The manuscript describes a novel approach to integrate multiple precipitation estimates from satellite soil moisture measurements into an existing precipitation data set. The validity of the approach is confirmed over six regions using triple collocation validation and cross-validation with rain-gauge measurements where available.
The manuscript is overall well written and mostly well structured. Title and abstract fit the topic and content. Data generated within the study is publicly available, the URL in the manuscript should be updated in order to work properly (omit ".XThcfHvOOUk").
The supplementary materials contain plots and tables that are addressed in the text and support the conclusions.
Using the SM2RAIN approach to improve existing precipitation products is a logical step in terms of deriving short-latency precipitation data and of interest for the community.
The underlying concepts of the manuscript (OLI, SM2RAIN, TC) are well established, original contributions are summarized in a comprehensible way and referenced properly.

We thank the reviewer for pointing this out and for the valuable suggestions. We will updated the link of the dataset.

More details on why the OLI method is preferred in the combination step over other approaches should be given.

Thanks for the comment. The ingestion of more SM2RAIN products implies that different satellite soil moisture products are used. Despite the potential benefits of one product with respect to another (i.e., active vs passive) the information we are ingesting is always soil moisture thus it is potentially redundant. The OLI method is particularly advantageous in this respect, as it accounts for both performance differences and error covariance between the rainfall products and is therefore insensitive to the addition of redundant information. We are aware that other more sophisticated methods exist (as for instance those based on data assimilation). However, it must be said that their better performance is not always guaranteed given the higher number of parameters upon which they rely. For instance, Brocca et al. (2017) found that simple integration methods performed equally well and in some cases better than data assimilation- based methods. We are also exploring new integration techniques (see Massari et al. 2019), but we think they are not mature enough for a continental scale application like this one. In the revised manuscript we will add more details in this respect.

Brocca, Luca, et al. "Rainfall estimation by inverting SMOS soil moisture estimates: A comparison of different methods over Australia." Journal of Geophysical Research: Atmospheres 121.20 (2016): 12-062.

Massari, C., Maggioni, V., Barbetta, S., Brocca, L., Ciabatta, L., Camici, S., ... & Todini, E. (2019). Complementing near-real time satellite rainfall products with satellite soil moisture-derived rainfall through a Bayesian Inversion approach. Journal of Hydrology, 573, 341-351.

The chosen calibration data set YREF within the combination step should show "homogeneous performance in space and time" globally, yet the chosen ERA5 reanalysis is likely not to provide this. ERA5 was found the best fitting of three candidates, yet potential issues of reanalysis products should be discussed and the (in)dependence of ERA5 from the used satellite data resp. raingauge data should be addressed thoroughly.

Thanks for this important comment. A perfect candidate is hard to find and even ERA5, which provided overall relatively good results, suffers from significant uncertainty in convective dominated systems (please refer to the reduced correlation in Western Africa and the Sahel in Figure 2). We considered three potential candidates and found that both GPCC and IMERG-FR strongly rely on rain gauge information, which for many regions may be scarce or absent altogether. This results in significant interpolation error of GPCC over Africa (Figure 2) and consequently potential detrimental effects of the gauge correction performed for IMERG-FR in the same regions. Thus, the problem of gauge absence seems much larger than the lower performance of the reanalysis dataset. Based on this consideration we selected ERA5 as calibration dataset.
We also tried to understand the impact of the quality of the calibration dataset on the integrated product. Based on the results reported in Table 3 and Figure 10, the integrated product seems slightly impacted by it (unless its performance is very bad) as:

1) **in the regions where ERA5 performs relatively bad (Sahel and Western Africa) the integrated product still provides relatively good results ; and**
2) **the integrated product generally performs better than the calibration dataset in the cross-validation analysis indicating that its impact on the quality of the integrated product is small.**

In the revised version of the manuscript we will discuss these aspects further trying to outline better the impact of Yref on the final results.

Some background for the chosen threshold of R < 0:4 between SMt2RAIN and reference data to perform OLI should be given.
This was simply a choice to exclude poor performance of SM2RAIN by accepting the risk of low correlations (<0.4) was due to the bad quality of ERA5. We performed some experiments over CONUS, Australia, Europe, and India and found that 0.4 is a reasonable threshold although its overall impact is very small and only limited to small regions (e.g., relatively high RFI regions for SMOS). We will add more details in the revised version of the manuscript trying to better explain the rationale behind the selection of this threshold.

The "classical" validation part and the "assessment of TC validity" could be shorter or it should be explained, why extensive verification of the TC approach within this study was found to be necessary. As referenced in the manuscript, TC has been used to validate precipitation from satellite SM in a previous study (Massari et al., 2017).

Massari et al. 2015 adopts a 1-deg resolution, so we thought it was important to validate the procedure at a finer scale (o.25deg). We will move the TC validation in the supplementary material.

Considering the short calibration/validation period, the potential impact of climate patterns or their absence (e.g. due to ENSO) on the calibration process should be discussed.

We are aware of potential problems related to this issue but at the time of writing IMERG (and SMAP) were only available since 2014 (2015). Now a reprocessed version of the IMERG products is available back to 2000. This is a potential opportunity to re-calibrate the integrated products with available soil moisture on specific years (i.e., ASCAT from 2007 onward, SMOS from 2010 onward, and SMAP from 2015 onward). We want to highlight however that this study shows already great potential in merging soil moisture-based rainfall observations with IMERG-ER and any limitation found at this stage (not directly related to the product development itself) represents additional room of improvement.
In the revised version of the manuscript specific sections/paragraphs will be devoted to the impact of the choice of a different calibration dataset on the final results.

The increase in FAR in Fig. 7 needs more explanation.

We observed an increasing trend in FAR from 60 to 80 percentiles and a drop above 85-90. This means that medium-high rainfall classes are increased more than necessary while high rainfall intensity are rightly reduced. While these increments in FAR are very small for Australia and CONUS, they are not negligible in India and Europe. Soil moisture-based rainfall has a tendency to generally provide higher FAR (deterioration) and higher POD (improvement) than IMERG-ER. When merged, SM2RAIN and IMERG-ER, result in a small increase of FAR and a significant increase of POD with an overall improvement of TS. Although the integration is overall optimal (TS increases), the POD increment -not surprisingly - causes an increase in FAR and viceversa . This is a classical multi-objective calibration exercise where, for the choice of the best configuration, one has to operate with a specific application in mind.
In addition, we found that the higher FAR for SM2RAIN datasets could be reduced by changing the calibration score to KGE, which is less affected by conditional biases , however, this also reduced the POD with an overall lower TS.

In the revised manuscript we will expand the discussion on these issues when discussing the figure and try to provide more insights and explanations on this issue.

The measures (median) that are shown in the (bar) plots should be defined in the plots or the caption (also for the tables).

It will be done.

Areas that bar plots refer to are not always clear from the figures alone (Fig. 4), box plots resp. tables instead of bar plots would provide more information resp. improve readability/comparability. Information on what box edges, whiskers represent in Fig.8 are especially necessary, to show that the impact of (single) outlier triples is not omitted in the plots.

We will try to use tables/box plots instead of bars in Figure 4. Also, all the missing information in the caption will be added.

P1, L5: "they are" instead of "they're" P4, L25: "Metop" instead of "METOP" P5, L12: A.M. - AM consistency P6, L27: "within" instead of "wihtin" P6, L29: "whereas suffer" missing words P8, L3: "weighting" instead of "weighing" P8, Eq. (5) : missing superscript "1" P11, L3: "satisfy" instead of "satisfies" P12, L23: duplicate section reference
P20, L6: "targeted" instead of "target" and duplicate "(" in L5

Thanks a lot. We will correct all these typos.