The paper introduces a potentially useful precipitation product and is overall quite well written. However, some serious issues need to be resolved before it can be published. "we minimised the daily root mean squared error (RMSE) between the SM2RAIN rainfall applied to the specific SM product and YREF during 2015-2017." This is problematic because of the noisy nature and highly skewed distribution of precipitation.
ERA5 already underestimates precipitation peaks, and using this approach, the obtained SM2RAIN estimates will underestimate precipitation peaks even more. The Kling-Gupta Efficiency is probably a better choice as it accounts for the variability.

We thank the reviewer for their valuable suggestion. The reviewer is theoretically correct. KGE would theoretically provide a better dynamic range although our attempts to use it instead of RMSE did not always show these improvements. Moreover, by looking at Figure 7 (which presents the results obtained using RMSE) it can be seen a particular benefit to rainfall peaks with a reduction of FAR, an increase in POD and TS for percentiles larger than 90% which contrasts with what suggested by the reviewer.
In this study, we would rather ensure homogeneity among all the calibration steps, keeping in mind the results obtained here can be further improved by a better calibration using for instance KGE. We will clearly underline this issue in the revised version of the manuscript specifically in Section 3.3.

"The final product is then composed of multiple rainfall datasets weighed according to Eq. 6." An averaging scheme like this causes underestimation of peaks and introduces spurious drizzle. I realize that zero values of IMERG were kept, but this does not eliminate all spurious drizzle issue. It will, however, probably introduce a spurious discontinuity in the precipitation distribution.

We investigated this issue and found no detrimental effects of the current integration scheme on either low rainfall regimes (below the 50 percentile) or high peaks (rainfall percentiles larger than 90). By contrast we observed a slight increase of FAR in Australia and CONUS and little higher increments of this score for Europe and CONUS at medium/high rainfall regimes (60% up to 85%). Larger FARs were also accompanied by a significant increase in POD, which in turn determined the improvement in TS. As highlighted above, the change of the calibration score for SM2RAIN did not provide always a better behaviour and resulted in a POD decrease with overall smaller TS.

"The continuous scores were the Pearson correlation coefficient (R), the Root Mean Squared Error (RMSE), and the additive bias (BIAS)." The RMSE statistic should not be used at the daily time scale because it yields "better" values for datasets which underestimate precipitation peaks (such as SM2RAIN and the dataset introduced here). The KGE (with its three independent components) is probably a better choice.
Overall, I think the authors should remove the RMSE from the evaluation and introduce metrics that evaluate the low and high tails of the precipitation distribution of the new product. Any issues revealed using these new metrics should be highlighted in the abstract.

We thank the reviewer for this important comment. Any error-based metric like Means Squared Error (MSE) or KGE includes a correlation component, a variability component (often known as multiplicative bias or conditional bias), and an additive bias component. This can be demonstrated with some simple mathematical manipulations (see Murphy et al. 1988 and Gupta et al. 2009):

$$MSE = 2\sigma_s\sigma_o(1-r) + (\sigma_s - \sigma_o)^2 + (\mu_s - \mu_o)^2$$
$$KGE = 1 - \sqrt{(r-1)^2 + (\frac{\sigma_s}{\sigma_o} - 1)^2 + (\frac{\mu_s}{\mu_o} - 1)^2}$$

where σ and μ refer to the standard deviation and the mean of the simulated "s" and the reference "o" time series whereas r is the correlation between them. So the three components are present also in the MSE or in its root version.
The notable difference of KGE with respect to MSE is the weight associated to the variability component which is larger for KGE with respect to MSE and the fact that it is a self consistent score as it varies from 0 to 1 (Gupta et al. 2009).

In the revised version of the manuscript we will provide the validation also in terms KGE score while the RMSE will be still maintained as:

1) **many past studies are based on this metric and this facilitates the comparison of this work with them;**
2) **it is a physical error measure, therefore easier to relate to the actual physics of precipitation. (e.g. mm/day vs some fraction of KGE);**
3) **it can be compared against the results obtained via TC in a more meaningful way;**

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of hydrology, 377(1-2), 80-91.

Murphy, A. H. (1988). Skill scores based on the mean square error and their relationships to the correlation coefficient. Monthly weather review, 116(12), 2417-2424.

"Note that, based on this choice, the integrated product is totally independent upon rain gauges" This not true as ERA5 assimilates precipitation gauge observations.

We agree with the reviewer ande have removed the sentence. However we want to highlight some points which we think would be interesting to discuss:

1) **From the document "Operational global reanalysis: progress, future directions and synergies with NWP" which describes in details the development of ERA5 reanalysis dataset it is clear that after 2009 "rain rate" is ingested (Figure 17 panel k) in the reanalysis but it is not clear which rainfall information is ingested. To our knowledge only the NCEP Stage IV analysis which combines rain gauges and radars estimates are ingested into the analysis and ERA5 reanalysis but only in United States while no other gauge information is present in the two products outside this area (see also Lopez et al. 2011).**

Future developments will ingest radars information from OPERA (https://www.ecmwf.int/en/elibrary/18765-operational-global-reanalysis-progress-future-directions-and-synergies-nwp) but currently this is not already done.

2) **It is unlikely that rain gauges will be also ingested over data scarce regions as rain gauges are mostly absent over these regions (see Figure 2 in the paper).**
3) **Over CONUS we found relatively good correlations of ERA5 with Stage IV but also found it lower than the one obtained with our integrated product (see Table 3). This highlights that the integrated product is not really so dependant on the calibration dataset (i.e., ERA5) as highlighted also from the results in Figure 10 and 11.**

Lopez, P. (2011). Direct 4D-Var assimilation of NCEP stage IV radar and gauge precipitation data at ECMWF. Monthly Weather Review, 139(7), 2098-2116.

"ERA5, which provides full coverage and generally homogeneous performance all over the world." Not sure I agree with this as atmospheric models tend to perform markedly worse in convection-dominated regions.

Our application of TC demonstrated that ERA5 is the best available calibration dataset among those selected, although it still suffers from uncertainty in convection dominated systems (i.e., Western Africa and Sahel see Figure 2). Despite this, the performance of the integrated product seems not to be too much impacted by its quality, as shown in Figure 10 (i.e.,compare results over Western Africa of ERA5 and the integrated product).

Page 10 line 27: Add "out" after "carried".
It will be corrected.

Figure 12: Can you add short titles to each subplot?
It will be done.