An Assessment of Surface Soil Temperature Products from Numerical Weather Prediction Models Using Ground-based Measurements

Thomas R.H. Holmes (Corresponding Author)  
Hydrology and Remote Sensing Lab.  
USDA ARS  
10300 Baltimore Ave,  
Rm 104 Bldg 007 BARC-West,  
Beltsville, MD, 20705, United States

T. J. Jackson  
Hydrology and Remote Sensing Lab.  
USDA ARS  
10300 Baltimore Ave,  
Rm 104 Bldg 007 BARC-West,  
Beltsville, MD, 20705, United States

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

J. Basara  
Oklahoma Climatological Survey  
The University of Oklahoma  
120 David L. Boren Blvd., Suite 2900  
Norman, OK 73072, United States
Abstract

Surface soil temperature estimates are needed to retrieve surface 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 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). SMAP requires soil temperature estimates at a nominal contributing depth of 0.05 m. Therefore, each NWP set is synchronized to match the mean phase of the in situ data at 0.05 m. Accuracy and precision are assessed as a function of time of day with specific attention directed to the SMAP early morning overpass time at around 6am local time. The ECMWF and MERRA temperature products have very similar performance metrics, with a root mean square error of 1.7 K at 6am local time, indicating that both products meet the error budget criteria as defined for SMAP.
1. Introduction

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.

The motivation for this particular investigation is the need for accurate surface soil temperature products for estimating the effective land surface temperature, as utilized for retrieving surface soil moisture from low-frequency passive microwave observations. Of particular concern is the proposed National Aeronautics and Space Administration (NASA) satellite called Soil Moisture Active Passive (SMAP) [Entekhabi, Njoku, et al., 2010], which will require that this information be provided by a dynamic ancillary resource.

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 multi-frequency sensors with the lowest frequency either at C-band (~6 GHz) or X-band (~10 GHz) [Jackson et al., 2009; 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.

However, these multi-frequency platforms have limitations in how much soil moisture information they can provide and as a result a new generation of lower frequency (L-band) satellites are in various stages of development and implementation. These offer an improved depth of sensing, reduced vegetation attenuation, and in one case 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 does not require ancillary surface temperature data because it exploits multiple angle observations to retrieve soil moisture and soil temperature. It does however, use modeled soil temperature from the European Center for Medium range Weather Forecasts (ECMWF) as an initial value in its iterative optimization scheme. The next L-band satellite is NASA’s Aquarius / SAC-D (2011 launch) satellite [Le Vine et al., 2007]. In addition to an L-band radiometer this satellite will also include a Ka-band radiometer from the Space Agency of Argentina 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, Njoku, et al., 2010]. Therefore, for SMAP it will be necessary to provide soil temperature as a dynamic ancillary data set. Some requirements for this dataset are that it

- include the temperature of the soil at a depth of 0.05 m below the surface
- have a spatial resolution of at least 0.25 degree so that the resolution requirements for soil moisture are not compromised,
Soil Temperature from Numerical Weather Prediction Models

- closely match the overpass time of SMAP, at 6 am/pm local time,
- be available within a few hours of the satellite observation, so that the latency goals for soil moisture retrievals can be achieved, and
- meet the error budget of SMAP. In the error budget for achieving the target soil moisture accuracy of 0.04 m$^3$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].

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.

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.

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 87% of the soil emission at L-band for an incidence angle of 50 degrees varies between 0.03 m for a wet soil to 0.3 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 can not 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 \cite{Choudhury:1982, J-P:2001, Holmes:2006}.

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 net radiation, latent and sensible heat flux 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 sixty to ninety minutes after solar noon \cite{Fiebrich:2003, Betts:1995}. 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.

The length of the time lag between soil temperature measurements at two different depths is determined by the vertical distance between the measurements depths and the thermal properties of the medium. As shown by Van Wijk & de Vries [1963] this phase
shift is accompanied by an exponential reduction in amplitude of the daily temperature cycle as the measurement depth is moved deeper into the soil. 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.

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. This procedure removes a large part of the systematic differences between the data sets and can be applied to any pair of datasets, 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

This study analyzes soil temperature data at a fifteen minute 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 (see below) 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 (Koppen climate classification Cfa) to dryer semi-arid, or steppe climate (BSk) towards the western panhandle of the state. Accordingly, the vegetation ranges from short grassland and shrubs in the West, to pasture land and forests in the East (see Figure 1). The 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

The Oklahoma Mesonet [Illston et al., 2010; McPherson et al., 2007] 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 seconds, only the average over 15 minutes 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. 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, 4 of the 79 stations were removed from the analysis completely.

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. In total, 75 stations were available for this study, covering 56 half degree grid boxes between 33-37°N, 94-103°S. For each of these grid boxes, one station was selected to represent the in situ data for that box (detailed in Section 3.3). The stations that are not used in this first selection are used to replicate the experiment and validate the results. The station codes for these two selections are shown in Appendix A.

----------------- Table 1 about here -----------------
2.3 Numerical Weather Prediction Products

The general features of the NWP products that will be evaluated are summarized in Table 1. The following sections provide additional detail.

2.3.1 Modern-Era Retrospective analysis for Research and Applications

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.]). 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 degrees 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 closely resembles the GMAO operational analysis with near real time availability that currently runs at a 0.25 degree resolution.

The surface processes are described by the NASA Catchment land surface model [Ducharne et al., 2000; Koster et al., 2000]. Each MERRA grid cell contains several irregularly shaped catchments, called tiles. For each tile, surface exchange processes and “surface” temperatures are represented separately for sub-tiles that are characterized by one of three unique hydrological states: saturated, unsaturated, and wilting. The sub-tile fractions of each catchment are modeled dynamically based on the total amount of water
in the catchment. The “surface” temperature of a grid cell is then obtained by area-weighted averaging of the “surface” temperatures of all sub-tiles within the grid cell. The above-mentioned “surface” temperatures are prognostic variables of the model and represent a bulk layer that includes the vegetation canopy and the top 0.05 m of the soil column. Below this surface layer, a single deeper soil temperature profile for each tile is modeled with a heat diffusion equation using 6 layers, with boundaries at 0.05 m, 0.15 m, 0.34 m, 0.73 m, 1.49 m, 3.00 m, and 13.00 m. In this study we analyzed the area-weighted surface temperatures that describe the temperature of the canopy/top 0.05 m soil layer over land.

2.3.2 ECMWF analysis and forecasting system

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 hour 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/.

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 and are not dynamically related to surface
state, as in MERRA. 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 0-0.07 m.

2.3.3 NCEP global data assimilation system (GDAS)

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 July 27, 2010. The LSM was replaced with the NCEP Noah LSM beginning with May
31, 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 fields were considered,
and it was found that the average of both fields yielded the best results. Therefore only
the results of this combination are presented.

2.4 Spatial scales and recording times

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

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 56 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 19 grid cells that are sampled by more than one station.

The times associated with the NWP outputs are defined in Coordinated Universal Time (UTC). The sparsely sampled temperature series are interpolated to the 15 minute interval of the in situ data by means of a piecewise cubic spline interpolation.

The observation times of the Mesonet data are recorded in Central Standard Time (CST), which is six hours 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 sun-synchronous satellite SMAP. To illustrate, the morning overpass of SMAP will occur at about 6:23 AM local solar time at the latitude of Oklahoma (for a descending Equator overpass at 6 AM [D. B. Johnson et al., 1994]). However, in CST this will be at 6:43 AM in Eastern Oklahoma, and at 7:13 AM at the end of the Western panhandle. For this reason, all time stamps are converted to local solar time 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 minutes) 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

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 ($\rho$), and the standard deviation of the in situ data ($\sigma_{\text{true}}$). Together they are used to calculate the standard error of estimate (SEE):

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

Note that we are using $\sigma_{\text{true}}$ instead of the standard deviation of the model data ($\sigma_{\text{est}}$), which is not the traditional approach. By using the $\sigma_{\text{true}}$, 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 $\sigma_{\text{true}}, \sigma_{\text{est}}, \rho$, and mean bias (b) as shown by Entekhabi et al. [2010]. For the discussion in the present paper it is useful to rewrite those expressions in terms of SEE:
The relationships between these three performance metrics (Equations 1 to 3) show that $\text{SEE} \leq \text{ubRMSE} \leq \text{RMSE}$ and may serve to quantify a decreasing level of bias removal (in the same units as the variable being assessed).

Equation 2 shows that the lowest ubRMSE is obtained when $\sigma_{est} = \sigma_{true}\rho$. Considering that in a real application $\rho$ 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 $\rho$ values in temperature comparisons are generally high (between 0.9 and 0.95), this effect will be limited to favoring an underestimation of $\sigma_{est}$ by 5 to 10%.

The performance metrics are affected by the accuracy of the in situ data, and how representative the sites are for their 0.5 x 0.5 degree 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 Theory

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 ($z_D$):

\begin{equation}
d\varphi = \frac{-dz}{z_D}
\end{equation}

\begin{equation}
A_{z_2} = A_{z_1}e^{\left(\frac{dz}{z_D}\right)}
\end{equation}

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 ($\alpha$, m$^2$/s), and indicates the distance ($z_D$) over which the amplitude of the wave is reduced by 63%:

\begin{equation}
z_D = \sqrt{\frac{2\alpha}{2\pi f}}
\end{equation}

where $f$ (1/s) is frequency of the temperature wave.

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 three soil types, Sand, Sandy Clay, and Clay soils 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. Based on these simulations the phase shift is rather constant at soil moisture levels above 0.10 m$^3$/m$^3$, with values ranging between 70 and 100 minutes
depending on soil type. Below this soil moisture level the phase shift may increase sharply to 180 minutes. 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

Because soil temperature harmonics change with depth, soil temperature records from different depths cannot be directly compared. 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 remove the phase difference between the temperature measurements using the heat flow principles as described in Section 3.2. Following Equation 4, the relative distance (the vertical distance between input and target depth, divided by the damping depth) can be replaced by $d\varphi$. This method avoids the modeling of the soil thermal properties by using their integrated effect over a given distance, as manifested in the mean phase shift of the daily temperature harmonic between two measurement depths. The temperature at the target depth is then modeled 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 B.
Soil Temperature from Numerical Weather Prediction Models

The phase synchronization method was first tested on the in situ data, by modeling the temperature at the depth of the shallow record ($T_{5}^*$) based on the data from the deeper sensor ($T_{10}$) and the observed mean phase difference between the two records:

$$T_{5}^* = f(T_{10}, d\varphi)$$  \hspace{1cm} (7)

The phase of each record is determined by optimizing $\varphi$ so that the RMSE is minimized between the mean diurnal cycle and the sine function with a maximum at noon:

$$T_{\text{sim}} = \bar{T}_z + A \sin((t - \varphi)2\pi - \pi/4)$$  \hspace{1cm} (8)

where $A$ is the amplitude of the diurnal cycle. Figure 3 shows the fitted values of $\varphi$ for $T_5$ in the top panel. 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 can not assume that $T_5$ represents exactly the 0.05 m soil depth. Furthermore, there may be differences in vegetation density between the plots 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 $T_5$, which is illustrated by the spread in phase of $T_5$, as determined for each station. The standard deviation is 30 minutes for a mean phase shift of 4 hours and 10 minutes (Figure 3, top panel).
If the sensors 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 m and 0.1 m (nominal) in situ temperature records is shown in Figure 3 (middle panel). The mean $\phi$ of 80 minutes fits comfortably within the theoretical range as given in Section 3.2, 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 90 minutes). For the stations with a phase shift outside these bounds the actual vertical distance between the sensors may deviate from the nominal 0.05 m.

For each station the RMSE was then calculated between $T_5^*$ and $T_5$ (see Figure 3, lower panel). The RMSE is generally well below the stated accuracy for the probes (0.5 K). Some stations have a higher RMSE, which we attributed to a change in the relative depth of the sensors over the year. To minimize possible errors in the in situ measurements, stations with an overall RMSE of more then 0.8 K were discarded from further analysis.

Figure 4 shows the daily cycle of the aggregated performance metrics as calculated 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 ~80
minutes later. The mean diurnal cycle of the phase synchronized record ($T_5^*$: blue line) is virtually indistinguishable from that of the actual in situ measurements at 0.05 m. As a result, all performance metrics improve significantly from the case when no synchronization is applied (middle panel), to the phase synchronized results (lower panel). The RMSE as calculated between $T_5$ and $T_{10}$ (middle panel, red line) has minima of ~0.4 K at 8 AM/PM but spikes to 1.5 K at other times of the day. Applying the phase synchronization reduces this error throughout the day, as is evident from the RMSE between $T_5$ and $T_5^*$ (lower panel, blue line). Still, the large gradients in the afternoon yield an increase in RMSE to 0.7 K, from a minimum of 0.3 K at 8 AM.

3.4 Implementation of phase synchronization

The phase synchronization method uses only the $d\phi$ as determined based on 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 is well below the stated accuracy of the probes (0.5 K) for most of the day. Therefore, it will be used in the following analysis to match the phase of each NWP product with that of the in situ data.

The uncertainty about the actual depth of the in situ sensors complicates a robust assessment of NWP temperature products. In order to minimize the phase coherence of the validation target, the in situ $T_5$ is synchronized for each station individually to match the average phase of all stations. In this way, the effect of varying sensor depths between stations is minimized and the resulting $T_5^*$ will better represent the actual temperature at
Soil Temperature from Numerical Weather Prediction Models

0.05 m depth under native sod. Furthermore, for each grid box with multiple stations, the station with the phase of T5 closest to the average of all stations was selected.

The difference between the mean phase of the in situ data and each NWP set is then determined. Each NWP data set is corrected for this mean phase shift to create data sets for which the phase of the mean daily cycle is in line with that of T5. The assessment of the NWP temperature will be based on these synchronized data sets, which should be regarded as the best estimate of the 0.05 m temperature to be obtained from these NWP models.

4. Inter comparison of NWP products

Three surface (soil) temperature products from different NWP models that are most relevant to the SMAP mission were compared. These are the 1) GMAO’s MERRA, 2) ECMWF’s operational integrated forecast system, and 3) the global data assimilation system (GDAS) as used by NCEP. Table 1 is an overview of the general specifications of these products. All products are resampled to a 0.5 by 0.5 degree 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 TME, TEC, and TNC. Table 2 lists the general results of the comparison of each NWP set with the in situ temperature. The applied phase adjustments are significant, from 50 minutes for TEC to almost three hours for TME. The mean daily cycle over the period is best described by the synchronized TNC, with a small mean bias and only a 10 % overestimation of the mean daily amplitude. Both TME and TEC have a 60 % overestimation of the daily amplitude.
As an input to soil moisture retrieval algorithms based on sun-synchronous 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 4.1. Section 4.2 will then analyze the results at the time of the morning overpass of SMAP (~6.23 AM local solar time over Oklahoma). The robustness of these results will be tested in Section 4.3, where several replications are discussed.

--- Figure 5 and Table 2 about here ---

### 4.1 Daily performance metrics

Figure 5 shows the mean daily cycle for the three products (blue lines) and the in situ data (black lines). This figure also displays their hourly performance metrics as discussed previously. As shown in the top row, all three original (without phase synchronization) NWP products (red dashed lines) 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 morning is reduced to less than 1 K.

The second row 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 $R^2 = 0.9$ to $R^2 = 0.93$ in the night and early morning.

For each product the three performance metrics (SEE, ubRMSE, and RMSE) are displayed in the same graph (Figure 5, third row) to illustrate the additive nature of these
metrics. $T_{EC}$ and $T_{ME}$ have a precision, as quantified by SEE, of around 1.3 K during the night through 9 AM in the morning. During the day the precision is poorer, reaching SEE values of 2 K. Although $T_{NC}$ does not reach as low a value of SEE, its performance is more stable over the day.

As shown in Section 3.1, the absolute error for the region is a combination of SEE, bias, and bias in standard deviation. The resulting RMSE is displayed in blue in the third row of Figure 5. The RMSE values calculated for $T_{EC}$ and $T_{ME}$ are very similar, varying from a low of ~1.7 K in the morning to 3.2 K in the afternoon. The $T_{NC}$ accuracy is more stable over the day, varying only from 1.8 to 2.3 K, which can be attributed to this product being able to accurately estimate the daily amplitude in temperature.

In summary, the synchronized products all have a high correlation with the in situ measurement throughout the day. Even though $T_{EC}$ and $T_{ME}$ overestimate the daily amplitude significantly compared to $T_{NC}$, the early morning errors are lowest for these two products. During the day, $T_{NC}$ shows the lowest errors.

4.2 Results for 6 AM local time

For the narrow time window of concern to the SMAP mission (6:23 AM local solar time for Oklahoma) the results are examined in greater detail. The aggregated statistics at this time of day for the region as a whole are given in Table 3. The standard error for $T_{EC}$ is 1.3 K, with a bias of $0.4 \pm 0.7$ K. The RMSE is 1.7 K, with 8 of the 56 grid cells having an error above the target accuracy of 2 K. The results for $T_{ME}$ are almost identical,
trailing in RMSE by a narrow margin of 0.05 K but with only 7 cells with an error above 2 K. The error at 6 AM is slightly higher for T_{NC}, with an SEE of 1.4 and an RMSE of 1.8 K. However, for T_{NC}, 14 of the 56 grid cells have an error above 2 K, which is much higher than found for T_{EC} and T_{ME}.

In Figure 6, maps of the RMSE for Oklahoma at this time of day are presented for each model product. The distribution of the RMSE over the state is far from homogenous, but no obvious relationship with spatially varying features such as vegetation density 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.

Because the bias is relatively small during the night, the possible gain from bias reduction techniques is only on the order of 0.2 K for T_{ME} and T_{EC}. During the day the bias is more directly related to overestimation of the daily amplitude and, as a result, better modeling techniques have the potential of reducing this error to values closer to the standard error.

Since the primary location for this analysis is Oklahoma, with a UTC offset of 6 hours, the 6 AM 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 6 hours analysis steps) is minimized at this time, for this particular location. If available, using the 3-hour forecast data to better interpolate the temperature between analyses steps might help improve the overall performance. On the other hand, the high temporal output resolution of MERRA (one hour steps) makes its results likely more independent of longitude.
For the morning overpass of SMAP, the 0.05 cm temperature can be estimated with an accuracy of 1.7 K from either $T_{EC}$ or $T_{ME}$. Because this error varies spatially we found RMSE errors below 2 K at 9 out of 10 stations in this analysis. The scope for bias reduction appears limited, but more optimized profile modeling may reduce this error further.

--- Figure 6 About here ---

### 4.3 Replication

To test the robustness of the results, the analysis was repeated for the set of grid cells with duplicate stations and the results are listed in Table 3. The overall performance metrics are almost identical for the two groups of stations, which gives us confidence that the retrieved results are not significantly influenced by individual stations.

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 results in an overall 0.1 K reduction in SEE levels relative to 2009, as shown in Figure 7. 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 this results in improved accuracy. However, because of the different dynamic range in temperature between 2004 and 2009, it is not possible to assess the impact of the
changes in the NWP systems on the performance of the temperature products in absolute terms.

The analysis was repeated for the primary set of Mesonet stations, but now for the entire year 2009. This period includes 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, model estimates below 273 K are excluded from the analysis. The performance over the entire year is summarized in Figure 8, showing an increase in error during the night and a decreased error level during the day. The exact metrics for 6 AM are listed in Table 4 and show a minor degradation of the results, with an RMSE of 2.1 K for $T_{NC}$, 1.8 for $T_{ME}$, and 1.8 K for $T_{EC}$. Perhaps because the original MERRA temperature refers to temperatures closest to the surface, its performance degrades most if temperatures below 273 K are considered. In that case the RMSE at 6AM would be 2.6 $T_{ME}$, but still only 2.2 K and 2.0 K for $T_{NC}$ and $T_{EC}$ respectively.

All three replications confirm the general level of attainable error levels for the soil temperature products. 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.
4.4 Which soil layer do the models represent?

Following the theoretical considerations detailed in Section 3.2, and as confirmed by the in situ data, the mean phase shift over a 0.05 m soil layer should be around 80 minutes. By using this as a metric for converting phase shift into an estimate of vertical distance it is possible to assess the equivalent soil depth that corresponds to each NWP temperature product. On average the phase of the temperature for the 1st soil layer of ECMWF is 51 minutes ahead of T5, which translates into a 0.03 m vertical distance (0.05·51/80). This indicates that this ECMWF soil temperature represents a depth of about 0.02 m, which is close to the middle of this soil layer (0-0.07 m).

The phase difference of the surface temperature of NCEP is slightly larger than that between the in situ probes at 0.05 and 0.1 m, on average 90 instead of 80 minutes. This means that the NCEP temperature, derived by averaging the skin temperature and that of the 0 to 0.1 m model layer, represents the very surface of the soil. This might be explained if the vegetation temperature is considered separately from the 0 cm soil depth and has its own heat capacity resulting in the extra phase shift.

For the surface temperature of MERRA the phase difference is much larger, on average twice as large as that found between the in situ probes at 0.05 and 0.1 m. This means that even if the NWP temperature is considered to represent the very surface of the soil, heat flow principles based on 0.05 m vertical depth can not explain the phase shift between the two measurements. An equivalent to the damping as caused by a 0.05 m soil layer is necessary to explain the phase difference found for MERRA. This additional damping might in fact be caused by the vegetation. The MERRA model specifically states that its surface temperature refers to the average temperature of the canopy and the
Soil Temperature from Numerical Weather Prediction Models

upper 0.05 m of the soil. Considering the large phase shift with the 0.05 m sensors, it appears that the canopy temperature is either weighted too heavily in this average, or that the damping effect of vegetation on the temperature cycle is underestimated. The results for MERRA might then be improved if the 0-0.05 m soil temperature can be separated from the canopy temperature.

Even though the temperature sets relate to quite different soil (or vegetation) layers, a simple phase synchronization of the records allows us to estimate the temperature in the early morning with relatively high accuracy. For better estimation of the temperatures in the late afternoon more complicated profile modeling techniques will be necessary that account for the latent heat loss from the surface soil layers or vegetation during the day. A possible approach was detailed in earlier work by Holmes et al. [2008], and could be adapted for each individual model structure.

5. Discussion and Conclusion

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.

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. This was adjusted for by synchronizing the phase of the daily cycles of the temperature records. This procedure should minimize the limitation of using single in situ measurements to validate areal averages of temperature data and give an indication
Soil Temperature from Numerical Weather Prediction Models

of attainable levels of precision and accuracy. Furthermore, in order to compare the various NWP products, representing different soil depths, each product is synchronized to match the mean phase of the in situ data. 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.

From these results, it can be concluded that for the morning overpass of SMAP, at 6am local time, the 0.05 cm temperature can be estimated with an SEE of 1.3 K and an RMSE of 1.7 K using the ECMWF and NASA/GMAO systems. These results suggest that the overall accuracy of these two NWP soil temperature products does not exceed the level allotted in the error budget for SMAP (2 K). However, this error varies spatially and for ten percent of the grid cells the absolute error does exceed 2 K, and this may reduce the area for which soil moisture can be achieved with the desired accuracy. The temperature from NCEP, although overall below the 2 K threshold, does not seem a suitable candidate for use in the SMAP soil moisture retrieval as 25 percent of the grid cells had errors in excess of 2 K at 6 AM.

Some of the error is likely related to how representative the ground data is for the 0.5 degree grid box and may overstate the actual value. 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. On the other hand, the phase synchronization was based on the regional average phase difference between NWP and in situ temperatures, and may not be suitable for every location.

Upon implementation of a temperature product into an L-band soil moisture retrieval algorithm the 0.05 m temperature will be used in the parameterization of the effective
temperature of the emitting surface, varying between 0.03 to 0.3 m. Although the 0.05 m temperature plays the dominant role in the modeled effective temperature, the error might be moderated further if the temperature of the deeper layers is known more accurately.

Prior to the launch of SMAP, soil temperature estimates from NWP systems are expected to be available at higher resolutions as the atmospheric modeling and analysis grid are refined. All NWP systems evaluated here are operating currently at a higher spatial resolution than what was available for this study. Custom tile-based output from the (geo-referenced) 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.

**Acknowledgements**

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 Oklahoma fund the Oklahoma Mesonet through the Oklahoma State Regents for Higher Education and the Oklahoma Department of Public Safety; the European Centre for Medium range Weather Forecasting (ECMWF) for the dissemination of their data (by way of the SMAP Testbed), and National Centers for Environmental Prediction for dissemination of GDAS data (by way of Hiroko Beaudoing of NASA’s Hydrological Science Branch).
Soil Temperature from Numerical Weather Prediction Models

682

**Literature**

684  irrigation on midsummer surface fluxes and temperature under dry synoptic
687
688  Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K.
689  Betts (2009), A revised hydrology for the ECMWF model: Verification from field
690  site to terrestrial water storage and impact in the Integrated Forecast System,
692
693  Betts, A. K., and J. H. Ball (1995), The FIFE surface diurnal cycle climate, *Journal of
694  Geophysical Research-Atmospheres*, 100(D12), 25,679-25,693.
695
696  Choudhury, B. J., T. J. Schmugge, and T. Mo (1982), A Parameterization of Effective
698
700  catchment-based approach to modeling land-surface processes in a general
702  Res.*, 105(D20), 24,809-24,822.
703
704  Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J.
705  D. Tarpley (2003), Implementation of Noah land surface model advances in the
706  National Centers for Environmental Prediction operational mesoscale Eta model,
707  *Journal of geophysical research*, 108(D22), 8851.
708
709  Entekhabi, D., E. G. Njoku, et al. (2010), The Soil Moisture Active Passive (SMAP)
711
713  for Soil Moisture Retrievals and Application Requirements, *J. Hydrometeor*,
715
716  Fiebrich, C. A., J. E. Martinez, J. A. Brotzge, and J. B. Basara (2003), The Oklahoma
717  Mesonet's Skin Temperature Network, *Journal of Atmospheric and Oceanic
718  Technology*, 20(11), 1496-1504.
719
720  Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the
721  mean squared error and NSE performance criteria: Implications for improving
722  hydrological modelling, *Journal of Hydrology*, 377(1-2), 80-91,
723  doi:10.1016/j.jhydrol.2009.08.003.
724
725  Holmes, T. R. H., R. A. M. de Jeu, M. Owe, and A. J. Dolman (2009), Land surface
726  temperature from Ka band (37 GHz) passive microwave observations, *J.
Soil Temperature from Numerical Weather Prediction Models


Soil Temperature from Numerical Weather Prediction Models


## Appendix A

Table A1. List of Mesonet stations used for each replication. Stations that are not used have either large data gaps after Mesonets internal quality control, or are removed after a manual inspection of the data.

<table>
<thead>
<tr>
<th>Station</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Mesonet QC data gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACME Rush Springs</td>
<td>34.80833</td>
<td>-98.0233</td>
<td>0 0 0</td>
</tr>
<tr>
<td>ALTU Altus</td>
<td>34.58722</td>
<td>-99.3381</td>
<td>1 0 0</td>
</tr>
<tr>
<td>ALV2 Alva</td>
<td>36.70823</td>
<td>-98.7097</td>
<td>1 0 0</td>
</tr>
<tr>
<td>ANTL Antlers</td>
<td>34.22438</td>
<td>-95.7006</td>
<td>1 0 0</td>
</tr>
<tr>
<td>APAC Apache</td>
<td>34.91418</td>
<td>-98.2922</td>
<td>1 0 0</td>
</tr>
<tr>
<td>ARNE Arnett</td>
<td>36.07204</td>
<td>-99.9031</td>
<td>1 0 0</td>
</tr>
<tr>
<td>BEAV Beaver</td>
<td>36.80253</td>
<td>-100.53</td>
<td>1 0 0</td>
</tr>
<tr>
<td>BESS Bessie</td>
<td>35.40185</td>
<td>-99.0585</td>
<td>0 1 0</td>
</tr>
<tr>
<td>BOIS Boise City</td>
<td>36.69256</td>
<td>-102.497</td>
<td>1 0 0</td>
</tr>
<tr>
<td>BREC Breckinridge</td>
<td>36.41201</td>
<td>-97.6939</td>
<td>1 0 0</td>
</tr>
<tr>
<td>BRIS Bristow</td>
<td>35.7805</td>
<td>-96.354</td>
<td>0 0 0</td>
</tr>
<tr>
<td>BURN Burneville</td>
<td>33.89376</td>
<td>-97.2692</td>
<td>1 0 0</td>
</tr>
<tr>
<td>BUTL Butler</td>
<td>35.5915</td>
<td>-99.2706</td>
<td>1 0 0</td>
</tr>
<tr>
<td>BYAR Byars</td>
<td>34.8497</td>
<td>-97.0033</td>
<td>0 1 0</td>
</tr>
<tr>
<td>CAMA Camargo</td>
<td>36.02866</td>
<td>-99.3465</td>
<td>0 0 0</td>
</tr>
<tr>
<td>CENT Centrahoma</td>
<td>34.60896</td>
<td>-96.3331</td>
<td>1 0 0</td>
</tr>
<tr>
<td>CHAN Sparks</td>
<td>35.65282</td>
<td>-96.8041</td>
<td>1 0 0</td>
</tr>
<tr>
<td>CHER Cherokee</td>
<td>36.74813</td>
<td>-98.3627</td>
<td>1 0 0</td>
</tr>
<tr>
<td>CHEY Cheyenne</td>
<td>35.54615</td>
<td>-99.7279</td>
<td>1 0 0</td>
</tr>
<tr>
<td>CHIC Chickasha</td>
<td>35.03236</td>
<td>-97.9145</td>
<td>0 0 0</td>
</tr>
<tr>
<td>CLAY Clayton</td>
<td>34.65657</td>
<td>-95.326</td>
<td>0 0 0</td>
</tr>
<tr>
<td>CLOU Cloudy</td>
<td>34.22321</td>
<td>-95.2487</td>
<td>1 0 0</td>
</tr>
<tr>
<td>CLRM Claremore</td>
<td>36.32112</td>
<td>-95.6462</td>
<td>1 0 0</td>
</tr>
<tr>
<td>COPA Copan</td>
<td>36.90987</td>
<td>-95.8855</td>
<td>0 1 1</td>
</tr>
<tr>
<td>DURA Durant</td>
<td>33.92075</td>
<td>-96.3203</td>
<td>1 0 0</td>
</tr>
<tr>
<td>ELRE El Reno</td>
<td>35.54848</td>
<td>-98.0365</td>
<td>1 0 0</td>
</tr>
<tr>
<td>ERIC Erick</td>
<td>35.20494</td>
<td>-99.8034</td>
<td>1 0 0</td>
</tr>
<tr>
<td>FAIR Fairview</td>
<td>36.26353</td>
<td>-98.4977</td>
<td>0 0 0</td>
</tr>
<tr>
<td>FITT Fittstown</td>
<td>34.55205</td>
<td>-96.7178</td>
<td>0 0 0</td>
</tr>
<tr>
<td>GOOD Goodwell</td>
<td>36.60183</td>
<td>-101.601</td>
<td>1 0 0</td>
</tr>
<tr>
<td>HASK Haskell</td>
<td>35.74798</td>
<td>-95.6405</td>
<td>1 0 0</td>
</tr>
<tr>
<td>HECT Hectorville</td>
<td>35.84162</td>
<td>-96.0024</td>
<td>1 0 0</td>
</tr>
<tr>
<td>HOLL Gould</td>
<td>34.6855</td>
<td>-99.8333</td>
<td>1 0 0</td>
</tr>
<tr>
<td>IDAB Idabel</td>
<td>33.83013</td>
<td>-94.8803</td>
<td>1 0 0</td>
</tr>
<tr>
<td>KENT Kenton</td>
<td>36.82937</td>
<td>-102.878</td>
<td>1 0 0</td>
</tr>
<tr>
<td>LAHO Lahoma</td>
<td>36.38435</td>
<td>-98.1114</td>
<td>1 0 0</td>
</tr>
<tr>
<td>LANE Lane</td>
<td>34.30876</td>
<td>-95.9972</td>
<td>0 0 0</td>
</tr>
<tr>
<td>MADI Lebanon</td>
<td>34.03579</td>
<td>-96.9439</td>
<td>0 0 0</td>
</tr>
<tr>
<td>MANG Mangum</td>
<td>34.83592</td>
<td>-99.424</td>
<td>0 0 0</td>
</tr>
<tr>
<td>MAYR Freedom</td>
<td>36.98707</td>
<td>-99.0111</td>
<td>1 0 0</td>
</tr>
<tr>
<td>MEDI Medicine Park</td>
<td>34.72921</td>
<td>-98.5694</td>
<td>1 0 0</td>
</tr>
<tr>
<td>MRSH Marshall</td>
<td>36.11685</td>
<td>-97.6069</td>
<td>0 0 0</td>
</tr>
</tbody>
</table>
Soil Temperature from Numerical Weather Prediction Models

<table>
<thead>
<tr>
<th>City</th>
<th>Temperature</th>
<th>Longitude</th>
<th>Days</th>
<th>Rainfall</th>
<th>Snowfall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newkirk</td>
<td>36.8981</td>
<td>-96.9104</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ardmore</td>
<td>34.2281</td>
<td>-97.2014</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Delaware</td>
<td>36.74374</td>
<td>-95.608</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Norman</td>
<td>35.23611</td>
<td>-97.4649</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oklton</td>
<td>36.03126</td>
<td>-96.4975</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oklahoma City</td>
<td>35.47226</td>
<td>-97.4641</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oklahoma City</td>
<td>35.5556</td>
<td>-97.5107</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oklahoma City</td>
<td>35.47112</td>
<td>-97.5825</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Okemah</td>
<td>35.43172</td>
<td>-96.2627</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pauls Valley</td>
<td>34.7155</td>
<td>-97.2292</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pawnee</td>
<td>36.36114</td>
<td>-96.7699</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Perkins</td>
<td>35.99865</td>
<td>-97.0483</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Clarksville</td>
<td>35.8257</td>
<td>-95.5598</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Adair</td>
<td>36.36914</td>
<td>-95.2714</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Red Rock</td>
<td>36.3559</td>
<td>-97.1531</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willow</td>
<td>35.12275</td>
<td>-99.36</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ringling</td>
<td>34.19365</td>
<td>-97.5881</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Shawnee</td>
<td>35.36492</td>
<td>-96.9482</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Skiatook</td>
<td>36.4153</td>
<td>-96.0371</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Slapout</td>
<td>36.59749</td>
<td>-100.262</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Spencer</td>
<td>35.54208</td>
<td>-97.3415</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stigler</td>
<td>35.26527</td>
<td>-95.1812</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stillwater</td>
<td>36.12093</td>
<td>-97.0953</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stuart</td>
<td>34.87642</td>
<td>-96.0698</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Talihina</td>
<td>34.7107</td>
<td>-95.0115</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tipton</td>
<td>34.43972</td>
<td>-99.1376</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tishomingo</td>
<td>34.33262</td>
<td>-96.679</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Vanoss</td>
<td>34.79146</td>
<td>-96.8438</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Centralia</td>
<td>36.77536</td>
<td>-95.2209</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Walters</td>
<td>34.3647</td>
<td>-98.3203</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Washington</td>
<td>34.98224</td>
<td>-97.5211</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Watonga</td>
<td>35.84185</td>
<td>-98.5262</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Waurika</td>
<td>34.16775</td>
<td>-97.9882</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Weatherford</td>
<td>35.5083</td>
<td>-98.7751</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Westville</td>
<td>36.011</td>
<td>-94.645</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Wister</td>
<td>34.98426</td>
<td>-94.6878</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Woodward</td>
<td>36.42329</td>
<td>-99.4168</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Wynona</td>
<td>36.51806</td>
<td>-96.3422</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Totals for 2009: 56 19 4

816
Appendix B: CMA series approach to temperature profile modeling

Instead of the classical Fourier analysis, the change in temperature with depth is modeled 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 the daily and annual harmonics, to capture harmonics with a period (P) of less than a day to a year: $P = [0.5, 1, 2, 3, 4, 8, 16, 32, 64, 365]$ days. This approach can be summarized as:

$$ T_i^* = \bar{T} + \sum_{n \in P} H_{n,j} e^{i\phi_n} $$  

where

$$ H_{n,j} = T_i - CMA_n(T_i) \quad \text{for} (n = 1), $$  

$$ H_{n,j} = \bar{T}_{n-1,j} - \bar{T}_{n,j} \quad \text{for} (n > 1), $$  

$$ \bar{T}_{n,j} = CMA_n(\bar{T}_{n,j}), $$  

$$ t_n = t - d\phi_n \cdot n/2\pi, $$  

$$ d\phi_n = d\phi/\sqrt{n}, $$

and CMA is the Central Moving Average over the time period $t-n/2:t+n/2$. The variable $d\phi$ is the phase shift of the daily temperature harmonic.
**Figure Captions**

Fig 1 Fig 1 Land cover classification for Oklahoma, according to the 2001 national landcover database for the United States (NLCD 2001, www.mrlc.gov [Homer et al., 2007]).

Fig 2. Theoretical phase shift of the temperature harmonic with a period of a day over a 0.05 m vertical distance. Three general soil types are considered: Sand (wilting point at 0.1 m$^3$/m$^3$, and 92 % quartz particles), Sandy Clay (wilting point at 0.2 m$^3$/m$^3$, and 52 % quartz particles), and clay (wilting point at 0.3 m$^3$/m$^3$, and 25 % quartz). The porosity for all soil types is set at 40 %.

Fig 3. Phase shift of the daily harmonic of T5 (top), phase difference between the T5 and T10 (middle), and RMSE for T5* (bottom).

Fig. 4. Results of phase synchronization method by hour of day. Panel (a) shows the mean diurnal cycle of the in situ measurements (red: T10, and in black: T5) and the modelled T5* in blue. Panel (b) and (c) show how the SEE, ubRMSE, and RMSE vary over the day; in panel b for the case when T10 is compared directly with T5 without any phase synchronization; and in panel (c) when the synchronized T5* and T5 are compared. Note that the stated accuracy of the sensors is 0.5 K.
Fig. 5. Daily cycle of performance metrics for the three phase synchronized temperature products, compared to in situ data at 0.05 m under sod soil, for April through September of 2009. On the top row (a) the mean daily temperature cycle is shown for the in situ data (black), the original NWP temperature (red dash) and the synchronized T5* product (blue). In row (b) the correlation coefficient is shown for the original (red dash) and the synchronized T5* product (blue). In row (c) each graph shows the SEE (grey), the ubRMSE (black), and the RMSE (blue line).

Fig. 6. RMSE at 6 AM local solar time for the temperature products of (from top to bottom) ECMWF, MERRA, and NCEP.

Fig. 7. Same as the third row in Figure 5, but now for April through September of 2004.

Fig. 8. Same as the third row in Figure 5, but now for the full year 2009, excluding temperatures below 273 K.
Soil Temperature from Numerical Weather Prediction Models

### Table 1. Specifications of NWP products.

<table>
<thead>
<tr>
<th>NWP Center / Model</th>
<th>GMAO / MERRA</th>
<th>ECMWF / IFS</th>
<th>NCEP / GDAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tcanopy/T(0-0.05 m) hourly</td>
<td></td>
<td>T(0-0.07 m) 6- hourly</td>
<td>(Tskin + T(0-0.1 m))/2</td>
</tr>
<tr>
<td>Output Interval (UTC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hourly</td>
<td></td>
<td>6- hourly</td>
<td>6- hourly (0z/6z/12z/18z), + 3 hour forecasts</td>
</tr>
<tr>
<td>Spatial Resolution</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004 0.5x0.67 deg</td>
<td>39 km</td>
<td>0.469 deg (T254)</td>
<td></td>
</tr>
<tr>
<td>2009 0.5x0.67 deg</td>
<td>25 km</td>
<td>0.313 deg (T382)</td>
<td></td>
</tr>
<tr>
<td>Regridding to 0.5 deg</td>
<td>Bilinear Interpolation</td>
<td>Linear Average</td>
<td>Bilinear Interpolation</td>
</tr>
<tr>
<td>Name after phase synchronization to Tz</td>
<td>T_ME</td>
<td>T_EC</td>
<td>T_NC</td>
</tr>
</tbody>
</table>

### Table 2. General comparison of the synchronized NWP temperature data sets with the Mesonet in situ temperature.

<table>
<thead>
<tr>
<th>MODEL</th>
<th>T_EC</th>
<th>T_NC</th>
<th>T_ME</th>
</tr>
</thead>
<tbody>
<tr>
<td>dφ (NWP– in situ, minutes)</td>
<td>-52</td>
<td>-93</td>
<td>-175</td>
</tr>
<tr>
<td>Mean Bias (NWP– in situ)</td>
<td>1.3</td>
<td>-0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Bias in Amplitude (NWP/in situ)</td>
<td>1.5</td>
<td>0.9</td>
<td>1.2</td>
</tr>
</tbody>
</table>

### Table 3. Overview of validation results at 6 AM for April 1, to October 1, 2009.

<table>
<thead>
<tr>
<th>MODEL</th>
<th>T_EC</th>
<th>T_NC</th>
<th>T_ME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>Principal Duplicate</td>
<td>Principal Duplicate</td>
<td>Principal Duplicate</td>
</tr>
<tr>
<td>SEE</td>
<td>0.4 ± 0.8 0.3 ± 0.7</td>
<td>-0.4 ± 0.8 -0.5 ± 0.6</td>
<td>0 ± 0.7 0 ± 0.7</td>
</tr>
<tr>
<td>ubRMSE</td>
<td>1.4 1.4</td>
<td>1.4 1.4</td>
<td>1.5 1.5</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.7 1.6</td>
<td>1.8 1.8</td>
<td>1.7 1.8</td>
</tr>
<tr>
<td>RMSE&gt;2K/N</td>
<td>8/56 1/19</td>
<td>14/56 3/19</td>
<td>7/56 3/19</td>
</tr>
</tbody>
</table>

### Table 4. Overview of validation results at 6 AM for the entire year 2009 (with T>273 K).

<table>
<thead>
<tr>
<th>MODEL</th>
<th>T_EC</th>
<th>T_NC</th>
<th>T_ME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.2</td>
<td>-0.3</td>
<td>-0.5</td>
</tr>
<tr>
<td>SEE</td>
<td>1.5</td>
<td>1.8</td>
<td>1.5</td>
</tr>
<tr>
<td>ubRMSE</td>
<td>1.7</td>
<td>1.9</td>
<td>1.8</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.9</td>
<td>2.1</td>
<td>1.9</td>
</tr>
<tr>
<td>RMSE&gt;2K/N</td>
<td>11/56 32/56</td>
<td>19/56</td>
<td></td>
</tr>
</tbody>
</table>
Soil Temperature from Numerical Weather Prediction Models

Fig 1. Land cover classification for Oklahoma, according to the 2001 national landcover database for the United States (NLCD 2001, www.mrlc.gov [Homer et al., 2007]).

Fig 2. Theoretical phase shift of the temperature harmonic with a period of a day over a 0.05 m vertical distance. Three general soil types are considered: Sand (wilting point at 0.1 m$^3$/m$^3$, and 92 % quartz particles), Sandy Clay (wilting point at 0.2 m$^3$/m$^3$, and 52 % quartz particles), and clay (wilting point at 0.3 m$^3$/m$^3$, and 25 % quartz). The porosity for all soil types is set at 40 %.
Soil Temperature from Numerical Weather Prediction Models

Fig 3. Phase shift of the daily harmonic of T_s (top), phase difference between the T_s and T_{10} (middle), and RMSE for T_{s*} (bottom).

Fig 4. Results of phase synchronization method by hour of day. Panel (a) shows the mean diurnal cycle of the in situ measurements (red: T_{10} and in black: T_s) and the modelled T_{s*} in blue. Panel (b) and (c) show how the SEE, ubRMSE, and RMSE vary over the day; in panel b for the case when T_{10} is compared directly with T_s without any phase synchronization; and in panel (c) when the synchronized T_{s*} and T_s are compared. Note that the stated accuracy of the sensors is 0.5 K.
Soil Temperature from Numerical Weather Prediction Models

Fig. 5. Daily cycle of performance metrics for the three phase synchronized temperature products, compared to in situ data at 0.05 m under sod soil, for April through September of 2009. On the top row (a) the mean daily temperature cycle is shown for the in situ data (black x), the original NWP temperature (red dash) and the synchronized T5* product (blue). In row (b) the correlation coefficient is shown for the original (red dash) and the synchronized T5* product (blue). In row (c) each graph shows the SEE (grey), the ubRMSE (black), and the RMSE (blue line).
Soil Temperature from Numerical Weather Prediction Models

Fig. 6. RMSE at 6 AM local solar time for the temperature products of (from top to bottom) ECMWF, MERRA, and NCEP.

Fig. 7. Same as row (c) in Figure 5, but now for April through September of 2004.

Fig. 8. Same as row (c) in Figure 5, but now for the full year 2009, excluding temperatures below 273 K.