Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-191-SC2, 2019

Open Access Earth System Science Data Discussions

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: OVERVIEW The study has developed a gap-filled, and downscaled with topography-derived information, long-term annual (15 km) global soil moisture dataset based on the ESA CCI satellite soil moisture products. The assessment of the dataset with respect to in situ observations has been carried out through annual comparisons as well as in terms of long-term trends (1991-2016).

REVIEWER COMMENT: GENERAL COMMENTS The paper is mostly well written and clear. The topic of the paper is interesting for the readership of Earth System Science Data as a global scale gap-filled annual soil moisture dataset is surely useful for many applications. However, I believe the paper needs major changes before the publication

as several parts are not properly described and other sections need to be improved or summarized. I have listed below my comments with the indication of their relevance.

AUTHORS RESPONSE: We appreciate the reviewer comments. We can clarify and improve the narrative of our methods in a revised version.

REVIEWER COMMENT:1) MAJOR: The factors used in the downscaling and gap-filling algorithm should be described in details. Two figures (Figures 2 and 3) are not mentioned in the text. It seems a part is missing. The reader needs to know the details on the methodology employed and which factors have been found to be more important. Which Digital Elevation Model is used?

AUTHORS RESPONSE: We fully agree with this comment and will include this information in the revised version of our manuscript.

REVIEWER COMMENT: Additionally, other static factors such as vegetation and soil types are not considered. Why?

AUTHORS RESPONSE: Our initial objective was to test the predictive capacity of topographic terrain parameters derived from a single source of information (elevation), considering that a satellite soil moisture pixel is representative of soil moisture of the spatial configuration of climate and ecological conditions (within each pixel) for a specific period of time.

In a revised version we will update our soil moisture prediction framework and include (as prediction factors) bioclimatic and soil classes static information. In a new version of our paper we will compare the predictive capacity of topographic patterns in relation to these bioclimatic and soil classes.

REVIEWER COMMENT: The discussion on the approach employed for downscaling and gap-filling needs to be included in the paper. Why do the authors select such approach?

AUTHORS RESPONSE: We will include more information on model selection in the revised version of our manuscript. Briefly, we used a kknn algorithm because it is fast and yielded reasonable good accuracy (r=>0.8 in all models) in comparison with other modeling approaches that are computationally more expensive. Algorithms such as Random Forests or Support Vector Machines are examples of conventional machine learning methods that can be used as regressors for satellite soil moisture gridded surfaces. We decided to use kknn to generate a baseline of predictions that can be reproduced in hours (annual 1991-2018) using conventional laptops (i.e., 6GB of RAM) in order to increase the users of soil moisture information.

In addition in a revised version we will include a bootstrapping technique applied to the kknn prediction framework to account for the sensitivity of the model to variations in datasets. Also, detailed information on variable importance for kknn models will be included in the revised version as well as an appendix with a performance comparison against other prediction algorithms.

REVIEWER COMMENT: 2) MAJOR: I have found particularly challenging performing gap-filling over dense forest regions in the Amazon and in Congo. Satellite soil moisture data cannot be used in such regions due to dense forest that mask the soil moisture signal. How is it possible to extend the signal there only based on topography?

AUTHORS RESPONSE: There are sparse pixels with soil moisture data across specific areas of the tropical rain forest such as the Amazon or Congo that are useful for modeling and calibrating soil moisture predictions using our approach (Fig 1). In a revised version we will include new prediction factors to account for ecological and climate variability (e.g., bioclimatic variables).

Soil type information (e.g., harmonized world soil database) could also be included to account for the capacity of soil to retain water across these areas with low availability of soil moisture data in the ESA-CCI soil moisture product. We will include a comparison of model performance using multiple combinations of prediction factors. We expect that with a better description of the methods and this revised approach we will provide

a better explanation of our alternative approach to downscale soil moisture across the world.

REVIEWER COMMENT: I would suggest to perform a more detailed validation in these areas. I strongly suggest to perform a comparison with modelled datasets (e.g., ERA5 soil moisture) to have an assessment of the performance over dense vegetated areas. A similar comment can be done for high latitude areas in which frozen soils and snow completely mask the soil moisture signal. Please perform a detailed validation over these areas, too.

AUTHORS RESPONSE: We appreciate the comment and we will improve our validation across these areas (e.g., tropical forests or high latitude areas) using in situ records of temperature, precipitation and ecological patterns synthesized in previous work (Bond-Lamberty and Thomson, 2018). For example, previous studies have described the coupling between soil moisture and precipitation (Koster, et al., 2004, McColl et al., 2017).

REVIEWER COMMENT: 3) MAJOR: The trend analysis is very interesting. However, as above, we need more details on how trends are computed. For instance, in situ stations are available only at some points over the Earth, are the same locations used with the satellite-derived datasets? If not, the comparison is wrong.

AUTHORS RESPONSE: We will improve the description of the methods in a revised version of our manuscript. We clarify that the comparison was done using only pixels where there was a spatial match with the sites of the ISMN.

REVIEWER COMMENT: Similarly, in situ stations are not available every year, and for the full year. How are the data aggregated in time and space? These details are needed. I expect the results are strongly impacted to these choices.

AUTHORS RESPONSE: We recognize that there is limited soil moisture field information for validating models and satellite soil moisture estimates across large areas of the world. We will provide more information about soil moisture (historic and current) data availability in the ISMN and its use for this study. We will also provide more detailed information on how this dataset was aggregated.

REVIEWER COMMENT: 4) MODERATE: The machine learning downscaling approach provides soil moisture data with a resolution higher than the original ESA CCI product. However, I am always doubtful on these downscaling approaches as instead of resolution it should be higher spatial sampling. The higher spatial resolution should be tested, but I am aware it is very hard to do (I have this comment for all downscaling studies).

AUTHORS RESPONSE: We fully agree with the reviewer comment and we recognize that there is a compromise between where and when to sample across scales. We also recognize that all global studies are limited with the available information of global networks and local studies (across multiple ecosystems and regions of the world) are needed to better test downscaling approaches. In a revised version we will include an improved validation including new information data across poorly represented areas.

REVIEWER COMMENT: The authors should demonstrate that the downscaled product is able to reproduce features at higher resolutions with respect to the parent ESA CCI product. It is not done in the paper, that's why I believe higher spatial sampling, and not spatial resolution, is more appropriate.

AUTHORS RESPONSE: We found a larger range of soil moisture predicted values compared with the original ESA-CCI soil moisture product. We found a temporal trend at the places of field stations that is more similar between the field data and our predictions compared with the ESA-CCI soil moisture product. Please note that the main purpose of the model is to reproduce the signal of satellite soil moisture using as reference the relationship that it maintains with topographic data. This is a regression problem were the satellite soil moisture measurements (for a specific time across an area, a pixel under a approximately the same vegetation type or general climate con-

dition) are statistically related to multiple quantitative topography surrogates.

We believe that a spatial resampling (Fig 2a) is just a change of spatial resolution by using simple algorithmic approaches across the orthogonal relationship of the variable itself with the latitude and longitude plane. In contrast, our models (Fig 2b) are replicated and there is a statistical learning process on each iteration for the selecting optimal model parameters and minimizing the prediction error given a specific spatial resolution defined by the topographic prediction factors. We understand that this represents a conceptual and semantic debate and we will elaborate in a revised version.

REVIEWER COMMENT: 5) MODERATE: Several performance scores have been used in the paper. However, I don't think it is necessary to use all of them. The authors should discuss what information each performance score is providing for the assessment of the dataset, not simply to list many numbers. Indeed, Tables 2 and 3 are hard to read and not informative. Please summarize only the more relevant scores in a figure.

AUTHORS RESPONSE: We can summarize the accuracy numbers of these reports and we can improve the discussion around them. We will also include a figure (i.e., from multivariate analysis) showing the main differences in the revised version of our manuscript.

REVIEWER COMMENT: 6) MAJOR: The range of values of ESA CCI soil moisture products has little value, as the satellite products are rescaled to match the range of variability of modelled soil moisture from GLDAS. Therefore, the range of values is that obtained from GLDAS. For the analysis shown in Figure 4, and similarly for the trend analysis, the soil moisture datasets should be rescaled between the minimum and maximum of each time series and expressed as relative soil moisture (between 0 and 1). Then the data should be aggregated and the range of values and the trends can be assessed.

AUTHORS RESPONSE: We can transform the values of the soil moisture maps to relative soil moisture and repeat the trend analysis and the comparison with field data. We will update trend results in a revised version of our approach.

7) MODERATE: I believe the discussion section must be rewritten. General results are mostly discussed, whereas it should be closely related to the results shown in the paper. I believe it should be shorter and better focused.

AUTHORS RESPONSE: We can improve the narrative of our discussion and main findings.

REVIEWER COMMENT: SPECIFIC COMMENT (L: line or lines) L307: Why the "angle between satellite sensors and the earth surface" is useful for determining soil moisture? It has no sense and I believe it is wrong.

AUTHORS RESPONSE: We meant to say that topography affect the distance between the satellite and the earth surface and therefore It could be correlated with the satellite soil moisture signal (which is a hypothesis proven with the data analyzed in this study).

REVIEWER COMMENT: RECOMMENDATION Based on the above comments, I suggest a major revision before the possible publication on Earth System Science Data.

AUTHORS RESPONSE: We appreciate the comments of the reviewer that will improve the overall revised manuscript.

References:

Bond-Lamberty, B.P., and A.M. Thomson. 2018. A Global Database of Soil Respiration Data, Version 4.0. ORNL DAAC, Oak Ridge, Tennessee, USA.Âăhttps://doi.org/10.3334/ORNLDAAC/1578

McColl, K. A., Alemohammad, S. H., Akbar, R., Konings, A. G., Yueh, S. and Entekhabi, D.: The global distribution and dynamics of surface soil moisture, Nature Geoscience, 10(2), 100–104, doi:10.1038/ngeo2868, 2017.

Koster, R. D.: Regions of Strong Coupling Between Soil Moisture and Precipitation,

Science, 305(5687), 1138–1140, doi:10.1126/science.1100217, 2004.

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

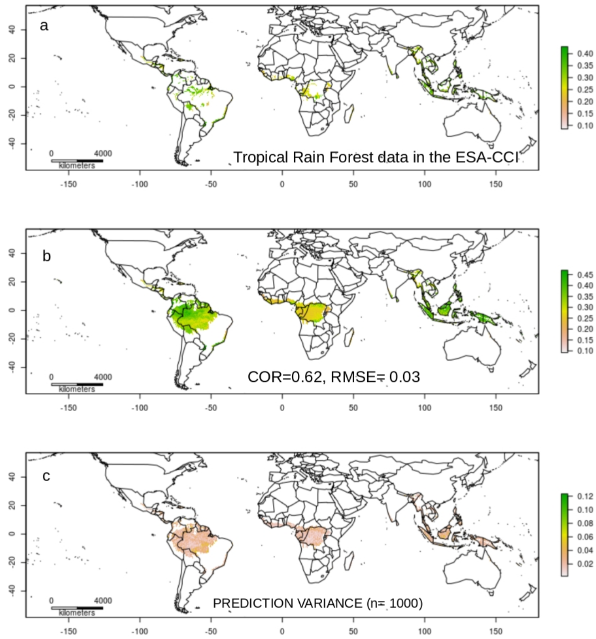

Fig. 1. Training data for 2018 across Tropical Rain Forests (a), the model prediction for 2018 (b) and the prediction variance after multiple realizations for 2018 (c).

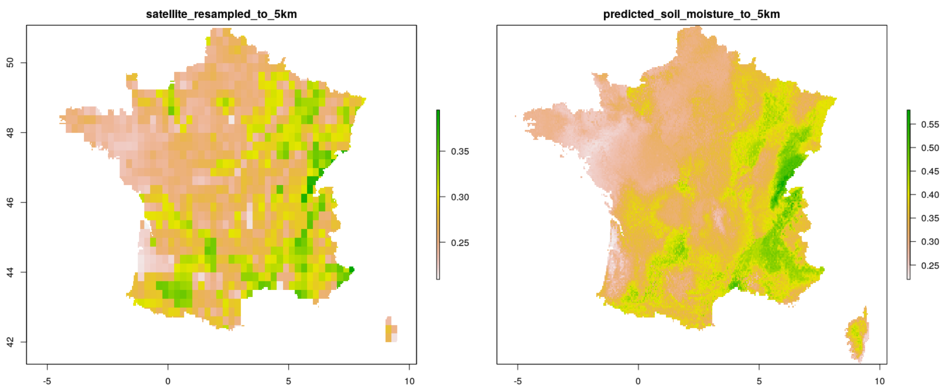

Fig. 2. Simple spatial resampling (bilinear) applied to the ESA-CCI product from 27 to 5 km grids (a) and our modeling output using 5km grids (b) across France.