g., how quickly the impact of land initial conditions on a forecast reduces with time) and forecast skill. Thus, model-to-model variations in predictability associated with ocean processes were avoided; with two exceptions, participants ran their forecast simulations with the same set of prescribed SSTs. The prescribed SSTs for each forecast period, provided to all participants, were constructed by applying simple persistence measures to the SST anomalies present on the forecast start date. These persistence measures were obtained through analysis of long-term SST datasets (e.g., Reynolds and Smith 1994) using cross validation and are appropriate for use given the short length of the forecast simulations (two months) relative to the long time scales of ocean variability. One model (CCCma) used a coupled atmosphere–ocean system with SSTs initialized on the forecast start date, allowing the SSTs to evolve with the other model fields. Another [Center for Ocean–Land–Atmosphere Studies (COLA)] did use observed SSTs throughout the forecast period in both series; again, given that land-derived skill is quantified by differencing the Series 1 and 2 results, the impact of using observed SSTs in this one model is expected to be minimal.

e. Output diagnostics

The analyses throughout this paper focus on forecast skill generated at subseasonal time scales but beyond short- to medium-range weather forecasts (0–10 days, for which intermodel differences in the approaches used to initialize the atmosphere muddy the analysis). In this paper, we examine in particular the models’ ability to predict rainfall and air temperature averaged over three 15-day forecast periods: days 16–30, 31–45, and 46–60. Each participating modeling group thus provided to the GLACE-2 organizers their total precipitation and air temperature data averaged over the four 15-day periods of each 60-day forecast simulation. Corresponding averages of evaporation, net radiation at the surface, vertically integrated soil moisture content, and near-surface relative humidity were also provided. The participants provided their data on their models’ native grids; these data were subsequently regridded, using a simple area weighting procedure, to a common (2° latitude \times 2.50° longitude) grid prior to joint model analysis.

3. Participants

Ten modeling systems performed the full suite of GLACE-2 experiments as described above. One additional system performed the experiment for a large subset of the forecast start dates. Together the 11 systems represent the state of the art in atmospheric modeling and prediction. Details on the systems are provided in Table 1.

Not included here are additional simulations performed by the European Centre for Medium-Range Weather Forecasts (ECMWF) group using a fully coupled land–atmosphere–ocean forecast system. These results were very similar to those of the ECMWF system using prescribed SSTs (which are included here)—the use of the coupled ocean had no marked impact on land-related skill.

4. Results

a. Forecast skill metric

To strengthen the potential for land impacts, our analysis focuses on the warmest forecast periods in the Northern Hemisphere; namely, those 15-day periods (excluding the initial 15 days of the forecast) that lie in June, July, or August. Thus, a forecast initialized on 15 April would provide data only for the 46–60-day period, a forecast initialized on 1 June would provide data for all three considered leads, and a forecast initialized on 1 August would provide data only for the 16–30-day period. Through this procedure, the analysis for each lead for a given model consists of 60 separate and independent forecasts (10 years of forecasts, with 6 forecasts per year).

The daily precipitation observations used for forecast evaluation were derived from a combined station–satellite pentad dataset (Xie et al. 2003; see ftp://ftp.cpc.ncep.noaa.gov/precip/GPCP_PEN). The closest set of three pentads to a given 15-day forecast period were used; errors of one or two days in the overlap period should have a small impact and, in any case, would only hinder the computed skill levels. Daily temperature observations were estimated by averaging the minimum and maximum daily temperatures (again derived from station observations) stored in the Hadley Centre archives (<http://hadobs.metoffice.com/hadghcnd/>; Caesar et al. 2006). This dataset includes large data gaps, particularly in South America and Africa. While a reanalysis-based dataset could be used to fill in these gaps, we decided to avoid any validation dataset that includes a model-based component, since the goal here is to evaluate the models strictly with observations—the model physics and diagnostic approaches underlying

a reanalysis affect both precipitation and air temperature products (e.g., Pitman and Perkins 2009), particularly in areas of sparse observations. Constructing a new submonthly gridded temperature dataset over South America and Africa from available observations is beyond the scope of this study.

The forecast skill metric used in this paper is the square of the correlation coefficient (r^2) between the predicted variable and the observed variable. The isolated contribution of realistic land initialization to forecast skill is computed as the difference of the r^2 values obtained for Series 1 and 2 ($r_{\text{ser1}}^2 - r_{\text{ser2}}^2$). We note that other metrics, such as root-mean-square error (RMSE), could be examined instead, but such metrics are clouded by biases in model climate; our goal here is rather to determine the degree to which the models forecast correctly the temporal variability and relative magnitudes of temperature or precipitation anomalies. Because the means and standard deviations of both the observational time series and the forecast results for any model are known, the r^2 metric can be directly transformed into an RMSE metric, or better yet, biased model outputs can be scaled to reproduce observations-based moments (Entekhabi et al. 2010).

b. Intermodel variability in skill

We discuss only briefly here the intermodel differences in quantified skill levels. Figure 1 shows the skill for precipitation and air temperature obtained for Series 1 (r_{ser1}^2) during days 16–30 of the forecasts at three representative $2^\circ \times 2.5^\circ$ grid cells in the continental United States. The bars in each histogram represent different models (in random order, though the same order is used in each panel). The skill levels are seen to vary greatly, and models that appear to do well for one field at one location may do poorly for another field or another location. While skill maps (not shown) suggest that some models do appear, at face value, to perform generally better than others, a quantitative, statistically valid comparative evaluation of the models (in the context of both inherent skill differences and sampling error) is beyond the scope of this paper, which focuses instead on the determination and characterization of an overall multimodel-consensus vision of skill. The intermodel variability illustrated in Fig. 1 should nevertheless be kept in mind when interpreting the consensus results.

c. Unconditional consensus skill levels

Koster et al. (2010) processed the GLACE-2 results into consensus estimates of land-derived forecast skill (r^2 for Series 1 – r^2 for Series 2) over the continental

TABLE 1. Subseasonal-to-seasonal forecast systems participating in GLACE-2.

System name, resolution, and No. of JJA forecasts submitted	References	Notes
Canadian Centre for Climate Modelling and Analysis system (CanCM3); $2.8^\circ \times 2.8^\circ$; 60 forecasts	Scinocca et al. (2008)	Land scaling: standard normal deviate scaling. Initialization of Series 2 land states: for the 10 ensemble members for a forecast start date in a given year, selection of soil moisture conditions at that start date from different years of an offline land simulation covering 1979–2007. Atmospheric initialization: 6-hourly assimilation of ERA reanalysis. Runs performed with a coupled atmosphere–ocean model.
COLA GCM v3.2; $1.4^\circ \times 1.4^\circ$; 60 forecasts	Misra et al. (2007)	Land scaling: standard normal deviate scaling. Initialization of Series 2 land states: for the 10 ensemble members for a forecast start date in a given year, use of Series 1 land initializations for that start date from the 10 different years of simulation. Atmospheric initialization: National Centers for Environmental Prediction (NCEP) reanalyses at 12-h intervals prior to forecast start date. Observed SSTs used throughout forecast period.
ECMWF integrated forecast system; $1.1^\circ \times 1.1^\circ$; 60 forecasts	Vitart et al. (2008), Balsamo et al. (2009), Jung et al. (2010), ECMWF (2010)	Land scaling: CDF matching. Initialization of Series 2 land states: for the 10 ensemble members for a forecast start date in a given year, use of Series 1 land initializations for that start date from the 10 different years of simulation. Atmospheric initialization: singular vectors, as in operational seasonal forecasting suite.
ECHAM5–JSBACH system; $1.9^\circ \times 1.9^\circ$; 60 forecasts	Roeckner et al. (2003), Raddatz et al. (2007)	Land scaling: standard normal deviate scaling, using GSWP2 multimodel output. Initialization of Series 2 land states: extraction of soil moisture states from different years (but at correct time of year) of free-running AMIP runs. Atmospheric initialization: atmospheric states from free-running AMIP runs.
FSU–COAPS model; $1.9^\circ \times 1.9^\circ$; 60 forecasts	Shin et al. (2005), Cocke and LaRow (2000)	Land scaling: not performed because of use of land data assimilation technique in coupled mode, involving use of Sheffield et al. (2006) atmospheric forcing. Initialization of Series 2 land states: for the 10 ensemble members for a forecast start date in a given year, use of Series 1 land initializations for that start date from the 10 different years of simulation. Atmospheric initialization: atmospheric states from NCEP–Department of Energy (DOE) reanalysis 2.
Geophysical Fluid Dynamics Laboratory (GFDL) global atmospheric model; $2^\circ \times 2.5^\circ$; 30 forecasts	GFDL Global Atmospheric Model Development Team (2004), Delworth et al. (2006)	Land scaling: standard normal deviate scaling. Initialization of Series 2 land states: use of SST-consistent states from parallel AMIP runs. Atmospheric initialization: simple version of NCEP’s iterative breeder mode approach.
NASA GMAO seasonal forecast system [pre–Goddard Earth Observing System (GEOS-5) version]; $2^\circ \times 2.5^\circ$; 60 forecasts	Bacmeister et al. (2000)	Land scaling: standard normal deviate scaling. Initialization of Series 2 land states: use of SST-consistent states from parallel AMIP runs. Atmospheric initialization: atmospheric states from parallel free-running AMIP runs. [One ensemble member in a given forecast used an atmospheric restart (and land restart, for Series 2) from a different year.]
NASA GMAO GEOS-5 system; $2^\circ \times 2.5^\circ$; 60 forecasts	Rienecker et al. (2011)	Land scaling: standard normal deviate scaling. Initialization of Series 2 land states: for the 10 ensemble members for a forecast start date in a given year, use of Series 1 land initializations for that start date from the 10 different years of simulation. Atmospheric initialization: atmospheric states from different years of an atmospheric reanalysis.
National Center for Atmospheric Research (NCAR) Community Atmospheric Model 3.0; $2.8^\circ \times 2.8^\circ$; 60 forecasts	Collins et al. (2006)	Land scaling: standard normal deviate scaling. Initialization of Series 2 land states: use of SST-consistent states from parallel AMIP runs. Atmospheric initialization: extracted from NCEP II reanalysis states spaced 6 h apart in neighborhood of forecast start time.

TABLE 1. (Continued)

System name, resolution, and No. of JJA forecasts submitted	References	Notes
NCAR Community Atmospheric Model 3.5–Community Land Model 3.5; $1.4^\circ \times 1.4^\circ$; 60 forecasts	Neale et al. (2008), Oleson et al. (2008)	Land scaling: standard normal deviate scaling. Initialization of Series 2 land states: for the 10 ensemble members for a forecast start date in a given year, use of Series 1 land initializations for that start date from the 10 different years of simulation. Atmospheric initialization: NCEP reanalyses at 12-h intervals prior to forecast start date.
NCEP Global Forecast System (GFS–Noah); $0.9^\circ \times 0.9^\circ$; 60 forecasts	Moorthi et al. (2001), Ek et al. (2003)	Land scaling: standard normal deviate scaling. Initialization of Series 2 land states: for the 10 ensemble members for a forecast start date in a given year, use of Series 1 land initializations for that start date from the 10 different years of simulation. Atmospheric initialization: standard NCEP “ensemble transfer” method.

United States, while Van den Hurk et al. (2011) present results for Europe. Here, we extend those results, as much as possible, to the globe. The skill metric is computed by first standardizing each model’s set of forecasts (using the model-dependent means and standard deviations for the time of year and lead in question) and then plotting the standardized forecasts for all models against the corresponding observations on the same scatterplot. The standardization removes model-dependent biases in mean and variance from each set of forecasts, thereby allowing them to be considered together. With ten models providing 60 forecasts each for the June–August (JJA) period and one providing 30 forecasts, a total of 630 forecasts are plotted against the 60 observations—a substantial number, allowing for robust statistics.

Of course, the different models may not be strictly independent; this is accounted for in the calculation of significance levels. Significance levels are computed via a brute-force Monte Carlo approach in which the time series of the combined (all models included) forecasts are compared to multiple reshufflings of the time series of standardized observational anomalies. At a given location, for a given variable and lead, the number of shuffles (out of a suitably large sampling) that produce a skill level (r^2 for Series 1 – r^2 for Series 2) at or above a certain value provides an estimate for the confidence with which the null hypothesis (zero land-related skill) is rejected at that value.

The land contributions to precipitation and air temperature skill levels are plotted in Fig. 2 for the three

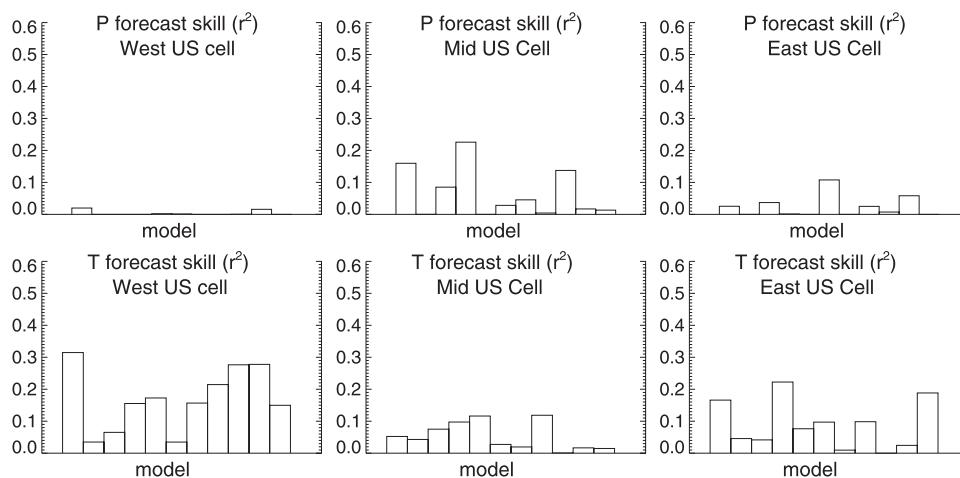


FIG. 1. Skill levels for the Series 1 experiment (r_{ser1}^2) for (top) precipitation and (bottom) air temperature forecasts at days 16–30 at 3 representative grid cells: (left to right) a western North American cell (40°N , 110°W), a central North American cell (40°N , 95°W), and an eastern North American cell (40°N , 80°W). Each histogram bar represents a different model, ordered randomly. (Each histogram uses the same order of models.)

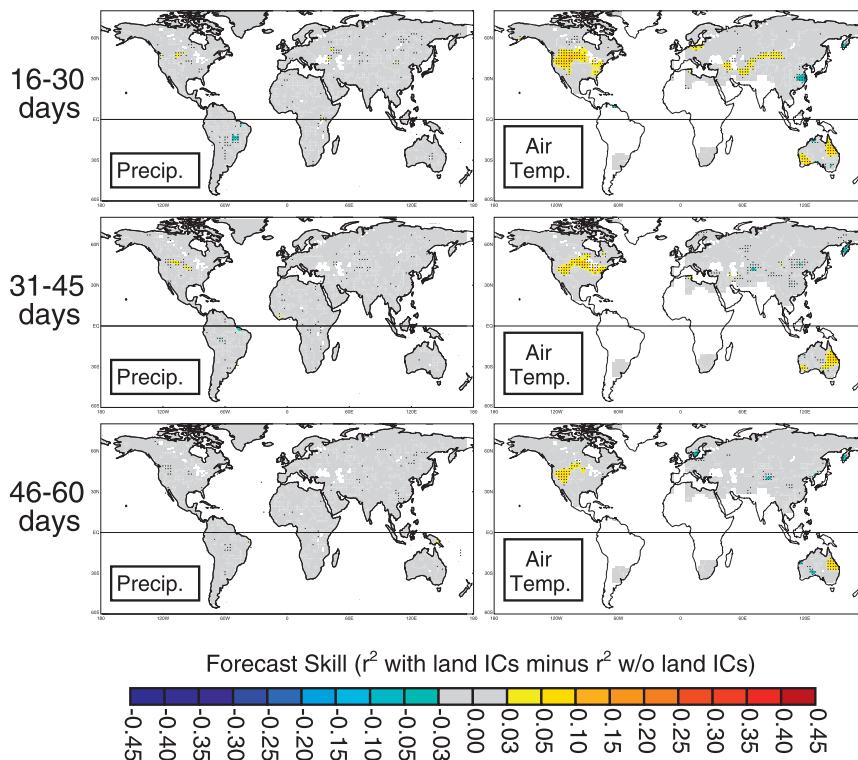


FIG. 2. Consensus (left) precipitation and (right) air temperature forecast skill (r^2 against observations for Series 1 minus that for Series 2) as a function of lead, considering (top to bottom) all 15-day forecast periods during JJA. (See text for details.) Dots are shown where the plotted results are statistically different from 0 at the 99% confidence level; white areas lack available validation data.

different averaging periods: the 15-day lead (days 16–30), 30-day lead (days 31–45), and 45-day lead (days 46–60). Areas for which validation data are unavailable are whitened out; dots in the figure indicate where the values are significantly different from zero at the 99% level. The salient result from this figure is that on the global scale, precipitation forecast skill is close to zero everywhere, and air temperature skill is limited to North America, eastern Australia, and parts of Europe. (Of course, we cannot say if air temperature forecast skill exists in South America and Africa, given the lack of verification data.)

Note that the precipitation skill values over North America differ slightly from those documented in Koster et al. (2010) because of the use of a different validation dataset (a global one that does not rely on as many stations for the United States) and a slightly larger set of forecast contributions. The patterns over the continental United States in Fig. 2 also differ somewhat in structure from those produced in the original GLACE experiment (Koster et al. 2004b), which showed the highest potential for land–atmosphere feedback along a north–south swath down the center of the country with a local maximum near the Texas coast. This latter

discrepancy may result from several factors. First, the original GLACE examined the ability of imposed soil moisture variations to influence the atmosphere, whereas GLACE-2 examines the full prediction problem, which also involves the ability of a model to retain an initial soil moisture anomaly through a forecast period. Seneviratne et al. (2006) show that the south-central United States indeed has a reduced soil moisture memory relative to the north-central United States, which presumably hinders the generation of skill there. Sampling error may also be a factor, as may the fact that the original GLACE experiment was purely synthetic—the patterns it produced necessarily reflect the biased climatologies of the models. The patterns produced in the nonsynthetic GLACE-2 experiment are controlled in large part by the character of land–atmosphere coupling in nature and the ability of the models to reproduce this character.

d. Skill levels in the context of predictability and observational network quality

The skill levels shown in Fig. 2 can be evaluated in the context of two necessary (though not sufficient) conditions

for skill: (i) an underlying “predictability” in the models (i.e., an ability of the models to extract a signal, right or wrong, from a forecast in the presence of chaotic noise), and (ii) some level of quality in the models’ land initialization. Here we examine the joint impact of these two requirements.

Predictability (sometimes called “potential predictability” in the literature) refers here to the degree to which the model’s initialization affects the fields forecasted by the model and, in particular, how this impact diminishes with time. Given enough time, the information content of the initial conditions is overwhelmed by chaotic atmospheric dynamics, at which point the ensemble members of a forecast differ from each other as much they do from any randomly chosen state of the atmosphere. At this point, the initial conditions are unable to provide skill to the forecast.

We provide for this discussion a model-consensus estimate of predictability. To compute it, the standardized precipitation (or air temperature) anomaly produced by each model with its first ensemble member is averaged across the models to produce a single synthetic “truth” for a given start date and lead. The corresponding individual model “forecasts” to be compared with this truth are the averages for each model over its remaining nine ensemble members for that start date and lead. The 630 synthetic “truth-forecast” pairs are plotted against each other in a manner equivalent to that used for the consensus skill calculations in section 4c above. The square of the correlation coefficient between the points (r_{ideal}^2) represents the multimodel-consensus estimate of predictability; to isolate the land contribution to this predictability from the contributions of SSTs, atmospheric initialization, and other prescribed boundary conditions, we subtract the r_{ideal}^2 produced for Series 2 from that produced by Series 1.

It is important to emphasize here that this estimate does not represent the intrinsic level of predictability in nature—it does not provide information on how much extra skill could be extracted from forecast systems if model formulations or initialization data were improved. Intrinsic levels of predictability in nature cannot be measured or even estimated with any confidence; the only accessible quantitative measure of nature’s predictability involves its lower bound, as determined from forecast verification studies. The predictability estimates presented here are specific to our definition of consensus skill and are useful for one purpose only: to quantify the upper limit of skill we can achieve given the character of our calculation and the current structures of the forecast models. That is, prior to looking at any forecast verification data, these predictability estimates indicate where consensus skill, given the way we have defined it

and given the suite of forecast models used, is possible. Note that predictability levels for the individual GLACE-2 models (not shown) differ substantially, with some being larger than the consensus values; removal or addition of additional models to the study could thus modify the patterns shown.

With this in mind, Fig. 3 shows, for the 16–30-, 31–45-, and 46–60-day leads, the global distributions of the r_{ideal}^2 differences for precipitation and air temperature. As expected, land contributions to precipitation predictability (left panels) decrease with lead. Note, however, that predictability values even for short leads are small, suggesting a low, though nonzero, potential for consensus forecast skill given the model frameworks used. The consensus predictability for air temperature is much larger, as shown in the right panels, with r_{ideal}^2 differences exceeding 0.25 in many places. Even so, hope for skillful consensus air temperature forecasts is highly limited in many parts of the world (e.g., Asia and Europe after day 30). Simply put, Fig. 3 shows that the low skill values seen in many parts of the world in Fig. 2 result from the character of predictability within the participating models—it would be impossible, with this set of models and this definition of consensus skill, to generate skill levels higher than those in Fig. 3, even if the observational networks used for the land initialization were greatly improved.

Of course, imperfections in land initialization do limit skill in regions having significant predictability. For this study, we key the realism of the soil moisture initialization to the quality of the meteorological forcing used to produce it. We focus in particular on the monthly precipitation estimates used, making the assumption that soil water initialization is most strongly controlled by the temporally averaged rainfall amounts contributing to it, more so than by other meteorological fields or by the high-frequency timing of the rainfall, which in these experiments are derived from reanalysis. We further assume that the quality of the monthly precipitation data (for the 1986–95 period) is best represented by the areal density of the rain gauges used to estimate it. The link between rain gauge density and errors in precipitation estimates has been reported in various studies (e.g., Zawadzki 1973; Seed and Austin 1990; Gebremichael et al. 2003; Huang et al. 2008), and the corresponding impact of rain gauge density on the performance of hydrological models was documented by Oki et al. (1999). The impact of density on forecast skill, however, has not, until now, been documented. Here, in analogy to the hydrological study of Oki et al. (1999), we examine the idea that rain gauge density affects forecast skill through its impact on the realism of the soil moisture initialization.

Figure 4, a variant of a figure from Zhao and Dirmeyer (2003) and Decharme and Douville (2006), shows the

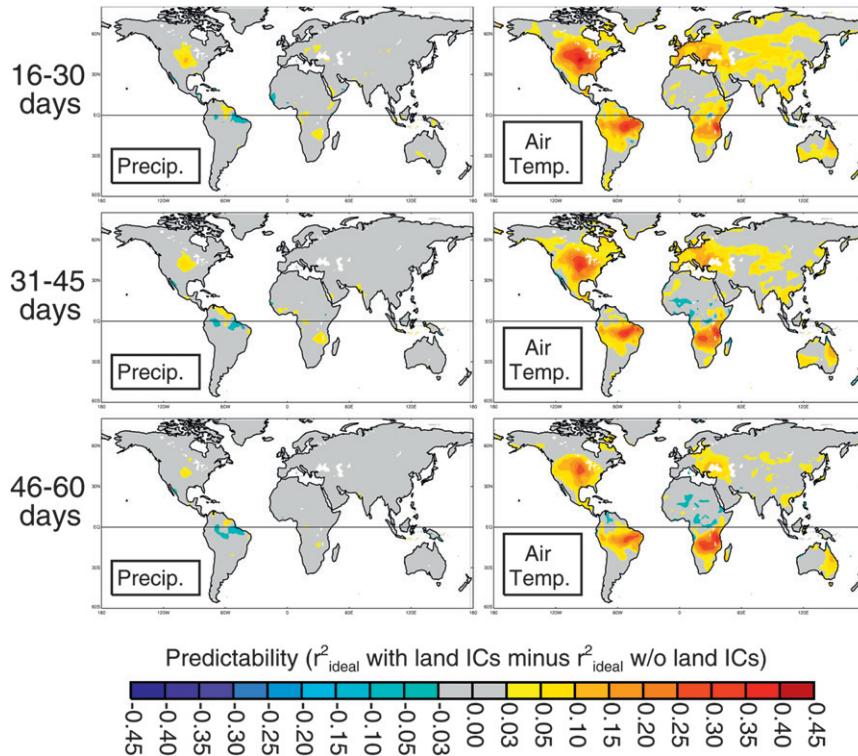


FIG. 3. Multimodel-consensus estimate of (left) precipitation and (right) air temperature predictability associated with soil moisture initialization—in essence a quantification of how one ensemble member in a given forecast reproduces the synthetic truth produced by the remaining ensemble members in that forecast: (top to bottom) all 15-day forecast periods.

average number of monthly rain gauges per $2^{\circ} \times 2.5^{\circ}$ grid cell contributing to the Global Precipitation Climatology Project (GPCP) during 1986–95 (D. Bolvin 2010, personal communication), the dataset underlying the GSWP-2 rainfall time series. Rainfall measurement stations are dense (and thus soil moisture initialization is likely to be reasonably accurate) in, for example, central North America, Europe, eastern South America, and the east coast of Australia. Measurement stations are sparser, and thus the soil moisture initialization is more questionable in most of the remainder of South America and Australia and in most of Africa and Asia. Note that while satellite information was used to “fill in the gaps” of the station record when producing the precipitation datasets that contributed to GSWP-2, satellite-based rainfall data are notoriously questionable, particularly over land for the 1986–95 period. Indeed, the fact that satellite data are included in the precipitation forecast verification dataset must be kept in mind when considering model performance for precipitation in areas of low rain gauge density.

Figure 5 provides, for air temperature forecasts, a useful joint look at the limitations imposed by underlying model predictability and by the accuracy of the initialization, as represented by gauge density. Within each

scatterplot (one panel for each forecast lead), each plotted dot corresponds to a $2^{\circ} \times 2.5^{\circ}$ grid cell that is at least 90% land. The abscissa assigned to a given dot is determined from its underlying model-consensus predictability (from Fig. 3), and the ordinate assigned to the dot is determined from the local gauge density (from Fig. 4). The size and color of the dots are keyed to the land-derived forecast skill uncovered by the GLACE-2 experiment.

The salient result from Fig. 5 is the strong impact of both underlying model predictability and gauge density on the ability to extract true skill from the experiments. As expected, skill appears only where the background predictability is sufficiently large. For the most part, skill also appears only for higher values of gauge density—the plots suggest that a density of about 10 gauges per $2^{\circ} \times 2.5^{\circ}$ cell is needed to produce the larger values of skill in the forecasts. The results for the 30-day lead (days 31–45) in particular imply a strong impact of gauge density on skill, though a similar impact is also seen for the 15-day lead results. [Similar suggestions of gauge density impact are also seen in corresponding scatterplots for precipitation (not shown), though the signal is weaker.] We caution that such plots show correlation rather than causation; they do not prove that higher densities are

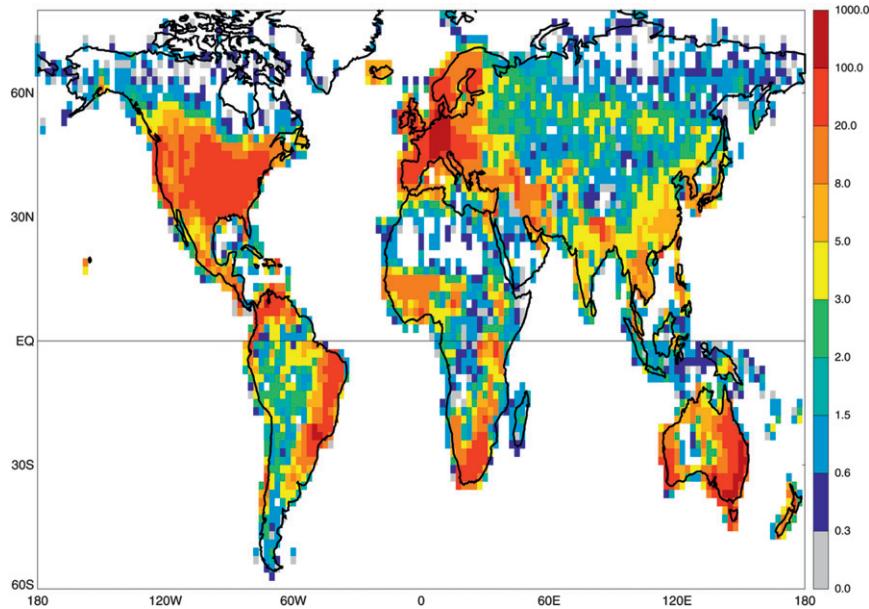


FIG. 4. Average number of rain gauges contributing to the GSWP-2 forcing data during the period 1986–95 (sum over all months divided by 120). The original GPCP rain gauge density data have a resolution of $2.5^\circ \times 2.5^\circ$; these data were regridded and shown here at a resolution of $2^\circ \times 2.5^\circ$ for consistency with the resolution used for the joint model analysis.

responsible for higher forecast skill. Furthermore, the quality of temperature verification data is probably poor in many locations with low rain gauge density, which further hampers skill scores, though presumably air temperature measurements, characterized by larger spatial correlations than those for precipitation, are relatively robust. In any case, Fig. 5 suggests that if precipitation (and perhaps air temperature, for verification) were monitored more comprehensively during the 1986–95 period, additional air temperature forecast skill might indeed have been computed in GLACE-2 at many of the grid cells for which the land-derived predictability is high but the gauge density is low. This result underscores the importance for the forecasting problem of observational networks, both in terms of rain gauge networks and, for present-day, satellite-based observations of precipitation and soil moisture.

Of course, improved observations will not help with air temperature forecasts everywhere. Many regions with poor networks are located either in extremely dry or wet climates; such areas, typically characterized by low land–atmosphere feedback (and thus low amounts of land-derived predictability), are not expected in any case to produce high land-derived skill.

e. Conditional consensus skill levels

In the initial documentation of GLACE-2 results, Koster et al. (2010) utilized the concept of conditional skill—the calculation of skill levels for a subset of the forecast start dates, a subset defined by the state of the

initial soil moisture. The idea parallels that used in coupled atmosphere–ocean seasonal forecasting; forecasters know that certain weather patterns are more predictable when the El Niño–Southern Oscillation sea surface temperature indices are strongly positive or negative rather than close to neutral. Here, we hypothesize that soil moisture impacts on skill will be larger and thus easier to diagnose when the initial soil moisture anomalies are particularly large.

Figure 6, in analogy to Fig. 2, shows the land-derived skill levels obtained for precipitation and air temperature at the three leads; these skill levels, however, were computed from only 40% of the forecast periods. The subset was chosen as follows. At each $2^\circ \times 2.5^\circ$ grid cell, a multimodel time series of root zone soil moisture was computed from a $1^\circ \times 1^\circ$ soil moisture analysis using the outputs of several land models participating in GSWP-2. Most of the models underlying this analysis are independent of those used in GLACE-2; see Koster et al. (2009a) for details on the models used. For the present study, the soil moisture time series for each GSWP-2 model was standardized at the $1^\circ \times 1^\circ$ resolution to remove seasonality in mean and variance before being averaged with the standardized results of the other GSWP-2 models and over a large area ($6^\circ \times 7.5^\circ$) centered on the $2^\circ \times 2.5^\circ$ grid cell being examined. For each 15-day period within JJA, and for each assumed lead, the 60 corresponding “initial” large-scale soil moisture anomalies (10 years times 6 start dates per year) extracted from

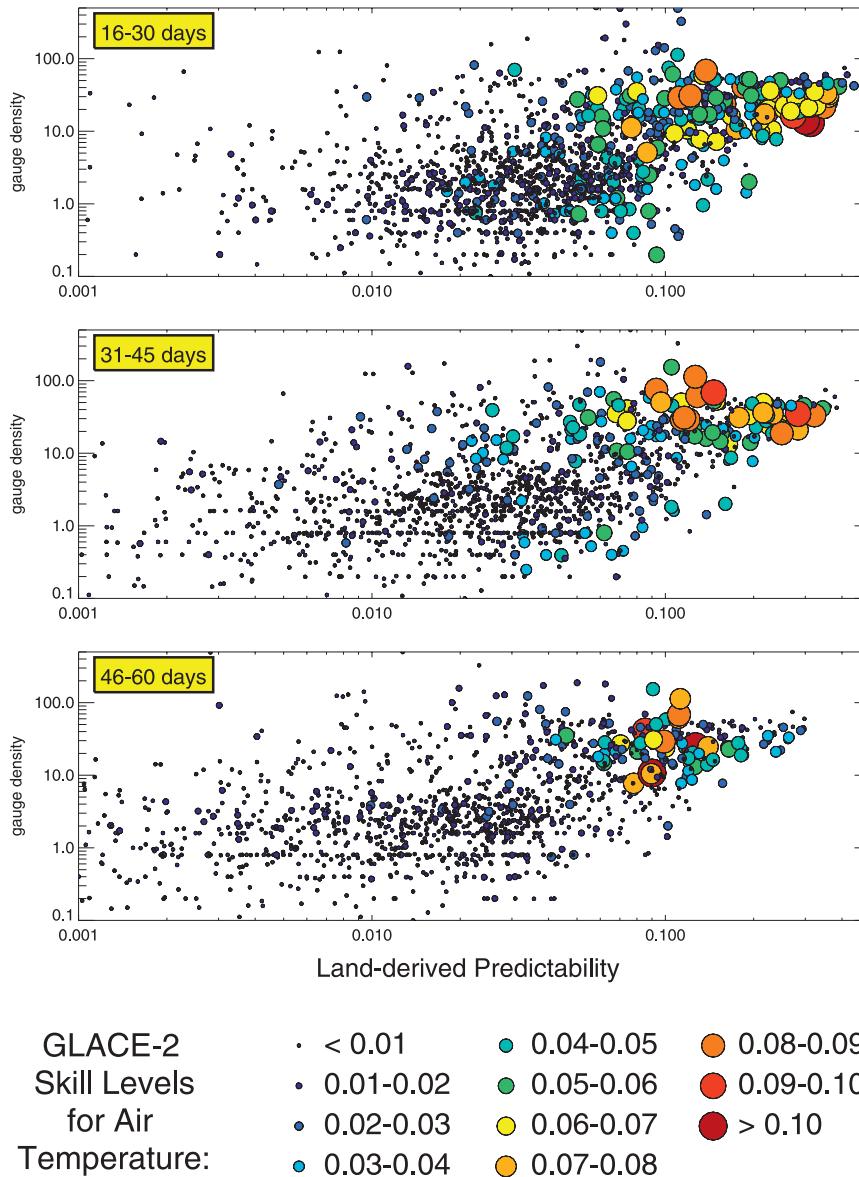


FIG. 5. (top) Air temperature forecast skill at 16–30 days as a function of background consensus predictability (x axis) and gauge density (y axis). Each dot corresponds to a $2^\circ \times 2.5^\circ$ land grid cell; the size and color of the dot (see legend) reflect the land-derived air temperature forecast skill computed for the location by GLACE-2. (middle) Same, but for 31–45-day forecasts. (bottom) Same, but for 46–60-day forecasts.

this time series were ranked, and the driest and wettest fifths (quintiles) of the soil moisture anomalies were established. The skill metric shown in Fig. 6 is for this subset of start dates—the subset of forecasts initialized with these very dry or very wet anomalies. Note that while the subset of start dates used for the calculation is different for each grid cell, spatial correlation in soil moisture patterns may lead to similar subsets for neighboring cells.

Figure 6 shows a slight increase in precipitation skill in North America for the extreme quintiles, in agreement

with results already shown in Koster et al. (2010). Precipitation skill results for the rest of the globe, however, appear unimpressive, particularly when considering field significance, which reflects the degree to which skill levels in a map exceed those that would appear by chance. The combined area of the small yellow (positive skill) patches appears to match roughly that of the blue (negative skill) patches, implying that neither are significant. Of course, from Fig. 4, the number of places where skill might be realized in the first place, given the underlying

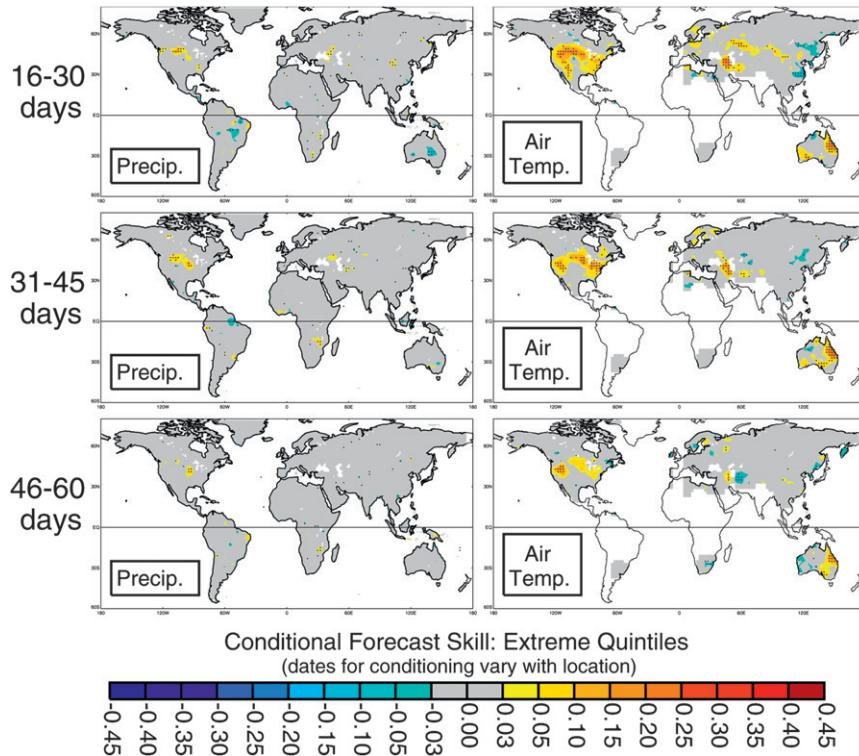


FIG. 6. (left) Precipitation and (right) air temperature forecast skill (r^2 against observations for Series 1 minus that for Series 2) as a function of lead for (top to bottom) a 40% subset of the 15-day forecast periods during JJA: those periods for which the local initial soil moisture content is in the top fifth or the bottom fifth of all realized values. (See text for details.) Dots are shown where the plotted results are statistically different from zero at the 99% confidence level; white areas lack available validation data.

observational network, is limited. For air temperature, the right panels in Fig. 6 show forecast skill levels that significantly exceed those in Fig. 2; the contribution of land initialization to skill is larger when only the driest and wettest quintiles are considered. Significant positive skill patches are seen in North America and in patches scattered across Europe, Asia, and Australia. Again, the skill shows up where the observational rain gauge network is relatively dense (Fig. 4).

A global average of the land-derived skill levels is telling. Here, by “global average,” we mean the average over all grid cells that are at least 90% land and have a rain gauge density of at least 10 gauges per $2^\circ \times 2.5^\circ$ grid cell, given considerations raised in section 4d above. The global averages for a number of different temporal subsettings are shown in Fig. 7. The x axis shows the fraction of start dates considered; the global averages from Fig. 6 are thus represented by the fraction 0.4, and those from Fig. 2 are represented by the fraction 1. Average skill values are also shown for the analysis of extreme terciles, quartiles, deciles, and twentieths (0.667, 0.5, 0.2, and 0.1 on the x axis, respectively). The figure

shows that as the initial conditions considered become more extreme, the globally averaged air temperature forecast skill increases. A much smaller effect is seen for precipitation forecasts. For air temperature, all globally averaged land-derived skill levels are significantly different from 0 at the 99% confidence level or higher.

What does this mean in practical terms? At the start of a given forecast, a forecaster would know whether or not the initialized soil moisture anomaly at a given location is at the dry or wet end of its distribution. If the initial anomaly is not particularly extreme, the forecaster may expect little help from the realistic soil moisture initialization. A more extreme initialized anomaly should lead to more land-derived skill and hence, to the extent that this soil moisture impact is not overshadowed by other complicating factors, a more confident forecast.

f. Asymmetric contributions of wet and dry initialization

GLACE-2 can begin to address the question of wet-dry asymmetry in prediction—to what degree can anomalously wet conditions contribute more to a forecast

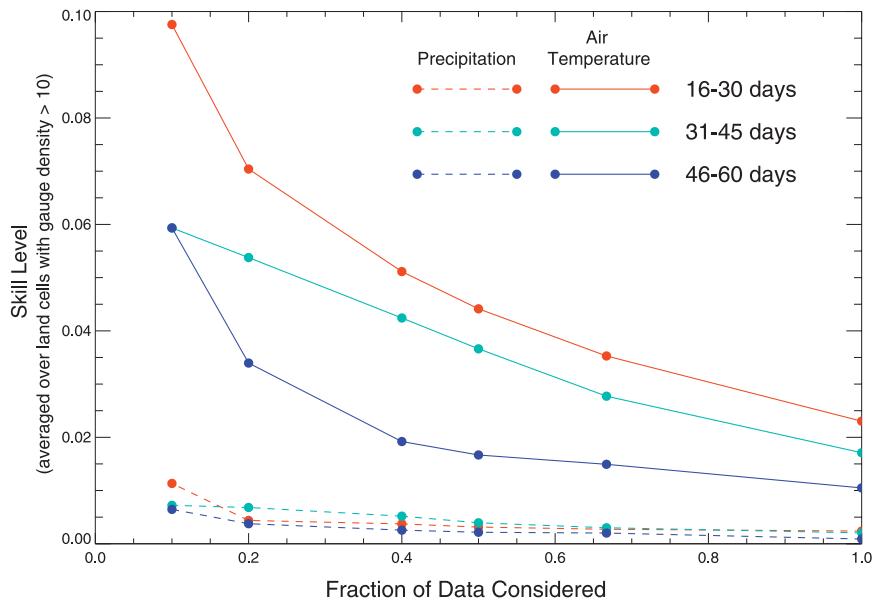


FIG. 7. Precipitation and air temperature forecast skill (r^2 against observations for Series 1 minus that for Series 2) averaged over all land grid cells for which validation data are available and for which the rain gauge density exceeds 10 gauges per $2^\circ \times 2.5^\circ$ grid cell, as a function of lead and the fraction of forecast periods considered. A fraction of 0.1 refers to the examination of the driest twentieth and wettest twentieth of forecast start dates, a fraction of 0.2 refers to the examination of the driest tenth and wettest tenth of forecast start dates, and so on. (See text for details.)

than anomalously dry conditions, or vice versa? The land-derived skill levels shown in Fig. 2 for air temperature prediction at the 30-day lead (days 31–45) in the continental United States are shown again in the top panel of Fig. 8. The middle panel of Fig. 8 shows the skill levels obtained for exactly half of the start dates: those for which the (standardized) initial soil moisture at the local grid cell lies in the lower half of all the values realized—that is, the times for which the local soil is initialized anomalously dry, in a median sense. [As before, the multimodel soil moisture products generated independently by Koster et al. (2009a) are used to partition the start dates.] The bottom panel of Fig. 8 shows the corresponding skill for the wettest start dates. The distinction between the middle and lower maps is clear: dry initial conditions lead to greater skill in the north-central United States, whereas wet initial conditions lead to more skill toward the Southwest.

This distinction, while not conclusive, is nevertheless consistent with basic hydrologic concepts regarding wet–dry asymmetry. Consider Fig. 9, which shows, in a highly idealized way, how evaporative fraction (the ratio of evaporation to net radiation) varies with soil moisture variations. At the dry end, evaporation is sensitive to soil moisture—soil moisture availability limits the evaporation rate. At the wet end, on the other hand, soil moisture availability no longer acts as a bottleneck

limiting evaporation; evaporation is thus insensitive to soil moisture variations, being instead sensitive to variations in “atmospheric demand,” or energy availability.

Consider first a climate for which the average soil moisture lies at the value A (top panel). Under the assumption that soil moisture influences the atmosphere through its impact on the surface energy budget, a dry anomaly at this location has more of a chance to contribute to a forecast than a wet anomaly because the soil moisture would influence the atmosphere only in the dry case. In other words, at this location, the structure of the idealized relationship in Fig. 9 suggests a potential wet–dry asymmetry in land-derived forecast skill. Consider next a location for which the average soil moisture lies at the value B. Soil moisture is, on average, very dry here—already near its lower limit—and evaporation is very low. Soil moisture cannot get too much lower during drier-than-average years, and its impact on evaporation and the overlying atmosphere is correspondingly limited for these years. During wet years, however, soil moisture can enter a regime where it can have a significant impact on evaporation. Thus we get the opposite wet–dry asymmetry in forecast skill: for this location, a relatively wet initialization is more likely to positively contribute to skill. Such asymmetric effects of soil moisture on temperature have been seen in experiments with a regional climate model in France (Jaeger

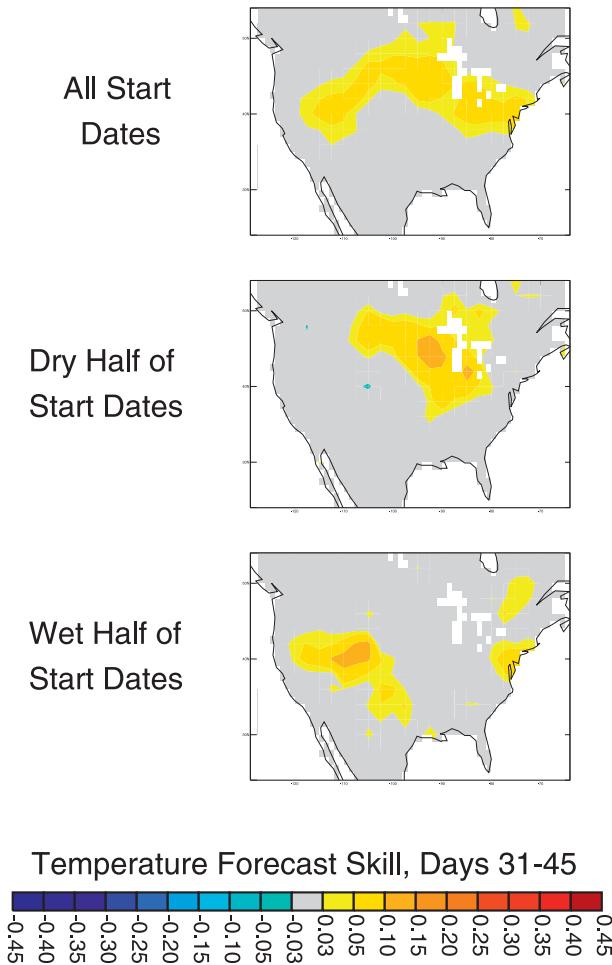


FIG. 8. Air temperature forecast skill (r^2 against observations for Series 1 minus that for Series 2) for the 30-day lead (days 31–45). (top) All start dates. (middle) Start dates for which the local initial soil moisture lies in the driest half of all values realized there. (bottom) Start dates for which the local initial soil moisture lies in the wettest half of all values realized there.

and Seneviratne 2011), for which prescribed anomalous soil moisture conditions mostly affect temperatures at the dry end, leading to longer tails of the temperature distributions for dry extremes.

The GLACE-2 results would support this interpretation if the high skill areas in the middle and bottom panels of Fig. 8 tended to be characterized by “A-type” and “B-type” points, respectively, in Fig. 9. This is certainly possible given the known dryness gradient across the continent. To demonstrate this more precisely, Fig. 10 shows maps of the correlation coefficient between 15-day evaporative fraction (EF; the ratio of latent heat flux to net radiative energy) and 15-day soil moisture (for the 30-day lead) for the dry and wet halves of the start dates (calculated locally), along with the differences (dry minus wet) in these correlations.

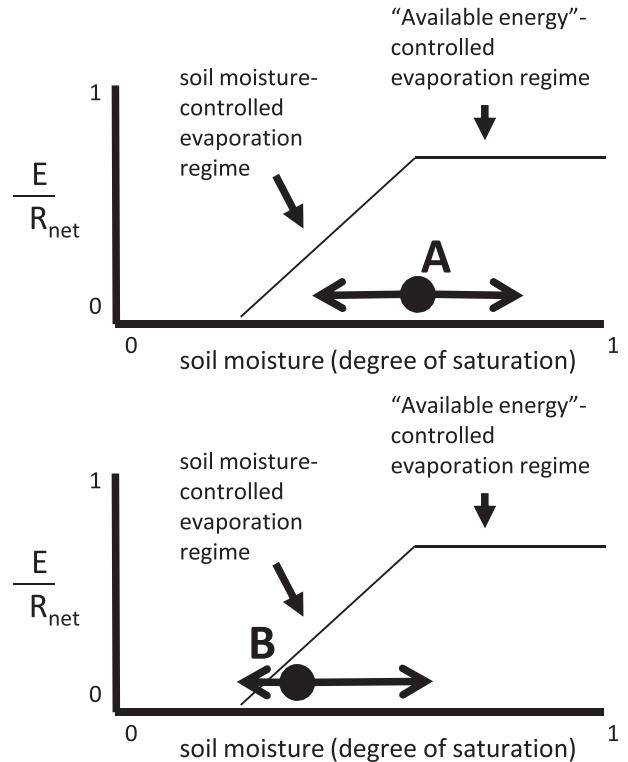


FIG. 9. Idealized breakdown of the soil moisture–evaporation relationship into two parts: a drier regime in which evaporative fraction (the ratio of evaporation to net radiation) is sensitive to soil moisture, and a wetter regime in which it is insensitive. (bottom) The location B has a mean soil moisture at the dry end, close to the dry limit; drier-than-average years cannot get too dry, whereas wetter-than-average years could get quite wet. (top) The wetter location A straddles the two regimes, with only the drier-than-average years showing sensitivity of evaporation to soil moisture.

The data represent averages over the nine GLACE-2 models that provided usable soil moisture information. For the region just southwest of the Great Lakes, the correlation is much higher for the dry case, just as the air temperature forecast skill is (Fig. 8), supporting the interpretation provided by the top panel in Fig. 9. The results for the wet case, while not inconsistent with the interpretation in Fig. 9, are not quite as clear; the southwest does show higher correlation between soil moisture and EF for the wetter half of start dates, but these locations do not precisely coincide with those in the lower panel of Fig. 8, and Fig. 10 does not fully explain the small patch of skill for the wet case in the mid-Atlantic area.

Results for precipitation (not shown) are similar. The land-derived skill levels are much smaller than those in Fig. 8 for both the dry and wet subsets, but the maximum for the dry case does appear to the southwest of the Great Lakes, roughly consistent with Fig. 10.

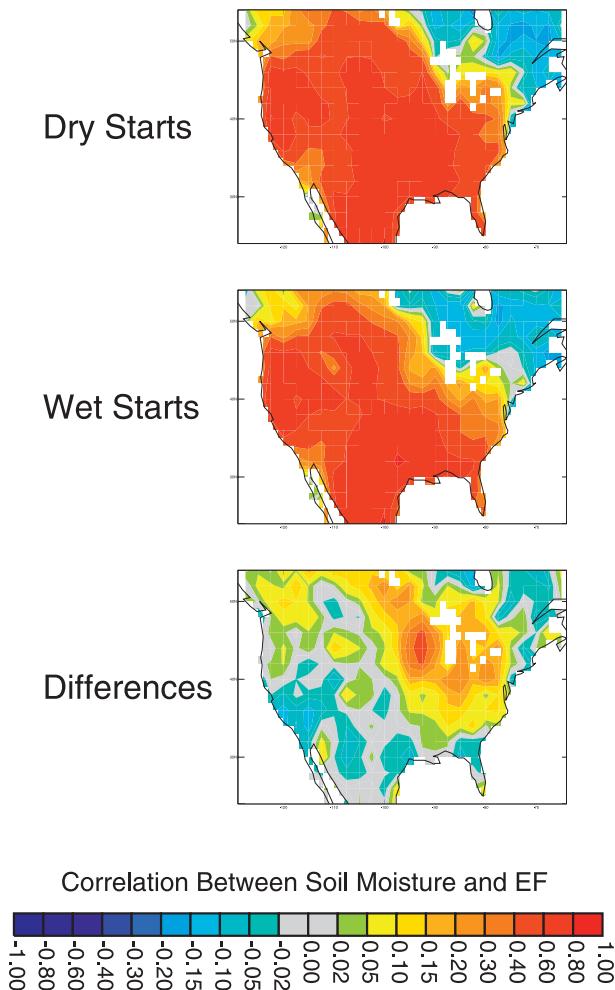


FIG. 10. (top) Map of the correlation between soil moisture and evaporative fraction for the drier half of start dates, as determined at each grid cell independently. (middle) Same, but for the wetter half of start dates. (bottom) Differences (dry – wet).

5. Conclusions

GLACE-2 provides a common experimental framework to quantify the impact of realistic soil moisture initialization on precipitation and air temperature forecast skill at subseasonal leads. Eleven modeling systems participated in the experiment, allowing the determination of a consensus view of this impact. In agreement with the North American results of Koster et al. (2010), but here extended to the globe, we find significant impacts of soil moisture on air temperature forecast skill in many regions. This skill is limited, however, to areas for which the underlying observational precipitation network provides adequate coverage and thus trustworthy initialization. The results in Fig. 5 indeed underscore the importance of observational networks for the forecast problem, suggesting that the real-time provision of

improved observations for initialization—as derived in modern times from both improved rainfall estimates and satellite-derived soil moisture estimates—would likely lead to increased forecast skill in some regions. Skill levels for precipitation are much weaker, but global averaging does show that skill levels for air temperature, and to some small degree for precipitation, increase as the start dates are subsetted toward more extreme initial conditions—larger soil moisture anomalies have a stronger impact on skill. Wet and dry initializations tend to generate skill in different areas, with dry initialization providing more skill at the transition between soil-moisture-and energy-availability-controlled evaporation.

One benefit of the GLACE-2 experiment is the identification of areas for which enhanced soil moisture (or antecedent precipitation) measurement may prove especially fruitful, according to the consensus of model behavior. The experimental framework is also suitable for additional sensitivity tests addressing various facets of the prediction problem, such as the importance of scaling the land variables (section 2b) prior to forecast initialization. Perhaps the framework's greatest value, however, is the provision of an objective means by which any forecasting group can evaluate, for its own model in isolation, the practical benefit of realistic land initialization—particularly if relevant additional factors associated with initialization in an operational setting (real-time data availability, etc.) are incorporated into the experimental design. Because the GLACE-2 forecasts are compared to actual observations, any skill levels generated by a given forecast system with such an operational version of GLACE-2 can be interpreted as “lower bounds” for what could be achieved with that system using improved model parameterizations and more complete measurement networks.

Acknowledgments. Coordinators of the GLACE-2 project gratefully acknowledge financial support from NOAA's Climate Prediction Program for the Americas and NASA's Terrestrial Hydrology program. The various participants in the project (see Table 1) were able to perform the GLACE-2 simulations thanks to financial and computational support from their home institutions and/or from the institutions hosting the numerical prediction systems. We thank George Huffmann and David Bolvin for valuable discussions on the precipitation gauge network, and we thank WCRP's GEWEX and CLIVAR projects for their sponsorship of this project. One of the authors (Tony Gordon) passed away during the review process for this manuscript. His contributions to the project were particularly significant, and we dedicate this paper to his memory.

REFERENCES

- 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, 194 pp.
- Balsamo, G., P. Viterbo, A. Beljaars, B. J. J. M. van den Hurk, M. Hirschi, A. K. Betts, and K. Scipal, 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *J. Hydrometeorol.*, **10**, 623–643.
- Berg, A. A., J. S. Famiglietti, M. Rodell, U. Jambor, S. L. Holl, R. H. Reichle, and P. R. Houser, 2005: Development of a hydrometeorological forcing data set for global soil moisture estimation. *Int. J. Climatol.*, **25**, 1697–1714.
- Betts, A. K., and J. H. Ball, 1995: The FIFE surface diurnal cycle climate. *J. Geophys. Res.*, **100**, 25 679–25 693.
- , —, A. C. M. Beljaars, M. J. Miller, and P. Viterbo, 1994: Coupling between land-surface, boundary-layer parameterizations and rainfall on local and regional scales: Lessons from the wet summer of 1993. Preprints, *Fifth Symp. on Global Change Studies*, Nashville, TN, Amer. Meteor. Soc., 174–181.
- Caesar, J., L. Alexander, and R. Vose, 2006: Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *J. Geophys. Res.*, **111**, D05101, doi:10.1029/2005JD006280.
- Cocke, S., and T. E. LaRow, 2000: Seasonal predictions using a regional spectral model embedded within a coupled ocean-atmosphere model. *Mon. Wea. Rev.*, **128**, 689–708.
- Collins, W. D., and Coauthors, 2006: The formulation and atmospheric simulation of the Community Atmosphere Model Version 3 (CAM3). *J. Climate*, **19**, 2144–2161.
- Cook, B. L., G. B. Bonan, and S. Levis, 2006: Soil moisture feedbacks to precipitation in southern Africa. *J. Climate*, **19**, 4198–4206.
- Decharme, B., and H. Douville, 2006: Uncertainties in the GSWP-2 precipitation forcing and their impacts on regional and global hydrological simulations. *Climate Dyn.*, **27**, 695–713.
- Delworth, T. L., and S. Manabe, 1989: The influence of soil wetness on near-surface atmospheric variability. *J. Climate*, **2**, 1447–1462.
- , and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674.
- Dirmeyer, P. A., 2000: Using a global soil wetness dataset to improve seasonal climate simulation. *J. Climate*, **13**, 2900–2922.
- , X. Gao, and T. Oki, 2002: *GSWP-2: The Second Global Soil Wetness Project Science and Implementation Plan*. IGPO Publication Series, No. 37, IGPO, 65 pp.
- , —, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki, 2006: GSWP-2: Multimodel analysis and implications for our perception of the land surface. *Bull. Amer. Meteor. Soc.*, **87**, 1381–1397.
- Douville, H., 2002: Influence of soil moisture on the Asian and African monsoons. Part II: Interannual variability. *J. Climate*, **15**, 701–720.
- , 2010: Relative contribution of soil moisture and snow mass to seasonal climate predictability: A pilot study. *Climate Dyn.*, **34**, 797–818, doi:10.1007/s00382-008-0508-1.
- , and F. Chauvin, 2000: Relevance of soil moisture for seasonal climate predictions: A preliminary study. *Climate Dyn.*, **16**, 719–736.
- , —, and H. Broqua, 2001: Influence of soil moisture on the Asian and African monsoons. Part I: Mean monsoon and daily precipitation. *J. Climate*, **14**, 2381–2403.
- ECMWF, cited 2010: IFS documentation CY33r1. [Available online at <http://www.ecmwf.int/research/ifsdocs/CY33r1/index.html>.]
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of the Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851, doi:10.1029/2002JD003296.
- Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow, 2010: Performance metrics for soil moisture retrievals and application requirements. *J. Hydrometeorol.*, **11**, 832–840.
- 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.
- Findell, K. L., and E. A. B. Eltahir, 1997: An analysis of the soil moisture-rainfall feedback, based on direct observations from Illinois. *Water Resour. Res.*, **33**, 725–735.
- , and —, 2003: Atmospheric controls on soil moisture-boundary layer interactions. Part I: Framework development. *J. Hydrometeorol.*, **4**, 552–569.
- Gebremichael, M., W. F. Krajewski, M. Morrissey, D. Langerud, G. J. Huffman, and R. Adler, 2003: Error uncertainty analysis of GPCP monthly rainfall products: A data-based simulation study. *J. Appl. Meteor.*, **42**, 1837–1848.
- GFDL Global Atmospheric Model Development Team, 2004: The new GFDL global atmosphere and land model AM2-LM2: Evaluation with prescribed SST simulations. *J. Climate*, **17**, 4641–4673.
- Guo, Z., and Coauthors, 2006: GLACE: The Global Land-Atmosphere Coupling Experiment. Part II: Analysis. *J. Hydrometeorol.*, **7**, 611–625.
- Hong, S. Y., and E. Kalnay, 2000: Role of sea surface temperature and soil-moisture feedback in the 1998 Oklahoma-Texas drought. *Nature*, **408**, 842–844.
- Huang, J.-C., S.-J. Kao, K.-T. Chang, C.-Y. Lin, and P.-L. Chang, 2008: Cost-effective raingauge deployment and rainfall heterogeneity effect on hydrograph simulation in mountainous watersheds. *Hydrol. Earth Syst. Sci. Discuss.*, **5**, 2169–2197.
- Jaeger, E. B., and S. I. Seneviratne, 2011: Impact of soil moisture-atmosphere coupling on European climate extremes and trends in a regional climate model. *Climate Dyn.*, **36**, 1919–1939.
- Jung, T., and Coauthors, 2010: The ECMWF model climate: Recent progress through improved physical parametrizations. *Quart. J. Roy. Meteor. Soc.*, **136**, 1145–1160, doi:10.1002/qj.634.
- Koster, R. D., and Coauthors, 2004a: Realistic initialization of land surface states: Impacts on subseasonal forecast skill. *J. Hydrometeorol.*, **5**, 1049–1063.
- , and Coauthors, 2004b: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138–1140.
- , and Coauthors, 2006: GLACE: The Global Land-Atmosphere Coupling Experiment. Part I: Overview. *J. Hydrometeorol.*, **7**, 590–610.
- , Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma, 2009a: On the nature of soil moisture in land surface models. *J. Climate*, **22**, 4322–4335.
- , S. D. Schubert, and M. J. Suarez, 2009b: Analyzing the concurrence of meteorological droughts and warm periods, with

- implications for the determination of evaporative regime. *J. Climate*, **22**, 3331–3341.
- , and Coauthors, 2010: Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophys. Res. Lett.*, **37**, L02402, doi:10.1029/2009GL041677.
- Mahanama, S. P. P., R. D. Koster, R. H. Reichle, and M. J. Suarez, 2008: Impact of subsurface temperature variability on surface air temperature variability: An AGCM study. *J. Hydrometeorol.*, **9**, 804–815.
- Misra, V., and Coauthors, 2007: Validating and understanding ENSO simulation in two coupled climate models. *Tellus*, **59A**, 292–308.
- Moorthi, S., H.-L. Pan, and P. Caplan, 2001: Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system. NWS Tech. Procedures Bull. 484, 14 pp.
- Neale, R. B., J. H. Richter, and M. Jochum, 2008: The impact of convection on ENSO: From a delayed oscillator to a series of events. *J. Climate*, **21**, 5904–5924.
- 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.
- Oleson, K. W., and Coauthors, 2008: Improvements to the Community Land Model and their impact on the hydrological cycle. *J. Geophys. Res.*, **113**, G01021, doi:10.1029/2007JG000563.
- Pitman, A. J., and S. E. Perkins, 2009: Global and regional comparison of daily 2-m and 1000-hPa maximum and minimum temperatures in three global reanalyses. *J. Climate*, **22**, 4667–4681.
- Raddatz, T. J., and Coauthors, 2007: Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century? *Climate Dyn.*, **29**, 565–574, doi:10.1007/s00382-007-0247-8.
- Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate*, **7**, 929–948.
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648.
- Roeckner, E., and Coauthors, 2003: The general circulation model ECHAM5. Part I: Model description. Max Planck Institute for Meteorology Rep. 349, 140 pp.
- Scinocca, J. F., N. A. McFarlane, M. Lazare, J. Li, and D. Plummer, 2008: The CCCma third generation AGCM and its extension into the middle atmosphere. *Atmos. Chem. Phys.*, **8**, 7055–7074.
- Seed, A. W., and G. L. Austin, 1990: Sampling errors for raingauge-derived mean areal daily and monthly rainfall. *J. Hydrology*, **118**, 163–173, doi:10.1016/0022-1694(90)90256-W.
- Seneviratne, S. I., and Coauthors, 2006: Soil moisture memory in AGCM simulations: Analysis of Global Land–Atmosphere Coupling Experiment (GLACE) data. *J. Hydrometeorol.*, **7**, 1090–1112.
- , T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010: Investigating soil moisture–climate interactions in a changing climate: A review. *Earth Sci. Rev.*, **99**, 125–161.
- Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**, 3088–3111.
- Shin, D. W., S. Cocke, T. E. LaRow, and J. J. O'Brien, 2005: Seasonal surface air temperature and precipitation in the FSU climate model coupled to the CLM2. *J. Climate*, **18**, 3217–3228.
- Shukla, J., and Y. Mintz, 1982: Influence of land-surface evapotranspiration on the earth's climate. *Science*, **215**, 1498–1501.
- Van den Hurk, B. J. J. M., and E. van Meijgaard, 2010: Diagnosing land–atmosphere interaction from a regional climate model simulation over West Africa. *J. Hydrometeorol.*, **11**, 467–481.
- , F. Doblas-Reyes, G. Balsamo, R. Koster, S. Seneviratne, and H. Camargo, Jr., 2011: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. *Climate Dyn.*, doi:10.1007/s00382-010-0956-2, in press.
- Vinnikov, K. Ya, and I. B. Yeserkepova, 1991: Soil moisture: Empirical data and model results. *J. Climate*, **4**, 66–79.
- Vitart, F., and Coauthors, 2008: The new VarEPS-monthly forecasting system: A first step towards seamless prediction. *Quart. J. Roy. Meteor. Soc.*, **134**, 1789–1799.
- 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.
- Xie, P., J. E. Janowiak, P. A. Arkin, R. Adler, A. Gruber, R. Ferraro, G. J. Huffman, and S. Curtis, 2003: GPCP pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. *J. Climate*, **16**, 2197–2214.
- Zawadzki, I. I., 1973: Errors and fluctuations of raingauge estimates of areal rainfall. *J. Hydrol.*, **18**, 243–255.
- Zhao, M., and P. A. Dirmeyer, 2003: Production and analysis of GSWP-2 near-surface meteorology data sets. Center for Ocean Land Atmosphere Studies Tech Memo. 159, 38 pp.