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.

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?

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.

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.

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)?

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.

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)

Given that these aspects require substantial consideration, I think that the current manuscript is not publishable in ESSD.