The main concerns are listed as follows:

The main aim of the paper is not clear. Several times it is stated that the aim of the paper is to explore the possibility of substituting the source of the NDVI data (Sentinel-3 instead of Sentinel-2), but this actually occupies a very small portion of the manuscript. The authors conclude that the performances are comparable, so that go in detail on comparisons between the S2MP output against other products and in-situ data. Hence, is this a paper aiming at exploring the use of Sentinel-3 instead of -2 in the processing chain or is it a validation study of the S2MP retrieval (hence an extension of Bazzi et al., 2019)? The paper should be better organized under the perspective of highlighting the main aim of this research.

We truly thank the reviewer for his/her pertinent comments on the manuscript. We understand that the exact goal was not very clear. There is a thoughtful evaluation of different soil moisture products but the goal of the study is not just an evaluation. The goal of the manuscript is to study how to produce 1-km soil moisture maps over large regions using the synergies of different Sentinel satellites. For that, it is needed to (i) extend the S^2MP to other land cover types than croplands and (ii) evaluate the use of S3 instead of S2. The S^2MP was originally designed to work at very high resolution (~10 m). Only afterwards, the relative performances of these maps are evaluated against other state-of-the-art datasets. In a revised version of the manuscript we will clarify the goal and improve the structure. In particular, taking into account this and the other reviewer comments, after comparing S2 and S3 NDVI and the S1+S2 and S1+S3 soil moisture estimations, the rest of the paper will present the S1+S3 maps.

The known issue of the S2MP-derived product consisting in unreliable SM estimates associated with NDVI > 0.7 (Bazzi et al., 2019). Is there any benefit in this sense by using Sentinel-3? Is the issue attenuated working at 1 km?

When the vegetation is well-developed, the estimation of the soil moisture becomes less reliable due to the penetration capabilities of the C-band SAR data in some developed vegetation covers. This limitation has been assessed by several studies showing that for NDVI values beyond 0.7 (NDVI > 0.7), the soil moisture estimation becomes less accurate. This threshold is only a good descriptor of the developed vegetation and is independent from the algorithm used to estimate soil moisture. In other words, when the vegetation is well-developed, the C-band SAR signal will fail to penetrate the vegetation cover and the soil moisture estimation will be less accurate whether the NDVI is calculated from S2 or S3.

Of course, it is more common to have NDVI lower than 0.7 with S3 data because the NDVI value for a 1 km x 1 km pixel includes land parcels with NDVI lower and higher than 0.7 and corresponds to a weighted mean of all NDVI in the pixel of 1 km resolution. But this does not change anything with respect to the fact that S1 can saturate locally. This will be an issue to extend the maps to forest dominated regions but it is not a significant problem for crops and herbaceous areas. In addition, NDVI is a predictor used as input to the neural network that helps to take into account the effect of the vegetation, but the main predictor to estimate the soil moisture is the microwave backscattering.

Additional minor issues are listed in the following:

Independently of the SM normalization expressed in eq. (1), I believe that the comparison with CoperSWI should be carried out by calculating the SWI (with same T value) for the S2MP SM as well.

This study is dedicated to surface soil moisture retrieval. In contrast to the so-called change-detection approaches, the S^2MP algorithm produces SM maps in physical units. Studying the soil moisture dynamics in deeper layers using different T values is out of the scope of the study. Therefore, using well-validated and published results, we took the T values that represent the surface layer before transforming the index SWI into physical units. Comparing the two datasets without scaling the indices will also prevent the inter-comparisons with other datasets since most of them are provided in volumetric units.

Potential impacts of a lower coverage of Sentinel-derived observations outside Europe should be discussed.

In contrast to change-detection approaches, which are by construction time-series based approaches, the S^2MP algorithm performs instantaneous retrievals, without using previous values. Therefore, the revisit frequency does not affect the retrievals. We thank the referee for this comment, we will add a short discussion on this topic in the discussion section of a revised version of the manuscript. Regarding the final soil moisture time series, of course the temporal frequency of S1+S2 or S1+S3 will be limiting for some applications but not for long-term soil moisture monitoring at high resolution.

The purpose of the Sentinel-2 VS Sentinel-3 NDVI comparison at 1 km is not clear to me, since Sentinel-2 data has been processed at higher resolution within the S2MP algorithm.

The S^2MP algorithm was designed to work at very high resolution and the S1 data is actually aggregated at high resolution within the 1 km^2 pixels only for croplands and herbaceous vegetation areas. When using S2 to estimate NDVI, only those areas are used while in the S3 NDVI all areas within the 1 km^2 pixels contribute to the NDVI estimation. It was then necessary to evaluate the potential effect of this mismatch on the way S2 and S3 are used. Therefore, we first compared the NDVI from S2 aggregated for croplands and herbaceous areas to that obtained with S3. One must bear in mind that NDVI is just a predictor used as input to the neural network, but it is not the main predictor (that with the highest weight). Hence, the impact of this possible mismatch is limited. However, it was still needed to evaluate the impact of replacing S2 by S3 in the final result, and this was done by comparing S1+S2 and S1+S3 retrievals to in situ measurements and other soil moisture products. We will clarify the interest of those two comparisons (NDVI first, and final SM estimates afterwards) in a corrected version of the manuscript.

In addition, it is noteworthy that even if the temporal revisit of the SM estimates derived from S^2MP_{S1S2} and S^2MP_{S1S3} is equivalent, the higher temporal revisit of S3 allows to retrieve more optical images without cloud conditions than S2. This results in a better estimation of NDVI.

Lines 273-276. Are the differences attributable to NDVI only? The aggregation to 1 km of Sentinel-1 VV backscattering and of the incidence angle in the S2MP adapted to Sentinel-3 has no impacts?

Of course, there are two reasons that explain the differences between the S1+S2 and S1+S3 soil moisture maps.
(i) Lower the correlation between S2 and S3 NDVI is, higher the difference in soil moisture estimation is. This is due to the fact that S3 NDVI at 1-km resolution integrates NDVI of all types of land parcels within the pixel while S2 NDVI at 1-km resolution only integrates parcels corresponding to croplands and herbaceous vegetation.
(ii) Backscattering coefficients at 10-m resolution are aggregated at 100-m and the neural network is only applied over croplands and herbaceous vegetation for the S1+S2 maps. The SM estimates are then aggregated at 1-km resolution.

In contrast, for the S1+S3 maps, backscattering coefficients over croplands and herbaceous vegetation are aggregated from 10-m to 100-m resolution, and then to 1-km resolution only over croplands and herbaceous vegetation. Finally, the SM estimation is performed at 1-km resolution.

These differences in aggregation (S1 aggregation + SM aggregation for S^2MP_{S1S2}, S1 double aggregation for the S^2MP_{S1S3}) result in differences between the S1+S2 and S1+S3 soil moisture maps.
These two sources of differences will be reminded in the discussion section.

The higher correlation between in-situ and CR data with respect to HR estimates. Can it be due to the higher temporal resolution of the CR data sets?

The temporal sampling of the CR datasets (SMAP, SMOS, ESA CCI) is roughly 5 times higher than those of the HR datasets (S^2MP_{S1S2}, S^2MP_{S1S3}, CoperSSM, SMAPS1) and might actually affect the evaluation results with respect to in-situ measurements. Hence, the evaluation of the CR time series has been performed a second time by taking into account only 1 observation out of 5. However, the performances are not affected significantly (see Tables below). In a revised version of the manuscript, we will add a sentence on this check in the results section.

Here are the results without reducing the temporal revisit:

Products	R	R^a	$Bias$	$STDD$
Sentinel-only high resolution data				
S^2MP_{S1S2}	0.59	0.36	-0.06	0.05
S^2MP_{S1S3}	0.56	0.37	-0.06	0.06
$CoperSSM$	0.53	0.18	0.04	0.08
Merged high resolution data				
$CoperSWI$	0.74	0.46	0.05	0.05
$SMAPS1$	0.64	0.35	-0.03	0.06
Coarse resolution data				
$SMAPL3$	0.76	0.58	-0.04	0.05
$SMAPL3E$	0.77	0.59	-0.04	0.05
$SMOSL3$	0.67	0.47	-0.03	0.07
$SMOSNRT$	0.68	0.46	-0.03	0.05
$CCISM$	0.71	0.50	0.03	0.05
High resolution data aggregated to coarse resolution				
$S^2MP^*_{S1S2}$	0.58	0.38	-0.06	0.06
$S^2MP^*_{S1S3}$	0.56	0.38	-0.05	0.06
$CoperSSM^*$	0.53	0.20	0.05	0.07
$CoperSWI^*$	0.73	0.47	0.05	0.05
$SMAPS1^*$	0.79	0.44	-0.02	0.04

Here are the results after reducing the temporal revisit of the coarse resolution datasets:

Table 4. Evaluation of the HR and CR SM time series against in-situ measurements in terms of Pearson correlation (R,R^a), bias (remotely sensed minus ground based SM in [$m^3\,m^{-3}$]) and standard deviation of the difference ($STDD$ in [$m^3\,m^{-3}$]). The metrics were computed by taking into account the 6 regions of study together and only the median values are shown here. The symbol * indicates the HR datasets averaged at 25-km resolution.The analysis was performed from January to December 2019.

Products	R	R^a	$Bias$	$STDD$
Sentinel-only high resolution data				
S^2MP_{S1S2}	0.59	0.36	-0.06	0.05
S^2MP_{S1S3}	0.56	0.37	-0.06	0.06
$CoperSSM$	0.53	0.18	0.04	0.08
Merged high resolution data				
$CoperSWI$	0.74	0.46	0.05	0.05
$SMAPS1$	0.64	0.35	-0.03	0.06
Coarse resolution data				
$SMAPL3$	0.81	0.55	-0.05	0.05
$SMAPL3E$	0.81	0.52	-0.05	0.05
$SMOSL3$	0.69	0.49	-0.03	0.06
$SMOSNRT$	0.76	0.45	-0.02	0.05
$CCISM$	0.73	0.48	0.03	0.05
High resolution data aggregated to coarse resolution				
$S^2MP^*_{S1S2}$	0.58	0.38	-0.06	0.06
$S^2MP^*_{S1S3}$	0.56	0.38	-0.05	0.06
$CoperSSM^*$	0.53	0.20	0.05	0.07
$CoperSWI^*$	0.73	0.47	0.05	0.05
$SMAPS1^*$	0.79	0.44	-0.02	0.04

Line 435. Maybe this is due to the fact that the CR component in such data sets is more conclusive than the HR one.

Yes, this is a good remark. Taking into account the results on the HR and CR comparisons with respect to in situ measurements this could actually be the case.

Figure 1. Please add an image explicating the location with respect to the countries.

Please find below a new version of Figure 1. As advised, it has been modified to help the reader to better locate the different regions of studies. Locations of in-situ stations have also been added. This new figure will replace the original Figure 1 into the revised manuscript.

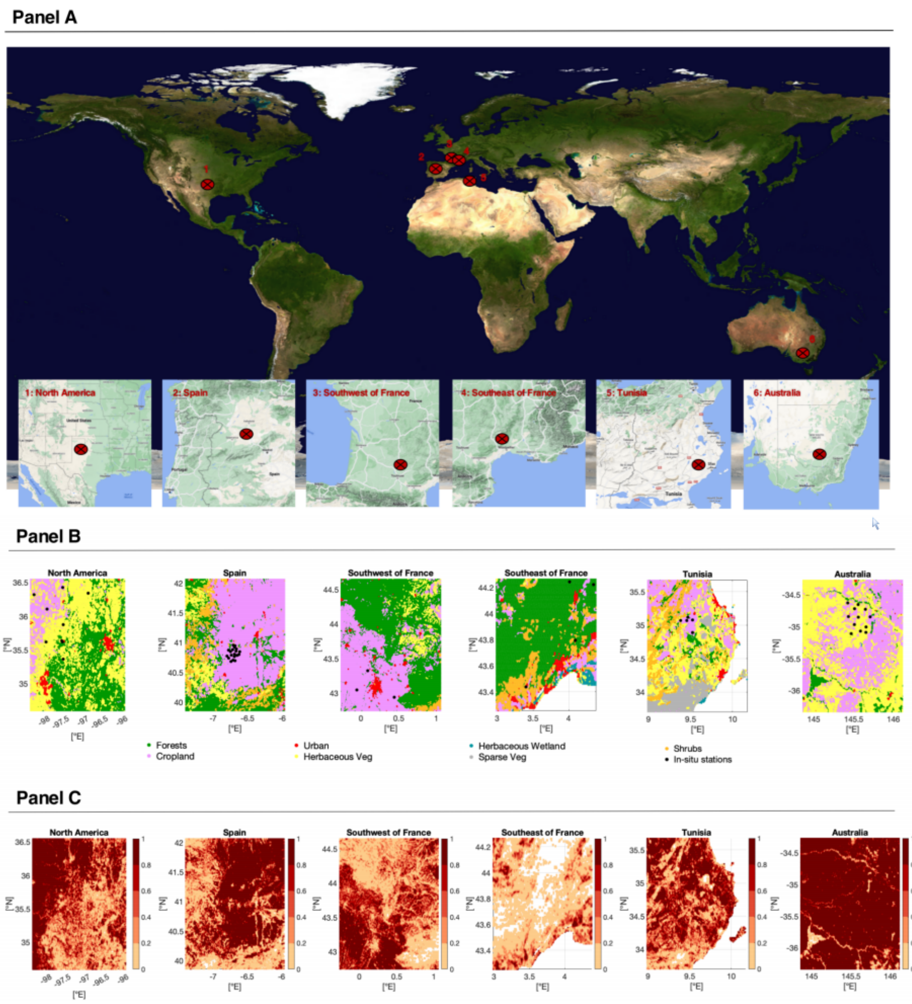

Figure 1. Panel A: Global locations of the 6 regions of study. **Panel B:** Copernicus land cover maps of the 6 regions of study aggregated at 1-km spatial resolution. Only the dominant land cover type within a 1-km^2 pixel is shown. For instance, a pixel characterised as forests can contain 27% of forests, 26% of croplands, 24% of herbaceous vegetation and 23% of shrublands, or 90% of forests and 10% of herbaceous vegetation. The in-situ stations are shown as black dots. One black dot can correspond to several sensors since some of them have the same coordinates. **Panel C:** Proportion of croplands and herbaceous vegetation within each 1-km^2 pixel for the 6 regions of study. The proportion is expressed as a percentage ranging from 0 to 1. Pixels with no cropland or herbaceous vegetation at all are shown as white areas.

Figure 4. The low spatial variability of the shown indices makes the figure not so informative.

Please find a new version of Figure 4 where the colorbar minimums and maximums have been adapted to better match the actual dynamical range of the metrics (R, bias and STDD) and to better show the spatial variations. This new figure will replace the original Figure 4 in the revised manuscript.

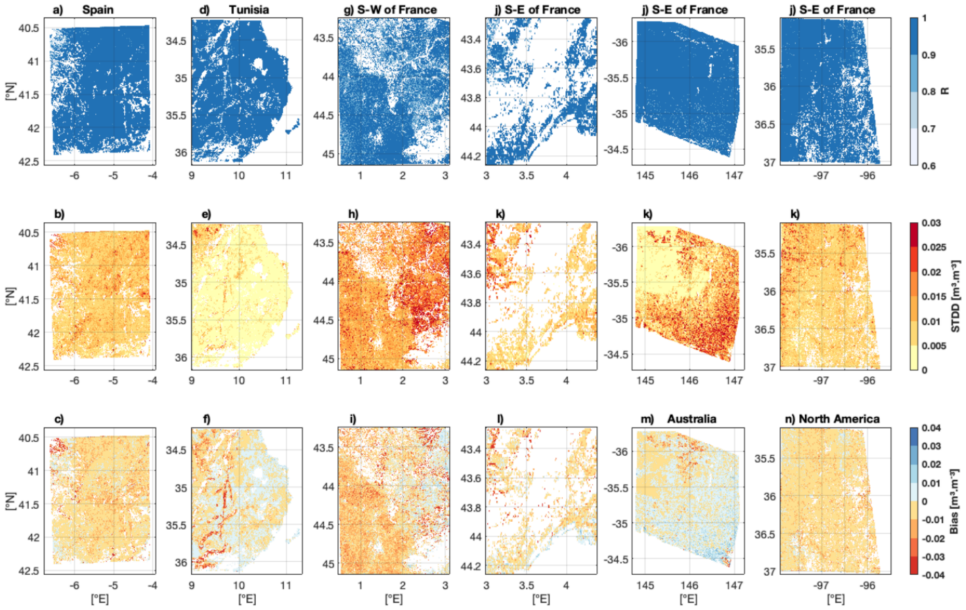

Figure 4. Comparison of S^2MP_{S1S3} with respect to S^2MP_{S1S2} over the regions of study in terms of Pearson correlation (R) as well as bias (S^2MP_{S1S2} minus S^2MP_{S1S3}) and standard deviation of the difference ($STDD$) in m^3 m^{-3}. The analysis was performed from January to December 2019.