Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-191-SC3, 2019

Interactive comment on "Gap-Free Global Annual Soil Moisture: 15km Grids for 1991–2016" *by* Mario Guevara et al.

Mario Antonio Guevara Santamaría

mguevara@udel.edu

Received and published: 31 December 2019

REVIEWER COMMENT: General comments: While seeking a higher resolution global soil moisture product is certainly a laudable goal, I find the methods used in this paper lack a credible connection between ground measurements and the remotely sensed data. The authors use machine learning to generate regressions between multiple topographic variables thought to be related to soil moisture that are available at higher resolution with coarse resolution satellite data. It is difficult for the reader to discern if actual new information results from the downscaling because there is not a clear connection made between insitu soil moisture measurements, their physical connection to the chosen topographic variables used, and resultant satellite measurements.

Printer-friendly version

Discussion paper

AUTHOR RESPONSE: We will improve the clarity of our methodological approach in a revised version of our manuscript to highlight why topographic terrain parameters are hydrologically meaningful. We would like to clarify that in situ observations are used for validating purposes only. In this study, the main purpose was to compare if the fusion of satellite soil moisture records and elevation-derived terrain parameters (by the means of an empirical modeling approach, not physical) was useful to fill gaps of satellite estimates. The physical connection between soil moisture and terrain attributes is that these attributes regulate the overland flow and the potential solar radiation income. Both process are controlled by topography and therefore should show a direct influence on soil moisture patterns. We will also include more information supporting the use of these hydrologically meaningful terrain parameters for predicting soil moisture patterns, an emergent research opportunity on Geomorphometry.

REVIEWER COMMENT: To do so, requires first demonstrating the rigor of the methodology over a much smaller and better measured area, such as the area of the International Soil Moisture Network.

AUTHOR RESPONSE: Our approach has already been tested across the contiguous United States (Guevara and Vargas, 2019). Thus, this study is an extension of that demonstrated methodology and now is applied to the global scale.

In a revised version of our paper, we will compare and test the accuracy of soil moisture predictions using country specific and global models to demonstrate the consistency of this approach across multiples scales of data availability. Using the North American Soil Moisture Database (Quiring, 2016) across the state of Oklahoma, in the Great Plains of the US (one of the areas with highest in situ soil moisture records for validating purposes, Llamas et al., 2019).

REVIEWER COMMENT: Extending the algorithm to areas that are hydrogeomorphically and climatically distinct from the existing soil moisture measurement networks cannot be result in credible data. I do not recommend publication of this work.

AUTHOR RESPONSE: As explained in previous work (i.e., Gessler et al., 2009, Florinsky, 2012, MacMillan et al., 2016), terrain parameters (such as those we used for predicting soil moisture) both influence the accumulation of surface geological materials and reflect this spatial distribution. Terrain parameters such as the wetness index or the relative slope position both influence the flow and accumulation of moisture and reflect it. Terrain parameters both influence the spatial patterns of distribution of vegetation/land use and reflect these patterns (MacMillan et al., 2016). Our methods are based on geomorphometry which differs from a purely physical hydrological model. That said the novelty of our study is that it takes the variables to a machine learning approach in the covariate space using as training data the satellite soil moisture pixels in the latest version of the ESA-CCI, which are soil moisture values representative of a mosaic of ecological and environmental conditions (for a specific time) within an area (pixel) of around 27km of spatial resolution. The strength of a machine learning model as kknn is to find non-linear relationships in complex spaces to recreate patterns.

These patterns are dependent of training data and there are sparse pixels with valid soil moisture data across dense vegetation areas or high latitudes that are useful to support the reliability of our soil moisture predictions (Fig 1). Replicating our models can give us information about modeling variance across areas with sparse training data . We re-analyzed our datasets and updated soil moisture predictions based on the most recent version of the ESA-CCI soil moisture product (to 2018) and included multiple sources of prediction factors (e.g., bioclimatic, soil type information) in order to maximize the reliability of our prediction framework and highlight the importance of terrain parameters modeling soil moisture.

In a revised version we will use data from the ISMN and other published values of soil moisture to expand our validation dataset. These new validation analyses can be performed across different biomes rather than at the global scale.

REVIEWER COMMENT: Major comment: Much more information is needed about the regression methods and parameter selection.

AUTHOR RESPONSE: We will improve the narrative of the kernel based machine learning algorithm we used to perform regression. The main parameters of this method (kknn) are the kernel type (used to convert distances between neighbors points in the statistical space, to weights) and the k parameter (the number of neighbors used to calculate a weighted average in regression). These parameters are selected in our modeling approach automatically using repeated cross validation (Table 1 of submitted paper). The terrain parameters derived by the means of Geomorphometry from the digital elevation model are another component of our modeling framework, used prediction factors (soil moisture covariates) in the kknn approach.

REVIEWER COMMENT: It would be helpful to discuss what each parameter is mathematically and why it can be useful in a soil moisture prediction context.

AUTHOR RESPONSE: We will explain with more detail why the terrain parameters derived from elevation data using Geomorphometry are hydrologically meaningful, and will include also mathematical formulations to identify the main impacts of topography in soil moisture. Detailed information on these terrain parameters is reported in our previous study across the conterminous United States (see table: https://doi.org/10.1371/journal.pone.0219639.s007).

REVIEWER COMMENT: Can it be demonstrated over smaller areas that there is a valid argument for using these variables, some or all of them.

AUTHOR RESPONSE: We tested our approach across continental and state scales and found consistent results. Across the conterminous US, we recently found an increase of nearly 25% of accuracy validating our predictions against the North American Soil Moisture Database (https://doi.org/10.1371/journal.pone.0219639.g005).

REVIEWER COMMENT: Explain in detail the cross validation process used and why.

AUTHOR RESPONSE: We will include more information and references about cross validation strategies for re-sampling and bootstrapping prediction models. Cross validation is a family of re-sampling techniques that are used to analyze the sensitivity of models (in this case) to variations of available data for training purposes. Multiple models are generated using multiple proportions of available data and validating with subsets of information that are leaved out of these models. This process is repeated several times until capturing the magnitude of variance associated with the models based on the data subsets.

REVIEWER COMMENT: It is unclear if there is a single model developed and applied to all years or separate models for each year. Are all topographic parameters used in the model in all years? What is their weighting?

AUTHOR RESPONSE: We generate and cross validate a model for each year of available data. Table 1 shows the results of cross validation and available data for each model/year. All topographic parameters are used on each model/year. During the cross validation, kknn uses a kernel form to convert distances in weights that are used for prediction. These distances are different from one place to another as the pattern recognition includes all the points and their k neighbors. Using the independent residuals of each model realization during the cross validation strategy, the optimal weights are selected using as indicators the root mean squared error and the correlation between observer and predicted (also included in Table 1 of submitted manuscript). After all model realizations including all possible kernel combinations (e..g, "rectangular", "triangular", "epanechnikov","gaussian", "rank", "optimal") and increasing the number of neighbors (k) systematically (e.g., from 2 to 25, See Table 1) we find for each year the optimal weights maximizing the correlation between observed and predicted and minimizing the root mean squared error. In our reanalysis, we have included a novel variable importance analysis (by permutation) for kknn, that allows to identify which are the most important variables in the overall prediction using this kernel based algorithm.

REVIEWER COMMENT: Major comment: Extending the results to areas without soil moisture measurements (whether in situ or remotely sensed values) without any validation is not an improvement over current spatial data. It is highly suspect and could lead to inappropriate applications using what is in effect non-data.

AUTHOR RESPONSE: We agree that any modeling output should be interpreted and used carefully as it is not a direct measurement. However, we performed a robust cross validation strategy including data available in the ESA-CCI representing all environmental conditions and provide information about data-model agreement. Then we validate with field data only at the places (pixels) containing in situ information in the International Soil Moisture Network. It is clear that contiguous information on soil moisture is challenging to obtain across large areas of the world such as polar or tropical rain forest conditions. However there are limited and sparse but representative satellite soil moisture records across these Polar (Fig 1a) or tropical areas (Fig 1b) that can be used to improve the spatial representation of global soil moisture patterns.

We will include in our reanalysis more datasets (Fig 1c) and more validation information to enrich the discussion about the reliability of our prediction framework for soil moisture (e.g., Bond-Lamberty and Thomson). Please note that we are following a conceptual and data-driven framework assuming that comparing and testing multiple modeling approaches is still required to better understand the implications of soil moisture on ecological patterns across areas where no soil moisture information is available. Thus, our results provide some evidence that our soil moisture product is useful to better understand ecological patterns in places with low availability of satellite soil moisture estimates.

REVIEWER COMMENT: Moderate comment: I found the paper very repetitive and in need of detailed editing.

AUTHOR RESPONSE: We will revise narrative and overall organization of the manuscript.

References

Bond-Lamberty, B.P., and A.M. Thomson. A Global Database of Soil Res-

Printer-friendly version

Discussion paper

piration Data, Version 4.0. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1578, 2018 Florinsky, I. V.: Influence of Topography on Soil Properties, Digital Terrain Analysis in Soil Science and Geology, 145–149, doi:10.1016/b978-0-12-385036-2.00008-0, 2012.

âĂŇGessler, P., et al., Chapter 28 The Future of Geomorphometry, Developments in Soil Science, 637–652, doi:10.1016/s0166-2481(08)00028-7, 2009.

Guevara, M. and Vargas, R.: Downscaling satellite soil moisture using geomorphometry and machine learning, edited by B. Poulter, PLOS ONE, 14(9), e0219639, doi:10.1371/journal.pone.0219639, 2019.

Llamas, R. M., Guevara, M., Rorabaugh, D., Taufer, M. and Vargas, R.: Spatial Gap-Filling of ESA CCI Satellite-Derived Soil Moisture Based on Linear Geostatistics, , doi:10.20944/preprints201909.0126.v1, 2019.

âĂŇâĂŇâĂŇMacMillan et al., A hydrologically explicit, spatially exact, classification of landforms for Canada at 1:500,000 scale., EGUGA, EPSC2016-13103 https://ui.adsabs.harvard.edu/abs/2016EGUGA..1813103M/abstract (Accessed 23 December 2019), 2016.

Quiring, Steven M., Trent W. Ford, Jessica K. Wang, Angela Khong, Elizabeth Harris, Terra Lindgren, Daniel W. Goldberg, and Zhongxia Li. 2015. "The North American Soil Moisture Database: Development and Applications." Bulletin of the American Meteorological Society, November. doi:10.1175/BAMS-D-13-00263.1. âĂŇ

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-191, 2019.

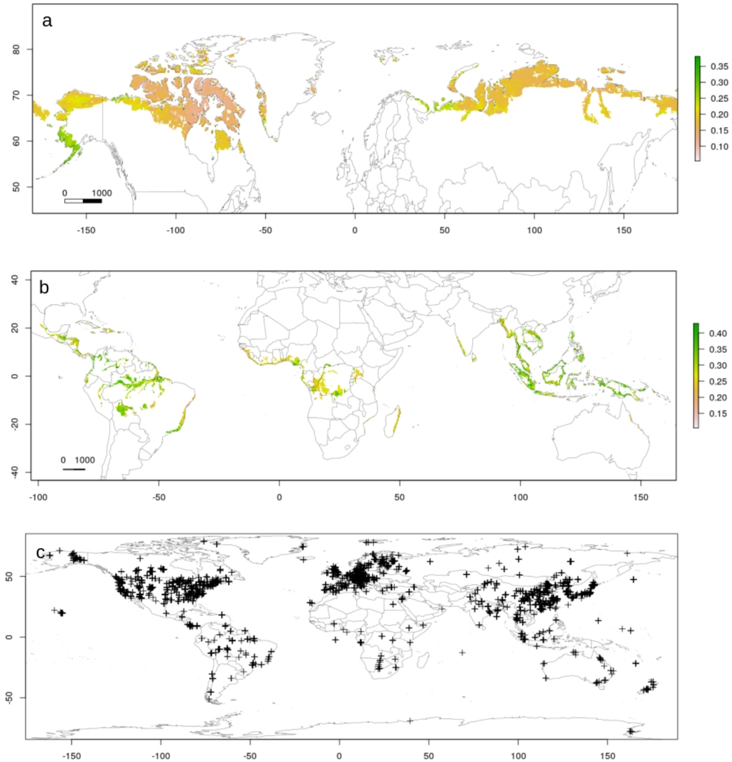

Fig. 1. Available information in the ESA-CCI (4.5, 2018) across Polar (a) and dense vegetated areas and tropical rain forests (b). In situ dataset that we will use to improve our validation effort (c).

Printer-friendly version

Discussion paper