AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2016JD024932

Key Points:
- Observed land and atmosphere initial states increase skill of both precipitation and 2 m temperature in CCSM4 hindcasts
- Perfect model predictability indicates skill to be gained should model errors, initial errors, or land-atmosphere coupling strength improve

Supporting Information:
- Supporting Information S1

Correspondence to:
J. M. Infanti,
jinfanti@rsmas.miami.edu

Citation:
Infanti, J. M., and B. P. Kirtman (2016), Prediction and predictability of land and atmosphere initialized CCSM4 climate forecasts over North America, *J. Geophys. Res. Atmos.*, *121*, 12,690–12,701, doi:10.1002/2016JD024932.

Received 9 FEB 2016
Accepted 14 OCT 2016
Accepted article online 15 OCT 2016
Published online 8 NOV 2016

Prediction and predictability of land and atmosphere initialized CCSM4 climate forecasts over North America

Johnna M. Infanti[1] and Ben P. Kirtman[1]

[1]Department of Atmospheric Sciences, University of Miami Rosenstiel School of Marine and Atmospheric Science, Miami, Florida, USA

Abstract Subseasonal-to-seasonal prediction is influenced by slowly varying surface fields such as sea surface temperature (SST) and soil moisture. Fully coupled hindcasts were recently completed in the Community Climate System Model version 4.0 (CCSM4) as part of the North American Multi-Model Ensemble project. Using similar land and atmosphere initialization strategies, but with prescribed climatological SSTs, we attempt to determine the isolated impact of combined observed atmosphere and land initialization and of observed atmosphere initialization on monthly precipitation and 2 m temperature prediction-estimated skill (i.e., skill assessed without SST variability) and predictability on monthly time scales. CCSM4 has been cited as having low land-atmosphere coupling, and while combined land and atmosphere initialization significantly increases the estimated skill of precipitation and temperature in the first month after initialization (lead 0), land initialization influence is weak, consistent with low land-atmosphere coupling in CCSM4. In contrast, atmosphere initialization is a stronger contributor to prediction skill and predictability. We find stronger influence of land and atmosphere initialization on precipitation in CCSM4 versus results from CCSM3. Predictability results show that there is potential skill to be gained for both precipitation and temperature should model errors, atmosphere or land initial state errors, and/or land-atmosphere coupling improve.

1. Introduction

A series of retrospective forecasts (or hindcasts) were recently completed with the Community Climate System Model version 4.0 (CCSM4) for use within phase 2 of the North American Multi-Model Ensemble System for intraseasonal to interannual prediction (North American Multi-Model Ensemble (NMME)) [*Kirtman et al.*, 2014]. The rationale behind the predictions themselves is that slowly varying surface fields such as sea surface temperature (SST) and soil moisture influence skill of precipitation and temperature on time scales of months to seasons [*Shukla*, 1998; *Paolino et al.*, 2011, among others]. CCSM4 hindcasts made for a period of 1982–2009 included in NMME are fully coupled, with atmosphere, land, and ocean domains initialized using Climate Forecast System Reanalysis (CFSR) data [*Saha et al.*, 2010]. Effectively, although CFSR is a data assimilation product, these hindcasts are initialized from data representative of observations and can be used to provide estimates of skill or past model performance. Our focus here is on the influence of land and atmosphere initialization on estimated hindcast skill and predictability in CCSM4. We utilize an idealized experiment design where land and atmosphere initialization is isolated from constructive or destructive ocean influence by prescribing observed, climatological SSTs, with land and/or atmosphere initialization techniques consistent with fully coupled CCSM4 hindcasts.

Fully coupled CCSM4 hindcasts follow from Community Climate System Model version 3.0 (CCSM3) hindcasts, which were included in NMME phase 1. CCSM3 hindcasts were previously described in *Kirtman and Min* [2009], who found that hindcasts initialized from ocean observational estimates from a separate data assimilation source [i.e., *Derber and Rosati*, 1989] were suitable for climate prediction studies. *Paolino et al.* [2011] studied the addition of land and atmosphere initialization strategies to seasonal fully coupled hindcasts in CCSM3, with focus on hindcasts with ocean-land-atmosphere initialization compared to hindcasts with only the ocean initialized. In CCSM3, *Paolino et al.* [2011] found no additional errors, better soil moisture predictions, and better surface temperature predictions in wintertime seasonal hindcasts using ocean-land-atmosphere initialization compared to ocean-only initialization. However, precipitation skill did not benefit from the addition of land and atmosphere initialization compared to ocean-only initialization on seasonal time scales. *Paolino et al*. [2011] speculate that a significant portion of the skill in the seasonal forecasts of temperature is due to initialization of the land surface using observational estimates.

Although CCSM3 hindcasts were studied in some detail in the above manuscripts, land and atmosphere initialization techniques in CCSM4 hindcasts have not yet been studied. Moreover, the impact of land and atmosphere initial states on monthly forecast skill remains an open question. Further study of CCSM4 hindcasts and land and atmosphere initialization is necessitated due to some key differences between CCSM3 and CCSM4. CCSM4 has many improvements over CCSM3, including better representation of surface air temperature and extreme events [*Gent et al*., 2011]; however, *Mei and Wang* [2012] found that land-atmosphere coupling in CCSM4 is much weaker than CCSM3 and that CCSM coupling strength is overall underestimated compared with both observational and reanalysis data as well. This comparatively weaker land-atmosphere coupling may impact CCSM4 prediction skill, especially for variables influenced by the land surface such as land-based 2 m temperature or precipitation. Formulation of initial CCSM4 land and atmosphere states is similar to CCSM3; however, both land and atmosphere initial states are taken from a different source, namely, ERA-40 [*Uppala et al*., 2005] and GSWP-2 [*Dirmeyer et al*., 2006] for CCSM3 and CFSR for CCSM4. While the CFSR soil moisture field produced is in agreement with current understanding, it is still an estimate [*Meng et al*., 2012]. The land initial states utilized in CCSM4 hindcasts are variance adjusted to agree with the Community Land Model version 4 (CLM4), but there still may be incompatibilities that lead to forecast drift or shock. The goal of this manuscript is to study the potential impact of land and atmosphere initialization on CCSM4 hindcast skill and predictability.

In regards to land and atmosphere initialization and their influence on climate predictions, atmospheric initialization adds to hindcast skill due to initialization of large-scale fields such as geopotential height but decays within ~2 weeks [*Paolino et al*., 2011]. The memory of an initial anomaly in the land surface influences predictions [*Koster and Suarez*, 2001] mainly through soil moisture. Soil moisture affects surface sensible/latent heating, thus local precipitation and temperature anomalies [*Eltahir*, 1998; *Seneviratne et al*., 2010]. Soil moisture memory is about 2–3 months [*Dirmeyer*, 2003; *Paolino et al*., 2011]. While *Paolino et al*. [2011] found a significant influence to 2 m temperature skill due to land and atmosphere initialization in CCSM3, the authors did not find a significant skill increase for precipitation. However, *Kim and Wang* [2007] found that soil moisture anomalies persisted long enough to influence seasonal precipitation.

Our focus is on the estimated monthly hindcast skill (i.e., idealized model hindcast experiments compared to observations) of North American land-based precipitation and 2 m temperature. Experiments are performed in an idealized setting where the influence of observed land and/or atmosphere initializations are isolated from that of the ocean by prescribing observed climatological SSTs. We also compare the estimates of monthly prediction skill to estimates of predictability (perfect model), which can offer guidance on months or regions in which forecasts could be improved given less forecast errors. As the monthly hindcast skill depicted in this manuscript is calculated for idealized simulations with climatological SSTs, it is an estimate of the impact of land and atmosphere initialization on skill and not a true representation of CCSM4 hindcast skill.

In this manuscript, the land and atmosphere components of CCSM4, the Community Atmosphere Model version 4 (CAM4) and the Community Land Model (CLM4), are initialized every June and every December from observed (data assimilated) atmosphere and land states, where initialization months are chosen due to summer/winter differences in evaporation and soil moisture [*Seneviratne et al*., 2010], although our main focus is on June initialization. Experiments are performed with prescribed climatological SST, effectively isolating the influence of land or atmosphere initialization on skill and predictability. Although these experiments are performed using the atmospheric general circulation model formulation of CCSM4, we utilize initialization strategies that are identical to those used in fully coupled CCSM4 hindcasts and forecasts, as this could provide further information and potentially improve/better understand these climate predictions. Given results from *Mei and Wang* [2012] indicating lower land-atmosphere coupling in CCSM4, this manuscript expands upon those from CCSM3 to include hindcasts of monthly precipitation and 2 m temperature, predictability assessments, and a comparison of experiments with observed land and atmosphere initialization to those with observed atmosphere initialization and land initialization taken from a randomly chosen CCSM4 initial state.

2. Methods

2.1. Modeling Framework

We consider the atmospheric component of CCSM4 coupled to the land component (CAM4-CLM4) with prescribed ocean/ice. We compare three experiments, described below.

1. No-Init: 30 year simulation where monthly climatological SST and sea ice coverage from observational estimates are prescribed. No initialization in land or atmosphere.
2. LA-Init: initialized hindcasts (1982–2009; 10-member ensembles) using observed atmospheric and land states (initialization strategy discussed below). Prescribed climatological SST and sea ice as above.
3. A-Init: initialized hindcasts (1982–2009; 10-member ensembles) using observed atmospheric states (initialization strategy discussed below). The land initial state is taken from a long run of CCSM4 (CCSM4 long run described below), as opposed to observations. Prescribed climatological SST and sea ice as above.

We loosely define experiments 2 and 3 as "hindcasts," as the initial states in the land and atmosphere (for experiment 2) and in the atmosphere (for experiment 3) are initialized from observed, data-assimilated, CFSR initial states. Land and atmosphere initial states are identical to CCSM4 fully coupled hindcasts when CFSR data are used for initialization. However, in all experiments, SST and sea ice are prescribed from 1982 to 2001 climatology using Hadley Centre Global Sea Ice and Sea Surface Temperature data [*Rayner et al*., 2003]. Using prescribed SSTs decouples the global ocean and atmosphere as the atmosphere is unable to influence SST variability (for more information on the implications of such exercises, see *Wu and Kirtman* [2007]). Because we utilize climatological SSTs, SST anomalies do not influence the global circulation. Effectively, this experiment design isolates the influence of land and atmosphere initialization from the influence of the ocean.

The No-Init experiment uses prescribed climatological SSTs and does not have any initialization in the land or atmosphere. No-Init is a free-running simulation corresponding to twentieth century forcing, with coupling between the atmosphere and land surface. The No-Init experiment does not specifically correspond to any observed period. The remaining experiments are initialized every June and every December for the period of 1982–2009 using the methodology described below.

Initialization of CCSM4 is similar to CCSM3 in concept, although the initial data sets differ (CCSM3 initialization discussed in *Paolino et al*. [2011]). All initial data are taken from CFSR [*Saha et al*., 2010]. CFSR includes data-assimilated estimates of observed initial states. In the interest of brevity, we refer to the experiments with land and/or atmosphere initialization from CFSR as initialized from observations. The atmospheric component (CAM4) is initialized from multilevel fields of temperature, zonal and meridional winds, specific humidity, cloud liquid water content, cloud ice water content, and cloud fraction and from single-level fields of surface pressure, surface geopotential, surface temperature, and planetary boundary layer height. CFSR data are regridded to the $0.9 \times 1.25°$ grid and 26 hybrid sigma pressure levels used by CAM4.

In the LA-Init experiment, the land component (CLM4) is initialized from daily fields of soil moisture, soil temperature, snow depth, snow temperature, vegetation temperature, and canopy moisture, again from CFSR. These fields are normalized by their standard deviations, and combined with the mean and standard deviation of soil climatology from 30 years of CLM4 output data, sampled after 100 year spin-up (i.e., variance corrected). Initial data south of 60S is set to model climatology. As observed initial states are not available, vegetation temperature and canopy moisture initial fields are produced from a 7 day CCSM4 spin-up forecast where CAM is initialized as above to produce fields influenced by the initial atmospheric state.

In the A-Init experiment, the land initial states are taken from a long run of CCSM4 and do not involve CFSR or any observational data. The atmosphere initial states are taken from CFSR as above. The CCSM4 long run is a freely evolving fully coupled CCSM4 control run that utilizes forcing corresponding to the year 2000. Although this run uses year 2000 forcing, the model output does not correspond specifically to any observed year(s). Approximately 300 model years are available in the CCSM4 long run. In order to allow some variance in the initial state, the land initial states are taken from model years 175 to 184, depending on ensemble member. The land initial states in the A-Init experiment are adapted from CLM4 output and are similar to the CFSR initial state, in that daily fields of soil moisture, temperature, etc., are included for initialization, but these states are not consistent with any specific time period of observations, beyond a general representation of the current period. Thus, while the land initial states here are "realistic" in the sense that the model output is representative of the natural climate, the land initial states in the A-Init experiment are not observed and are not in agreement with the atmosphere initial states. Therefore, A-Init shows the influence of observed atmosphere states, in the absence of observed land initial states.

In order to initialize each ensemble member, daily data from the end of the prior month is used. For example, a hindcast initialized on 1 June uses data from May 26 00:00 Z to initialize ensemble member 1, data from May 26 12:00 Z to initialize ensemble member 2, and so on to May 30 12:00 Z to initialize ensemble member 10. We refer to the first month after initialization as lead 0 (i.e., June-initialized hindcast predicting June), second month after initialization as lead 1 (i.e., June-initialized hindcast predicting July), and so on.

2.2. Methods of Analysis

Observational data considered for model verification is the Climate Prediction Center Merged Analysis of Precipitation (CMAP) [*Xie and Arkin*, 1997] and the Global Historical Climatology Network/Climate Anomaly Monitoring System 2 m temperature (T2m) [*Fan and van den Dool*, 2008]. All data, including model output, are regridded to a 1° latitude by 1° longitude grid.

Our attention is on anomaly correlation, and the ensemble mean (if available) is considered the most probable outcome and treated as one hindcast [*Stefanova et al*., 2012; *Infanti and Kirtman*, 2013]. Probabilistic analysis can offer more information, such as prediction skill for upper or lower tertile events, etc. [*Mason and Graham*, 1999; *Kirtman*, 2003], however, to simplify our comparison to *Paolino et al*. [2011] and in the interest of space we use deterministic measures.

Correlation of model anomalies with observed anomalies is used as an estimate of deterministic hindcast skill. However, it should be noted that the skill depicted in this manuscript is only an estimate, as the hindcasts do not incorporate SST variability; thus, the figures are not representative of the true realizable hindcast skill. This estimate is defined using anomaly correlation of the experiments with observations, where the ensemble mean anomaly (or model output anomaly) of the given experiment is correlated with observations for a period of 1982–2009. Anomalies are calculated with respect to the 1982–2009 monthly mean. For experiments 2 and 3, climatology is removed from each ensemble member and the resulting anomalies are averaged assuming equal weights to form the ensemble mean.

We are specifically interested in a comparison of experiments, for example, if LA-Init correlation with observations significantly differs from No-Init correlation with observations. Because correlation is not normally distributed, comparison of correlation coefficients is not straightforward as simply subtracting. We use Fisher's R-to-Z transformation [see *Wilks*, 2011; *Infanti and Kirtman*, 2013] to transform the correlations to approximately normal. This allows us to easily difference or average correlation values and assign significance levels. Differences are shown using transformed correlation values, and significance is shown at the 95% confidence level, assuming a normal distribution of values (contours). While Fisher's R-to-Z transformation is a practical choice for comparison of correlation coefficients, some caveats to this are discussed in *DelSole and Tippett* [2014].

A second consideration for this manuscript is predictability or the "ability to predict" a given variable in an idealized situation [e.g., *Boer*, 2004]. Predictability is the expected prediction skill if the model hindcasts were without systematic error or bias, indicating regions or time periods in which we would expect variability to be predictable [e.g., *Cheng et al*., 2011]. However, this is only true to the extent that the model variability is consistent with observed variability, thus is only an estimate. The approach used here to estimate predictability assumes a perfect model framework where one ensemble member is assumed "truth" and the remaining nine form the ensemble mean. Anomaly correlation is then defined as above using the nine-member ensemble mean with the chosen verification ensemble member. To minimize sampling bias, we calculate all possible combinations of nine ensemble members predicting the remaining ensemble member. For example, the mean of members 1 to 9 predicting ensemble member 10 and the mean of members 1 to 8 and 10 predicting ensemble member 9. These correlations are averaged after conversion to approximately normal following the transformation above. Robustness of this calculation is estimated given 7/10 predictability correlations agreeing that the change is greater than zero (stippling), highlighting where there is a robust increase in predictability given the relevant comparison. The results of this differ from a traditional significance test as they are more relaxed but are appropriate for our purposes of determining if there is a robust skill increase.

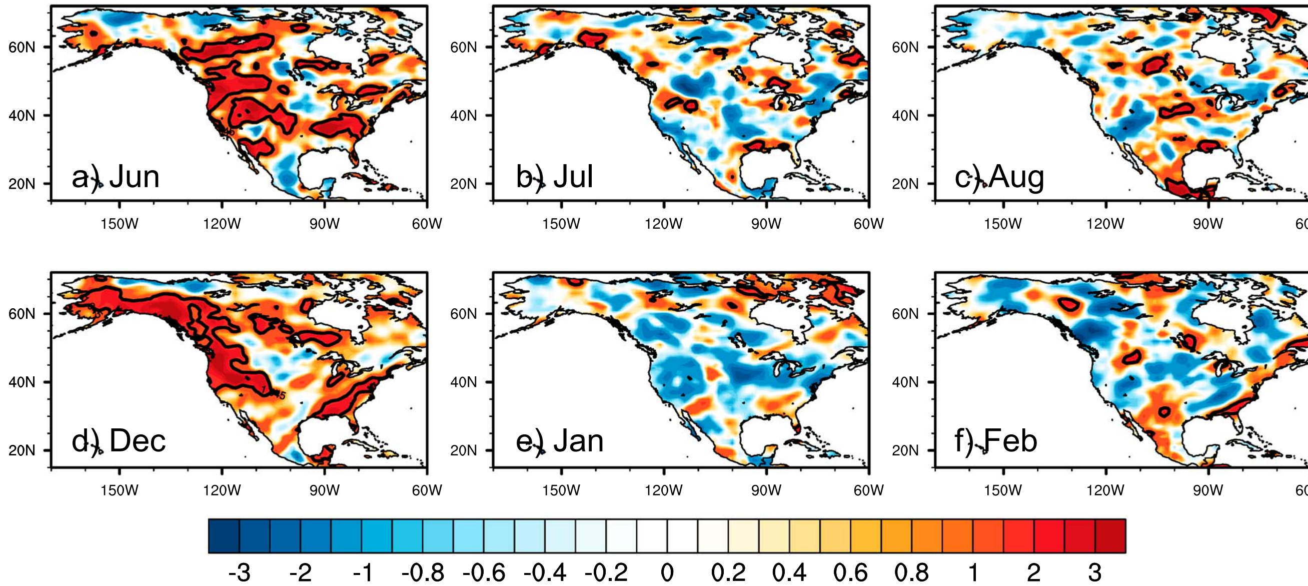

Figure 1. Difference in estimated deterministic skill of precipitation for model output correlated with observations. The difference is defined as the anomaly correlation of LA-Init minus No-Init, where correlations have been transformed to approximately normal using Fisher's *R*-to-*Z* transformation. The red (blue) shading indicates the regions in which land and atmosphere initialization from observations improve (decrease) skill. The contours indicate the significance of difference at 95% confidence level. (a–c) June-initialized hindcasts (or No-Init output) verifying in June, July, and August. (d–f) December-initialized hindcasts (or No-Init output) verifying in December, January, and February.

3. Results

3.1. Prediction Skill

As stated in section 2.2, we are mainly interested in a comparison of the experiments. The first comparison we show is LA-Init versus No-Init, which highlights any estimated skill increase or decrease due to combined observed land and atmosphere initialization (Figures 1 and 2). The second comparison we show is LA-Init versus A-Init, which highlights any estimated skill increase or decrease due to combined observed land and atmosphere initialization over forecasts with only the atmosphere initial state taken from observations (Figure 3). Again, we caution the reader that while we refer to the results of this section as "prediction skill" due to their comparison with observations as opposed to homogeneous predictability, the depicted skill is an estimate given an idealized experiment environment in which climatological SSTs are prescribed. Thus, these results should not be treated as the "true" skill of CCSM4 hindcasts.

The first comparison (LA-Init and No-Init) allows us to easily deduce regions where combined observed land and atmosphere initialization significantly increase or decrease hindcast skill compared to an uninitialized run. June (December) initialized hindcasts predicting June, July, and August (December, January, and February) precipitation are shown in Figures 1a–1c (Figures 1d and 1e). Figure 2 is similar but for 2 m temperature comparison. In this comparison, anomaly correlations have been transformed to approximately normal using Fisher's *R*-to-*Z* transformation, and the resulting transformed correlations are subtracted. The full anomaly correlation fields for LA-Init and No-Init (nontransformed and for each experiment) are shown in supporting information S1 and S2 for precipitation and 2 m temperature, respectively. As these experiments are performed in the absence of SST anomalies (SSTA), this comparison isolates the influence of land and atmosphere initialization from that of SSTA by design. If we were to include SSTA, it is possible that SSTA could reinforce atmosphere and land initial anomalies and improve the forecast skill associated with the initial condition.

For precipitation (Figure 1), combined observed land and atmosphere initialization largely adds to skill over North America in the first month after initialization (June and December), sometimes significantly, when compared to No-Init. There are few grid points where initializing decreases skill of lead 0 predictions. Approximately 76% of grid points show skill increase in June and 86% in December. This ~10% difference in June compared to December may be surprising due to larger expected impact of land initialization on

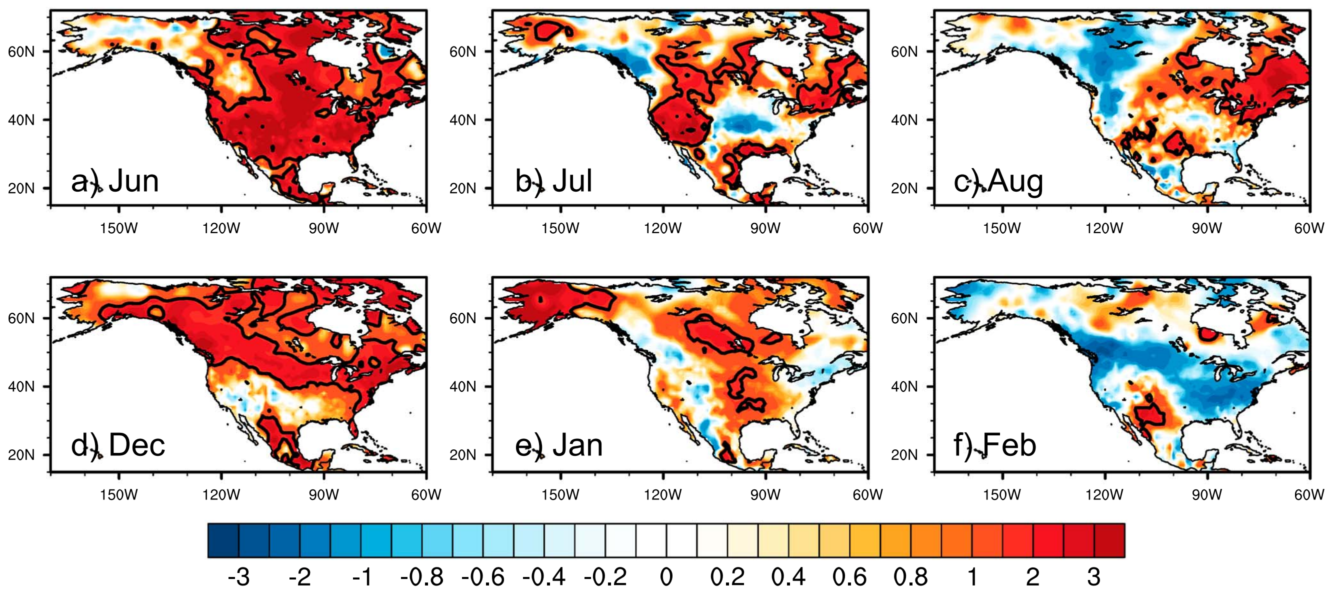

Figure 2. (a–f) Same as in Figure 1 but for 2 m temperature.

skill in summer months [*Koster et al.*, 2004; *Seneviratne et al.*, 2010; etc.]. However, July and August show comparatively more grid points (50% and 56%) with skill increase than January and February (41% and 47%), and initialization negatively affects skill in many regions in January and February, though not significantly. This indicates stronger persistence of initial anomalies in the June-initialized hindcasts compared to December-initialized hindcasts.

Results for 2 m temperature are similar (Figure 2), but the LA-Init hindcasts show much larger regions of significant difference in the first month after initialization compared to precipitation. For June-initialized 2 m temperature hindcasts, 92%, 78%, and 65% of grid points in June, July, and August show increased skill

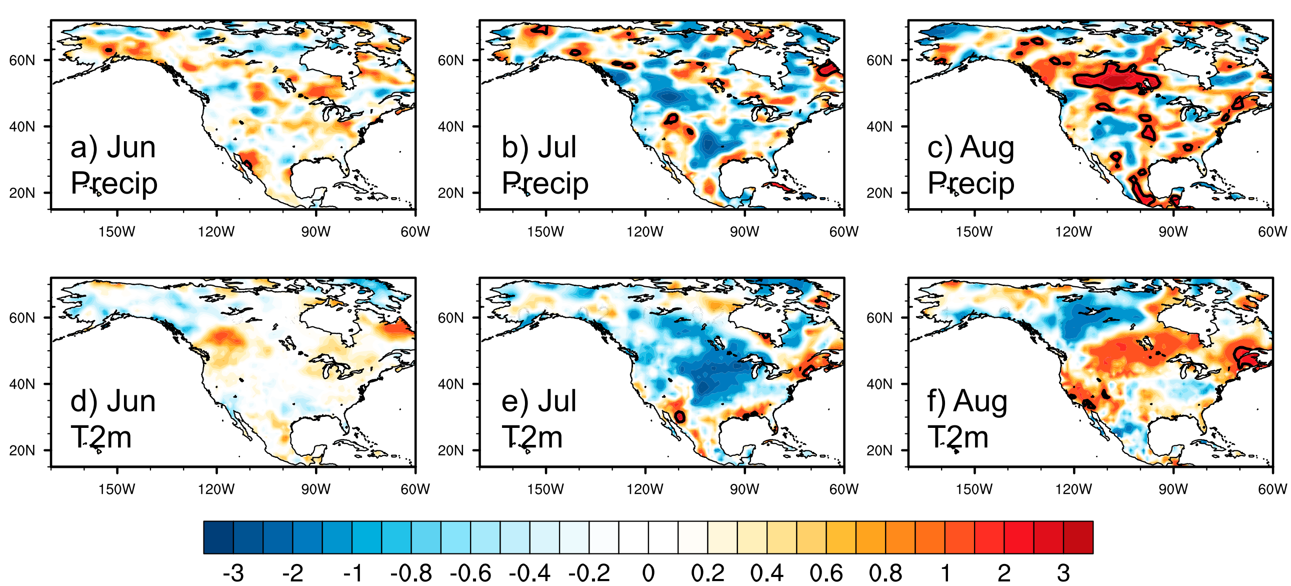

Figure 3. Difference in estimated deterministic skill of precipitation for model output correlated with observations. The difference is defined as the anomaly correlation of LA-Init minus A-Init, where correlations have been transformed to approximately normal using Fisher's *R*-to-*Z* transformation. The red (blue) shading indicates the regions in which *combined* land and atmosphere initialization from observations improves (decreases) skill versus a hindcast initialized from only observed atmosphere states. The contours indicate the significance of difference at 95% confidence level. (a–c) June-initialized precipitation hindcasts verifying in June, July, and August. (d–f) June-initialized 2 m temperature hindcasts verifying in June, July, and August.

in LA-Init. For December-initialized 2 m temperature hindcasts, 96%, 81%, and 37% of grid points in December, January, and February show increased skill over No-Init. These results are expected, as 2 m temperature skill is generally stronger than precipitation due to more spatial coherence and temporal persistence in temperature fields.

We have also computed the anomaly correlation between forecasted and observed precipitation for the first and second 2 weeks separately in the LA-Init experiment (not shown). There is a distinct difference in the anomaly correlation for the first 2 weeks of the hindcasts versus the second 2 weeks, with higher anomaly correlation in the first 2 weeks. *Paolino et al*. [2011] performed a repeat analysis of seasonal skill of the fully coupled CCSM3 OLA-initialized hindcasts with the first 2 weeks of the hindcasts removed in order to isolate the effects of land initialization. On seasonal time scales (3 month mean), there was no significant change to the anomaly correlation, which indicates that land initialization is important on seasonal time scales. It is possible that the land surface initialization in CCSM4 is only affecting the prediction skill at very short leads (~2 weeks). We question whether the increase in skill of lead 0 hindcasts seen in Figures 1 and 2 is due to short-lived land surface anomalies related to the initial state or due to atmospheric initialization. Thus, we further decompose the CCSM4 hindcasts into those with combined land and atmosphere initialization and those with only atmosphere initialization. This allows us to determine how much of this skill increase is coming from observed land initial states versus observed atmosphere initial states. Because there is typically more land-atmosphere interaction in the summer months, we focus the remaining analyses on June-initialized hindcasts.

Figure 3 shows a comparison of transformed anomaly correlations for the LA-Init experiment minus A-Init for both precipitation (Figures 3a–3c) and 2 m temperature (Figures 3d–3f). The nontransformed anomaly correlation fields for June-initialized A-Init hindcasts are shown in supporting information S3. While A-Init does have land initialization (required by model formulation), the initial state is taken from a long run of CCSM4 as opposed to CFSR. Thus, this comparison indicates regions where land initialization in combination with atmosphere initialization based on *observed* states significantly enhances or weakens prediction skill. For precipitation in the first month after initialization (June; Figure 3a), there is very little difference between LA-Init and A-Init experiments (similar results are found for 2 m temperature). Thus, land initialization from observed states does not significantly add or subtract from hindcast skill. In the remaining months, differences are larger, but again overall insignificant, excepting select regions for August precipitation hindcasts (Figure 3c). Similar results are found for 2 m temperature. The largest positive influence on skill when combined observed land and atmosphere initial states are used is seen at lead 2 for both precipitation and 2 m temperature (although this is only sometimes significantly more positive). This indicates that initialization from observed land states influences long-term persistence; however, we also note that there was an insignificant gain in overall skill at this lead time, so the influence is quite small.

Overall, the comparison of LA-Init to A-Init indicates that using observed land initial states in conjunction with observed atmosphere initial states does not affect skill at lead 0, *decreases* skill at lead 1, and moderately *increases* skill at lead 2. However, differences are rarely significant. While the LA-Init experiment significantly increased lead 0 prediction skill over an uninitialized run, inclusion of observed land initial states does not significantly affect skill at this lead time (positively or negatively). There is some indication of more skillful predictions when the land surface is initialized from observations at lead 2. The weak influence of observed land surface initialization on skill is due to two possible sources: soil moisture initial data may be inaccurate, and/or the documented weak land-atmosphere coupling strength in CCSM4 [*Mei and Wang*, 2012]. In other words, model systematic error or errors in the initial land surface states may overshadow any positive influence on skill. In the remaining section, we discuss differences in predictability. This minimizes any model errors as it considers a perfect model framework. The predictability results in the remaining section intend to offer guidance on regions/months in which we could skillfully predict variability given less model or initial state errors.

3.2. Predictability

Our aim for this section is to determine the predictability of monthly precipitation and 2 m temperature in the hindcasts, again focusing on summertime predictions as this season shows larger latent heating (evaporation) and traditionally stronger coupling of land and atmosphere, in comparison to winter months. As the experiments we consider here do not include SST variability, this assessment isolates the predictability stemming from observed land and/or atmosphere initial states [e.g., *Koster and Suarez*, 2003], rather than any predictability stemming from SSTA [e.g., *Shukla et al*., 2000].

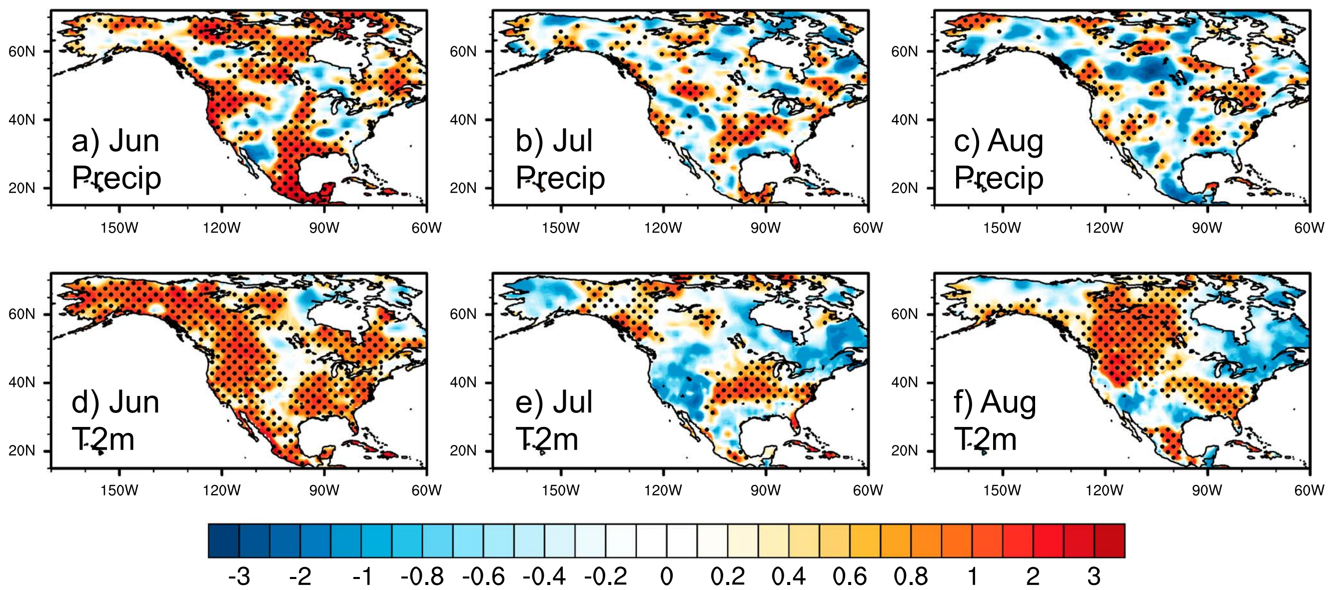

Figure 4. Average difference of homogeneous predictability minus prediction skill for LA-Init experiment. Correlations are transformed using Fisher's *R*-to-*Z* transformation. Differences are defined as predictability minus prediction skill, and the transformed average difference is shown by shading. The red (blue) shading indicates the regions in which predictability of the LA-Init experiment is larger (smaller) than prediction skill. The stippling indicates where 7/10 possible differences show a positive increase in predictability. (a–c) June-initialized precipitation hindcasts verifying in June, July, and August. (d–f) June-initialized 2 m temperature hindcasts verifying in June, July, and August.

As discussed in section 2.2, the predictability measure considered here is homogeneous anomaly correlation or the mean of nine ensemble members predicting the remaining ensemble member. In order to avoid sampling bias in the chosen verification ensemble member, we consider all possible combinations of 9 ensemble members predicting the remaining (10 possible combinations). All anomaly correlations are again transformed to approximately normal using Fisher's *R*-to-*Z* transformation. We first consider the difference of predictability minus prediction skill of the LA-Init experiment (Figure 4). We further decompose this into the predictability of the LA-Init experiment minus the predictability of the A-Init experiment (Figure 5). As opposed to a traditional significance test, we consider the robustness of the difference, defined as when 7/10 of the differences are positive (stippling). As these results are depicted in a "perfect model" framework, they are specific to CCSM4 and cannot be generalized to other models.

Comparing LA-Init predictability to estimated prediction skill for precipitation (Figures 4a–4c), it is not surprising that the predictability is quite higher, notably in the Central U.S. at leads 0 and 1. While there are some small regions of robust increase at lead 2, there does not appear a coherent predictability increase at this lead time; thus, the influence of land and atmosphere initialization on predictability has decreased. Interestingly, the increased predictability is located in the Central U.S., which has been previously defined as a "hot spot" of soil moisture-precipitation coupling by *Koster et al*. [2004]. This indicates that despite the weaker land-atmosphere coupling in CCSM4, some influence can still be seen on local precipitation, and we speculate that predictability would increase if coupling strength were improved. For 2 m temperature (Figures 4d–4f), there is a (comparatively) large and robust increase in predictability primarily at leads 0 and 2. The enhanced predictability of these variables over prediction skill suggests that there is potential skill to be gained should model errors be minimized, either with respect to the initial states or the coupled land-atmosphere system. More attention on initial states is discussed below.

We also consider the predictability of LA-Init compared to predictability of A-Init (Figure 5). Figure 5 is similar in concept to Figure 3 but for predictability. While there are some regions of robust increase in precipitation predictability in the LA-Init experiment compared to the A-Init experiment, the differences are weak compared to 2 m temperature. Predictability of precipitation is not benefitting from the use of observed initial land states in this model. Thus, observed land initial states affect 2 m temperature predictability more than precipitation, including at long leads (lead 2). The predictability results suggest that should the initial states be improved, the bulk of the improvement will be seen with 2 m temperature. This also means that 2 m

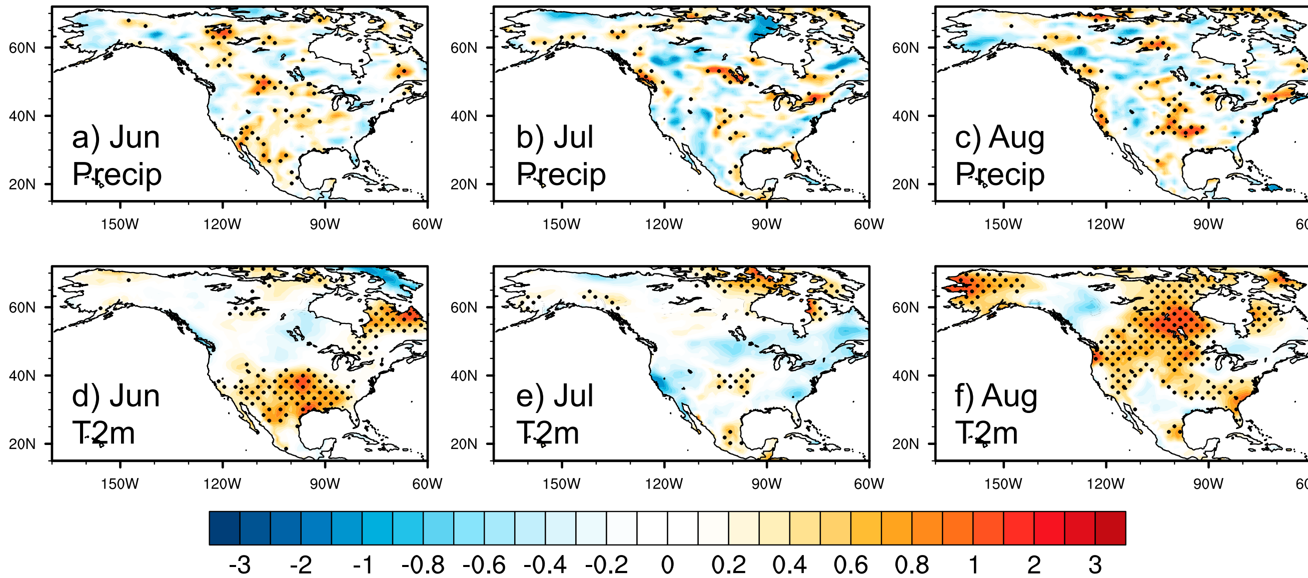

Figure 5. Average difference in homogeneous anomaly correlations (predictability) of LA-Init compared to A-Init. Homogeneous anomaly correlation is defined as the correlation of all possible combinations of the nine ensemble member mean predicting the remaining. Correlations are transformed using Fisher's *R*-to-*Z* transformation. Differences are defined as LA-Init minus A-Init, and the transformed average difference is shown. The red (blue) shading indicates the regions in which *combined* land and atmosphere initialization from observations improves (decreases) predictability versus a hindcast initialized from only observed atmosphere states. The stippling indicates where 7/10 possible differences show a positive increase in predictability. (a–c) June-initialized precipitation hindcasts verifying in June, July, and August. (d–f) June-initialized 2 m temperature hindcasts verifying in June, July, and August.

temperature is more strongly influenced by initial state error than precipitation. This agrees with recent results considering influence of initialization on precipitation and 2 m temperature [for example, *Koster et al*., 2006]. Again, we caution that this result should be interpreted with the caveat that CLM4 in CCSM4 in know to have weak atmosphere-land coupling strength.

Overall, these results imply that while using combined observed land and atmosphere initialization does lead to a significant increase in (estimated) monthly skill of *both* precipitation and 2 m temperature (particularly at lead 0), predictability results performed in a perfect model indicate regions where there is still potential skill to be gained, such as the Central U.S. for precipitation at short leads and in many regions for 2 m temperature. The predictability comparison of LA-Init to A-Init shows that 2 m temperature is more strongly affected by errors in the initial land state than precipitation.

4. Discussion

This analysis highlights the influence of land and atmosphere initialization on estimated monthly deterministic prediction skill and predictability, isolating this influence from any SSTA constructive or destructive interference by utilizing prescribed, climatological SSTs. We discuss differences in estimated prediction skill when observed (data-assimilated) land and atmosphere initial states are used compared to an uninitialized run and also compare estimated prediction skill of observed land and atmosphere initialized hindcasts to those with only the atmosphere initialized from observed states. Finally, we compare the predictability of land and atmosphere initialized hindcasts to atmosphere-only initialized hindcasts.

Compared to an uninitialized hindcast, there is a significant increase in precipitation skill over North America at lead 0, and out to lead 2 for 2 m temperature (in select regions), when observed land and atmosphere initial states are used. This result is based on an idealized framework in which SSTA are not included. In a study comparing ocean-only initialized hindcasts to ocean-land-atmosphere initialized hindcasts using the fully coupled CCSM3 model, the increase in precipitation skill due to land and atmosphere initialization was not as apparent on seasonal time scales [*Paolino et al*., 2011]. CCSM3 showed larger influence of land and atmosphere initialization on 2 m temperature. Similar assessment using the multimodel Global Land-Atmosphere Coupling Experiment (GLACE-2) found that realistic initialization had small effect on precipitation skill past 30 days and a larger effect on 2 m temperature that persisted to almost day 60 [*Koster et al*., 2010]. Although this was a multimodel assessment, the CCSM4 results presented here agree in many ways with the results from GLACE-2 and other studies involving precipitation/2 m temperature [e.g., *Mei et al*., 2013]; however, weaker land-atmosphere coupling in CCSM4 may be influencing the results [*Kim and Wang*, 2007].

We further decompose the skill influence into the contributions from combined observed land and atmosphere initialization and contribution from observed atmosphere initialization (with realistic but not observed land states). Interestingly, while we may have expected the influence of initial states on skill to be linearly decreasing with lead due to persistence, this is not the case. This comparison shows no significant difference at lead 0, decreased (in many regions) skill at lead 1, and increased skill at lead 2 (or a return of skill) when observed land states are included, although these differences are rarely significant. Despite correction of the initial states to match the mean and variance of CLM4, the decrease in skill at lead 1 when observed land initial states are included may be related to “rejection” of the initial land state or initial shock, which is realized as a decrease in skill. The return in skill at lead 2 indicates a delayed influence of the land initial states after the initial shock has occurred. The reader should note that while we find differences in skill when considering LA-Init to A-Init experiments at leads 1 and 2, overall skill (correlation of LA-Init with observations) is weak at these two lead times.

We also consider measures of predictability, or the models ability to predict a given variable given a perfect model framework, i.e., initial state and model systematic error minimized. Predictability compared to prediction skill of the LA-Init experiments indicates that both precipitation and 2 m temperature have regions in which there is skill to be gained should model or initial errors be minimized. For precipitation, robustly increased predictability exists for lead 0 and 1, with the Central U.S. highlighted as a region where predictability is larger than prediction skill. For 2 m temperature, predictability is increased over prediction skill primarily at leads 0 and 2. Predictability of LA-Init versus A-Init shows very little influence of observed land initial states on predictability of precipitation, compared to 2 m temperature. This indicates that land initialization (and errors in the initial state) have a smaller influence on precipitation and a stronger influence on 2 m temperature.

The predictability results indicate that prediction skill of precipitation could be weakly improved given improvements to model errors or land-atmosphere coupling strength, and prediction skill of 2 m temperature could be strongly improved given improvements to initial land states. However, we base these results on predictability analysis intrinsic to the model used, assuming that model errors are minimized (homogeneous predictability). This does not take into account that the model may overestimate, or underestimate, the actual predictability associated with natural variations stemming from, for example, soil moisture precipitation coupling. Despite this caveat, this comparison of skill and predictability can give a first-order estimate of regions that may potentially benefit from improved accuracy of land and atmosphere initial states or model systematic error. The predictability analysis can thus be used as guidance for regions or seasons that would benefit from greater accuracy in initial states, such as 2 m temperature.

Although these CCSM4 results agree with recent work, *Koster et al*. [2011] found that stratifying skill based on anomalously wet and anomalously dry soil initial states lead to differing levels of skillfulness depending on region and initial soil state. We did not stratify our results but expect that CCSM4 would behave similarly. By design, we have isolated the effects of observed land and atmosphere initialization from that of SSTs; however, the inclusion of ocean initialization, ocean-atmosphere coupling, or SSTs will affect skill, particularly on seasonal time scales and longer leads [*Kirtman and Min*, 2009; *Paolino et al*., 2011; *Infanti and Kirtman*, 2015]. We expect that both prediction skill would improve and predictability would be enhanced if SSTs were included, as SST variability can cause an increase in persistence of initial anomalies given coincident initial states [*Schubert et al*., 2007]. The results presented here include the monthly time scale in order to study the evolution of influence of land and atmosphere initial states but also benefit the climate prediction community by considering these shorter lead monthly predictions. Although we did not include the seasonal time scale in this analysis, these results and those from *Koster et al*. [2011] and *Paolino et al*. [2011] indicate that the influence of land and/or atmosphere initialization declines quickly and will not extend past 1 season. As 2 m temperature benefits more from land initialization, influence lasts longer, but we expect it to be strongest in the first season.

Overall, these results show that for CCSM4, both precipitation and 2 m temperature benefit from observed initialization in the land and/or atmosphere (to varying degrees), agreeing with results from other models, despite having weaker land-atmosphere coupling strength compared to CCSM3. Predictability results indicate that there is still potential skill to be gained for both variables, should initial errors be minimized (2 m temperature) or model systematic error be minimized (precipitation). We caution that these results are specific to CCSM4 and the initialization strategies used, but note that this experiment framework and analysis technique could be used for other models to isolate the influence of observed land and/or atmosphere initialization on predictability and skill. This study was designed specifically to isolate the influence of land and atmosphere initialization from SSTs; however, we cannot discount the very important influence of SSTs on skill and predictability. Future studies will include assessment of prediction skill and predictability in relation to fully coupled CCSM4 hindcasts with ocean, land and atmosphere initialization in comparison to uncoupled hindcasts, and those with "perfectly predicted (observed)" SSTs.

Acknowledgments
The authors would like to thank Dughong Min and the Center for Ocean-Land-Atmosphere studies for creating and maintaining the initial states for CCSM4 climate prediction studies. The free software NCAR Command Language was used to create the plots and analyze data. Observational data used for this analysis and NMME data can be found at http://www.cpc.ncep.noaa.gov/products/NMME/. Model code is available from NCAR at http://www.cesm.ucar.edu/models/ccsm4.0/. The prescribed SST CAM4 simulation data are archived at the Center for Computational Science at the University of Miami and are available upon request. We would like to thank two anonymous reviewers for their comments and suggestions that greatly improved the quality of this manuscript. The authors acknowledge support from NA15OAR4320064 and NA10OAR4320143.

References

Boer, G. J. (2004), Long time-scale potential predictability in an ensemble of coupled climate models, *Clim. Dyn.*, *23*(1), 29–44, doi:10.1007/s00382-004-0419-8.

Cheng, Y., Y. Tang, and D. Chen (2011), Relationship between predictability and forecast skill of ENSO on various time scales, *J. Geophys. Res.*, *116*, C12006, doi:10.1029/2011JC007249.

DelSole, T., and M. K. Tippett (2014), Comparing forecast skill, *Mon. Weather Rev.*, *142*(12), 4658–4678, doi:10.1175/MWR-D-14-00045.1.

Derber, J., and A. Rosati (1989), A global oceanic data assimilation system, *J. Phys. Oceanogr.*, *19*(9), 1333–1347, doi:10.1175/1520-0485(1989)019<1333:AGODAS>2.0.CO;2.

Dirmeyer, P. A. (2003), The role of the land surface background state in climate predictability, *J. Hydrometeorol.*, *4*(3), 599–610, doi:10.1175/1525-7541(2003)004<0599:TROTLS>2.0.CO;2.

Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki (2006), GSWP-2: Multimodel analysis and implications for our perception of the land surface, *Bull. Am. Meteorol. Soc.*, *87*(10), 1381–1397, doi:10.1175/BAMS-87-10-1381.

Eltahir, E. A. B. (1998), A soil moisture–rainfall feedback mechanism: 1. Theory and observations, *Water Resour. Res.*, *34*(4), 765–776, doi:10.1029/97WR03499.

Fan, Y., and H. van den Dool (2008), A global monthly land surface air temperature analysis for 1948–present, *J. Geophys. Res.*, *113*, D01103, doi:10.1029/2007JD008470.

Gent, P. R., et al. (2011), The Community Climate System Model version 4, *J. Clim.*, *24*(19), 4973–4991, doi:10.1175/2011JCLI4083.1.

Infanti, J. M., and B. P. Kirtman (2013), Southeast U.S. Rainfall prediction in the North American Multi-Model Ensemble, *J. Hydrometeorol.*, doi:10.1175/JHM-D-13-072.1.

Infanti, J. M., and B. P. Kirtman (2015), North American rainfall and temperature prediction response to the diversity of ENSO, *Clim. Dyn.*, 1–17, doi:10.1007/s00382-015-2749-0.

Kim, Y., and G. Wang (2007), Impact of initial soil moisture anomalies on subsequent precipitation over North America in the coupled land–atmosphere model CAM3–CLM3, *J. Hydrometeorol.*, *8*(3), 513–533, doi:10.1175/JHM611.1.

Kirtman, B. P. (2003), The COLA anomaly coupled model: Ensemble ENSO prediction, *Mon. Weather Rev.*, *131*(10), 2324–2341, doi:10.1175/1520-0493(2003)131<2324:TCACME>2.0.CO;2.

Kirtman, B. P., and D. Min (2009), Multimodel ensemble ENSO prediction with CCSM and CFS, *Mon. Weather Rev.*, *137*(9), 2908–2930, doi:10.1175/2009MWR2672.1.

Kirtman, B. P., et al. (2014), The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction, *Bull. Am. Meteorol. Soc.*, *95*(4), 585–601, doi:10.1175/BAMS-D-12-00050.1.

Koster, R. D., and M. J. Suarez (2001), Soil moisture memory in climate models, *J. Hydrometeorol.*, *2*(6), 558–570, doi:10.1175/1525-7541(2001)002<0558:SMMICM>2.0.CO;2.

Koster, R. D., and M. J. Suarez (2003), Impact of land surface initialization on seasonal precipitation and temperature prediction, *J. Hydrometeorol.*, *4*(2), 408–423, doi:10.1175/1525-7541(2003)4<408:IOLSIO>2.0.CO;2.

Koster, R. D., et al. (2004), Regions of strong coupling between soil moisture and precipitation, *Science*, *305*(5687), 1138–1140, doi:10.2307/3837616.

Koster, R. D., et al. (2006), GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview, *J. Hydrometeorol.*, *7*(4), 590–610, doi:10.1175/JHM510.1.

Koster, R. D., S. Mahanama, T. Yamada, G. Balsamo, A. Berg, M. Boisserie, P. Dirmeyer, F. Doblas-Reyes, G. Drewitt, and C. Gordon (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.

Koster, R. D., et al. (2011), The second phase of the Global Land–Atmosphere Coupling Experiment: Soil moisture contributions to subseasonal forecast skill, *J. Hydrometeorol.*, *12*(5), 805–822, doi:10.1175/2011JHM1365.1.

Mason, S. J., and N. E. Graham (1999), Conditional probabilities, relative operating characteristics, and relative operating levels, *Weather Forecast.*, *14*(5), 713–725, doi:10.1175/1520-0434(1999)014<0713:CPROCA>2.0.CO;2.

Mei, R., and G. Wang (2012), Summer land–atmosphere coupling strength in the United States: Comparison among observations, reanalysis data, and numerical models, *J. Hydrometeorol.*, *13*(3), 1010–1022, doi:10.1175/JHM-D-11-075.1.

Mei, R., G. Wang, and H. Gu (2013), Summer land–atmosphere coupling strength over the United States: Results from the Regional Climate Model RegCM4–CLM3.5, *J. Hydrometeorol.*, *14*(3), 946–962, doi:10.1175/JHM-D-12-043.1.

Meng, J., R. Yang, H. Wei, M. Ek, G. Gayno, P. Xie, and K. Mitchell (2012), The land surface analysis in the NCEP Climate Forecast System reanalysis, *J. Hydrometeorol.*, *13*(5), 1621–1630, doi:10.1175/JHM-D-11-090.1.

Paolino, D. A., J. L. Kinter, B. P. Kirtman, D. Min, and D. M. Straus (2011), The impact of land surface and atmospheric initialization on seasonal forecasts with CCSM, *J. Clim.*, *25*(3), 1007–1021, doi:10.1175/2011JCLI3934.1.

Rayner, N., D. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. Kent, and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, *108*(D14), 4407, doi:10.1029/2002JD002670.

Saha, S., et al. (2010), The NCEP Climate Forecast System Reanalysis, *Bull. Am. Meteorol. Soc.*, *91*(8), 1015–1057, doi:10.1175/2010BAMS3001.1.

Schubert, S., R. Koster, M. Hoerling, R. Seager, D. Lettenmaier, A. Kumar, and D. Gutzler (2007), Predicting drought on seasonal-to-decadal time scales, *Bull. Am. Meteorol. Soc.*, *88*(10), 1625–1630, doi:10.1175/BAMS-88-10-1625.

Seneviratne, S. I., 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*(3–4), 125–161, doi:10.1016/j.earscirev.2010.02.004.

Shukla, J. (1998), Predictability in the midst of chaos: A scientific basis for climate forecasting, *Science*, *282*(5389), 728–731, doi:10.1126/science.282.5389.728.

Shukla, J., et al. (2000), Dynamical seasonal prediction, *Bull. Am. Meteorol. Soc.*, *81*(11), 2593–2606, doi:10.1175/1520-0477(2000)081<2593:DSP>2.3.CO;2.

Stefanova, L., V. Misra, J. J. O'Brien, E. P. Chassignet, and S. Hameed (2012), Hindcast skill and predictability for precipitation and two-meter air temperature anomalies in global circulation models over the southeast United States, *Clim. Dyn.*, *38*(1-2), 161–173, doi:10.1007/s00382-010-0988-7.

Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, *131*(612), 2961–3012, doi:10.1256/qj.04.176.

Wilks, D. S. (2011), *Statistical Methods in the Atmospheric Sciences*, *Int. Geophys. Ser.*, vol. 100, 3rd ed., Elsevier/Academic Press, Amsterdam, Boston.

Wu, R., and B. Kirtman (2007), Regimes of seasonal air–sea interaction and implications for performance of forced simulations, *Clim. Dyn.*, *29*(4), 393–410, doi:10.1007/s00382-007-0246-9.

Xie, P., and P. A. Arkin (1997), Global precipitation: A 17-year monthly analysis based on gauge estimates, and numerical model outputs, *Bull. Am. Meteorol. Soc.*, *78*, 2539–2558, doi:10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2.