"Gap-Free Global Annual Soil Moisture: 15km Grids for 1991–2016" by Mario Guevara et al. Responses to anonymous reviewer #1

REVIEWER COMMENT: This paper presents a 15 km annual-average soil moisture product that is generated by machine-learning the relation between 0.25 degree ESA CCI soil moisture estimates and topographic indices derived from a higher-resolution DEM.
I have several major concerns regarding the hypothesis/assumptions on which the methodology is based as well as the employed validation methodology, and consequently also the conclusions drawn from the presented analysis:

AUTHORS RESPONSE: We appreciate the reviewer comments as they provide valuable feedback to increase the impact of our work. We have revised our work and provided new analyzes to clarify our framework and improve the validation methodology. We have updated our work with new datasets (e.g., eco-climatic and soil type classes) and updated our results with the recently released version of the ESA-CCI soil moisture product (4.5, up to 2018). These results have been uploaded as an updated version of our data repository. We clarify that the original data-product has not substantially changed, but we have added new model outputs and further analyzes to address the concerns from the reviewers.

To increase reproducibility and transparency of our prediction framework we compiled a set of R functions (link available with the new version of the manuscript). This compilation is useful to predict soil moisture based on remote sensing and machine learning, from the global to country-specific scales. Furthermore, we now include a bootstrapping approach given a user defined sample size (if not specified, it will use 1/3 of available data/pixels for each year) to analyze the variance of model predictions as a function of variations in training data. Thus, in the revised version of the manuscript we report a spatial explicit metric of model-based uncertainty for the fusion of soil moisture satellite estimates and topographic constraints. We believe that these advances (including new model estimates and uncertainty estimates) constitute a substantial improvement from the previous version of the manuscript.

REVIEWER COMMENT: The methodology is based on the hypothesis that topography is a main driving factor for soil moisture patterns. However, the reference used to support this claim (Mason et al., 2016) presents only a very local analysis of differences between soil moisture values at low-slope and high-slope areas over grasslands only, and only in a small region over the UK. The observed relation is relatively low (Rˆ2 = 0.21) and the authors conclude "[...] a topographic signal can be seen in high resolution remotely sensed surface soil moisture data [...]. Unfortunately this signal is relatively weak." Moreover, Mason et al. (2016) uses 1 km SAR data for which topographic corrections are applied in the pre-processing, which likely induces a sensitivity of the measurements to topography parameters. These topography corrections are usually not applied to coarse resolution measurements such as the sensors used within the ESA CCI SM, because topographic effects average out at these scales.

AUTHORS RESPONSE: We have improved the introduction with more references and examples. We also clarified the main hypothesis driving this effort (i.e., that topographic terrain parameters are good predictors for downscaling satellite-derived soil moisture).

We argue that satellite-derived soil moisture data is retrieved by a direct measurement of the dielectric constant of soils that is representing specific vegetation types and climate conditions (within each pixel) that are intrinsically influenced by topographic patterns. We highlight that there is a high correlation (Table 1) between topographic patterns and the ESA-CCI soil moisture product, which provides evidence that topography has information that could be used for downscaling satellite soil moisture. Our machine learning approach is able to reproduce non-linearities between satellite-derived soil moisture and topographic variables, and generate predictions of soil moisture across higher resolution spatial grids.

In the new version we compared the prediction capacity of multiple prediction factors of soil moisture (topographic terrain parameters, bioclimatic features and soil types) to identify the value of terrain parameters predicting soil moisture. We highlight that our main purpose was to generate a soil moisture downscaled product that is independent of climate or vegetation variables and that provides cross validated hypothesis of a more local (nearly 50% improvement) spatial resolution compared with the original satellite soil moisture signal. In the updated version of our soil moisture datasets, we explicitly account for the spatial relationships of soil moisture available data and the geographical space following a digital soil mapping spatial prediction framework (McBratney et al., 2003, Hengl et al., 2018, Møller et al., 2019) to increase the reliability of our predictions. Finally, we clarify that the use of a soil moisture product where vegetation or climate data are not used as predictors could be relevant for decreasing potential spurious correlations in the further use of our soil moisture data-product.

REVIEWER COMMENT: The presented paper itself also does not analyze the predictive power of the used topographic indices for soil moisture (e.g., the godness-of-fit for the obtained regression, variable importance, etc.). Hence, there is no evidence supporting the reliability of topographic indices as predictor for soil moisture, especially on a global scale.

AUTHORS RESPONSE: We highlight that we have provided information about the predictive power of our models. In Table 1 we reported indicators of accuracy and predictive power including the correlation between observed and predicted, the root mean squared error, the number of pixels with soil moisture data available for each year and the best parameters (kernel and k) for each kknn model/year. In the revised version, the predictive power of our modeling approach increased thanks to the use of a more extensive set of prediction factors. We have also bootstrapped our statistical models and now they provide uncertainty estimates associated with the number of available data for modeling soil moisture patterns.

In the revised version we include a simple variable importance analysis (by permutation) to support/generate new hypothesis about the spatial variability of soil moisture, in relation to land surface characteristics represented by multiple sources of environmental information. In addition, we provide a bootstrapping approach to estimate uncertainty associated with the number of available data and predictors for modeling soil moisture patterns.

REVIEWER COMMENT: Even more doubtable is the assumption that the developed regression function can be used to extrapolate soil moisture to regions not covered by the ESA CCI SM, which are mainly the arctic ice sheet and tropical forests. Tropical rainforests, for example, have a quite unique moisture regime that is expected to be largely rainfall dependent. It is very questionable to use a soil moisture - topography relation that is trained over non-tropical regions to predict soil moisture there. Moreover, no in situ measurements are available in these regions to verify the validity of these predictions.

AUTHORS RESPONSE: We agree that soil moisture estimates across the arctic ice sheet are not informative; therefore, these areas have been removed from our analysis and the model outputs in the revised version.

We recognize that in-situ global soil moisture field datasets are limited for validating our soil moisture predictions across tropical areas, but we have done efforts to bring more available in-situ measurements in this revised version. That said, we demonstrate the reliability of our predictions by comparing the original satellite and our predictions against in-situ records of rainfall patterns across the tropical areas of the world (171 sites across tropical areas of the world for the years 2008 to 2018). We now report and improvement with our predictions and in-situ precipitation records across multiple biomes of the world (e.g., r=031 to r=0.38 across tropical biomes and r=40 to r=51 across temperate biomes) using information from previous studies.

We argue that our approach can be used for predicting soil moisture (and associated uncertainty) across tropical forests. We highlight that our framework is an empirical approach using machine learning algorithms that are able to reproduce patterns extracted from a multivariate space. We highlight that there are pixels with satellite soil moisture values across the tropical rain forests of the world, including the Amazon and Congo regions that could be used for prediction within our framework (Figure R1_1).

In the revised version we compare a model using terrain parameters as prediction factors of soil moisture and a model including information on bioclimatic features (FAO, 2010, Fick and Hijmans, 2017) and soil type variability (Wieder et al., 2014). We use geo-spatial information of available data/pixels in order to explicitly account for soil moisture spatial relationships between soil moisture available values for training the models. Finally, we report uncertainty estimates to provide spatial information about model performance and to identify regions with high model uncertainty. We belive that this new addition (reporting uncertainty) is a much needed effort in local-to-global predictions.

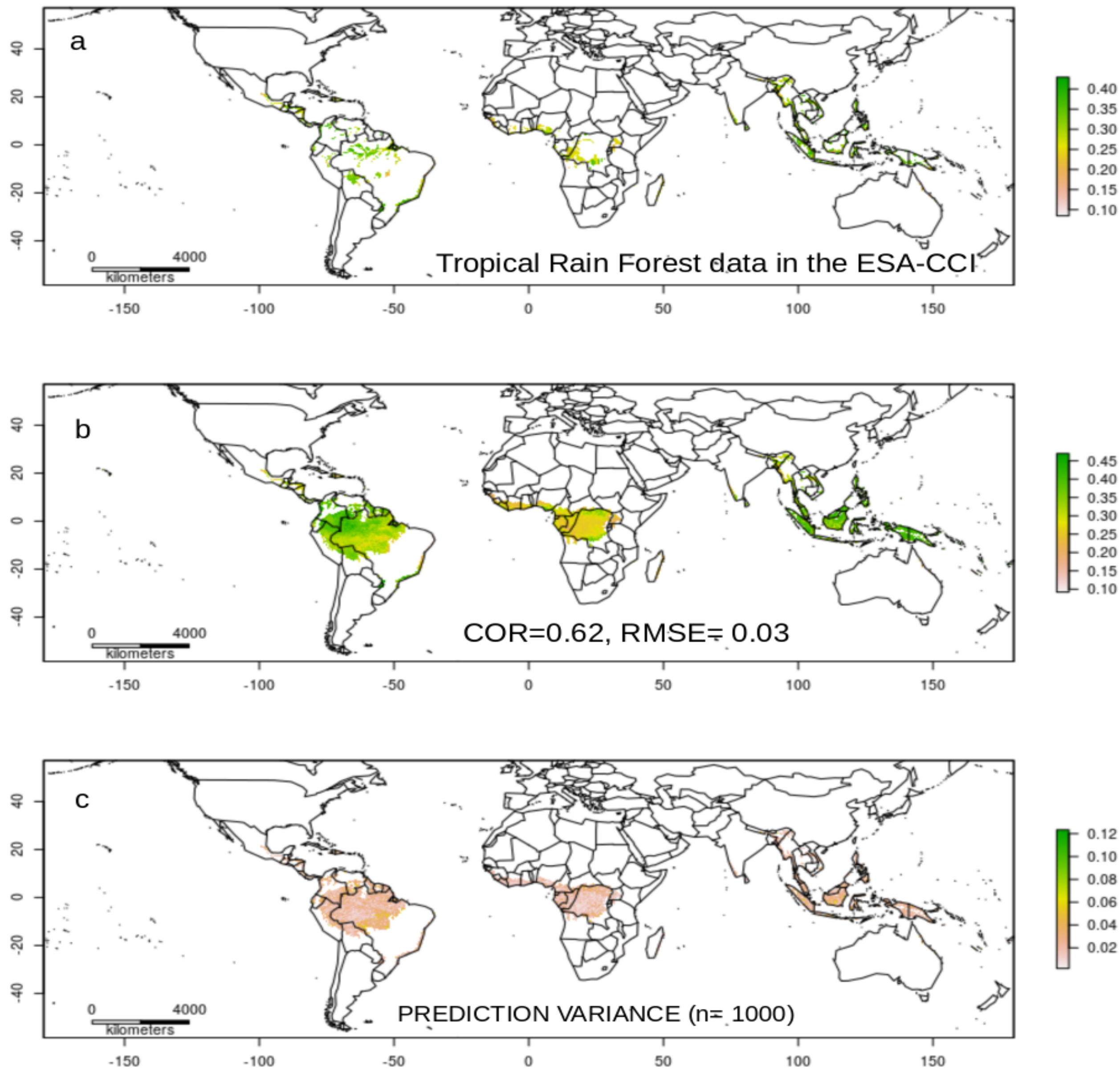

Figure R1_1Soil moisture across Tropical Rain Forests of the world based on the data available in the ESA-CCI soil moisture product (4.5) for the year 2018 (a). We show the soil moisture prediction (b), the soil moisture prediction variance using only the data available for Tropical Rain Forests (c). Note that the correlation between observed and predicted decreased to 0.62, most likely due to the limited information for modeling these ecosystems, however the root mean squared error is comparable with a model using all global data (e.g., <0.04).

In the revised version we compare a model using terrain parameters as prediction factors of soil moisture and a model including information on bioclimatic features (FAO, 2010, Fick and Hijmans, 2017) and soil type variability (Wieder et al., 2014). We use in our modeling approach geo-spatial information of available data/pixels in order to explicitly account for soil moisture spatial relationships between soil moisture available values for training the models.

REVIEWER COMMENT: Also, the presented validation does not support the conclusions. First, the statement (L234) "In all cases, the evaluation statistics are equal or better for the downscaled soil moisture predictions based on digital terrain analysis (Table 3) than the original ESA-CCI soil moisture product (Table 2)" is wrong. In fact, results are quite balanced, sometimes the downscaled product is "better", sometimes the original is "better", but most likely results are not actually distinguishable within reasonable confidence limits (which should be estimated). The authors do indeed acknowledge (L252): "The downscaled predictions based on digital terrain analysis are not significantly different compared with the ESA-CCI soil moisture product [...]", but the subsequent conclusions are not supported. Specifically, "[...] but they provide (1) gap free soil moisture-related information ": while they are provided, there is no evidence that they are of any reasonable accuracy (for the earlier discussed regions), and "(2) higher resolution (from 27 to 15 km grids)": This is a mix-up of resolution and sampling.

AUTHORS RESPONSE: We highlight that this is an empirical model (i.e., a statistical learning process driven by a machine learning algorithm) and we used a cross validation to show the correlation of the change of resolution between the original satellite estimate and the prediction; thus it is not only a spatial re-sampling exercise. It is an empirical model that relies on satellite soil moisture data and its spatial relationships with topographic data. Then, independent in-situ data is used to validate the model output.

The reviewer is correct to point that the validation against filed information are balanced between the CCI-ESA product and our downscaled product. There are no significant differences in the relationship between the downscaled estimate and the original product; therefore, this demonstrates that the downscaled product is preserving the statistical properties of the ESA-CCI soil moisture product. This is expected as the information that was used for building predictions of soil moisture were the pixel/values of the ESA-CCI soil moisture product without adding any in-situ information (again, the in-situ data was only used for independent validation). In the revised version we have improved the narrative and the explanation on this comparison for clarity.

Our results comparing the predictive capacity of terrain parameters on soil moisture (in relation to bioclimatic and soil type features), suggest that terrain parameters are useful to fill-gaps in the ESA-CCI and maintain its accuracy against field observations. We now include a full model (including bioclimatic and soil type information) that suggests a slightly improvement (but not significant) in all model evaluation metrics (Figure R1_2). However, we highlight that our main purpose was to generate a parsimonious product independent of climate, biological and other sources of information with the main motivation of minimizing spurious correlations in the further use of soil moisture data (in Earth models and other ecological or geo-scientific analysis). Therefore, we conclude that a parsimonious model based only on terrain parameters performs similarly to a more complex model for predicting soil moisture at 15Km resolution.

REVIEWER COMMENT: Improved resolution would imply that there is different / more information in the downscaled product, but the indistinguishability of performance metrics (see above) suggests that this is not the case.

AUTHORS RESPONSE: We argue that the ‘indistinguishability of performance metrics’ is a proof of the reliability of our prediction framework. Please note that we are using a data driven model were the selection of best parameters (e.g., distances, kernels, neighbors, predictors) is made by cross validation. Cross validation allows to generate unbiased residuals and identify the linear relationship between observed and predicted data. Therefore, it is expected that our predictions maintain the general pattern (i.e., the similar mean and standard deviation) of the ESA-CCI product (Figure R1_2). No significant differences were observed between a model based on terrain parameters or the full model (using bioclimatic and soil type classes), but both of these models were better correlated against field data.

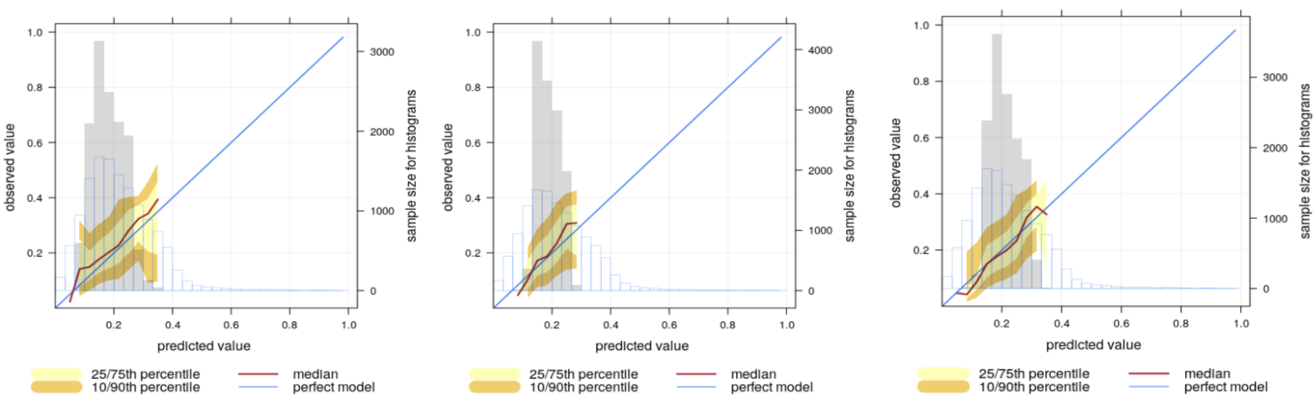

Figure R1_2 Evaluation of soil moisture predictions based on quantiles. The relationship between the ESA-CCI and the ISMN in an annual basis (a). We show the relationship between the ISMN field soil moisture and our predictions based on terrain parameters (b) in relation with a model using bioclimatic and soil type classes as prediction factors (c). Blue line is a perfect model. Blue histogram is from training data and gray histogram are from model predictions.

The soil moisture variability estimated within each 15km pixel is meant to maintain the numerical integrity of the ESA-CCI product at the global scale. We highlight that the gain of information between the original satellite and the downscaled predictions when visualizing soil moisture predictions across smaller areas (i.e., countries). We believe that the value of our downscaled product will be better recognized when the user sees the detail gained within a smaller region of the world (i.e., a country) (Figure R1_3).

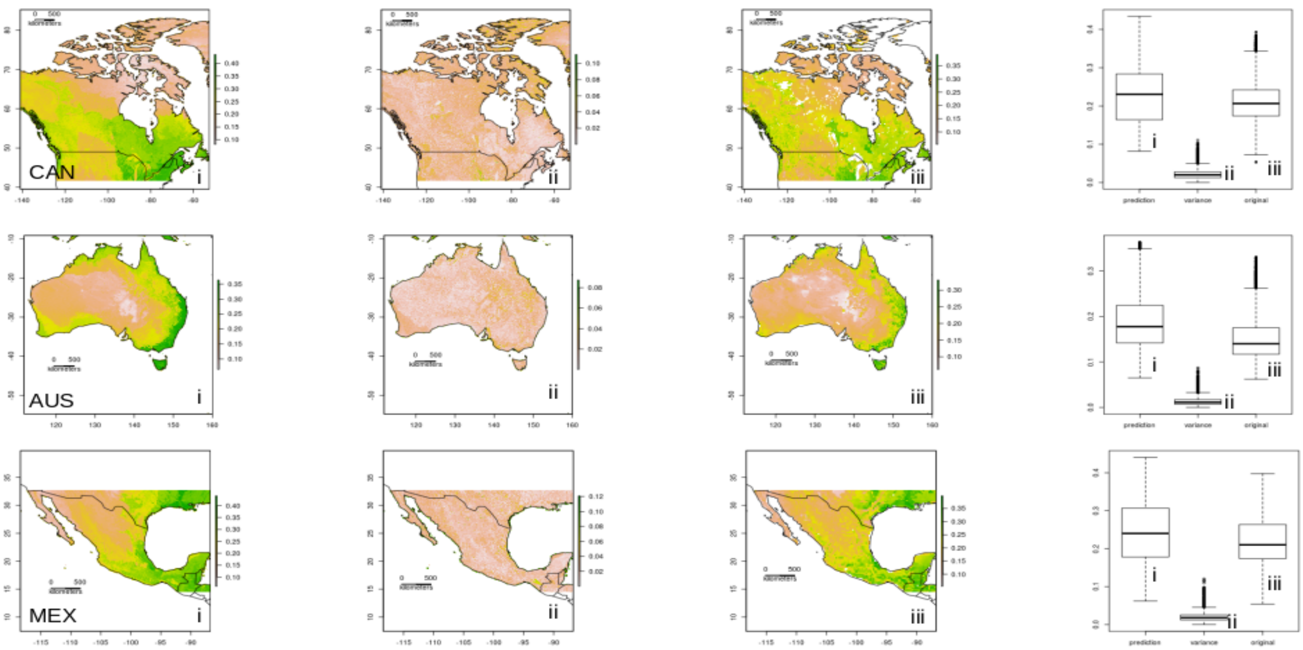

Figure R1_3 Example of soil moisture predictions based on the extent of countries with different sizes (e.g., Canada, Australia and Mexico) to show that there is a consistent increase in the range of soil moisture values in the predicted soil moisture maps of 15km grids compared with the original satellite estimate (~27km grids). The first column of maps contains the predictions, the second column the prediction variances and the third column the training data for each country specific model. The last column shows boxplots of the three maps for each country (note that in all cases our predictions reveal a larger range of values compared with the original estimates).

REVIEWER COMMENT: Also the comparison in Figure 6 (b and c) shows that the original and the downscaled products exhibit the exact same behaviour with a slightly lower overall variability in the original product. It is, however, not clear whether this different overall magnitude reflects an actual improvement, because soil moisture varaibility and trends are actually supposed to be different at a point scale and at a satellite scale (see e.g. Famiglietti et al. 2008).

AUTHORS RESPONSE: We would like to highlight that one clear improvement of our soil moisture predictions is that the soil moisture trend reported by our predictions is closer to the trend detected from soil moisture field stations, when compared with the trend from the ESA-CCI (Figure 6 of submitted manuscript version). We clarify that this comparison is made only with information at the places of field stations and associated pixels in the ESA-CCI and our downscaled product.

Although these trends are different, they are all negative and significant (as the confidence intervals are not overlapping zero values) in the three datasets. We found that the ISMN is shows the strongest negative trend, followed by the trend of our downscaled predictions and then the ESA-CCI soil moisture product.

In terms of spatial variability, the scale dependent variance of soil moisture analysis by Famiglietti (et al. 2008) is focused on shorter spatial scales, but we argue that there is no clear evidence of significant differences in soil moisture at the global scale when comparing the ESA-CCI and our downscaled product. Famiglietti (et al. 2008) provide evidence of scale dependent variances of soil moisture across shorter distances and relatively smaller spatial extents compared with our global effort. Recent efforts also have described scale invariant properties of soil moisture across scales (Mascaro and Vivoni, 2019). Thus, multiscale predictions of soil moisture using multiple modeling approaches could be useful to strive to overcome the main limitations (i.e., spatial gaps) of current soil moisture spatial information and solve the multiscale variability of soil moisture patterns from the plot to the country and global scales. This important discussion is now included in a revised version of the manuscript.

REVIEWER COMMENT: Also, the soil moisture mean is supposed to be different at different scales, hence the negative bias between point and satellite measurements cannot be reliably interpreted as error, and a reduction of this bias may as well be a going in the wrong direction with respect to the true areal-average mean. In other words, even though the generated product is sampled on a higher-resolution grid, it can not be concluded that this product contains higher-resolution information. Given the low amount of evidence that topography (alone) is a good predictor for soil moisture, observed differences may well be a result of the smoothing-nature of the KKNN approach, and any spatial-window resampling approach may lead to a seemingly "higher-resolution" (which is truly only a higher-sampling) product with the same (or even better) performance, but this is not tested.

AUTHORS RESPONSE: This is an important comment and we argue that is a science frontier that still needs active research. In the revised version we discuss the challenge of representing the scale-variance of soil moisture. We highlight that our analyzes between datasets have been done at the global scale and therefore any potential regional or country scale-variance are not included, but could be an important use of the dataset as a follow-up study.

To address the reviewer's comments we have done a reanalysis of our approach using the recently released ESA-CCI soil moisture version 4.5 and we can confirm a large correlation between the ESA-CCI and our soil moisture predictions (>0.92) at the global scale. Continental to global scales may be consistent in the overall range of values and spatial patterns, however smaller regions may highlight potential larger differences (Figure R1_3).

REVIEWER COMMENT: Therefore, I recommend to reject this publication. However, I do believe that topography may well be an imprtant complementary predictor for soil moisture at higher-resolution when combined with other dominant factors. I therefore encourage the authors to pursue this approach addressing the concerns outlined above.

AUTHORS RESPONSE: We believe that our revised version addresses most of the concerns of this reviewer. We have included new model outputs (i.e., full model including soil and bioclimatic factors), developed uncertainty estimates, and revised the analyses to improve clarity. We now clarify several sections of the main text, use the recently released version 4.5 of the ESA-CCI, and demonstrate the applicability of our methods across smaller areas and spatial extents. Please note that our previous work has demonstrated the effectiveness of our approach at the continental scale of CONUS using 1km grids (Guevara and Vargas, 2019).

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.Âahttps://doi.org/10.3334/ORNLDAAC/1578

Fick, S. E. and Hijmans, R. J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land area, International Journal of Climatology, 37(12), 4302–4315, doi:10.1002/joc.5086, 2017.

Food and Agriculture Organization (FAO). Global Ecological Zones for FAO Forest Reporting: 2010 Update; FAO: Rome, Italy, 2012.

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.

Mascaro, G., Ko, A. and Vivoni, E. R.: Closing the Loop of Satellite Soil Moisture Estimation via Scale Invariance of Hydrologic Simulations, Scientific Reports, 9(1), doi:10.1038/s41598-019-52650-3, 2019.

Møller, A. B., Beucher, A. M., Pouladi, N. and Greve, M. H.: Oblique geographic coordinates as covariates for digital soil mapping, , doi:10.5194/soil-2019-83, 2019.

Wieder, W.R., J. Boehnert, G.B. Bonan, and M. Langseth. 2014. Regridded Harmonized World Soil Database v1.2. Data set. Available on-line [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. http://dx.doi.org/10.3334/ORNLDAAC/1247.