We thank all reviewers for their suggestions. The original comments are in black normal fonts. The answers are in blue italic fonts. Modified text is underlined. Figure, page and line numbers generally refer to the old manuscript. An indication of anticipated new figure, page and line numbers are provided in brackets.

Referee #2

General comments

Thank the authors for this interesting work. Based on a catchment land surface model and the advanced EnKF data assimilation technique, this study employed several ex- periments trying to answer the question on "how to make the best use of L-band mi- crowave satellite observations through DA". Prior to the assimilation, a sophisticated data quality control, model perturbation, and bias mitigation in both TB and SM retrievals are applied. Finally, DA outputs are carefully evaluated by comparing with in-situ measurements and with special attention on DA innovation and increments. The findings provide important inspirations for further SMAP DA and the manuscript is overall well organized. However, some statements within this manuscript remains unclear to me and more details are needed. I therefore recommend this manuscript being published in Hydrology and Earth System Sciences by taking care of the following minor comments.

We thank the reviewer for the encouraging review and for all suggestions.

Specific comments

1. P1, Line 9-10: soil moisture evaluations are based on anomaly rather than the absolute values. Thus I would suggest rephrasing this sentence as ". . . to model-only simulations in terms of unbiased root mean square difference and anomaly correlations during the period . . ."

This will be edited accordingly (p.1, L9-10 [p.1, L9-10]).

2. P3, Line 5: what does the "treatment" exactly refer to? Do the authors mean RFI and uncertainty screening and re-gridding as depicted in section 2.2? if yes, I would not consider this as a major difference compared with previous studies.

The "treatment" mainly refers to the interpretation of the spatial support of the SMOS retrievals. For the Tb observations and observation prediction, the "treatment" refers to both the spatial support and the improved RTM forward simulation. This will be edited for clarification: (p.2, L28 [p.2, L29]), (p.3 L5 [p.3, L6]), (p.5, L6 [p.5, L13-15]).

3. P5, Line 1-6: what is a "footprint scale"? As I read from Table 1 of De Lannoy et al. 2014b, the RTM calibrated parameters are assigned to the same IGBP vegetation class, but how do you manage to make "all 36-km grid cells within one footprint area are assigned the same set of RTM parameters". When you practically do assimilation for a specific

"footprint", does all the 36-km grids use the same RTM parameters or they are vegetation-class-dependent? Please elaborate.

The footprint scale refers to the area effectively observed by the satellite. We will rephrase the statement and further elaborate in the text to explain that each 36-km pixel uses its own RTM parameters in forward modeling of the data assimilation (p.5, L2 [p.4, L8-11]).

With regard to Table 1 in De Lannoy et al. (2014b): there is a little misunderstanding here, as this table showed the imposed background information (not calibrated) per vegetation class, whereas the calibration definitely happens for each pixel separately.

4. P5, Line 30: what criterion do you use when excluding frozen soil and snow cover during assimilation?

We will edit the text to explicitly include the thresholds (p.5, L30 [p.6, L6]).

5. P8, Line 23-25: it looks the representativeness error during the upscaling process as described in P3, Line 26-29, is not considered.

We agree that the text needs a clarification (p.5, L25 [p.9, L12]).

6. P8, Line 6-7: the work of Reichle and Koster (2004, GRL) looks a better reference on CDF-matching. Besides, on what temporal scale did the authors conduct this CDF-matching? Is it for each month, year, or the entire study period (2010-2015)? As also stated by the authors in P10, Line 5-6, could the SM innovation be seasonally corrected as well if do CDF-matching on a monthly basis? Please clarify.

We agree; this was a bibtex/latex glitch. The reference will be updated.

The text already mentioned that the retrievals were not corrected seasonally, but we will rephrase it slightly to further clarify the issue (p.8, L8 [p.8, L27-30]).

7. P10, Line 13-19: another reason for the degradation of TB assimilation might be the modeled vegetation. Meanwhile, Leaf area index, other than vegetation water content, has been found to be reliable in estimating vegetation optical depth at global scale (Kerr et al. 2012, IEEE-TGRS). To ingest real-time dynamic vegetation observations (e.g., LAI from MODIS) might help mitigate the TB assimilation. In any case, RTM over highly vegetated land cover is always tough, and the authors may consider excluding areas with vegetation water content over 5 kg m-2.

We agree with this thought and we believe that it was already articulated in our submitted manuscript, but we will add an extra note for clarification. We already had stratified the data assimilation results into favorable and non-favorable areas (Fig. 7-8), with favorable areas excluding grid cells with vegetation water content greater than 5 kg/m^2. We will further clarify this (p.10, L18 [p.11, L1]).

8. The in-situ soil moisture is usually measured at the 5 cm depth whereas the model output represents the first layer's average (0-5 cm), and this vertical depth-mismatch can potentially introduce biases in soil moisture evaluation given that the topsoil moisture usually have larger variations. Horizontally, the direct comparison between model estimate of a gridcell average and point-scale in-situ observations can also be ques- tioned due to high sub-grid heterogeneity. For the former, it could be alleviated by configure the land model to have denser soil layers at the top. For the latter, a way of mitigating this spatial representativeness issue is to compare their spatial averages (e.g., Xia et al. 2014, JH). However, I realize it might be difficult for the authors to reconfigure the land model or redesign the evaluation scheme within this paper but it can be considered in future studies.

We agree and we will keep this in mind, but indeed the model cannot be easily re-configured, because it is part of a larger operational system.

The issue of the spatial (vertical and horizontal) mismatch between in situ measurements and model or assimilation results is partly circumvented in our paper by using bias-free validation metrics, as clarified on p.5, L.27 [p.6, L4]. Also, we prefer to first calculate the metrics at each site and then take a spatial average, instead of switching these two operations, because we believe that our sequence of operations is more conservative and has less risk of hiding local problems.

9. Similar DA framework has already been used in the SMAP_L4 algorithm to produce a value-added root zone soil moisture. Thus in the conclusion section I would like to see a short paragraph of the authors' speculation on possible improvements in the future SMAP TB and SM assimilation as well as the feasibility of their joint assimilation given that these two products complementarily have different spatial coverage and content different land surface information.

Our reference to Reichle et al., 2016 gives the latest available update on the SMAP L4 product. We will further add a few thoughts in line with the suggestion by the reviewer ([p.16, last paragraph]).

Technical corrections

1. P7, Line 14: "as well as surface soil temperature. . ."

The original sentence is actually correct, but for clarification we will adjust it (p.7, L14 [p.7, L1]).

2. P22, Figure 3: should the captions of g, h, and i be for subplots j, k, and l?

(Figure 3 [Figure 4]) Good catch, many thanks.

Additional references

Reichle, R. H., and R. D. Koster (2004), Bias reduction in short records of satellite soil moisture, Geophysical Research Letters, 31(19), L19501, doi:10.1029/2004GL020938.

Kerr, Y. H., et al. (2012), The SMOS Soil Moisture Retrieval Algorithm, Geoscience and Remote Sensing, IEEE Transactions on, 50(5), 1384-1403, doi:10.1109/TGRS.2012.2184548.

Xia, Y., J. Sheffield, M. Ek, J. Dong, N. Chaney, H. Wei, J. Meng, and E. F. Wood (2014), Evaluation of Multi-Model Simulated Soil Moisture in NLDAS-2, Journal of Hydrology, 512, 107-125, doi:10.1016/j.jhydrol.2014.02.027.