

An assessment of surface soil temperature products from numerical weather prediction models using ground-based measurements

Thomas R. H. Holmes,¹ Thomas J. Jackson,¹ Rolf H. Reichle,² and Jeffrey B. Basara³

Received 10 February 2011; revised 21 November 2011; accepted 12 January 2012; published 25 February 2012.

[1] Surface soil temperature estimates at approximately 0.05 m depth are needed to retrieve soil moisture from the planned Soil Moisture Active Passive (SMAP) L-band (1.4 GHz) satellite. Numerical weather prediction (NWP) systems as operated by various weather centers produce global estimates of soil temperature. In this study in situ data collected over the state of Oklahoma are used to assess surface (soil) temperature from three NWP systems: (1) the integrated forecast system from the European Center for Medium range Weather Forecasts (ECMWF), (2) the modern-era retrospective analysis for research and applications (MERRA) from the NASA Global Modeling and Assimilation Office, and (3) the global data assimilation system used by the National Center for Environmental Prediction (NCEP). The results are presented by hour of day with specific attention directed to the SMAP early morning overpass time at around 6 A.M. local time, and the period of 1 April to 1 October 2009. It was found that the NWP systems estimate the 0.05 m soil temperature at this time of day with an overall root mean square error of 1.9 to 2.0 K. It is shown that this error can be reduced to 1.6 to 1.8 K when differences between the modeling and measurement depth are accounted for by synchronizing each NWP set to match the mean phase of the in situ data and adjusting the amplitude in accordance with heat flow principles. These results indicate that with little calibration all products meet the SMAP error budget criteria over Oklahoma.

Citation: Holmes, T. R. H., T. J. Jackson, R. H. Reichle, and J. B. Basara (2012), An assessment of surface soil temperature products from numerical weather prediction models using ground-based measurements, *Water Resour. Res.*, 48, W02531, doi:10.1029/2011WR010538.

1. Introduction

[2] Numerical weather prediction (NWP) models developed by various weather centers produce estimates of a wide range of land, atmosphere, and ocean variables. Some of these variables are tied directly to the intended applications of the models, while others are intermediate variables that have not been carefully scrutinized for accuracy and reliability. Here we examine a rarely evaluated product, surface soil temperature.

[3] The motivation for this particular investigation is the need for accurate surface soil temperature products to retrieve surface soil moisture from low-frequency passive microwave observations. Of particular concern is the planned National Aeronautics and Space Administration (NASA) satellite called Soil Moisture Active Passive (SMAP) [Entekhabi *et al.*, 2010a], which will require global estimates of surface soil temperature as a dynamic ancillary resource.

[4] The type of soil moisture retrieval algorithm that will be used by SMAP requires the effective temperature of the emitting soil layer, a value related to the physical temperature of all soil layers and weighted by the proximity to the surface and their dielectric properties [Wilheit, 1978]. Until recently, all passive microwave satellite soil moisture products were derived from multifrequency sensors with the lowest frequency either at C-band (~6 GHz) or X-band (~10 GHz) [Jackson *et al.*, 2010; Li *et al.*, 2010; Njoku *et al.*, 2003; Owe *et al.*, 2008]. For these frequencies the soil moisture sensing depth is relatively shallow (~0.01–0.02 m). These same sensor systems have all included a Ka-band (~37 GHz) vertically polarized sensor and researchers had been able to establish good relationships between this channel and the effective temperature for C- and X-band based emission [Holmes *et al.*, 2009]. As a result, there has not been a need for ancillary surface temperature data.

[5] However, the existing multifrequency platforms have limitations in how much soil moisture information they can provide, and as a result a new generation of lower frequency L-band (1.4 GHz) satellites are in various stages of development and implementation. These offer an improved depth of sensing, reduced vegetation attenuation, and in the case of SMAP improved spatial resolution products. The first of these is the European Space Agency Soil Moisture Ocean Salinity (SMOS) satellite [Kerr *et al.*, 2001] launched in 2009, which currently uses modeled soil

¹Hydrology and Remote Sensing Lab, USDA Agricultural Research Service, Beltsville, Maryland 20705, USA.

²Global Modeling and Assimilation Office, NASA GSFC, Greenbelt, Maryland 20771, USA.

³Oklahoma Climatological Survey, University of Oklahoma, Norman, Oklahoma 73072, USA.

temperature from the European Center for Medium range Weather Forecasts (ECMWF) as an input to the soil moisture retrieval scheme. The next L-band satellite is NASA's Aquarius/SAC-D satellite [Le Vine *et al.*, 2007] launched in June 2011. In addition to an L-band radiometer this satellite also includes a Ka-band radiometer from Argentina's National Space Activities Commission (CONAE) which can be used for estimating soil temperature. The third satellite is the NASA's Soil Moisture Active Passive (SMAP) mission (2014 launch), which will have a higher spatial resolution but no onboard source of soil temperature information [Entekhabi *et al.*, 2010a]. For SMAP it will thus be necessary to provide soil temperature as a dynamic ancillary data set. Some requirements for this data set are that it

[6] 1. includes the temperature of the soil at a depth of 0.05 m below the surface,

[7] 2. has a spatial resolution of at least 0.25 deg so that the resolution requirements for soil moisture are not compromised,

[8] 3. closely matches the overpass time of SMAP, at 6 A.M./P.M. local time,

[9] 4. will be available within a few hours of the satellite observation, so that the latency goals for soil moisture retrievals can be achieved, and

[10] 5. meets the error budget of SMAP. In the error budget for achieving the target surface soil moisture accuracy of $0.04 \text{ m}^3 \text{ m}^{-3}$, the SMAP project has assigned an absolute root mean square error of 2 K to the temperature input [O'Neill *et al.*, 2010].

[11] One approach to providing the effective temperature for SMAP is to use analysis or forecast output from global NWP systems that are run operationally or in research mode at weather centers such as ECMWF, the National Center for Environmental Prediction (NCEP), or NASA's Global Modeling and Assimilation Office (NASA/GMAO). This approach would satisfy the requirements for spatio-temporal resolution and latency noted above. The remaining error budget requirement is the focus of this investigation. It should also be noted that in NWP systems that the latency can be traded off against accuracy; recognizing that the accuracy of NWP-based soil temperatures presumably degrades with forecast lead time.

[12] To date, very little analysis has been performed to assess the accuracy of the NWP soil temperature data products. In addition, the relationship between the soil temperature provided by the NWP system and that required for passive microwave radiative transfer modeling and soil moisture retrieval, specifically at L-band, requires further study.

[13] A benefit of using L-band for soil moisture remote sensing is the deeper depth that contributes to the measurement. Theoretical models predict that the thickness of the soil layer that contributes 63% of the soil emission at L-band for an incidence angle of 50 deg varies between 0.015 m for a wet soil to 0.15 m for a dry soil [Ulaby *et al.*, 1986]. Regardless of the actual wetness, this represents a larger part of the root zone than the previous generation of instruments measured. At the same time, this deeper sensing depth means that the originating layer cannot be assumed to have a homogeneous temperature profile and will require the need for the parameterization of the effective temperature to account for the dependence of the

sensing depth on soil moisture. Simple parameterizations exist that are based on the weighting of the temperature of a surface layer, typically at 0.05 m depth, and a deeper soil layer, typically at 0.5 m depth [Choudhury *et al.*, 1982; Holmes *et al.*, 2006; Wigneron *et al.*, 2001].

[14] Related to this discussion of L-band effective depth is the uncertainty regarding the actual depth of the soil layer that the available soil temperature products represent. The daily temperature cycle is determined by the surface energy balance between the net radiation forcing, the latent and sensible heat fluxes into the atmospheric boundary layer, and the ground heat flux into the soil. Although the incoming solar radiation reaches its maximum at solar noon, the net energy input into the soil remains positive for some hours longer resulting in a continued warming of the soil layers. The skin temperature, as measured by thermal infrared sensors, is generally found to reach its maximum at 60 to 90 min after solar noon [Betts and Ball, 1995; Fiebrich *et al.*, 2003]. The further away a specific soil layer is from the surface, the longer the lag between its daily maximum temperature and solar noon. This is also generally true for the air temperature, as it is warmed from the surface, but the near-surface air temperature profile is largely determined by turbulent dynamics.

[15] The length of the time lag between soil temperature at two different depths is determined by the vertical distance between the two depths and the thermal properties of the medium. As shown by Van Wijk and de Vries [1963] this phase shift is accompanied by an exponential reduction in amplitude of the daily temperature cycle with increasing depth. The combined effect of the phase shift and amplitude reduction makes it difficult to compare temperature estimates from different depths directly. As will be shown in this paper, it is possible to synchronize measurements from different sources and minimize the uncertainty related to differences in measurement (or model) depths. This method not only enables the comparison of various temperature products, but may also facilitate the modeling of the effective soil temperature for L-band.

[16] In order to perform a robust assessment of the NWP soil temperature products it would be useful to have high quality and consistent in situ observations for a wide range of climate, vegetation, and soil conditions. This network would also take into consideration scale differences between the NWP products and point samples. There are no data sets that meet all of these criteria. The Oklahoma mesonet [McPherson *et al.*, 2007] is one of the few, if not the only, network that may meet most of these requirements. In this investigation, we use in situ data from the Oklahoma mesonet to assess the near surface soil temperature output from the three NWP systems hosted at ECMWF, NCEP, and GMAO. The analysis is focused on 2009, the latest year for which all data are available to us for the entire growing season. In order to facilitate the statistical analysis of the temperature records, the NWP soil temperature data are synchronized to match the phase of the in situ data while proportionally adjusting the amplitude of its underlying harmonics. This procedure removes a large part of the systematic differences between the data sets and can be applied to any pair of data sets, without ancillary information. The results will be discussed in terms of the requirements for the L-band microwave retrievals.

2. Materials

2.1. Time and Location

[17] This study analyzes soil temperature data at a 15 min resolution for a year of data over the entire state of Oklahoma. Although the analysis is performed for both 2004 and 2009, the focus will be on data from the year 2009, as this is the most recent full year for which data are available and for which the NWP model versions are closest to current specifications. The location was determined by the availability of Oklahoma mesonet data (section 2.2) and also because of the dynamics of temperature and moisture in this region. Oklahoma spans the geographical region of 33–37°N (~400 km), and 94–103°W (~800 km). The climate ranges from subtropical-Mediterranean (Köppen climate classification Cfa) to dryer semiarid, or steppe climate (BSk) toward the western panhandle of the state. Accordingly, the vegetation ranges from short grassland and shrubs in the west, to pastureland and forests in the east (see Figure 1). The annual average temperature in the center of the state is about 16°C, with an average air temperature of 3°C in January and 27°C in July.

2.2. Ground Data

[18] The Oklahoma mesonet [McPherson *et al.*, 2007; Illston *et al.*, 2008] is a statewide network of meteorological stations. At each location the soil temperature is measured with thermistor probes installed horizontally at depths of 0.05, 0.1, and 0.3 m under native sod and at depths of 0.05 and 0.1 m under bare soil. Although the sampling rate is 30 s, only the average over 15 min is reported with an accuracy of 0.5°C for a temperature range of –30 to 55°C. Various automated and manual quality control checks are performed by the Oklahoma mesonet, including a site visit at least three times a year [Shafer *et al.*, 2000]. For this study all data that are not labeled “good” were removed from the analysis. This leaves us with 106 stations for which at least 100 days of soil temperature data were available during 2009. An additional comparison of the 0.05 and

0.1 m temperature records was performed to establish that there was consistency between these depths. Stations with clear discontinuities that can be attributed to a change in sensor depth after reinstallation during the site visit were identified and only the measurements after such an event are used in this study. In all, 5 of the 106 stations were removed from the analysis completely.

[19] The purpose of this paper is to evaluate how well the 0.05 m soil temperature can be estimated from NWP products. Of the two types of mesonet soil temperature data, under native sod or bare soil, the measurements under the vegetated plot were expected to be a better approximation of the mean soil temperature at 0.05 m for the grid cell since most of the land surface is covered by vegetation. For this reason only the two shallowest measurements under native sod will be used, and are referred to as T_5 and T_{10} with the subscript indicating the nominal measurement depth in centimeters. The 101 stations that were available for this study cover 63 half degree grid boxes within 33–37°N, 94–103°W, and their locations are indicated in Figure 1. For each of these 63 grid boxes, one station was selected to represent the in situ data for that box (detailed in section 3.3). For the 27 grid boxes that contain more than one station a second station is selected to replicate the experiment and validate the results. The remaining 11 of 101 stations are included in the analysis of the in situ observations (section 3.3). The names of all stations used in this study are listed in Table 1. The exact location and specific information for each site can be found on <http://www.mesonet.org/>.

2.3. Numerical Weather Prediction Products

[20] Three NWP products, ECMWF, NCEP, and MERRA, are evaluated in this study. The NCEP and ECMWF outputs represent operational near real time (NRT) products, whereas MERRA is a reanalysis product generated with the system that was operational at GMAO during 2010. All three centers are able to provide NRT products very similar to the ones evaluated in this paper.

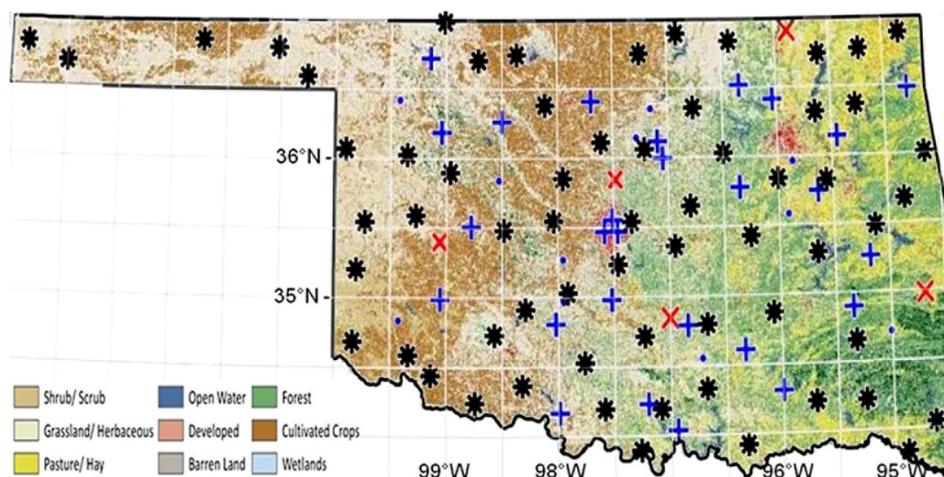


Figure 1. Mesonet station locations indicated on top of a land cover classification for Oklahoma (from the 2001 national land cover database for the United States: NLCD 2001, www.mrlc.gov [Homer *et al.*, 2007]). Mesonet station locations of the Primary set are indicated by an *, the Replication set with a blue +, and the stations that failed our quality control are indicated with a red x. Other stations that are only used in Figure 3 are indicated with blue dots.

Table 1. List of 106 Mesonet Stations With at Least 100 Days of Soil Temperature Data for the Year 2009^a

Primary set (63)	Ada, Altus, Alva, Antlers, Apache, Ardmore, Arnett, Beaver, Blackwell, Boise City, Broken Bow, Burneyville, Butler, Camargo, Chandler, Cherokee, Cheyenne, Chickasha, Clayton, Cloudy, Claremore, Cookson, Durant, El Reno, Erick, Eufaula, Foraker, Grandfield, Hectorville, Hinton, Hollis, Hooker, Idabel, Kenton, Ketchum Ranch, Kingfisher, Lahoma, Marena, May Ranch, Medicine Park, Miami, Marshall, Newkirk, Nowata, Norman, Oilton, Okemah, Pauls Valley, Pawnee, Porter, Pryor, Putnam, Ringling, Shawnee, Slapout, Spencer, Stuart, Tipton, Tishomingo, Vinita, Walters, Webbers Falls, Westville
Replication set (27)	Acme, Breckinridge, Bristow, Centrahoma, Fairview, Freedom, Haskell, Hobart, Inola, Jay, Lane, Madill, Newport, Oklahoma City East, Oklahoma City North, Oklahoma City West, Perkins, Seiling, Skiatook, Tigler, Stillwater, Vanoss, Washington, Waurika, Weatherford, Wilburton, Wynona
Rest set (11)	Bixby, Lake Carl Blackwell, Fittstown, Mangum, Minco, Ninnekah, Talihina, Watonga, Woodward, Okmulgee, Red Rock
Removed by our QC (5)	Bessie, Byars, Copan, Guthrie, Wister

^aStations with insufficient data are not listed. The stations of the primary set are used in the main analysis. The replication set is used for the replication exercise. The rest set is only used in Figure 3.

These products are currently considered for use by the SMAP mission. The general features of the three NWP products are summarized in Table 2. Sections 2.3.1 through 2.3.3 provide additional detail.

2.3.1. ECMWF Analysis and Forecasting System

[21] The ECMWF analysis and forecasting system, called the integrated forecasting system (IFS), includes a comprehensive earth-system model, the deterministic atmospheric (DA) model. The spatial resolution of the DA improved over time; for 2004, IFS version CY25R1 has an average distance between grid points of 39 km, from 2006 onwards this was improved to 25 km and in January 2010 to 16 km. The routine global atmospheric analyses are produced at the synoptic hours 0 and 12 h UTC, and output is provided at 6-hourly intervals. More details about ongoing resolution improvements can be found at: http://www.ecmwf.int/products/data/operational_system/.

[22] The surface processes are described by TESSEL, the tiled ECMWF scheme for surface exchanges over land. In 2007 this land surface model was changed to improve the description of hydrological processes with HTESSEL [Balsamo *et al.*, 2009]. The tiles are based on static land cover information. The skin temperature is defined for each tile, and is in thermal contact with a single four-layer soil profile (or one layer if snow is present). The soil heat budget follows a Fourier diffusion law, modified to take into account the thermal effects of soil water phase changes. The soil temperature represents the layer from 0 to 0.07 m and is analyzed at 00, 06, 12, 18 UTC using screen level (2 m) air temperature increments [Mahfouf *et al.*, 2000].

2.3.2. Modern-Era Retrospective Analysis for Research and Applications

[23] The modern-era retrospective analysis for research and applications (MERRA) is generated by the NASA GMAO

(<http://gmao.gsfc.nasa.gov/research/merra>, [Rienecker *et al.*, 2011]). The MERRA products are generated using Version 5.2.0 of the GEOS-5 DAS [Goddard Earth Observing System (GEOS) Data Assimilation System (DAS)] with the model and analysis each at 0.5 by 0.67 deg resolution in latitude and longitude, respectively, and with a 6-hourly analysis cycle. Two-dimensional diagnostics, describing the radiative and physical properties of the surface, are available as hourly averages. Currently, MERRA data are available from 1979 and are updated through the present with approximately two month latency. MERRA is a frozen system but resembles the GMAO quasi-operational analysis with near real time availability that currently runs at a 0.25 deg resolution.

[24] The surface processes are described by the NASA catchment land surface model [Ducharme *et al.*, 2000; Koster *et al.*, 2000]. Each MERRA grid cell contains several irregularly shaped catchments, called tiles. Within each tile, surface exchange processes and skin temperatures are represented separately for subtiles that are characterized by one of three unique hydrological states: saturated, unsaturated, and wilting. The subtile areal fractions are modeled dynamically based on the total amount of water in the tile. The skin temperature of a grid cell is then obtained by area-weighted averaging of the skin temperatures of all subtiles within the grid cell. The subtile skin temperatures are prognostic variables of the model and represent a bulk surface layer with a small but finite heat capacity. For all vegetation classes except broadleaf evergreen trees (which are not present in Oklahoma), this bulk surface layer represents the vegetation canopy and a skin layer at the top of the soil column (effective layer depth <1 mm). Below the bulk surface layer, a single deeper soil temperature profile for each tile is modeled with a heat diffusion equation using six layers, with layer depths (from top) of 0.10, 0.20, 0.39, 0.76, 1.51, and 10.00 m.

Table 2. Specifications of NWP Products

NWP Center/Model	ECMWF/IFS		GMAO/MERRA		NCEP/GDAS	
Output interval (UTC)	6-hourly (0z/6z/12z/18z)		Hourly average centered on the half h		6-hourly (0z/6z/12z/18z), + 3 h forecasts	
Spatial resolution						
2004	39 km		0.5 × 0.67 deg		0.469 deg (T254)	
2009	25 km		0.5 × 0.67 deg		0.313 deg (T382)	
Regridding to 0.5 deg	Linear average		Bilinear interpolation		Bilinear interpolation	
Parameters	$T(\text{skin})$	$T(0-0.07 \text{ m})$	$T(\text{skin})$	$T(0-0.1 \text{ m})$	$T(\text{skin})$	$T(0-0.1 \text{ m})$
Name after phase synchronization to T_5	$T_{\text{skin}}^{\text{EC}}$	$T_{\text{soil}}^{\text{EC}}$	$T_{\text{skin}}^{\text{ME}}$	$T_{\text{soil}}^{\text{ME}}$	$T_{\text{skin}}^{\text{NC}}$	$T_{\text{soil}}^{\text{NC}}$

In this study we analyzed the gridded skin temperatures and the temperature of the first soil layer (0–0.1 m) over land. This last parameter was only available from a land-only reprocessing of the MERRA data for the year 2009 [Reichle *et al.*, 2011].

2.3.3. NCEP Global Data Assimilation System (GDAS)

[25] GDAS is NCEP's operational forecast system [Ek *et al.*, 2003] (<http://www.emc.ncep.noaa.gov/GFS/>). The land surface model, for the year 2004, was the Oregon State University (OSU) LSM, with 2 soil layers (0–0.1 m and 0.1–2 m). It has an average distance between grid points of 50 km, but this has been increased to 27 km as of 27 July 2010. The LSM was replaced with the NCEP Noah LSM starting in 31 May 2005, and included an increase in soil layers from two to four by dividing the second soil layer in three. Both the skin temperature and the 0–0.1 m soil temperature are analyzed in this study. The 6-hourly analysis is supplemented with 3 h forecasts to provide output at 3-hourly intervals.

2.4. Spatial Scales and Recording Times

[26] The analysis was performed at 0.5 by 0.5 deg resolution after regridding the NWP data to a common regular grid. The regridding method was tailored to the native resolution of each set; the MERRA and NCEP data were regridded by means of a bilinear interpolation; the higher resolution ECMWF data was averaged to the lower target resolution.

[27] In the analysis, a single ground station is selected to represent the “truth” for each grid box. Obviously there is a large-scale discrepancy between the spatial resolution of the NWP data and the single point observation. The effect of this difference should be mitigated in the overall analysis by the size of the in situ network that allows for 63 separate grid boxes to be evaluated, which averages out some of the potential bias. In addition, the analysis is repeated for a second group of stations for the 27 grid cells that are sampled by more than one station.

[28] The times associated with the NWP outputs are defined in coordinated universal time (UTC). The more sparsely sampled NWP soil temperature series are interpolated to the 15 min interval of the in situ data by means of a piecewise cubic spline interpolation. As shown by Aires *et al.* [2004], this interpolation method preserves the amplitude and the time of maximum and minimum temperatures of the underlying cycle. The observation times of the mesonet data are recorded in central standard time (CST), which is 6 h behind UTC for Oklahoma and has its meridian at 90°W. It is very important in this study that we assure that all observations are aligned relative to the position of the sun, to facilitate the assessment at the overpass times of the unsynchronous satellite SMAP. To illustrate, the morning overpass of SMAP will occur at about 6:23 A.M. *local solar time* (LST) at the latitude of Oklahoma (for a descending equator overpass at 6 A.M. [Johnson *et al.*, 1994]). However, in CST this will be at 6:43 A.M. in eastern Oklahoma, and at 7:13 A.M. at the end of the western panhandle. For this reason, all time stamps are converted to LST by adding the time correction factor, $longitude/360$, to the time in UTC (for the time in decimal days). For practical reasons, the small deviations throughout the year (of up to

15 min) that are caused by the eccentricity of the Earth's orbit and the Earth's axial tilt are not corrected for.

3. Methods

3.1. Performance Metrics

[29] The soil temperature products were systematically compared to the in situ data for different periods of the year and for different hours of the day. Two bias independent performance metrics are used; the Pearson's correlation (ρ) and the standard error of estimate (SEE):

$$SEE = \sigma_{\text{true}} \sqrt{1 - \rho^2}, \quad (1)$$

where σ_{true} is the (time series) standard deviation of soil temperature. Note that we are using the standard deviation of the in situ data to approximate σ_{true} instead of the standard deviation of the model data (σ_{est}), which is not the traditional approach. Assuming that the standard deviation of the in situ data is an unbiased estimate of σ_{true} , the SEE represents only the random error and indicates the lowest attainable error level in an application framework when all systematic differences are removed. In practice, the systematic differences will not always be known and two metrics that quantify the absolute error will be more relevant: the root mean square error (RMSE) and the unbiased RMSE (ubRMSE). These well known metrics can be expressed in terms of σ_{true} , σ_{est} , ρ , and mean bias (b) [Entekhabi *et al.*, 2010b]. For the discussion in the present paper it is useful to rewrite those expressions in terms of SEE:

$$\text{ubRMSE} = \sqrt{SEE^2 + (\sigma_{\text{true}}\rho - \sigma_{\text{est}})^2}, \quad (2)$$

$$\text{RMSE} = \sqrt{\text{ubRMSE}^2 + b^2}. \quad (3)$$

[30] The relationships between these three performance metrics (equations (1) to (3)) show that $SEE \leq \text{ubRMSE} \leq \text{RMSE}$ and may serve to quantify a decreasing level of bias removal (in the same units as the variable being assessed).

[31] Equation (2) shows that the lowest ubRMSE is obtained when $\sigma_{\text{est}} = \sigma_{\text{true}}\rho$. Considering that in a real application ρ will always be below unity, this implies that the lowest ubRMSE (and RMSE) is obtained when the estimated variability is lower than the true variability. This feature of the RMSE metrics is discussed in more detail by Gupta *et al.* [2009]. Because the ρ values in temperature comparisons are generally high (between 0.9 and 0.95), this effect will be limited to favoring an underestimation of σ_{est} by 5% to 10%.

[32] The performance metrics are affected by the accuracy of the in situ data, and how representative the sites are for their 0.5×0.5 deg pixel. For example, an unrepresentative ground site can result in both a constant systematic bias (offset) and a proportional error (slope), both of which will directly affect the RMSE.

3.2. Soil Temperature Dynamics

[33] The diurnal and seasonal cycles of heating of the land surface result in distinct periodic temperature variations that propagate downward below the surface. Assuming

only conductive heat transfer and a long-term average temperature that is constant with depth, the propagation of the temperature waves to deeper layers can be described by an exponential decrease in amplitude (A) and an increase in phase shift ($d\varphi$) (e.g., *Van Wijk and de Vries* [1963]). Both modulations are parameterized as a function of vertical distance (dz) and damping depth (z_D):

$$d\varphi = \frac{-dz}{z_D}, \quad (4)$$

$$A_{z_2} = A_{z_1} \exp\left(\frac{dz}{z_D}\right), \quad (5)$$

where dz ($dz = z_2 - z_1$) is positive in the upward direction. The damping depth is an expression of the thermal properties of the medium, in particular the thermal diffusivity (a , $\text{m}^2 \text{s}^{-1}$), and indicates the distance (z_D) over which the amplitude of the wave is reduced by 63%:

$$z_D = \sqrt{\frac{2a}{2\pi f}}, \quad (6)$$

where f (1 s^{-1}) is the frequency of the temperature wave.

[34] The thermal properties in a soil are mainly determined by soil moisture content and soil type. To assess the impact of these factors on the variability and size of the phase shift, the thermal diffusivity was calculated for a range of soil types from sand to silt and clay according to *Peters-Lidard et al.* [1998]. The damping depth for a harmonic with a period of a day is calculated according to equation (6), and the associated phase shift over 0.05 m vertical distance then follows from equation (4). The simulation results are displayed in Figure 2, and the results relating to the main soil types as found in Oklahoma are indicated in the legend. Based on these simulations the phase shift is nearly constant at soil moisture levels above

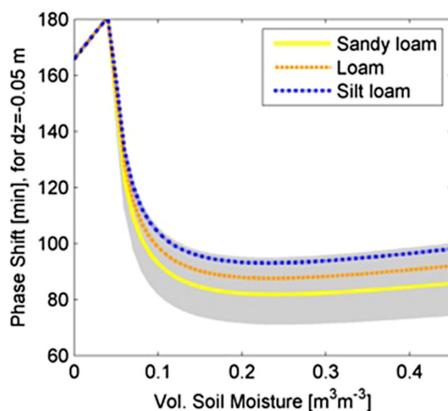


Figure 2. Theoretical phase shift of the temperature harmonic with a period of a day over a 0.05 m vertical distance. The gray area indicates the range of values for a porosity of $0.45 \text{ m}^3 \text{ m}^{-3}$, and quartz contents from 0 to 100%. The three main soil types as found in Oklahoma are indicated as well: Sandy loam (62% sand, 10% clay), loam (43% sand, 15% clay), and silt loam (26% sand, 13% clay).

$0.10 \text{ m}^3 \text{ m}^{-3}$, with values ranging between 70 and 100 min depending on soil type. The increase in thermal conductivity with moisture content is matched by an increase in heat capacity, making the thermal diffusivity insensitive to soil moisture in this range. At low moisture levels on the other hand, the diffusivity is so low that the heat dissipates only slowly in the soil and large vertical gradients are sustained, resulting in the large phase shifts of up to 180 min in the dry range of Figure 2. Over the year and between different localities the propagation of temperature harmonics into the soil may thus be described by a single set of equations that is only weakly affected by variations in soil moisture, if the soil is not very dry.

3.3. Phase Synchronization and Amplitude Adjustment

[35] Because soil temperature harmonics change with depth, soil temperature records from different depths cannot be compared directly. Even a slight vertical misalignment will result in an artificial increase in the error as calculated between the two records. The calculated error will then not only depend on the accuracy of the assessed records, but also on the time of day and represented soil depths. To better compare two temperature records, we can apply the heat flow principles as described in section 3.2 to remove the phase difference and reduce the bias in amplitude. Following equation (4), the relative distance (the vertical distance between input and target depth, divided by the damping depth) can be replaced by $d\varphi$, the phase shift between the mean daily temperature harmonic of two temperature series. This phase shift captures in a single number the integrated effect of the soil thermal properties and can be calculated between any two time series of soil temperature (see below). The temperature at the target depth is then estimated by applying both the phase shift and the exponential amplitude decay to the underlying harmonics of the original temperature record. Decomposing the temperature signal in the underlying harmonics is done in a way similar to the classic Fourier analysis as described by *Van Wijk and De Vries* [1963]. The exact approach used here is described in Appendix A.

[36] The phase synchronization method was first tested on the in situ data as follows. We estimate the temperature at the depth of the shallow record (T_5) from the deeper sensor measurements (T_{10}). This new estimate, labeled T_{5-10} , is based on the observed mean phase difference $d\varphi$ between T_5 and T_{10} :

$$T_{5-10} = f(T_{10}, d\varphi). \quad (7)$$

[37] The phase of each record is determined by optimizing φ so that the RMSE is minimized between the mean diurnal cycle and the sine function T_{sim} :

$$T_{\text{sim}} = \bar{T} + A \sin[(t - \varphi)2\pi - \pi/2], \quad (8)$$

where \bar{T} and A are the mean and amplitude of the diurnal cycle. A value of $\varphi = 0$ would have a maximum at local solar noon. Figure 3 shows the fitted values of φ for T_5 in the top panel. There is poor phase coherence with a range of φ from 3 to 5 h. The standard deviation is 30 min for a mean phase shift of 4 h and 10 min. If all sensors were installed

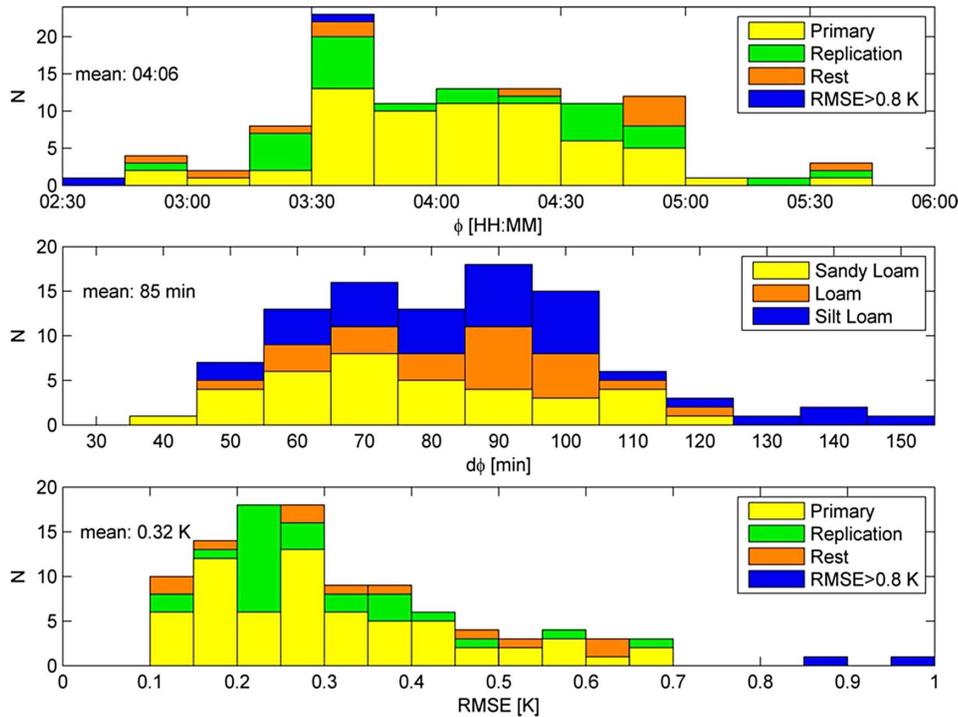


Figure 3. (Top) Phase shift from local solar noon of the daily harmonic of in situ soil temperature (T_5), (middle), phase difference between the T_5 and T_{10} in situ soil temperatures, and (bottom) RMSE for the estimated in situ soil temperature at 0.05 m (T_{5-10}). In the top and bottom graph stations that are selected as the primary set are indicated in yellow, the replication set in green, the rest set in orange, and the stations with $\text{RMSE} > 0.9$ in blue. In the middle graph the stations are grouped by soil texture class.

at the same depth, a narrower distribution of φ would be expected. But installing and maintaining the temperature sensors at a constant shallow depth is difficult. The topsoil can be affected by rainfall, freeze-thaw heaving, vegetation, and animal activity, which can all lead to erosion or sedimentation of several centimeters. Therefore, we cannot assume that T_5 represents exactly the 0.05 m soil depth. Furthermore, there may be differences in vegetation density between the measurement sites that can cause an apparent damping of the temperature harmonics. Both effects may explain the poor phase coherence of the temperature records as found for T_5 (Figure 3, top panel).

[38] If the sensors for T_5 and T_{10} are installed vertically above each other, disturbances at the surface should not affect the distance between them. However, settling of the sensors after installation may still affect the distance between the sensors. The phase difference between the 0.05 and 0.1 m (nominal) in situ temperature records is shown in Figure 3 (middle panel). The mean $d\varphi$ of 85 min fits comfortably within the theoretical range as given in section 3.2 (Figure 2), and the bulk of the stations have a phase shift that is within the expected range for wet soils and a vertical difference of 0.05 m (70 to 100 min).

[39] When we group the stations by soil type into three groups the differences in mean phase again confirm the predictions of the theoretical model. The sandy loam group has the lowest average $d\varphi$ of 78 min. The phase shift increases for the loam group ($d\varphi = 86$ min) and the silt loam group ($d\varphi = 90$ min). It is clear, however, that the soil texture differences alone cannot explain the wide range

of phase differences as measured at the stations. Therefore, the variation in phase shift between the stations is largely attributed to deviations in the actual vertical distance between the sensors from the nominal 0.05 m.

[40] Using these values for $d\varphi$ we estimate the temperature at the depth of T_5 from T_{10} for each site (T_{5-10}). The RMSE as calculated between T_{5-10} and T_5 for each site is generally well below the stated accuracy for the probes (0.5 K), see the lower panel of Figure 3. Some stations have a higher RMSE, which we attribute to a change in the relative depth of the sensors over the year. To minimize possible errors in the in situ measurements, two stations (Bessie and Guthrie) with an overall RMSE of more than 0.8 K were discarded from further analysis, in addition to three other stations that were removed outright because of large discontinuities in the data set (see Table 1).

[41] Figure 4 shows the daily cycle of the aggregated performance metrics for the selected stations. The top panel of Figure 4 illustrates the effect of the temperature phase synchronization. The original temperature record (T_{10} : red dash), as measured at a nominal depth of 0.1 m, has a smaller amplitude for the daily temperature cycle than the in situ measurement at 0.05 m (T_5 : black x), and the daily maximum occurs 85 min later. The mean diurnal cycle of the phase synchronized record (T_{5-10} : blue line) is virtually indistinguishable from that of the actual in situ measurements at 0.05 m. As a result, all performance metrics improve dramatically from the case when no synchronization is applied (middle panel), to the phase synchronized results (lower panel). The RMSE of T_{10} (versus T_5 ; middle

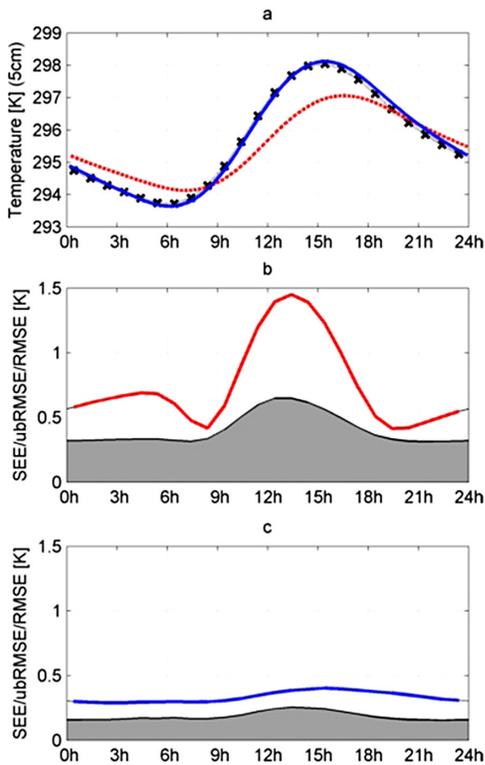


Figure 4. Results of phase synchronization method by hour of day. (a) Mean diurnal cycle of in situ measurements (red: T_{10} ; black x: T_5) and the estimated T_{5-10} (blue). (b) SEE (top of gray shading), ubRMSE (top of black shading), and RMSE (solid line) when T_{10} is compared with T_5 ; (c) same as but for T_{5-10} versus T_5 . Note that the stated accuracy of the sensors is 0.5 K. Note that in (b) and (c) ubRMSE is only marginally greater than SEE.

panel, red line) has minima of ~ 0.4 K at 8 A.M./P.M. but spikes to 1.6 K at other times of the day. Applying the phase synchronization reduces this error throughout the day, as is evident from the RMSE of T_{5-10} (versus T_5) which ranges between 0.3 and 0.4 K (lower panel, blue line).

[42] The phase synchronization method uses only the $d\varphi$ determined from the mean diurnal temperature cycle as measured over the growing season; it implicitly assumes constant soil thermal properties throughout the year. Despite this simplification, the RMSE of T_{5-10} (versus T_5) is well below the stated accuracy of the probes (0.5 K) throughout the day. Therefore the phase synchronization method will be used in the following analysis to match the phase of each NWP product with that of the in situ data.

3.4. Implementation of Phase Synchronization for NWP Soil Temperature Products

[43] The uncertainty about the actual depth of the in situ sensors, as shown in section 3.3, complicates a robust assessment of NWP temperature products. In order to maximize the phase coherence of the validation target, the station with the phase of T_5 closest to the average of all stations was selected if more than one station was available for a particular grid box.

[44] The mean phase difference between the T_5 in situ data and the NWP output is then determined. The NWP data set is corrected for this mean phase shift to create a data set for which the phase of the mean daily cycle is in line with that of T_5 . This operation does not affect the temporal and spatial differences in phase of the NWP data. The assessment of the NWP temperature will be based on both the original and these synchronized data sets. The synchronized sets should be regarded as the current best estimate of the 0.05 m temperature to be obtained from these NWP models.

[45] As a test on how well the phase synchronization can reproduce the NWP model physics we modeled the NWP soil temperature from the NWP skin temperature using the phase shift and amplitude adjustment approach. For example, we estimated the MERRA 0–0.1 m soil temperature from the MERRA skin temperature. When compared to the actual MERRA 0–0.1 m temperatures, the resulting estimates have an RMSE of between 0.2 and 0.4 K and an R^2 above 0.997 throughout the day showing that the phase synchronization reproduces the MERRA soil physics with great accuracy, even though the amplitude is underestimated by 10%.

[46] The same procedure for NCEP and ECMWF yields higher RMSE values of 0.6 at night to 1.2 K during the day. The lower accuracy of estimating the NCEP and ECMWF soil temperatures can be explained by the lower output interval (3 and 6 h, respectively) as opposed to the hourly output from MERRA. We confirmed this explanation by subsampling MERRA at 6-hourly intervals and found that the characterization of the higher frequency dynamics through the cubic spline interpolation becomes much poorer, which in turn complicates the phase synchronization and increases the RMSE. For applications such as SMAP, all NWP centers should be able to provide output at a custom sampling frequency. However, higher sampling was not available for all products in this paper. We therefore approximate the higher sampling by averaging the skin and soil temperatures before performing the interpolation. Because the phase difference between skin and soil temperatures is 2 to 4 h, this averaging effectively doubles the sampling frequency of the temperature series. For this reason we include the average of the skin and soil temperatures from each model as an additional data product in the following analysis (where the average is computed prior to phase synchronization). As will be shown below, this average temperature provides the best performance.

4. Intercomparison of NWP Products

[47] Surface (soil) temperature products from different NWP models that are most relevant to the SMAP mission were compared. These are the (1) ECMWF's operational integrated forecast system, (2) GMAO's MERRA, and (3) the global data assimilation system (GDAS) as used by NCEP (section 2.3). Table 2 is an overview of the general specifications of these products. All products are resampled to a 0.5 by 0.5 deg box (see section 2.4) and synchronized to match the phase of the in situ data as described in section 3.4, and hereinafter referred to as T^{EC} , T^{ME} , and T^{NC} to indicate the NWP model, and with either the subscript skin to indicate the skin temperature, soil for the temperature of

the models first soil layer, and avg for the average of the skin and soil temperatures (section 3.4).

4.1. General Characteristics of the NWP Temperature Products

[48] As a first step these products are compared to the in situ data in terms of mean, amplitude, and phase of the average diurnal cycle for April through October 2009. These results give an indication of the characterization of the daily temperature cycle by each model, and are listed in Table 3.

[49] The theoretical considerations on the relationship between the phase and the depth of temperature measurements, as detailed in section 3.2, may help us to interpret the phase difference with in situ data of the NWP temperature products. Based on theory and the empirical evidence of the in situ data (section 3.3) it was found that a phase difference of 85 min corresponds to a 0.05 m vertical distance in soil temperature measurements. It was further found that the in situ data have, on average, a peak temperature at 4 h and 6 min after solar noon (section 3.3). Assuming that the mean sensor depth is equal to the nominal depth of 0.05 m we can then interpret the phase difference between the NWP temperature products and the in situ data T_5 .

[50] 1. The phase difference with T_5 for all NWP skin temperatures (Table 3) exceeds the 85 min that is expected for just the 0.05 m soil layer, which may be expected for a skin temperature corresponding to the top of a vegetation layer (with its own heat capacity). MERRA's T_{skin} has the largest phase difference with T_5 , it peaks just an hour after noon. The T_{skin} of ECMWF and NCEP are 1 h 45 min and 2 h after solar noon, respectively.

[51] 2. The phase of the soil temperature product shows a greater variety between the three models. The phase of MERRA's T_{soil} is almost exactly the same as that of T_5 and therefore corresponds to the center of the soil layer (0–0.10 m). Likewise; the 43 min phase difference for ECMWF's T_{soil} agrees with the center of that models first

soil layer (0–0.07 m). The phase of NCEP's T_{soil} , more than 2 h after the in situ data, would correspond to a depth of 0.14 m which is well below the NCEP topsoil temperature layer (0–0.1 m).

[52] 3. Averaging the skin and soil temperatures results in phase differences of between 86 and 115 min. This is a smaller variation than recorded for the individual products and corresponds to a soil depth at the top of the soil (0 m).

[53] From this analysis it is clear that the NWP soil temperature data sets relate to quite different soil (or vegetation) model layers. Besides the misalignment in phase these differences in model depth result in large biases in amplitude of the diurnal temperature variation. By applying the phase synchronization the effect of these depth-related differences can be reduced significantly. For example, for MERRA's T_{avg} the bias in amplitude is reduced from 2.7 times to 1.5 times the amplitude of the T_5 in situ data.

[54] Judging by bias in mean and in amplitude, the mean daily cycle over the period is best described by the synchronized T_{avg} for each model. Particularly $T_{\text{avg}}^{\text{NC}}$ has only a small mean bias and just a 14% overestimation of the mean daily amplitude. In contrast, $T_{\text{avg}}^{\text{ME}}$ and $T_{\text{avg}}^{\text{EC}}$ still have a 50% to 70% overestimation of the daily amplitude after the synchronization.

[55] As an input to soil moisture retrieval algorithms based on sunsynchronous satellites, the accuracy of the temperature estimate at a specific time of day is of importance. For this reason the performance metrics are calculated by time of day and are presented in section 4.2. Section 4.3 will then analyze the results at the time of the morning overpass of SMAP (~6.23 A.M. LST over Oklahoma). The robustness of these results will be tested by analyzing several replication experiments in section 4.4.

4.2. Diurnal Variation in Performance Metrics

[56] The diurnal variation in performance metrics is studied in more detail for the T_{avg} of each model since these performed better than the individual T_{skin} and T_{soil} . Figure 5 shows for each NWP model the mean daily cycle of the phase-synchronized T_{avg} (blue lines) in comparison to the in situ data (black lines). Also shown are the unsynchronized NWP products (red dashed lines); all of which overestimate the amplitude of the daily temperature cycle. This overestimation is only partly removed by the phase synchronization. But more importantly, the bias in the early morning is reduced to less than 1 K.

[57] Figure 5 also displays the diurnal variation in the performance metrics as discussed in section 3.1. The second row of the figure displays the coefficient of determination for both the original (red) and the synchronized products (blue). The size of the improvement in correlation coefficient is directly related to the phase difference between the original product and in situ data. After the phase synchronization the three products all have a high correlation of between $R^2 = 0.91$ – 0.94 in the night and early morning.

[58] For the synchronized products the three performance metrics (SEE, ubRMSE, and RMSE) are displayed in the third row of Figure 5. $T_{\text{avg}}^{\text{EC}}$ and $T_{\text{avg}}^{\text{ME}}$ have an SEE of around 1.3 K during the night through 9 A.M. in the morning. During the day the SEE reaches values of 2 K.

Table 3. Characterization of the Mean Daily Temperature Cycle by NWP Temperature Estimates Versus In Situ Measurements for 1 April to 1 October 2009

Phase Synchronization	T_{skin}		T_{avg}		T_{soil}	
	No	Yes	No	Yes	No	Yes
<i>ECMWF</i>						
$d\phi$ (minutes relative to in situ)	-138	-12	-104	-9	-43	-3
Mean bias (K)	2.0	1.5	2.0	1.7	2.0	1.9
Amplitude bias $A^{\text{NWP}}/A^{\text{InSitu}}$ (K K ⁻¹)	3.9	1.9	2.9	1.7	2.1	1.7
<i>MERRA</i>						
$d\phi$ (minutes relative to in situ)	-169	-14	-115	-10	-1	2
Mean bias (K)	1.5	1.0	1.2	0.9	1.0	1.0
Amplitude bias $A^{\text{NWP}}/A^{\text{InSitu}}$ (K K ⁻¹)	3.8	1.6	2.7	1.5	1.9	1.8
<i>NCEP</i>						
$d\phi$ (minutes relative to in situ)	-128	-12	-86	-9	148	9
Mean bias (K)	1.0	0.6	0.5	0.2	0.0	0.4
Amplitude bias $A^{\text{NWP}}/A^{\text{InSitu}}$ (K K ⁻¹)	3.2	1.7	1.8	1.1	0.7	1.3

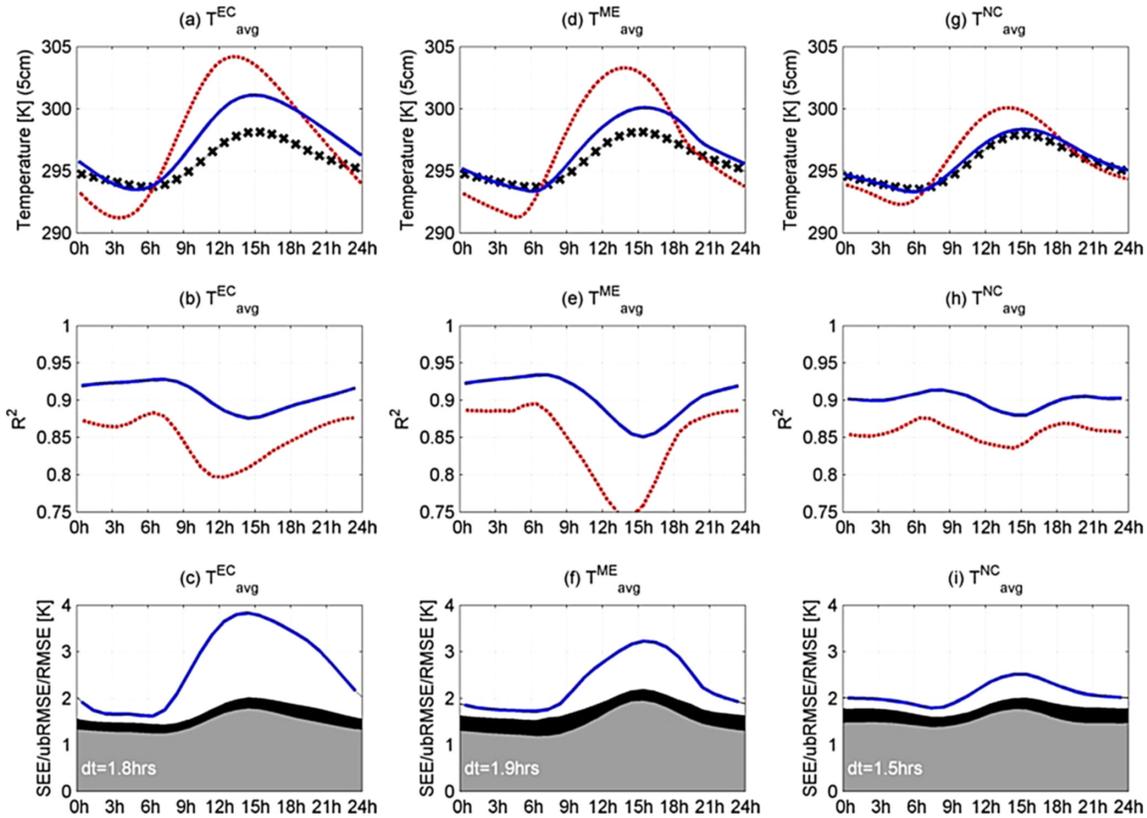


Figure 5. Daily cycle and performance metrics for T_{avg} from (a,b,c) ECMWF, (d,e,f) MERRA, and (g,h,i) NCEP, compared to in situ data at 0.05 m under sod soil, for April through September of 2009. (a,d,g) the mean daily temperature cycle for in situ data (black x), original NWP temperature (red dash) and synchronized NWP product (blue). (b,e,h) correlation coefficient for original (red dash) and synchronized product (blue). (c,f,i) SEE (top of gray shading), ubRMSE (top of black shading), and RMSE (blue line).

Although $T_{\text{avg}}^{\text{NC}}$ does not reach as low a value of SEE, its performance is slightly more stable over the day.

[59] As discussed in section 3.1, the RMSE is a combination of SEE, bias, and bias in standard deviation and is shown in the third row of Figure 5 (blue lines). The RMSE values calculated for $T_{\text{avg}}^{\text{EC}}$ and $T_{\text{avg}}^{\text{ME}}$ are very similar, varying from a low of ~ 1.6 K in the morning to a high between 3 and 4 K in the afternoon. The accuracy of $T_{\text{avg}}^{\text{NC}}$ is more stable over the course of the day, varying only from 1.8 to 2.5 K, which can be attributed to this product being able to accurately estimate the daily amplitude in temperature.

[60] In summary, the synchronized products all have a high correlation with the in situ measurement throughout the day. Even though $T_{\text{avg}}^{\text{EC}}$ and $T_{\text{avg}}^{\text{ME}}$ overestimate the daily amplitude significantly, their early morning errors are lower than $T_{\text{avg}}^{\text{NC}}$. On the other hand, the accurate estimate of the diurnal cycle makes $T_{\text{avg}}^{\text{NC}}$ the best choice during the day.

4.3. Results for 6 A.M. Local Time

[61] For the narrow time window of concern to the SMAP mission (6:23 A.M. LST for Oklahoma) the results are now examined in greater detail. The metrics at this time of day, aggregated for the region as a whole, are listed in

Table 4. The standard error for $T_{\text{avg}}^{\text{EC}}$ is 1.3 K, with almost no bias. The RMSE is 1.6 K, with only 5 of the 63 grid cells having an error above the target accuracy of 2 K. The results for $T_{\text{avg}}^{\text{ME}}$ are very similar, trailing in RMSE by a narrow margin of 0.1 K and with nine cells with an error above 2 K. Table 4 also shows that without removing the phase difference and the corresponding amplitude adjustments, the RMSE would be significantly higher (2.1 K for $T_{\text{avg}}^{\text{EC}}$ and 2.4 K for $T_{\text{avg}}^{\text{ME}}$).

[62] Looking at the individual products of ECMWF we see that the RMSE for both the $T_{\text{skin}}^{\text{EC}}$ and $T_{\text{soil}}^{\text{EC}}$ is only a little higher than for the averaged product, but that the number of grid cells with errors above 2 K increases from 5 to 13 and 12, respectively. For MERRA the same analysis shows that even though the $T_{\text{soil}}^{\text{ME}}$ did not require phase synchronization, its results are poorer than for the $T_{\text{skin}}^{\text{ME}}$ that needed a phase adjustment of almost 3 h (RMSE of 2.0 K versus 1.8 K).

[63] The RMSE for $T_{\text{avg}}^{\text{NC}}$ at 6 A.M. is 1.8 K, with 12 of the 63 grid cells with an error above 2 K. This shows that the accurate description of the mean diurnal cycle of the in situ data by $T_{\text{avg}}^{\text{NC}}$ does not result in the lowest error at 6 A.M. For $T_{\text{skin}}^{\text{NC}}$ the effect of the phase synchronization is a

Table 4. Performance Metrics for NWP Temperature Estimates Versus In Situ Measurements for 1 April to 1 October 2009, Considering Only the Data at the Time of the Morning Overpass for SMAP

Phase Synchronization	T_{skin}		T_{avg}		T_{soil}	
	No	Yes	No	Yes	No	Yes
<i>ECMWF</i>						
Bias (K)	-0.7	-0.4	-0.1	0.1	0.4	0.6
R^2	0.85	0.92	0.88	0.93	0.90	0.92
SEE (K)	1.8	1.3	1.6	1.3	1.4	1.3
ubRMSE (K)	2.4	1.6	1.9	1.4	1.6	1.4
RMSE (K)	2.6	1.8	2.1	1.6	1.9	1.7
$N(\text{RMSE} > 2 \text{ K})/N(\text{all})$	62/63	13/63	38/63	5/63	14/63	12/63
<i>MERRA</i>						
Bias (K)	-0.7	-0.4	-0.8	-0.3	-0.8	-0.7
R^2	0.85	0.93	0.9	0.93	0.92	0.92
SEE (K)	1.8	1.2	1.5	1.2	1.3	1.3
ubRMSE (K)	2.8	1.5	2.2	1.5	1.7	1.7
RMSE (K)	3.0	1.8	2.4	1.7	2.0	2.0
$N(\text{RMSE} > 2 \text{ K})/N(\text{all})$	63/63	10/63	57/63	9/63	34/63	31/63
<i>NCEP</i>						
Bias (K)	-1.6	-0.8	-0.6	-0.2	0.4	-0.3
R^2	0.81	0.88	0.88	0.91	0.90	0.84
SEE (K)	2.1	1.6	1.7	1.4	1.5	1.9
ubRMSE (K)	2.6	1.9	2.0	1.6	1.7	2.3
RMSE (K)	3.2	2.2	2.2	1.8	1.9	2.5
$N(\text{RMSE} > 2 \text{ K})/N(\text{all})$	63/63	53/63	48/63	12/63	21/63	53/63

significant reduction in RMSE from 3.2 to 2.2 K. On the other hand, removing a similar but opposite phase difference from $T_{\text{soil}}^{\text{NC}}$ results in an increase in RMSE from 1.9 to 2.5 K. This result is curious because the characterization of the diurnal cycle actually improves, and may be explained by difficulties in recreating higher frequency harmonics from a dampened and poorly sampled signal at a deeper depth.

[64] Even though for the area as a whole an RMSE at 6 A.M. of below 2 K was found for $T_{\text{avg}}^{\text{EC}}$, $T_{\text{avg}}^{\text{ME}}$, and $T_{\text{avg}}^{\text{NC}}$, the percentage of individual grid cells with errors above this threshold amounted to 8%, 14%, and 19%, respectively. In Figure 6 maps of the RMSE at 6 A.M. LST over Oklahoma are presented for the T_{avg} of each model. The distribution of the RMSE over the state is far from homogenous, but no obvious relationship with spatially varying features such as vegetation density, average soil moisture content, or soil texture maps can be identified. The lack of a spatial structure in the error might indicate that it is associated with installation and/or scaling errors associated with the in situ data more than with the model structure itself. The error varies in time as well as in space, but to a slightly lesser degree. Further analysis (not shown) indicates that there is no relationship between temperature error and soil moisture at 6 A.M. When calculating the RMSE for 6 A.M. across Oklahoma, the percentage of days with errors in excess of 2 K is 9% for $T_{\text{avg}}^{\text{EC}}$, 11% for $T_{\text{avg}}^{\text{ME}}$, and 16% for $T_{\text{avg}}^{\text{NC}}$.

[65] Because the bias is relatively small during the night, the possible gain from bias reduction techniques is limited. During the day the bias is more directly related to an overestimation of the daily amplitude and, as a result, better modeling techniques have the potential of reducing the absolute error to values closer to the SEE.

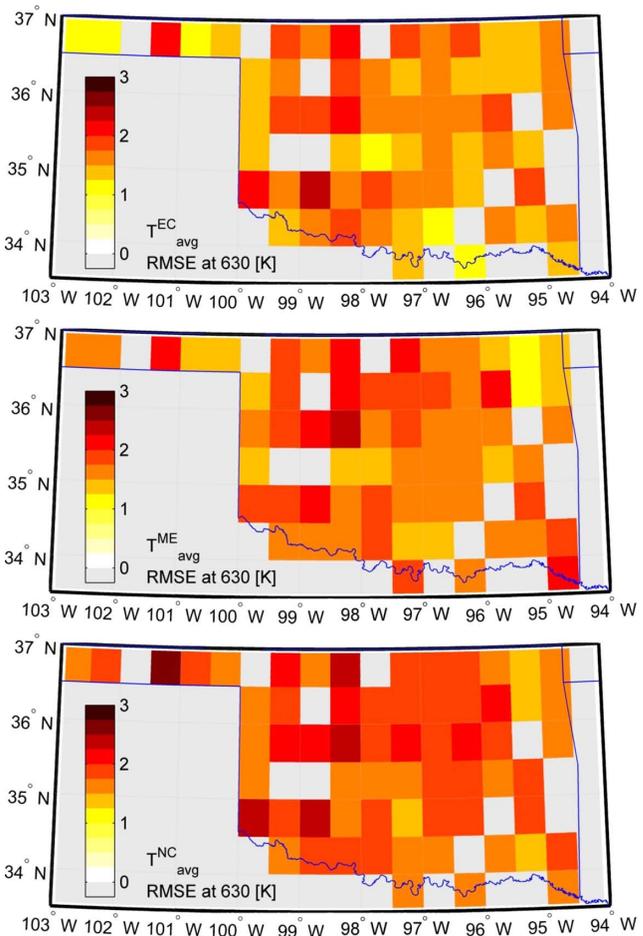


Figure 6. RMSE at 6 A.M. LST for the phase synchronized temperature products from (top) ECMWF, (middle) MERRA, and (bottom) NCEP.

[66] Since the primary location for this analysis is Oklahoma, with a UTC offset of 6 h, the 6 A.M. local time coincides closely with the noon analysis of ECMWF and NCEP. As a result, the penalty resulting from the low temporal resolution of ECMWF and NCEP (both provide analysis output at 6 h time steps) is minimized at this time of day and for this particular location. If available, using the 3 h forecast data to better interpolate the temperature between analyses steps might help improve the overall performance. Even so, a customized output stream with higher temporal resolution might have to be requested if NCEP or ECMWF were to be used for the SMAP algorithms. On the other hand, because the hourly resolution of the MERRA output is already high, its results are likely more independent of longitude.

[67] In summary, for the morning overpass of SMAP, the 0.05 m temperature can be estimated with an RMSE below 2.0 K from either NWP model, but best results were found using the ECMWF data, followed by MERRA. Because the RMSE varies in space and time, even the best temperature product had errors above 2.0 K at 1 out of 10 stations, and on 1 out of 10 days in this analysis. The scope for bias reduction appears limited, but acquiring output at a higher temporal resolution may reduce this error further.

4.4. Replication

[68] To test the robustness of the results, the analysis was repeated for the set of 27 grid cells with duplicate stations. The overall performance metrics are almost identical for the two groups of stations, the RMSE values are all within 0.07 K of the values as shown for the primary set. This gives us confidence that the retrieved results are not significantly influenced by individual stations.

[69] As another replication, we conducted the same analysis for the year 2004. For this year the standard deviation of temperature as recorded by the in situ sensors was slightly smaller. This lower variation may explain the SEE levels that are about 0.2 K lower than in 2009. For ECMWF and MERRA this in turn results in lower RMSE values, but not for NCEP. As mentioned above, ECMWF and NCEP have modified their land surface models between 2004 and 2009 and improved their spatial resolution, whereas MERRA represents a frozen framework. It seems that especially for NCEP the modifications have resulted in improved accuracy of the soil temperature. However, because of the different dynamic range in temperature between 2004 and 2009, it is not possible to isolate the impact of the changes in the NWP systems on the performance of the temperature products.

[70] Finally, the analysis was repeated for the primary set of mesonet stations, but now for the entire year 2009 which includes long periods with frozen soil and possibly snow. At such times no soil moisture retrieval will be possible and, therefore, the quality of the soil temperature estimate will not affect the SMAP error budget. Therefore, the times for which model skin (or soil; or avg) temperatures are below 273.2 K are excluded from the analysis. The performance over the entire year is summarized in Figure 7, showing an increase in error during the night and a decreased error level during the day. The exact metrics for 6 A.M. are listed in Table 5 and show a minor increase in error levels, with an RMSE of 2.1 K for $T_{\text{avg}}^{\text{NC}}$, 2.0 for $T_{\text{avg}}^{\text{ME}}$, and 1.9 K for $T_{\text{avg}}^{\text{EC}}$. Even when considering all data and including temperatures below 273.2 K, the RMSE at 6 A.M. would still only be 2.3 and 2.0 K for $T_{\text{avg}}^{\text{NC}}$ and $T_{\text{avg}}^{\text{EC}}$, respectively. For $T_{\text{avg}}^{\text{ME}}$, however, including these low temperatures results in a larger increase in RMSE to 2.5 K. The temperature of MERRA's 0–0.1 m soil layer shows an identical increase in RMSE when subfreezing temperatures are considered.

[71] All three replications confirm the general level of attainable accuracies for the soil temperature products.

Table 5. Validation Results at 6 A.M. for the Entire Year 2009 (With Model Temperature > 273.2 K)

Model	$T_{\text{avg}}^{\text{EC}}$	$T_{\text{avg}}^{\text{ME}}$	$T_{\text{avg}}^{\text{NC}}$
Bias	-0.2	-0.8	-0.2
R^2	0.95	0.95	0.93
SEE	1.5	1.4	1.8
ubRMSE	1.7	1.8	1.9
RMSE	1.9	2.0	2.1
$N(\text{RMSE} > 2 \text{ K})/N(\text{all})$	17/63	36/63	41/63

When considering a larger temperature range up to the freezing point the error will increase only by a few tenths of a degree, but subfreezing temperatures pose a problem for MERRA.

5. Discussion and Conclusion

[72] In this paper, several NWP model-based soil temperature products were validated using in situ observations from the Oklahoma mesonet. The objective of the study was to provide an error assessment that supports the selection of the product that is best suited for use in satellite-based soil moisture retrieval algorithms. For this reason it is the accuracy of the temperature products at a specific time of the day that is our primary focus (6 A.M. for SMAP), rather than the daily average temperature. This focus makes the validation framework highly dependent on the depth of the in situ sensors and the model levels.

[73] We addressed the uncertainty in the actual depth of the sensors. Variations in the exact placement and subsidence/erosion over time will cause the depth to vary between points and over time. We adjusted for this effect by removing stations with clear and persistent mismatches between the two topsoil temperature sensors. For the primary set that was used in the assessment we selected stations for which the timing of the temperature wave was closest to the mean value of all stations. Furthermore, in order to compare the various NWP products, depth differences are accounted for by synchronizing the phase of each temperature product (and proportionally adjusting its amplitude) according to the mean phase difference measured between in situ data and model data over Oklahoma. An improved representation by the models of the average daily temperature cycle of the in situ data and a clear improvement in precision and absolute error demonstrate the usefulness and validity of this approach.

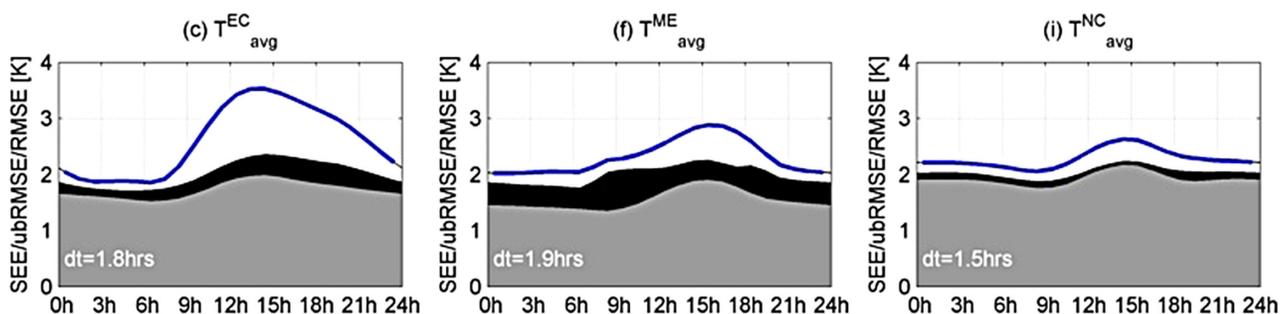


Figure 7. Same as bottom row of Figure 5 but for the full year 2009, excluding model temperatures below 273 K.

[74] By applying these transformations to make the data sets more comparable, it is shown that at 6 A.M. local time, the 0.05 cm temperature can be estimated with an SEE of 1.3 K and an RMSE of 1.6 K using the ECMWF and NASA/GMAO systems. These results reflect the lowest error levels measured during the day and increase to an RMSE of above 3 K from noon to 7 P.M. The temperature from NCEP appears to have a better agreement of the full diurnal cycle which leads to a much more stable performance across the day. Its error at 6 A.M. is only slightly higher with an RMS of 1.8 K. However, this error varies spatially and 19% of the grid cells had errors in excess of 2 K at 6 A.M., compared to only 8% for ECMWF and 14% for MERRA.

[75] These results suggest that the overall accuracy of the best NWP soil temperature products from each center does not exceed the level allotted in the error budget for SMAP (2 K), during the morning overpass of 6 A.M. local time. However, the spatial variation in error suggests that the actual area for which the temperature products perform within the desired accuracy for SMAP will be in the order of 90% of the land surface of Oklahoma for ECMWF and MERRA, and only 80% for NCEP. Some of the error is likely related to how representative the point based validation data are for the 0.5 deg grid box; if a particular site is not representative of the grid cell average conditions, the true error in the NWP soil moisture estimates may well be smaller than our analysis indicates. In addition, more expansive studies of the bias over longer time periods may yield a better mitigation of those errors, as will more optimized temperature modeling methods.

[76] The above results represent the current best estimate of NWP soil temperature performance using the most expansive and consistent set of in situ temperature data currently available. Although the area covered by this study is large and covers some of the main vegetation types (grassland, crop land) of interest for soil moisture remote sensing, one has to be careful in extrapolating the results to other areas of the world. The main reason is that the performance of the NWP models is determined by its input data, the quality of which varies considerably across the globe. Indeed, the NWP products generally perform best in the continental US and western Europe, and if this is also true for soil temperature then the presented results must be regarded as the best case scenario.

[77] A second caveat is that the temporal resolution of the model output may result in longitude dependent interpolation errors. When considering a 6 A.M. LST across the globe the output time of ECMWF and NCEP can be as much as 3 h removed, whereas for MERRA with an hourly output interval the time difference is never more than a half hour. These types of errors happened to be minimal over Oklahoma for all NWP models. A final factor is that the phase synchronization was based on the regional average phase difference between NWP and in situ temperatures, and may not be suitable for other parts of the world.

[78] Prior to the launch of SMAP, soil temperature estimates from NWP systems are expected to be available at higher resolutions as the land and atmospheric modeling and analysis grids are refined. At the time of this writing, the three NWP systems evaluated here are already operating at a higher spatial resolution than what was available

for this study. Custom tile-based output from the (georeferenced) tiles of the GEOS-5 system might further improve the resolution. Moreover, gains in accuracy are expected from improvements in the atmospheric data assimilation components and through the addition of land data assimilation modules in operational systems. For example, an enhanced soil moisture analysis based on the extended Kalman filter was implemented in the operational ECMWF model in November 2010 [Rosnay *et al.*, 2011].

[79] Although the 0.05 m temperature plays the dominant role in the modeled effective temperature for L-band microwave emission, the error might be moderated further if the temperature of the deeper layers is known more accurately. Future studies should investigate the errors of NWP soil temperature estimates at deeper depths up to around 0.3 m and the potential to use the deeper layer information in the parameterization of the effective temperature of the emitting surface.

Appendix A: CMA Series Approach to Temperature Profile Modeling

[80] Instead of the classical Fourier analysis, the change in temperature with depth is estimated based on a summation of residuals after removing the central moving average (CMA), and was earlier used to implement the Van Wijk model [Holmes *et al.*, 2008]. To more fully capture seasonal variation, this approach is expanded from considering only two harmonics (daily and annual) to capture ten harmonics periods (P) of less than a day to a year with $P = [0.5, 1, 2, 3, 4, 8, 16, 32, 64, 365]$ days. For example, $P(n = 1) = 0.5$ days and $P(n = 10) = 365$ days. Given a time series of temperature at a single depth (T_t), and an estimate of the phase shift of the daily temperature harmonic between this input depth and that of the target depth ($d\varphi$), the temperature at the output depth (T_t^*) can be estimated as

$$T_t^* = \bar{T}_{n=10,t} + \sum_{n=1}^{10} H_{n,t} e^{d\varphi_n}, \quad (\text{A1})$$

where n is the subscript into the array of harmonics P and

$$d\varphi_n = d\varphi / \sqrt{P(n)}, \quad (\text{A2})$$

$$t_n = t - d\varphi_n \cdot P(n) / 2\pi, \quad (\text{A3})$$

and for $n = 1$

$$\bar{T}_{n=1,t} = \text{CMA}_{n=1}(T_t), \quad (\text{A4})$$

$$H_{n=1,t} = T_t - \bar{T}_{n=1,t},$$

and for $n > 1$

$$\bar{T}_{n,t} = \text{CMA}_n(\bar{T}_{n-1,t}), \quad (\text{A5})$$

$$H_{n,t} = \bar{T}_{n-1,t} - \bar{T}_{n,t}.$$

[81] The CMA is the central moving average over the time period $[t - P(n)/2, t + P(n)/2]$ and the harmonics are

interpolated to the phase corrected time (t_n) by means of a cubic spline interpolation.

[82] **Acknowledgments.** The authors would like to thank the Global Modeling and Assimilation Office (GMAO) and the GES DISC (both at NASA Goddard Space Flight Center) for the dissemination of MERRA; the taxpayers of the State of OK for funding the OK mesonet through the OK State Regents for Higher Education and the OK Department of Public Safety; the European Centre for Medium range Weather Forecasting (ECMWF) for the dissemination of their data (by way of the SMAP Test bed), and the National Centers for Environmental Prediction for dissemination of GDAS data (by way of Hiroko Beaudoin of NASA's Hydrological Science Branch). Rolf Reichle was funded by the NASA SMAP Science Definition Team. The USDA is an equal opportunity provider and employer.

References

- Aires, F., C. Prigent, and W. B. Rossow (2004), Temporal interpolation of global surface skin temperature diurnal cycle over land under clear and cloudy conditions, *J. Geophys. Res.*, *109*, D04313, doi:10.1012/2003JD004173.
- Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts (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*(3), 623–643.
- Betts, A. K., and J. H. Ball (1995), The FIFE surface diurnal cycle climate, *J. Geophys. Res. (Atm.)*, *100*(D12), 25,679–25,693.
- Choudhury, B. J., T. J. Schmugge, and T. Mo (1982), A parameterization of effective soil temperature for microwave emission, *J. Geophys. Res.*, *87*, 1301–1304.
- Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, and K. Praveen (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*(D20), 24,809–24,822.
- 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*(D22), 8851, doi:10.1029/2002JD003296.
- Entekhabi, D. et al. (2010a), The soil moisture active passive (SMAP) mission, *Proc. IEEE*, *98*(5), 704–716.
- Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow (2010b), Performance metrics for soil moisture retrievals and application requirements, *J. Hydrometeorol.*, *11*(3), 832–840, doi:10.1175/2010JHM1223.1.
- Fiebrich, C. A., J. E. Martinez, J. A. Brotzge, and J. B. Basara (2003), The Oklahoma mesonet's skin temperature network, *J. Atmos. Oceanic Technol.*, *20*(11), 1496–1504.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, *377*(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003.
- Holmes, T. R. H., P. de Rosnay, R. de Jeu, J.-P. Wigneron, Y. H. Kerr, J.-C. Calvet, M. J. Escorihuela, K. Saleh, and F. Lemaitre (2006), A new parameterization of the effective temperature for L band radiometry, *Geophys. Res. Lett.*, *33*, 7405.
- Holmes, T. R. H., M. Owe, R. A. M. De Jeu, and H. Kooi (2008), Estimating the soil temperature profile from a single depth observation: A simple empirical heatflow solution, *Water Resour. Res.*, *44*(2), W02412, doi:10.1029/2007WR005994.
- Holmes, T. R. H., R. A. M. De Jeu, M. Owe, and A. J. Dolman (2009), Land surface temperature from Ka band (37 GHz) passive microwave observations, *J. Geophys. Res.*, *114*(D4), D04113, doi:10.1029/2008JD010257.
- Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerron, J. N. VanDriel, and J. Wickham (2007), Completion of the 2001 national land cover database for the conterminous United States, *Photogr. Eng. Remote Sens.*, *73*(4), 337–341.
- 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*(2), 167–182, doi:10.1175/2007JTECHA993.1.
- Jackson, T. J., M. H. Cosh, R. Bindlish, P. J. Starks, D. D. Bosch, M. Seyfried, D. C. Goodrich, M. S. Moran, and J. Du (2010), Validation of Advanced Microwave Scanning Radiometer soil moisture products, *IEEE Trans. Geosci. Remote Sens.*, *48*(12), 4256–4272.
- Johnson, D. B., P. Flament, and R. L. Bernstein (1994), High-resolution satellite imagery for mesoscale meteorological studies, *Bull. Am. Meteor. Soc.*, *75*, 5–34.
- Kerr, Y. H., P. Waldteufel, J.-P. Wigneron, J.-M. Martinuzzi, J. Font, and M. Berger (2001), Soil moisture retrieval from space: The soil moisture and ocean salinity (SMOS) mission, *IEEE Trans. Geosci. Remote Sens.*, *39*(8), 1729–1735.
- Koster, R. D., M. J. Suarez, A. Ducharme, 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*(D20), 24,809–24,822.
- Le Vine, D. M., G. S. Lagerloef, F. R. Colomb, S. H. Yueh, and F. A. Pellerano (2007), Aquarius: An instrument to monitor sea surface salinity from space, *IEEE Trans. Geosci. Remote Sens.*, *45*(7), 2040–2050.
- Li, L., P. W. Gaiser, B. C. Gao, R. M. Bevilacqua, T. J. Jackson, E. G. Njoku, C. Rudiger, J. C. Calvet, and R. Bindlish (2010), WindSat global soil moisture retrieval and validation, *IEEE Trans. Geosci. Remote Sens.*, *48*(5), 2224–2241.
- Mahfouf, J.-F., P. Viterbo, H. Douville, and A. C. M. Beljaars (2000), A revised land-surface analysis scheme in the integrated forecasting system, *ECMWF Newsl.*, edn. 88, 8–13.
- McPherson, R. A. et al. (2007), Statewide monitoring of the mesoscale environment: A Technical update on the Oklahoma mesonet, *J. Atmos. Oceanic Technol.*, *24*(3), 301–321.
- Njoku, E. G., T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem (2003), Soil moisture retrieval from AMSR-E., *IEEE Trans. Geosci. Remote Sens.*, *41*, 215–229.
- O'Neill, P. E., E. G. Njoku, T. J. Jackson, J. Shi, and S. Chan (2010), *SMAP Level 2 & 3 Soil Moisture (Passive) Algorithm Theoretical Basis Document*, Preliminary Version 2.5, SMAP Science Document 024.
- Owe, M., R. de Jeu, and T. Holmes (2008), Multisensor historical climatology of satellite-derived global land surface moisture, *J. Geophys. Res.*, *113*, F01002, doi:10.1029/2007JF000769.
- Peters-Lidard, C. D., E. Blackburn, X. Liang, and E. F. Wood (1998), The effect of soil thermal conductivity parameterization on surface energy fluxes and temperatures., *J. Atm. Sci.*, *55*, 1209–1224.
- Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. Mahanama, and A. Toure (2011), Assessment and enhancement of MERRA land surface hydrology estimates, *J. Clim.*, *24*(24), 6322–6338, doi:http://dx.doi.org/10.1175/JCLI-D-10-05033.1.
- Rienecker, M. M., et al. (2011), MERRA - NASA's modern-era retrospective analysis for research and applications, *J. Clim.*, *24*(14), 3624–3648, doi:http://dx.doi.org/10.1175/JCLI-D-11-00015.1.
- Rosnay, P. de, M. Drusch, G. Balsamo, C. Albergel, and L. Isaksen (2011), Extended Kalman filter soil-moisture analysis in the IFS, *ECMWF Newsl.*, edn. 127, 12–15.
- Shafer, M. A., C. A. Fiebrich, D. S. Arndt, S. E. Fredrickson, and T. W. Hughes (2000), Quality assurance procedures in the Oklahoma mesonet network, *J. Atmos. Oceanic Technol.*, *17*(4), 474–494.
- Ulaby, F. T., R. K. Moore, and A. K. Fung (1986), *Microwave Remote Sensing: Active and Passive. Vol. III. From Theory to Applications*, Artech House, Norwood, MA.
- Van Wijk, W. R., and D. A. de Vries (1963), Periodic temperature variations in a homogeneous soil., in *Physics of the Plant Environment*, pp. 102–143, North-Holland Publ. Co., Amsterdam.
- Wigneron, J.-P., L. Laguerre, and Y. H. Kerr (2001), A Simple parameterization of the L-band microwave emission from rough agricultural soils, *IEEE Trans. Geosci. Remote Sens.*, *39*, 1697–1707.
- Wilheit, T. T. (1978), Radiative transfer in a plane stratified dielectric, *IEEE Trans. Geos. Electr.*, *16*(2), 138–143.

J. B. Basara, Oklahoma Climatological Survey, University of Oklahoma, 120 David L. Boren Blvd., Suite 2900, Norman, OK 73072, USA.

T. R. H. Holmes and T. J. Jackson, Hydrology and Remote Sensing Lab, USDA Agricultural Research Service, 10300 Baltimore Ave., Rm. 104, Bldg. 007, BARC-West, Beltsville, MD 20705, USA. (thomas. & holmes@ars.usda.gov)

R. H. Reichle, Global Modeling and Assimilation Office, NASA GSFC, Mail Code 610.1, 8800 Greenbelt Rd., Greenbelt, MD 20771, USA.