Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-12-RC1, 2018

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Estimating time-dependent vegetation biases in the SMAP soil moisture product" *by* Simon Zwieback et al.

A. Konings (Referee)

konings@stanford.edu

Received and published: 24 February 2018

This paper presents a new, extended Bayesian methodology for estimating errors of remotely sensed soil moisture. The model is inspired by triple collocation approaches (and their assumed linear error model), and in some sense, extends triple collocation to allow time-varying multiplicative and additive errors. This new methodology is then applied to show that the sensitivity of the SMAP soil moisture product is influenced by its mis-specification of the vegetation optical depth, and that this could artificially inflate estimates of vegetation, soil moisture coupling. This paper could become an important contribution to the literature – the point about SMAP is quite informative given the broad use of this dataset. Furthermore, the new error characterization technique is an important advance and could (or should) become widely used. I applaud the authors

for the careful testing of the method through a simulation study and several sensitivity analyses However, as currently written, the paper is frequently lacking in sufficient detail of the methodology employed to derive its results, as I've outlined below. In particular, for each figure in the paper, what is shown in each figure and especially how it was described must be explicitly described in the text. This is not currently the case for a majority of figures. These, and a few other major concerns outlined below, need to be addressed before it can be published.

Major Comments:

A) Figure 1b lists the soil moisture as an output. If I understand correctly from the text, an explicit best guess 'true' soil moisture timeseries is never determined. This is probably the conservative thing to do – I am sure the uncertainty would be quite wide. Nevertheless, some explicit discussion/warning about the fact that this Bayesian approach is primarily for determining error statistics, and that accompanying posterior true soil moisture timeseries may not be useful (or if the authors disagree with me, some justification on that, as that would obviously be very intriguing!), is warranted.

B) Figure 2 is unclear. How is the bias defined? And how can the RMSE be greater than posterior in right-most column of Figure 2b if sigma simulation values (Table 1) are positive?

C) Even though the units are the same, it is a little confusing to have both the RMSE/bias and posterior on the same axes in Figure 2b, since the former represent a *difference*. I suggest splitting this into two rows. Than in the row where you show the posterior, it would also be useful to include the uncertainty of the posterior (through violin pots if necessary) and how it compares to the prior uncertainty. Is it actually much tighter, or has the mean just shifted? The bottom of page 7 mentions that "Fig. 2b shows that the posterior standard deviations are. . ." but I only see the posterior represented by a single point.

D) How is Figure 3a calculated? Is this assuming perfect retrieval? It must be influenced by the type of soil (influencing the dielectric mixing model) in some way. . .Also, are the different lines different average levels of true tau or something else? Please mention this also in the caption and clarify the text. What happened to the tau = 0.1 line in figure b? Did you decide to no longer use it? All of these things should be explained!

E) Fig 3b: The small clarification on the definition of L and M (which falls out of the model equations pretty easily) is negated by how long it takes to understand the figure because what it shows is barely described in the text. I suggest just removing this part of the figure.

F) Looking at Figure 4a, it is not clear visually that L is actually more closely related to delta tau than to tau itself. Can the authors check the statistics on this (preferably at all sites)? As evidenced by the sensitivity analyses the authors needed to do, estimating tau a priori is pretty difficult. If indeed L is a better match to tau directly than to delta tau, it would be easier for the understandability of the paper, and arguably more useful for future researchers' intuition about spatio-temporal variations in SMAP baseline soil moisture sensitivity.

G) More on Figure 4: The caption mentions "The magnitude of the dependence for a unit change in delta tau, lambda* is consistent with predictions by tau-omega". This is a strong-ish claim to casually throw into a caption. First of all, I'm guessing that the grey bar is some sort of model prediction from tau-omega? This needs to be explained in the caption though. It's particularly unclear since the color between the word 'model' is different than that of the grey bar. As mentioned elsewhere, the paper does not explain how it arrives at these model predictions. This has to be explained somewhere for it to be a paper that has any chance of being reproducible. Also, presumably it would not be hard to make these model predictions site-dependent (e.g. changing soil texture, estimated albedo, mean tau) – why are they constant with time? Lastly, it's unclear exactly what's going on in the right-hand column. Is it just the left hand column divided by the average delta tau at each site? If so, given that delta tau is probably as uncertain as the performance of the new methodology in this application and given

that the resulting model – estimate mismatch is actually not particularly encouraging, I suggest just leaving this out. Lastly, it would be useful if there was some discussion about what the sites mean. Are the trends in lambda and mu across sites consistent with e.g. vegetation density or canopy type characteristics?

H) Page 10, L27: I don't see why the re-analysis data error should depend significantly on delta tau at all. Why is this assumption made?

I) The baseline SMOS VOD product is known to have significant issues, because it relies heavily on an LAI-based prior (see discussion in Fernandez-Moran et al, Remote Sensing 2017). The SMOS-IC product has been developed specifically to get around this and early results are looking favorable. It is not yet publicly available to my knowledge, but the authors are quite willing to share. However, I am not sure SMOS VOD is the best 'true' VOD here – it will differ from the underlying ideal SMAP values due to differences in footprint, orbit, etc between the two satellites. Thus, I suggest using VOD from the dual-channel algorithm (either the O'Neill et al once currently used in the sensitivity analysis or I'd be happy to share our MT-DCA retrievals, which have somewhat less high-frequency noise and spatially variable albedo) instead of the SMOS VOD. The point in Figure 6 about the role of using optical data vs using a climatology for VOD would work just as well even without the first column in the figure.

J) The discussion section would benefit from some more discussion about the greater implications of this new methodology. For example, this technique might work particularly well for triple collocation of land surface fluxes of water and carbon, where it is easy to imagine significant seasonality in the error terms. Do the authors agree?

K) Similarly, can the authors discuss the implications of the normalization in Eq. 6 for the interpretation of the results?

Minor Comments:

L) Page 2, line 32: See also Momen et al, JGR-B 2017

M) Page 3, line 5: You haven't defined delta tau here

N) Figure 2: it would be helpful to explicitly explain somewhere why there are no RMSE values in the no mu, no lambda, no kappa case. It would also be easier to read the axes if there were more horizontal tick marks in each row, and if the tick labels were repeated between part a and part b.

O) Section 2.1.3: You assume quite specific priors. Would be helpful to show these distributions in the supplementary material to give the reader a sense of what they look like

P) Page 7: I suggest defining the RMSE error with equation or at least separate symbol for clarity. It's easy to miss this definition in the middle of the writing, but integral to following the rest of the discussion

Q) Page 7, line 29: How is this calculated?

R) Figure 2: Suggest splitting this into three columns: one with posterior vs. prior distribution (in violin plots if necessary), then third column with bias and RMSE.

S) Figure 2: Need to make it clearer that the 'no kappa' and no mu, lambda, kappa' simulations are cases where still have that in forward model. This is very difficult to pick out from text as is.

T) Page 11, line 5: note that this reference is broken

U) Page 16, line 1 : The authors might want to cite Crow et al, GRL 2015 here, which showed this point quite convincingly for soil moisture –latent heat coupling

V) I don't think the subscript p is ever defined. Is this an index for the number of explanatory variables?

References Cited:

Crow, W. T., Fangni, L., Hain, C., Anderson, M. C., Scott, R. L., Billesbach, D., &

Arkebauer, T. (2015). Robust Estimates of soil moisture and latent heat flux coupling strength obtained From Triple Collocation. Geophysical Research Letters, 8415–8423. https://doi.org/10.1002/2015GL065929

Fernandez-Moran, R., Al-Yaari, A., Mialon, A., Mahmoodi, A., Bitar, A. Al, Lannoy, G. De, ... Wigneron, J.-P. (2017). SMOS-IC: An alternative SMOS soil moisture and vegetation optical depth product, (March), Remote Sensing, 1–26. https://doi.org/10.20944/preprints201703.0145.v1

Momen, M., Wood, J. D., Novick, K. A., Pangle, R., Pockman, W. T., Mcdowell, N. G., & Konings, A. G. (2017). Interacting Effects of Leaf Water Potential and Biomass on Vegetation Optical Depth. Journal of Geophysical Research: Biogeosciences. https://doi.org/10.1002/2017JG004145

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-12, 2018.