Hydrology and
Earth System
Sciences

Discussions

Interactive comment on "Blending SMAP, Noah and In Situ Soil Moisture Using Multiple Methods" *by* Ning Zhang et al.

Anonymous Referee #1

In this manuscript, authors have analyzed the performance of different merging methodologies using NLDAS NOAH, API-derived, SMAP, and in-situ soil moisture estimation/observations. The study considered different merging techniques (simple averaging, least squares weighting) while the error variances needed for the least squares weighting are estimated using TCA and REV. However, I have serious concerns about the added innovation obtained from the study:

The last paragraph of the introduction section lists the innovation in the study as "(1) of the lack of in-situ soil moisture inclusion in product blending (2) There is no comprehensive evaluation of different data blending methods, (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? (4) A simple and operational methodology is still needed for accurate daily soil moisture mapping with high spatial resolution."

On the other hand, out of these listed items: 1) I am not sure about the utility of using in-situ soil moisture datasets in operational applications. If there are station-based observed datasets, then why not directly use them rather than using other datasets? Afterall, station-based datasets are used in soil moisture validation studies and arguably they represent the best moisture conditions on the ground. If the added benefit from blending in-situ datasets is about the regions lacking in-situ observations while dense-networks are present around, then this significantly limits the applicability of the introduced study to regions with dense soil moisture networks (e.g., Oklahoma). On the other hand, there are not many such dense networks that facilitate retrieval of soil moisture by blending information coming from station-based observations and other ancillary datasets (e.g., remote sensing-based precipitation or soil moisture estimates). If this is the case, then 1.a) the motivation of the study & the structure of the introduction section should be given accordingly (i.e., why soil moisture retrieval over Oklahoma-like regions is very important) and 1.b) the study area should be re-selected accordingly by excluding the regions not having ground-stations (i.e., the study area could be limited to state of Oklahoma having $\sim$ 180,000 km2 area + region laying between 100W-103W & 33N-36N having $\sim$90,000km2 area, rather than total 1,150,000km2 area used in this study). For example, there are regions without any station data over south of Texas; I cannot imagine the sparse networks outside of these regions will add any useful information to these regions lacking any in-situ data.

2) Here in this study only simple merging and least squares merging methodologies are compared. REV and TCA are error variance estimation-related methodologies (i.e., not blending), while kriging is more like a spatial interpolation rather than being blending methodology that facilitates merging of two, three, four datasets, unlike least squares or simple merging methodologies. Given earlier studies have already performed least squares - simple merging methodology comparison (e.g., as cited by authors, Yilmaz

et al., 2012 have already done it), I don't agree with the statement that "comprehensive evaluation of different data blending methods" has not been implemented. Here, if the contribution is about "investigation of the impact of estimated error variance information over the blending", then perhaps this should be reflected in the title as well as the motivation given in the introduction.

3) 3.a) It seems apples and oranges are being compared. Even though native datasets and anomaly information are utilized separately in the merging methodology, their comparisons on an equal ground have not been performed. It does not make much sense to compare the errors of the native datasets and errors of the anomaly components of the same datasets (i.e., native time series are composed of two independent components, called in this study anomaly and climatology). If the goal is to obtain a product that is more like raw-product in nature, then the merged anomaly product should be converted back to the space of native product by adding the already-subtracted climatology component, and then the errors of this anomaly + climatology merged product could be directly compared against the errors of the native product-like merged estimate. 3.b) I am not convinced that merging native products without a proper rescaling methodology is justifiable. For the last two decades numerous of studies have clearly shown that there are systematic differences between the statistics (e.g., mean and standard deviation) of soil moisture estimates (e.g., Dirmeyer et al., 2004; and Reichle & Koster, 2004). These systematic differences should be removed via certain rescaling methodologies before they could be merged (Afshar et al., 2019). The results shown in this manuscript are also very consistent with these earlier studies: absolute value merging (i.e., no rescaling before merging) yields the worst performance (nash value 0.25) compared with rescaled versions via anomalies and percentiles (nash values 0.60 or higher). Accordingly, all "absolute value" related investigations are redundant, hence should be removed from the study.

Dirmeyer et al., (2004). Comparison, validation, and transferability of eight multiyear global soil wetness products. Journal of Hydrometeorology, 5(6), 1011–1033.

Reichle & Koster (2004). Bias reduction in short records of satellite soil moisture. Geophysical Research Letters, 31, L19501.

Afshar et al. (2019). Impact of rescaling approaches in simple fusion of soil moisture products. Water Resources Research,55, 7804–7825.

4) The manuscript states "A simple and operational methodology is still needed for accurate daily soil moisture mapping with high spatial resolution". I strongly believe "simple methodology for accurate daily soil moisture mapping with high spatial resolution for operational applications" exists. Authors should clearly state what is missing in the established literature with more detail.

5) There are many unjustified/redundant statements in the manuscript all over. They hinder the impact of the delivered messages. - Lines 66-79, following earlier studies (Dirmeyer et al., (2004) and Reichle & Koster, 2004), there is no use in stating the facts that there are systematic differences between the model-based soil moisture products. - Lines 121-123, "However, none of them, at least by themselves, are adequate for providing accurate soil moisture data at high temporal and spatial resolutions.". Noah model runs at 1km spatial resolution and 1-hour temporal resolutions exist (e.g., LIS model-based runs can simulate soil moisture at 1km spatial resolution and 1-hour temporal resolution globally). Is it not adequate in terms of spatial and temporal resolutions? - Lines 123-125, "Therefore, it is useful to combine these three independent data sources to capitalize on the strengths of each and to generate an optimal soil moisture product to facilitate real-world applications.". Without merging datasets real-world applications can not be facilitated? So, does it mean model runs with 1km spatial resolution and 1-hour temporal resolution is not sufficient? - Lines 137-138, "Current studies mainly focus on combining modeled and RS soil moisture, rather than combining all three sources (modeled, RS and in-situ)", over which locations? Given such in-situ datasets are limited what is the added benefit obtained from in-situ observations globally? Should we only focus on local studies? But, if we focus on local studies and know the in-situ data, then why merge such very high-accuracy products with datasets

with much lower accuracy? - Lines 148-149, "Current methods to generate gridded soil moisture data products cannot produce data with sufficient spatial resolution for many agricultural and hydrological applications.", I don't believe current methodologies lack the ability of producing sufficient spatial resolution datasets for the applications. Again, 1km spatial resolution model runs are already globally available (i.e., current manuscript considers only 4km resolution product). - Lines 515-516, "Our study also demonstrates that the measurement units (Fig. S4) do not impact the relative relationship (error ranking) between the different datasets". I don't really see how this result is obtained from this figure. - Line 685, "... sites that are less temporally representative sites ...". "Temporally representative" phrase should be elaborated.

6) "In this study, a set of k values (from 0.80 to 0.99) is tested to determine the optimal k value that results in the highest correlation between API and soil moisture based on 215 stations." But this is clearly overfitting: k values are fit to yield API values that give highest correlations, while these API products are later used to obtain gridded products which are later validated using the same in-situ datasets that the API is calibrated against.

7) Equation 14 introduces a rescaling step before the triple collocation given in equations 15-17. On the other hand, Stoffelen (1998) clearly introduced another methodology to rescale the datasets before the error estimation (equation 2 in the study of Stoffelen, 1998). Without proper rescaling step, the triple collocation methodology (i.e., equations 15-20) would be void. Accordingly, the use of equation 14 before equations 15-20 is absolutely not acceptable (i.e., if the original rescaling steps given by Stoffelen, 1998, is used, then equation 14 is redundant). The original methodology introduced by Stoffelen (1998) must be followed.

8) "Our preliminary results showed that the choice of reference dataset did not impact the final results, thus the RK-gridded soil moisture is selected as the reference dataset in this study." There is plenty of literature available showing that reference dataset selection matters in rescaling soil moisture products (Afshar et al., 2019). Accordingly,

C5

the manuscript should show which "final results" are not impacted.

9) In order to show the RK-gridding impact, manuscript also show the performance of merging SMAP, NLDAS, API products (i.e., API itself without blending with in-situ data using kriging). Only after this step the manuscript can claim some added benefit obtained from kriging methodology (i.e., otherwise the added benefit seems to stem from API dataset).

10) It seems "displacement of autocorrelation" lies at the hearth of the "Relative error variance" methodology, while equations 21-23 does not really show how exactly it is calculated. Frankly I did not understand "The displacement of the autocorrelation \gamma(\tau) at \tau = 0". I am puzzled with "autocorrelation with 0-lag" (i.e., to me autocorrelation at 0-lag should be equal to 1; this clearly shows I am missing something while the manuscript does not help me). Step by step instructions are needed (i.e., given I have soil moisture time series, how exactly I should code this? The details should give this much information).

11) Giving too much emphasis to the differences between the raw products (Lines 465-480) does not make much sense to me given earlier studies already says it (Dirmeyer et al., 2004, and Reichle & Koster, 2004). I think this paragraph is redundant.

12) Fig 4 shows four different TCA results using different products (assuming different reference datasets are selected in each of these TCA calculations). Accordingly, comparison of absolute value TCA errors retrieved using different reference datasets does not make much sense. I strongly recommend authors to have a look at the study of Draper et al., (2013) in this context.

Draper, et al. (2013). Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. Remote Sensing of Environment 137, 288-298.

13) The word "significance" has been written > 25 times, while not a single sentence is

C6

written how exactly these "significance tests" are applied (i.e., at what confidence level, which significance test?).

14) Variance of products must be given explicitly to convert REV values into error variance estimates. Without variance of products explicitly given, the results obtained from this study may not be compared with other studies (i.e., 0.15 – 0.18 REV values for SMAP and NLDAS would imply 0.03 – 0.034 error standard deviation if variance of these two products are assumed 0.08, [sqrt(0.15*0.08ˆ2)=0.03]).

15) The abbreviations "RK-gridded" and "K-API" are used interchangeably. Consistent use of abbreviations is needed.

16) "The in-situ measurements cannot be considered the "truth" because they are point measurements that may not reflect the soil moisture value for each 4 km grid cell." But almost all soil moisture validation efforts (e.g., Jackson et al., 2010, https://doi.org/10.1109/TGRS.2010.2051035) and this present manuscript are done using in-situ based observations. I don't agree with this statement.