Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-12-RC2, 2018

Hydrology and Earth System Sciences
Open Access
Discussions

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

W.T. Crow (Referee)

wade.crow@ars.usda.gov

Received and published: 9 March 2018

This paper describes and applies a new analysis technique to identify time-dependent biases present in remotely sensed soil moisture products. This represents a very significant methodological advancement in the tools available to examine the error structure of these products (indeed any remotely sensed product). The authors offer a compelling motivation their approach (i.e., as we start to use remotely sensed soil moisture data products for coupling applications, it is important that we develop a more sophisticated understanding of their underlying errors). In my view, this paper represents a major step in that direction and has the potential to impact a great deal of on-going research plans (including my own).

However, as is often the case with highly novel manuscripts, there are some important questions regarding the presentation and interpretation of results that needed to be cleared up prior to publication.

Wade Crow

1) What happens if there is error correlation between the explanatory variable (w) and the products (y)? There are credible reasons to suspect that this arises between the SMOS "tau" product and the SMAP L3 soil moisture product - particularly in agricultural areas. Both products suffer from a common dependence of the zero-order tau-omega emission equation and the assumption of temporally constant surface roughness. These assumptions are particularly problematic over cropland agriculture and their violation could easily induce correlated errors into both products.

A related issue is that the interpretation of SMOS tau products is known to be complicated in agricultural areas (see e.g., https://lib.dr.iastate.edu/agron_pubs/115). In fact, the "reference" SMOS tau time series shown in Figure 4a demonstrates questionable features. First, corn crop canopies (responsible for ∼60% of the land cover in the South Fork water shed) typically demonstrate a biomass plateau between growth stages R2 and R6, which in Iowa which corresponds (roughly) to between August 1 and September 15 later. This expected "plateau" is actually somewhat more consistent with the SMAP "input tau" than the SMOS "reference tau" plotted in Figure 4a. Second, the rise in SMOS tau after October 1 is almost certainly a roughness artifact associated with post-harvest tillage and not a real vegetation opacity signal. So, there are credible reasons to suspect that (at least some) of the dynamics in the "delta tau" results actually reflect error in the SMOS tau "reference" (versus the SMAP tau input).

I'm probably overstating the problems with SMOS tau product here, but the broader question is how results are impacted by the presence of (potentially non-independent) errors in the explanatory variable? Is it possible that the diagnosed time dependent vegetation bias is due (in part) to the presence of error in in the SMOS tau product?

2) Section 2.1.1 – While the notation presented here which suggests that all three soil moisture products are subject to the same error model, I couldn't find any discussion of retrieved error parameters for the other two soil moisture products (i.e., in situ and MERRA). In addition, there seems to be a break in symmetry in that the selected explanatory variable is relevant for only one product (SMAP L3) and Figure 1 seems to indicate that no explanatory is applied to the in situ product. One of the appealing facets of triple collocation is the symmetry in its treatment of all three products. Does the break in symmetry applied here (via the selection of a single explanatory variable) preclude the objective cross comparison of error results across all three products? Discussion of error results for the other two products would also help establish credibility of the approach (e.g., were in time dependent biases found in the MERRA product and did that analysis reflect the known superiority of the core network relative to the other two products?).

3) Section 2.1.2 - The auto-regressive nature of a soil moisture time series signal is (arguably) its most defining characteristic. Therefore, the application of a transformed white noise process in (5) as a temporal soil moisture model is jarring. Some discussion regarding the sensitivity of results to the lack of serial correlation in (5) is needed. It is hard to imagine that the retrieval of time-dependent bias parameters is not impacted at least somewhat by the neglect of serial auto-correlation in the soil moisture model.

4) Section 2.2 - I understand that the Bayesian interference applied here is a fairly standard statistical procedure; however, I think it would help the (general earth science) reader if the authors provided more expository detail on exactly how the MC chain is implemented to solve the Bayesian problem. I'm a little unclear, for example, on how time is handled in the analysis (i.e. the analysis conducted sequentially or as a batch process across all time?).

On a related point, I'm also not quite clear on how effective the triple collocation analogy is. For example, the decision to use N=3 products seem almost arbitrary (e.g., later on the analysis, the MERRA product is dropped with apparently minimal consequence).

Presumably, larger N equates to tighter posterior distributions; however, this is never clarified.

5) Section 4.1.1 – I had to read this section a couple of time before I realized that the in situ observations were directly used as one of the three products in the Bayesian analysis (and not withheld as some type of independent verification). Presumably, the in situ observations correspond to the "y_o" product in described in Figure 1; however, I'm not sure if that link is ever explicitly made. More clarification on this point would be helpful.

6) Section 4.2.1 – Here I missed something fairly basic. What exactly is meant by the "model" referenced in the 3rd paragraph of the section and the vertical shading in parts b) and c) of Figure 4? Presumably, the authors are referring to the tau-omega model sensitivity results shown in Figure 3. However, this is never quite made clear. In addition, it isn't clear to me exactly how the (site-independent) "model" bias parameters are calculated. As a result, I'm missing some of the insight provided by Figures 4b and 4c. Is the take-away message that, despite not being given explicit access to the tau-omega model, the Bayesian model recovers the same bias parameter results predicted by the tau-omega model? I recommend that the authors spend a little more time outlining the context behind (and the interpretation of) Figure 4.

7) Section 4.2.2 - The authors provide a nice sensitivity analysis which describes the impact of using a different tau reference on results (in the first two columns of Figure 6). In theory, this should go a long way in addressing my first point; however, (as with the case in Figure 4 above) I did not take away as much from this figure as I had hoped. The lack of sensitivity in the time-variation bias parameters to the use of a second tau references is reassuring. However, I don't quite follow why the large changes observations when using a contemporary MODIS tau indicates a lack of sensitivity to the use of MODIS tau climatology in the SMAP L3 retrievals. The delta tau generated by the MODIS contemporary minus climatology differences leads to significantly non-zero lambda and mu estimates - just not the same estimates as the application of "delta tau"

results generated relative to SMOS tau. How exactly does this support the conclusion that inter-annual tau anomalies are not a significant source of error? Some additional discussion on this point would be very helpful. I also think a fuller sensitivity discussion of results in Figure 6 here would likely go a long way towards addressing concerns I raised in my first point.

8) Section 4.2.4 – The author's link the results in Figure 7c to the presence of time-dependent errors identified in Figure 7a and 7b. However, there is a major difference in that Figure 7c results reflect climatological anomalies (lacking any seasonality) while results in 7a and 7b reflect time-dependent biases which (almost certainly) have a fixed seasonal component (which, of course, would not be reflected in an anomaly). Therefore, a substantial(?) fraction of the time dependent biases reflected in Figures 7a and 7b have no impact on anomaly results in Figure 7c. Given this, I'm unclear exactly what the relevance of Figures 7a and 7b is for the interpretation of Figure 7c (although, admittedly there does appear to be some spatial consistency across the sub-figures).

9) I also have two general comments concerning Figure 7. I'll present them as "comments" to reflect that I'm inclined to give the authors some latitude with how they respond to them:

A) The authors discuss spatial representative issues; however, the impact of upscaling a single, point-scale observation to the SMAP footprint scale should not be underestimated. While the point is never explicitly made in Chan et al. [2016]; however, a comparison of TC-based results in (their) Figures 7 and 9 suggests that the correlation between a single-point ground observation and grid-scale truth is approximately equal to that between ASCAT soil moisture retrievals and the same grid-scale truth. Given that there is strong reason to suspect that SMAP soil moisture products are significantly more precise than ASCAT products, a priori, I'd expect single-point ground observations to be a noisier source of grid-scale soil moisture than SMAP L3 retrievals over a great deal of the United States. Combined with the fact that there is likely some

error cross-correlation between SMAP L3 products and SMOS tau products (especially over agricultural sites. . .see my point #1 above), it seems possible that results in Figure 7c can be explained without the need to invoke the presence of time-dependent vegetation biases in the SMAP L3.

Chen, F., Crow, W.T., Colliander, A., Cosh, M.H., Jackson, T.J., Bindlish, R., Reichle, R.H., Chan, S.K., Bosch, D.D., Starks, P.J. and Goodrich, D.C. Application of triple collocation in ground-based validation of Soil Moisture Active/Passive (SMAP) level 2 data products. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 99:1-14. 10.1109/JSTARS.2016.2569998. 2016.

B) Point-to-grid upscaling issues associated with ground-based soil moisture observations are particularly daunting for agricultural landscapes. Most of the time the actual site isn't even located in a cultivated field (instead that are typically shunted into non-cultivated areas at the edges of the field). As a result, these measurements have no hope of capturing (often significant) inter-annual soil moisture variability associated with changes in planting, canopy development and crop development. Given the soil moisture ground measurement expertise among the co-authors, I'll defer to their judgment on this issue - but it does seem relevant to the interpretation of Figure 7c.

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