Soil Moisture Active Passive (SMAP)

Ancillary Data Report

Surface Temperature

Version 2
SMAP Science Document no. 051

Contributors:
SMAP Algorithm Development Team
SMAP Science Team

March 3, 2015
JPL D-53064

Jet Propulsion Laboratory
California Institute of Technology

Preface

The SMAP ancillary data reports provide descriptions of ancillary data sets used with the science algorithm software in generation of SMAP science data products. The Ancillary Data Reports are updated as new data or processing methods become available. Current versions of the ancillary data reports are available along with the Algorithm Theoretical Basis Documents (ATBDs) via the SMAP web site at http://smap.jpl.nasa.gov/documents/.

List of Acronyms

ATBD	Algorithm Theoretical Basis Document
ECMWF	European Centre for Medium-range Weather Forecasts
ESA	European Space Agency
FLUXNET	Flux Network (Micrometeorological Tower Sites)
GDAS	Global Data Assimilation System
GEOS-5	Goddard Earth Observing System Model, Version 5
GES DISC	Goddard Earth Sciences Data and Information Services Center
GFS	Global Forecast System
GMAO	Goddard Modeling and Assimilation Office
IFS	Integrated Forecast Center
IGBP	International Geosphere Biosphere Program
LST	Land Surface Temperature
MERRA	Modern Era Retrospective-analysis for Research and Applications
NCCS	NASA Center for Climate Simulation
NCEP	National Centers for Atmospheric Prediction
NOAA	National Oceanic and Atmospheric Administration
NRCS	Natural Resources Conservation Service
NWP	Numerical Weather Prediction
SCAN	Soil Climate Analysis Network
SDS	Science Data System
SMAP	Soil Moisture Active Passive
TOPMODEL	Topography-based hydrological Model
USDA	United States Department of Agriculture

Contents

1. Introduction

Surface temperature information is needed by the SMAP passive soil moisture retrieval algorithms to determine the soil surface emissivity from brightness temperature (T_B) observations. The emissivity is then used to estimate the soil moisture. This procedure is described in the Algorithm Theoretical Basis Document (ATBD) for Level 2 & 3 Soil Moisture (Passive) Data Products[1].

Since the SMAP observations typically view mixed scenes containing components of bare soil, vegetation and open water, the physical temperature of each of these components needs to be considered. First, a water temperature T_w is needed to estimate the brightness temperature of the open water fraction within the footprint. This enables a "water-corrected" or land-only brightness temperature (T_B) to be computed for the land-fraction part of the footprint. The land surface emissivity can then be determined by dividing this T_B by the composite or footprint-averaged land surface temperature (including soil and vegetation components of the scene). The vegetation and soil temperatures are assumed to be equal in the SMAP soil moisture retrieval algorithm, which is a reasonably good assumption for the half-orbits with 6:00 am (descending) Equator crossing, so that the terms soil temperature and land surface temperature can be used interchangeably and referred to as T_s. Furthermore, T_s as used by the soil moisture retrieval algorithms is considered to be an average or effective temperature of the top 5 cm layer of the soil.

1.1 Purpose

The purpose of this report is to document the evaluation of a number of candidate ancillary datasets that were used to select the best dataset to use for deriving T_s values that best represent globally what the SMAP soil moisture retrieval algorithms require. An assessment of data quality is provided along with a description of the selected dataset.

Three primary sources of ancillary surface temperature data for SMAP can be considered: (1) in situ monitoring stations; (2) other (non-SMAP) satellite remote sensing observations; and, (3) land surface models including data assimilation and re-analyses. The in situ source is not preferred because SMAP is a global mission, and in situ stations provide sparse and spatially inconsistent coverage. Satellite observation sources are not preferred due to difficulties in obtaining observations at the same overpass time as SMAP and incorporating them into the data processing stream, and also due to errors associated with T_s retrieval (e.g., cloud/atmospheric contamination). Therefore, the assessment presented here focused on evaluating operational land surface model surface temperature products and selecting the most appropriate product as an ancillary dataset for SMAP.

The Surface Temperature dataset is one of a suite of ancillary datasets required by the SMAP L1-L3 science data processing algorithms. The complete list of datasets is provided in Appendix A. Ancillary data required by the L4 processing algorithms (data assimilation products) are described in the L4 ATBDs.

1.2 Scope and Objectives

A comparison was performed of in situ observations and surface temperature outputs from three land surface models: (1) the Global Data Assimilation System (GDAS) and Global Forecast System (GFS) produced by the U. S. National Ocean and Atmospheric Administration (NOAA) Centers for

[1] O'Neill, P. et al., Algorithm Theoretical Basis Document (ATBD), Level 2 & 3 Soil Moisture (Passive) Data Products, http://smap.jpl.nasa.gov/documents/.

Environmental Prediction (NCEP); (2) the Modern Era Retrospective-analysis for Research and Applications (MERRA) product produced by the U. S. National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO); and (3) the Integrated Forecasting System (IFS) produced by the European Centre for Medium-range Weather Forecasts (ECMWF). The objectives are two-fold:

1. Determine whether the surface temperature outputs from the three models above differ significantly between the models, and if so, where and when (spatial and temporal patterns of the differences)

2. Compare and validate the model outputs using in situ measurements, and perform an error analysis.

2. Parameter Description and Requirements

Attributes of the T_s data required by the SMAP algorithms, to be supplied from model output data with pre-processing as needed, are:

- represents the mean soil temperature (Kelvins) of the top 5 cm of the soil column
- represents soil conditions at 06:00 am (and 06:00 pm) local solar time
- available on the global 9 km and 36 km SMAP EASE grids
- available with latency required by the SMAP SDS processing system
- includes all regions and seasons globally

The SMAP algorithms that use T_s are:

- Level 2 soil moisture passive, L2_SM_P (36 km EASE-grid, 24 hour latency)
- Level 2 soil moisture active-passive, L2_SM_AP (9 km EASE-grid, 24 hour latency)

A description of the SMAP products including grid resolutions and latencies, and product algorithms, is available in the SMAP Handbook (http://smap.jpl.nasa.gov/mission/description).

3. Evaluation of Source Data

Much of this section builds on earlier work by Holmes et al. [2012] who compared model output surface temperature data with in situ data collected over Oklahoma in the U.S. This analysis seeks to extend the approach to a more globally distributed set of in situ data, with results more representative of global conditions. Table 1 provides a summary of the surface temperature-related variable attributes provided by the three models evaluated in this report.

Table 1. Summary of source models and temperature variables relevant to SMAP.

	NWP Center / Model		
	GMAO / MERRA	**ECMWF / IFS**	**NCEP / GDAS**
Parameters	T (skin) T (0-10 cm)	T (skin) T (0-7 cm)	T (skin) T (0-10 cm)
Output Interval (UTC)	Hourly average, centered on the half-hour	6-hourly (0z/6z/12z/18z)	6-hourly (0z/6z/12z/18z) + 3 hour forecasts
Spatial Resolution (2009 data)*	0.5 x 0.67 deg	25 km	0.313 deg (T382)

* The data sets were interpolated to 1° x 1° for subsequent analysis and intercomparison.

3.1 Error Sources

In assessing modeled and in situ surface temperature there are four main sources of error: (1) in situ sensor error; (2) upscaling error; (3) depth correction error; and (4) model error. Sensor error may be due to calibration error (both instantaneous and over time), uncertainty in depth of measurement, and sensor disturbance. Assessments for theoretically implausible phase, and significant shifts in record at specific times, can help in identifying problematic in situ sites for exclusion. Upscaling error includes error in the assumption that an observed value at a point can be used to represent an average over a larger pixel area, as well as error of any remote sensing data used in an upscaling method. Depth correction error involves the shifting or extrapolation of soil temperature values from the model or measurement, which are provided for a given vertical depth, to values representing a different vertical depth. Finally, model error is what we aim to quantify. More detail on the error sources is provided in the sections below.

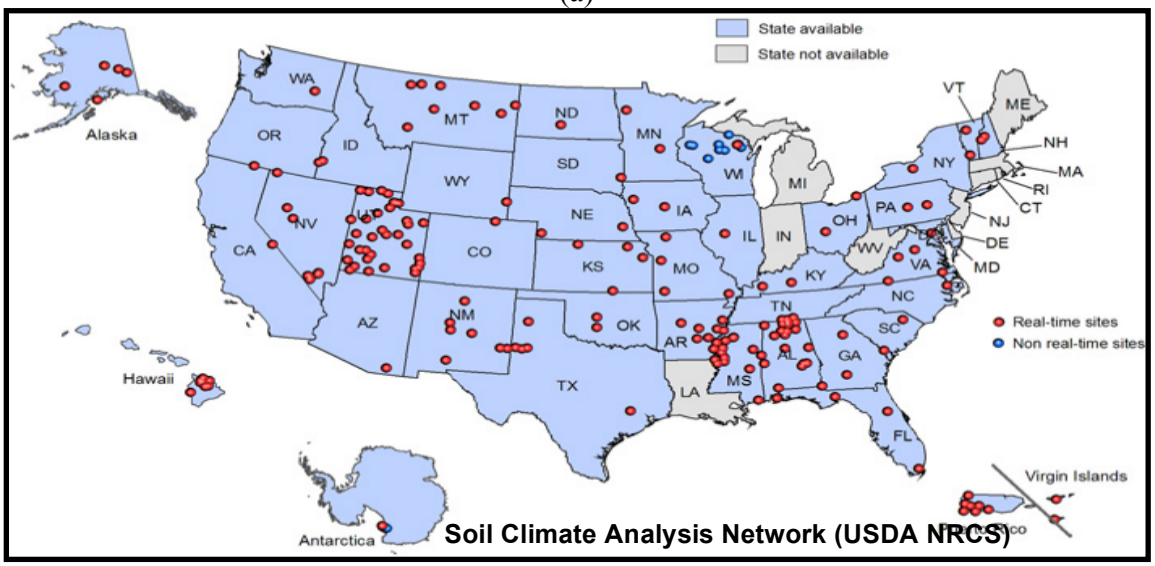
3.2 In Situ Data

In situ soil temperature (T_s) data at 5 cm depth were compiled for analysis from: (1) 198 sites located within the U.S. that are provided by the U.S. Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS) Soil Climate and Analysis Network (SCAN)[2], and (2) 61 sites within the FLUXNET tower network[3] that includes CarboEurope (34 sites), AmeriFlux (24 sites), Canada-Fluxnet (2 sites), and CarboAfrica (1 site) [*Baldocchi et al.*, 2001]. The SCAN and FLUXNET sites are shown in Figure 1. Of these, a subset of 129 sites was selected based on requirements for the depth correction procedure (section 3.4), a further 64 of which were selected as suitable for the upscaling procedure (section 3.3). The selected sites represent most land cover types as defined by the International Geosphere-Biosphere Programme (IGBP) and most climates, though

[2] http://www.wcc.nrcs.usda.gov/scan/
[3] http://www.fluxdata.org

with a few noteworthy absences in hot and wet areas (e.g., tropical rainforests), hot and dry areas (e.g., deserts), and very cold regions (e.g., tundra and ice) (Figure 2). Further work should include in situ data for deserts.

(a)



(b)

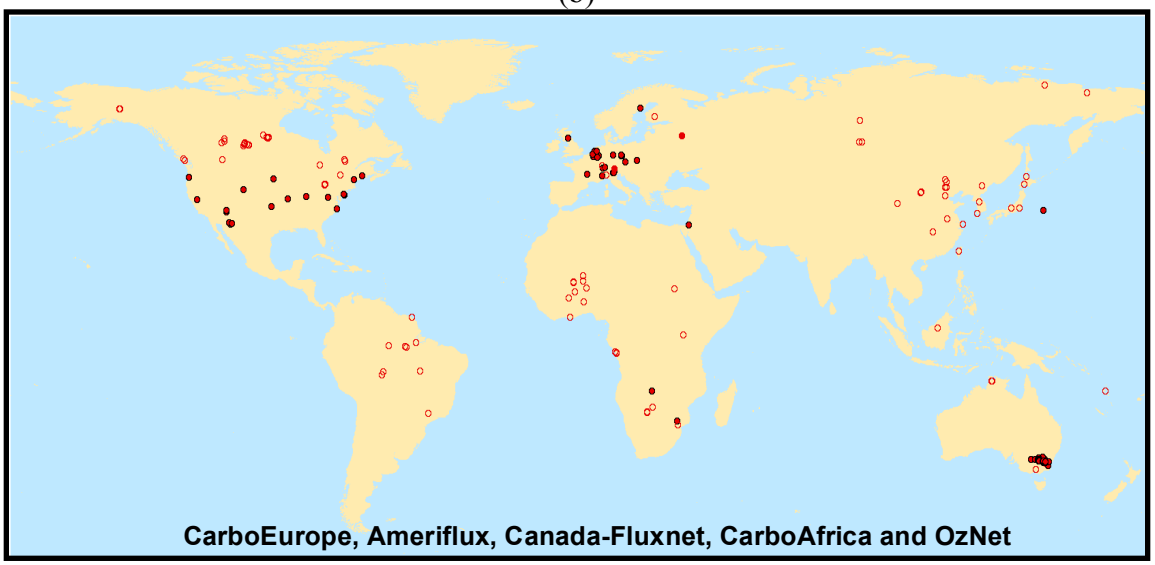


Figure 1. Sites used for in situ measurements of soil temperature:
(a) SCAN (http://www.wcc.nrcs.usda.gov/scan/SCAN_brochure.pdf)
(b) FLUXNET (http://www.fluxdata.org)

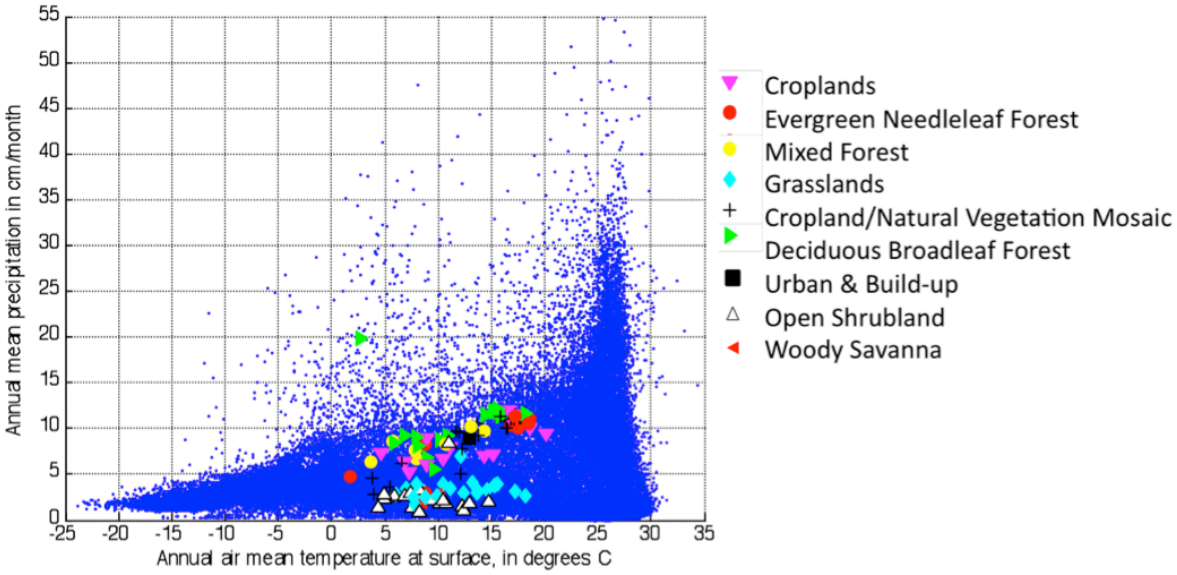


Figure 2. Climate and land cover representation of ground sites where in situ observations were collected, shown as symbols, on top of blue points representing all land points (from MERRA) in precipitation and temperature space.

3.3 Upscaling

In situ measurements may be representative of the surrounding conditions if the soil and land cover properties are spatially uniform (e.g., FLUXNET sites are often relatively homogeneous for up to 1 km^2). In this analysis comparisons were made of in situ data with model output data for a grid cell, which ranges from 25 km^2 (ECMWF) to 0.5x0.67 degrees (MERRA). A direct point-to-pixel comparison contains representation error in that the model pixel averages together all land conditions within the pixel, but the in situ site may not be representative of the average of the larger area. For example, a point measurement made at an oasis in a desert would have a dramatically different characteristic signature than the larger model pixel that includes primarily desert. Nonetheless, visual examination in Google Earth for each site showed that such extreme case do not occur in our dataset, and rather most sites are visually and qualitatively "similar" to the surrounding areas on average. However, no upscaling was applied in the analysis beyond this inspection.

3.4 Depth Correction

The soil layer depth assumed for T_s varied among the models, as well as the in situ validation data. The surface layer T_s depth is 10 cm for MERRA and ECMWF, 7 cm for NCEP, and 5 cm for the in situ measurements. Hence, a normalization routine was necessary to make the depth of T_s comparable across models and with the in situ measurements. This involves a depth-correction that leverages temporal phase information to ensure all model estimates and in situ measurements are in phase with one another. The depth correction procedure applied here is based on the earlier work of *Holmes et al.* [2012] who used in situ sites in Oklahoma, USA. The depth correction procedure used in this study was extended to a more globally distributed set of in situ sites. The following is excerpted from the *Holmes et al.* [2012] paper to describe the theory and implementation of the depth correction approach (Box 1):

Box 1 (*from Holmes et al., 2012*)

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 e-folding damping depth (z_D):

$$d\varphi = \frac{-dz}{z_D} \tag{4}$$

$$A_{z2} = A_{z1} e\left(\frac{dz}{z_D}\right) \tag{5}$$

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, m^2/s), and indicates the distance (z_D) over which the amplitude of the wave is reduced by 63%to 1/e of its original value:

$$z_D = \sqrt{\frac{2a}{2\pi f}} \tag{6}$$

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. 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.

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 apply the heat flow principles 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. 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 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:

$$T_5^* = f(T_{10}, d\varphi) \quad (7)$$

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 with a maximum at noon:

$$T_{sim} = \overline{T} + A\sin((t-\varphi)2\pi - \pi/2) \quad (8)$$

where $\overline{T}$ and A are the amplitude of the diurnal cycle. 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 plots that can cause an apparent damping of the temperature harmonics.

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. For each station the RMSE was then calculated between T_5^* and T_5. 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.

To determine how to apply the depth correction procedure globally using a method that has only been applied to sites in Oklahoma, there were three options for how to synchronize phases: (1) use the same phase differences found in Oklahoma, (2) determine the phase locally for each site, and fit for each model, and (3) determine the average global phase shift using the global database of validation sites. The problem with option (1) is that sites in Oklahoma do not represent the global diversity of ecosystems and climates. The problem with option (2) is that the phase would need to be calculated operationally for each site, but the model phase dependency on vegetation would be lost. Therefore option (3) was followed. The spatial variation in phase within each model was first qualitatively observed, which was relatively stable for each model though with differences between models both in time and space. The model differences in phase are related to how the models couple moisture and temperature with land cover. Each model's phase was shifted by the difference between the average phase of the models and the in situ data. This removed structural (i.e., depth) phase difference between the models without modifying the spatial phase variation inherent to each model. In preparing the globally corrected data a normal moving average was used instead of a central moving average, as the latter was not possible operationally, only in reanalysis.

3.5 Model vs. In Situ Data

When the models were compared against in situ data without depth correction all three models showed RMSEs of less than 3.6 K, with an average RMSE across all three models of 3.2 K (Figure 3). NCEP had the lowest RMSE at 2.98 K. Negative temperature values displayed noticeable "flooring" in that the linear trend exhibited by positive temperatures was lost at negative values. It was unclear whether the problem was in the models or in the in situ data, as sometimes one showed realistic values while the other did not, and vice versa. Removal of negative values, as was done in *Holmes et al.* [2012], led to further reduction in RMSE in NCEP (to 2.79 K) and MERRA (from 3.56 to 3.43 K), but increased RMSE in ECMWF (from 3.14 to 3.17 K). The data were skewed towards over-representation by some land cover types, so a random selection was made from over-represented sites to equilibrate with other land covers, producing an equal IGBP representation

analysis. This led to significant RMSE reduction in all models, with NCEP at 2.67 K, MERRA at 3.39 K, and ECMWF at 3.00 K.

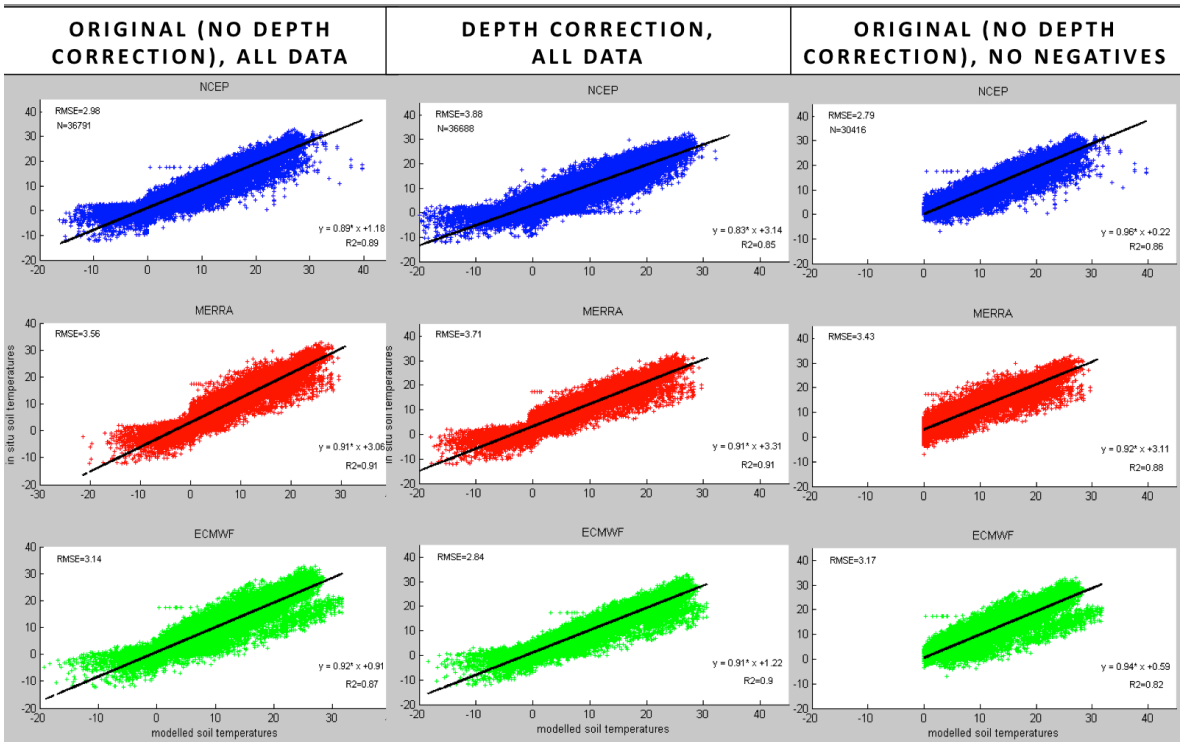


Figure 3. Scatterplots of model (x-axis) versus in situ (y-axis) soil temperature (C). The left column shows all data without depth correction. The middle column shows all data with depth correction. The right column shows all data without depth correction and without negative values.

Depth correction improved the RMSE of only ECMWF, bringing the RMSE down to between 2.77–2.90 K (all sites: 2.84; excl. negatives: 2.82; equal IGBP: 2.77; equal IGBP, excl. negatives: 2.90). RMSE's with depth correction were larger for NCEP (3.35–3.88 K) and MERRA (3.51–3.71 K), with NCEP gaining the most error due to depth correction. Thus the improvements gained by using the depth correction procedure were mixed. It should be noted that "best" errors for all models were in the range of 2.7–3.4 K, close to the 2 K target for SMAP, and these errors still include upscaling error and in situ measurement error. Further analysis of RMSE and bias by biome, climate space, latitude and season was not performed as part of this study.

4. Selection of Source Dataset

4.1 Overview

To select the primary source data set (ECMWF, NCEP, or MERRA) for generating the T_s ancillary data required by the SMAP algorithms, the analysis results in Section 3 were considered along with the following criteria for each model:

1. Compatibility with SMAP ancillary data requirements for T_s discussed in Section 2 (e.g., spatial resolution, time-of-day, etc.)

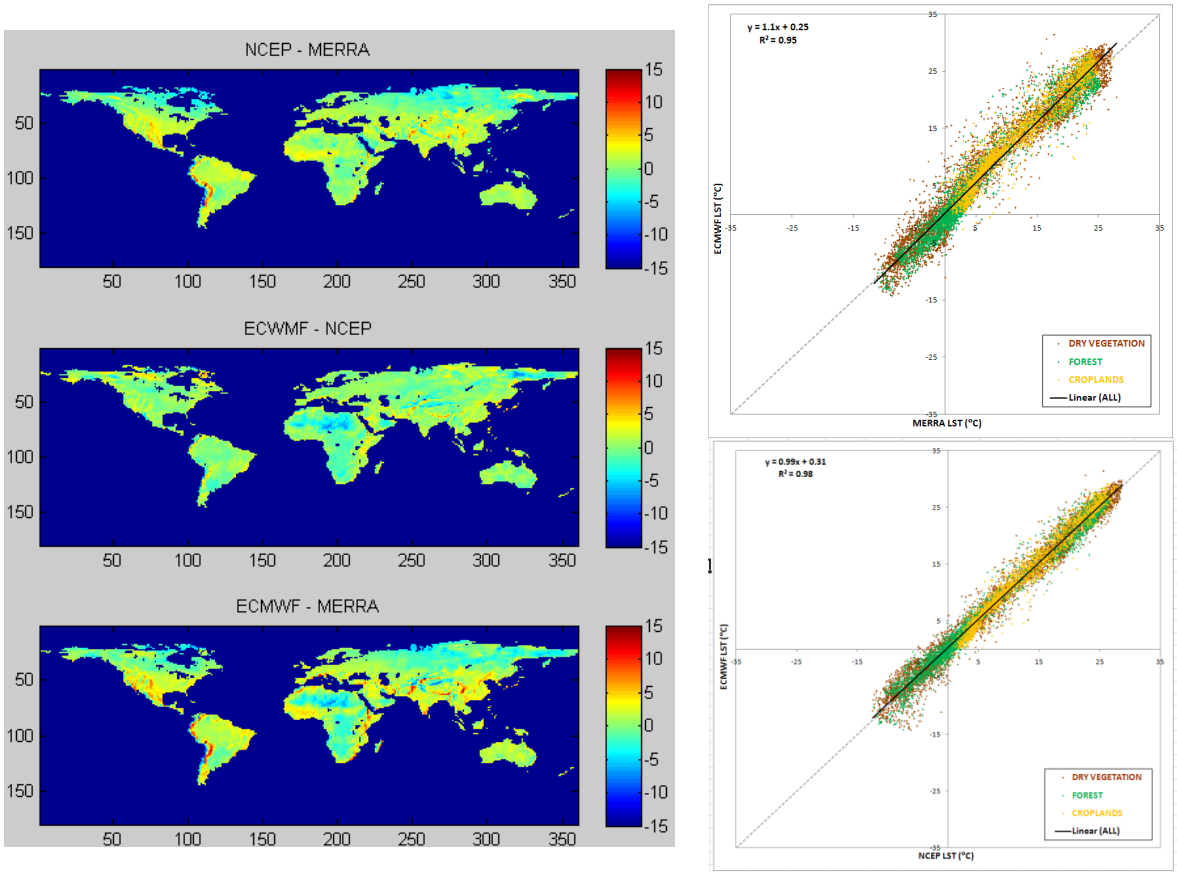
2. Accuracy and performance against in situ measurements (analysis of Section 3)

3. Availability, including latency, data loss, etc.

4. Responsiveness of the source data production team to SMAP needs

5. Public availability and permission to redistribute ancillary data along with the SMAP data

Based on the above criteria, the MERRA data set was the prime candidate for selection. Since MERRA is a reanalysis data set, the actual source data set was selected to be the NASA Goddard Earth Observing System Model, Version 5 (GEOS-5) forward processing (FP) system. GEOS-5 is a near-real time atmospheric modeling and assimilation system [Lucchesi, 2013] that is of the same lineage as MERRA. In the analysis of Section 3, MERRA showed slightly higher RMSEs than both NCEP and ECMWF though all three models showed similar values (Figures 3, 4). Therefore, the choice of model based on RMSE alone is somewhat arbitrary, and other logistical factors have higher weight in the selection decision. Since GEOS-5 is NASA-based, and the GEOS-5 land model developers are part of the SMAP team, the GEOS-5 model output products are likely to be the most suitable and responsive to SMAP needs. GEOS-5 provides data on an hourly time-step rather than the 6-hourly ECMWF and NCEP data, so less processing is needed to temporally interpolate. GEOS-5 currently provides data at 0.25 x 0.3125 degree spatial resolution, which is similar to the SMAP resolution. Finally, GEOS-5 ancillary data can be freely redistributed with the SMAP data.



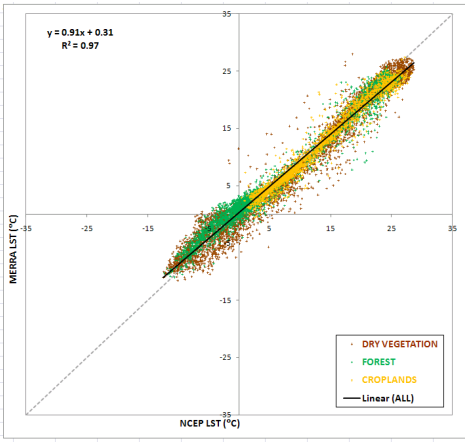


Figure 4. Mean global average soil temperature differences between models.

4.2 GEOS-5 Data Description

Below are details regarding the shared lineage between the MERRA and GEOS-5 data assimilation systems. The processing steps taken by SMAP on GEOS-5 outputs for subsequent soil moisture data processing are explained later in Section 6.

- MERRA is the NASA/GMAO global atmospheric reanalysis [Rienecker et al., 2008, 2011; Trenberth et al., 2009]. This dataset covers the time period from 1979 through to the present day, with applications for both climate and meteorological studies, and includes a number of data fields such as cloud, ozone, surface turbulent fluxes and land surface diagnostics [Rienecker et al., 2008; Trenberth et al., 2009]. The reanalysis is produced using version GEOS-5.2.0 of the Goddard Earth Observing Data Assimilation System. This system assimilates approximately 4 million observations every 6 hours [Lucchesi, 2012], using an incremental analysis update (IAU) procedure [Bloom et al., 1996; Trenberth et al., 2009]. This approach attempts to combine the benefits of both intermittent and continuous data assimilation, by using the analysis increments continuously to force the GCM [Bloom et al., 1996]. The main characteristic of MERRA is its focus on the hydrological cycle. This feature of the reanalysis is reflected in its use of the Catchment Land Surface Model (CLSM) [Koster et al., 2000] to derive four land surface temperature dynamics. In this model, the land surface is classified horizontally into categories (saturated, unsaturated, wilted and snow-covered zone), based on modeled root zone moisture distribution derived from a topography-based hydrological model [TOPMODEL, Beven and Kirkby, 1979].

- The MERRA data used in this analysis come from the 'Land-related surface quantities' collection ("lnd" collection or "tavg1_2d_lnd_Nx"). In this collection, four MERRA surface temperature parameters are needed to derive global land surface (skin) temperature together with their fractional cover. These parameters are: the top snow layer, the unsaturated zone, the saturated zone and the wilted zone surface temperatures ('TPSNOW', 'TUNST', 'TSAT' and 'TWLT' respectively), and their respective fractional areas ('FRSNO', 'FRUNST', 'FRSAT' and 'FRWLT'; Lucchesi, 2012; Reichle, 2012). By calculating the area-weighted average of each surface temperature from the four categories, we derived the land surface temperature for each pixel, as shown in the following equation: TSKIN = TPSNOW * FRSNO + TUNST * FRUNST + TSAT * FRSAT + TWLT * FRWLT

- Holmes et al. (2012) also used the top-layer soil temperature (TSOIL1) provided by GMAO using an off-line replay of MERRA. For all vegetation classes except "broadleaf evergreen", TSOIL represents the soil temperature in the 0-10 cm layer. For "broadleaf evergreen" land cover (which is of minimal relevance to SMAP), the model surface (skin) temperature includes the temperature in the 0-5 cm layer, and TSOIL1 thus represents the temperature in the 5-15 cm soil layer. Holmes et al. [2012] (see also Box 2 above) report results for TSKIN, TSOIL (=TSOIL1) and TAVG, which is the average of TSKIN and TSOIL1 (and thus represents most closely the temperature in the 0-5 cm layer). For details see Reichle [2012] and Holmes et al. [2012].

- The operational SMAP data products will use ancillary soil temperature derived from the GEOS-5 forward processing NWP system [Lucchesi, 2013]. As of February 2015, the GEOS-5 FP system is using version 5.13 of the atmospheric modeling and data assimilation system. Output is provided hourly on a grid with 0.25 degree latitude spacing and 0.3125 degree longitude spacing. While the 5.13 system includes many enhancements since the 5.2.0 version of MERRA, the data format and variable definitions are nearly identical to those of MERRA.

- GEOS-5 FP spatial and temporal resolution is currently 0.25 x 0.3125 degrees, hourly

- GEOS-5 FP spatial coverage: global

- GEOS-5 FP years of available data: version GEOS-5.11: Jun 11, 2013-Aug 20, 2014; GEOS-5.13: Aug 20, 2014-present.

- GEOS-5 FP latency: NWP near-real time (< 12 hours); medium-range forecast data are also available.

- GEOS-5 FP accuracy: <3.4 K (based on MERRA data).

4.3 Metadata

- unit of measurement: Kelvin (K)
- range of measurement: 0 to 340 K
- projection: lat/lon
- spatial resolution: 0.25 x 0.3125 degrees
- temporal resolution: hourly averages, centered at 00:30, 01:30, 02:30 GMT
- spatial extent: global
- start date time: near real-time
- end data time: near real-time
- number of bands: not applicable
- data type: float
- min value: 0
- max value: 340
- no data value: 1.0e15

4.4 Quality Control

- bad data values: 1.0e15
- missing data values: 1.0e15
- flags: not applicable

5. Mask/Flag Derivation

(Not applicable)

6. Data Processing

6.1 Customized GEOS-5 Collection

The GEOS-5 data are delivered regularly to the JPL SMAP SDS in the form of a customized GEOS-5 "smp" (SMAP) collection [Reichle, 2012]. Data are available in near-real time (with ~10 hours latency from the time for which the assimilation data are valid) in hourly intervals. In addition to the "assimilation" (analysis) data, medium-range "forecast" data are also delivered so that processing and dissemination of SMAP products are not delayed due to dependency on T_s data availability. Unless otherwise stated, only the analysis fields are used in SMAP surface temperature processing.

6.2 Preprocessing Algorithm and Re-gridding

The required SMAP surface temperature (T_s) parameter (effective soil temperature) is obtained from the GEOS-5 data by taking the arithmetic mean of the two parameters TSURF (skin temperature) and TSOIL1 (temperature of the 0-10 cm layer) at their native 0.25 x 0.3125-deg grids. This averaged value most closely represents the temperature in the 0-5 cm layer as supported by the analysis of *Holmes et al (2012)*. Only values of these parameters over land grid cells are used by SMAP Level 2 soil moisture algorithms. Water temperature values T_w (see Section 1, Introduction) are taken directly from the TS parameter without modification and used for water fraction T_B correction.

The above processing results in a composite temperature array on a 0.25 x 0.3125-deg grid. Two-dimensional interpolation (bilinear interpolation) is then applied to create the required surface temperature fields on 3-, 9-, and 36-km EASEv2 projections.

7. Acknowledgements

This work was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Contributions to this report by Josh Fisher of the Jet Propulsion Laboratory (JPL) and Thomas Holmes of the U.S. Dept. of Agriculture, are gratefully acknowledged. Irene Garonna of the Environmental Change Institute, University of Oxford, UK performed much of the data processing and analysis for this study while visiting JPL as an exchange student.

Thanks are expressed to the NASA/GSFC Global Modeling and Assimilation Office (GMAO) and GES DISC for providing the MERRA data, to the European Centre for Medium-range Weather Forecasting (ECMWF) for the IFS data, and to the National Centers for Environmental Prediction (NCEP) for the GDAS data.

In situ data were provided courtesy of the CarboEurope PIs for FLUXNET data, the USDA NRCS for SCAN data, and the International Soil Moisture Network (ISMN).

8. References

Baldocchi, D., E. Falge, L. H. Gu, R. J. Olson, D. Hollinger, S. W. Running, P. M. Anthoni, C. Bernhofer, K. Davis, R. Evans, J. Fuentes, A. Goldstein, G. Katul, B. E. Law, X. H. Lee, Y. Malhi, T. Meyers, W. Munger, W. Oechel, K. T. P. U, K. Pilegaard, H. P. Schmid, R. Valentini, S. Verma, T. Vesala, K. Wilson, and S. C. Wofsy (2001), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, Bulletin of the American Meteorological Society, 82(11), 2415-2434.

Beven, K. J., and M. J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology, Hydrological Sciences Journal, 24(1), 43-69.

Bloom, S. C., L. L. Takacs, A. M. Da Silva, and D. Ledvina (1996), Data assimilation using incremental analysis updates, Monthly Weather Review, 124, 1256-1271.

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

Jung, M., M. Reichstein, and A. Bondeau (2009), Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model, Biogeosciences, 6, 2001-2013.

Koster, R. D., M. J. Suárez, A. Ducharne, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a GCM, Part 1, Model Structure, Journal of Geophysical Research, 105, 24809-24822.

Lucchesi, R. (2012), File Specification for MERRA Products, *NASA GMAO Office Note, No. 1 (Version 2.3)*, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 87pp. Available at https://gmao.gsfc.nasa.gov/pubs.

Lucchesi, R. (2013), File Specification for GEOS-5 FP (Forward Processing), *NASA GMAO Office Note, No. 4 (Version 1.0)*, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 63pp. Available at https://gmao.gsfc.nasa.gov/pubs.

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, Journal of Atmospheric Sciences, 55, 1209-1224.

Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Toure (2011), Assessment and enhancement of MERRA land surface hydrology estimates, *Journal of Climate,* 24, 6322-6338, doi:10.1175/JCLI-D-10-05033.1.

Reichle, R. H. (2012), The MERRA-Land Data Product, *NASA GMAO Office Note, No. 3 (Version 1.2)*, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 38pp. Available at https://gmao.gsfc.nasa.gov/pubs.

Reichle, R. (2012), GEOS-5 Ancillary Data for the SMAP Project, Data Delivery Memorandum to the SMAP Project.

Rienecker, M. M., and Coauthors (2011), MERRA - NASA's Modern-Era Retrospective Analysis for Research and Applications, Journal of Climate, 24, 3624-3648, doi:10.1175/JCLI-D-11-00015.1.

Rienecker, M. M., M. J. Suarez, R. Todling, J. Bacmeister, L. Takacs, H.-C. Liu, W. Gu, and M. Sienkiewicz (2008), The GEOS-5 Data Assimilation System – Documentation of Versions 5.0.1, 5.1.0, and 5.2.0Rep., NASA.

Seyfried, M. S., L. E. Grant, E. Du, and K. Humes (2005), Dielectric loss and calibration of the hydra probe soil water sensor, Vadose Zone Journal, 4, 1070-1079.

Trenberth, K. E., R. Dole, Y. Xue, K. Onogi, D. Dee, M. Balmaseda, M. Bosilovich, S. Schubert, and W. Large (2009), Atmospheric reanalyses: A major resource for ocean product development and modelingRep.

Van Wijk, W. R., and D. A. de Vries (1963), Periodic temperature variations in a homogeneous soil, in Physics of the Plant Environment, edited, North-Holland Publ. Co., Amsterdam.

Appendix A: Ancillary Datasets Used in SMAP L1-L3 Processing

	Dataset Name	Static/Dynamic	Report Number	
1	Soil Attributes	Static	JPL D-53058	
2	DEM	Static	JPL D-53056	
3	Landcover Classification	Static[1]	JPL D-53057	
4	Crop Type	Static[1]	JPL D-53054	
5	Urban Area	Static	JPL D-53060	
6	Static Water Fraction	Static	JPL D-53059	
7	Surface Temperature	Dynamic[2]	JPL D-53064	
8	Vegetation Water Content	Dynamic[3]	JPL D-53061	
9	Vegetation & Roughness Parameters	Static	JPL D-53065	
10	Permanent Ice	Static	JPL D-53062	
11	Snow	Dynamic[4]	GSFC-SMAP-007	
12	Precipitation	Dynamic[2]	JPL D-53063	

[1] Updated yearly
[2] Updated six-hourly
[3] Updated monthly
[4] Updated daily