First of all we would like to thank the reviewer for the detailed and helpful review to further improve the quality of our paper. Below you can find the answers to the reviewing comments.

Validation strategy & L9 P14793:

We agree that the number of stations for verification is very limited. As there are no more stations available according to our knowledge for the target region we were interested in, we see no possibility to overcome this problem although we agree that from the statistical point of view the results are of low significance. It is also true that the two stations are not within the same pixel. Again, this is a problem that cannot be solved due to the low number of in-situ measurements on the one hand and the given model grid on the other hand. Anyway, the lack of a large verification data set (usually due to financial constraints) should not prevent us from testing new approaches. For sure the shortcomings of the small verification data set have to be stated in the paper, and if not done yet we will point that out more clearly.

Lack of in-situ data is also the reason why there is no spatial comparison within this study as it is supposed that the investigation of spatial structures with just two verification stations is of low significance. Investigating the spatial component is interesting due to the ensemble approach: The factor which is applied during the downscaling is temporally constant, but due to the several forecast realizations for each time step, the mean value (which has been used for verification) is temporally not constant, which is evident when comparing Fig. 2 e and f.

Results for the two stations are not consistent:

This is true just for the RMSE (not significant) but not for the PCC, which is improved (not significantly) for both stations. Anyway, the direct model output is improved for all of the seven seasonal forecast runs and both stations due to the calibration + downscaling, which is a good result. The fact that the downscaling is slightly degrading output compared to the calibration in some cases for one station might be further investigated, e.g. by testing the spatial variance of the ASAR downscaling signal in the vicinity of the COSMOS stations.

Innovative aspect & L1 P14784:

To our knowledge, CDF matching is usually used to match measurements to model reality, but not vice versa. If it is also common to bias correct model output to satellite measurements, it would be helpful if the reviewer could name some examples from literature to compare them to the approach presented here. Furthermore it has to be mentioned that the application of the calibration to get rid of the model bias (in relation to ASCAT measurements) is indispensable to apply the downscaling, which is per definition designed for ASCAT. So it is the new combination of two existing and tested approaches which is the main innovative aspect of the paper. For sure, this can and will be stated more clearly throughout the paper.

The validation strategy is indeed weak from the statistical point of view; the reasons are already discussed in the answer to the validation strategy.

Downscaling method in Wagner et al., 2008 & L22 P14790:

The suggestion to create scatterplots comparing ASAR and ASCAT data for the target region is appreciated by the authors. Anyway, the added value of the 1km product is shown by the improved

statistical measures (RMSE and PCC) for one and improved (RMSE) the other verification stations. Again, from statistical point of view it is a weak proof.

<u>L21 P14784:</u>

The temporal stability assumption is described on P14790, L5-10. Soil moisture is varying in time, but the variations are usually very similar across wide ranges of scales as they are often influenced by the same parameters (e.g. precipitation).

<u>Fig. 2:</u>

The intention of Fig. 2 is to demonstrate the effect of the calibration and downscaling on the data set. For this purpose, one station seems to be enough from our point of view.

<u>L23 P14793:</u>

It is true that not all possible sources of errors are mentioned in this statement. Therefore it will be reformulated, adding the proposed additional sources.

<u>L5 P14784:</u>

Indeed this sentence is formulated unclear and will be rephrased. For sure there are several sources of uncertainty which are influencing soil moisture forecasting performance.

<u>L3 P14787 & L18 P14788:</u>

In principle the approach proposed by the reviewer is the preferred one, but there is one major problem to use it. According to Wagner et al. (2013, Meteorologische Zeitschrift, Vol. 22, No. 1, 5‑33), the backscattered measurements are used to estimate the surface soil moisture content (their Eq. 1), which is a number ranging between 0 (dry) and 1 (wet), usually expressed in %. To convert it to volumetric soil moisture content (being comparable to ECMWF model output), the soil porosity is necessary (their Eq. 2). As we have no porosity data for the target region in Kenya, this conversion is not possible, thus model output and COSMOS measurements had to be converted to the index with values between 0 and 100% to make all data sets comparable. So it was possible to avoid the introduction of an additional source of error (the unknown soil porosity) in the comparison.