NASMo-TiAM 250m 16-day North America Surface Soil Moisture Dataset

Get Data

Documentation Revision Date: 2024-03-21

Dataset Version: 1

Summary

This NASMo-TiAM (North America Soil Moisture Dataset Derived from Time-Specific Adaptable Machine Learning Models) dataset holds gridded estimates of surface soil moisture (0-5 cm depth) at a spatial resolution of 250 meters over 16-day intervals from mid-2002 to December 2020 for North America. The model employed Random Forests to downscale coarse-resolution soil moisture estimates (0.25 deg) from the European Space Agency Climate Change Initiative (ESA CCI) based on their correlation with a set of static (terrain parameters, bulk density) and dynamic covariates (Normalized Difference Vegetation Index, land surface temperature). NASMo-TiAM 250m predictions were evaluated through cross-validation with ESA CCI reference data and independent ground-truth validation using North American Soil Moisture Database (NASMD) records. The data are provided in cloud optimized GeoTIFF format.

This dataset holds 1109 files in cloud optimized GeoTIFF format.

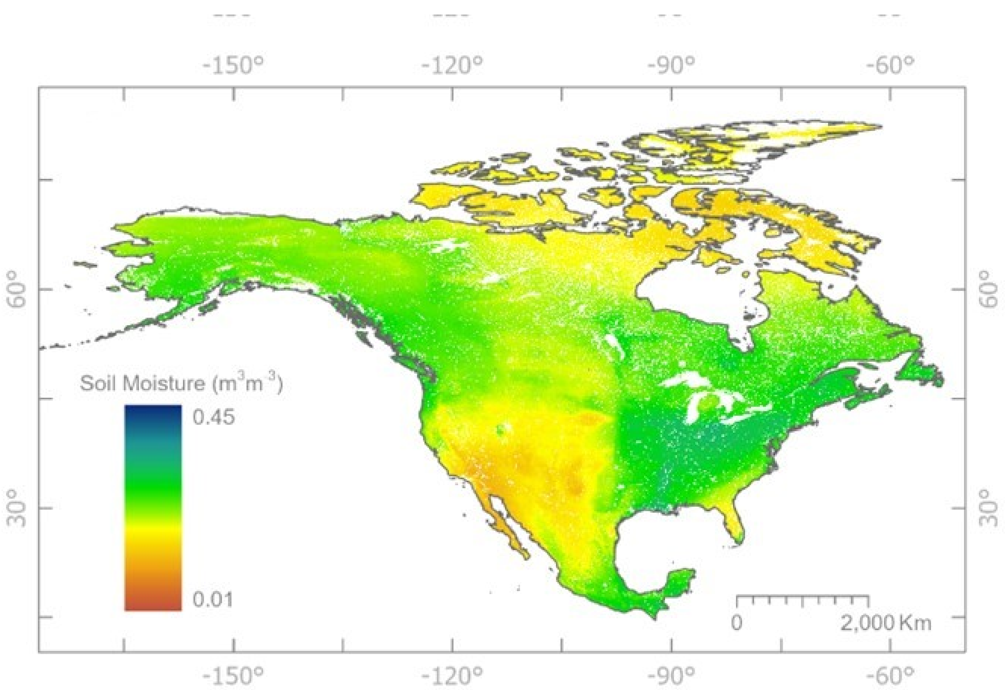

Figure 1. Random Forests predictions of mean volumetric soil moisture values for North America derived from 426 biweekly periods from 2002 to 2020.

Citation

Llamas, R., P. Olaya, M. Taufer, and R. Vargas. 2024. NASMo-TiAM 250m 16-day North America Surface Soil Moisture Dataset. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/2326

Table of Contents

1. Dataset Overview

This NASMo-TiAM (North America Soil Moisture Dataset Derived from Time-Specific Adaptable Machine Learning Models) dataset holds gridded estimates of surface soil moisture (0-5 cm depth) at a spatial resolution of 250 meters over 16-day intervals from mid-2002 to December 2020 for North America. The model employed Random Forests to downscale coarse-resolution soil moisture estimates (0.25 deg) from the European Space Agency Climate Change Initiative (ESA CCI) based on their correlation with a set of static (terrain parameters, bulk density) and dynamic covariates (Normalized Difference Vegetation Index, land surface temperature). NASMo-TiAM 250m predictions were evaluated through cross-validation with ESA CCI reference data and independent ground-truth validation using North American Soil Moisture Database (NASMD) records.

Project: Carbon Monitoring System

The NASA Carbon Monitoring System (CMS) program is designed to make significant contributions in characterizing, quantifying, understanding, and predicting the evolution of global carbon sources and sinks through improved monitoring of carbon stocks and fluxes. The System uses NASA satellite observations and modeling/analysis capabilities to establish the accuracy, quantitative uncertainties, and utility of products for supporting national and international policy, regulatory, and management activities. CMS data products are designed to inform near-term policy development and planning.

Related Publication

Llamas, R., P. Olaya, M. Taufer, and R. Vargas. 2024. North America Soil Moisture Dataset derived from Time-specific Adaptable Machine learning models (NASMo-TiAM 250m). In Preparation for Scientific Data, 2024.

Acknowledgement

This study was funded by NASA’s Carbon Monitoring System program (grant 80NSSC21K0964) and the National Science Foundation's Office of Advanced Cyberinfrastructure (grants 2103845, 2103836, and 2334945).

2. Data Characteristics

Spatial Coverage: North America (Canada, United States of America, Mexico)

Spatial Resolution: 0.002548 deg (~250 m); 0-5 cm soil depth

Temporal Coverage: 2002-06-26 to 2020-12-31

Temporal Resolution: 16-day periods for each year

Study Areas: Latitude and longitude are given in decimal degrees.

Site	Westernmost Longitude	Easternmost Longitude	Northernmost Latitude	Southernmost Latitude
North America	-180.000	-39.995	82.720	14.531

Data File Information

This dataset holds 1109 files in cloud optimized GeoTIFF format (*.tif). These files hold estimates of volumetric water content in the upper layer of soil, typically 0 to 5 cm of soil surface.

The file naming convention is *northamerica_<metric>_rf_esacci71_<year>_<period>.tif*, where

- *<metric>* = "ssm" (surface soil moisture), "cv_residuals" (cross validation residuals), or "gt_residuals" (ground truth validation residuals). See Table 1.
- *<year>* = year from 2002 to 2020
- *<period>* = the 16-day period within the year numbered 1 to 23.
- "*rf*" represents Random Forests, the modeling methodology
- "*esacci71*" denotes the ESA CCI soil moisture product used as input data for modeling and validation.

Example file name: "northamerica_cv_residuals_rf_esacci71_2002_12.tif"

GeoTIFF characteristics:

- Coordinate system: geographic coordinates in WGS 1984 datum (EPSG 4326)
- Spatial resolution: 0.002548 degrees (~250 m)
- 26,761 rows x 54,946 columns
- Pixel values: volumetric soil moisture in units of volume water per volume soil (e.g., cm ^{3} cm^{-3})
- No data value: -9999

Table 1. File types and description. Units for all file types is volumetric soil moisture (e.g., cm ^{3} cm^{-3}).

Metric	Description	Number of Files
ssm	Surface soil moisture: the water content present in the upper layer of the soil, typically within 5 cm of the soil surface. Values predicted by the *NASMo-TiAM 250m* model.	426
cv_residuals	Residuals from cross-validation procedure using Random Forests models.	426
gt_residuals	Residuals from ground truth validation. Values are differences between model-predicted *ssm* values and in situ soil moisture measurements from the NASMD.	257

3. Application and Derivation

Soil moisture plays a crucial role in the Earth's ecosystems and has substantial implications in different scientific fields such as hydrology (Jackson et al., 1996; Robinson et al., 2008), ecology and climate science (Davidson et al., 1998; Falloon et al., 2011; Legates et al., 2010; Ward, 2008). A greater understanding of soil moisture processes and its spatial and temporal distribution can lead to improvements in different fields, such as agriculture (Engman, 1991; Hunt, 2015; Pablos et al., 2017), water resources management (Jacobs et al., 2003), natural disasters related to flooding (Tuttle et al., 2017), landslides (Crow, 2019), and drought events (Pablos et al., 2017).

This NASMo-TiAM 250 m dataset provides fine spatial resolution soil moisture dataset across the North American region. Other continent-scale datasets have resolutions ranging from 0.25 degrees (O. and Orth, 2021) to 25 km (Skulovich and Gentine, 2023) and 1 km (Han et al., 2023). While the dataset from Vergopolan et al. (2021) has 30-m resolution, its spatial coverage is limited to the conterminous United States.

4. Quality Assessment

Two time-specific validation approaches were performed for the biweekly prediction outputs; (i) a cross-validation approach with the 30% of ESA CCI soil moisture sample points set aside during the generation of the training matrices and (ii) an independent ground-truth validation using in situ 0-5 cm soil moisture measurements from the NASMD.

Cross-validation showed an overall mean correlation coefficient of 0.9 and an RMSE of 0.03 m^{3} m^{-3}. In this validation, 2,885,010 points were used across all 426 biweekly periods.

In the ground-truth validation, an overall mean correlation of 0.4 and RMSE of 0.11 m^{3} m^{-3} were obtained; 96,217 points were used in relation to the availability of data in all NASDM stations with valid records during the study's time frame.

5. Data Acquisition, Materials, and Methods

The North America Soil Moisture dataset derived from Time-specific Adaptable machine learning Models (*NASMo-TiAM 250 m*), represents a fine spatial resolution soil moisture dataset for the North American region. It holds soil moisture values on 16-day periods from mid-June 2002 to December 2020, at 250 m resolution. Soil moisture estimates were downscaled from course scale measurements (0.25 degrees) in satellite-derived soil moisture from the European Space Agency Climate Change Initiative (ESA CCI, version 7.1) product (Dorigo et al., 2023).

A Random Forests (RF) machine learning technique was chosen to predict soil moisture in fine resolution due to its performance in previous studies (Llamas et al., 2022; Rorabaugh et al., 2019) and its utility with varying spatial and temporal resolutions (Han et al., 2023; Vergopolan et al., 2021). A set of static and dynamic covariates were used to downscale the coarse scale ESA CCI data to 250 m. Hydrologically meaningful terrain parameters and soil bulk density data are used as static covariates, while dynamic covariates were represented by Normalized Difference Vegetation Index (NDVI) and land surface temperature (LST). Data on snow- and Ice-cover were used to prevent the RF models from training and predicting soil moisture values over those areas.

The *NASMo-TiAM 250 m* workflow (Figure 2) was developed based on standardized input data, wherein all prediction covariates were preprocessed to allocate the same temporal and spatial characteristics. Input data was standardized at a spatial resolution of 250 m, aligning with the finest spatial resolution available among the input spatial data (i.e., MODIS NDVI, topographic information, and soil bulk density). This standardization allowed for the disaggregation of coarser resolution datasets (i.e., LST and snow cover).

The time-specific component of the *NASMo-TiAM 250m* dataset accounts for seasonal variation. The 16-day time steps represented the lowest temporal resolution of the input data (i.e., MODIS NDVI) and enabled the aggregation of higher temporal resolution datasets (i.e., ESA CCI soil moisture, MODIS LST, MODIS snow cover) within the same 16-day period. Time-specific models were trained every 16 days across the North American region.

Two accuracy assessments of predictions were performed using a cross-validation procedure and an independent validation with ground-truth data from NASMD (Quiring et al., 2016). Cross validation involved setting aside 30% of ESA CCI soil moisture sample points during model training, running the RF model, then comparing RF predictions to the set aside reference data. Separately, RF model predictions were compared to in situ soil moisture measurements from NASMD for 2002-2013. The accuracy results are reported as residuals, the difference between predicted and reference values.

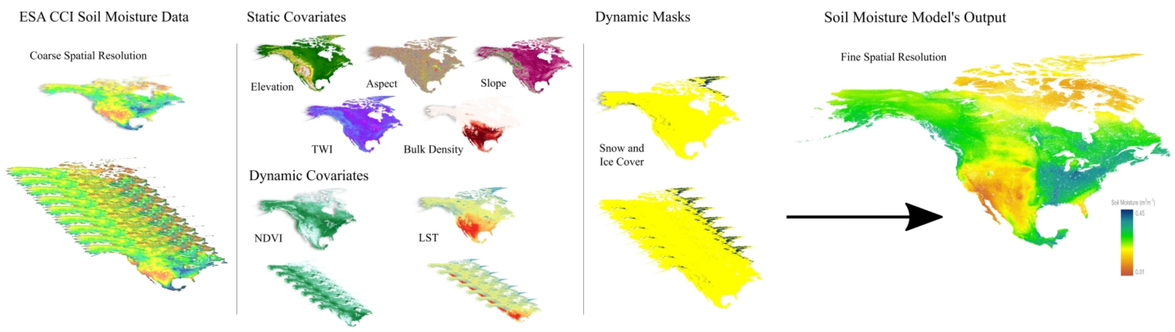

Figure 2. The *NASMo-TiAM 250 m* workflow involved standardizing input data to common spatial and temporal resolutions, integrating static and dynamic covariates, and training a Random Forest model to output fine scaled soil moisture across North America.

Additional details of methods are available in Llamas et al. (2024).

6. Data Access

These data are available through the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC).

NASMo-TiAM 250m 16-day North America Surface Soil Moisture Dataset

Contact for Data Center Access Information:

- E-mail: uso@daac.ornl.gov
- Telephone: +1 (865) 241-3952

7. References

Crow, W.T. 2019. Utility of soil moisture data products for natural disaster applications. Extreme Hydroclimatic Events and Multivariate Hazards in a Changing Environment:65-85. https://doi.org/10.1016/B978-0-12-814899-0.00003-1

Davidson, E.A., E. Belk, and R.D. Boone. 1998. Soil water content and temperature as independent or confounded factors controlling soil respiration in a temperate mixed hardwood forest. Global Change Biology 4:217–227. https://doi.org/10.1046/j.1365-2486.1998.00128.x

Dorigo, W., W. Preimesberger, L. Moesinger, A. Pasik, T. Scanlon, S. Hahn, R. Van der Schalie, M. Van der Vliet, M., R. De Jeu, R. Kidd, N. Rodriguez-Fernandez, and M. Hirschi. 2023. ESA Soil Moisture Climate Change Initiative (Soil_Moisture_cci). Version 07.1. NERC EDS Centre for Environmental Data Analysis. https://catalogue.ceda.ac.uk/uuid/e235d2980d6a441895f7221ff4787a6f

Engman, E.T. 1991. Applications of microwave remote sensing of soil moisture for water resources and agriculture. Remote Sensing of Environment 35:213–226. https://doi.org/10.1016/0034-4257(91)90013-V

Falloon, P., C.D. Jones, M. Ades, and K. Paul. 2011. Direct soil moisture controls of future global soil carbon changes: An important source of uncertainty. Global Biogeochemical Cycles 25:GB3010. https://doi.org/10.1029/2010GB003938

Han, Q., Y. Zeng, L. Zhang, C. Wang, E. Prikaziuk, Z. Niu, and B. Su. 2023. Global long term daily 1 km surface soil moisture dataset with physics informed machine learning. Scientific Data 10:101. https://doi.org/10.1038/s41597-023-02011-7

Hunt., E.D. 2015. Analyzing the relationship of soil moisture and biophysical variables in wet and dry seasons at a rainfed and irrigated field in eastern Nebraska. Dissertation. ETD collection for University of Nebraska-Lincoln. AAI3716537. https://digitalcommons.unl.edu/dissertations/AAI3716537

Jacobs, J.M., D.A. Myers, and B.M. Whitfield. 2003. Improved rainfall/runoff estimates using remotely sensed soil moisture1. Journal of the American Water Resources Association 39:313–324. https://doi.org/10.1111/j.1752-1688.2003.tb04386.x

Jackson, T.J., J. Schmugge, and E.T. Engman. 1996. Remote sensing applications to hydrology: soil moisture. Hydrological Sciences Journal 41:517–530. https://doi.org/10.1080/02626669609491523

Legates, D.R., R. Mahmood, D.F. Levia, T.L. DeLiberty, S.M. Quiring, C. Houser, and F.E. Nelson. 2010. Soil moisture: A central and unifying theme in physical geography. Progress in Physical Geography: Earth and Environment 35:65–86. https://doi.org/10.1177/0309133310386514

Llamas, R.M., L. Valera, P. Olaya, M. Taufer, and R. Vargas. 2022. Downscaling satellite soil moisture using a modular spatial inference framework. Remote Sensing 14:3137. https://doi.org/10.3390/rs14133137

Llamas, R., P. Olaya, M. Taufer, and R. Vargas. 2024. North America Soil Moisture Dataset derived from Time-specific Adaptable Machine learning models (NASMo-TiAM 250m). In Preparation for Scientific Data, 2024.

O, S., and R. Orth. 2021. Global soil moisture data derived through machine learning trained with in-situ measurements. Scientific Data 8. https://doi.org/10.1038/s41597-021-00964-1

Pablos, M., J. Martínez-Fernández, N. Sánchez, and Á. González-Zamora. 2017. Temporal and spatial comparison of agricultural drought Indices from moderate resolution satellite soil moisture data over northwest Spain. Remote Sensing 9:1168. https://doi.org/10.3390/rs9111168

Quiring, S.M., T.W. Ford, J.K. Wang, A. Khong, E. Harris, T. Lindgren, D.W. Goldberg, and Z. Li. 2016. The North American Soil Moisture Database: development and applications. Bulletin of the American Meteorological Society 97:1441–1459. https://doi.org/10.1175/BAMS-D-13-00263.1

Robinson, D.A., C.S. Campbell, J.W. Hopmans, B.K. Hornbuckle, S.B. Jones, R. Knight, F. Ogden, J. Selker, and O. Wendroth. 2008. Soil moisture measurement for ecological and hydrological watershed-scale observatories: a review. Vadose Zone Journal 7:358–389. https://doi.org/10.2136/vzj2007.0143

Rorabaugh, D., M. Guevara, R. Llamas, J. Kitson, R. Vargas, and M. Taufer. 2019. SOMOSPIE: A modular SOil MOisture SPatial Inference Engine based on data-driven decisions. 2019 15th International Conference on eScience (eScience); September 2019, San Diego, California, USA. https://doi.org/10.1109/eScience.2019.00008

Skulovich, O., and P. Gentine. 2023. A long-term consistent artificial intelligence and remote sensing-based soil moisture dataset. Scientific Data 10:154. https://doi.org/10.1038/s41597-023-02053-x

Tuttle S.E., E. Cho, P.J. Restrepo, X. Jia, C.M. Vuyovich, M.H. Cosh, and J.M. Jacobs. 2017. Remote sensing of drivers of Spring snowmelt flooding in the North Central U.S. Pp. 21-45 in V. Lakshmi (ed), Remote Sensing of Hydrological Extremes. Springer. https://doi.org/10.1007/978-3-319-43744-6_2

Vergopolan, N., N.W. Chaney, M. Pan, J. Sheffield, H.E. Beck, C.R. Ferguson, L. Torres-Rojas, S. Sadri, and E.F. Wood. 2021. SMAP-HydroBlocks, a 30-m satellite-based soil moisture dataset for the conterminous US. Scientific Data 8:264. https://doi.org/10.1038/s41597-021-01050-2

Ward, S. 2008. The Earth Observation Handbook – Climate Change Special Edition 2008. European Space Agency (ESA) Communication Production Office; The Netherlands. https://eohandbook.com/eohb2008/CEOS_EOHB_2008.pdf