Impact of Land Model Calibration on Coupled Land-Atmosphere Prediction

Joseph A. Santanello, Jr.[1], Sujay V. Kumar[2,1], Christa D. Peters-Lidard[1],

Ken Harrison[3,1], and Shujia Zhou[4,1]

[1] NASA-GSFC Hydrological Sciences Branch, Greenbelt, MD

[2] Science Applications International Corporation, McLean, VA

[3] University of Maryland, College Park, MD

[4] Northrop Grumman Information Systems, Chantilly, VA

Submitted to *Journal of Hydrometeorology*

Summer 2012

ABSTRACT

Land-atmosphere (L-A) interactions play a critical role in determining the diurnal evolution of both planetary boundary layer (PBL) and land surface heat and moisture budgets, as well as controlling feedbacks with clouds and precipitation that lead to the persistence of dry and wet regimes. Recent efforts to quantify the strength of L-A coupling in prediction models have produced diagnostics that integrate across both the land and PBL components of the system. In this study, we examine the impact of improved specification of land surface states, anomalies, and fluxes on coupled WRF forecasts during the summers of extreme dry (2006) and wet (2007) land surface conditions in the U.S. Southern Great Plains. The improved land initialization and surface flux parameterizations are obtained through calibration of the Noah land surface model using the new optimization and uncertainty estimation subsystem in NASA's Land Information System (LIS-OPT/UE). The impact of the calibration on the a) spinup of the land surface used as initial conditions, and b) the simulated heat and moisture states and fluxes of the coupled WRF simulations is then assessed. Changes in ambient weather and land-atmosphere coupling are evaluated along with measures of uncertainty propagation into the forecasts. In addition, the sensitivity of this approach to the period of calibration (dry, wet, average) is investigated. Results indicate that the offline calibration leads to systematic improvements in land-PBL fluxes and near-surface temperature and humidity, and in the process provide guidance on the questions of what, how, and when to calibrate land surface models for coupled model prediction.

1. Introduction

Despite evidence of the importance of land-atmosphere (L-A) interactions in weather and climate prediction (e.g. Betts 2009; Seneviratne et al. 2010), the systematic impact of land surface parameterizations on coupled mesoscale modeling has proven difficult to quantify in a robust manner. The role of the land in modulating water and energy cycling has been well-documented in terms of land-atmosphere coupling strength and the support of hydrological anomalies and extremes such as flood and drought (van den Hurk et al. 2011; Koster et al. 2010), which includes immediate effects of the land on the temperature and humidity structure in the boundary layer, convective initiation, and mesoscale circulations (Di Giuseppe et al. 2011). In addition, the influence of soil moisture on precipitation has been under community-wide investigation in a range of studies from local (Santanello et al. 2011b) to global (Koster et al. 2004) scales. What is less understood is how specific land surface models (LSMs), parameterizations, datasets, and initialization approaches impact coupled mesoscale model predictions on diurnal timescales, and how each could be improved.

One confounding factor in quantifying LSM impact on coupled prediction lies in the varying and non-standard approaches to land surface spinup and initialization of mesoscale models. The impetus for the development of offline North American and Global Land Data Assimilation Systems NLDAS (Mitchell et al. 2004) and GLDAS (Rodell et al. 2004) was to be able to provide improved land initial conditions for numerical weather prediction and reanalysis systems. During this time, approaches to land spinup and initialization have diverged significantly among modeling groups and application. Recent studies have demonstrated the importance of a performing LSM spinups for mesoscale prediction (Chen et al. 2007; Kumar et al. 2008; Case et al. 2008, 2011; Wen et al. 2012; Di Giuseppe et al. 2011), and show marginalto-significant improvements over cruder initialization practices based solely on coarse resolution atmospheric models or reanalysis products. It still remains, though, that a great majority of coupled prediction studies do not make use of rigorous spinup or initialization methods, thereby limiting the potential impact of the land on those simulations before coupled integration even begins.

Adding to the non-uniformity in the treatment of the land surface for coupled modeling is that the complexity of LSM physics rely heavily on diverse parameter sets corresponding to soil, vegetation, and other land-specific conditions and are not treated consistently across LSMs or even within the same community. The accuracies of these parameters on regional scales are strongly limited by their coarse resolution datasets and inability to capture local-scale heterogeneity in parameters such as soil hydraulic properties. As a result, attempts have been made to calibrate parameters based on observations of land surface conditions in order to ultimately improve prediction of state variables such as soil moisture (Santanello et al. 2007; Harrison et al. 2012). To date, LSM calibrations have been typically performed offline (uncoupled) and evaluated in terms of offline or 1-D (single-column) model predictions, and have shown promise in improving state and flux prediction based on an array of observed variables (Liu et al. 2003, 2004, 2005; Santanello et al. 2007; Peters-Lidard et al. 2008). The results of these calibration studies are highly specific to the model, resolution, parameter set, and region, however, so applicability and transferability to other offline or coupled models is strongly limited (Hogue et al. 2005).

Unifying the LSM spinup and calibration issues is the fact that, in essence, the atmospheric component of a coupled model is connected to the land solely through the fluxes. As a result, the atmosphere only responds and is sensitive to the turbulent (sensible, latent heat and shear stress or momentum flux) and radiative fluxes coming from the land surface at each timestep. From an atmospheric perspective, all the specificity and complexity of an LSM, including its parameters and the spinup approach, are hidden during the execution of a coupled simulation. A key question can therefore be asked: 'What is the potential impact of providing 'optimal' fluxes from the land surface to an atmospheric model versus those generated from default or coarse resolution initialization approaches?' The answer would provide insight as to the first-order influence of the land surface on accurate prediction of ambient weather (e.g. temperature, humidity, precipitation) as well as the behavior of particular scheme components (e.g. planetary boundary layer (PBL) height, convective initiation) in response to the optimal partitioning of surface fluxes. It would also provide a methodology to control for the inter and intra-LSM variability in spinup and parameterization approaches by focusing solely on providing the best lower boundary condition to the coupled system.

In this study, we address these questions using NASA's Land Information System (LIS; Kumar et al. 2006; Peters-Lidard et al. 2007). LIS supports a suite of LSMs under the generalized modeling framework and facilitates the ability to utilize diverse and high-resolution input data and data assimilation from local to global scales. The sensitivity of land surface spinups to methods and forcing data has already been addressed under this framework (Rodell et al. 2005; Kato et al. 2007). The recently developed LIS optimization and uncertainty estimation subsystem (LIS-OPT/UE) provides the ability to calibrate the LSM parameters (Kumar et al. 2012) and evaluate the impact of parameter uncertainties on LSM outputs (Harrison et al. 2012). Finally, the coupling of LIS and the Weather Research and Forecasting model (WRF-ARW; Skamarock et al. 2005) has been demonstrated in a number of land-atmosphere coupling studies

(Santanello et al. 2009; 2011a, 2012). For these reasons, LIS is an ideal platform from which to quantify the impact of LSM calibrations on coupled mesoscale prediction.

The focus of these experiments will be on LSM calibration over a range of surface conditions (dry to wet) in the U. S. Southern Great Plains (SGP) where the land is known to have a strong modulating impact on the atmosphere (Koster et al. 2004; Dirmeyer et al. 2006). In the process, these experiments will shed light on the following issues: 1) what to calibrate, 2) how to calibrate, and 3) when to calibrate. LIS-WRF will then be evaluated using coupling diagnostics already developed to simultaneously assess the land-PBL system as a whole in terms of water and energy cycling. Section 2 of this paper provides some background on recent land model calibration and spinup studies, as well as the coupling diagnostics developed to assess the land-PBL system. The model, LIS optimization and uncertainty subsystems (LIS-OPT/UE), and experimental design are then described in Section 3. Results are presented in Section 4, with discussion and conclusions on the role of the land surface in coupled prediction following in Section 5.

2. Background

a. LSM Spinup

Because in-situ and remotely sensed observations of soil temperature and moisture states or fluxes are not available at the resolution of a mesoscale model grid (horizontally or vertically), LSMs are used to produce flux and state estimates based on sound physics and constrained by forcing (based on traditional atmospheric meteorological data such as precipitation) and parameter data (based on static maps of vegetation and soil properties at high spatial resolutions). The practice of long-term spinup of offline LSMs to equilibrate soil moisture and temperature states has been in place for some time. Rodell et al. (2005) looked specifically at the sensitivity (and in turn, requirements) of equilibration to the length of the spinup run, which was found to vary based on climate regime (e.g. cold and dry regions tend to take longer to equilibrate than warm and moist locales) and soil type. They found that spinup time is typically more than 1 year, but no more than 3-4 years is required for most locations and conditions.

Spinup time has also been shown to be dependent on initial values of soil moisture, atmospheric forcing, and vegetation conditions (Yang et al. 1995; Chen and Mitchell 1999; Cosgrove et al. 2003; de Goncalves et al. 2006). Overall, LSMs use either manual or automated approaches to spinup based on reaching a minimum threshold of memory to the initial condition of the run (which can range from horizontally-uniform to climatologically-distributed). The particular threshold values are rather arbitrary, however, and have produced spinup times varying from a few weeks to over a decade in different studies. Also a factor is whether forcing data is available to run an offline LSM for the period leading up to the coupled simulation of interest, or whether cyclical data from a single annual cycle must be used to equilibrate the states (e.g. Cosgrove et al. 2003). For these reasons, the overall practice of spinup for coupled initialization has typically been inconsistent, leaving unanswered the question of the overall impact of LSM spinup on mesoscale prediction.

Recent case studies have been able to shed more light on this question, and, while limited in a quantitative assessment, do indicate specific impacts and improvements in coupled models as a result of improved specification of the land initial condition. Using LIS and LIS-WRF (described in Section 3), Kumar et al. (2008) found significant differences in prediction of fluxes, boundary layer structure, and temperature and humidity versus using default WRF initialization. Their studies also revealed improvements in precipitation forecasts using LIS-WRF due solely to the higher-resolution soil states from a long-term spinup run using LIS.

Following this work, Case et al. (2008) used LIS to show that spun-up initial conditions in LIS-WRF led to improved sea-breeze circulation and 2-meter temperature forecasts over Florida, particularly due to drier and more accurate soil moisture conditions generated by a 2-year spinup. Case et al. (2011) also investigated the impact of a LIS spinup on summertime precipitation simulated by LIS-WRF over the southeastern United States. They found that the near-surface soil moisture was improved, and that there was measureable impact and improvement of the spinup on the coupled near-surface and PBL conditions relative to that using the default land initialization via WRF. Small improvements were also seen in hourly precipitation forecasts that were initialized with a LIS spinup, but impact was limited due to the dominance of the atmospheric schemes in controlling these types of airmass-generated events.

In a similar vein to LIS, the High-Resolution Land Data Assimilation System (HRLDAS, Chen et al. 2007) was developed to provide improved land initialization for WRF simulations. Holt et al. (2006) and others have likewise demonstrated a large potential impact on coupled forecasts from using high-resolution (and assumed to be improved) representation of soil states and fluxes. They also show how the combined use of a spinup approach and mesoscale modeling can be used to simultaneously test and develop new LSM physics and parameterizations by evaluating both the impact on offline spinups and the coupled forecast. Trier et al. (2008, 2011) also used HRLDAS and WRF to show that the initial soil moisture for a coupled forecast is significantly more important than the evolution of soil moisture during a 1-2 week simulation. They also showed that sensitivity to the choice of LSM complexity could be minimized by calibrating the initial soil condition.

Using a different combination of land surface and atmospheric models, Di Giuseppe et al. (2011) analyzed three approaches used for initializing soils for mesoscale modeling. Their intercomparison of soil initialization using a) downscaling from a coarse resolution global parent model, b) results from a previous mesoscale coupled run, and c) nudging of soil moisture based on screen-level temperature observations indicate strongly that consistency in the physics and configuration between the offline and coupled models is paramount when choosing a source for initial values of soil moisture and temperature profiles. Therefore, the approach of using a previous run (i.e. spinup) of the same LSM to initialize the coupled forecast produced the best results, while the other two approaches were discouraged in practice. They also highlighted the importance of the soil temperature profile initialization (typically ignored in previous studies).

The impact of improved initialization of land surface states in WRF short-term prediction was also demonstrated by Wen et al. (2012). Although a spinup was not used, they updated the initial condition with in-situ observations of soil moisture and temperature and new land cover data measured from satellite and found significant impacts on all coupled components of the WRF simulation across a heterogeneous (dry/wet) region, including the atmospheric circulation enhanced by the surface conditions.

Overall, these studies have demonstrated an impact of LSM spinups on coupled prediction and are focused on short-term (diurnal) forecasts over mesoscale domains (1-10 km horizontal resolution), as will be the case performed here using LIS-WRF. Further, the consistent use of the same model and configuration to generate the soil initial conditions in the spinup and the coupled run is specifically what LIS and LIS-WRF has been designed for as a testbed, and follows with what these studies have suggested as best practice for maximizing the positive impact of the land on coupled prediction.

b. Calibration of Offline and Coupled LSMs

As mentioned, the physics of LSMs are highly dependent on specification of a large number of parameter values representing soil, vegetation, and other surface conditions. To simplify things, lookup tables are commonly associated to a particular soil or vegetation type that relates a number of parameters to each classification. Lookup tables are only as accurate as the available soil or vegetation information, however, and attempt to provide a representative value of each parameter for each soil or vegetation type. High-resolution maps that accurately capture the observed heterogeneity in parameter values are difficult to obtain on the scales of land surface and mesoscale models (particularly for regions outside the U. S. and on global scales), and there is little flexibility between soil or vegetation classes (e.g. for mixed crops or soil types). This can be a problem, particularly for soils where larger differences in soil parameters have been observed *within* a soil type than between types (Feddes et al. 1993; Soet and Stricker 2003; Gutmann and Small 2005; Santanello et al. 2007).

In order to combat these limitations, numerous attempts have been made to calibrate (or 'optimize') LSM parameters using observations of state variables such as soil moisture and surface temperature as constraints (Gupta et al. 1999; Hess 2001; Hogue et al. 2005; Liu et al. 2003, 2004, 2005; Santanello et al. 2007; Peters-Lidard et al. 2008; Harrison et al. 2012). Such approaches can improve matches of state variables to observations during the calibration period (and beyond), and in the process address LSM systematic biases. However, it remains difficult to derive parameter information that could be evaluated independently as most studies have focused on techniques that derive large sets of ‘effective’ parameters. Such studies also require a great deal of computational time and limit assessment of larger-scale applicability, and as a result little has been gained in terms of quantifying the effectiveness of calibrated parameters in improving coupled simulations.

For example, Hogue et al. (2005) investigated the transferability of large calibrated parameter sets in an offline LSM across varying surface conditions and time periods. They conclude that optimization should be site-specific for best results, and should be recalibrated for changes in seasons or over longer time intervals even if the surface and climatic features of the region remain the same. This suggests that if a spinup is to be used to initialize a coupled model, the calibration performed offline needs to be tailored (e.g. domain, resolution, LSM) specifically for the experiment of interest. In turn, this supports the idea that a testbed such as LIS and LIS-WRF is ideal for such investigations.

Liu et al. (2003) extended parameter estimation to coupled systems by examining the pathways by which limitations in the LSM physics impact both offline and 1-D (single-column) model simulations. Their results show that offline calibration is well-constrained due to the realistic forcing applied and is able to identify and correct deficiencies in evaporative physics, but in coupled mode some parameter sets acted to amplify flux errors due to occurrence of land-atmosphere feedbacks. Liu et al. (2004) and (2005) then included atmospheric parameters in the calibration, and highlight the computational difficulty in calibrating large parameter sets in coupled models (which has precluded the calibration of a full 3-D mesoscale model to date). As a result, they suggest a stepwise procedure of offline before coupled calibration as an alternative. Overall, their results found that calibrated parameter values are particularly sensitive to the surface latent heat flux as the lower boundary condition, and all simulations were found to be highly sensitive to the initial soil moisture value (prescribed uniformly in their study rather than spun up), stressing the importance of an accurate LSM spinup for coupled simulations.

Overall, these studies have highlighted that the land initialization for coupled models is important, and that the methodology of an offline spinup with calibrated parameters shows promise in providing the most accurate initial condition consistent with best surface physics and parameterizations. Performing fully coupled (3-D) land surface and atmospheric parameter calibration remains a daunting task, but we are now in a position to quantify the impact of an optimal *and* physically meaningful LSM spinup for coupled prediction models.

c. Evaluation of Land-Atmosphere Coupling

The quantification of land-atmosphere interactions in coupled models is a complex task that involves a great number of processes and feedbacks. For example, in terms of accurately representing the relationship between soil moisture (SM) and precipitation (P) in coupled models, a full understanding will only come by careful examination and quantification of a series of interactions and feedbacks (i.e. 'links in the chain') that can be summarized as follows (from Santanello et al. 2011a):

$$\Delta SM \underset{\textbf{(a)}}{\rightarrow} \Delta EF_{sm} \underset{\textbf{(b)}}{\rightarrow} \Delta PBL \underset{\textbf{(c)}}{\rightarrow} \Delta ENT \underset{\textbf{(d)}}{\rightarrow} \Delta EF_{atm} \;\blacktriangleright\; \Delta P/Clouds \qquad (1)$$

where EF is the evaporative fraction, defined as

$$EF = \frac{Qle_{sfc}}{Qh_{sfc} + Qle_{sfc}} \quad . \qquad (2)$$

and is a function of the sensible (Qh_{sfc}) and latent (Qle_{sfc}) heat fluxes at the land surface. From Eq. 1, the impact of soil moisture (ΔSM) on clouds and precipitation (ΔP) is therefore dependent on the sensitivities of: **a)** the surface fluxes (EF_{sm}) to soil moisture, **b)** PBL evolution to surface fluxes, **c)** entrainment fluxes at the PBL-top (ENT) to PBL evolution, and **d)** the collective feedback of the atmosphere (through the PBL) on surface fluxes (EF_{atm}) (Santanello et al. 2007; van Heerwaarden et al. 2009). As a result, there are numerous pathways composed of positive and negative feedback loops in this chain, including the influence of additional inherent and external factors (e.g. canopy interception, large-scale convergence).

The initial communication between the land and atmosphere occurs on local scales, and therefore a community effort has been launched to diagnose and quantify local L-A coupling in coupled models, called 'LoCo' (Hurk and Blythe 2008; Santanello et al. 2009, Santanello et al. 2011b). The realm of LoCo has been defined by GLASS as "*The temporal and spatial scale of all land-surface related processes that have a direct influence on the state of the PBL*". Therefore, the fundamental processes that fall into this realm correspond directly to the question of the role of offline LSM spinup on coupled mesoscale prediction. This research is a core component of the Global Energy and Water Cycle Study (GEWEX) Land Atmosphere System Study (GLASS; Hurk et al. 2011), which coordinates community working groups and intercomparison studies related to offline and coupled land surface modeling. A thorough review of LoCo research and the related diagnostic framework can be found in Santanello et al. (2009, 2011a, 2011b, 2012; hereafter referred to as S09, S11a, S11b, S12).

LIS and LIS-WRF have served as a core testbed to develop and implement LoCo diagnostics utilizing the range of LSM and PBL scheme options available in each. Under this framework, a methodology that simultaneously addresses the components of Eq. 1 was tested by S09 and extended by S11, and employs the 'mixing diagram' approach as introduced by Betts (1992). This power of this diagnostic lies in its ability to exploit the co-variance of 2-meter potential temperature (θ) and humidity (q) to quantify the components of the LoCo process-chain, and is based only on routine variables that can be applied to any model or observations and across a range of scales. From this analysis, the full PBL budgets of heat and moisture, relationship of EF to PBL height (PBLH), and the evolution of the lifting condensation level (LCL) deficit (PBLH minus LCL) can be derived and used to understand the nature of and sensitivity of a particular land-PBL coupling. For a full description of this approach and implementation for LoCo studies, the reader is again referred to S09 and S11a.

The LoCo approach diagnoses the land and PBL fluxes simultaneously, and therefore provide the components of the full budgets of heat and moisture in the coupled system. LoCo diagnostics can therefore be used to quantify the joint evolution of coupled variables, such as those that showed strong sensitivities in earlier studies, but only independently (e.g. θ and q in the work of Trier et al. (2008)). As shown in S09 and S11a, how anomalies and/or errors in the surface fluxes computed by a particular LSM-PBL coupling are then translated into the atmospheric water and energy cycle can then be quantified using this approach. Differences in soil moisture differences strongly impact the signatures of heat and moisture evolution and diagnosis of coupling behavior. For example, results from S12 during dry/wet extremes show that the choice of LSM is critical for dry regimes, but that both PBL and LSM are comparable influences on the coupled behavior during wet regimes. LoCo diagnostics are therefore well-suited to capture the first-order impact of land spinup and specification on the PBL and atmosphere as a whole.

3. Model and Site Description

a. LIS and LIS-OPT/UE

NASA's Land Information System (LIS) consists of a suite of LSMs under the same software framework and provides a detailed representation of land surface physics and states, which can then be directly coupled to an atmospheric model. More recently, new subsystems have been added to LIS that allow sophisticated optimization and uncertainty estimation (LIS-OPT/UE) algorithms to be applied to the LSMs to exploit further the information content from observations. The algorithms (e.g. Levenberg-Marquardt (Levenberg 1944; Marquardt 1963), Genetic Algorithm (Holland 1975), Shuffled Complex Evolution from University of Arizona (Duan et al. 1993)) calibrate the model parameters to the remote sensing observations, thereby enabling improved model forecasts and enhancing the efficiency of data assimilation approaches (Santanello et al. 2007, Peters-Lidard et al. 2008, Kumar et al. 2012a). The uncertainty estimation subsystem also includes Bayesian approaches based on Markov Chain Monte Carlo (Gilks et al. 1996) to estimate the uncertainty in model parameters given calibration datasets, which enables probabilistic prediction.

Overall, the high-performance computing infrastructure in LIS provides an advantage over previous parameter estimation studies which were limited to trial and error, manual, and lower-dimensional (i.e. smaller parameter sets) calibration approaches, and have been demonstrated by Kumar et al. (2012) and Harrison et al. (2012) for offline spinup and data assimilation applications. The evaluation of offline, coupled, and LIS-OPT/UE experiments is performed using a LIS-based tool called the Land surface Verification Toolkit (LVT; Kumar et al. 2012b). LVT provides a standardized platform for intercomparing model output (from LIS or other sources) with observations and offers a range of statistical and benchmarking approaches.

b. NU-WRF

Derived from the Fifth-Generation NCAR/Penn State Mesoscale Model (MM5; Anthes and Warner 1978), WRF-ARW has been designated as the community model for atmospheric research and operational prediction and is ideal for high-resolution (e.g. 1-10 km) regional simulations on the order of 1-10 days. WRF-ARW has a Eulerian mass dynamical core and includes a wide array of radiation, microphysics, and PBL options as well as 2-way nesting and variational data assimilation capabilities.

Recently, work has been performed to develop a NASA-Unified WRF (NU-WRF; https://modelingguru.nasa.gov/community/atmospheric/nuwrf) modeling system at NASA's Goddard Space Flight Center (GSFC). NU-WRF is built upon the WRF-ARW model, and incorporates and unifies NASA's unique experience and capabilities by fully integrating LIS, the WRF/Chem enabled version of the Goddard Chemistry Aerosols Radiation Transport (GOCART; Chin et al. 2000) model, GSFC radiation and microphysics schemes, and the Goddard Satellite Data Simulation Unit (SDSU; Matsui et al. 2009) into a single modeling framework. In turn, NU-WRF provides the modeling community with an observation-driven integrated modeling system that represents aerosol, cloud, precipitation and land processes at satellite-resolved scales.

The land-atmosphere coupling is a core component of NU-WRF, and has been performed through the coupling of LIS and WRF by Kumar et al. (2008). The advantages of coupling LIS and WRF include the ability to spin-up land surface conditions on a common grid from which to initialize the regional model, flexible and high-resolution (satellite-based) soil and vegetation representation, additional choices of LSMs that continue to expand in range and complexity, and direct coupling of the atmospheric model to the LIS subsystems (including LIS-OPT/UE). The work of S09, S11a, and S12 has demonstrated NU-WRF as a testbed for L-A interaction studies and LoCo due to its land-PBL scheme flexibility and high resolution. Hereafter we refer to NU-WRF as the coupled prediction system that includes the LIS-WRF coupling for these experiments.

The continuous development and support of NU-WRF ensures that the most recent versions of LIS (currently V 6.2) and WRF-ARW (currently V 3.2) are coupled and tested, and are used in this study. The experiment described below are run on a single 500x500 domain at 1 km spatial resolution (see below), and include a 5-second timestep, GSFC microphysics, longwave, and shortwave radiation, and the Monin-Obukhov surface layer scheme. The North American Regional Reanalysis (NARR; Mesinger 2006) data was used for atmospheric initialization and lateral boundary conditions using 3-hourly nudging, and the vertical resolution of NU-WRF was specified as 43 vertical levels, with the lowest model level ~24m above the surface.

The LSM employed in LIS for this study is the Noah LSM (Noah; Ek et al. 2003), and was originally developed from the land component of the Oregon State University 1-D PBL model (Troen and Mahrt, 1986). The Noah model employed in this study is Version 3.2 and is identical to the version of Noah packaged in the original version of WRF-ARW Version 3.2. Noah is used operationally by the National Center for Environmental Prediction as the LSM for the North American Mesoscale (NAM) model and the Global Forecasting System (GFS). As such, Noah is a well-supported, developed, and utilized LSM for both offline and coupled applications. Particularly important for the LIS-OPT/UE calibration (see below), the soil type specification in LIS is based on the STATSGO (Miller and White 1998) database over the U. S., while vegetation type is assigned based on the UMD landcover dataset (Hansen et al. 2000).

The PBL scheme chosen in NU-WRF is the Yonsei University (YSU; Hong et al. 2006) PBL, based on non-local K theory and includes explicit treatment of entrainment and counter gradient fluxes. The combination of Noah LSM and YSU PBL is a common selection in WRF, and has served as the default configuration for test cases involving NU-WRF. While some results suggest other PBL schemes in WRF sometimes perform better than YSU under certain conditions (e.g. stable/nighttime periods), a universally-accepted hierarchy of PBL scheme usage has not been developed as of yet and it is beyond the scope of this study to engage in further study of PBL and LSM scheme sensitivities (which can be seen in S12).

c. 2006-7 Dry/Wet Extremes

The SGP region has been identified as a hotspot for land-atmosphere coupling in terms of the strength of interactions and impact of soil moisture anomalies on clouds and precipitation (e.g. Koster et al. 2004). Because of this, and the large record of observational data from the Atmospheric and Radiation Measurement testbed (ARM-SGP), S09, S11a, and S12 have focused WRF studies on the SGP region to develop and test the LoCo diagnostics described in Section 2c. In particular, S12 looked at the extreme conditions observed during the 2006-7 period and the impact on LoCo. Low anomalies of clouds and precipitation in 2006 (October-September) were immediately followed by conditions of high cloudiness and rainfall in 2007, with 2006 being the second driest and 2007 the seventh wettest year on record. This period was followed by a relatively normal summer season in 2008, with soil moisture conditions in between that of the 2006 and 2007 extremes (as confirmed by ARM-SGP observations and Noah simulations).

As described in S12, ideal case studies were chosen for each regime. The 14-20 July 2006 experiment consists of a lengthy dry-down period with little synoptic disturbance in which the land was free to interact and evolve with the atmosphere on primarily local scales. The case study of 14-20 June 2007 focuses on a period with scattered precipitation every 1-2 days in portions of the ARM-SGP domain, interspersed with brief dry-downs in which conditions were clear and/or cloudy and culminating in a large mesoscale convective system (MCS) traversing the domain on the final nighttime period.

d. Experimental Design: Default Spinups

Forcing data from the North American Land Data Assimilation System (NLDAS-2; Xia et al. 2012) project was used to drive the spinup simulations. Noah was run offline in LIS beginning 1 January 2003, thus producing a ~3.5-4.5 year spinup prior to the start time of the 2006 and 2007 case studies. This is longer than the recommended spinup length based for similar moisture regimes (soil and precipitation), and is consistent with previous studies using this LSM, location, and time period (S09, S11a, S12) in ensuring a fully-equilibrated soil condition that is insensitive to the initial condition of the spinup (horizontally homogeneous in this case).

Using the resultant spun-up surface fields as initial conditions for the 2006-7 case studies, NU-WRF simulations were then performed over a single high-resolution 1km domain centered over Oklahoma and Kansas. Figure 1 shows the upper layer (0-10 cm) soil moisture values over the ARM-SGP domain as generated by Noah spinups valid at 00Z on 1 July 2006, 2007, and 2008. The advantages of using LIS for this purpose are evident in the high spatial resolution seen in Fig. 1 as a reflection of the inputs of vegetation and soil properties. Soil moisture varies significantly from dry and heterogeneous (generally < 25 percent volumetric) in 2006 to extremely wet (near saturation) and more uniform conditions in 2007, with 2008 showing more moderate soil moisture and heterogeneity.

e. Experimental Design: LIS-OPT/UE Case Studies

The offline calibration experiments were performed using the GA algorithm in LIS-OPT/UE, and applied to a set of 29 parameters describing soil, vegetation, and general characteristics in the Noah model (Table 1). The goals of calibration are to provide the best possible surface fluxes for NU-WRF simulations. Therefore, the observations employed are measurements of sensible (Qh), latent (Qle), and soil (Qg) heat fluxes from the ARM-SGP network of sites over the domain, including 6 Energy Balance Bowen Ratio (EBBR; Qh, Qle, and Qg) and 12 Eddy Correlation (ECOR; Qh and Qle only) tower locations. The GA was applied using an objective function that minimizes RMSE with no discrimination of flux type (i.e. Qh, Qle, and Qg flux observations are weighted equally). The calibration was performed over the periods of 1 May – 1 Sept of 2006, 2007, and 2008 to produce separate calibrated parameter sets for the dry, wet, and normal regimes. Having three separate calibration periods allows for the study of the impact of calibration period and varying atmospheric and land surface conditions on the calibration results.

The number of observations of Qle, Qh and Qg that used in the GA optimization are comparable, but vary slightly from 2006 (Qle: 48546, Qh: 48822, Qg: 32218) to 2007 (Qle: 37936, Qh: 39063, Qg: 30100), and to 2008 (Qle: 45767, Qh: 48353, Qg: 31344). As a result, the objective function is skewed towards the fluxes with the greater number of observations in each case and therefore is weighted more heavily towards Qh and Qle than Qg. The GA integrations use a population size of 50 and employ an elitism strategy to ensure that the current best solution is not overwritten during GA evolution, with a mutation rate of 0.005 and a recombination rate of 0.9. The GA parameters (including the mutation and recombination rates) are chosen largely from experience and the success of the optimization simulations in Kumar et al. (2012). The algorithm was found to converge after approximately 200 generations, when the fitness of the best solution was found not to improve in the last 30 generations.

From these simulations, a unique calibrated value of each of the 29 Noah parameters was obtained at each of the 18 grid cells pertaining to the flux sites. In order to obtain calibrated values covering the full model domain, the values from each site then were grouped and averaged by common vegetation and soil types and assigned to the full domain based on the vegetation and soil classification at each grid cell. Note that Noah parameters were designated into soil (15 parameters, 5 classes in the SGP domain), vegetation (11 parameters, 3 classes in the SGP domain), and general (3 parameters, no classification) categories as based on their functionality and most direct impact on the model physics. For example, for a soil-related variable such as porosity, the calibrated values of porosity from each flux site with a 'clay' classification were averaged, and then applied as the porosity value to the remainder of the domain where 'clay' was also the soil type. Also, if a soil/vegetation class occurs in the domain but was not represented at one of the observation sites, default table values are used. General parameters are constant across the domain and do not have a classification, and therefore were averaged across all the sites.

Using the calibrated parameters, new soil, vegetation, and general lookup tables for Noah were then generated. Spinup runs (as described in the previous section for the default case) were repeated using the new tables based on the 2006, 2007, and 2008 calibration results, thereby producing spun-up and initial conditions that are optimized for dry, wet, and average conditions, respectively, over this region. To examine the impact of calibrated spinups on coupled forecasts, four targeted NU-WRF case studies were then chosen from the larger 7-day periods described above, with characteristics as follows:

- 14 July 2006: 24 hours, dry regime; NU-WRF test case
- 18-19 July 2006: 48 hours, dry regime; peak of dry-down
- 16-17 June 2007: 48 hours, wet regime; limited/scattered precipitation
- 19-20 June 2007: 48 hours, wet regime; scattered/MCS precipitation

NU-WRF was then run for each case study above using four different combinations of parameter values/lookup tables, as shown in Table 2. The array of simulations was designed to capture the impact on NU-WRF forecasts from using a combination of a) default spinup (uncalibrated) and default parameters in the coupled run (DEF), b) default spinup with calibrated parameters in the coupled run (CPL), c) calibrated spinup with default parameters in the coupled run (SPN), and d) calibrated spinup with calibrated parameters in the coupled run (SCP). Note that the focus of the results presented here will be on the differences between the DEF (no calibration) and SCP (fully calibrated) cases, but CPL and SPN offer the ability to parse out the relative impacts of using optimal parameters during the spinup vs. coupled simulation period, and will be included in the discussion when relevant.

f. Observation Data

The ARM-SGP program provides a long-standing record of quality-controlled surface flux, meteorological, and hydrological observations along with atmospheric profiles for a network of sites across the domain shown in Fig. 1. This includes co-located soil moisture, net radiation, sensible, latent, and soil heat, along with co-located surface meteorology data that provide the full set of variables needed to calculate the LoCo diagnostics discussed in Section 2c and evaluate against model results. For the calibration experiments, ARM-SGP data was collected from ECOR and EBBR towers, and the LoCo evaluation was performed using the co-located surface meteorology, flux towers, and available radiosonde profile data.

4. Results

The performance of the offline calibration experiments will be evaluated first, followed by the impact of spinup calibration and initialization on NU-WRF predictions and LoCo, the sensitivity of the coupled results to the period of calibration, and concluding with the uncertainty introduced into the forecasts by different parameter sets.

a. Offline Calibration

Before examining the coupled cases, it is important to quantify the impact of the calibrated parameters on the offline spinup. Figure 2 shows the flux components simulated using default and calibrated Noah parameters during the dry regime (2006) versus observations at each of the ARM-SGP sites and over the full domain. Both Qh and Qle show improvement at nearly all sites, with RMSE values reduced by up to 25.7 Wm-2 (10.5 Wm-2 on average) in Qle, and up to 45.3 Wm-2 (19.1 Wm-2 on average) in Qh. Note that the 95 percent confidence interval for the domain averages are ~ 4-7 Wm-2, so the improvements are statistically significant. The improvement due to the calibration is also clearly evident in the mean diurnal cycle behavior of Qh and Qle across all sites. Focusing on the daytime when the turbulent fluxes are large and positive, Qh matches observations almost exactly and improves over the high bias present in the default simulations. Analogously, daytime Qle increases due to calibration and matches observations more closely than when default parameters are used in Noah. The Noah model has often been shown to produce systematic over/underestimation of surface fluxes, and the GA calibration successfully improves upon the biases exhibited for the SGP and study period demonstrated here.

Qg shows more mixed results, with 5 of the 11 EBBR sites showing slight degradation after calibration, but the magnitudes for Qg are small overall and this does not present a concern for this study. The mixed results are partially a reflection of the reduced number of observations of Qg available for the GA and the heavier weighting towards Qh and Qle. In addition, phase errors in Qg are well documented (Robock et al. 2003, Reichle et al. 2010) and could possibly be corrected if joint calibration approaches including soil temperature and Qg were conducted.

Figure 3 shows the offline calibration results for the wet regime (2007), and once again Qh and Qle are improved at nearly all ARM-SGP sites (and in the case of Qh, all sites show improvement). In this calibration, Qh improvements are more modest than in 2006 (up to 25.9 Wm-2 and 12.3 Wm-2 on average), while Qle improvements are larger than during the dry regime (up to 54.9 Wm-2 and 12.3 Wm-2 on average). Interestingly, Site E24 shows the largest improvement in this case, opposite of the 2006 calibration. The mean diurnal cycles show marked improvement (decrease) in daytime Qle over the default simulations, while Qh is only very slightly impacted (and also decreased). This suggests an available energy bias and overestimation in the offline Noah runs in 2007. Once again, Qg shows mixed results as 5 of 11 sites show degradation; though in this case there is a noticeable increase in Qg after calibration that improves afternoon simulations, but does not impact the phase error where Qg peaks too early (as in the 2006 case).

Overall, the largest impact and improvement due to calibration of Noah is seen in Qh in 2006 and in Qle in 2007. Physically, this can be explained by the fact that during the dry regime, Noah has a dry bias and produces too little evaporation thereby overestimating Qh. In the wet regime, Noah has a wet bias and produces too much Qle (partially due to too much net radiation). The LIS-OPT/UE calibration has thus adjusted the parameter values accordingly, to correct for the dry bias in 2006 by increasing soil moisture and modifying the efficiency of the evaporative physics in Noah (and vice-versa in 2007) that compliments the new soil moisture levels to produce the optimal fluxes. These results are also consistent in that during a dry regime which is water-limited, the primary adjustment in fluxes would be towards the higher magnitude flux (Qh), and during a saturated regime the largest impact would be felt in Qle.

b. Coupled Simulations

In order to assess the impact of offline LSM calibration on the coupled system, LoCo diagnostics are used to simultaneously evaluate the land (LSM) and atmospheric (PBL) component evolution and interaction.

1) 14 JULY 2006

The mixing diagram analysis for the 14 July 2006 case at the ARM-SGP E4 site is shown in Fig. 4. Focusing first on the comparison of the DEF and SCP simulations, it is shown that the default Noah parameters produce the poorest simulation of heat and moisture states and fluxes in NU-WRF. Visually, the DEF curve is drier (and slightly warmer) than observed throughout the daytime period. This is improved significantly in the SCP simulation which matches closely with observed T2 and Q2 throughout. Table 3 provides error statistics of simulated versus observed T2 and Q2 co-evolution, and because mixing diagrams are in energy-space these can be represented in units of J kg-1 and used to describe a total RMSE and MAE of heat and moisture combined (i.e. quantifying the spatial differences between the model and observed curves in Fig. 4). These metrics confirm that the DEF run performs worst of all the simulations, while the SCP improves all aspects of the temperature and moisture states (T2 and Q2) by 15-26 percent in RMSE and 8-30 percent in bias.

The fluxes in the coupled system can be evaluated via the Bowen and entrainment ratios (as defined by S09 and in Fig. 4). As expected, SCP produces a β_{sfc} (=Qh_{sfc}/Qle_{sfc}) nearly identical to that observed due to the calibration to surface fluxes performed, which produced the parameters used in the SCP simulation. DEF overestimates β_{sfc}, consistent with the dry bias observed in the offline spinup and the coupled T2 and Q2 results. The entrainment fluxes (as reflected by Bent) are also impacted by the LSM calibration by ~15 percent and slightly closer to observations. Likewise, the heat and moisture entrainment ratios (Ale and Ah) show substantial improvement in SCP over default, where the higher Qle and lower Qh as a result of correcting the dry bias at the surface produce better ratios of land to PBL fluxes.

Focusing on the remaining two simulations, CPL and SPN, indicates how calibrated parameters impact coupled simulations when used in either offline spinups or the coupled run only. It is first evident that SPN does well with T2 and Q2 state estimation, correcting the dry bias of Noah, and producing the best overall error metrics in Table 3. The fluxes of SPN are severely overcompensated, however (e.g. β_{sfc} very low), and produce too much evaporation. Because the calibrated parameters in this simulation are used only for the spinup, these results indicate that the default parameters still employed in the coupled run produce too high of evaporation rates for the given initial soil moisture state. The CPL simulation performs poorly both in terms of T2 and Q2 (with comparable or worse metrics in Table 3 to the DEF simulation) and surface and PBL fluxes, indicating that using calibrated parameters only for the coupled simulation along with a default spinup does not impact or improve the coupled forecast at all. These results are also consistent with those of Trier et al. (2008), who showed that initial soil moisture (i.e. fluxes calibrated in SPN) has a much larger influence on forecasts than the evolution of soil moisture during the coupled run (i.e. fluxes calibrated in CPL).

The full heat and moisture budgets of the coupled system can be derived from the mixing diagram analysis and are shown in Fig. 5. The calibration of the surface fluxes to observations in SCP is most evident, as is the overestimation of Qle and Qh in the SPN and CPL simulations, respectively. Less impact of different calibration approaches is seen in the PBL components of the budget, where all are relatively close to observed. The total budgets do, in turn, directly reflect the improvement of surface fluxes in the SCP and SPN simulations.

Another related diagnostic of the coupled system performance is the relationship of evaporative fraction (EF) and PBL height (PBLH), as shown in Fig. 6. Once again, the best combination of land and atmospheric behavior is exhibited by the SCP simulation, which closely matches both the EF (which integrates the land surface condition) and PBLH (which integrates the atmospheric response). SPN and CPL are the extremes in terms of EF and PBLH, while the dry bias in the DEF simulation is evident and leads to slightly higher PBL growth.

From the full suite of simulations and diagnostics in Figs. 4-6 and Table 3, it is clear that offline LSM calibration can improve coupled simulation components significantly and in a consistent fashion in terms of correcting a bias and the impact of that correction (e.g. soil moisture) on the coupled components (e.g. T2 and Q2). It is also evident that employing calibrated parameters in both the offline spinup and the coupled run is required to achieve optimal improvement in coupled prediction. It is the combination of a spinup produced with calibrated parameters that support a wetter initial condition along with those same parameters that support lower evaporation rates in the coupled simulation that are actually compensatory. Therefore, if the calibrated parameters are only used in either the spinup or coupled run, significant and overreaching impacts will be seen in the prediction of coupled states and/or fluxes (as seen in SPN and CPL).

A robust measure of the impact of LSM spinup and calibration on NU-WRF simulations can be found in the performance of T2 and Q2 across the entire model domain. Figure 7 shows the domain average statistics computed using the Model Evaluation Tools statistical software package (MET; developed by the National Center for Atmospheric Research (NCAR): www.dtcenter.org/met/users/docs/overview.php and incorporating NCEP Automated Data Processing (ADP) atmospheric and surface data), and based on 214 site observations at 6-hourly intervals on 14 July 2006 which provides a true independent evaluation of the model. In particular, the RMSE and Bias statistics are largely improved in SCP versus DEF and are consistent in terms of lowering the dry/warm bias of the default simulation. Also plotted are the results from a NU-WRF simulation that does not use LIS nor a spinup of the Noah LSM (as a true 'off the shelf' WRF-default case comparison). Overall, by introducing a spinup (DEF vs. WRF) there is a definite increment of improvement over a default or coarse atmospheric-based initial condition (e.g. NARR in this case). Performing offline calibration for a spinup then increases the accuracy of the simulation even further (SCP vs. DEF vs. WRF). Likewise, the land surface energy balance (Qh, Qle, and Qg) components across the entire suite of 19 ARM-SGP sites are shown in Fig. 8, where improvement is seen across the board in terms of reducing the RMSE and Bias. Overall, these results provide strong evidence that spinup and calibration improves coupled forecasts across the entire NU-WRF domain, as well as the individual site details shown in Figs. 4-6.

2) 18-19 JULY 2006

The other dry regime case study results are shown in Fig. 9 and Table 4. As the dry-down has progressed over the period, there is a larger diurnal range in 2m temperature observed (~20K) than the 14 July case (~13 K), while the humidity ranges are comparable on 18 July but reach a much drier condition on 19 July as the surface begins nears desiccation. On both days in Fig. 9, the DEF simulation shows a more extreme dry bias now versus observations, as reflected in Q2 and the surface Bowen ratio. Despite this, the calibration in SCP still produces consistent improvement in heat and moisture states and fluxes, particularly on 18 July. β_{sfc} on 19 July is observed to be much higher than the previous day, and supports a sharp diurnal decrease in Q2 due to lack of surface evaporation (and is similar to the mixing diagram signature seen in the dry soils results of S09 and S11). Overall, the SPN simulation (not shown) produces the lowest T2 and Q2 errors, but as was the case for 14 July this occurs for the wrong reasons, as β_{sfc} is vastly underestimated while CPL remains close to the DEF results.

That SCP doesn't match or improve Bsfc observations as well as the previous cases is because the overall nature of the calibration is to correct the dry bias in Noah thereby increasing the soil moisture and Qle. The calibration works well overall, but for extreme conditions like on 19 July the DEF simulation just so happens to produce better β_{sfc} due to its inherent dry bias. The limits of calibrating the spinup are also evident here, as the shift due to higher initial soil moisture is felt in the coupled simulation to the degree of the shift in DEF to SCP curves, and suggests there is still significant uncertainty and limitations in LSM physics that prevent even a detailed calibration of large parameter sets from improving upon.

3) 16-17 JUNE 2007

The wet regime cases show a vastly different signature in the mixing diagrams that is reflective of much higher evaporation rates at the surface and limited PBL growth and entrainment above. Fig. 10 and Table 5 show that the DEF simulations generally perform well relative to observations in terms of T2 and Q2 evolution, and that there actually is some degradation in results after calibration on 16 June (note that the calibration performed for these cases was appropriately based on the 1 May- 1 September 2007 period). The degradation/improvement seen in T2 and Q2 in the SCP simulation on June 16/17 is due to the DEF simulation being too wet/dry on these days, and due to the dry bias correcting nature of the calibration has a positive impact only on the day when an initial dry bias exists.

Overall, there is very little impact of using calibrated vs. default parameters, though the patterns are consistent in that CPL performs worst and SPN performs best in terms of T2 and Q2 metrics. The calibration does improve β_{sfc} in SCP over DEF and very close to observations, as designed by the calibration. There is not any translation of this improvement to the PBL fluxes or 2m states, however. This is consistent with the results of S12, who showed that the impact of a particular LSM is dampened during wet regimes when the PBL scheme and atmosphere-dominated regime takes over. It can also be summarized that when the LSM and coupled model perform well (as 16 June MAE, RMSE, Bias, and N-S metric suggest), there is little to be gained in calibrating large sets of parameters because the inherent predictability in the system has already been maximized.

4) 19-20 JUNE 2007

At the end of the wet regime, much poorer performance is seen in both the DEF and SCP simulations (Fig. 11 and Table 6) in terms of the diurnal evolution of T2 and Q2. Particularly on 19 June when DEF has a wet bias in the morning, there is degradation across all metrics (with the exception of the Q2 bias), which is again consistent with the calibration attempt to correct the overall dry bias that is not evident on this particular day. As also evident from the comparisons of all the case studies thus far, there is a noticeable shift on 19 June to a very wet regime (high Q2) that is reflective of frequent precipitation events in the days prior (including the passage of a MCS over the study region).

20 June is much similar to 16-17 June in that there is very little impact of calibration on the results. Overall, the wet regime is dominated by low β_{sfc} and relatively high Qle, along with lower net radiation (due to clouds and precipitation), and reduced PBLH, entrainment, and diurnal cycles of T2 and Q2. This makes the potential impact from LSM adjustments (such as calibration, spinup and initialization approaches) on the coupled system much lower than in the dry regime. In addition, the attempt of calibration to systematically reduce inherent LSM biases

works least well for the extremes of regimes (e.g. just after frequent rainfall; end of a severe dry-down) as opposed to the more benign, moderate, and transitional periods (as reflected in the overall offline and domain-average results presented above).

c. Period of Calibration

The second part of this analysis addresses the question of 'what is the impact of the period of calibration on coupled predictions?'. The 2006 case studies above were performed using parameters calibrated during summer 2006 period, and the 2007 cases with parameters calibrated during 2007. For broader applicability of this methodology, it is important to address the impact of data availability and limitations on the calibration. For example, if observed fluxes are only available for a limited time, certain year, or season (as is often the case for field experiments) that does not coincide with the forecast period of interest there likely will not be as optimal results seen in the offline calibration or coupled simulations.

Table 7 lists the experiments conducted to determine the impact of having observations only during dry, wet, or average years, or having all three years available. These simulations are each conducted using calibrated parameters in the spinup and during the coupled run, and therefore C06 is identical to SCP in Figs. 4 and 9, C07 is the same as SCP in Figs. 10 and 11, and DEF is the same as in all previous analyses.

The land surface energy balance components for the 2008 offline calibration are shown in Fig. 12. Improvement in RMSE of Qle and Qh is seen at all but 3 and 5 sites, respectively, but to a much lesser degree overall (~5-10 Wm-2) than was seen in 2006 and 2007. Likewise, the impact of calibration on the diurnal cycle fluxes is very small, particularly for Qle (which is already simulated quite well by default), although Qg shows more impact and degradation during daytime than either 2006 or 2007.

The results for the offline calibration using all three years of data (2006, 2007, and 2008) combined are then shown in Fig. 13. Once again, the GA algorithm performs well in improving the flux components nearly at nearly all sites (with the exception of only 2 in Qle and Qh), and overall improvement in RMSE is on the order of 15-20 Wm-2. The diurnal cycles show marked improvement in both Qle and Qh, nearly matching observations in each and lowering the daytime magnitude of each. Some degradation is seen in Qg where it is overestimated during the daytime, therefore compensating somewhat for the reduction in Qh and Qle.

The 14 July 2006 case study results for the suite of simulations with different year calibrations are shown in Figs. 14-16 and Table 8. DEF and C06 are the same as in Fig. 4, but what is now evident is the spread in results introduced by different calibration periods. C07 performs nearly as well as C06 despite that this is a 2006 case (Fig. 14), with both the T2 and Q2 evolution and error metrics almost identical (Table 8). The similarity of C06 and C07 follow in the PBL budget (Fig. 15) and EF vs. PBLH analysis (Fig. 16) as well. The worst performing simulation by far is that with the calibrated parameters from the average year (C08), which is too dry and significantly overestimates β_{sfc} as a result (low Qle, high Qh). This translates into entrainment and total PBL budgets that are too large in Fig. 15, and reflected in low EF and large PBL growth in Fig. 16. The calibration using all three years of data (C678) generally performs well, but less so than either C06 or C07 which is as expected given the performance and weighting of the individual years.

These results suggest that calibration using observations that capture the dry and wet sides of the soil moisture distribution is critical to coupled prediction improvement. Similar results are also seen for the 18-19 July 2006 case study (ranked as C06, C07, C678, C08 from most to least improvement), and similar mixed/limited impacts seen in the 2007 cases. This may be due to the calibration correction of the Noah dry bias through the new parameter sets, but only is possible during extreme conditions when the model biases are significant. It is also an important result that using 'average' calibrated parameters (C08) during an extreme condition actually degrades the coupled results due to a now slightly drier soil moisture condition and less evaporative Noah overall (thus enhancing the bias).

d. Uncertainty Propagation

An interesting question that is inherent in parameter estimation studies is how to quantify the sensitivity of LSMs to calibrated parameter sets generated by algorithms such as GA. In a similar vein, tools have been developed for LIS-OPT/UE that can be extended to quantify how uncertainty in LSM spinups and initial conditions is translated to coupled forecasts. To address this issue, an additional suite of simulations was conducted using a simple Monte Carlo simulation (MC-SIM) sampling algorithm implemented in LIS-OPT/UE in order to propagate uncertainty from inputs (e.g. soil, vegetation, and general parameters) to model outputs (e.g., offline spinup, coupled prediction). As such, this algorithm allows for an assessment of LSM uncertainty, and can be used to gauge the relative sensitivity of the coupled system to LSM inputs. A small sample size (5) was applied given that WRF does not have a true ensemble mode, and essentially requires independent integrations for each set. As in Kumar et al. (2012), uniform distributions were applied to all parameters given the limits of the ranges also based on Kumar et al. (2012). The result is a sense of the spread in simulations prior to calibration.

Figure 17 shows the results of the DEF and C06 simulations (as in Fig. 14) for the 14 July 2006 case, along with the simulations using the 5 parameter sets sampled with MC-SIM (used in both the spinup and coupled run, as for C06). The large spread in results (shaded area) highlights the importance of LSM parameter sets in the coupled forecast of heat and moisture states and fluxes. That MC-SIM randomly sampled these sets suggests the full spread, using physically reasonable bounds on parameter values as was done here, could actually be much larger than shown here as well. Nearly all of the MC-SIM simulations are on the dry side of observations, an indication of the dry bias in the Noah model that is only circumvented when using the full C06 calibration with observations. The fluxes in MC-SIM vary quite a bit as well, where β_{sfc} ranges from 0.733-4.960 and large errors versus observed are carried into the entrainment and ratio components.

Overall, these results show the potential uncertainty in LSM parameter specification and substantial impact on the coupled system. The next phase of this research will further explore uncertainty propagation, and quantify how the spread in predictions is narrowed after incorporating observations into the system via calibration. For this task, LIS-OPT/UE has been augmented to include recent algorithmic advances in Markov chain Monte Carlo (MCMC) and will be used to evaluate trade-offs in observation quality and frequency on reducing uncertainty in coupled forecasts.

5. Discussion

The questions addressed in this study of improving coupled prediction using LSM calibration have shed light on the following issues: 1) what to calibrate, 2) how to calibrate, and 3) when to calibrate. Because fluxes are the most important aspect of LSMs for atmospheric models, the largest impact will be seen in calibrating a LSM to Qle and Qh observations. In the approach presented here, in contrast to Santanello et al. 2007, we calibrate only fluxes and therefore, soil states such as moisture and temperature are by-products without observational constraints. Current and future missions such as SMOS and SMAP will provide soil moisture state observations that can be used to calibrate soil hydraulic properties as shown in Santanello et al, etc. However, based on the work presented here, and given the interaction between the soil hydraulics and the canopy conductance, it will be most beneficial to land-atmosphere prediction if both state and flux measurements can be used simultaneously to calibrate LSM parameters.

In terms of how to calibrate, it is not so much the algorithm choice (e.g. similar performance has been seen in LIS-OPT/UE intercomparisons of the three methods therein; Harrison et al. 2012) so much as the parameter sets and mapping approach that is employed that is important for coupled prediction. NU-WRF is fully 3-D and communicates horizontally between grid cells through the atmospheric flow. This is in contrast to LIS and most LSMs, which operate in 1-D. This makes it particularly important that parameter calibration and assignment be considered carefully for coupled studies. The approach performed in this study entailed the assignment of soil, vegetation, and general parameter types, followed by averaging across observation sites for like classes of each and assignment to the full domain. With the exception of a few sites in the offline calibration results, this approach seemed to work well overall as evidenced by the independent assessment of 214 locations of T2 and Q2 performance in the coupled run. A next step in this regard is to investigate the classification at those ARM-SGP sites that degraded after calibration to see if the soil type and land cover representation at those flux towers was represented accurately by the datasets (STATSGO and UMD) chosen for this study.

The final question of when to calibrate has been addressed directly as well, and found some interesting results that should be taken into account in future studies. That the calibration in the wet regime worked nearly as well as the dry regime parameters suggests that in order to improve simulations during extremes, the calibration should at least include a period of extreme soil moisture conditions. Clearly, this is not a one-size-fits-all approach, and depends on the seasonality of a particular location/climate regime, but also suggests that the model physics be tested outside of 'average' conditions in order to maximize LSM improvement due to calibration. (i.e. to capture wings of the distribution (dry-downs and wet-ups) and model biases). There are many more experiments that could be performed in terms of period sensitivity (e.g. seasonal, application to average condition coupled cases, etc.) that will be a part of future research.

Another issue rarely addressed in studies of LSM calibration is that of the physical meaningfulness of the calibrated parameter values. It is important to consider what the calibrated values look like and actually represent, relative to the default lookup tables. Santanello et al. (2007) was successful in achieving both goals of reducing model bias and maintaining parameter realism amongst soil hydraulic properties through the use of pedotransfer functions. Here, the parameter set is large such that it remains difficult to ensure or even evaluate inter-parameter consistency and applicability to real world (or measured) properties, not to mention that not all parameters in Noah LSM are observable. For most calibration studies, the ends (i.e. improved flux output) justify the means (i.e. limited parameter realism). However, we can still take a closer look at the evaporative physics in Noah and two of the commonly modified and 'tuned' parameters in previous studies.

The FXEXP parameter is the exponent for bare soil evaporation in Noah, which is a function of soil moisture and vegetation amount. Lower values of FXEXP increase the bare soil component of Qle for a given soil moisture/vegetation amount, and the default value is 2.0. Table 9 shows the calibrated values from the different period experiments, and there is a definite downward shift in FXEXP due to calibration towards 1.0. In fact, Santanello et al. (2007) modified the FXEXP parameter in their study to be 1.0, due to the semi-arid region and inability of Noah to produce enough Qle. The calibration here has acted in the same manner in order to increase Qle to match observations.

The other parameter of interest is part of the evaporative/flux calculations in Noah. CZIL is the Zilitinkevich coefficient relating surface fluxes to the roughness length for heat (Z_{oh}) and the exchange coefficient (C_h). There has been recent work in Noah model development to modify this from its default value of 0.1 to something higher or lower dependent on vegetation coverage (e.g. Mitchell et al. 2004, LeMone et al. 2010, Trier et al. 2011). Higher values of CZIL decrease Z_{oh}, C_h, and flux magnitudes overall. Table 9 shows the values of CZIL from DEF lookup table of Noah along with calibrated values from different periods and the prior study estimates. The value has been raised to 0.6 in the calibrations that perform best (C06, C07, C678) versus 0.1 in the DEF and the poor calibration of C08.

These results are consistent with tests of the Noah model over the ARM-SGP domain by LeMone et al. (2010) who found that CZIL should be larger in this region. The SPN vs. CPL results here also support those of Trier et al. (2008) in terms of consistency in calibrated parameter sets, and suggest that the results of Trier et al. (2011) would have shown even greater sensitivity of land-PBL coupling to CZIL if the same modified values were used both in the spinup and coupled runs (their CZIL modifications were applied to the coupled run only). Overall, the calibrated values of both CZIL and FXEXP appear to be physically consistent with previous studies' manual tuning of parameters, and while they by no means guarantee the same for the other 27 parameters involved at least suggest some physical consistency and model improvement that produces the right answer for the right reasons.

6. Conclusions

This study examines the impact of LSM spinup and calibration on the land-PBL coupling in regional model forecasts. Sensitivities to dry/wet regimes, period of calibration, and parameter sets were quantified using diagnostics of land-atmosphere coupling and applied to the NU-WRF coupled modeling system. Key findings from this work include the following:

- Offline calibration using a surface flux network is successful in reducing LSM biases and improving diurnal cycles of Qle and Qh.
- Calibrated parameter sets can improve fluxes and states during both dry and wet regimes, and extend their impact to PBL fluxes and ambient weather (T2 and Q2).
- Largest impacts of offline calibration on coupled runs are seen during the dry regime when the turbulent fluxes are larger and atmospheric and precipitation forcing is weak.
- A calibrated spinup by itself can produce more accurate temperature and humidity forecasts, regardless of the parameter sets used in the coupled simulation; though consistency in parameter sets between spinup and coupled runs is critical to improving performance and maintaining physical consistency in *both* states and fluxes
- Including periods of dry and/or wet extremes for a particular region in the calibration process leads to better offline and coupled simulations.
- Significant variability in hydrometeorological prediction can result from LSM parameter uncertainty, but can be reduced using observations and calibration approaches.

These experiments were also designed as a prototype testbed for future satellite missions (e.g. SMAP). Using LIS-OPT/UE, the tradeoffs of data availability vs. accuracy and uncertainty in prediction can be quantified systematically. The classification strategy relates to the spatial tradeoffs of satellite sensors, while the period of calibration relates to the satellite overpass return time. In the future, simultaneous development of Earth science technologies (e.g. microwave soil moisture sensors) and methodologies (e.g. thermal evapotranspiration retrievals) will warrant the LIS-OPT/UE approach in assessing the impact of observations on coupled forecasts, for both calibration and data assimilation studies alike.

Acknowledgements. This work was supported by the NASA's Advanced Information System Technology program (AIST) and the Earth Science Technology Office (ESTO). The NU-WRF team was also instrumental in providing support related to LIS-WRF coupling and a stable and updated version of the coupled system. The data for the MET analysis are from the Research Data Archive (RDA), which is maintained by the Computational and Information Systems Laboratory (CISL) at the National Center for Atmospheric Research (NCAR). The original data are available from the RDA (http://dss.ucar.edu) in dataset number ds337.0.

7. References

Anthes, R.A., and T.T. Warner, 1978: Development of Hydrodynamic Models Suitable for Air Pollution and Other Mesometerological Studies. *Mon. Wea. Rev.*, **106**, 1045–1078.

Betts, A.K., 1992: FIFE atmospheric boundary layer budget methods. *J. Geophys. Res.,* **97,** 18523–18532.

Betts, A. K., 2009: Land-surface-atmosphere coupling in observations and models. *J. Adv. Model Earth Syst.,* **Vol. 1,** Art. #4, 18 pp., doi: 10.3894/JAMES.2009.1.4.

Case, J. L., W. L. Crosson, S. V. Kumar, W. M. Lapenta, C. D. Peters-Lidard, 2008: Impacts of High-Resolution Land Surface Initialization on Regional Sensible Weather Forecasts from the WRF Model. *J. Hydrometeor*, **9**, 1249–1266.

Case, J. L., S. V. Kumar, J. Srikishen, G. J. Jedlovec, 2011: Improving Numerical Weather Predictions of Summertime Precipitation over the Southeastern United States through a High-Resolution Initialization of the Surface State. *Wea. Forecasting*, **26**, 785–807.

Chen, F., and K. Mitchell, 1999: Using GEWEX/ISLSCP forcing data to simulate global soil moisture fields and hydrological cycle for 1987-1988. J. *Meteor. Soc. Japan*, **77,** 167-182.

Chen, F., Manning, K.W., Lemone, M.A., Trier, S.B., Alfieri, J.G., 2007: Description and evaluation of the characteristics of the NCAR high-resolution land data assimilation system. *Journal of Applied Meteorology and Climatology*, **46**, 694-713, 10.1175/JAM2463.1.

Chin, M., R. B. Rood, S.-J. Lin, J. F. Muller, and A. M. Thomspon, 2000: Atmospheric sulfur cycle in the global model GOCART: Model description and global properties. *J. Geophys. Res.*, **105,** 24,671-24,687.

Cosgrove, B. A., et al., 2003: Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS), *J. Geophys. Res.*, **108,** 8845, doi:10.1029/2002JD003316.

de Goncalves, L. G. G., W. J. Shuttleworth, E. J. Burke, P. Houser, D. L. Toll, M. Rodell, and K. Arsenault, 2006: Toward a South America Land Data Assimilation System: Aspects of land surface model spin-up using the Simplified Simple Biosphere, *J. Geophys. Res.*, **111,** D17110, doi:10.1029/2005JD006297.

Di Giuseppe, Francesca, Davide Cesari, Giovanni Bonafé, 2011: Soil Initialization Strategy for Use in Limited-Area Weather Prediction Systems. *Mon. Wea. Rev.,* **139**, 1844–1860.

Dirmeyer, P., R. Koster, Z. Guo, 2006: Do Global Models Properly Represent the Feedback Between Land and Atmosphere?. *J. Hydrometeorol.*, **7,** Issue 6. DOI:10.1175/JHM532.1.

Duan, Q., V. Gupta, and S. Sorooshian, 1993: A shuffled complex evolution approach for effective and efficient global minimization. *J. Optim. Theory. Appl.*, **76(3),** 501-521.

Ek M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of 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.

Feddes, R.A., M. Menenti, P. Kabat, and W.G.M. Bastiaanssen, 2003: Is large-scale inverse modeling of unsaturated flow with areal average evaporation and surface soil moisture as estimated from remote-sensing feasible. *J. Hydrol.*, **143**(1-2), 125-152.

Gilks, W.R., S. Richardson, and D.J. Spiegelhalter, 1996: Introducing Markov chain Monte Carlo", *In Markov chain Monte Carlo in Practice*, Edited by W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, Chapman and Hall.

Gupta, H.V., L.A. Bastidas, S. Sorooshian, W.J. Shuttleworth, and Z.L. Yang, 1999: Parameter estimation of a land-surface scheme using multicriteria methods. J. Geophys. Res., 104(D16), 19,491-19,503.

Gutmann, E.D., and E.E. Small, 2005: The effect of soil hydraulic properties vs. soil texture in land surface models. *Geophys. Res. Let.*, **32**(2), L02402, doi:10.1029/2004GL021843.

Hansen, M. C., Defries, R. S., Townshend, J. R. G., & Sohlberg, R., 2000: Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, **21,** 1331−1364.

Harrison, K.W., S.V. Kumar, C.D. Peters-Lidard and J.A. Santanello, 2012: Reducing soil moisture modeling uncertainty with remote sensing, *Water Resources Res.*, submitted.

Hess, R., 2001: Assimilation of screen level observations by variational soil moisture analysis. *Meteor. & Atmos. Phys.*, **77**(1-4), 145-154.

Hogue, T. S., L. Bastidas, H. Gupta, S. Sorooshian, K. Mitchell, and W. Emmerich, 2005: Evaluation and Transferability of the Noah Land Surface Model in Semiarid Environments. *J. Hydrometeorol.*, **6,** 68-84.

Holland, J., 1975: *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Michigan.

Holt, T., Niyogi, D., Chen, F., Manning, K.W., Lemone, M.A., 2006: Effect of land-atmosphere interactions on the IHOP 24-25 May 2002 convection case. *Monthly Weather Review*, **134**, 113-133, 10.1175/MWR3057.1.

Hong, S.Y., Y. Noh, and J. Dudhia, 2006: A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. *Mon. Wea. Rev.*, **134**, 2318–2341.

Hurk, B.J.J.M. van den and E.M. Blyth, 2008: WATCH/LoCo workshop report.

GEWEX Newsletter, 12-14.

Hurk, B. J. J. M. can den, M. Best, P. Dirmeyer, A. Pitman, J. Polcher, J. Santanello, Jr. 2011: Acceleration of Land Surface Model Development over a Decade of Glass. *Bull. Amer. Meteor. Soc.*, **92**, 1593–1600.

Kato, H., M. Rodell, F. Beyrich, H. Cleugh, E. van Gorsel, H. Liu, and T.P. Meyers, 2007: Sensitivity of Land Surface Simulations to Model Physics, Parameters, and Forcings, at Four CEOP Sites, *J. Meteor. Soc. Japan,* **85A,** 187-204.

Koster R. D., Coauthors, 2004: Regions of strong coupling between soil moisture and precipitation. *Science,* **306**, 1138–1140.

Koster , R. D., S. Mahanama, T. Yamada, G. Balsamo, A.A. Berg, M. Boisserie, P. Dirmeyer, F. Doblas-Reyes, G. Drewitt, C.T. Gordon, Z. Guo, J.H. Jeong, D.M. Lawrence, W.-S. Lee, Z. Li, L. Luo, S. Maleyshev, W.J. Merryfield, S.I. Seneviratne, T. Stanelle, B.J.J.M. van den Hurk, F. Vitart and E.F. Wood, 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.

Kumar, S. V., C. D. Peters-Lidard, Y. Tian, P. R. Houser, J. Geiger, S. Olden, L. Lighty, J. L. Eastman, B. Doty, P. Dirmeyer, J. Adams, K. Mitchell, E. F. Wood and J. Sheffield, 2006: Land Information System - An Interoperable Framework for High Resolution Land Surface Modeling. *Environmental Modelling & Software,* **Vol. 21,** 1402-1415.

Kumar, S. V., C. D. Peters-Lidard, J. L. Eastman, and W.-K. Tao, 2008: An integrated high resolution hydrometeorological modeling testbed using LIS and WRF. *Environmental Modelling and Software*, **23,** 169-181.

Kumar, S., R. H. Reichle, K. W. Harrison, C. D. Peters-Lidard, S. Yatheendradas, and J. A. Santanello, 2012a: A comparison of methods for a priori bias correction in soil moisture data assimilation.*Water Resour. Res.*, doi:10.1029/2010WR010261, **in press**.

Kumar, S. V., C. D. Peters-Lidard, J. A. Santanello, K. Harrison, Y. Liu, and M. Shaw, 2012b: Land surface Verification Toolkit (LVT) - A generalized framework for land surface model evaluation. *Geosci. Model Dev.*, **in press**.

LeMone, Margaret A., Fei Chen, Mukul Tewari, Jimy Dudhia, Bart Geerts, Qun Miao, Richard L. Coulter, Robert L. Grossman, 2010: Simulating the IHOP_2002 Fair-Weather CBL with the WRF-ARW–Noah Modeling System. Part I: Surface Fluxes and CBL Structure and Evolution along the Eastern Track. *Mon. Wea. Rev.*, **138**, 722–744.

Levenberg, K., 1944: A method for the solution of certain non-linear problems in least squares. *The Quarterly of Applied Mathematics*, **2,** 164-168.

Liu, Y., L.A. Bastidas, H.V. Gupta, and S. Sorooshian, 2003: Impacts of a parameterization deficiency on offline and coupled land surface model simulations. *J. Hydrometeor.*, **4**(5), 901-914.

Liu, Y., H.V. Gupta, S. Sorooshian, L.A. Bastidas, and W.J. Shuttleworth, 2004: Exploring parameter sensitivities of the land surface using a locally coupled land-atmosphere model. *J. Geophys. Res. Atmos.*, **109**(D21), D21101, doi: 10.1029/2004JD004730.

Liu, Y., H.V. Gupta, S. Sorooshian, L.A. Bastidas, and W.J. Shuttleworth, 2005: Constraining land surface and atmospheric parameters of a locally coupled model using observational data. *J. Hydrometor.*, **6**(2), 156-172.

Marquardt, D., 1963: An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, **11,** 431-441.

Matsui, T., W. Tao, H. Masunaga, C. D. Kummerow, W. S. Olson, N. Teruyuki, M. Sekiguchi, M. Chou, T. Y. Nakajima, X. Li, J. Chern, J. J. Shi, X. Zeng, D. J. Posselt, K. Suzuki, 2009: Goddard Satellite Data Simulation Unit: Multi-Sensor Satellite Simulators to Support Aerosol-Cloud-Precipitation Satellite Missions, *Eos Trans. AGU,* **90(52),** Fall Meet. Suppl., Abstract A21D-0268

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. Bull. Amer. Meteor. Soc., 87, 343–360.

Miller, D. and R. White, 1998. A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. *Earth Interactions,* **2,** 1-26.

Mitchell, K. E., et al., 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.,* **109,** D07S90,doi:10.1029/2003JD003823.

Nash, J. E. and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology,* **10 (3),** 282–290.

Peters-Lidard, C. D., D. M. Mocko, J. A. Santanello, M. Tischler, M. S. Moran, M. Garcia, and Y. Wu, 2008: The role of precipitation uncertainty for soil property estimation using soil moisture retrievals in a semi-arid environment. *Water Resour. Res.,* **44,** W05S18.

Peters-Lidard, C. D., P. R. Houser, Y. Tian, S. V. Kumar, J. Geiger, S. Olden, L. Lighty, B. Doty, P. Dirmeyer, J. Adams, K. Mitchell, E. F. Wood and J. Sheffield, 2007: High-performance Earth system modeling with NASA/GSFC's Land Information System. *Innovations in Systems and Software Engineering,* **Vol. 3(3),** 157-165.

Reichle, R.H., S.V. Kumar, S.P.P. Mahanama, R.D. Koster and Q. Liu, 2010: Assimilation of satellite-derived skin temperature observations into land surface models. *J. Hydrometeorol.,* **11,**1103-1122,doi:10.1175/2010JHM1262.1.

Robock, A., L. Luo, E.F. Wood, F. Wen, K.E. Mitchell, P.R. Houser, J.C. Schaake, D. Lohmann, B. Cosgrove, J. Sheffield, Q. Duan, R.W. Higgins, R.T. Pinker, J.D. Tarpley, J.B. Basara, and K.C. Crawford, 2003: Evaluation of the North American Land Data Assimilation System over the southern Great Plains during the warm season. *J. Geophys. Res.*, **108(D22),** 8846, doi:10.1029/2002JD003245.

Rodell, M., P.R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J.K. Entin, J.P. Walker, D. Lohmann, and D. Toll, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.,* **85(3),** 381-394, 2004.

Rodell, M., P.R. Houser, A.A. Berg, and J.S. Famiglietti, 2005: Evaluation of 10 Methods for Initializing a Land Surface Model. *J. Hydromet.*, **6(2),** 146-155

Santanello, J. A., C. Peters-Lidard, M. Garcia, D. Mocko, M. Tischler, M. S. Moran, and D. P Thoma, 2007: Using Remotely-Sensed Estimates of Soil Moisture to Infer Spatially Distributed Soil Hydraulic Properties. *Rem. Sens. Env*., **110,** 79-97.

Santanello, J. A., Christa D. Peters-Lidard, Sujay V. Kumar, Charles Alonge, Wei-Kuo Tao, 2009: A Modeling and Observational Framework for Diagnosing Local Land–Atmosphere Coupling on Diurnal Time Scales. *J. Hydrometeor*, **10**, 577–599.

Santanello, J. A., C. D. Peters-Lidard, and S. V. Kumar, 2011a: Diagnosing the Sensitivity of Local Land–Atmosphere Coupling via the Soil Moisture–Boundary Layer Interaction. *J. Hydrometeor*, **12**, 766–786.

Santanello, J. A., 2011b: Results from Local Land-Atmosphere Coupling (LoCo) Project. *GEWEX Newsletter*, **21(4),** 7-9.

Santanello, J. A., C. Peters-Lidard, A. Kennedy, and S. Kumar, 2012: Diagnosing the Nature of Land-Atmosphere Coupling During the 2006-7 Dry/Wet Extremes in the U. S. Southern Great Plains. *J. Hydromet.*, under revision.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner,, B Orlowsky, A. J. Teuling, 2010: Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Rev.,* **99,** 125-161.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers, 2005: A Description of the Advanced Research WRF Version 2, NCAR Tech Note, NCAR/TN–468+STR, 88 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO, 80307; on-line at: http://box.mmm.ucar.edu/wrf/users/docs/arw_v2.pdf]

Soet, M., and J.N.M. Stricker, 2003: Functional behaviour of pedotransfer functions in soil water flow simulation. *Hydro. Proc.,* **17**(8), 1659-1670.

ter Braak, C.J.F., 2006: A Markov chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces. *Stat Comput*, **16,** 239-249.

Trier, S. B., F. Chen, K. W. Manning, M. A. LeMone, C. A. Davis, 2008: Sensitivity of the PBL and Precipitation in 12-Day Simulations of Warm-Season Convection Using Different Land Surface Models and Soil Wetness Conditions. *Mon. Wea. Rev.*, **136**, 2321–2343.

Trier, S. B., M. A. LeMone, F. Chen, K. W. Manning, 2011: Effects of Surface Heat and Moisture Exchange on ARW-WRF Warm-Season Precipitation Forecasts over the Central United States. *Wea. Forecasting*, **26**, 3–25.

Troen, I., and L. Mahrt, 1986: A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation. *Bound.-Layer Meteor.*, **37,** 129–148..

U.S. Department of Agriculture, Natural Resources Conservation Service, 1994. State Soil Geographic (STATSGO) database for Arizona. State College, PA: Penn State University Earth Systems Science Center. http://www.essc.psu.edu/soil_info/index.cgi?soil_data&statsgo

van Heerwaarden, C. C., Vilà-Guerau de Arellano, J., Moene, A. F. and Holtslag, A. A. M., 2009: Interactions between dry-air entrainment, surface evaporation and convective boundary-layer development. *Quarterly Journal of the Royal Meteorological Society*, **135,** 1277–1291. doi: 10.1002/qj.431.

Wen, X., S. Lu, J. Jin, 2012: Integrating Remote Sensing Data with WRF for Improved Simulations of Oasis Effects on Local Weather Processes over an Arid Region in Northwestern China. *J. Hydrometeor*, **13**, 573–587.

Xia, Y., Ek, M., Wei, H. and Meng, J., 2012: Comparative analysis of relationships between NLDAS-2 forcings and model outputs. *Hydrol. Processes,* **26,** 467–474.

Yang, Z., Dickinson, R.E., Henderson-Sellers, A. and Pitman, A.J., 1995: Preliminary study of spin-up processes in land surface models with the first stage data of Project for Intercomparison of Land Surface Parameterization Schemes Phase 1(a). *Journal of Geophysical Research* **100,** doi: 10.1029/95JD01076. issn: 0148-0227.

LIST OF TABLES

LIST OF FIGURES

Figure 14: Mixing diagrams for the 14 July 2006 case study showing the default (DEF) and suite of experiments using parameters calibrated during 2006 (C06), 2007 (C07), 2008 (C08), and all three years combined (C678), along with observations at the ARM-SGP E4 site.

Figure 15: Full PBL budget components of surface, PBL, and total sensible and latent heat flux derived from the mixing diagram analysis in Fig. 14.

Figure 16: Daytime mean evaporative fraction versus PBL height for the simulations in Fig. 14.

Figure 17: Mixing diagrams for the 14 July 2006 case study showing the default and suite of experiments using parameters calibrated during 2006 (C06) and 5 randomly sampled set generated from MC-SIM (MCSIM; shaded), along with observations at the ARM-SGP E4 site.

Noah Parameter	Minimum	Maximum
SMCMAX	0.30	0.50
PSISAT	0.01	0.70
DKSAT (ms-1)	0.05 E-5	3.00 E-5
DWSAT	5.71 E-6	2.33 E-5
BEXP	3	9
QUARTZ	0.10	0.90
RSMIN (m)	40	1000
RGL	30	150
HS	36	55
Z0 (m)	0.01	0.99
LAI	0.05	6.00
CFACTR	0.10	2.00
CMCMAX (m)	1.00 E-4	2.00 E-3
SBETA	-4.00	-1.00
RSMAX (m)	2000	10000
TOPT (K)	293	303
REFDK	5.00 E-7	3.00 E-5
FXEXP	0.20	4.00
REFDT	0.10	10.00
CZIL	0.05	0.80
FRZK	0.10	0.25
SNUP	0.025	0.08
SMCREF	0.00	0.50
SMCDRY	0.00	0.15
SMCWLT	0.00	0.15
F1	-11	0
CSOIL	1.26 E6	3.56 E6
SLOPE	0.00	1.00
EMISS	0.80	1.00

Table 1: Minimum and maximum values of the Noah parameters used in the LIS-OPT experiments.

	Exp.	Description	Spinup Parameters	Coupled Parameters
1	**DEF**	Default run w/uncalibrated params in LIS & NU-WRF	**Default**	**Default**
2	CPL	Impact of calibrated parameters in NU-WRF ONLY	Default	Calibrated
3	SPN	Impact of calibrating LIS spinup (ICs) ONLY	Calibrated	Default
4	**SCP**	Impact of full calibration (LIS and NU-WRF)	**Calibrated**	**Calibrated**

Table 2: Description of calibration approaches and parameter sets used in NU-WRF simulations.

		DEF	CPL	SPN	SCP
Cum RMSE		6288.60	6161.24	4665.10	**5314.07**
Cum MAE		5231.25	5181.39	4044.50	**4541.69**
BIAS	**Q2**	-6022.76	-5743.49	-3159.91	**-4196.35**
BIAS	**T2**	4244.72	4458.54	3336.54	**3919.27**
N-S Efficiency		-1.78	-1.67	-0.53	**-0.98**

Table 3: Error statistics for Fig.4, where the co-evolution of 2m-specific humidity (Q2) and temperature (T2) are from each simulation is evaluated against observations in time in terms of RMSE, MAE, Bias, and the Nash-Sutcliffe Efficiency (Nash and Sutcliffe, 1970).

		DEF	CPL	SPN	SCP
Cum RMSE		6018.59	5992.34	3977.58	**5086.32**
Cum MAE		4921.32	4992.19	3050.16	**4129.53**
BIAS	**Q2**	-7889.19	-7859.74	-5002.86	**-6663.78**
BIAS	**T2**	1953.45	2124.63	818.18	**1595.27**
N-S Efficiency		-0.385	-0.373	0.394	**0.011**

		DEF	CPL	SPN	SCP
Cum RMSE		5916.36	5464.83	4031.29	**5116.14**
Cum MAE		4638.54	4450.96	2475.01	**3970.43**
BIAS	**Q2**	-6905.71	-6541.11	-3709.10	**-5976.76**
BIAS	**T2**	2371.36	2360.82	416.55	**1964.09**
N-S Efficiency		-0.128	0.038	0.476	**0.157**

Table 4ab: Error statistics from a) Fig. 9a and b) Fig. 9b for all four simulations.

		DEF	CPL	SPN	SCP
Cum RMSE		**1380.69**	1731.27	1539.36	1718.66
Cum MAE		**1190.26**	1421.29	1280.70	1386.36
BIAS	**Q2**	**436.17**	-478.81	1283.31	938.37
BIAS	**T2**	**1412.82**	1920.64	1155.18	1485.82
N-S Efficiency		**0.809**	0.699	0.762	0.704

		DEF	CPL	SPN	SCP
Cum RMSE		1788.06	2480.89	1240.10	**1498.29**
Cum MAE		1644.65	2280.67	1119.14	**1338.25**
BIAS	**Q2**	-1761.03	-2627.25	-977.38	**-1164.02**
BIAS	**T2**	1528.27	1934.09	1237.55	**1240.91**
N-S Efficiency		0.183	-0.573	0.607	**0.426**

Table 5ab: Error statistics from a) Fig. 10a and b) Fig. 10b for all four simulations.

		DEF	CPL	SPN	SCP
Cum RMSE		**4177.31**	4963.27	4263.40	4611.42
Cum MAE		**3501.51**	4383.16	3576.48	3987.41
BIAS	**Q2**	-257.51	-1412.37	1159.99	**142.81**
BIAS	**T2**	**2361.73**	3213.09	2043.18	2811.18
N-S Efficiency		**-1.193**	-2.096	-1.285	-1.673

		DEF	CPL	SPN	SCP
Cum RMSE		**1598.93**	1898.51	2301.55	1632.62
Cum MAE		**1412.15**	1708.75	2026.01	1497.77
BIAS	**Q2**	-467.35	-1119.43	2471.04	**-195.45**
BIAS	**T2**	**1373.55**	1948.36	1144.36	1639.91
N-S Efficiency		**0.672**	0.538	0.321	0.658

Table 6ab: Error statistics from a) Fig. 11a and b) Fig. 11b for all four simulations.

	Exp.	Description	Spinup Parameters	Coupled Parameters
1	**DEF**	Default run w/uncalibrated params	**Default**	**Default**
2	**C06**	Impact of calibrating during 2006 only	**2006**	**2006**
3	**C07**	Impact of calibrating during 2007 only	**2007**	**2007**
4	**C08**	Impact of calibrating during 2008 only	**2008**	**2008**
5	**C678**	Impact of calibrating to all three years combined	**2006-7-8**	**2006-7-8**

Table 7: Description of calibration approaches and parameter sets used in NU-WRF simulations.

	DEF	C07	C08	C06	C678
Cum MAE	5231.25	**4538.32**	5707.05	**4541.69**	4630.35
Cum RMSE	6288.60	**5371.56**	6851.72	**5314.07**	5490.36
Q2 BIAS	-6022.76	**-4249.04**	-7044.01	**-4196.35**	-4492.11
T2 BIAS	4244.73	**3977.18**	4370.09	**3919.27**	3998.27
N-S Efficiency	-1.782	**-1.030**	-2.303	**-0.987**	-1.121

Table 8: Error statistics from Fig.14 for each of the simulations.

	DEF	**C06**	**C07**	**C08**	**C678**	**LeMone et al. (2008)**	**Trier et al. (2011)**
FXEXP	2	1.06	1.34	0.969	1.19	-	-
CZIL	0.1	0.6	0.6	0.1	0.6	0.5	0.1-1.0

Table 9: Values of the Noah CZIL and FXEXP parameters used in each of the simulations and the CZIL studies of LeMone et al. (2008) and Trier et al. (2011).