National Aeronautics and Space Administration

Goddard Earth Sciences Data and Information Services Center (GES DISC)

README Document for SMERGE Root-zone Soil Moisture Data Product Version 2.0

Last Revised March 01, 2021

Goddard Earth Sciences Data and Information Services Center (GES DISC)
http://disc.gsfc.nasa.gov
NASA Goddard Space Flight Center
Code 610.2
Greenbelt, MD 20771 USA

Prepared By:

Hualan Rui	Jianzhi Dong
Name	Name
GES DISC	USDA ARS Hydrology and
GSFC Code 610.2	Remote Sensing Laboratory

11/27/2018

Date

Reviewed By:

Bill Teng	09/11/2019
Reviewer Name	Date
GES DISC	
GSFC Code 610.2	

Goddard Space Flight Center
Greenbelt, Maryland

Revision History

Revision Date	***Changes***	***Author***
2018-11-08	Add “What are the differences between SMERGE V1.0 and SMERGE V2.0”	Bill Teng
2018-11-15	Add “Known issues regarding the new CCI anomaly variable”	Jianzhi Dong
2019-09-11	Update the temporal coverage related content	Hualan Rui
2020-04-03	Update the reference citation for (Tobin et al. 2019)	Hualan Rui
2021-03-01	Update the email address of GES DISC Help Desk	Hualan Rui

Table of Contents

Contents

1.0 Introduction

The SoilMERGE (SMERGE) product combines long-term (January 1979 – May 2019) satellite-based soil moisture retrievals with land surface model estimates acquired from Phase-2 of the North American Land Data Assimilation System (NLDAS-2) to produce a 0.125-degree, daily, root-zone soil moisture (RZMS) product within the conterminous United States (CONUS). This document contains basic information on the use of SMERGE RZSM for large-scale climate and hydrological studies.

1.1 Data Set Summary Characteristics

Table 1. Basic characteristics of the SMERGE Root Zone (0-40 cm) Soil Moisture Data Product.

Contents	Outputs from SMERGE 0-40 cm soil moisture
Format	netCDF
Spatial extent	25^0 to 53^0 latitude, -125^0 to -67^0 longitude;
Spatial resolution	All: 0.125^0;
Temporal resolution	Daily
Temporal coverage	Jan 2, 1979 – May 10, 2019;
Dimension	All: 224 (lat) x 464 (lon)
Origin (1^{st} grid center)	All: (52.9375, -124.9375)

1.2 Data Set Algorithm Description

As noted above, SMERGE is based on combining satellite and land surface model (LSM) soil moisture (RZSM) estimates to produce an improved analysis of root-zone (0-40 cm) soil moisture. The key focus of SMERGE is on the reduction of anomaly, and not climatology, errors. Hence, RZSM anomaly estimates (obtained via the merger of satellite and modelling information) are directly added back onto an LSM-based climatology to derive a final RZSM product (see Figure 1).

To start, satellite-based soil moisture retrievals were taken from the European Space Agency (ESA) Climate Change Initiative (CCI) 0.25-degree combined (active/passive) dataset (Liu et al. 2012). This 0.25-degree product was linearly downscaled into 0.125-degree spatial resolution soil moisture. Likewise, the 0.125-degree soil moisture estimates obtained from NLDAS-2 Noah simulations (NLDAS_NOAH0125_H.002) for the top 10 cm and 40 cm to provide model-based estimates of surface and root zone soil moisture, respectively.

Next, CCI-based RZSM anomalies were calculated using CCI surface soil moisture anomalies on a grid-by-grid base. The CCI surface soil moisture anomaly was estimated by removing its long-term seasonality (Chen et al. 2017) and converting resulting surface anomalies into a RZSM anomaly using an exponential filter (Albergel et al. 2008):

$$SWI_{(n)} = SWI_{(n-1)} + K_{(n)}\left(\theta_{(n)} - SWI_{(n-1)}\right) \tag{1}$$

$$K_{(n)} = \frac{1}{1+\sum_{i}^{n-1} e^{\frac{t_i - t_n}{T}}} \tag{2}$$

where SWI is the CCI-based 0-40 cm RZSM anomaly; θ is CCI soil moisture anomaly; T is a time scale parameter and n is a particular time step.

The parameter T was calibrated using NLDAS-based surface (noted as NS) and root zone (noted as NR) soil moisture anomalies on a grid-by-grid base. For each grid, RZSM anomalies were estimated by exponentially filtering NS using an initialize guess for T (the resulting RZSM anomaly is noted as NR_{ef}). The parameter T was then optimized by maximizing the Kling–Gupta efficiency (Gupta et al. 2009) of NR_{ef} and NR:

$$KGE = 1 - ED \tag{3}$$

$$ED = \sqrt{(r-1)^2 + (\alpha - 1)^2} \tag{4}$$

where KGE is the Kling-Gupta efficiency, r is the linear correlation between NR_{ef} and NR and α the ratio of NR_{ef} and NR standard deviations. Since soil moisture anomalies are zero-mean, the bias term in the KGE number was neglected.

The anomaly of SMERGE product was derived by weighted averaging this CCI- and NLDAS-based (i.e., NR) RZSM anomaly:

$$\mathrm{SMERGE} = w\mathrm{NR} + (1-w)\mathrm{SWI} \tag{5}$$

where w is a spatially-varied weighting factor. The weights are determined by maximizing the temporally lag-1 [month] Spearman rank correlation between monthly SMERGE and NDVI for each spatial grid cell. However, it should be noted that this weight optimization can be problematic when CCI or NLDAS has extremely low data qualities. To provide the most robust estimates of w, prescribed w values are used for the following cases: (1) $w = 1$ if the CCI monthly coverage is less than 90%; (2) $w = 1$ if the Spearman rank correlation between SMERGE v2.0 and Normalized Difference Vegetation Index (NDVI) based on Advanced Very High-Resolution Radiometer (AVHRR) Global NDVI Version 3.1 (1992 to 2015) is less than 0.20; (3) $w = 0.8$ if the predominant land cover type (from National Land Cover Database 2006) is either forest or wetlands; and (4) $w = 0.5$ if the predominant land cover type is either urban, herbaceous, or cultivated, and the Spearman rank correlation between SMERGE v2.0 and NDVI-AVHRR (1992 to 2015) is less than 0.2. With all these factors considered, the spatial distribution of *w* is as shown in Figure 1. Final SMERGE RZSM estimates were derived by adding the NLDAS RZSM climatology to the SMERGE anomaly (Fig. 2).

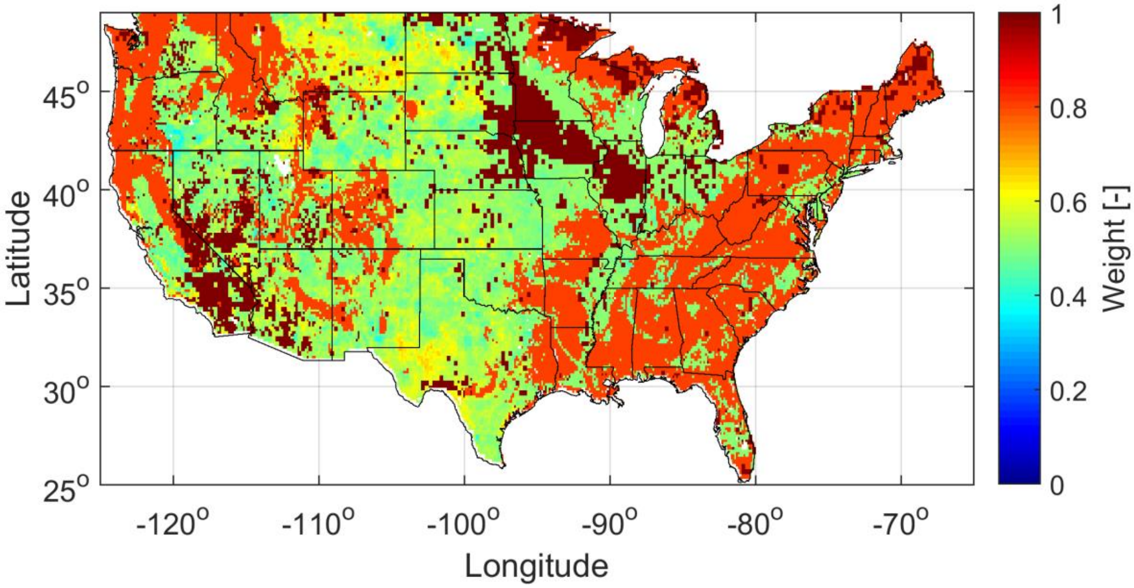

Figure 1. Spatial distribution of the weighting factor for merging NLDAS- and CCI-based 0 – 40 cm soil moisture anomalies.

It should be noted that CCI soil moisture retrievals contain temporal gaps. To provide a continuous SMERGE product, the CCI- and NLDAS-based RZSM anomaly differences were linearly interpolated on days lacking remote-sensing observations. The interpolated CCI values were derived by adding the interpolated CCI-NLDAS differences to NLDAS RZSM anomalies. Days with and without this interpolation are noted in the "smflag" data quality flag as 0 (interpolated) and 1 (recommended), respectively.

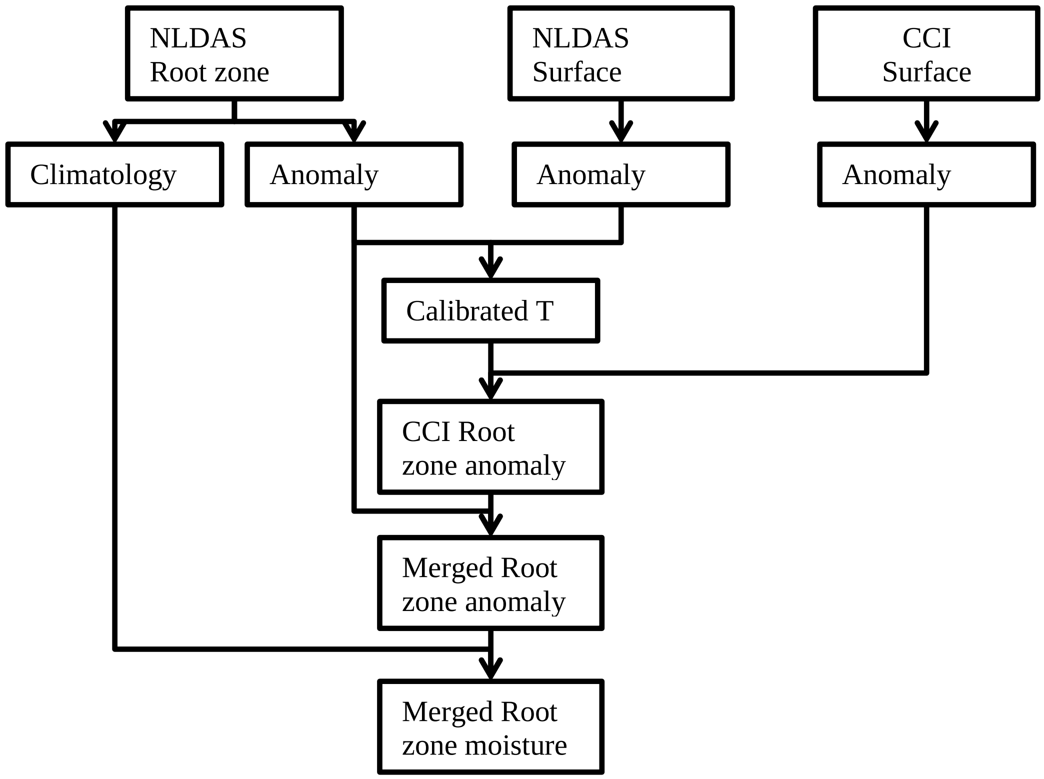

Figure 2. Flowchart of the main steps in the SMERGE RZSM algorithm.

1.3 Data Disclaimer

The SMERGE Root Zone Soil Moisture Data Products are made available with no warranty, explicit or implied, to the extent permitted by applicable law.

1.4 Digital Object Identifier (DOI) and Citation

Users of SMERGE Root Zone Soil Moisture Data Products should cite, in research papers, the data used, along with their associated Digital Object Identifiers (DOIs) (Table 2). A DOI is a unique alphanumeric string used to identify a digital object and provide a permanent link online. DOIs are often used in online publications in citations.

Table 2. DOIs for SMERGE Root Zone Soil Moisture Data Products.

Product Short Name	Product Description	DOI
SMERGE_RZSM0_40CM.2.0	SMERGE 0-40 cm root zone soil moisture	10.5067/PAVQY1KHTMUT

The "Product Short Name" in Table 2 is linked to its corresponding data product page. In the latter, the tab "Data Citation" provides the recommended citation for that product.

1.5 What's New?

1.5.1 What are the differences between SMERGE V1.0 and SMERGE V2.0?

The main differences between SMERGE v1.0 and SMERGE v2.0 are the addition of a new variable, "CCI derived soil moisture anomalies of 0-40 cm layer," and additional years of data for 2016 to 2019.

1.5.2 Known issues regarding the new CCI anomaly variable

In some of the earlier (before around 1990) CCI surface soil moisture estimates, there are gaps because of a lack of satellite data. Due to continual improvement in soil moisture remote sensing techniques and data availability (Fig. 3), more recent CCI estimates have temporally reduced data gaps and improved accuracy. In particular, significant reductions in temporal data gaps and retrieval errors are expected at points in time when AMSR-E (2002) and ASCAT (2007) data products are incorporated into the CCI. These factors may also lead to temporal variations in the accuracy of SMERGE V2.0 root-zone soil moisture estimates.

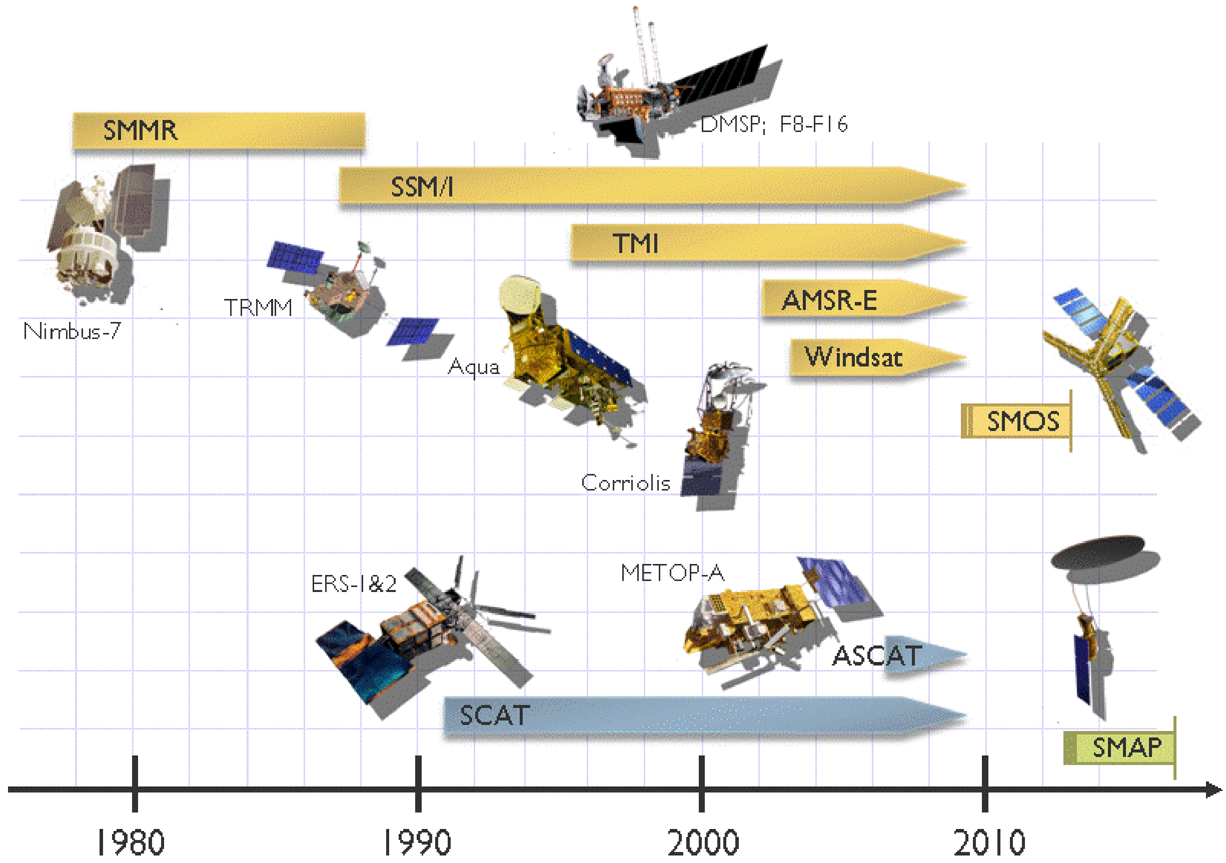

Figure 3. Available satellites and their active and passive microwave sensors used for the generation of the CCI soil moisture data set. (Figure courtesy https://www.esa-soilmoisture-cci.org/node/93)

2.0 Data Organization

The SMERGE Root Zone Soil Moisture Data Product (Table 2) is gridded at a 0.125-degree resolution. The first grid cell is centered at a (latitude, longitude) of (-124.9375, 52.9375) or (-124.9375, 25.0625) in the northwest corner of the grid.

2.1 File Naming Convention

File names are formatted as follows and with <fields> described in Table 3:

Smerge_ Noah_CCI_L4_RZSM0_40cm_V< product version >_<date of data>.nc4

where

- Smerge is soil moisture merging project
- Noah_CCI means soil moisture from Noah and CCI products are used
- RZSM0_40cm indicates root zone soil moisture representative depth
- < product version > indicates the version of the SMERGE product
- < date of data > indicates the date of the soil moisture data
- .nc indicates netCDF file format

Table 3. Description of file name attributes.

Attribute	**Description**
<product version>	2.0
<date of data >	YYYYMMDD

Filename examples:

Smerge_Noah_CCI_L4_RZSM0_40cm_V2.0_19790102.nc4 – SMERGE 0-40 cm root zone soil moisture, version 2.0, for January 2, 1979

2.2 File Format and Structure

The SMERGE Root Zone Soil Moisture Data Products (Table 2) are in netCDF (http://www.unidata.ucar.edu/software/netcdf/docs/), which facilitates the creation, access, and sharing of array-oriented data in a form that is self-describing and portable.

Each daily file contains geolocation information (latitude and longitude of grid box centers), a flag, and 1 data field (Sec. 3.0).

3.0 Data Contents

The following section and their corresponding tables provide descriptions of the SMERGE data variables. The valid flag variable only contains 0 and 1.

3.1 SMERGE Data Variables

Short Name	Description	Unit
smflag	Data Quality Flag	
RZSM	Volumetric soil moisture	m^3m^{-3}
CCI_ano	CCI-based 0 – 40 soil moisture anomalies	m^3m^{-3}

SMERGE bit flag values and their meaning are as follows:

Flag value	Meaning
0	Interpolated
1	Recommended

4.0 Options for Reading the Data

The following are a few of the many command line and visualization tools available for reading netCDF format data, such as the SMERGE Root Zone Soil Moisture Data Products. For more comprehensive lists of tools, please see the following:

https://www.unidata.ucar.edu/software/netcdf/docs/

https://www.hdfgroup.org/products/hdf5_tools/

Most of the following tools (e.g., GrADS, NCO, CDO, NCL, IDL) can subset variables or subset data within specified temporal and/or spatial ranges. These tools can also calculate statistics like mean, standard deviation, maximum, minimum, etc.

4.1 Command Line Utilities

4.1.1 ncdump (free)

The ncdump tool generates the CDL (Common Data Language) text (ASCII) representation of a netCDF or compatible file and writes to standard output. The tool can also be used as a simple browser for netCDF files, to display the dimension names and lengths; variable names, types, and shapes; attribute names and values; and, optionally, the values of data for all variables or selected variables. A common use of ncdump is with the –h option, with which only the header information is displayed. The ncdump tool comes with the netCDF library as distributed by Unidata.

http://www.unidata.ucar.edu/downloads/netcdf/

4.1.2 h5dump (free)

The h5dump tool enables users to examine the contents of an HDF5 file and dump those contents to an ASCII file or, optionally, as XML or binary outputs. It can display the contents of the entire HDF5 file or selected objects, which can be groups, data sets, a subset of a data set, links, attributes, or datatypes. Please note h5dump may not work with older netCDF formats. The h5dump tool is included with the HDF5 distribution from The HDF Group.

https://www.hdfgroup.org/HDF5/release/obtain5.html

4.1.3 NCO (free)

The netCDF Operator (NCO) (http://nco.sourceforge.net/) toolkit manipulates and analyzes data stored in netCDF-accessible formats, including DAP, HDF4, and HDF5.

4.1.4 CDO (free)

CDO (Climate Data Operators) (https://code.zmaw.de/projects/cdo) is a collection of command line operators to manipulate and analyze Climate and Numerical Weather Prediction (NWP) model Data.

4.2 Visualization Tools

4.2.1 Ncview (free)

Ncview is a quick and easy way to visualize the contents of netCDF files.

http://meteora.ucsd.edu/~pierce/ncview_home_page.html

4.2.2 ncBrowse (free)

ncBrowse is a Java application that provides flexible, interactive graphical displays of data and attributes from a wide range of netCDF data file conventions.

http://www.epic.noaa.gov/java/ncBrowse/

4.2.3 Panoply (free)

Panoply is a Java application, developed by the NASA Goddard Institute for Space Studies (GISS), that plots geo-referenced and other arrays from netCDF, HDF, GRIB, and other data types. Among other capabilities, Panoply enables one to slice and plot geo-referenced latitude-longitude, latitude-vertical, longitude-vertical, time-latitude, or time-vertical arrays from larger multidimensional variables; combine two geo-referenced arrays in one plot by differencing, summing, or averaging; plot maps using various map projections; and access remote catalogs to retrieve data files.

http://www.giss.nasa.gov/tools/panoply/

The How-To's of NASA GES DISC provides a recipe for How to View Remote Data in OPeNDAP with Panoply

4.2.4 HDFview (free)

HDFView is a Java-based visual tool created by The HDF Group for browsing and editing HDF4 and HDF5 files. It allows users to view all objects in an HDF file hierarchy, which is represented as a tree structure, and create, add, delete, and modify object contents and attributes.

https://www.hdfgroup.org/products/java/hdfview/

4.2.5 IDL netCDFtools (commercial)

Users familiar with the IDL programming language (http://www.exelisvis.com/ProductsServices/IDL.aspx) can use the netCDF functions available with the IDL software package to read and visualize the data.

4.2.6 GrADS netCDFtools (free)

Users familiar with the GrADS programming language (http://iges.org/grads/) can use the netCDF functions available with the GrADS software package to read and visualize the data.

4.2.7 NCL (free)

The NCAR Command Language (NCL) (http://www.ncl.ucar.edu/) is a free interpreted language designed specifically for scientific data processing and visualization.

5.0 Data Services

Access to GES DISC data requires all users to be registered with the **NASA Earthdata Login System** (as of August 1, 2016). Data continue to be free of charge and accessible via HTTP. Access to data via FTP is no longer available (as of October 3, 2016). Detailed instructions on how to register and receive authorization to access GES DISC data are provided at https://disc.sci.gsfc.nasa.gov/data-access.

GES DISC users who deploy scripting methods to list and download data in bulk via anonymous FTP are advised to review the How to Download Data Files from HTTP Service with wget recipe, which provides examples of GNU wget commands for listing and downloading data via HTTP.

If you need assistance or wish to report a problem:
Email: gsfc-dl-help-disc@mail.nasa.gov
Voice: 301-614-5224

The following link is the Data Product Page for SMERGE root-zone soil moisture of 0 – 40 cm: https://disc.gsfc.nasa.gov/datasets/NLDAS_FORA0125_H_V002/summary

5.1 HTTP

For access to the online archive data via HTTP.

For SMERGE root zone soil moisture of 0 – 40 cm:
https://hydro1.gesdisc.eosdis.nasa.gov/data/SMERGE/SMERGE_RZSM0_40CM.1.0/

5.2 Earthdata Search

Use the Earthdata Search Client (EDSC) to find and retrieve data sets archived at the GES DISC or across multiple data enters.

For SMERGE root zone soil moisture of 0 – 40 cm:
https://search.earthdata.nasa.gov/search?q=SMERGE_RZSM0_40CM

5.4 OPeNDAP

The Open-source Project for a Network Data Access Protocol (OPeNDAP) provides a means for requesting and accessing data across the internet, in a form usable by OPeNDAP clients, i.e., clients that can remotely access OPeNDAP-served data (e.g., Panoply, IDL, MATLAB, GrADS, IDV, McIDAS-V, Ferret). OPeNDAP provides the ability to retrieve subsets of files and to aggregate data from several files in one transfer operation.

For SMERGE root zone soil moisture of 0 – 40 cm:
https://hydro1.gesdisc.eosdis.nasa.gov/data/SMERGE/SMERGE_RZSM0_40CM.1.0/

5.5 Giovanni

The NASA GES-DISC Interactive Online Visualization ANd aNalysis Interface (Giovanni) is a Web-based tool that allows users to interactively visualize and analyze data.

https://giovanni.gsfc.nasa.gov/giovanni/#dataKeyword=SMERGE_RZSM0_40CM_1.0

Figure 4 shows a sample map of SMERGE root zone soil moisture generated by NASA Giovanni.

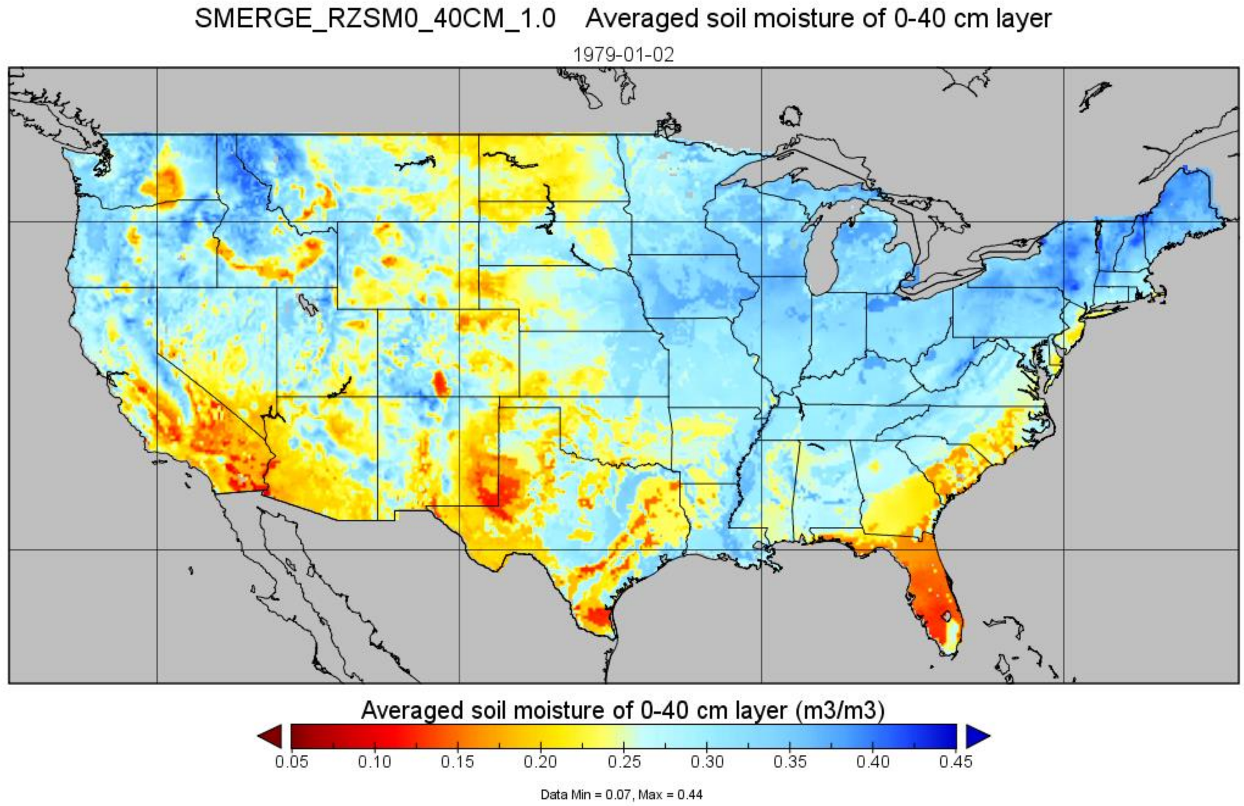

Figure 4. Map of SMERGE root zone soil moisture of 0 – 40 cm for January 02, 1979.

6.0 More Information

6.1 Other Soil Moisture Resources

For other soil moisture and related data available at the GES DISC, please see

https://disc.gsfc.nasa.gov/datasets?page=1&keywords=soil%20moisture

For other soil moisture and related data available elsewhere, please search NASA's Global Change Master Directory (GCMD) at http://gcmd.nasa.gov/.

6.2 Points of Contact

Name: GES DISC Help Desk
URL: http://disc.sci.gsfc.nasa.gov/
E-mail: gsfc-dl-help-disc@mail.nasa.gov
Phone: 301-614-5224
Fax: 301-614-5268
Address: Goddard Earth Sciences Data and Information Services Center
Attn: Help Desk
Code 610.2
NASA Goddard Space Flight Center
Greenbelt, MD 20771
USA

For specific questions about the SMERGE algorithm, please contact the data producer, Wade Crow (Wade.Crow@ars.usda.gov), Jianzhi Dong (Jianzhi.Dong@ars.usda.gov), and Kenneth Tobin (ktobin@tamiu.edu).

6.3 Acronyms

CCI:	Climate Change Initiative
ESA:	European Space Agency
DOI:	Digital Object Identifier
NLDAS:	North American Land Data Assimilation System
SMERGE:	Root-zone Soil moisture MERGE Project

7.0 Acknowledgments

The SMERGE is an awarded project under ROSES-2014, Program Element A.29: Climate Indicators and Data Products for Future National Climate Assessments. The production of the SMERGE Data Products and related research was supported by the research team of Crow, Wade and J. Dong (USDA ARS), and K. Tobin (Texas A&M IU).

References

Tobin, K. J., W. T. Crow, J. Dong, and M. E. Bennett, 2019. "Validation of a New Root-Zone Soil Moisture Product: Soil MERGE." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3351-3365, doi:10.1109/JSTARS.2019.2930946

Tobin, K.J., R. Torres, W. Crow, and M.E. Bennett. 2017. "Multi-decadal analysis of root-zone soil moisture applying the exponential filter across CONUS." *Hydrol. Earth Syst. Sci.*, 21, 4403-4417, doi:10.5194/hess-21-4403-2017.

Albergel, C., C. Rüdiger, T. Pellarin, J.-C. Calvet, N. Fritz, F. Froissard, D. Suquia, a. Petitpa, B. Piguet, and E. Martin. 2008. "From near-Surface to Root-Zone Soil Moisture Using an Exponential Filter: An Assessment of the Method Based on in-Situ Observations and Model Simulations." *Hydrology and Earth System Sciences Discussions* 5 (3): 1603–40. doi:10.5194/hessd-5-1603-2008.

Chen, Fan, Wade T. Crow, Andreas Colliander, Michael H. Cosh, Thomas J. Jackson, Rajat Bindlish, Rolf H. Reichle, et al. 2017. "Application of Triple Collocation in Ground-Based Validation of Soil Moisture Active/Passive (SMAP) Level 2 Data Products." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10 (2): 489–502. doi:10.1109/JSTARS.2016.2569998.

Gupta, Hoshin V., Harald Kling, Koray K. Yilmaz, and Guillermo F. Martinez. 2009. "Decomposition of the Mean Squared Error and NSE Performance Criteria: Implications for Improving Hydrological Modelling." *Journal of Hydrology* 377 (1–2). Elsevier B.V.: 80–91. doi:10.1016/j.jhydrol.2009.08.003.

Liu, Y. Y., W. A. Dorigo, R. M. Parinussa, R. A M De Jeu, W. Wagner, M. F. McCabe, J. P. Evans, and A. I J M Van Dijk. 2012. "Trend-Preserving Blending of Passive and Active Microwave Soil Moisture Retrievals." *Remote Sensing of Environment* 123 (October 2006). Elsevier Inc.: 280–97. doi:10.1016/j.rse.2012.03.014.