

Confronting Weather and Climate Models with Observational Data from Soil Moisture Networks over the United States

PAUL A. DIRMEYER,* JIEXIA WU,* HOLLY E. NORTON,* WOUTER A. DORIGO,^{+,#}
 STEVEN M. QUIRING,[@] TRENTON W. FORD,[&] JOSEPH A. SANTANELLO JR.,**
 MICHAEL G. BOSILOVICH,** MICHAEL B. EK,⁺⁺ RANDAL D. KOSTER,**
 GIANPAOLO BALSAMO,^{##} AND DAVID M. LAWRENCE^{@@}

* *George Mason University, Fairfax, Virginia*

+ *Vienna University of Technology, Vienna, Austria*

Laboratory of Forest and Water Management, Ghent University, Ghent, Belgium

@ *Texas A&M University, College Station, Texas*

& *Southern Illinois University, Carbondale, Illinois*

** *NASA Goddard Space Flight Center, Greenbelt, Maryland*

++ *NOAA/National Centers for Environmental Prediction, College Park, Maryland*

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

@@ *National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 19 October 2015, in final form 14 January 2016)

ABSTRACT

Four land surface models in uncoupled and coupled configurations are compared to observations of daily soil moisture from 19 networks in the conterminous United States to determine the viability of such comparisons and explore the characteristics of model and observational data. First, observations are analyzed for error characteristics and representation of spatial and temporal variability. Some networks have multiple stations within an area comparable to model grid boxes; for those it is found that aggregation of stations before calculation of statistics has little effect on estimates of variance, but soil moisture memory is sensitive to aggregation. Statistics for some networks stand out as unlike those of their neighbors, likely because of differences in instrumentation, calibration, and maintenance. Buried sensors appear to have less random error than near-field remote sensing techniques, and heat-dissipation sensors show less temporal variability than other types. Model soil moistures are evaluated using three metrics: standard deviation in time, temporal correlation (memory), and spatial correlation (length scale). Models do relatively well in capturing large-scale variability of metrics across climate regimes, but they poorly reproduce observed patterns at scales of hundreds of kilometers and smaller. Uncoupled land models do no better than coupled model configurations, nor do reanalyses outperform free-running models. Spatial decorrelation scales are found to be difficult to diagnose. Using data for model validation, calibration, or data assimilation from multiple soil moisture networks with different types of sensors and measurement techniques requires great caution. Data from models and observations should be put on the same spatial and temporal scales before comparison.

1. Introduction

Coupled land–atmosphere model development has lagged behind coupled ocean–atmosphere model development for a variety of reasons. Top among them is that the necessary measurements for assessing land–atmosphere feedback processes have been largely lacking. In recent

years, collocated measurements of surface fluxes, near-surface meteorology, and land surface states like soil moisture have begun to cross a critical threshold of quantity and coverage, largely because of the maturation of the global FLUXNET set of environmental measurements (Baldocchi et al. 2001). Systematic benchmarking of land surface models (LSMs) has begun based on simulation of daily mean surface fluxes (Best et al. 2015). However, for soil moisture alone there are even more widespread data in the form of many independent networks of in situ measurements (Dorigo et al. 2011; Quiring et al. 2016). They span a tremendous

Corresponding author address: Paul A. Dirmeyer, Center for Ocean–Land–Atmosphere Studies, George Mason University, 4400 University Dr., Mail Stop 6C5, Fairfax, VA 22030.
 E-mail: pdirmeye@gmu.edu

range of station densities, down to subgrid scales relative to current weather and climate models, making them enticing for model calibration and validation.

Bringing observational data to bear on model improvement requires not just the datasets and models themselves, but also sound methods of analysis and processes understanding to guide the approach. Comparing models with observations can easily become misguided, if not actually unfair, if the basic differences between how models represent the world and how instruments measure the world are not carefully considered and accounted for. For a quantity like soil moisture, this is a particularly significant issue (Dirmeyer 2004; Koster et al. 2009). Xia et al. (2015) spatially averaged both model and observed data to coarse scales to facilitate comparison. Stillman et al. (2014) assessed the ability of multiple soil moisture instruments in a catchment to represent area-averaged soil moisture, using a higher-density rain gauge network to infer smaller-scale variations. Gruber et al. (2013) examined random errors in soil moisture at spatial scales comparable to global models using a triple collocation method combining remotely sensed and modeled soil moisture estimates with in situ soil moisture measurements, highlighting the care that must be taken in applying in situ measurements as ground truth. Such approaches hold promise to evaluate remote sensing products (Dorigo et al. 2015) and improve estimates of soil moisture–atmosphere interactions (Crow et al. 2015).

In this study, we confront 12 unique model configurations using four different land surface models with soil moisture measurements from 19 networks across the conterminous United States. However, we first address the observational datasets themselves to estimate their error characteristics in a distinctive way based on lagged autocorrelation statistics and their representativeness of temporal variability, spatial, and temporal scales.

Section 2 describes the observational and model data used. The metrics evaluated are introduced in section 3, and section 4 presents an evaluation of observational error. Scaling issues are addressed in section 5. Section 6 gives an evaluation of soil moisture variance and memory in observations and models. Spatial scales of soil moisture variability are considered in section 7, and conclusions and a summary are offered in section 8.

2. Data

In this comparison, point observations and model gridbox estimates of soil moisture data at daily time intervals or daily time means are used. The domain of observations for this study is confined to the conterminous United States, and model comparisons are performed over roughly the same area. Table 1 lists all networks used,

the data collections (described below) from which data were taken, the location of the networks (many are regional), and the type of instrumentation each uses.

a. International Soil Moisture Network

The International Soil Moisture Network (ISMN) is a data synthesis effort focused on collecting in situ soil moisture measurements and associated collocated observations of relevant meteorological data from all available international sources (Dorigo et al. 2011, 2013). The resulting quality-controlled database of raw observations is meant to provide ground-truth calibration and validation for satellite observations as well as for the calibration and validation of land surface models. It is coordinated by the Global Energy and Water Cycle Experiment (GEWEX) in cooperation with the Group on Earth Observations (GEO) and the Committee on Earth Observation Satellites (CEOS).

Data from many different networks and extended field campaigns are archived by ISMN. Those networks used in this experiment are listed in Table 2. ISMN archives data at the highest available temporal resolution up to hourly from each reporting instrument, allotting one file for each instrument and level. Basic quality control is performed, and records suspected to be out of range or otherwise untrustworthy are flagged, but no data are omitted. Figure 1a shows the locations of stations used from ISMN.

b. North American Soil Moisture Database

The North American Soil Moisture Database (NASMD; Quiring et al. 2016) is a collection of harmonized daily soil moisture data from in situ measurements across North America. Networks used in this study from NASMD are listed in Table 3. The motivation for NASMD is to provide a dataset to investigate processes by which soil moisture variability influences climate on seasonal to interannual time scales over North America. Unlike ISMN, NASMD provides processed data from each station location in each network. Daily values are calculated from stations with subdaily data using a simple average. Interpolation is used to fill gaps of less than 10 days using a monthly average replacement method, which has been shown to work well with daily observations (Ford and Quiring 2014).

Only one time series is provided at each station and sensor depth, regardless of how many instruments are in place. When stations have multiple instruments at a single depth, usually the first reported sensor is used. However, if data from the first instrument are flagged by the quality-control routine or if there are excessive missing observations, the next reported instrument is considered and may be used instead. Because of microscale variability in soil texture, averaging observations from multiple

TABLE 1. Observational networks used in this study.

Collection	Network	Full name	Location	Probe type(s)	Reference(s)
NASMD	AMERIFLUX	AmeriFlux	United States	Various dielectric	Baldocchi et al. (2001)
Both	ARM	U.S. Department of Energy Atmospheric Radiation Measurement	Oklahoma and Kansas	Campbell 229-L and SMP1	Bond (2005)
Both	AWDN	Automated Weather Data Network	Nebraska	Vitel and Thetaprobe ML2X	Hubbard et al. (1983)
NASMD	CHILI	Center for Hurricane Intensity and Landfall Investigation	Alabama	Hydraprobe	Kimball et al. (2010)
Both	COSMOS	Cosmic-Ray Soil Moisture Observing System	United States	Cosmic-ray probe	Zreda et al. (2012)
NASMD	DEOS	Delaware Environmental Observing System	Delaware	Campbell 616L	Legates et al. (2005), (2007)
NASMD	ECONet	North Carolina Environment and Climate Observing Network	North Carolina	Thetaprobe	Pan et al. (2012)
NASMD	MAWN	Michigan Automated Weather Network	Michigan	Campbell 616	Andresen et al. (2011)
NASMD	MAW-MO	Missouri Agricultural Electronic Bulletin Board	Missouri	Campbell 616	Guinan and Travlos (2008)
NASMD	NOAA HMT	NOAA Hydrometeorology Testbed	United States	Campbell 616 and Stevens Hydraprobe	Zamora et al. (2011)
NASMD	OK-MESO	Oklahoma Mesonet	Oklahoma	Campbell 229-L	Illston et al. (2008)
ISMN	PBO H ₂ O	Xenon Plate Boundary Observatory (PBO) H ₂ O	Western United States	GPS	Larson et al. (2008)
Both	SCAN	U.S. Department of Agriculture (USDA) Soil Climate Analysis Network	United States	Hydraprobe (analog and digital)	Schaefer et al. (2007)
NASMD	SDAWN	South Dakota Automated Weather Data Network	South Dakota	Stevens Hydraprobe	—
Both	SNOTEL	Snowpack Telemetry	Western United States	Hydraprobe (analog and digital)	Schaefer and Paetzold (2001)
ISMN	SoilSCAPE	Soil Moisture Sensing Controller and Optimal Estimator	United States	EC-5	Moghaddam et al. (2010)
Both	USCRN	NOAA U.S. Climate Reference Network	United States	Stevens Hydraprobe II Sdi-12	Bell et al. (2013)
ISMN	USDA ARS	USDA Agricultural Research Service	United States	Hydraprobe analog	Jackson et al. (2010)
NASMD	WTX-MESO	West Texas Mesonet	Texas	Campbell 616L	Schroeder et al. (2005)

sensors was considered unjustifiable. [Figure 1b](#) shows the distribution of stations in the NASMD repository.

c. Observational data processing

The data from both collections were further processed for this study. Each data file is scanned for basic statistics including the time range of available data, depths of instrument readings, and data reporting intervals so the data from each network can be synthesized into a single file spanning the maximum time range of the network's observations. For the ISMN data, daily means are first calculated.

For each station and profile of sensors in the soil [or across the reporting depth for remote sensors like those in the Cosmic-Ray Soil Moisture Observing System (COSMOS) or Xenon Plate Boundary Observatory (PBO) H₂O (PBO H₂O)], the observational data are vertically interpolated to the model levels for each of the four land surface schemes in this study, following the procedure used in the second Global Soil Wetness Project ([Dirmeier et al. 2006](#)). The lowest model layer to encompass the depth of the deepest reporting sensor is the lowest model layer to contain interpolated data;

TABLE 2. Details of ISMN networks used in this study. Local extent means the network spans parts of one (or for ARM, two) states; regional networks span many states, and often all of the conterminous United States. Interval indicates typical interval; for some networks labeled “hourly” some subset of instruments may report 3- or 6-hourly. Years and days indicate the smallest span that incorporates all data. The instruments column indicates the sum of the max number of each unique type or uniquely labeled type of sensor deployed at any site. SM levels indicates the total number of unique depths of sensor placement for the network.

Network	Extent	Interval	Stations	Years	Days	SM levels	Instruments
ARM	Local	Hourly	29	22	7740	10	15
AWDN	Local	Daily	50	13	4749	4	2
COSMOS	Regional	Hourly	101	7	2308	1 ^a	1
PBO H ₂ O	Regional	Daily	108	8	2865	1 ^a	3
SCAN	Regional	Hourly	211	19	6877	24	52
SNOTEL	Regional	Hourly	415	19	6425	16	45
SoilSCAPE	Regional	Hourly	135	4	7256	29	6
USCRN	Regional	Hourly	114	15	5029	5	5
USDA ARS	Regional	Hourly	4	8	2618	1	4

^a Sensors that sit above ground and sense soil moisture remotely—the depth of the measurement is not static and varies with location, soil moisture, and other conditions.

layers below are set to missing. Model layers above the shallowest sensor depth are set to the soil moisture value of that shallowest sensor. Otherwise, it is assumed that the observed data of buried sensors are representative of a layer whose top is exactly halfway between it and the next shallowest sensor (or the surface if it is the shallowest sensor), and whose bottom is exactly halfway between it and the next deepest sensor (or if it is the deepest sensor, to the same distance below as the top boundary was determined to be above it). It is assumed that the soil moisture across this thickness is uniform, as is typically supposed for land surface model layers. Then the interpolated value for any model layer is a simple weighted average of all observation “layers” that overlap the model layer, preserving water content. This process has the advantage that the final observed time series is on each land surface model’s vertical coordinate, facilitating comparison.

Where there are multiple instruments at the same station in a network in ISMN, the data are sorted based on the number of days without missing data so that the most complete time series can be accessed easily. Data are gathered by network so that analyses and comparisons can be performed on a network-by-network basis.

An initial concern was whether the differences in the processing of data from the same network taken from each data collection would affect the results, particularly the fact that some gap filling had been applied to the NASMD time series, but not to ISMN. Furthermore, it is evident from a comparison of Tables 2 and 3 that the number of stations and period of data collected is not the same between the two collections for the same network. Comparison of network statistics, some of which are shown in section 4, suggests there is no significant difference between the two versions of data for the same networks. Finally, because of varying data availability, for each calculation a fixed period was identified subjectively

for each network where most stations have data. Within that period, a station is removed if more than half the period is outside the range of data for that station.

d. Models

Four LSMs are confronted with the observational data from ISMN and NASMD: the Catchment model from the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC; [Koster et al. 2000](#); [Ducharne et al. 2000](#)); Noah, version 2.7 (Noah2.7), from the National Oceanic and Atmospheric Administration (NOAA; [Ek et al. 2003](#)); the Hydrology Tiled European Centre for Medium-Range Weather Forecasts (ECMWF) Surface Scheme for Exchange over Land (HTESSEL; [Balsamo et al. 2009](#)); and the Community Land Model, version 4 (CLM4), which is sponsored by the National Science Foundation (NSF) and DOE ([Lawrence et al. 2011](#)). Catchment parameterizes an idealized hillslope in each grid box to estimate from bulk water prognostic variables the fractional areas of saturated, unstressed, and dry surfaces with respect to evapotranspiration; it then calculates soil moisture profiles as a diagnostic. The other three LSMs calculate soil moisture in each layer as a balance between gravitational drainage and downgradient conduction in the vertical only. All models treat infiltration of precipitation as a water input at the top, direct evaporation from the top soil layer as an output to the atmosphere, transpiration drawing water out of all soil layers containing roots, and baseflow drainage removing water from the bottom of the soil column. CLM4 includes deep interaction with a water-table parameterization below the soil column. Each model uses its own distributed global map of soil properties on the model grid to determine saturated hydraulic conductivity, porosity, and other necessary hydraulic parameters. None of the models as used here consider vertical variations in basic soil properties.

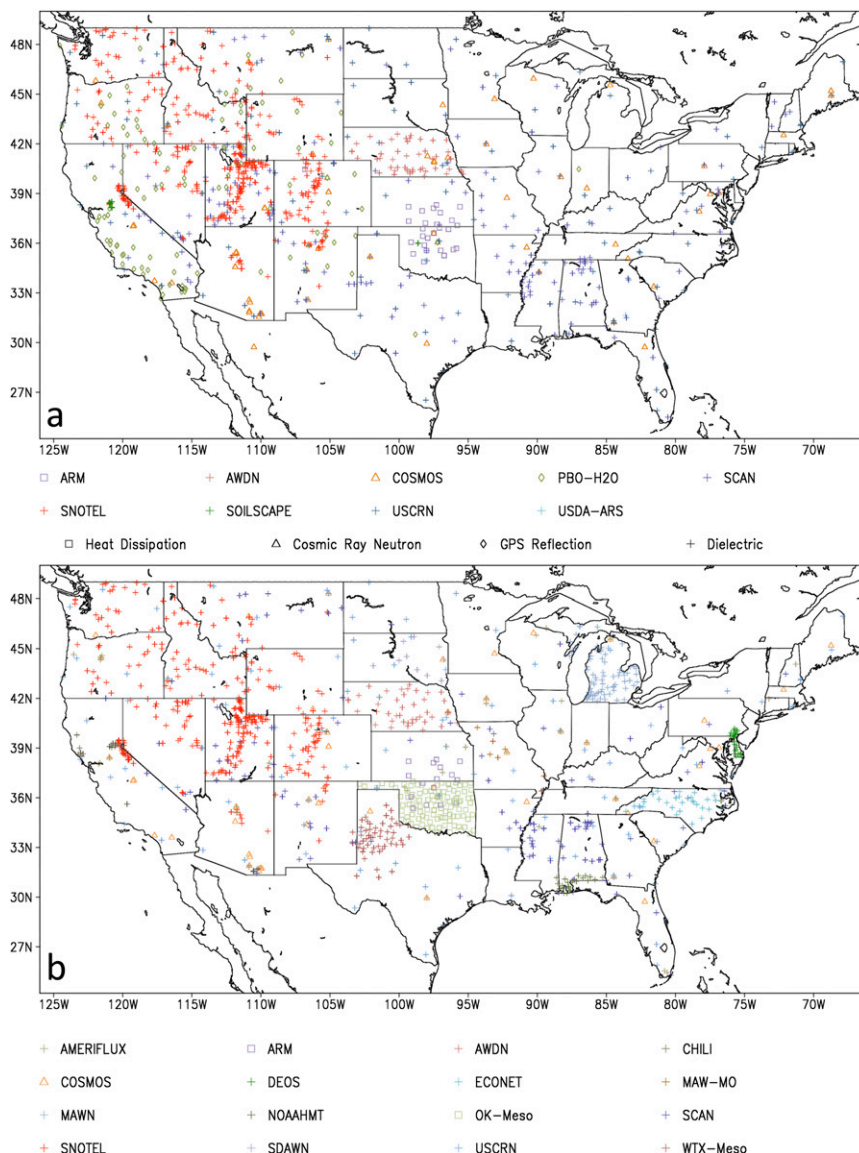


FIG. 1. Station locations for the networks of (a) ISMN and (b) NASMD.

Multiple sets of soil moisture data from each of four modeling centers above have been collected and compared. For contributions from a given modeling center, the LSM used is nearly or exactly the same, but the way time series of soil moisture have been produced varies. The contributions from each modeling center include an offline simulation with the LSM driven by gridded global observationally based meteorological analyses and a simulation or set of simulations with the LSM coupled to its corresponding global atmospheric model in a free-running (unconstrained or forecast) mode. In the case of Noah2.7 and CLM4, an ocean general circulation model was also coupled to the atmospheric model, but that has little consequence for this study, and for CLM4 predicted vegetation phenology was

enabled. For all but CLM4, there is also a reanalysis where the atmosphere and the land surface states, to varying extents, are constrained by data assimilation. Noah2.7 is the LSM in two reanalyses investigated here. Table 4 outlines the various configurations and the spatial resolutions of these models. When compared to observed data, the model grid box containing the site of observed station is used.

3. Metrics

The model simulations described in section 2d are confronted with three metrics from the observed soil moisture networks' data. Additionally, the observations

TABLE 3. As in Table 2, but for NASMD.

Network	Extent	Interval	Stations	Years	Days	SM levels	Instruments
AMERIFLUX	Regional	Daily	55	13	5844	37	1
ARM	Local	Daily	17	16	5845	10	1
AWDN	Local	Daily	41	5	1827	4	1
CHILI	Local	Daily	25	3	1097	1	1
COSMOS	Regional	Daily	54	5	1828	1 ^a	1
DEOS	Local	Daily	26	12	3289	1	1
ECONet	Local	Daily	31	15	5479	1	1
MAWN	Local	Daily	80	17	6209	2	1
MAW-MO	Local	Daily	8	12	4018	1	1
NOAA HMT	Regional	Daily	25	16	4384	3	1
OK-MESO	Local	Daily	104	13	4749	4	1
SCAN	Regional	Daily	123	17	6210	13	1
SNOTEL	Regional	Daily	351	13	4750	16	1
SDAWN	Local	Daily	11	7	2558	5	1
USCRN	Regional	Daily	113	4	1462	27	1
WTX-MESO	Local	Daily	53	11	4018	4	1

^a Sensors that sit above ground and sense soil moisture remotely—the depth of the measurement is not static and varies with location, soil moisture, and other conditions.

themselves are evaluated to assess their likely measurement error based on the methodology of Vinnikov et al. (1996), as described in the next section.

The first metric assessed is the variance or standard deviation of daily soil moisture for each month, grouped by season [December–February (DJF), March–May (MAM), June–August (JJA), and September–November (SON)]. No attempt is made to remove a climatological annual cycle, as the in situ networks have, by and large, not been in place long enough to calculate stable climatological mean annual cycles for most stations. As we are concerned with linkages between land and atmosphere at subseasonal time scales, the mean of each month is removed from all data in that month so that no interannual variability enters the calculation, but some seasonal

trends within months may still be present that may affect statistics.

Second, the soil moisture memory is assessed for each station and vertical level in the soil by computing lagged autocorrelations of the daily time series. Lagged autocorrelations indicate soil moisture behaves as a first-order Markov process (Schlosser and Milly 2002). As a result, we can estimate time scales of correlation, that is, memory, as the time it takes the lagged autocorrelation of soil moisture to drop to $1/e$. We have found that linear extrapolation between the values of $\ln(r)$, where r is autocorrelation at lags of 1 and 2 days, to the lag where $\ln(r) = -1$ provides an estimate that is not significantly different from using a linear fit through $\ln(r)$ at a larger number of lags (cf. Robock et al. 1995). This is also

TABLE 4. Specifications for the four land and atmosphere models, including source of data and spatial resolution.

LSM	Offline	Free running	Reanalysis
Catchment (three layers spanning 2.01 m)	MERRA-Land (MERRA + GPCP forcing) $0.67^\circ \times 0.5^\circ$ Reichle et al. (2011)	GEOS-5 simulation on MERRA-2 mode $0.67^\circ \times 0.5^\circ$	MERRA $0.67^\circ \times 0.5^\circ$ Rienecker et al. (2011)
Noah2.7 (four layers spanning 2.00 m)	GLDAS $1^\circ \times 1^\circ$ Rodell et al. (2004)	CFSv2 seasonal forecasts $0.94^\circ \times 0.95^\circ$	CFSR $0.31^\circ \times 0.37^\circ$ Saha et al. (2010); 20CR, version 2 $1.88^\circ \times 1.91^\circ$ Compo et al. (2011)
HTESSEL (four layers spanning 2.89 m)	ERA-Interim Land $0.75^\circ \times 0.75^\circ$ Balsamo et al. (2015)	IFS in Athena project $0.14^\circ \times 0.14^\circ$ Kinter et al. (2013)	ERA-Interim $0.75^\circ \times 0.75^\circ$ Dee et al. (2011)
CLM4 (12 layers spanning 3.43 m)	Qian et al. (2007) forcing $1.25^\circ \times 0.9^\circ$ Lawrence et al. (2011)	CCSM4 seasonal forecasts $1.25^\circ \times 0.9^\circ$ Dirmeyer et al. (2013)	—

calculated on a seasonal basis as there is a pronounced annual cycle of memory time scales in most locations. These are then compared to model grid boxes at the same locations. Third, we perform a similar calculation between time series from pairs of stations, and between their corresponding model grid boxes, to assess spatial correlation and length scales.

Finally, we are concerned about the representativeness of point soil moisture measurements for model gridbox averages. There are inherent problems with direct comparison between point observations and LSM gridbox output (cf. Gruber et al. 2013; Dirmeyer et al. 2013). The densities of some of the in situ networks are high enough to allow us to assess the sensitivity of an area-average soil wetness at model grid scales to the number of measurements contributing to the average. Thus, we attempt to address the issue of scale mismatch and account for it when comparing models and observations. We address the scaling issue in section 5 before showing model performance on the metrics listed above.

4. Observational error

Vinnikov and Yesserkepova (1991), following the proposition of Delworth and Manabe (1988), showed that soil moisture time series behave like first-order Markov processes, such that the autocorrelation of soil moisture at a location at lag τ decreases as lag grows:

$$r(\tau) = \exp(-\lambda\tau), \quad (1)$$

where λ is the decay frequency, or $1/\lambda$ is the time scale. Robock et al. (1995) showed that for actual observed data, a linear best fit of $\ln(r)$ versus τ for a range of lags does not cross zero at a value of $r = 1$ [i.e., $\ln(r) = 0$], but rather at some correlation $r < 1$. The displacement a of the correlation at $\tau = 0$ [i.e., $r(\tau = 0) = 1 - a$] is an indicator of measurement error.

Vinnikov et al. (1996) noted that the variance in any time series of observed measurements is composed of the sum of the actual variance of the measured quantity and the noise contributed from random observational error. The ratio of error variance δ^2 to real variance σ^2 is related to the displacement of the extrapolated autocorrelation:

$$\frac{\delta^2}{\sigma^2} = \left(\frac{a}{1+a} \right). \quad (2)$$

This relative error can also be derived from a different perspective using triple collocation (Gruber et al. 2016). Given a sufficient number of measurements, observational error from a station or network of stations can be estimated without specific validation or comparison to

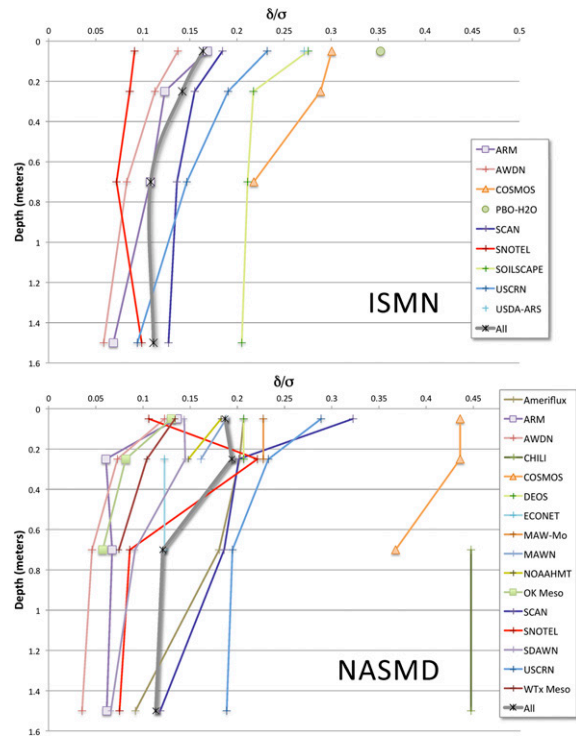


FIG. 2. Estimated relative random observational error as a fraction of estimated real soil moisture variability as a function of depth (interpolated to Noah2.7 LSM layers) for each observational network in (top) ISMN and (bottom) NASMD. “All” is a weighted average of all networks by the number of stations in each network.

independent data. This is a very powerful result that can provide a measure of uncertainty for data that have red noise spectra, such as soil moisture.

We have applied this approach to estimate the error in the networks represented in ISMN and NASMD. Figure 2 shows the relative random error as the square root of the ratio in Eq. (2) estimated across all stations in each network for the two databases. The data from all networks are interpolated to the four Noah2.7 model layers: 0–10, 10–40, 40–100, and 100–200 cm. Recall that for the same networks ISMN and NASMD do not always contain the same stations or span of years. As a result, the estimated observational errors for the same networks in the two databases do not match exactly—the differences may be taken as representative of the uncertainties in applying this method.

Certain features are apparent nevertheless. As found by Gruber et al. (2013), observational errors are generally largest in the surface layer and decrease with depth. There is also a distinct difference between networks, and in fact between types of instrumentation. The GPS reflection method of the PBO H₂O network appears to result in a large relative random error of measurement

of 0.35 for surface soil moisture. Relative random error in the cosmic-ray neutron method of the COSMOS network is nearly as large at 0.32. Some of this may not be truly random error but rather due to the fact that the effective measurement depth varies with soil moisture content, so the static station measurement depths used here introduce additional error.

Dielectric probes are inexpensive and thus the most widely used. They have a relative random error of about 0.18 for near-surface measurements, dropping to 0.12 below 1-m depth. Heat-dissipation instruments appear to be the most stable, with a surface relative random error of 0.15 dropping to around 0.07 at depth. There is a great deal of variation among networks using the same class of instrumentation. For the Center for Hurricane Intensity and Landfall Investigation (CHILI) network, which places dielectric instruments only at 1-m depth, random error appears exceptionally large. The Soil Moisture Sensing Controller and Optimal Estimator (SoilSCAPE) network also appears to have unusually large random errors.

A further aspect of this approach to error estimation is that we can define representative profiles of a for classes of instruments, networks, or stations, allowing us to estimate a “corrected” soil moisture memory for comparison to models, which by their nature do not suffer from random error in their reported state variables. This can be accomplished by shifting the linear best fit of $\ln(r)$ by a so as to intersect 0 at $\tau = 0$; the corrected estimate of memory from observations is then the lag at which the adjusted $\ln(r) = -1$.

5. Spatial consistency between models and observations

The high-density SoilSCAPE network contains several sets of instruments or “nodes” clustered in groups within areas of $<1 \text{ km}^2$ in several locations. At these locations it may be assumed that the stations are so close together that their separations are well within the meteorological spatial scale over which precipitation time series decorrelate. Variations in soil moisture time series from node to node should be due to variations in the hydrologic properties of the sites of each node and random measurement error.

There are also parts of the United States where the Soil Climate Analysis Network (SCAN) and Snowpack Telemetry (SNOTEL) have multiple stations separated by distances larger than SoilSCAPE, but within a typical global climate model grid box $O(100) \text{ km}$. Variations on these scales may begin to be determined by differences in the meteorological time series at each station that are not represented explicitly by global models

because they are at the subgrid scale. Thus, data from these sites may allow us to see how this bridging of scales affects the representativeness of climate model soil moisture time series and shed light on how to compare models to observations across scales.

First, we consider how averaging together the time series of multiple proximate stations affects the statistics of the time series. The assumption is that a model grid-box time series is more representative of the average of multiple stations within that grid box than single stations. If statistics are found to converge as more stations are added to the average, and a general scaling factor is found to apply, we may have a means to translate single-station statistics to model gridbox statistics and vice versa. If the statistics do not appear to be sensitive to the number of stations included in the average, then scaling may not be necessary to compare point measurements to model gridbox values.

We examine data from four of the SoilSCAPE sites that include the most nodes and longest time series. These are located near Canton, Oklahoma (nodes numbered in the 100s); Tonzi Ranch, California (400s); and New Hogan Lake, California (500s and 700s). Across all sites the maximum separation of any two nodes is 503 m, and the median separation ranges from 166 to 214 m. Data span 3 years from 2011 to 2014. We include 19 nodes from the 400s site, 18 from the 500s, 21 from the 100s, and 13 from the 700s. Also examined are data from a cluster of 12 SCAN sites in and around northern Alabama that span 2002–14 and range in separation from 6 to 118 km, with a median separation of 52 km.

Calculations were performed using data vertically interpolated to the Noah2.7 model levels—results shown here are confined to the top 10-cm layer results, but lower-level results are consistent with these. We also examined data for the entire year, and separately by season. We find the mean and standard deviation in time for time series from each node at a site, then for each combination of two nodes averaged together only for days when each has no missing data, then for combinations of three, four, etc., up to the series where all nodes at a site are averaged together. We also calculate the variance among each statistic calculated with the same number of nodes in the combination. Because of missing data for different dates at various nodes, the total amount of data in the calculations dwindles as we go to larger and larger combinations. Figure 3a shows how data completeness drops from nodes considered one at a time through larger and larger combinations. We also constructed an abbreviated complete time series by taking a subset of the soil moisture data from SCAN stations in and around northern Alabama—10 stations

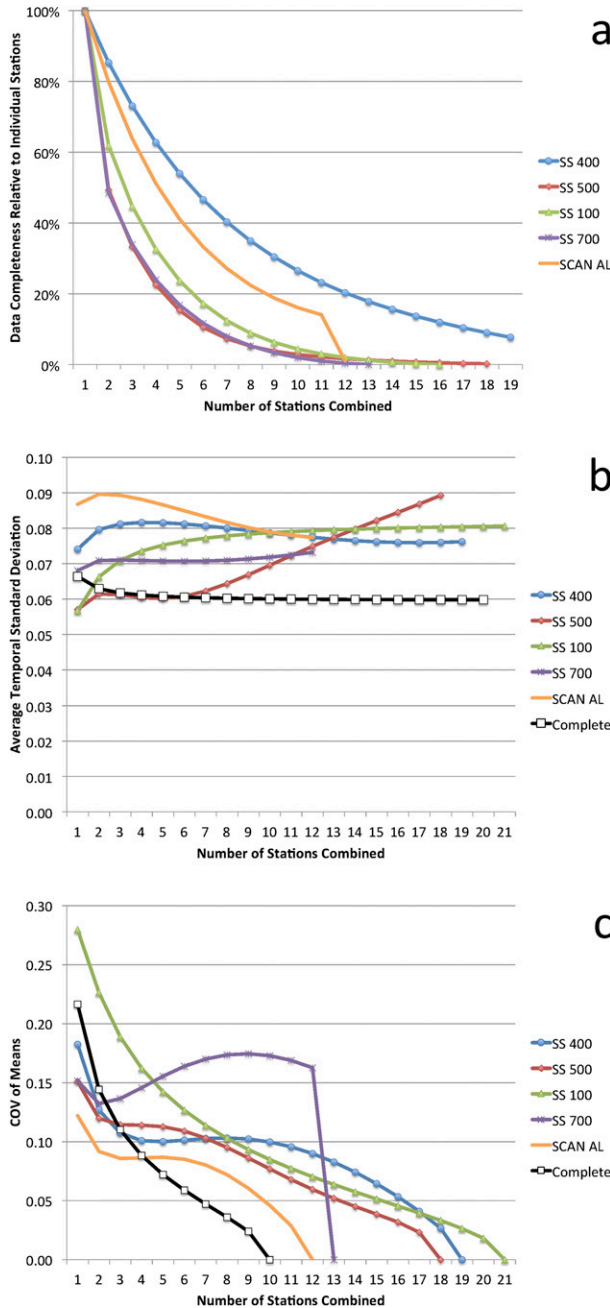


FIG. 3. Aggregation statistics for observational networks (SS are sets of nodes from SoilSCAPE, 100 is a set of stations in Oklahoma, the others are in California; SCAN AL is a set of SCAN stations across northern Alabama) with closely located stations. Values for each number of stations averaged together (abscissa) is the mean for all combinations with the indicated count: (a) number of days of data available when any station with missing data results in missing data for the combined average, (b) the estimated std dev in time of daily soil moisture (the mean for each month is first subtracted to remove interannual variability), and (c) the COV across combinations of the combination means. Black line is from a subset of SCAN AL with no missing data from any station for 306 consecutive days.

a that have complete data from 22 March 2007 through 21 January 2008. This is used to compare the impact of missing data on statistics.

Figure 3b shows how the average standard deviation in time for daily surface soil moisture from all seasons changes as more nodes are averaged together. Aside from a small uptick when going from single stations to combinations of two, there is little systematic change and the curves are remarkably flat. The complete data suggest a slight drop in the average standard deviation of ~10% from nodes considered individually to all considered together. Station data during summer only is more apt to show a slight rise in standard deviations with more combined nodes, whereas it is flatter in spring and fall (not shown). Overall, it would seem that the variability of daily soil moisture time series is not very sensitive to spatial scaling and aggregation over $O(100)$ m to $O(<100)$ km, and it appears model gridbox soil moisture can be reasonably validated against single-site data for this metric.

This is heartening, since there is certainly sensitivity of the time mean to aggregation of nodes. Figure 3c shows the coefficient of variation (COV) of mean soil moisture across the various numbers of combinations of nodes. There is a general drop in COV as more nodes are included in the averaging. The flattening of the slope of the curve around the middle values of combined nodes followed by an inflection and steepening of the curves again as nearly all nodes are included appears not to be solely an artifact of the drop in data completeness, although it appears to be less pronounced in the complete data subset.

Soil moisture memory as defined in section 3 is also examined. Missing data affect the estimation of memory as lagged autocorrelations can only be estimated when data are not missing on consecutive days (or 2 days apart for lag-2 estimates). As we average more stations together, there are more days when at least one station is missing data and the sample size decreases more steeply than for mean or variance.

Figure 4 shows estimates of top 10-cm soil moisture memory as a function of the number of stations combined for four SoilSCAPE locations, SCAN stations over northern Alabama, and the complete subset of those same SCAN stations. There are two curves for each set of stations. The solid curves show the median value of soil moisture memory calculated separately across all combinations of stations taken N at a time, N indicated on the abscissa. The dotted line is the memory calculated from the average 1- and 2-day lagged autocorrelations across all combinations.

There is a clear separation between the three California sites and those east of the Rockies. The California

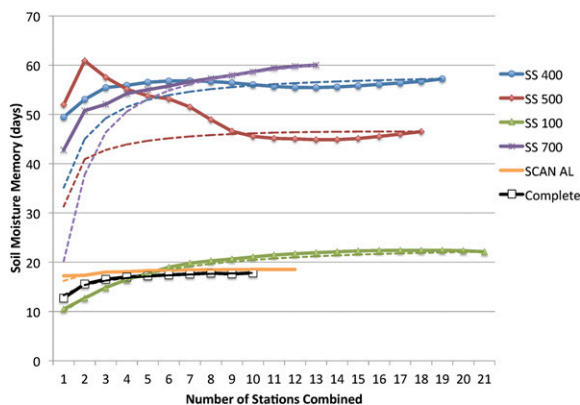


FIG. 4. As in Fig. 3, but for soil moisture memory (see text for definition). Solid lines are the median value of memory among all combinations, and dotted lines show the harmonic mean across all combinations.

sites have the longer memories, which is logical as California has a prolonged dry season when soil moisture anomalies may persist for months relative to the climatological annual cycle (though we caution that some of the memory may reflect submonthly facets of the climatological seasonal cycle, which were not removed by the aforementioned subtraction of monthly means from the data). They also show discrepancies between the two approaches to estimating memory for small numbers of combined stations. The estimates converge in all cases for greater numbers of combined stations. The Oklahoma and Alabama sites have shorter memories, consistent with their year-round likelihood for precipitation and general lack of a dry season. They also show very high agreement between the median memory and the memory calculated from the mean lagged autocorrelations. In all cases the memory calculated from the mean lagged autocorrelations increases as more stations are averaged together, by 14% for the SCAN Alabama sites to nearly 200% for the SoilSCAPE 700 nodes. Behavior of the medians is less consistent, as values go up or down as the number of combined stations increases.

Across this limited number of sites there is not an obvious relationship between characteristics of the station data (e.g., completeness shown in Fig. 3a) and the type or degree of change from single stations to inclusion of all stations. It would not be advisable based on these results to propose a method to scale memory calculated from individual stations to gridbox averages. It may be more judicious to compare a number of stations averaged over a certain area to model output averaged over the same area, although as shown later there are suspicious systematic differences between observational networks as well.

6. Variance and memory

In this section, we begin comparing model estimates of soil moisture statistics across the conterminous United States with statistics from collocated in situ measurements. For each model configuration, stations are compared to the model grid box that contains them. Table 5 shows the correlations calculated for each of the 19 networks listed in Table 1 between model and observed intraseasonal (calculated monthly and averaged for seasons) standard deviations of daily soil moisture. Correlations are then averaged across the networks, weighted by the number of stations in each network that went into the calculation. Because of the varying numbers of stations and areal extents of the different networks, it is not feasible to assign statistical significance to the averaged correlations. Networks are separated into local and regional extents because we have noticed a rather systematic separation in the correlations: uniformly, the models verify poorly with the local (mostly state level) networks in terms of the spatial pattern of soil moisture variability, but verify relatively well with the regional and national networks. This suggests that patterns of variability driven by the varying climate regimes across the United States are somewhat well reflected by the models, but smaller-scale variations over a few hundred kilometers or less are not captured. These smaller-scale patterns are likely more determined by variations in soils and landscape that are poorly represented by LSMs at gridbox scales.

Other patterns are evident in Table 5. Correlations between models and observations are generally highest for the shallow layer below the surface layer, which ranges in thickness from ~ 2 (Catchment and CLM4) to 10 cm (Noah2.7). There is no indication that the model output where near-surface meteorology is dictated by observations (offline and reanalysis) is better than free-running models. This is somewhat surprising, as there are acknowledged problems with global model simulations of precipitation, and precipitation quality is the major control on soil moisture variations (Guo et al. 2006; Wei et al. 2008). The implementation of HTESSEL in ERA-Interim used a single loamy soil texture globally, which may mute discrepancies because of soil property disagreement with observational sites and greater resemblance to forcing (precipitation) patterns, thus increasing correlation. Last, the various configurations of the Catchment LSM represent this particular metric of soil moisture variability more poorly than any of their counterparts. We should note, however, that soil moisture products from the Catchment LSM have been tested extensively against in situ measurements in other studies (Liu et al. 2011; de Lannoy

TABLE 5. Spatial pattern correlations of subseasonal JJA std dev of daily soil moisture between stations and model grid boxes for indicated models. Estimates are grouped by extent of network (local vs regional) and averaged across networks weighted by number of stations used in each network. Surface refers to the top LSM soil level, except CLM4, where it is the total for the top three layers; shallow is the second layer except for CLM4, where it is layers 4–6; deep is the third layer for Noah2.7 and HTESSEL and layers 7–8 for CLM4.

		Local				Regional			
		GLDAS	CFSR	CFS	20CR	GLDAS	CFSR	CFS	20CR
Noah2.7	Surface	-0.13	0.06	-0.01	0.06	0.38	0.40	0.30	0.25
	Shallow	0.08	0.09	-0.06	0.06	0.39	0.34	0.33	0.30
	Deep	0.09	0.00	-0.09	0.00	0.25	0.21	0.22	0.18
Catchment		MERRA Land	MERRA	GEOS-5		MERRA Land	MERRA	GEOS-5	
	Surface	-0.13	-0.11	-0.09		0.08	-0.04	0.21	
	Shallow	0.07		0.10		0.29		0.21	
HTESSEL		ERA-Interim Land	ERA Interim	IFS		ERA-Interim Land	ERA Interim	IFS	
	Surface	0.07	0.03	0.12		0.40	0.50	0.47	
	Shallow	0.09	0.03			0.34	0.43		
CLM4		CLM4	CCSM4		CLM4	CCSM4			
	Surface	0.14	0.07		0.41	0.39			
	Shallow	0.11	0.05		0.33	0.34			
Multimodel		Offline	Reanalysis	Free		Offline	Reanalysis	Free	
	Surface	-0.01	0.01	0.09		0.31	0.28	0.35	
	Shallow	0.09	0.06	0.12		0.34	0.36	0.29	
	Deep	0.03	0.00	0.04		0.27	0.24	0.22	

et al. 2014) and, in terms of capturing the time variability of soil moisture variation at a wide variety of sites, the model performs better (average time correlations of about 0.5) than suggested by the present metric, which focuses instead on the spatial

correlation against observations of the temporal standard deviation.

Figure 5 displays the network means and model biases in the daily standard deviation of JJA surface volumetric soil moisture (top three layers for CLM4) for each

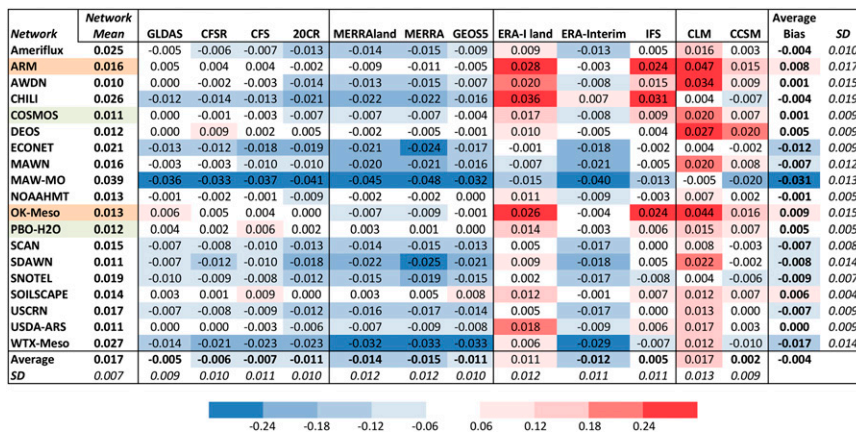


FIG. 5. Average std dev of daily surface volumetric soil moisture (dimensionless) during JJA for stations in each network (first column after network names), and the average bias of std dev for each model across the station locations of each network. Average and std dev of each network and model are shown at the right and bottom of the grid, respectively. Color shading is to aid recognition of negative (blue) and positive (red) biases. Two networks that employ heat-dissipation sensors are shaded orange, two networks that use near-field remote sensing techniques are shaded green, and all others use some form of dielectric probe.

TABLE 6. As in Table 5, but for soil moisture memory.

		Local				Regional			
		GLDAS	CFSR	CFS	20CR	GLDAS	CFSR	CFS	20CR
Noah2.7	Surface	0.11	0.11	0.16	0.17	0.30	0.20	0.18	0.22
	Shallow	-0.07	0.05	0.06	0.15	0.41	0.30	0.21	0.21
	Deep	-0.03	-0.07	-0.04	-0.01	0.15	0.17	0.12	0.10
Catchment		MERRA Land	MERRA	GEOS-5		MERRA Land	MERRA	GEOS-5	
	Surface	0.23	0.18	0.20		0.17	0.20	0.16	
	Shallow	0.05		0.04		0.04		0.09	
HTESSEL		ERA-Interim Land	ERA Interim	IFS		ERA-Interim Land	ERA Interim	IFS	
	Surface	-0.09	0.19	0.16		0.14	0.08	0.06	
	Shallow	0.09	0.06			0.15	0.25		
	Deep	0.02	-0.07			0.10	0.29		
CLM4		CLM4	CCSM4			CLM4	CCSM4		
	Surface	0.08	-0.02			0.15	0.08		
	Shallow	-0.03	-0.09			0.25	0.08		
	Deep	-0.02	-0.01			0.15	0.02		
Multimodel		Offline	Reanalysis	Free		Offline	Reanalysis	Free	
	Surface	0.08	0.16	0.16		0.19	0.17	0.12	
	Shallow	0.01	0.09	0.11		0.21	0.25	0.13	
	Deep	-0.01	-0.05	0.05		0.09	0.19	0.07	

model configuration and network in a color-coded tabular form. There are clear systematic biases in the representation of soil moisture variability among the models. Offline versions of both HTESSEL (ERA-Interim Land) and CLM4 exhibit excessive variance of soil moisture. For the Noah2.7 LSM the offline version (GLDAS) also has the highest variance, but still has a low overall bias across all networks. The reanalyses tend to exhibit the lowest variability, although the Climate Forecast System Reanalysis (CFSR) is essentially undistinguishable from the offline or free-running (CFS) simulations. Among the four model groups, CLM4 has the strongest positive biases and Catchment (MERRA and GEOS-5) has the strongest negative biases.

The vagaries of validation with multiple observational networks are also evident when one compares the different rows of Fig. 5. Models exhibit the strongest positive biases for the two networks that employ heat-dissipation sensors, ARM and OK-MESO, suggesting these sensors may behave differently than other types of measurements. All models show negative biases for the Missouri network, and most also show negative biases for the West Texas network. Meanwhile, biases are generally positive for the Delaware Environmental Observing System (DEOS); WTX-MESO and DEOS employ the same model of dielectric soil moisture probe at the same depth for surface soil moisture (5 cm), so the systematic differences are more likely because of disparities between the gridded soil parameter datasets commonly used by models and actual local conditions.

Soil moisture memory is calculated as described in section 3. The extrapolation procedure from data during a season can result in e -folding time scales for lagged autocorrelations that vary over two or more orders of magnitude in some cases. Accuracy of long-memory estimates is particularly tenuous, so all averaging is done in terms of frequency rather than time, and the inverse of the result is taken to give a memory time scale in days (i.e., the harmonic mean is used).

Table 6 shows how well the spatial patterns of soil moisture memory agree between models and observations. As in Table 5, the network results are grouped by extent: local versus regional/national. As with soil moisture variance, models represent large-scale patterns of memory better than intrastate variations. However, correlations are generally lower for memory than for the standard deviation of soil moisture. The highest skill is exhibited for surface soil moisture memory among the local networks and for shallow (~10–50-cm depth) soil moisture memory at larger scales. Free-running land-atmosphere models perform worst at simulating large-scale patterns of soil moisture memory—this could be because of errors in the temporal spectrum of precipitation in models (cf. Wei et al. 2010; Dirmeyer 2013). Interestingly, free-running models do the best at representing local network variations, but although the correlations are generally statistically significant because of the large number of stations included, they do not suggest practical

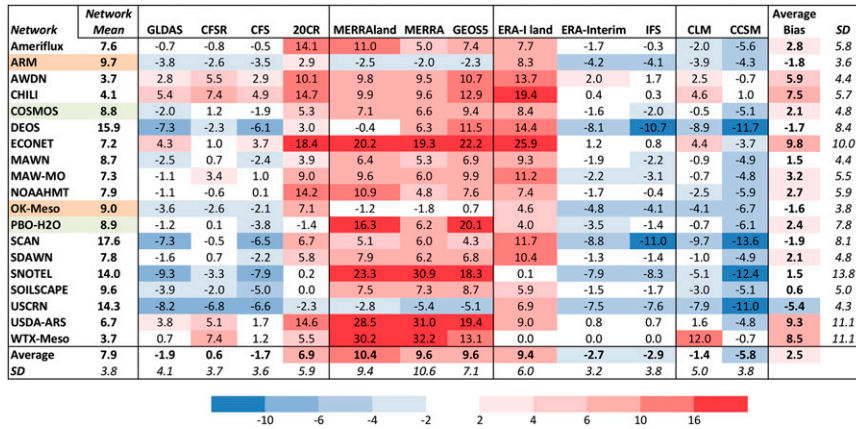


FIG. 6. As in Fig. 5, but for surface soil moisture memory (days).

usefulness. Various configurations with the Noah2.7 LSM show the best pattern correlations.

Figure 6 presents network-by-network comparisons of JJA surface soil moisture memory biases in the same manner as Fig. 5. The mean memory for different networks varies from less than 4 to more than 17 days, but there is considerable variation within each network. Model biases can be substantial. Model configurations using the Catchment land surface scheme predominantly exhibit strong positive biases, suggesting that the model is overly persistent in soil moisture anomalies, despite the fact it has the thinnest surface layer of the four LSMs. As noted in section 2d, Catchment has a distinctly nontraditional structure; its implementation of an explicit treatment of subgrid soil moisture variability is known to tie together strongly its diagnosed surface and subsurface soil moisture variables (Kumar et al. 2009), which can in fact produce artificially high values of memory in surface and shallow layers. On the other hand, CLM4 (for which the top three layers have been combined to represent the surface) largely underestimates soil moisture memory. Noah2.7 largely underestimates memory as well, except in the Twentieth Century Reanalysis (20CR), where the spatial resolution is considerably coarser than the other implementations, and the large ensemble approach to production (Compo et al. 2011) may have a bearing on hydrologic variables. Another model that shows inconsistent behavior among implementations is HTESSEL—the offline ERA-Interim Land simulation has consistently negative biases in soil moisture, mirrored also in ECMWF atmospheric coupled integration with data assimilation (Albergel et al. 2012, 2013), while coupled IFS simulations from the Athena project (Kinter et al. 2013) and ERA-Interim exhibit a slightly positive bias, highlighting the impact of both modeling and assimilation system changes in determining biases.

Finally, error profiles from network to network are fairly consistent, suggesting again that discrepancies exist between gridded datasets of land surface parameters used by the models and conditions at the station sites. An exception is SNOTEL, in which stations are largely positioned in high mountain locations across the western United States that may tend toward thinner and rockier soils than global gridded soil datasets specify or LSMs represent. Catchment has some of its largest positive biases over this network where other models generally have some of their strongest negative biases.

7. Spatial scales

Finally, we examine the spatial decorrelation scales in the observational networks and models. This approach is highly analogous to the temporal scaling we defined as “memory” in section 6, and follows similar principles (Vinnikov et al. 1999; Entin et al. 2000). In the case of decorrelation of soil moisture time series over space, we have three factors: decorrelation over meteorological scales of tens to hundreds of kilometers related to the spatial scales of the forcing of soil moisture variability, particularly precipitation; decorrelation over catchment hydrologic scales of meters to hundreds of meters brought about by variations in soil properties and sampling of different regimes along hillslopes; and random measurement errors as characterized in section 4.

As a check, we look first at the SoilSCAPE sites, whose nodes are close enough together to be well within the meteorological scales. We find essentially no relationship in any season between the correlation of time series of soil moisture from pairs of stations and their separation distances, which range from about 20 to 500 m. The implication is that evidence for the “catchment hydrologic scale” of Vinnikov et al. (1999) is swamped by the random error in measurements.

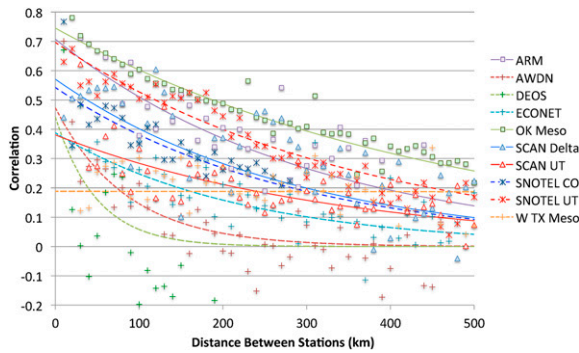


FIG. 7. Spatial autocorrelation of daily surface layer soil moisture during JJA as a function of separation between stations for several networks, binned in intervals of 10 km. Lines indicate linear best fits through the binned log of correlation (negative values omitted).

For other networks with greater distance between stations, a systematic relationship between station separation and correlation emerges. Figure 7 shows correlation as a function of station separation during summer for several state networks and larger networks that have numerous stations in a single state (“Delta” refers to all stations in a region centered on the Mississippi River, spanning 32°–35.1°N, 93°–85.5°W, and “UT” and “CO” refer to stations within Utah and Colorado, respectively). For all networks except the West Texas Mesonet there is a clear decrease in correlation with distance. The two networks using heat-dissipation sensors, ARM and the OK-MESO, have the two largest values of extrapolated correlation at distance 0, suggesting in a way similar to Fig. 2 that these sensors have small random errors. They also show higher correlations at larger distances than neighboring Automated Weather Data Network (AWDN) (Nebraska) or other dielectric sensor networks in the Midwest or eastern United States. The western networks (SCAN UT, SNOTEL CO, and SNOTEL UT) tend to show high

correlations at large separations like ARM and OK-MESO, likely indicative of the relatively rare precipitation in those areas during summer. The two networks overlapping in Utah (SCAN UT and SNOTEL UT) appear to show very different spatial correlation scales, but when they are adjusted for their different apparent random errors (discussed below), they become quite comparable. SCAN stations are located mainly in agricultural valleys and lowlands while SNOTEL stations are predominantly in mountains at higher altitudes; it is unclear how this may affect error estimates.

Figure 8 presents a color-coded table of model comparisons to network estimates of the spatial decorrelation distance (km) for surface soil moisture during JJA, defined as the separation between stations where zero-lag temporal correlation of time series of daily soil moisture drops to $1/e$ based on a best-fit regression of $\ln(r)$ against distance. For the stations, the lines are shifted so that the intercept at distance 0 is at a correlation of 1, adjusting for the effect of random measurement error on correlations. This has the effect of increasing the spatial decorrelation distance. No such adjustment is necessary for model output, where distances are calculated for each grid box relative to the eight surrounding grid boxes with distances calculated between the centers of each grid box. Both standard and harmonic means are given across models, and the standard deviation is relative to the standard mean.

Many difficulties in estimating spatial decorrelation scales consistently across networks, and comparing models to them, are evident. The spatial decorrelation distances for each network follow from what was shown in Fig. 7. Memories are longer in the west (typically several weeks as opposed to ~1 week in the central and eastern United States), and they are long for the two networks with heat-dissipation sensors (shaded orange). AWDN (Nebraska) and DEOS (Delaware) show much

Network	Network Distance	Network								Model				Harmonic		
		GLDAS	CFSR	CFS	20CR	MERRAland	MERRA	GEOS5	Era-I land	ERA-Interim	IFS	CLM4	CCSM	Mean	Mean	SD
ARM	307	164	546	536	1007	607	703	533	608	538	1054	751	858	659	529	230
AWDN	85	209	489	444	579	509	691	320	603	558	792	755	617	547	480	162
DEOS	49	326	512	421	670	635	737	387	635	607	590	959	547	585	541	162
ECONET	224	329	580	445	557	719	781	325	646	628	625	1004	530	591	538	175
OK-Meso	471	150	506	553	953	675	679	480	751	656	867	795	667	658	522	207
SCAN-Delta	285	468	734	501	637	601	660	303	928	813	860	1034	671	684	616	197
SCAN-UT	341	791	264	476	656	410	364	272	487	582	750	660	521	519	460	167
SNOTEL-CO	283	264	289	398	629	356	239	224	518	557	639	570	344	419	366	149
SNOTEL-UT	360	791	264	476	656	410	364	272	487	582	750	660	521	519	460	167
WTX-Meso	61518	163	493	519	845	531	689	476	819	686	748	725	691	615	509	182
Average		365	468	477	719	545	583	360	648	621	767	791	613	580	502	
Corr (no WTX)		0.22	-0.23	0.67	0.52	-0.12	-0.37	0.15	0.04	0.16	0.45	-0.35	0.31	0.14	-0.11	
Probability (1-tailed)		57%	#N/A	5%	15%	#N/A	#N/A	70%	91%	68%	23%	#N/A	42%	72%	#N/A	

FIG. 8. Average spatial decorrelation distance (km) where zero-lag temporal correlation of time series of daily soil moisture drops to $1/e$ for surface volumetric soil moisture during JJA for stations in each network (first column after network names), and models. Colors indicate percentage deviation of each model (averaged over domain of network) relative to in situ network estimates with estimated random measurement error removed. SD is std dev among models.

shorter decorrelation distances than other networks. The contrast between AWDN and both ARM and OK-MESO is especially troubling, since they are all close to each other in the Great Plains. Here the different performance of various types of sensors is obvious. Compare to the SCAN UT and SNOTEL UT networks over Utah, which use similar instruments and are within 6% of each other once adjusted for random measurement error.

Looking at the models, there is a tendency for overestimation of the decorrelation distance. This may be related to the differences in scaling between the point observations and model grid boxes. Even though many pairs of stations are used to estimate the correlations as a function of separation shown in Fig. 7, they still represent point-versus-point correlations, and not area-versus-area as is implied in model gridbox values. Comparing sets of stations averaged over areas comparable to adjacent model grid boxes might provide a more proportionate estimate, but we immediately face the problem of a lack of station density over all but a few areas. We saw in section 6 that the estimation of memory, which is a closely related calculation to decorrelation distance (Vinnikov et al. 1996), is not scale resilient like variance.

The row of Fig. 8 labeled “Corr (no WTX)” is the correlation between model and network estimates across the first nine networks listed. The curve fit for WTX-MESO is so flat (see Fig. 7) that a ridiculously long distance is projected for this network, so it is not included in the correlation. The probability shown is the likelihood the correlation could be arrived at by chance (with 7 degrees of freedom); a one-tailed test is made as there is no inherent value in a negative correlation. The free-running CFS simulation would appear to perform well, with a correlation of 0.67 and only a 5% chance of arriving at this correlation randomly. However, out of 12 models, it is not surprising to see one with a probability below 8%–10%, and the slope of the fit through the scatter is nearly 1:4, far from 1:1. The harmonic mean of all models produces a lower positive bias than a straight mean, but a negative correlation across the networks.

Among the models there is a great deal of variation. There is not a consistent behavior of offline LSMs relative to coupled models, or reanalyses relative to free-running GCMs. However, the suite of models using the Catchment LSM tends to have shorter decorrelation distances than either the HTESSEL or CLM4 sets of models. The Noah2.7 set spans a wide range, varying by a factor of 2 from offline GLDAS to the low-resolution 20CR simulation.

8. Conclusions and discussion

In this study, we have confronted a number of LSMs in both coupled and uncoupled modes with in situ soil

moisture measurements from a number of independent networks over the conterminous United States to 1) determine the feasibility and pitfalls of such a comparison and 2) see what can be learned about model and observational data. We first investigate characteristics of the observational data, with particular attention to how they differ between networks (possibly because of differences in instrumentation), in space, and over time. We then test approaches to compare model output to the observational data.

We examine three statistical metrics: variability as measured by the standard deviation of daily soil moisture, memory represented by the time it takes lagged autocorrelation of daily soil moisture to drop to $1/e$, and spatial scale calculated like memory as the distance over which unlagged correlations between soil moisture measurements or model estimates drop to $1/e$. For measurement networks with many closely located stations within the area of a typical model grid box, we find that aggregation of many stations (arguably more representative of gridbox-average values represented by LSMs) has little effect on the standard deviation, but does change estimates of memory in nonsystematic ways. Data completeness can affect aggregation, but not in a clearly predictable way. Although not directly investigated here, there is evidence that spatial scale estimates are also sensitive to the combination of stations. We conclude that modeled soil moisture variability can be safely validated against data from single stations, but other metrics cannot.

Another caveat regarding in situ soil moisture data is that there can be clear differences between the statistical properties of data from different networks that are in the same or adjacent locations. This mirrors what is already established for soil moisture products between models (Koster et al. 2009). In some cases this seems to be caused by the type of sensor used. Buried probes seemed to exhibit less random error than near-field remote sensing techniques, although we do find networks with dielectric sensors that appear to have large random errors. Heat-dissipation sensors have generally low random error, but also lower day-to-day variability and connote longer soil moisture memory than dielectric probes, even after random errors are accounted for. Random measurement errors generally decrease with depth for buried sensors. There are also differences in estimated random errors between networks with essentially the same instrumentation, suggesting differences in calibration and maintenance may also be a factor.

Models show systematic biases in near-surface soil moisture metrics (Figs. 5, 6, and 8). The model configurations with the Catchment LSM all show too little variability and long memory, but in the case of

free-running GEOS-5 simulations, an unusually short spatial scale in some cases. [Bosilovich \(2013\)](#) showed some precipitation comparisons for MERRA indicating too little interannual variability. CLM4 tends toward excessive variability and short memory. The characteristics of variability and memory are not always in opposition—ERA-Interim Land has positive biases in both while ERA-Interim has negative in both. These features can be attributed to the modeled soil moisture range lacking spatial variability in ERA-Interim, as documented in the soil hydrology revision ([Balsamo et al. 2009](#)). An interesting dependency on spatial resolution can be deduced from the IFS integrations within the Athena project that used a 4-times-higher spatial resolution, enhancing the match to in situ standard deviations of daily surface volumetric soil moisture ([Fig. 5](#)), memory ([Fig. 6](#)), and spatial decorrelation distance ([Fig. 8](#)).

The various implementations of the Noah2.7 LSM range between low and high biases, but often have the lowest biases across all networks. For both variability and memory the models show higher spatial correlations across stations within regional to national networks than to state-level networks ([Tables 5, 6](#)), suggesting they reflect large-scale hydroclimate patterns better than local ones. The specification of soil hydrologic properties in models is much coarser than natural heterogeneity of the soils in most regions, reflecting another aspect of subgrid variability in large-scale models that is poorly represented and confounds validation. Free-running coupled land-atmosphere models perform worst at simulating large-scale patterns, possibly because of the poor simulation of precipitation spectra by GCMs.

Spatial scales were found to be particularly difficult to diagnose. For the SoilSCAPE network where stations are separated by only $O(100)$ m, there appears to be no dependence of interstation correlation on distance between stations. Theory suggests there is a “catchment hydrologic scale” in this range ([Vinnikov et al. 1999](#)), but measurements seem to be dominated by random errors that obscure evidence of it. Meteorological-scale decorrelation over ranges of $O(10\text{--}1000)$ km is clear in observations and models, but we find the heat-dissipation sensors seem to imply longer spatial scales, and different networks of dielectric sensors can give very wide ranges of estimates that appear unrealistic. CLM4, Catchment, and HTESSEL each represent subgrid surface variations to differing extents, but this does not translate well to representation of observed subgrid variability (cf. [Bosilovich 2002](#)). To put comparisons between models and observations on the same footing for this metric, stations should first be aggregated into average time series over areas comparable to model grid boxes before estimating spatial decorrelation distances.

Overall, statistical vagaries between different soil moisture networks using different types of sensors and measurement techniques suggest great caution is needed when using these data for validation, calibration, or data assimilation. The typical assumption that model errors are large while observational errors are small may not apply readily for soil moisture. This is particularly important as the LSM community moves toward a more rigorous benchmarking approach (e.g., [Best et al. 2015](#)) for fluxes and other variables such as soil moisture. The results here also suggest statistical considerations that should be applied when extending model evaluation or benchmarking to two dimensions rather than at a single point. One must be very careful about scaling issues—everything possible should be done to put data from different sources on the same footing before comparison. Only temporal variability seemed to be insensitive to the differences in scale between point measurements and model gridbox values.

Note that this study is largely exploratory and we opt to go wide rather than deep; instead of giving a highly detailed examination of a particular soil moisture metric, observational network, or model, we traverse data and metrics from multiple observational networks and models. Our aim is to explore the problems and pitfalls, as well as bring to light the areas of promise for validation of models with observed soil moisture data. The indicated biases potentially indicate deficiencies in the land models. However, they may also reflect the model-specific character of a given land models' soil moisture representation. Given the necessity of computing fluxes averaged over large and complex domains with limited spatial resolution in soil moisture description, soil moisture in land models is arguably better interpreted as a model-specific index of wetness than a variable that can be directly compared to observations ([Dirmeyer 2004](#); [Koster et al. 2009](#)). For this reason, modeled soil moisture is known to have model-specific magnitudes (and yet still function well in climate models); by the same token, standard deviations of soil moisture will also necessarily be model specific. This point underscores a major difficulty faced when confronting land models with such observations.

The work presents several means to approach the assessment of model soil moisture behavior with in situ observations with particular focus on spatiotemporal inconsistencies. The problem is analogous to that faced in operational data assimilation, where observations from a wide range of sources with different spatiotemporal coverage and error characteristics must be harmonized to generate useful analyses. Key to such approaches is a large and robust set of calibration and validation data; none of the networks examined here are due to be

discontinued, and more networks are coming online and being synthesized into NASMD and ISMN every year, so the situation should only improve. Furthermore, satellites can provide spatially continuous measurements at scales comparable to model grid boxes. Current missions are beginning to provide such information, but better temporal coverage at higher spatial resolution, maintained uninterrupted over decades, will provide, in combination with in situ measurements and synthesis through data assimilation, the best overall monitoring and initialization for forecasts. Despite this, models can still be improved using the growing observational record to identify and improve processes and parameterizations that contribute to errors in the surface water cycle.

Acknowledgments. This work has been primarily supported by National Aeronautics and Space Administration Grant NNX13AQ21G. Funding for W.D. has come from SMOS Soil Moisture Network Study—Operational Phase (ESA ESTEC Contract 4000102722/10). Support for the Twentieth Century Reanalysis Project dataset is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office. We thank G. Compo for making 20CR data available to us.

REFERENCES

- Albergel, C., P. de Rosnay, G. Balsamo, L. Isaksen, and J. Muñoz-Sabater, 2012: Soil moisture analyses at ECMWF: Evaluation using global ground-based in situ observations. *J. Hydrometeorol.*, **13**, 1442–1460, doi:10.1175/JHM-D-11-0107.1.
- , and Coauthors, 2013: Skill and global trend analysis of soil moisture from reanalyses and microwave remote sensing. *J. Hydrometeorol.*, **14**, 1259–1277, doi:10.1175/JHM-D-12-0161.1.
- Andresen, J., L. Olse, T. Aichele, B. Bishop, J. Brown, J. Landis, S. Marquie, and A. Pollyea, 2011: Enviro-weather: A weather-based pest and crop management information system for Michigan. *Seventh Int. Integrated Pest Management Symp.*, Memphis, TN, IPM, 16.1. [Available online at http://www.ipmcenters.org/ipmsymposium12/16-1_Andresen.pdf.]
- Baldocchi, D., and Coauthors, 2001: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor and energy flux densities. *Bull. Amer. Meteor. Soc.*, **82**, 2415–2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.
- Balsamo, G., P. Viterbo, A. Beljaars, B. van den Hurk, M. Hirschi, A. K. Betts, and K. Scipal, 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *J. Hydrometeorol.*, **10**, 623–643, doi:10.1175/2008JHM1068.1.
- , and Coauthors, 2015: ERA-Interim/Land: A global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.*, **19**, 389–407, doi:10.5194/hess-19-389-2015.
- Bell, J. E., and Coauthors, 2013: U.S. Climate Reference Network soil moisture and temperature observations. *J. Hydrometeorol.*, **14**, 977–988, doi:10.1175/JHM-D-12-0146.1.
- Best, M. J., and Coauthors, 2015: The plumbing of land surface models: Benchmarking model performance. *J. Hydrometeorol.*, **16**, 1425–1442, doi:10.1175/JHM-D-14-0158.1.
- Bond, D., 2005: Soil Water and Temperature System (SWATS) handbook. Tech. Rep. ARM TR-063, U.S. Department of Energy, 24 pp.
- Bosilovich, M. G., 2002: On the use and validation of mosaic heterogeneity in atmospheric numerical models. *Geophys. Res. Lett.*, **29**, 1174, doi:10.1029/2001GL013925.
- , 2013: Regional climate and variability of NASA MERRA and recent reanalyses: U.S. summertime precipitation and temperature. *J. Appl. Meteor. Climatol.*, **52**, 1939–1951, doi:10.1175/JAMC-D-12-0291.1.
- Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, doi:10.1002/qj.776.
- Crow, W. T., F. Lei, C. Hain, M. C. Anderson, R. L. Scott, D. Billesbach, and T. Arkebauer, 2015: Robust estimates of soil moisture and latent heat flux coupling strength obtained from triple collocation. *Geophys. Res. Lett.*, **42**, 8415–8423, doi:10.1002/2015GL065929.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- de Lannoy, G. J. M., R. D. Koster, R. H. Reichle, S. P. P. Mahanama, and Q. Liu, 2014: An updated treatment of soil texture and associated hydraulic properties in a global land modeling system. *J. Adv. Model. Earth Syst.*, **6**, 957–979, doi:10.1002/2014MS000330.
- Delworth, T. L., and S. Manabe, 1988: The influence of potential evaporation on the variabilities of simulated soil wetness and climate. *J. Climate*, **1**, 523–547.
- Dirmeier, P. A., 2004: Soil moisture—Muddy prospects for a clear definition. *GEWEX News*, Vol. 14, No. 3, International GEWEX Project Office, Silver Spring, MD, 11–12.
- , 2013: Characteristics of the water cycle and land–atmosphere interactions from a comprehensive reforecast and reanalysis data set: CFSv2. *Climate Dyn.*, **41**, 1083–1097, doi:10.1007/s00382-013-1866-x.
- , X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki, 2006: The Second Global Soil Wetness Project (GSWP-2): Multimodel analysis and implications for our perception of the land surface. *Bull. Amer. Meteor. Soc.*, **87**, 1381–1397, doi:10.1175/BAMS-87-10-1381.
- , S. Kumar, M. J. Fennessy, E. L. Altshuler, T. DelSole, Z. Guo, B. Cash, and D. Straus, 2013: Evolution of land-driven predictability in a changing climate. *J. Climate*, **26**, 8495–8512, doi:10.1175/JCLI-D-13-00029.1.
- Dorigo, W. A., and Coauthors, 2011: The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrol. Earth Syst. Sci.*, **15**, 1675–1698, doi:10.5194/hess-15-1675-2011.
- , and Coauthors, 2013: Global automated quality control of in situ soil moisture data from the International Soil Moisture Network. *Vadose Zone J.*, **12**, doi:10.2136/vzj2012.0097.
- , and Coauthors, 2015: Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sens. Environ.*, **162**, 380–395, doi:10.1016/j.rse.2014.07.023.

- Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model: 2. Parameter estimation and model demonstration. *J. Geophys. Res.*, **105**, 24 823–24 838, doi:10.1029/2000JD900328.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarplay, 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.
- Entin, J. K., A. Robock, K. Y. Vinnikov, S. E. Hollinger, S. Liu, and A. Namkhai, 2000: Temporal and spatial scales of observed soil moisture variations in the extratropics. *J. Geophys. Res.*, **105**, 11 865–11 877, doi:10.1029/2000JD900051.
- Ford, T. W., and S. M. Quiring, 2014: Comparison and application of multiple methods for temporal interpolation of daily soil moisture. *Int. J. Climatol.*, **34**, 2604–2621, doi:10.1002/joc.3862.
- Gruber, A., W. A. Dorigo, S. Zwieback, A. Xaver, and W. Wagner, 2013: Characterizing coarse-scale representativeness of in-situ soil moisture measurements from the International Soil Moisture Network. *Vadose Zone J.*, **12**, doi:10.2136/vzj2012.0170.
- , C.-H. Su, S. Zwieback, W. Crow, W. Dorigo, and W. Wagner, 2016: Recent advances in (soil moisture) triple collocation analysis. *Int. J. Appl. Earth Obs. Geoinf.*, **45**, 200–211, doi:10.1016/j.jag.2015.09.002.
- Guinan, P. E., and J. S. Travlos, 2008: Missouri's transition to a near real-time mesonet. *17th Conf. on Applied Climatology*, Whistler, BC, Canada, Amer. Meteor. Soc., 1.2. [Available online at https://ams.confex.com/ams/13MontMet17AP/techprogram/paper_140960.htm.]
- Guo, Z., P. A. Dirmeyer, Z.-Z. Hu, X. Gao, and M. Zhao, 2006: Evaluation of GSWP-2 soil moisture simulations: 2. Sensitivity to external meteorological forcing. *J. Geophys. Res.*, **111**, D22S03, doi:10.1029/2006JD007845.
- Hubbard, K. G., N. J. Rosenberg, and D. C. Nielsen, 1983: Automated Weather Data Network for agriculture. *J. Water Resour. Plann. Manage.*, **109**, 213–222, doi:10.1061/(ASCE)0733-9496(1983)109:3(213).
- Illston, B. G., J. B. Basara, C. A. Fiebrich, K. C. Crawford, E. Hunt, D. K. Fisher, R. Elliott, and K. Humes, 2008: Mesoscale monitoring of soil moisture across a statewide network. *J. Atmos. Oceanic Technol.*, **25**, 167–182, doi:10.1175/2007JTECHA993.1.
- Jackson, T. J., M. H. Cosh, R. Bindlish, P. J. Starks, D. D. Bosch, M. S. Seyfried, D. C. Goodrich, and M. S. Moran, 2010: Validation of advanced microwave scanning radiometer soil moisture products. *IEEE Trans. Geosci. Remote Sens.*, **48**, 4256–4272, doi:10.1109/TGRS.2010.2051035.
- Kimball, S. K., M. S. Mulekar, S. Cummings, and J. Stamates, 2010: The University of South Alabama Mesonet and coastal observing system: A technical and statistical overview. *J. Atmos. Oceanic Technol.*, **27**, 1417–1439, doi:10.1175/2010JTECHA1376.1.
- Kinter, J. L., III, and Coauthors, 2013: Revolutionizing climate modeling with Project Athena: A multi-institutional, international collaboration. *Bull. Amer. Meteor. Soc.*, **94**, 231–245, doi:10.1175/BAMS-D-11-00043.1.
- Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *J. Geophys. Res.*, **105**, 24 809–24 822, doi:10.1029/2000JD900327.
- , Z. Guo, P. A. Dirmeyer, R. Yang, K. Mitchell, and M. J. Puma, 2009: On the nature of soil moisture in land surface models. *J. Climate*, **22**, 4322–4335, doi:10.1175/2009JCLI2832.1.
- Kumar, S. V., R. H. Reichle, R. D. Koster, W. T. Crow, and C. D. Peters-Lidard, 2009: Role of subsurface physics in the assimilation of surface soil moisture observations. *J. Hydrometeorol.*, **10**, 1534–1547, doi:10.1175/2009JHM1134.1.
- Larson, K. M., E. E. Small, E. D. Gutmann, A. L. Bilich, J. J. Braun, and V. U. Zavorotny, 2008: Use of GPS receivers as a soil moisture network for water cycle studies. *Geophys. Res. Lett.*, **35**, L24405, doi:10.1029/2008GL036013.
- Lawrence, D. M., and Coauthors, 2011: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *J. Adv. Model. Earth Syst.*, **3**, M03001, doi:10.1029/2011MS000045.
- Legates, D. R., D. J. Leathers, T. L. DeLiberty, G. E. Quelch, K. Brinson, J. Butke, R. Mahmood, and S. A. Foster, 2005: DEOS: The Delaware Environmental Observing System. *21st Int. Conf. Interactive Information Processing Systems*, San Diego, CA, Amer. Meteor. Soc., 18.10. [Available online at https://ams.confex.com/ams/Annual2005/techprogram/paper_87687.htm.]
- , —, —, —, and —, 2007: Delaware Environmental Observing System: An update. *23rd Int. Conf. Interactive Information Processing Systems*, San Antonio, TX, Amer. Meteor. Soc., P1.2. [Available online at https://ams.confex.com/ams/87ANNUAL/techprogram/paper_120626.htm.]
- Liu, Q., and Coauthors, 2011: The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system. *J. Hydrometeorol.*, **12**, 750–765, doi:10.1175/JHM-D-10-05000.1.
- Moghaddam, M., and Coauthors, 2010: A wireless soil moisture smart sensor web using physics-based optimal control: Concept and initial demonstration. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **3**, 522–535, doi:10.1109/JSTARS.2010.2052918.
- Pan, W., R. P. Boyles, J. G. White, and J. L. Heitman, 2012: Characterizing soil physical properties for soil moisture monitoring with the North Carolina Environment and Climate Observing Network. *J. Atmos. Oceanic Technol.*, **29**, 933–943, doi:10.1175/JTECH-D-11-00104.1.
- Qian, T., A. Dai, and K. E. Trenberth, 2007: Hydroclimatic trends in the Mississippi River basin from 1948 to 2004. *J. Climate*, **20**, 4599–4614, doi:10.1175/JCLI4262.1.
- Quiring, S. M., T. W. Ford, J. K. Wang, A. Khong, E. Harris, T. Lindgren, D. W. Goldberg, and Z. Li, 2016: North American Soil Moisture Database: Development and applications. *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-13-00263.1, in press.
- Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Touré, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate*, **24**, 6322–6338, doi:10.1175/JCLI-D-10-05033.1.
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:10.1175/JCLI-D-11-00015.1.
- Robock, A., K. Ya. Vinnikov, C. A. Schlosser, N. A. Speranskaya, and Y. Xue, 1995: Use of midlatitude soil moisture and meteorological observations to validate soil moisture simulations with biosphere and bucket models. *J. Climate*, **8**, 15–35, doi:10.1175/1520-0442(1995)008<0015:UOMSMA>2.0.CO;2.
- Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, doi:10.1175/BAMS-85-3-381.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:10.1175/2010BAMS3001.1.

- Schaefer, G. L., and R. F. Paetzold, 2001: SNOTEL (SNOWpack TELEmetry) and SCAN (Soil Climate Analysis Network). *Automated Weather Stations for Applications in Agriculture and Water Resources Management: Current Use and Future Perspectives*, K. G. Hubbard and M. V. K. Sivakumar, Eds., AGM-3, WMO/TD 1074, WMO, 187–194. [Available online at <http://www.wamis.org/agm/pubs/agm3/WMO-TD1074.pdf>.]
- , M. H. Cosh, and T. J. Jackson, 2007: The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN). *J. Atmos. Oceanic Technol.*, **24**, 2073–2077, doi:10.1175/2007JTECHA930.1.
- Schlosser, C. A., and P. C. D. Milly, 2002: A model-based investigation of soil moisture predictability and associated climate predictability. *J. Hydrometeorol.*, **3**, 483–501, doi:10.1175/1525-7541(2002)003<0483:AMBIOS>2.0.CO;2.
- Schroeder, J. L., W. S. Burgett, K. B. Haynie, I. Sonmez, G. D. Skwira, A. L. Doggett, and J. W. Lipe, 2005: The West Texas Mesonet: A technical overview. *J. Atmos. Oceanic Technol.*, **22**, 211–222, doi:10.1175/JTECH-1690.1.
- Stillman, S., J. Ninneman, X. Zeng, T. Franz, R. L. Scott, W. J. Shuttleworth, and K. Cummins, 2014: Summer soil moisture spatiotemporal variability in southeastern Arizona. *J. Hydrometeorol.*, **15**, 1473–1485, doi:10.1175/JHM-D-13-0173.1.
- Vinnikov, K. Ya., and I. B. Yeserkepova, 1991: Soil moisture, empirical data and model results. *J. Climate*, **4**, 66–79, doi:10.1175/1520-0442(1991)004<0066:SMEDAM>2.0.CO;2.
- , A. Robock, N. A. Speranskaya, and C. A. Schlosser, 1996: Scales of temporal and spatial variability of midlatitude soil moisture at different levels. *J. Geophys. Res.*, **101**, 7163–7174, doi:10.1029/95JD02753.
- , —, S. Qiu, and J. K. Entin, 1999: Optimal design of surface networks for observation of soil moisture. *J. Geophys. Res.*, **104**, 19 743–19 749, doi:10.1029/1999JD900060.
- Wei, J., P. A. Dirmeyer, and Z. Guo, 2008: Sensitivities of soil wetness simulation to uncertainties in precipitation and radiation. *Geophys. Res. Lett.*, **35**, L15703, doi:10.1029/2008GL034494.
- , —, and —, 2010: How much do different land models matter for climate simulation? Part II: A temporal decomposition of land–atmosphere coupling strength. *J. Climate*, **23**, 3135–3145, doi:10.1175/2010JCLI3178.1.
- Xia, Y., M. Ek, Y. Wu, T. Ford, and S. Quiring, 2015: Comparison of NLDAS-2 simulated and NASMD observed daily soil moisture. Part I: Comparison and analysis. *J. Hydrometeorol.*, **16**, 1962–1980, doi:10.1175/JHM-D-14-0096.1.
- Zamora, R. J., F. M. Ralph, E. Clark, and T. Schneider, 2011: The NOAA Hydrometeorology Testbed soil moisture observation networks: Design, instrumentation, and preliminary results. *J. Atmos. Oceanic Technol.*, **28**, 1129–1140, doi:10.1175/2010JTECHA1465.1.
- Zreda, M., W. J. Shuttleworth, X. Zeng, C. Zweck, D. Desilets, T. Franz, and R. Rosolem, 2012: COSMOS: The Cosmic-Ray Soil Moisture Observing System. *Hydrol. Earth Syst. Sci.*, **16**, 4079–4099, doi:10.5194/hess-16-4079-2012.