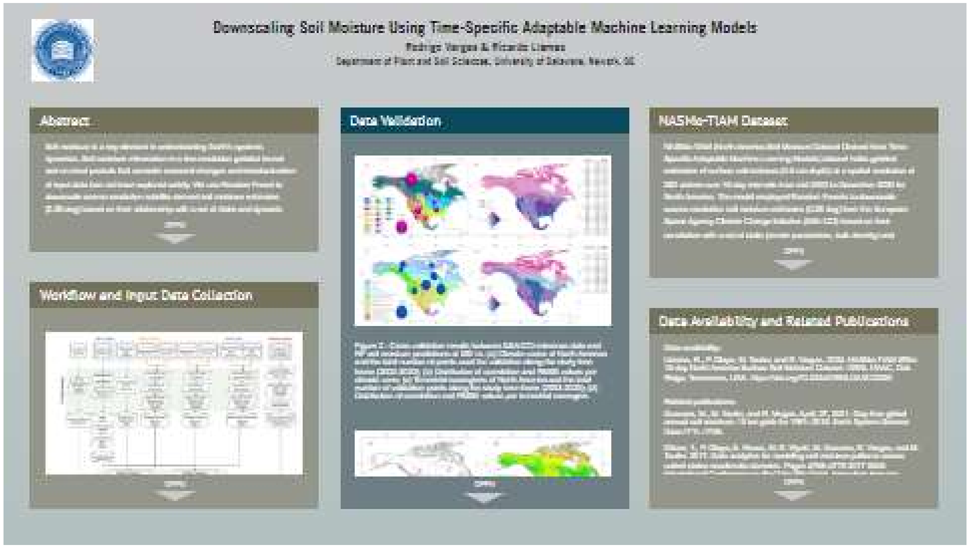
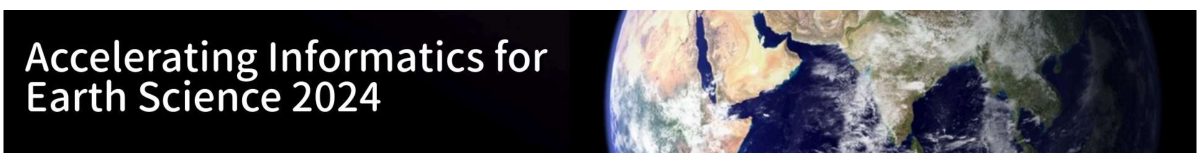
Downscaling Soil Moisture Using Time-Specific Adaptable Machine Learning Models



Rodrigo Vargas & Ricardo Llamas

Department of Plant and Soil Sciences, University of Delaware, Newark, DE

PRESENTED AT:



ABSTRACT

Soil moisture is a key element in understanding Earth's systems dynamics. Soil moisture information in a fine-resolution gridded format and on short periods that consider seasonal changes and standardization of input data has not been explored widely. We use Random Forest to downscale coarse-resolution satellite-derived soil moisture estimates (0.25 deg) based on their relationship with a set of static and dynamic covariates used as predictors. We provide surface soil moisture (0-5cm depth) estimates at 250 m of spatial resolution on 16-day periods from mid-2002 to December 2020 at a subcontinental level through the North America Soil Moisture Dataset Derived from Time-Specific Adaptable Machine Learning Models (NASMo-TiAM 250 m). NASMo-TiAM 250 m predictions are evaluated through cross-validation with ESA CCI reference data and independent ground-truth validation using North American Soil Moisture Database (NASMD) records. We found a correlation coefficient and RMSE derived from cross-validation of 0.91 and 0.03m $^{3}m^{-3}$ respectively. For ground-truth validation, we found an overall correlation of 0.4 and an RMSE of 0.11 m^3m^{-3} Additionally, we observed a correlation of 0.38 and RMSE of 0.12 m^3m^{-3} between reference ESA CCI data and NASMD. NASMo-TiAM provides a curated soil moisture dataset and a flexible workflow with the potential for executing alternative machine--learning approaches with different sets of predictors.

WORKFLOW AND INPUT DATA COLLECTION

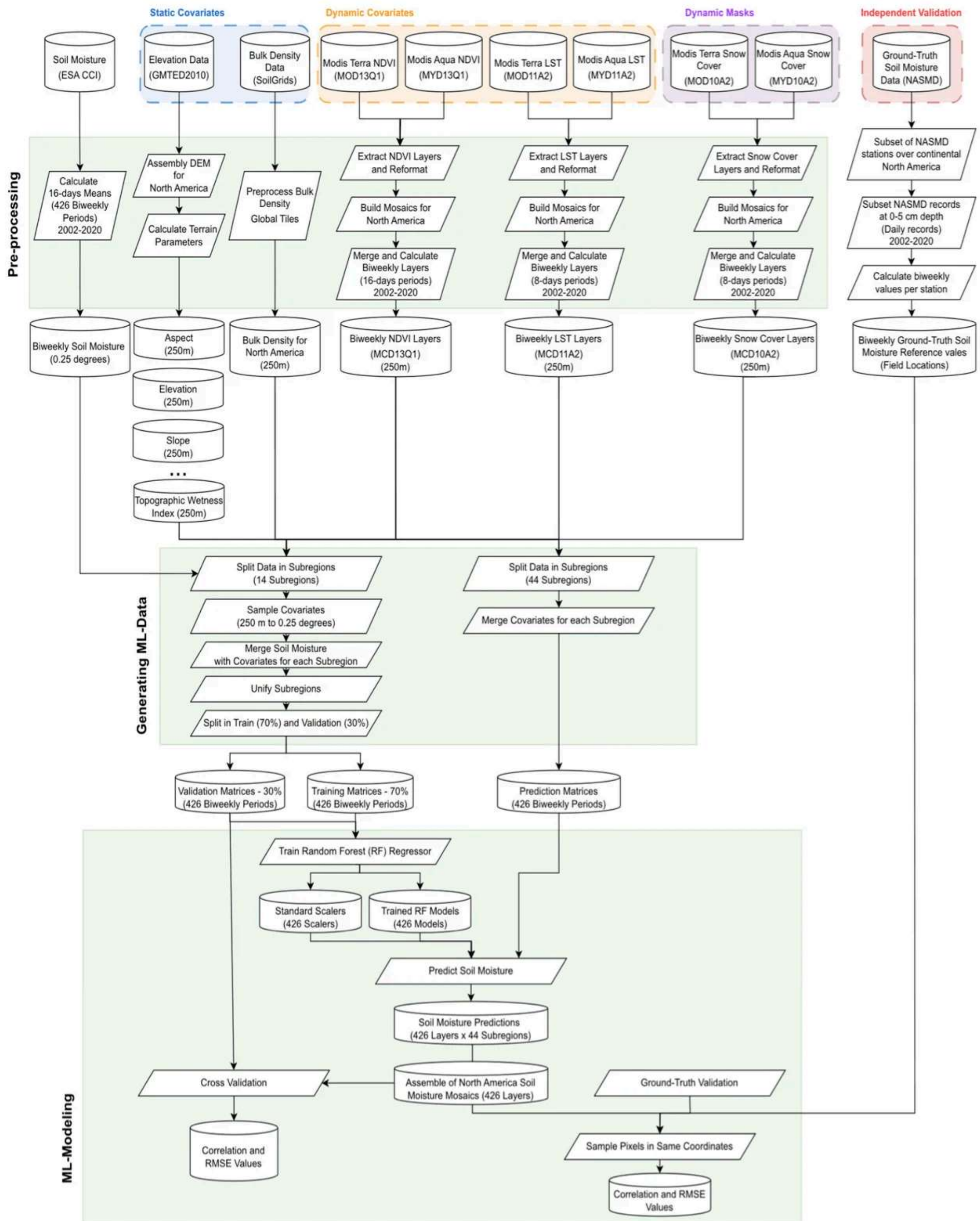


Figure 1 - Workflow for downscaling soil moisture

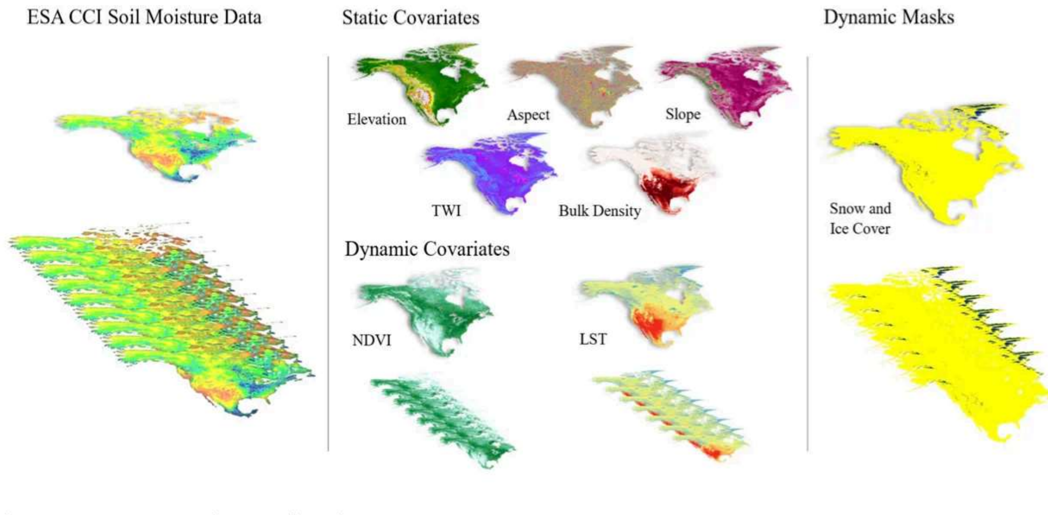
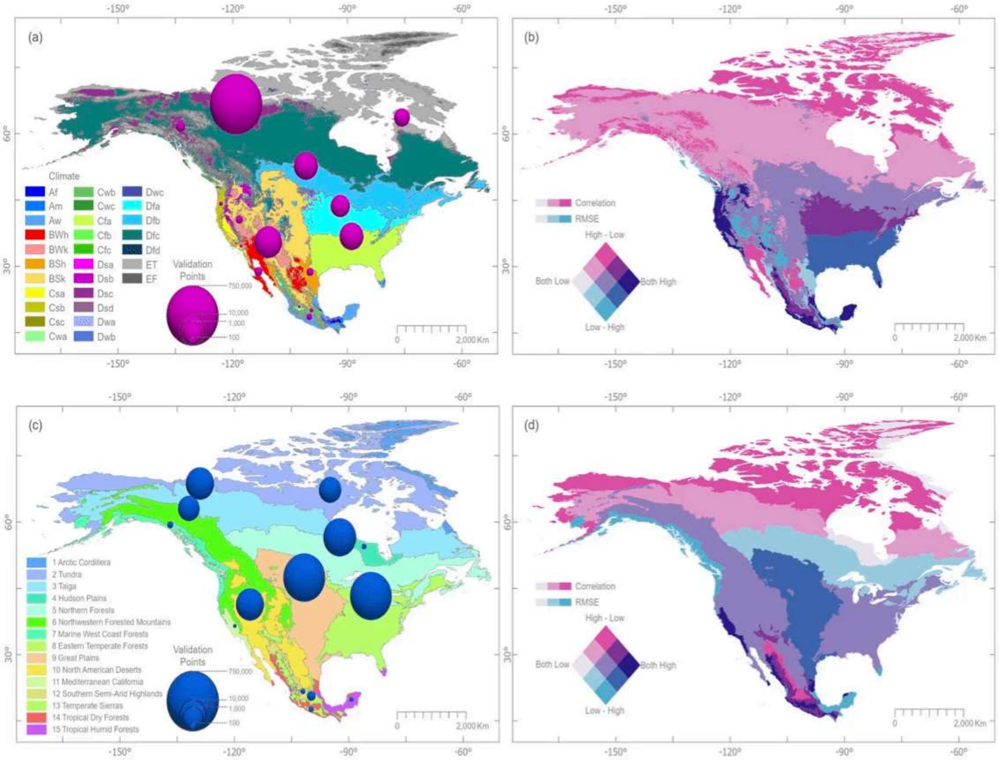


Figure 2 - Input data collection

DATA VALIDATION



Climate	r	RMSE
Af	0.85	0.03
Am	0.82	0.04
Aw	0.89	0.03
BWh	0.90	0.02
BWk	0.84	0.03
BSh	0.84	0.03
BSk	0.85	0.03
Csa	0.91	0.03
Csb	0.92	0.03
Csc	0.75	0.02
Cwa	0.90	0.03
Cwb	0.92	0.03
Cfa	0.85	0.03
Cfb	0.80	0.03
Cfc	0.65	0.03
Dsa	0.86	0.03
Dsb	0.88	0.03
Dsc	0.85	0.02
Dwa	0.86	0.03
Dwb	0.87	0.02
Dwc	0.86	0.03
Dfa	0.89	0.03
Dfb	0.85	0.03
Dfc	0.84	0.02
ET	0.89	0.02

Ecoregion	r	RMSE
1	0.81	0.02
2	0.91	0.02
3	0.84	0.02
4	0.82	0.02
5	0.78	0.03
6	0.85	0.03
7	0.78	0.03
8	0.85	0.03
9	0.87	0.03
10	0.88	0.03
11	0.89	0.03
12	0.94	0.02
13	0.92	0.03
14	0.91	0.03
15	0.83	0.04

Figure 3 - Cross-validation results between ESA CCI reference data and RF soil moisture predictions at 250 m. (a) Climate zones of North America and the total number of points used for validation along the study time frame (2002-2020); (b) Distribution of correlation and RMSE values per climate zone; (c) Terrestrial ecoregions of North America and the total number of validation points along the study time frame (2002-2020); (d) Distribution of correlation and RMSE values per terrestrial ecoregion.

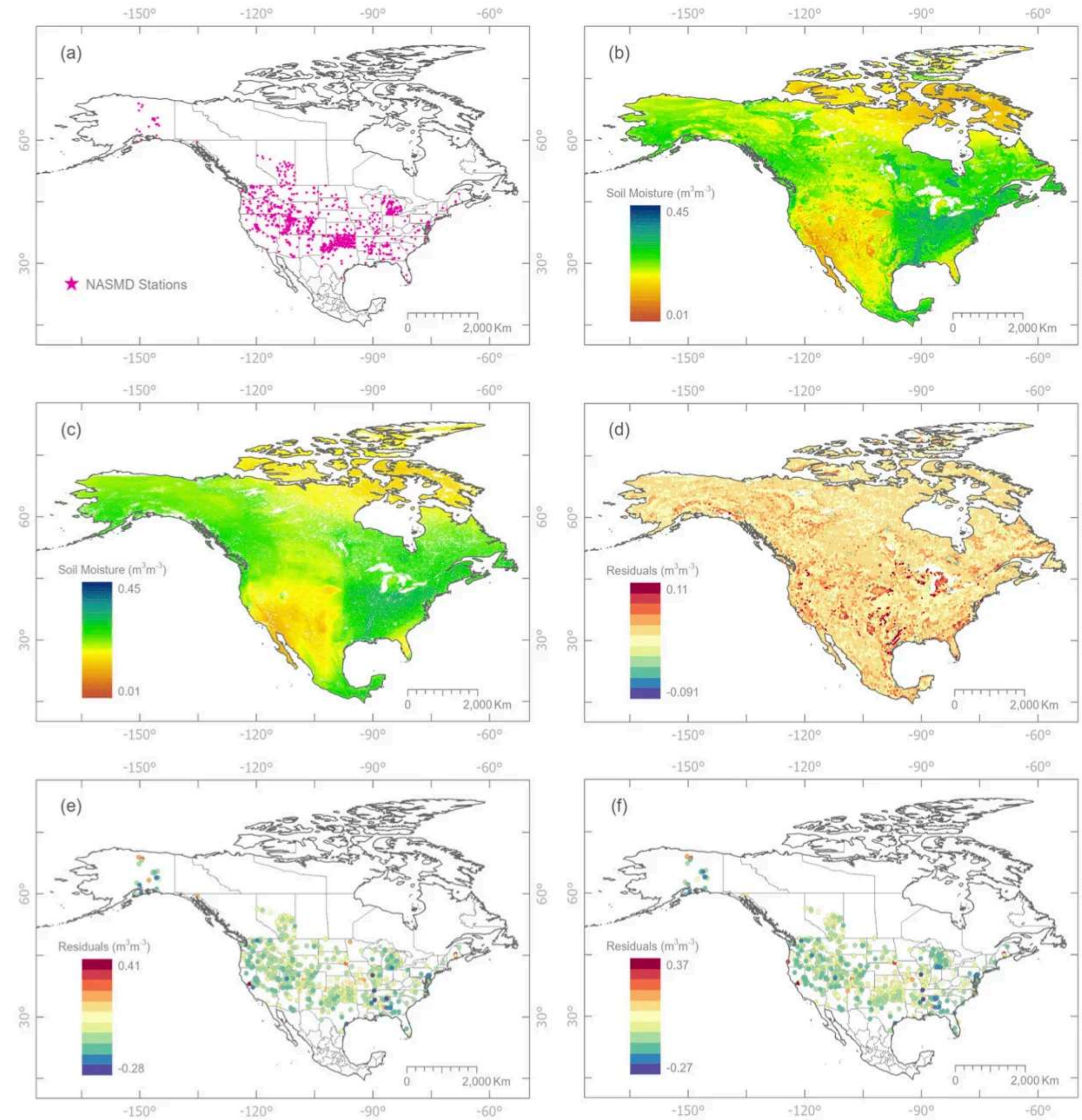


Figure 4 - (a) Distribution of 864 NASMD stations used for ground-truth validation; (b) ESA CCI mean soil moisture values derived from 426 biweekly layers; (c) RF predictions mean soil moisture values derived from 426 biweekly period; (d) Mean cross-validations residuals; (e) Mean residuals for 257 biweekly periods between ESA CCI and NASMD data; (f) Mean residuals for 257 biweekly periods between RF predictions and NASMD data.

NASMO-TIAM DATASET

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.

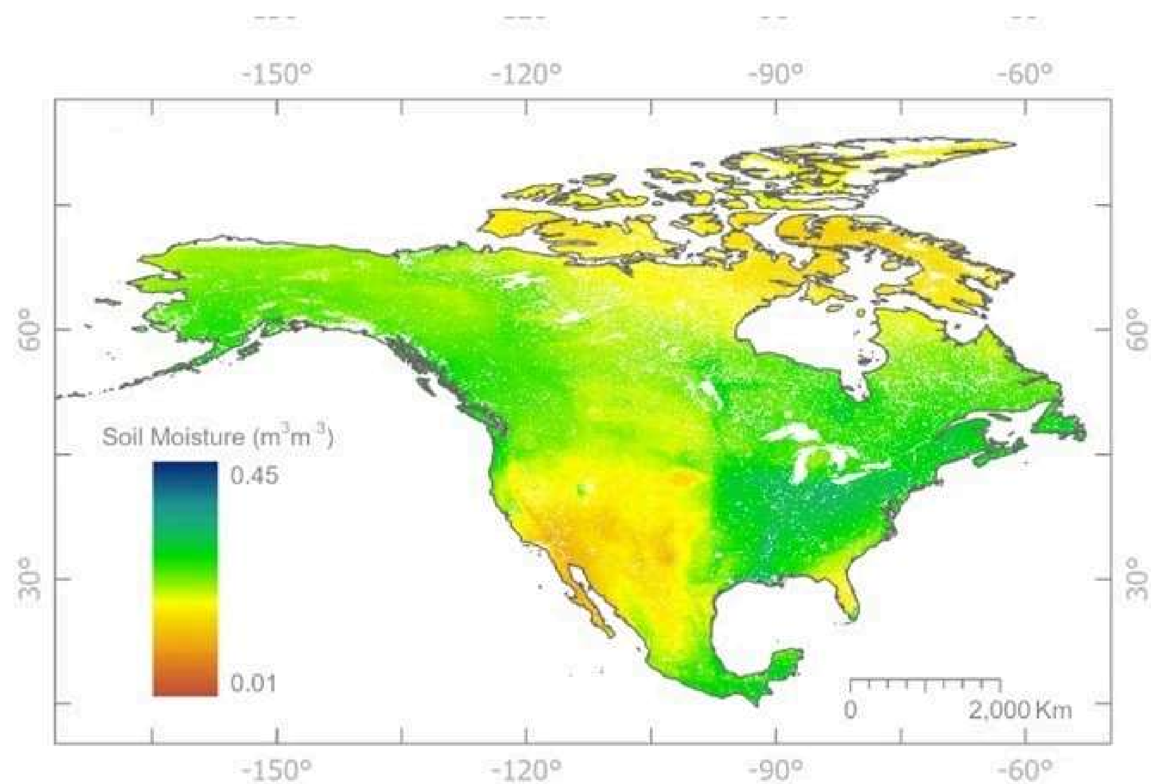


Figure 5 - Soil moisture across North America at 250 meters resolution biweekly.

DATA AVAILABILITY AND RELATED PUBLICATIONS

Data availability:

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

Related publications:

Guevara, M., M. Taufer, and R. Vargas. April, 27, 2021. Gap-free global annual soil moisture: 15 km grids for 1991–2018. Earth System Science Data:1711–1735.

Kitson, T., P. Olaya, E. Racca, M. R. Wyatt, M. Guevara, R. Vargas, and M. Taufer. 2017. Data analytics for modeling soil moisture patterns across united states ecoclimatic domains. Pages 4768–4770 2017 IEEE International Conference on Big Data (Big Data). ieeexplore.ieee.org.

Llamas, R. M., M. Guevara, D. Rorabaugh, M. Taufer, and R. Vargas. 2020. Spatial Gap-Filling of ESA CCI Satellite-Derived Soil Moisture Based on Geostatistical Techniques and Multiple Regression. Remote Sensing 12:665.

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.

McKinney, R., V. K. Pallipuram, R. Vargas, and M. Taufer. 2015a. From HPC performance to climate modeling: transforming methods for HPC predictions into models of extreme climate conditions. IEEE 11th International Conference on eScience:DOI 10.1109/eScience.2015.33.

McKinney, R., V. K. Pallipuram, R. Vargas, M. Taufer, and IEEE. 2015b. From HPC Performance to Climate Modeling: Transforming Methods for HPC Predictions Into Models of Extreme Climate Conditions. Pages 108–117 2015 Ieee 11th International Conference on E-Science.

Olaya, P., D. Kennedy, R. Llamas, L. Valera, R. Vargas, J. Lofstead, and M. Taufer. 2023. Building Trust in Earth Science Findings through Data Traceability and Results Explainability. IEEE Transactions on Parallel and Distributed Systems 34:704–717.

Roa, C., P. Olaya, R. Llamas, R. Vargas, and M. Taufer. 2023. GEOtiled: A Scalable Workflow for Generating Large Datasets of High-Resolution Terrain Parameters. Pages 311–312 Proceedings of the 32nd International Symposium on High-Performance Parallel and Distributed Computing. Association for Computing Machinery, New York, NY, USA.

Roa, C., M. Rynge, P. Olaya, K. Vahi, T. Miller, J. Griffioen, S. Knuth, J. Goodhue, D. Hudak, A. Romanella, R. Llamas, R. Vargas, M. Livny, E. Deelman, and M. Taufer. 2024. End-to-end Integration of Scientific Workflows on Distributed Cyberinfrastructures: Challenges and Lessons Learned with an Earth Science Application. Pages 1–9 Proceedings of the IEEE/ACM 16th International Conference on Utility and Cloud Computing. Association for Computing Machinery, New York, NY, USA.

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. Pages 1–10 2019 15th International Conference on eScience (eScience). arxiv.org.

TRANSCRIPT