**Response to Short Comments from Alan Frank**

Note that the reviewer's comments are in italic black, and responses in blue.

We thank the reviewer for his/her helpful comments. Here, we carefully address the issues on knowledge gaps raised by the reviewer.

*I thought this paper would be interesting while I saw the title. But after reading the entire article, the value is not apparent. In particular, two major things are lacking. First is any type of theoretical motivation. Second, the empirical motivation is not sufficient to demonstrate that this idea is particularly valuable. The latter is because none of the current multi-source blended soil moisture products uses either of the methods in the article for soil moisture blending. At least in this study, the authors failed to demonstrate their "simple method" can produce one dataset with higher (even similar) accuracy than any current dataset.*

We respectfully disagree with the reviewer's statements. The theoretical motivation of this study is clearly articulated in the introduction. The advances and limitations of recent research related to this topic culminates in the four knowledge gaps that are addressed by our study.

With regard to the empirical motivation, least-square weighting and error estimation from Triple Collocation (TC) have been adopted by several studies to blend different soil moisture products. For example, Yilmaz et al. (2012) adopted the least square framework to achieve an objective blending of satellite and modeled soil moisture. The same methodology was adopted by Zeng et al. (2016) to merge the satellite and reanalysis soil moisture data.

In this study, the authors have demonstrated the "simple method" can produce one dataset with similar and higher accuracy than current datasets (e.g., SMAP and NLDAS soil moisture). For example, Fig.8 illustrates that dataset generated by simple average, AVE(K, N), has significant ($p<0.05$) smaller MAE, RMSE than those of SMAP and NLDAS products. Fig. 10 also demonstrated that the AVE(K, N) has similar and even smaller errors than K-API in places where in-situ measurements are sparse, such as the central to the south of Texas and east Arkansas. Therefore, the dataset produced by "simple method" is an improvement over currently available soil moisture products due to its reduced uncertainty.

Yilmaz, M., Crow, W., Anderson, M., and Hain, C. An objective methodology for merging satellite‐and model‐based soil moisture products, Water Resources Research, 48, 2012.

Zeng, Y., Su, Z., van der Velde, R., Wang, L., et al. Blending satellite observed, model simulated, and in situ measured soil moisture over Tibetan Plateau, Remote sensing, 8, 268, 2016.

*This article lacks any meaningful science: there is no hypothesis statement (why should we expect these simple methods to be interesting or useful?), and the control method is insufficient.*

For operational soil moisture mapping, computation-efficiency is of great concern. The statistical methods require fewer input variables, parameterization process, and computation resources, thus are ideal for operational usage. Focusing on statistical methods, the primary question of this study is whether the simplest blending method (e.g., simple average) can perform similarly to the relative complicated method (e.g., least-square weighting using error variances from TC, REV or RK) to achieve higher computation efficiency without losing accuracy. We will add this clarification to the introduction section in our revised manuscript.

In addition, different controls on soil moisture formats (e.g., absolute soil moisture, anomalies, and percentiles) and error estimation methods (e.g., TC, REV and RK) were conducted to find out the impact of data formats and error estimation methods on final results.

*In line 136 to 149, the authors listed four knowledge gaps and they claimed that they addressed all four gaps. (1) "of the lack of in-situ soil moisture inclusion in product blending. Current studies mainly focus on combining modeled and RS soil moisture, rather than combining all three sources (modeled, RS and in-situ). In-situ measurements can be useful for improving the accuracy of hybrid soil moisture datasets".*
*Some widely used soil moisture products, such as ESA CCI Soil Moisture (https://www.esa-soilmoisture-cci.org/) and SMAP L4 (https://smap.jpl.nasa.gov/data/), merged satellite, in-situ and model data. If the authors want to demonstrate in-situ soil moisture is critical to improving the data quality, they need to at least compare the results with those using satellite and model data only.*

It is a good point to compare the in-situ soil moisture with the datasets that only using satellite and model data to demonstrate in-situ soil moisture is critical to improving the data quality. And it is exactly what this study did.

ESA CCI soil moisture products did not incorporate any in situ data except for validation. Although SMAP L4 include some model simulations, they do not incorporate any in-situ soil moisture measurements. NLDAS soil moisture is a pure model simulation result. By comparing the blended products with the products without in-situ soil moisture (e.g., SMAP and NLDAS), Fig.8 illustrates as long as the in-situ based soil moisture is included, the averaged products, either AVE(K, N) or AVE(K, S), has significant ($p<0.05$) smaller errors than those of SMAP and NLDAS products alone. However, the averaged product from SMAP and NLDAS, AVE(S, N), did not show any improvement compared with neither SMAP nor NLDAS. Therefore, Fig. 8 indicates the necessity of incorporating in-situ based soil moisture to improve blending accuracy.

*If they want to demonstrate their simple methods has better (or even similar) performance than more advanced data assimilation approaches, they need to include at least one data assimilation approach, such as 3D-Var. Comparing those "simple methods" with simpler method (averaging) does not make sense at all.*

Operational soil moisture mapping favors the simple methods. The primary purpose of this study is to propose and test two different, yet simple methods (e.g., simple average and least-square weighting) for blending soil moisture data. That way we don't compare with a much more complex data assimilation approach.

*(2) "There is no comprehensive evaluation of different data blending methods . . . compare the accuracy of different blending methods to identify the optimal approach for soil moisture"*
*This is a red flag. Regardless whether there is any document comparing different blending methods, being the first to compare some methods is not justification for publishing a study on these methods. Simply, I don't care if you are the first to do something; I only care if you have an interesting idea. It is implied that you are the first to do whatever you are doing, because otherwise it is not publishable (with certain rare exceptions). So instead of stating the obvious (that you think what you are doing is novel), tell me what problem you want to solve or what question you want to answer.*
*The conclusion from this article cannot convinced me the approach is optimal. It can only indicate a best one among these four approaches. Even so, in the abstract, the authors stated "no significant differences between blended soil moisture datasets using errors estimated from TC, REV or RK. Moreover, the LSW did not outperform AVE", which means nothing.*

Thank you for your comment. We will rephrase the statement in the revised manuscript as "(2) lack of understanding on the impact of estimated error variance over soil moisture blending. Currently, there are a variety of methods for estimating error variance of different datasets, such as TC, Kriging and Relative Error Variance (REV) (Vinnikov et al., 1996; Ford and Quiring, 2019). It would be helpful to compare these methods and identify their impact on soil moisture blending."

We agree the word "optimal" was used imprecisely in a number of places in the paper. The purpose of this study is not to find the optimal or absolute best method for blending datasets, but instead to propose and test multiple methods to find a simple, elegant solution for blending soil moisture data across a large area.

*(3) The impact of measurement units (e.g., volumetric water content, soil moisture anomalies, and percentiles) is unknown. For example, is it better to convert all of the soil moisture measurements to anomalies or percentiles before blending?*
*This might be the only question that is answered in this article. However, the answer to this question seems cherry-picked. Through standardizing volumetric water content, you will get a better result. When the length of record is short (which is in this study), the standardized anomalies or percentiles would be a proxy of ranking.*

Thank you for your agreement on our third knowledge gap. However, we don't agree this is a cherry-picked answer, nor do we agree that it is inherently proven that standardizing volumetric water content necessarily results in better results. However, our findings do provide empirical evidence to this assumption, and are therefore important additions to the soil moisture literature.

With regard to record length, Ford et al. (2016) demonstrated that sample sizes of 93 to 186 daily soil moisture observations are sufficient to generate robust percentiles. In our case, SMAP has the shortest data record (3 years) among the three products, with a total of 93 data points (31 days in window × 3 years). Therefore, three-year data have met the requirement to generate robust percentiles. We would confirm this using longer data records in future studies.

Ford, T. W., Wang, Q., and Quiring, S. M.: The observation record length necessary to generate robust soil moisture percentiles, Journal of Applied Meteorology and Climatology, 55, 2131-2149, 2016.

*(4) A simple and operational methodology is still needed for accurate daily soil moisture mapping with high spatial resolution. Current methods to generate gridded soil moisture data products cannot produce data with sufficient spatial resolution for many agricultural and hydrological applications.*
*Both remote sensing and model simulation are discussed in the Introduction. Regarding remote sensing observations, it is true that the resolution is poor. In this article, all the evaluations are performed at station-level. If you have in-situ data available and assume that that is the most accurate data, why do you need to blend it with other datasets?*

Although station-based soil moisture observations are usually assumed as the most accurate data, this does not mean station measurements are perfect and have no limitations. In fact, each source of soil moisture information has its advantages and disadvantages. The advantage of in situ observations is that they are the only source of ground soil moisture information, and thus are often used as a benchmark for models. The primary disadvantages of in situ observations are 1) sparse spatial density and 2) limited spatial representativeness. Similarly, model and satellite remote sensing soil moisture have the advantage of representing a larger spatial area and, for the most part, a finer spatial resolution. Of course, the primary disadvantage of models and satellites is that they are indirect representations of soil moisture variations and not direct measurements. Therefore, these three soil moisture data sources are complementary and, as we show in this study, can be combined to improve soil moisture monitoring.

We will add this clarification to the introduction section in our revised manuscript. We will also be more specific with this knowledge gap and revise it to "(4) A simple and operational methodology is still needed to leverage multiple, diverse sources of soil moisture information for accurate daily soil moisture mapping at high spatial resolution".

*If the authors aim to create a dataset with higher spatial resolution than remote sensing and model datasets, please show it. It is unfortunate that I did not see any improvement in terms of spatial resolution. What I saw is that the spatial resolution is determined by Noah or SMAP L4.*

The blended products do have a higher spatial resolution (4-km) compared with the remote sensing and model datasets. As listed in Table 1, the spatial resolution of SMAP data is 9-km, 1/8° for the NLDAS data. In this study, the nearest neighbor assignment is used to resample/downscale the SMAP L4 and NLDAS data to 4-km, which matches the spatial resolution of PRISM precipitation and interpolated soil moisture. Finally, the blended soil moisture products have a spatial resolution of 4-km.

*I also have some other serious concerns: 1. SMAP L4 is an assimilated product, which includes some model simulations. When this article does the blending with Noah model, do you think you double weighted the model simulations?*

This is a good question. The answers to this question are threefold:

(1) Although SMAP L4 includes model simulation, it is not a pure model simulation. Therefore, blending SMAP with NLDAS may not double weight the model simulation.

(2) Fig. 6 showed different weighting schemes calculated using TC, REV and RK-based methods for the three products (K-API, SMAP, NLDAS). However, no significant differences are observed when different weighting schemes are applied in Fig. 8. This indicates the weighting scheme has little impact on the blending results. Therefore, even we assume blending SMAP with NLDAS may give a little higher weight on model simulation, this still has a minor impact on the blending results.

(3) Fig. 8 also showed blending SMAP with NLDAS presented the largest errors among all blended products. In contrast, the inclusion of in-situ based soil moisture can significantly reduce errors. Therefore, the weighting on SMAP or NLDAS is not the critical impact factor on blending results, but the inclusion of the in-situ based soil moisture.

*2. Three-year data is too short to sufficiently calculate soil moisture anomalies and percentiles.*

We agree that a long record of data is ideal for any analysis. The data length of this study is limited by the SMAP data, which is only available since 2015. In addition, Ford et al. (2016) demonstrated that sample sizes of 93 to 186 daily soil moisture observations are sufficient to generate robust percentiles. In our case, SMAP has the shortest data record (3 years) among the three products, with a total of 93 data points (31 days in window × 3 years). Therefore, three-year data have met the requirement to generate robust percentiles.

Thanks to the reviewer's comment, we will include the short data period as the current limitation in the revised manuscript and propose to use longer record in future work.

Ford, T. W., Wang, Q., and Quiring, S. M.: The observation record length necessary to generate robust soil moisture percentiles, Journal of Applied Meteorology and Climatology, 55, 2131-2149, 2016.

*3. I wondering how this work can be applied in other regions with sparsed soil moisture sites. Figure 10 shows that MAE is larger over those regions with less sites (which is a common situation over the world).*

This is a good question. Our study (Fig. 10) also showed averaged product AVE(K, N) (blue circles), had similar and sometimes smaller errors than K-API (red circles) in places where in-situ measurements are sparse, such as the central to the south of Texas and east Arkansas. This indicates when in-situ measurements are sparse, using additional sources of soil moisture information (such that from NLDAS) may help to increase the accuracy (lines 705-710). Fig. 12 further confirmed this point by assessing K-API and AVE(K, N) with varying numbers of sampling points and different sampling schemes. Therefore, when applied in other regions with fewer soil moisture sites, it is suggested to adopt the simple averaged product AVE(K, N) according to this work.