This study develops a new global long-term soil moisture dataset by extending the SMAP data back into the ESA CCI era. To do this, the authors train a random forest with historical CCI data and apply the trained model to estimate SMAP soil moisture under the assumption that CCI and SMAP soil moisture have similar temporal variability. Such a dataset is valuable and relevant for a variety of climatological and hydrological applications.

Dear Referee

Thank you for your comments. The comments are undoubtedly helpful to improve the quality of the paper. Accordingly, we have analyzed the comments carefully and provided the response below.

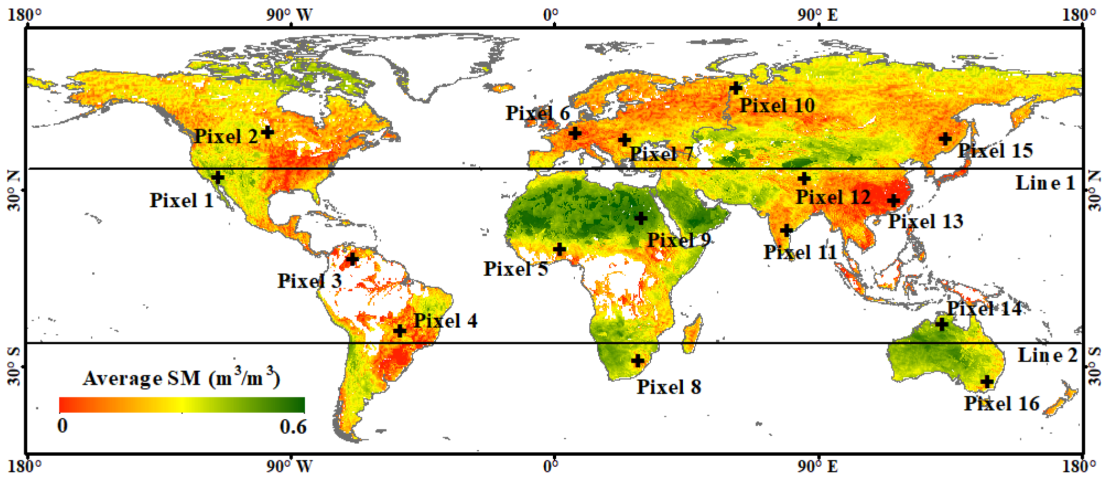
However, the main assumption made in this study requires more careful investigation. The authors present the temporal variability of CCI and SMAP over an overlapping period of 4-5 years, but most of the selected sites appear to be in arid regions (Fig. 2); comparison needs to be made in a more comprehensive way, including humid and/or high latitude regions with relatively high variability of soil moisture. Moreover, how can we ensure that this similarity is preserved also during the previous ~30 years of the CCI era?

Reply:

In fact, these five pixels in Figure2 were randomly selected. ***We consider that this issue can be solved by providing a more complete description in Section 2.2.3. Specifically, based on different humid regions and latitudes, we are going to supply more pixels to exhibit the change pattern of SM in Figure 2.***

As for the preservation of similarity during the previous ~30 years, we can explain this point from three aspects. First, the similarity of the CCI and SMAP data from 2015 to 2019 has been exhibited in Figure 2, which shows great similarity already. Second, we have found that both the CCI and predicted SMAP data can preserve consistent similarity to the in-situ data from a large period of about 10 years (i.e., from 1996 to 2015, as the earliest in-situ data began in 1996). Third, for the period before 1996, although there were no in-situ and SMAP data available for comparison with CCI data, the experimental results indicated that the general temporal profiles of CCI and predicted RF_SMAP are similar. Thus, we believe that the similarity can also be preserved over the 30 years.

To support the main assumption (similarity of the CCI and SMAP datasets), Figure 2 was modified and provided here in advance. Based on the original 6 pixels, we supplied additional 12 pixels (16 pixels in total). The pixels have random distribution, which include arid regions (e.g., Pixel 1 and 9), high latitudes (e.g., Pixel 10), and high altitudes (e.g., Pixel 12).



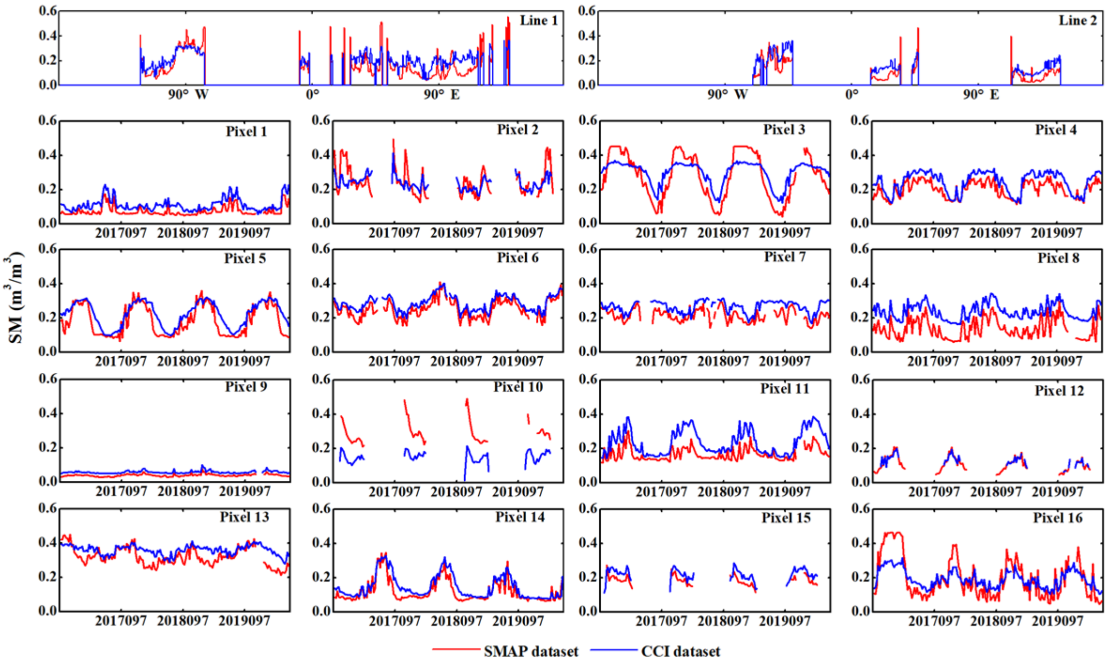


Figure 2. The patterns of changes in SMAP and CCI SM in temporal and spatial domains (16 pixels selected randomly at the global scale)

Second, the method described in Sect. 2.2.2 and 2.2.3 needs more explanation/clarification here and there. For instance, what is the purpose of having two separate experiments? If Experiment 1 was to evaluate the performance of RF_SMAP during the period of SMAP, the SMAP should be included for the model evaluation, e.g. in Fig. 6, Fig. 7, Table 5.

Reply:

*As for this issue, **we are going to supply more detailed description in Section 2.2.3**. **Meanwhile, the main purpose of the two experiments will also be further clarified in Section 2.2.3.** In fact, Experiment 1 aimed to demonstrate the predicted method based on the in-situ as well as the real SMAP data. Specifically, by using the CCI and SMAP data from 2016 to 2019 (2016105 to 2019361), the predicted data (i.e., RF_SMAP) in 2015 (2015105 to 2016097) were generated. In Experiment 1, both the real SMAP and in-situ data in 2015 (2015105 to 2016097) were available for validation. Accordingly, Figure 6, Figure 7 and Table 5 were the model evaluation results by referring to the in-situ data. Figure 5 and Table 4 are the evaluation results that used the SMAP data as reference. The reason of designing the experiment using the real SMAP as reference is that the reference in this case is known perfectly, avoiding the uncertainty introduced by other factors (e.g., the uncertainty in spatial support and geographical location in in-situ data).*

As for Experiment 2, it aimed to validated the RF_SMAP dataset from 1979 to 2015 based on the in-situ data, as in this period only the real SMAP data are not available.

Why is original soil moisture data (I guess you mean actual soil moisture time-series) unsuitable for model training? The authors stated that this is because 1) SM data has spatial gaps and 2) abundant precipitation can lead to abnormal change in SM (Lines 147-149). However, RF can only be trained with grid pixels where SM data is available, and the diverse relationship between precipitation-soil moisture should be included in the training data.

Reply:

It should be illustrated that the original SMAP time-series were used for model training in the first version (ESSD_2022_137). However, in the process of revision, we found the shortcomings of this training method. Specifically, in high latitude regions, the original SMAP time-series data contain unavoidable gaps (i.e., the missing data) in a year because of the snow cover and other factors. Theoretically, these spatially missing data cannot be involved in the training process, as you mentioned exactly. If we want to directly use the SMAP time-series data for training, we need to mask the regions with gaps. However, the usage of the mask can significantly harm the reliability of the RF_SMAP dataset in terms of spatial coverage. Also, the number of training data in the RF model can be reduced greatly. Hence, the hctsa characteristics-based training method was adopted in the manuscript. Since the hctsa-extracted temporal characteristics are spatial seamless, the interference of missing data in the SMAP time-series on model training can be eliminated.

To predict RF_SMAP, the trained RF model uses characteristics extracted from SMAP as input (Lines 185-194) at each grid pixel, is this correct? Then, how did you generate RF_SMAP for pixels and periods that do not have SMAP data (and thus unable to extract characteristics from SMAP)?

Reply:

We need to clarify that the trained RF model did not use the characteristics extracted from SMAP as input.

In fact, the construction of model is based on the core assumption that the CCI and SMAP datasets have similar pattern of temporal changes. Specifically, the model at a time t was trained by the label (CCI_t) and the characteristics (extracted ***from the CCI time series by the hctsa method****, coupled with the DEM and location data). In the prediction process, the characteristics (extracted* ***from the SMAP time series by the hctsa method****, coupled with the DEM and location data) were imported into the trained model, and the $SMAP_t$ data at time t was predicted. With the continuous change of CCI_t data from 1979001 to 2015097 (i.e., t, t+1, t+2, t+3, ...), different RF models were continuously trained and corresponding RF_SMAP_t data were predicted in turn.* ***We are going to rewrite this point and add key information in the new version of Figure 3****.*

To clearly illustrate the prediction process, Figure 3 is modified in advance.

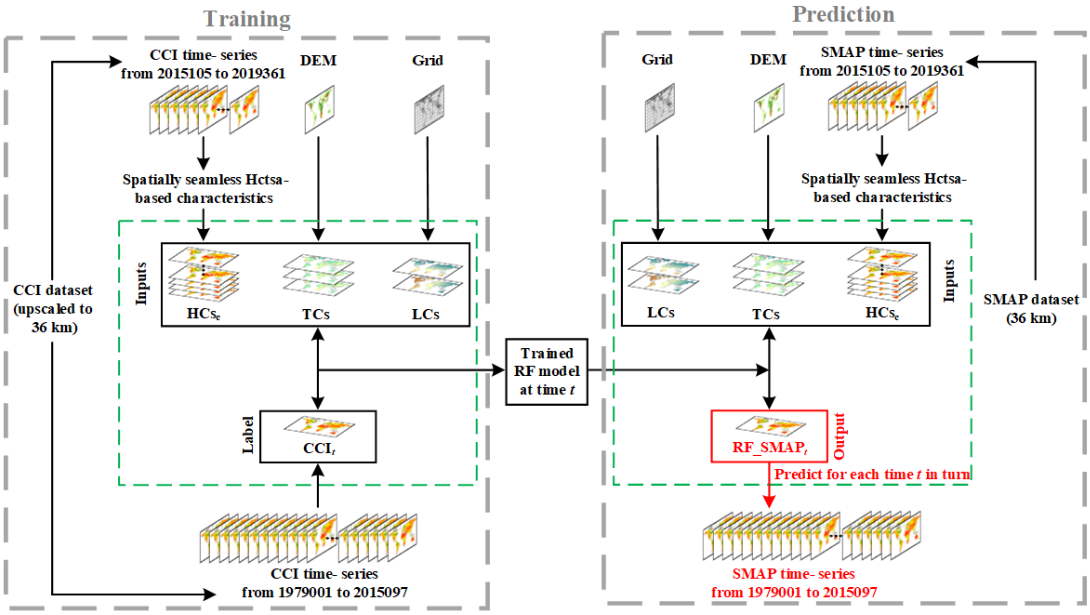


Figure 3. The prediction process of the RF_SMAP dataset at a time.

Lastly, the validation of RF_SMAP over the CCI era is highly limited due to the lack of ISMN before 2000.The validation of RF_SMAP over diverse climate regimes also seems limited, as most ISMN data are obtained from the US. I also wonder if there are any systematic biases between the RF_SMAP (historical SMAP before 2015) and the actual SMAP data from 2015.

Reply:

As you mentioned exactly, due to the uneven distribution and lack of in-situ stations, it is difficult validate the dataset based on diverse climate regions. However, the comparison based on different periods (e.g., from 2000 to 2005, and 2010 to 2015) is possible, ***we are going to analyze the systematic biases between the RF_SMAP (historical SMAP before 2015) and the actual SMAP data from 2015. We agree that this point is valuable, and we will revise.***

It is not clear why e.g. Fig. 6 shows only one time series per network and Fig. 7 shows a very small number of samples (dots) given that each ISMN network has >400 stations according to Table 2. Moreover, the comparison between the gridded datasets could be done from more diverse perspectives, e.g. comparison by season, during extreme (drought) conditions; SoMo is global, long-term data, but the comparison is done only for 4 years at three locations (Sect. 4.4)

Reply:

First of all, we need to clarify that Figure 6 aimed to show the change pattern of 11 networks (11 sub-figures) at 46 prediction times (i.e., from 2015105 to 2016097). Figure 7 provided the scatter plot of the corresponding 11 networks at 46 prediction times, based on the results in Figure 6. We need to clarify that the validation was at network level, that is, all stations in a network were averaged. In fact, the number of samples in Figure 7 is 46 (i.e., the prediction times) rather than the number of stations. ***The authors are going to revise the corresponding description to clarify this confusion.***

Additionally, the comparison of datasets in terms of different seasons is interesting. We will provide the results accordingly in the new version.

As for the comparison with the SoMo.ml dataset in Section 4.4, we need to clarify the purpose of this section first. That is, Section 4.4 aimed to exhibit the differences between the SoMo.ml and RF_SMAP dataset and provide a potential way to improve the RF_SMAP dataset in future. Specifically, the production of the SoMo.ml dataset used the in-situ data as model inputs to improve the accuracy. However, the in-situ data are always used as the reference for validation, which is undoubtedly beneficial for accuracy evaluation of the SoMo.ml dataset. In Section 4.4, we admitted the difference in accuracy between the SoMo.ml and RF_SMAP dataset, and proposed to use in-situ data to further enhance the predicted RF_SMAP dataset in future research. Thus, we considered that using longer time series of the SoMo.ml data and more in-situ data will not add anything to the current points in Section 4.4 (i.e., the conclusions will also be the same as the current version).