The manuscript presents a composited, SMAP-like soil moisture dataset derived with a random forest approach from historical CCI data. While a global soil moisture product for such a long period (1979 to 2015) has its merits, the dataset stands and falls with the successful evaluation of the derived product. Its spatial resolution (36 km) is interesting for large scale analyses.

It should be noted that the approach a) assumes stationarity of the CCI and SMAP data and b) generalises its global applicability of the model. Both are rather strong assumptions for a "simple" random forest approach. In addition, the evaluation refers to averaged network-level data, which introduces further uncertainty of the scaling. The authors present some limited evaluation, which does not really exemplify the validity of the derived model data.

There are some similar studies and datasets, the authors did not refer to: https://www.nature.com/articles/s41597-023-02053-x , https://www.nature.com/articles/s41597-021-00925-8 In which way does their dataset (and approach) advance these?

Moreover, there is a recent paper in HESS, which uses Sentinel data to estimate soil moisture: https://hess.copernicus.org/articles/27/1221/2023/

Obviously, the temporal extent of this approach has very little overlap with the data presented in this manuscript (since the sentinel satellites have been operational just from 2015 onwards). However, the authors might find inspiration for further evaluation in this?

After all, it is very difficult to evaluate if the dataset and its presentation justify publication in ESSD. If the authors could really corroborate the validity of their data product, this would be a clear yes. However given the open questions and despite the meticulous effort which went into the compilation of this data, it remains too unclear, how the dataset from a rather simple approach can advance already existing SMAP-like soil moisture products.

Reply:

Dear Referee

Thank you very much 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.

The major comments include three aspects. ***First, the referee concerns the stationarity of the CCI and SMAP datasets. Second, the referee concerns the global applicability for a simple RF model. Third, the referee concerns that an averaged network-level validation can introduce further uncertainty.*** *To address the above questions, we have provided responses and preliminary modifications, respectively.*

First, both the CCI and SMAP datasets are used to characterize the surface soil moisture, which have the same units (i.e., m^3/m^3) and physical meaning. Meanwhile, the original data of the CCI and SMAP datasets are derived from satellite sensors. Thus, it is reasonable to assume that the CCI and SMAP datasets have similar change pattern in a long-time series. Additionally, we have provided the basis for the stability of the CCI and SMAP datasets in the original manuscript

(Figure 2). ***For this concern, we are going to add more pixels in different regions to clearly illustrate the stability of the CCI and SMAP datasets.***

To clearly illustrate this point, we revise Figure 2 and supply more pixels to exhibit the stationarity of the CCI and SMAP datasets in advance.

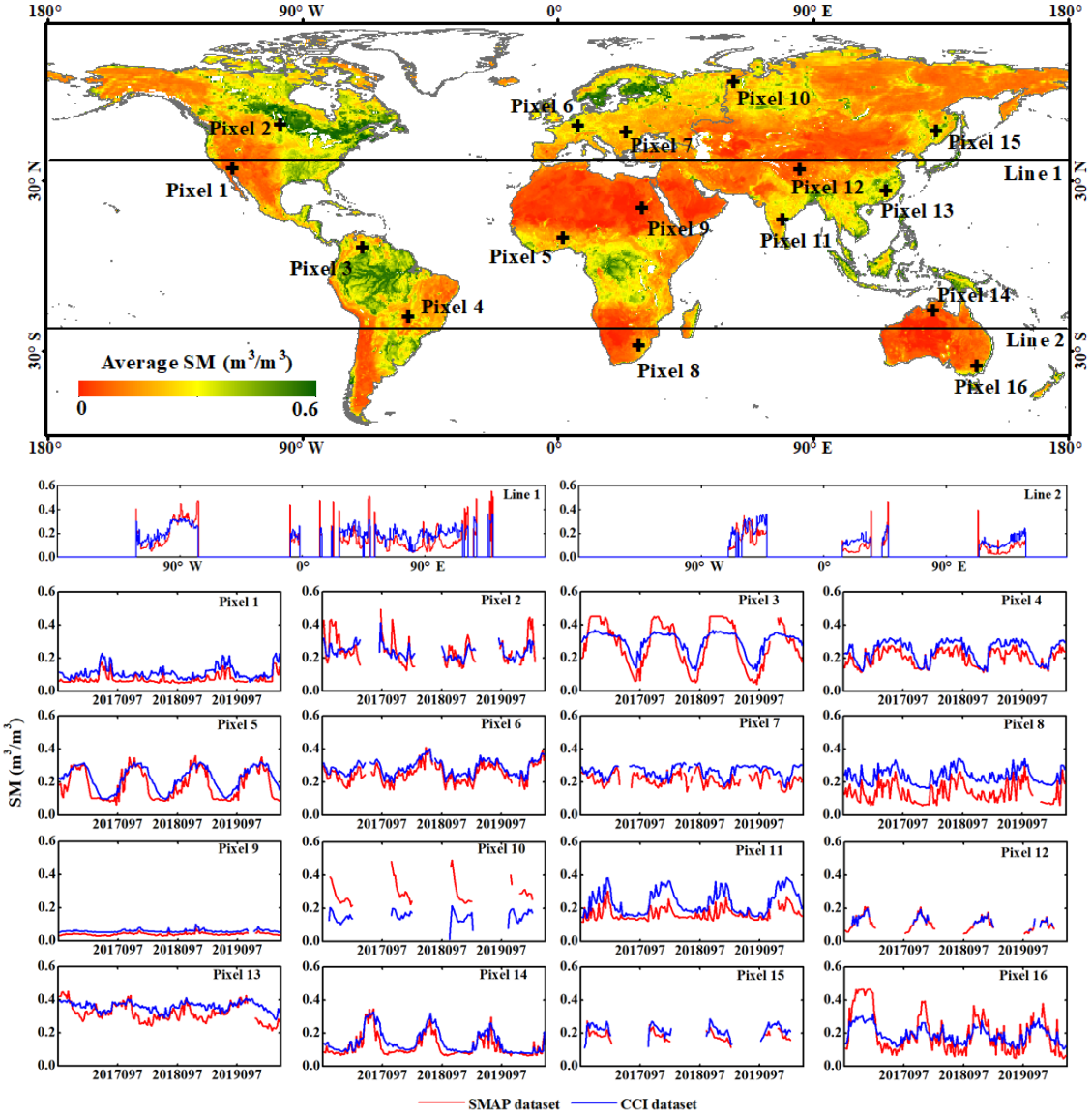

Figure 2. The patterns of changes in SMAP and CCI SM in temporal and spatial domains (16 pixels selected randomly at the global scale)

Second, it should be clarified that the proposed RF model is dynamic, and needs to continuously train for different scenes in a long-time series. Moreover, for high latitudes and high altitudes, spatial gaps usually affect the data quality. To solve this issue, the hctsa-based method was adopted in the manuscript to extract spatially seamless characteristics, suggesting that the spatially seamless RF_SMAP dataset can be generated.

The construction of model is based on the core assumption that the CCI and SMAP datasets have similar patterns 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 in the updated manuscript and provide more detailed information in the new version of Figure 3.***

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

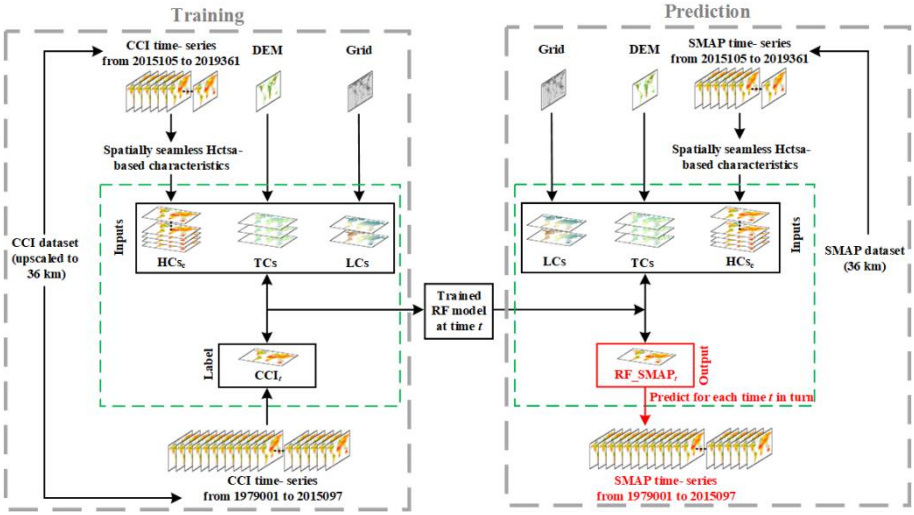

Figure 3. The prediction process of the RF_SMAP dataset at a time.

Currently, the training method is developed using the globe-based data. ***Considering the concern of the referee, we will add an additional training method*** *(i.e., continent-based data are used for training to generate datasets in different continents)* ***for comparison in Section 3.3.***

The comparison using different training data in Section 3.3 is provided in advance. *Specifically, 46 scenes in each continent from 2015105 to 2016097 were predicted using the different training data (including the globe- and continent-based data). By referring to the true SMAP data in each continent, it can be seen from Table 7 that the predicted results using the globe-based data (used in the manuscript) is more accurate than those using the continent-based data. More precisely, using the globe-based data in AF can provide the CC of 0.941 and the RMSE of 0.041, which are more satisfactory than using the continent-based data. Moreover, the CC values using the globe-based data in EU and OC are 0.871 and 0.881, which are 0.002 and 0.018 higher than those using the continent-based data. In all, using the globe-based data is more effective than using the continent-based data. This means that the spatially adjacent information in the training data is useful to improve the prediction accuracy of the dataset.*

Table 7. Comparison between the use of different train data.

Continent	*Training data*	*CC*	*RMSE*	*Bias*	*ubRMSE*
AF	*Globe-based*	***0.941***	***0.041***	***0.001***	***0.041***
	Continent-based	*0.935*	*0.043*	*0.001*	*0.042*
AS	*Globe-based*	***0.921***	***0.058***	***0.010***	***0.057***
	Continent-based	*0.920*	*0.059*	*0.010*	*0.057*
EU	*Globe-based*	***0.871***	***0.061***	***0.006***	***0.058***

NA	*Continent-based*	*0.869*	*0.063*	*0.007*	*0.059*
	Globe-based	***0.913***	***0.064***	***0.015***	***0.061***
OC	*Continent-based*	*0.911*	*0.065*	*0.016*	*0.063*
	Globe-based	***0.881***	***0.048***	***-0.001***	***0.047***
SA	*Continent-based*	*0.863*	*0.052*	*-0.011*	*0.051*
	Globe-based	***0.910***	***0.060***	***0.010***	***0.059***
	Continent-based	*0.909*	*0.061*	*0.010*	*0.059*

Third, ***we are going to supply a station-level evaluation for different climate types in Experiment 2 to further demonstrate the advantage of the RF_SMAP dataset.*** *Meanwhile, the supplement can clearly illustrate the validity of the RF_SMAP dataset and can provide a basis for applications. In addition, the uncertainty of validation can also be further reduced.*

To further validate the RF_SMAP dataset, we provide the station-level evaluation in Figure 12 based on the Köppen-Geiger climate classification in advance. *Generally, we found that the station-level evaluated results are similar to the network-level evaluated results. According to the KGE, RMSE, and ubRMSE, the RF_SMAP dataset can provide a more satisfactory performance than the CCI and GLEAM datasets in the arid steppe and temperate regions. The KGE (RMSE) of the GLEAM dataset is always smaller (higher) than that of the CCI and RF_SMAP datasets in the cold regions. Furthermore, the Bias of the RF_SMAP dataset is usually closer to the reference than that of the CCI and GLEAM datasets in addition to in BSh, Cfa, and Csa regions. Besides, the CC and SRC of the GLEAM dataset are always higher than those of the CCI and RF_SMAP datasets.*

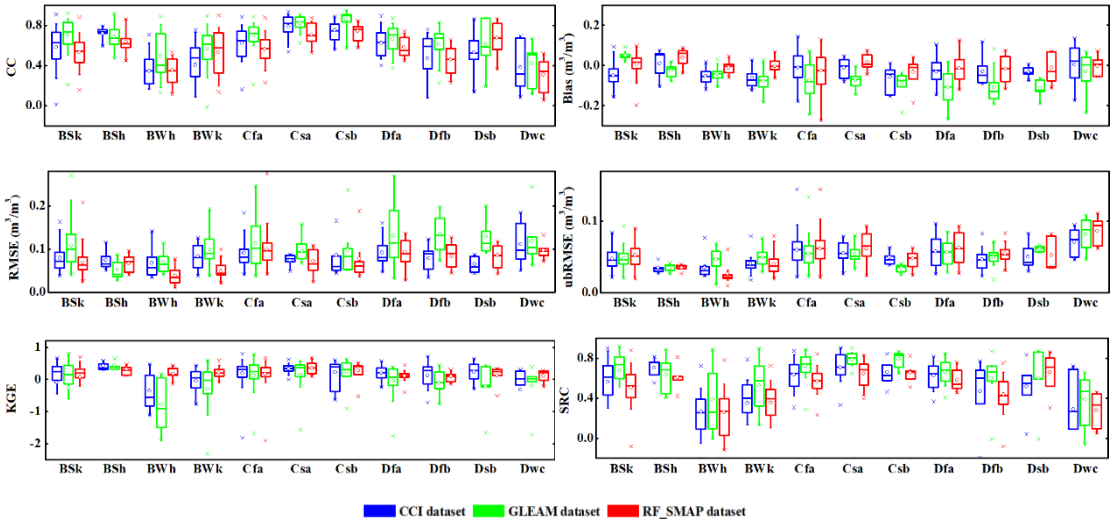

Figure 12. Accuracy comparison of the historical CCI, GLEAM, and RF_SMAP datasets in different climate types (each climate type contains the different number of stations, BSk contains 118 stations, BSh contains 7 stations, BWh contains 20 stations, BWk contains 39 stations, Cfa contains 83 stations, Csa contains 16 stations, Csb contains 11 stations, Dfa contains 35 stations, Dfb contains 18 stations, Dsb contains 7 stations, Dwc contains 8 stations, respectively. Climate types with less than 5 stations are not counted).

After revising, the accuracy validation of the RF_SMAP dataset is relatively comprehensive, including the network- and station-level. Meanwhile, six statistical metrics (i.e., CC, RMSE, Bias,

ubRMSE, KGE, and SRC) has been adopted in the manuscript, two widely used SM datasets (i.e., CCI, and GLEAM) have been used as benchmark, and the validation for different networks, continents and climate types has been provided to support the RF_SMAP dataset.

In addition, the three mentioned papers (Yao et al., 2021; Madelon et al., 2023; Skulovich and Gentine, 2023) will be added to the introduction, which can provide important theoretical support. According to the validated strategy in Skulovich and Gentine. (2023), we have added a new station-level validation for different climate types (i.e., Figure 12). Besides, the time period of the RF_SMAP dataset is longer than that in these papers.

Madelon, R., Rodríguez-Fernández, N.J., Bazzi, H., Baghdadi, N., Albergel, C., Dorigo, W., & Zribi, M. (2023). Soil moisture estimates at 1 km resolution making a synergistic use of sentinel data. Hydrology and Earth System Sciences, 27, 1221-1242.

Skulovich, O., & Gentine, P. (2023). A long-term consistent artificial intelligence and remote sensing-based soil moisture dataset. Sci Data, 10, 154.

Yao, P., Lu, H., Shi, J., Zhao, T., Yang, K., Cosh, M.H., Gianotti, D.J.S., & Entekhabi, D. (2021). A long term global daily soil moisture dataset derived from amsr-e and amsr2 (2002-2019). Sci Data, 8, 143.