

Dear Dr Graham Jewitt,

We would like to express our thanks to you and to the reviewers for the insightful comments.

We believe the revised manuscript has improved in quality thanks to the review process. The most important changes are:

- **Model predictions:** Much more detailed description of how we derive the predicted biases from the tau-omega model and how we compare the data-driven estimates to those predictions
- **Sensitivity analyses:** We have put a greater emphasis on the results obtained using the SMAP tau as reference product (one additional figure; greatly extended discussion on potential errors in the tau product). We have also added a robustness check in which we additionally controlled for changing tau in the estimation.
- **Extended discussion:** We now have separate discussion sections on the interpretation of the results (confounding and errors), on what questions the technique could be applied to in the future, and on the implications of our findings for hydrological studies.
- **Error parameters of other products:** We now provide more details on the estimated error parameters of the in-situ and reanalysis data, but the focus remains firmly on the SMAP error parameters.
- **Improved figures:** We have revised the captions and legends of several figures.

Our point-by-point responses have already been published online. To facilitate the review process, we have also attached them to this document, followed by a copy of the manuscript with tracked changes.

Yours sincerely,

Simon Zwieback

On behalf of all authors.

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.

We are grateful to Alexandra Konings for the insightful review. We have added numerous clarifications, as we outline in our response.

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.

The algorithm estimates the posterior distribution of the soil moisture at each time step, as indicated in Fig. 1. What it does not yield is a single best guess, but rather a posterior distribution, although an estimate of the location (e.g. mean) could easily be derived from the posterior distribution. We now describe this in more detail in the section on MCMC sampling:

Each sample consists of draws from the posterior distribution, or actually an approximation thereof, of all the unobserved random variables (Output in Fig. 1b). They comprise the parameter random variables (e.g. the time-dependent biases) as well the soil moisture time series, i.e. one value of θ for each SMAP observation.

For the future, we agree that the application of this technique to product merging (i.e. estimating soil moisture by combining several products) is an interesting avenue to explore, thus building on related triple collocation results (e.g. Yilmaz, M. T., W. T. Crow, M. C. Anderson, and C. Hain (2012), An objective methodology for merging satellite- and model-based soil moisture products, *Water Resour. Res.*, 48, W11502, doi:10.1029/2011WR011682.)

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?

We have now defined the bias in Eq. 7, and similarly the RMSE is now defined in a separate equation (Eq. 6). The text has similarly been extended, and so has the caption.

The dot refers to the posterior standard deviation, as we now make clear in the legend. It is also mentioned in the caption and in the text.

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. Then in the row where you show the posterior, it would also be useful to include the uncertainty of the posterior (through violin plots 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.

As we state in our response to point B), we believe there has been a misunderstanding due to our insufficiently clear wording: the dot represents the posterior standard deviation. We believe this confusion arose due to the bad wording in the legend, which we have fixed. We now denote the posterior standard deviation by s_p throughout (text and figure). We contend that these quantities are directly comparable: for instance, asymptotically the posterior standard deviation of a parameter coincides with its RMSE (in a frequentist setting), provided certain regularity assumptions apply.

The posterior standard deviation is indeed considerably smaller than the prior standard deviation, i.e. the data tighten the distribution of a given parameter. For instance, the posterior standard deviation of μ shown in Fig. 2, is $<0.01 \text{ m}^3\text{m}^{-3}$ and thus more than an order of magnitude smaller than the prior standard deviation of $0.3 \text{ m}^3\text{m}^{-3}$. We hope the new figure that shows the prior distributions will help readers to gauge this difference (see point O).

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 τ 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!

We have amended the figure accordingly. We make clear that the two τ levels are the prescribed τ in the forward simulation, which is now described in much more detail:

To compute the predicted biases in Fig. 3a), we assumed the τ - ω model applied and was correctly specified (temperature, dielectric mixing model [Dobson; silt loam], single-scattering albedo $\omega = 0.05$, etc.). For a given value of τ_{true} , we simulated the V-polarized brightness temperatures for dry and wet soil moisture conditions. These brightness temperatures were in turn the basis for estimating soil moisture by inverting the τ - ω model using the wrong τ_{inv} as a function of $\Delta \tau$. For both dry and wet soil moisture conditions, the deviation was an estimate of the retrieval bias: their mean was taken to be an estimate of the offset M , whereas their difference allowed us to estimate L . When plotted against $\Delta \tau$,

M and L are increase nearly linearly and only show a weak dependence on τ_{true} . The slope of this relation is thus well but not perfectly defined. We refer to the slopes as μ^{star} (for M) and λ^{star} (for L), respectively. To account for the spread due to the slight curvature and dependence on τ , we estimated the likely range of values by computing the slopes from the differences in L or M between five equally spaced values of $\Delta \tau$ (between -0.1 and 0.1), repeated for equally many values of τ_{true} (between 0.1 and 0.6). The range of these values was $\lambda^{\text{star}}_{\text{pred}}$ in [2.0, 3.8] and $\mu^{\text{star}}_{\text{pred}}$ in [0.33, 0.65] m³m⁻³. These ranges will later be compared to data-driven estimates, thus providing a first-order assessment of the agreement between predictions and observations, despite the neglect of other retrieval errors.

Further, we state explicitly that no τ value was assumed to produce figure b. We still have kept the same colour and linestyle across the two subfigures, as the caption should make it sufficiently clear that there is no direct link between the lines.

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.

The reason we included this figure in the first place is because the interpretation of L caused no small degree of puzzlement when we presented preliminary results of this study. As we suspect that some readers will skip Sec. 2 and 3, we have included this figure and we also recapitulate the meaning of the parameters in the text. While we have kept the figure, we have amended the caption in the hope that it will facilitate its interpretation. The relevant part reads

Explanation of the bias terms, illustrated for a time-changing sensitivity $L(t)$ and offset $M(t)$. A varying sensitivity changes the response of the SMAP retrieval to a unit change in the true soil moisture. When it is larger than one, the SMAP data have a larger dynamic range than the true soil moisture (illustrated by the slope > 1 in the inset). The time-average value of L is 1, and the temporal standard deviation of L is given by $|\lambda|$ (length of arrow). A variable M induces non-constant offsets, and the magnitude of its temporal variability is given by $|\mu|$. $M > 0$ corresponds to a positive offset (shown in the inset).

F) Looking at Figure 4a, it is not clear visually that L is actually more closely related to $\Delta \tau$ than to τ 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 τ a priori is pretty difficult. If indeed L is a better match to τ 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.

We believe a potential confounding by tau is an important concern, and we have included an additional scenario in Fig. 7 to address it.

This new scenario uses two explanatory variables for L and M, namely Delta tau and tau itself. As we write in the methods: “To account for a potential confounding of \tau itself, which may also have an effect on the bias estimates, we included the smoothed SMOS \tau as second explanatory variable for L and M , referred to as \tau control.” As we subsequently describe in the results, the estimates of the Delta tau lambda and mu change very little for all stations but one. This indicates that the standard inference results are not strongly influenced by confounding from this source. Also, the lambda parameter corresponding to tau is considerably smaller than that of Delta tau: medians of 0.00 and 0.16, respectively (25th/75th percentiles: -0.06/0.02 vs. 0.05/0.33). We hence do not believe that an additional dependence on tau, given delta tau, is a major issue here. However, we have completely revised the discussion section and now talk at length about confounding.

We have also computed estimates using only tau as explanatory variable, as suggested above, but we do not show them in the revised manuscript. The results for the tau parameters are potentially subject to confounding due to Delta tau (see above). For the South Fork site, the impression that there is a stronger relation to tau is borne out by the data to only a limited extent. The lambda parameter estimates turned out to be (10-90% posterior interval):

- standard model, i.e. only delta tau: Delta tau lambda: 0.25-0.36
- only tau: tau lambda: -0.02 - 0.11

Across all network sites, the Delta tau lambda are consistent in the sense that the posterior medians are all positive, whereas for the only tau configuration they are almost equally distributed between positive (4/7) and negative (3/7) values.

G) More on Figure 4: The caption mentions “The magnitude of the dependence for a unit change in delta tau, λ^* is consistent with predictions by tau-omega”. This is a strongish 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?

To clarify these issues, we have made several changes to the text and figure.

We have already described the extended description of the model predictions. There, we outline how we arrive at the range of model predictions displayed in the figure as well as the limitations. The reason for using time-independent model estimates is that these estimates are based on time series, i.e. multiple time instances (with changing $\Delta \tau$) are required to estimate a time-independent parameter like μ/μ_{star} .

We have changed the colour of the word 'model' and amended the caption:

The decent model-estimation match only pertains to λ , i.e. subfigure b), and we have revised the caption to make this crystal clear. About λ , we write that the magnitude are "broadly consistent with predictions by the τ - ω model of Sec. 4.'. Conversely, "the unnormalized quantities μ^{\star} smaller than predicted by the model."

The right-hand column shows the comparison of the un-normalized quantities to the model predictions. To make it easier for the reader to understand this panel, we now show how the un-normalized estimates are computed in a separate equation (10), and we have greatly extended the discussion of the model predictions.

Finally, we briefly discuss the apparent dependence on potential controls (like land cover). We discuss spatial patterns and the relation to land cover at much greater length in the subsection on the sparse sites. In this subsection, we write:

There is no clear apparent dependence of λ^{\star} on location or land cover properties; for instance, Monte Buey and Bell Ville are within < 100 km of one another, and despite the similarity in planted crops the latter's λ^{\star} is considerably larger.

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?

We now discuss our rationale for including the $\Delta \tau$ explanatory variable in the bias model of the re-analysis data.

The inclusion of a $\Delta \tau$ -dependent bias for the reanalysis product is not driven by physical reasoning, but for statistical reasons. By controlling for the same explanatory variables for both products, the impact of potential confounders - e.g. a seasonal bias that is correlated with $\Delta \tau$ - on the bias estimates of the remotely sensed product can be reduced. If this were not done, the model would try to partially adjust the time-variable bias term of the remote sensing product to minimize the systematic differences to the re-analysis product, thus distorting these bias estimates.

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.

We share the reservations with respect to the tau products. To better address them, we have made a number of changes. First, we discuss the SMAP DC results obtained over the network sites in more detail. In particular, we mention some of the issues associated with either product. Second, we now also show the SMAP DC results over the contiguous US, i.e. over the sparse sites (Fig. ?). As with the network sites, the results are very similar. Third, in response to Wade Crow's remarks, we have included a separate discussion section where we discuss errors in the tau products and their impact on interpreting the results in a descriptive and a causal framework.

Note that we continue to use the SMOS L3 product, as we hope that we are able to paint a more complete picture by showing the results obtained with two different products. Unfortunately, the other products mentioned are currently not publically available. We hope that the techniques developed in the manuscript will in the future contribute to elucidating the error structure of novel products such as the MT-DCA soil moisture.

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?

This is a good point. We have included a separate discussion section, where we dwell on the implications for error characterization more generally.

Geophysical products in general are potentially also subject to time-variable errors, so that the presented approach could be applied to variables such as wind speed, land surface fluxes and leaf area index. The issue of non-constant error sources, be they associated with environmental conditions or varying observational parameters, likely pertains to many such variables. Extensions of our approach could in the future shed light on the error properties of a wide range of products, thus contributing to the development of improved retrieval approaches.

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

We do so by comparing the normalized results with absolute (unnormalized ones: the quantities with an asterisk). We detail the associated changes to this point in our reply to point G).

Minor Comments:

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

We have added a reference to this paper.

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

At the beginning of the paragraph, we now write 'We hypothesize that seasonal changes in the error structure arise due to an inaccurate vegetation correction in the retrieval, so that the biases relative to the in-situ data track the misspecification in the vegetation optical depth $\Delta \tau$.' This is not a precise definition, but it should suffice for the introduction.

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.

We have amended the caption accordingly. We have also changed the ticks and labels as suggested. Note that we have slightly redesigned the figure in line with other suggestions.

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?

Good idea, we have added a new figure (supplement).

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

Done.

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

We now state explicitly the dynamic range on which these calculations were based: "The sensitivity coefficients λ are retrieved with comparable precision: the RMSE of 0.05 corresponds to a differential bias between dry and wet conditions of around 0.01 m³ m⁻³ (assuming a soil moisture dynamic range sim 0.25 m³m⁻³)"

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.

We believe this comment is to do with our bad wording in that we referred to the posterior standard deviation as the posterior (uncertainty), see points B) and C).

While we agree in principle that showing the entire posterior distribution is a good idea, we believe the well-behaved unimodal distributions, which we have exclusively encountered in our analyses, warrant the restriction to location and dispersion parameters to summarize those posterior distributions.

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.

done

T) Page 11, line 5: note that this reference is broken
We have added the year

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
We are grateful for this remark, as we were not aware of the paper.

V) I don't think the subscript p is ever defined. Is this an index for the number of explanatory variables?
We now also define it explicitly ("sensitivity to the p th explanatory variable")

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

We thank Wade Crow for his insightful review. We have made numerous modifications to the manuscript: apart from two new figures, we discuss several key results in considerably more detail.

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?

The impact of errors in the tau product on the estimated bias parameters is an important concern that we now address in more detail throughout the manuscript. The main changes are extended discussions and a greater focus on an alternative tau product (sensitivity analysis based on the SMAP tau product) through extended discussions and a new figure. We are aware that we cannot resolve the issues mentioned, a fact we acknowledge openly, but we hope the extended discussions will provide a balanced picture.

There is now a separate discussion section that deals with the interpretation of the inferred biases. There, we posit that the role of errors when analysing the results is contingent on what general view one adopts. One of these views is purely descriptive, the other tries to establish a causal link. The purely descriptive view is easier to uphold because it is only concerned with associations rather than the mechanisms of these associations. As associations can be misleading if interpreted causally (errors in the tau product, confounding, etc.), we have clearly stressed the largely

descriptive nature of our analyses by employing phrases such as "errors associated with the vegetation correction" rather than induced or even caused.

It is the causal view that is more directly affected by errors in the input tau product. While we focus on a largely descriptive view, we do engage in analyses towards establishing a causal link, chiefly via the comparison to tau omega predictions. These comparisons rely on the assumptions of no confounding and no errors in the input tau product. We now mention these assumptions explicitly (see below), and we discuss three important points in this context

- the definition of the errors: in the context of soil moisture retrieval, we believe that it is mainly a model-internal effective parameter (that can partially account for e.g. changes in effective roughness or for an inappropriate choice of the effective scattering albedo). It is this effective parameter that should serve as reference in the computation of delta tau, rather than a purely vegetation-based proxy.
- the nature of the errors, which have both systematic and random components
- possible confounders

The distinction between a hypothetical true tau and an effective tau is, we believe, an important one to make, both for interpreting the estimated biases but also with the view of diagnosing of vegetation-water interactions that forms part of our motivation for studying time-varying biases. For single-channel retrievals, such a value typically exists: for a given soil moisture value (and forward model, single scattering albedo, etc.) it is the value that aligns the error-free brightness temperature with the true soil moisture. For dual-channel or multi-angular algorithms, such a value may not exist, in which case a retrieval of both tau and soil moisture would yield a wrong soil moisture estimate. However, it may be a good approximation, as hinted at by Parrens et al. 2017. They found that a joint retrieval of a single vegetation & roughness parameter, i.e. one effective tau parameter, yielded good soil moisture estimation results; this should work even better for a constant incidence angle (because in that case, the value of N_r in their model is immaterial).

However, even if such an effective tau did exist, it would be dependent on the algorithm, incidence angle, etc. We stress this view in the discussion. We also highlight the limitations of this view in Sec. 5.2, where we discuss the diagnosing of vegetation-water interactions for which one tends to consider tau to be a vegetation proxy.

Owing to the complexities, we do not hazard a guess as to what influence deviations in the SMOS or SMAP DC tau from either a "true" or an "effective" tau have on the estimated biases. Based on analogies to regression modelling, we would expect causally biased estimates in the presence of random or systematic errors. We state this openly in the completely revised discussions. The two key paragraphs in the new discussion section read:

A mechanistic interpretation of the inferred biases is complicated by a number of poorly understood factors. First, the time-variable biases are relative to the in-situ data. The results over the sparse sites should hence be interpreted with caution due to

representativeness error, even if they are similar to those at the dense high-quality network sites. Even at the network sites, residual time-dependent biases of the in-situ data cannot be ruled out completely. Another major uncertainty are errors in the satellite-derived contemporaneous τ products, which are not accounted for in the estimation. One important reason for why these errors are difficult to quantify is that in the context of soil moisture retrieval τ can be considered as essentially a model-internal effective quantity (Parrens17). As such, an observation-based estimate of τ reflects not only the vegetation conditions but also inaccuracies of the τ - ω model itself, the way it is parameterized and other environmental conditions. An instance for the latter are roughness changes associated with harvest in croplands (Patton13) which likely contribute to the autumnal increase in SMOS τ in Fig. 6a. To a good degree of approximation, roughness changes will be captured by the effective τ that the SMOS or SMAP DC algorithms retrieve from the brightness temperatures (Parrens17). Nevertheless, the estimates used in this study will still be affected by systematic and random errors with respect to this effective quantity. Systematic differences between the effective τ for the SMAP retrievals and that of the SMOS satellite are, for instance, due to different incidence angles and model parameters. The impact of such errors on the estimated biases is unknown, but analogies to simple regression models suggest that they can distort these estimates in either direction.

While it is premature to attribute the inferred biases completely to an imperfect vegetation correction, there are two lines of reasoning that suggest that the inferred biases are not spurious. First, they are fairly consistent across croplands, and also between sites with sparse and dense in-situ networks. Also, they tend to be large both in absolute terms (e.g. $\lambda > 0.1$) and compared to the posterior uncertainties. Further, they are also robust to the specification of the input τ product (SMAP DC instead of SMOS τ) and to several model modifications (Sec.5). However, these results are purely descriptive in that they only quantify associations, rather than establishing a causal link. A first step towards such a mechanistic interpretation is the comparison of the time-variable biases with predictions by the τ - ω model. This second line of reasoning suggests that the magnitude of the multiplicative biases λ^* is largely consistent with theoretical expectations (Fig. 4a). However, this analysis is contingent on i) the τ - ω model being appropriate and correctly specified (e.g. known ω), ii) there being no confounding biases such as seasonal inundation, and iii) the sufficient accuracy of the input τ product. It is difficult to dispel these concerns, and indeed the deviations from the predictions (for μ^*) indicate that unconsidered phenomena also contribute to the time-varying biases in addition to those resulting from the vegetation correction.

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?).

We now discuss some of the estimated parameters pertaining to the other products. First, we have added a new figure (Fig. 8) that shows the estimated noise level for all three products, or more precisely a normalized version that facilitates inter-product comparisons. In particular, the discussion addresses the spatial representativeness issues raised at several points in the referee report:

To analyse the estimated noise level for all three products, we computed a normalized version σ/l , where the division by l accounts for the different dynamic ranges of the three products by scaling the noise level with respect to the in-situ data (Fig. 8). SMAP achieves a median value of 0.045 m^3m^{-3} , a higher value than that of the in-situ data or MERRA-2 (0.029 and 0.040 m^3m^{-3} , respectively). For all three products, the corresponding values over the network sites are smaller by around 50%. For the in-situ data, the larger noise level at the sparse sites is not surprising, owing to their limited representativeness. However, direct comparisons could be misleading. For instance, the larger noise level estimates (and greater spread of these estimates) may be partially accounted for by the small number of available networks and by the heterogeneous land cover and vegetation conditions across the sparse sites in the contiguous US.

We also discuss MERRA μ/λ parameters in a bit more detail, both in terms of the rationale and the results. With respect to the rationale, we have added that 'the inclusion of a delta tau dependent bias for the reanalysis product is not driven by physical reasoning, as the MODIS NDVI climatology that gives rise to a non-zero delta tau plays no role in the generation of the MERRA reanalysis product. However, there is a compelling statistical reason to include the same explanatory variables as in the remotely sensed product. By controlling for the same explanatory variables, the impact of potential confounders - e.g. a seasonal bias that is correlated with Delta tau - on the bias estimates of the remotely sensed product can be reduced. If this were not done, the model would try to partially adjust the time-variable bias term of the remote sensing product to minimize the systematic differences to the re-analysis product, thus distorting these bias estimates.' With respect to the results, we mention in the results that 'Reanalysis bias parameters were estimated as well, but they are considerably smaller in magnitude than those of the SMAP product.' More precisely, the μ parameter of the MERRA product is 0.000 on average (compared to 0.007 for the SMAP product), whereas for λ they are 0.04 and 0.18 respectively. Further, their direction is highly heterogeneous, whereas the SMAP bias parameters are all of the same sign.

Finally, we do not share the view that the notation in Section 2.1.1. suggests that the error models are identical for all products. To include the possibility of different error models, we indexed the explanatory variables and the number of explanatory variables by the product. To better highlight this dependence, we now write "The explanatory variables can depend on the product n as well as on the parameter (μ , λ). We also highlight that we use a reference product that is assumed unbiased.

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.

We share those concerns, and we devised the simulation study to address some of them: in the simulations, the simulated soil moisture is auto-correlated, whereas the standard inference model implementation prescribes independent soil moisture values. However, we do agree that we did not discuss these aspects in sufficient detail. We now discuss the issue of time scales in the simulation section 3:

The other crucial assumption in the model is the probability distribution for the soil moisture. Also here, the changes are typically small (up to 10% improvement in the RMSE, but a decrease in bias) when replacing the standard time-invariant model by a seasonally variable model. The improvement suggests that the model-internal soil moisture distribution can have an impact on the estimated bias parameters, in particular when the actual soil moisture is correlated with the explanatory variable, as it was in the simulated data. We would hence expect that for most applications it is the seasonal and sub-seasonal time scales that the soil moisture model should be able to capture. For comparison, autocorrelation on the inter-storm time scale that is not captured by our model but present in the simulated data did not seem to have a major impact (sufficient fidelity for the full model, Fig 2)

To expand on our discussion in the manuscript, we believe it is important to distinguish different time scales. The temporal structure of soil moisture time series is of course complex. To simplify it, we isolated two important time scales: long (i.e. seasonal) time scales, and short time scales

As outlined in our discussion, we believe that representing the seasonal time scales, on which also the biases vary in our application, is more important for accurate bias parameter inference than correctly representing the very short time scales. We do want to explore the shorter time scales in the future, though. It is definitely possible to represent such time scales in the model itself, but it requires a clever implementation (initially, we had tried to implement an AR-1 model for soil moisture, but it was impractical because the MCMC sampling was very slow, which is usually thought to indicate an issue with the way the model is set up. Internally, there are different ways to parameterize the same model, and they are not equivalent in terms of MCMC sampling efficiency).

4a) Section 2.2 - I understand that the Bayesian inference 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?).

To paint a clearer picture of the Bayesian approach, we have extended the description of the MCMC sampling.

... Here, we rely on Hamiltonian Monte Carlo as implemented using the adaptive No-U-Turn Sampler in pymc3. The No-U-Turn Sampler produces successive, dependent samples of the posterior distribution that are called a chain. Each sample consists of draws from the posterior distribution, or actually an approximation thereof, of all the unobserved random variables (Output in Fig1b). They comprise the parameter random variables (e.g. the time-dependent biases) as well the soil moisture time series, i.e. one value of θ for each SMAP observation. For each location, we sample two independent chains with 2000 samples each, which standard quality controls (divergences, chain mixing) indicate is sufficient. Following common practice, the first 1000 samples are discarded

4b) 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.

We agree that there are limitations to the analogy. Similar issues commonly arise when comparing classical inference approaches with Bayesian approaches. Classical approaches, including method-of-moments-type estimators that classical triple collocation can be thought an instance of, are plagued by problems of identifiability. In order for them to be applied successfully, the data must provide sufficient information to estimate all parameters at the same time, loosely speaking. In case of classical triple collocation, these are $N=3$ three error variances and often also $N-1=two$ sets of bias parameters (often called additive and multiplicative bias), and they can be uniquely identified when $N = 3$ (but not when $N = 2$).

Conversely, for Bayesian approaches this issue does not arise in this form, owing to the prior information. A proper prior distribution (which is what we adopt in our approach) ensures valid posterior distributions even in the extreme case that no data are available, in which prior = posterior. A relevant publication in this context is Bayarri and Berger, *The Interplay of Bayesian and Frequentist Analysis*, Statistical Science, 2004.

In such data-poor situations the specific choice of prior plays a crucial role. It is to minimize the importance of the prior that we assumed 3 products, in analogy to triple collocation, as 3 products provide enough information even without any priors (classical case). For the scenario with only two products, we prescribed a much stronger prior on the in-situ data in an attempt to make up for the reduced information content.

We do realize that these qualitative arguments can only partially address the valid concern raised by the reviewer. However, we do not have theoretical results to bolster these views. The simulations studies indicate that $N = 3$ products is sufficient to estimate the parameters of interest (more precisely: to substantially narrow the posterior distribution compared to the prior). With respect to the appropriate choice of products (number, type, etc.), our manuscript leaves a lot of questions open.

To better address these concerns, we have provided an abridged summary of our rationale in Sec. 2:

We focus on a setting inspired by triple collocation studies, i.e. we for the most part assume that $N = 3$ independent and noisy products are available Gruber 16. In regular triple collocation, three independent products provide sufficient information to estimate the random errors of all three products and bias parameters of two of the three products. In a Bayesian setting, the presence of prior information allows one to reduce the number of independent products, but the results will strongly depend on the prior distributions.

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_0 ” 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.

We have made two changes. First, we now list the three input products in the very first sentence of this section, and then describe them in more detail. Second, we now explicitly state that the in-situ data constitute the reference product y_0 (before, we had written that product $n = 0$ is the reference product).

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.

We have greatly extended the description and discussion of this aspect. The methods are now described in much greater detail. For M and L , we write

To compute the predicted biases in Fig. 3a), we assumed the τ - ω model applied and was correctly specified (temperature, dielectric mixing model [Dobson; silt loam], single-scattering albedo $\omega = 0.05$, etc.). For a given value of τ_{true} , we simulated the V-polarized brightness temperatures for dry and wet soil moisture conditions. These brightness temperatures were in turn the basis for estimating soil moisture by inverting the τ - ω model using the wrong τ_{inv} as a function of $\Delta \tau$. For both dry and wet soil moisture conditions, the deviation was an estimate of the retrieval bias: their mean was taken to be an estimate of the offset M , whereas their difference allowed us to estimate L . When plotted against $\Delta \tau$, M and L increase nearly linearly and only show a weak dependence on τ_{true} . The slope of this relation is thus well but not perfectly defined. We refer to the slopes as μ^{\star} (for M) and λ^{\star} (for L), respectively.

To account for the spread due to the slight curvature and dependence on τ , we estimated the likely range of values by computing the slopes from the differences in L or M between five equally spaced values of $\Delta \tau$ (between -0.1 and 0.1), repeated for equally many values of τ_{true} (between 0.1 and 0.6). The range of these values was $\lambda^{\text{star}}_{\text{pred}}$ in [2.0, 3.8] and $\mu^{\text{star}}_{\text{pred}}$ in [0.33, 0.65] m³m⁻³. These ranges will later be compared to data-driven estimates, thus providing a first-order assessment of the agreement between predictions and observations, despite the neglect of other retrieval errors.

We have also amended Fig. 3 (showing the star parameters explicitly). In the results, we have extended the description of the model-estimate comparison:

When converted into absolute quantities (λ^{star}), the inferred dependence of L on $\Delta \tau$ matches the model predictions reasonably well (Fig. [\ref{fig:networksites}](#)b). In other words, the data-derived, completely independent estimate is broadly consistent with the predicted impact of a τ misspecification in the retrieval, despite limitations in the estimates (e.g. issues with the reference τ) and the model predictions (e.g. assumed knowledge of the land surface temperature) of λ^{star} . There is no clear apparent dependence of λ^{star} on location or land cover properties; for instance, Monte Buey and Bell Ville are within < 100 km of one another, and despite the similarity in planted crops the latter's λ^{star} is considerably larger.

We now also revisit this issue in the discussions; the associated changes to the appropriate section are described in our reply to point 1.

7) Section 4.2.2 - The authors provide a nice sensitivity analysis which describes the impact of using a different τ 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 τ references is reassuring. However, I don't quite follow why the large changes observations when using a contemporary MODIS τ indicates a lack of sensitivity to the use of MODIS τ climatology in the SMAP L3 retrievals. The $\Delta \tau$ generated by the MODIS contemporary minus climatology differences leads to significantly non-zero λ and μ estimates - just not the same estimates as the application of "delta τ " results generated relative to SMOS τ . How exactly does this support the conclusion that inter-annual τ 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.

We have made two changes. First, we have added a new figure that also shows the results of the sparse sites obtained with the SMAP-based $\Delta \tau$. This figure also features in the second change, namely the extended descriptions. We also point out under what assumptions the smaller estimates obtained with the contemporaneous MODIS τ can be interpreted to indicate that the biases are not only due to outdated NDVI-derived data in the retrieval. In the results section, we write:

Our sensitivity analyses focus on the reference τ product. When the SMAP dual channel result is used as the reference τ product, the bias parameters change little for the vast majority of sites (Fig. 6). When the posterior uncertainties are taken into account, the λ and μ values tend to overlap with those obtained using the SMOS τ product, indicating that the results are not sensitive to the choice of microwave-derived reference τ product. Also the spatial patterns across the sparse study sites are very similar (Fig. 8).

We have also extended the analysis of the MODIS-derived Delta τ .

By contrast, the estimates can change substantially when τ is derived from contemporaneous NDVI data, and predominantly they are smaller in magnitude. If the problem with the use of the NDVI climatology in the retrieval were the use of a climatology alone, we would expect similar estimates. Conversely, we would expect the estimates to be smaller if it was the link between NDVI and τ that led to an inaccurate vegetation correction. The smaller estimates that were actually observed may thus indicate that the use of a climatology is not a dominant error source in the SMAP vegetation input data.

As outlined in our reply to point 1), there is now a separate discussion section that deals with errors in τ in the context of interpreting the results.

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).

We have extended the discussion of these results. Our main point in the discussions is not that there is a clear-cut link between the estimated biases and the R^2 values, but rather that the spatial patterns suggest that there may be one that deserves attention in future studies.

In the results:

While the spatial patterns largely match those of the time-variable biases, the link between them is not clear and not necessarily uniform across all sites. The computation of anomalies largely removes seasonal offsets, which constitute a major fraction of the estimated additive biases. However, it does not remove higher-frequency variations or inter-annual differences, although the record is too short to reliably study those. Neither can it account for the changes in sensitivity, which are particularly large over croplands. Finally, the in-situ soil moisture anomalies, predominantly derived from single probes, are subject to major uncertainties. All these factors likely contribute to the elevated associations between the τ and the SMAP soil moisture anomalies (Delta R^2), but the

precise impact of time-variable biases on our ability to diagnose such interactions remains an open question.

In the discussion:

The spurious vegetation signal in the soil moisture data may distort estimates of water-vegetation coupling. We find inflated values of R^2 between the SMOS vegetation optical depth and SMAP soil moisture, whereas purely random noise would decrease the R^2 (Fig. 7c). While the spatial patterns largely match those of the estimated biases, this does not imply a causal link between the two. However, the inflated R^2 values hint at potential pitfalls in using remotely sensed soil moisture to study global hydrology.

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.

We agree with the limitations of the sparse sites. This is also why we discuss the network sites in considerably more detail. Our motivation for including Figure 7c) is the similarity of the spatial patterns, which do suggest a connection. However, we make clear that the specifics of this link are currently unknown. In the future, we hope that our work on biases will inform the interpretation of correlation coefficients, regression models and similar statistics.

We hope the extended discussions, see in particular our reply to the previous point, clarify this stance.

Estimating time-dependent vegetation biases in the SMAP soil moisture product

Simon Zwieback^{1,2}, Andreas Colliander³, Michael H. Cosh⁴, José Martínez-Fernández⁵, Heather McNairn⁶, Patrick J. Starks⁷, Marc Thibeault⁸, and Aaron Berg¹

¹Department of Geography, University of Guelph, Guelph, Ontario, Canada

²Department of Environmental Engineering, ETH Zurich, Zurich, Switzerland

³NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA

⁴USDA-ARS Hydrology and Remote Sensing Laboratory, Beltsville, Maryland, USA

⁵Instituto Hispano Luso de Investigaciones Agrarias, Universidad de Salamanca, Salamanca, Spain

⁶Science and Technology Branch, Agriculture and Agri-Food Canada, Ottawa, Ontario, Canada

⁷USDA-ARS Grazinglands Research Laboratory, El Reno, Oklahoma, USA

⁸Comisión Nacional de Actividades Espaciales, Buenos Aires, Argentina

Correspondence to: Simon Zwieback (zwieback@uoguelph.ca)

Abstract. Remotely sensed soil moisture products are influenced by vegetation and how it is accounted for in the retrieval, which is a potential source of time-variable biases. To estimate such complex, time-variable error structures from noisy data, we introduce a Bayesian extension to triple collocation in which the systematic errors and noise terms are not constant but vary with explanatory variables. We apply the technique to the SMAP soil moisture product over croplands, hypothesizing that errors in the vegetation correction during the retrieval leave a characteristic fingerprint in the soil moisture time series. We find that time-variable offsets and sensitivities are commonly associated with an imperfect vegetation correction. Especially the changes in sensitivity can be large, with seasonal variations of up to 40%. Variations of this size impede the seasonal comparison of soil moisture dynamics and the detection of extreme events. Also, estimates of vegetation-hydrology coupling can be distorted, as the SMAP soil moisture has larger R^2 values with a biomass proxy than the in-situ data, whereas noise alone would induce the opposite effect. We conclude that complex biases can be present in soil moisture products and that they should be accounted for in observational and modelling studies.

1 Introduction

Soil moisture products derived from satellite measurements are subject to errors. These are not constant, but vary in space and time. For any given location, they may depend on variable factors such as vegetation phenology, atmospheric conditions or measurement characteristics like the incidence angle (Loew and Schlenz, 2011; Entekhabi et al., 2010; Su et al., 2016). Vegetation is commonly considered to be the most delicate factor to control when retrieving soil moisture from the raw satellite measurements (Dorigo et al., 2017; Chan et al., 2016). An imperfect vegetation correction during the retrieval will induce time-variable errors in the soil moisture product (Konings et al., 2017). As such structural errors potentially distort estimates

of vegetation-soil moisture coupling in unexpected ways (cf. Doherty and Welter, 2010), they are more pernicious than simple quasi-random noise or than time-constant biases.

However, the time-variable error properties of soil moisture products are poorly understood and rarely considered in practice (Loew and Schlenz, 2011). Arguably one reason for this is that there is currently no consistent way of estimating such rich error structures from data. If an exact reference soil moisture measurement were available, this would be an easy enterprise (Su and Ryu, 2015). Dense in-situ measurements are widely considered to be the closest one can get to perfect reference data, but they are rare, and non-negligible uncertainties remain (Colliander et al., 2017). In absence of perfect reference data, any useful error estimation procedure must cope with errors in all input data. Triple collocation and its various extensions can provide consistent error estimates in these circumstances (Gruber et al., 2016; Zwieback et al., 2016). However, similar to the standard RMSE metric they cannot directly separate non-constant systematic errors from quasi-random measurement noise. Conversely, common pre-processing steps (forming anomalies, analysis of short time periods) provide a simple means to deal with certain non-constant errors but lack generality (Loew and Schlenz, 2011; Gruber et al., 2016).

Here, we extend triple collocation to estimate non-constant error structures (Sec. 2). The central idea is to express systematic errors such as an offset and random errors in terms of explanatory variables like a vegetation index (cf. Xu et al., 2017; Doherty and Welter, 2010). The choice of explanatory variables should be informed by the measurement principle and the retrieval algorithm. In our case study of the SMAP product, we will specify it in terms of a misspecification of the vegetation [..] correction during the retrieval process. We generally assume that three independent and noisy products are available. Once their hypothesized error structure has been specified, our Bayesian Triple Collocation approach proceeds in two steps. First, the specified error structure is embedded in a probabilistic model that links the unknown soil moisture θ with observed soil moisture products y_n . Second, Bayesian inference is applied, and one thus obtains estimates and uncertainties of the error parameters of interest. A simulation study indicates that time-variable systematic errors can be estimated reliably for as few as 250 samples (Sec. 3).

We apply this procedure to estimate time-variable biases in the SMAP soil moisture product that are associated with an imperfect vegetation correction (Sec. 4). This is not to say that other soil moisture products are not subject to similar biases; on the contrary, the SMAP baseline product is widely considered to provide the most reliable global soil moisture data record available (Chan et al., 2017; Colliander et al., 2017). To estimate soil moisture from the passive microwave measurements, the retrieval algorithm has to correct for the vegetation influence (Kurum et al., 2011; Chan et al., 2017). Specifically, the SMAP algorithm assumes that the vegetation optical depth τ , a dimensionless measure of how much the microwaves interact with the vegetation, is known a priori.

We focus on croplands, as they present a particular challenge to the vegetation correction approach using an a priori τ . Over crops, the input τ , which is derived from the Normalized Difference Vegetation Index (NDVI), only provides an incomplete picture of the vegetation influence at microwave frequencies. The NDVI, being strongly influenced by leaf chemistry, can only indirectly account for the dominant controls on τ , namely the vegetation water content and the canopy structure, both of which are particularly diverse and dynamic in crops [..] (Lawrence et al., 2014; Momen et al., 2017). The NDVI-based input τ also does not account for inter-annual variability in vegetation conditions, which cannot be neglected over croplands (Patton and

Hornbuckle, 2013). Consequently, agricultural regions have been identified as a weak spot for the SMAP product (Colliander et al., 2017). However, the time-average metrics analysed so far cannot distinguish seasonal biases from quasi-random noise, and the magnitude of vegetation-induced systematic errors thus remains unknown.

We hypothesize that seasonal changes in the error structure arise due to an inaccurate vegetation correction in the retrieval, so that the biases relative to the in-situ data track the misspecification in the vegetation optical depth $\Delta\tau$. We specify the SMAP offset and sensitivity as a function of $\Delta\tau$, based on predictions by the τ - ω radiative transfer model (Kurum et al., 2011). We estimate the associated error parameters with the Bayesian triple collocation approach, using in-situ and re-analysis data as additional soil moisture products. We find that SMAP soil moisture biases that track $\Delta\tau$ are widespread and large over croplands. This is especially so for the sensitivity, resulting in a time-dependent dynamic range of the SMAP soil moisture product that impedes seasonal comparisons of soil moisture dynamics. We attribute the time-variable biases to the imperfect vegetation correction, as the inferred bias characteristics largely match those predicted by the τ - ω model. To illustrate the potential influence of these biases on estimates of vegetation-hydrology coupling (Adegoke and Carleton, 2002), we show that the coefficient of determination between SMOS τ anomalies and SMAP soil moisture anomalies is inflated compared to in-situ soil moisture measurements. In summary, our analyses suggest that soil moisture products can be subject to previously neglected time-variable biases that should be accounted for in observational and modelling studies.

2 Bayesian triple collocation

We now present a general overview of the approach (Fig. 1), which consists of two components. First, a probabilistic model that links the unknown soil moisture θ with the observed soil moisture products y_n . The link itself characterizes the error structure: it depends on error parameters γ_{y_n} and explanatory variables w . Second, a Bayesian inference approach that provides estimates of all the unknown quantities, in particular the error parameters. By conditioning on the observed soil moisture data (input), we get a posterior distribution over the unobservable quantities (output). We focus on a setting inspired by triple collocation studies, i.e. we for the most part assume that $N = 3$ independent and noisy products are available (Gruber et al., 2016). In regular triple collocation, three independent products provide sufficient information to estimate the σ of all three products and bias parameters of two of the three products. In a Bayesian setting, the presence of prior information allows one to reduce the number of independent products, but the results will strongly depend on the prior distributions.

Our approach has several characteristics that make it useful in a wide range of applications. It is widely applicable, as no soil moisture product is assumed to be free of errors. This is particularly critical for estimating the noise magnitude and the sensitivity, which cannot be estimated consistently by standard regression approaches when the reference product is subject to errors (Yilmaz and Crow, 2013). Also, it provides principled uncertainty estimates through the posterior distribution. It is flexible, as it can be adapted to many functional relations and error structures. Finally, it is transparent because all modelling assumptions are explicit. Owing to its flexibility, the model can be modified to test the sensitivity of the results to certain assumptions. We will exploit these advantages by modifying the prior distribution and the likelihood; for instance, we will test several models for the unknown soil moisture θ .

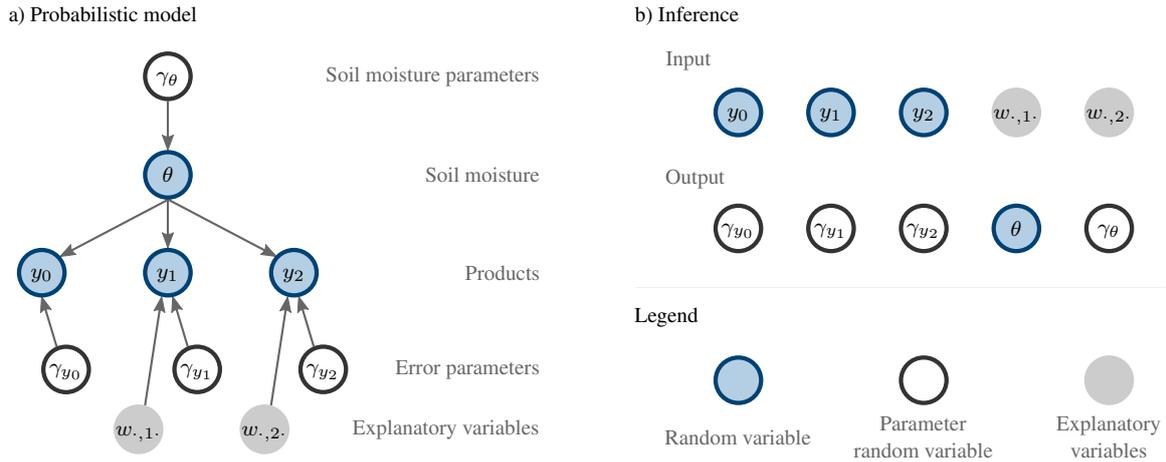


Figure 1. The two components of Bayesian triple collocation. a) The probabilistic model expresses the observable products in terms of the unknown soil moisture, error parameters and explanatory variables. The set of error parameters of the reference product, γ_{y_0} , does not contain variable bias terms. b) The soil moisture products and explanatory variables serve as input to the inference, which produces estimates (posterior probability distributions) of, amongst other things, the error parameters.

2.1 Probabilistic model

The probability distribution comprises the observable products y_0 to y_{N-1} , the unknown soil moisture θ and numerous parameters: the set of parameters γ_{y_n} characterizes product n , and γ_θ does the same for the soil moisture. The structure of the model is summarized in Fig. 1a, which depicts the corresponding directed acyclic graph (MacKay, 2003). This graph expresses how the distribution over all the random variables is factorized into smaller components. Starting with an observable product y_n , its distribution is modelled conditional on the unknown θ , the associated parameters γ_{y_n} (including the time-independent and time-dependent bias parameters) and the explanatory variables $w_{.,n}$. We refer to this conditional distribution as the error model. It is conditional on the soil moisture θ , whose distribution (conditional on parameters γ_θ) is referred to as the soil moisture model. The final component are the parameters γ_{y_n} and γ_θ , which are assigned prior distributions. These prior distributions are integral to the Bayesian approach, and they express the initial belief about the parameters' likely values. We now describe each of these components in turn. To facilitate future applications of the approach, we do this using very general notation, which we will later in the SMAP case study simplify by dropping subscripts.

2.1.1 Error model

Each product's error model is governed by a set of parameters γ_{y_n} which quantify the error component such as the biases (systematic errors). We consider an affine error model according to which the dynamic range or sensitivity of the product can differ from that of the true soil moisture, governed by the scaling parameter $L_n(t)$. We also include an additive offset or bias

$M_n(t)$ (Zwieback et al., 2012; Yilmaz and Crow, 2013), yielding

$$y_n(t) = L_n(t) (\theta(t) - \theta_0) + (\theta_0 + M_n(t)) + \epsilon_n(t), \quad (1)$$

where we essentially relate the deviation of soil moisture from a typical, prescribed soil moisture θ_0 to the observed product.

Our key extension compared to previous triple collocation studies is to make the bias terms and the noise magnitude vary with time-dependent explanatory variables, $w_{\cdot,np}(t)$:

$$L_n(t) = l_n + \sum_{p=1}^{P_{\lambda,n}} \lambda_{np} w_{\lambda,np}(t) \quad (2)$$

$$M_n(t) = m_n + \sum_{p=1}^{P_{\mu,n}} \mu_{np} w_{\mu,np}(t) \quad (3)$$

where l_n and m_n are time-invariant values specific to product n , and the parameters λ_{np} and μ_{np} are coefficients that quantify the [..]dependence on the p th explanatory variable $w_{\lambda,np}$ and $w_{\mu,np}$, respectively. [..]The explanatory variables can depend on the product n as well as on the parameter (μ , λ). They are external quantities that are part of the model input. To facilitate the interpretation of the parameters, we always assume the explanatory variables to have zero mean and unit standard deviation (Gelman et al., 2008). l_n and m_n thus represent the time-average biases. The magnitude of λ_n and μ_n represents the magnitude of the associated temporal changes in L_n and M_n , respectively.

The quasi-random errors are characterized by their variance and further distributional assumptions. For the variance $S_n^2(t) = E(\epsilon_n(t)^2)$ we suggest a multiplicative model that is commonly employed in regression studies (Harvey, 1976)

$$S_n^2(t) = \sigma_n^2(t) \prod_{p=1}^{P_{\sigma,n}} w_{\sigma,np}(t)^{\kappa_{np}} \quad (4)$$

where κ governs the sensitivity of the error variance to the explanatory variable, which has to be positive. We always assume it to be normalized so its geometric mean is one, as this simplifies the interpretation of the typical variance $\sigma_n^2(t)$. A positive value $\kappa > 0$ indicates a larger variance as the explanatory variable increases, and a negative value corresponds to a smaller variance. To complete the specification of the errors, a probability distribution has to be assumed. We focus on a normal distribution with zero mean and variance given by S_n^2 , as our robustness checks indicated that the results were not very sensitive to this assumption (cf. Sec. 3). Finally, two further properties have to be specified. First, we assume that the errors at different times are independent. We will later show that violations of this assumption commonly have negligible impact on the estimated parameter values. Second, we postulate that the errors of the various products are independent. This is generally a key assumption in triple collocation-type studies (Gruber et al., 2016).

We generally specify y_0 to be the reference product (Yilmaz and Crow, 2013; Gruber et al., 2016). Its error magnitude is assumed to be constant over time, and so are its additive bias ($M = 0 \text{ m}^3 \text{ m}^{-3}$) and its sensitivity $L = 1$. This constraint ensures that the unknown soil moisture distribution (its mean and scale) can be inferred from data, as it essentially specified what the reference product was a noisy estimate of (Zwieback et al., 2016). The bias terms are thus always relative to this reference product.

2.1.2 Soil moisture model

The second piece of the probabilistic model concerns the soil moisture θ , the distribution of which also has to be specified. Our default representation is a simple logistic model. The physical quantity θ , which is constrained to between 0 and the porosity ϕ , is expressed as a function of the non-dimensional unbounded soil moisture Θ

$$5 \quad \theta(t) = \phi \frac{1}{1 + \exp(-A - B\Theta(t))} \text{ with } \Theta \sim \mathcal{N}(0, 1). \quad (5)$$

Θ is modelled as a standard normal random variable. The site-specific parameters A , B and ϕ are inferred in the Bayesian inference. We summarized the parameters of the soil moisture model under γ_θ .

One drawback of this model is that it cannot account for the autocorrelation and seasonality of soil moisture. To test for the importance of temporal characteristics, we also generalize the model by making A and B time dependent. We do this by
10 expanding A and B in a spline basis with 12 monthly basis functions. While this model cannot completely account for the complex temporal characteristics of soil moisture, it captures the seasonal trends.

2.1.3 Prior distributions

To complete the full probability distribution, one has to specify the prior distributions of the parameters γ_θ and γ_{y_n} . As we work with normalized explanatory variables, we can use a problem-independent prior distribution (Gelman et al., 2008). We
15 choose the priors to be weakly informative, thus partially ruling out unreasonable values but still letting the data speak for themselves (Gelman et al., 2008). Our default choices are summarized in Tab. 1.

For all products y_n , we put a very weak prior on the error magnitude variance σ^2 . It is given by an exponential distribution with mean $0.1 \text{ (m}^3 \text{ m}^{-3})^2$, plotted in Fig. S1. In other words, we barely constrain this quantity. For the product error parameters, m and $\mu \text{ [m}^3 \text{ m}^{-3}]$ are assumed distributed according to a t distribution $T(0, 0.3^2; 4)$; Fig. S1. It is centred at 0 with a
20 standard deviation of 0.3 and very heavy tails due to its four degrees of freedom. These values barely constrain the estimation of the additive bias m , as they do not rule out biases as large as $0.5 \text{ m}^3 \text{ m}^{-3}$. Similarly for l and $\lambda [-]$, whose prior is $T(1, 0.3^2; 4)$.

The standard prior distribution for the soil moisture parameters γ_θ follows a similar logic. The porosity $\text{[m}^3 \text{ m}^{-3}]$ is given by $T(0.4, 0.1^2; 4)$. The parameters A and B in the standard logistic model are assigned $T(0, 3.0^2; 4)$ and an exponential distribution with mean 3.0, respectively.

25 2.2 Bayesian inference

The Bayesian inference takes the observed products and explanatory variables as input and outputs posterior probability distributions over the unknown quantities (Fig. 1). The posterior distributions are obtained from the probabilistic model by conditioning on the input data. Conditioning is a well-defined mathematical operation, but analytical solutions are infeasible for complicated models like ours (MacKay, 2003). Instead, one has to resort to approximations. Monte Carlo methods are arguably
30 the most popular. Their output is an approximation to the posterior distribution that consists of samples drawn from this distribution. Here, we rely on Hamiltonian Monte Carlo as implemented using the adaptive No-U-Turn Sampler in pymc3 (Hoffman

Table 1. The default model specification used in both the simulation study and the SMAP case study, and the baseline configuration for the simulation runs. The bias terms for the reference product y_0 were assumed known.

	Model specification	Simulation parameters
Errors		
σ [$\text{m}^3 \text{m}^{-3}$]	exponential prior	[0.02, 0.04, 0.05]
m [$\text{m}^3 \text{m}^{-3}$]	Student prior $m_0 = 0$	[0.00, 0.03, -0.05]
l [-]	Student prior $l_0 = 1$	[1.0, 1.1, 0.9]
μ [$\text{m}^3 \text{m}^{-3}$]	Student prior $\mu_0 = 0$	[0.0, 0.02, -0.02]
λ [-]	Student prior $\lambda_0 = 0$	[0.0, 0.06, 0.0]
κ [-]	Student prior $\kappa_0 = 0$	[0.0, 0.2, -0.2]
Ancillary components		
ϵ distribution	Normal	Normal
θ distribution	logistic (A, B, ϕ)	logistic

and Gelman, 2014; Salvatier et al., 2016). [..]The No-U-Turn Sampler produces successive, dependent samples of the posterior distribution that are called a chain. Each sample consists of draws from the posterior distribution, or actually an approximation thereof, of all the unobserved random variables (Output in Fig. 1b). They comprise the parameter random variables (e.g. the time-dependent biases) as well as the soil moisture time series, i.e. one value of θ for each SMAP observation. For each location, we sample two independent chains with 2000 samples each, which standard quality controls (divergences, chain mixing) indicate is sufficient. Following common practice, the first 1000 samples are discarded (Brooks and Gelman, 1998)

3 Simulation study

We now study the applicability of Bayesian triple collocation using a simulation study. We used three simulated products with realistic error properties (Tab. 1). y_0 was taken to be the reference product in both the simulation and in the probabilistic models. For the other two products the biases and error magnitudes were assumed dependent on a normalized explanatory variable that varied seasonally (Tab. 1). The soil moisture was prescribed using a simple antecedent precipitation model driven by a seasonally-dependent Hidden Markov Model rainfall generator, which gave rise to autocorrelated soil moisture. Mimicking the SMAP sensor, an observation was made every 2-4 days.

We first analysed the fidelity with which the error parameters could be estimated. To this end, we simulated $R = 25$ time series and computed the posterior distribution using the default probability model of the previous section, summarized in Tab. 1. We computed an [..]aggregated RMSE error for each parameter π by comparing the prescribed parameters π_n for all

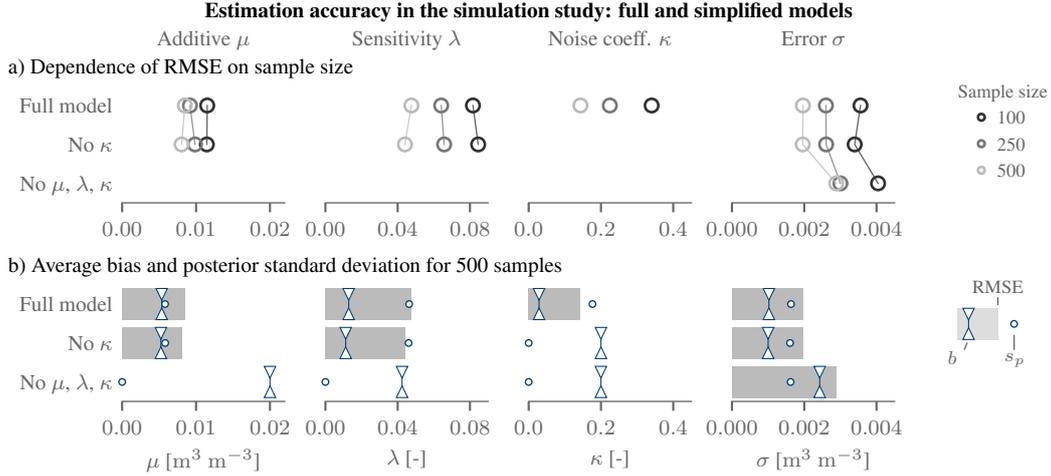


Figure 2. Simulation results illustrating the estimation fidelity. a) Dependence on sample size. b) Comparison of RMSE and the simulation bias magnitude b to the posterior standard deviation s_p . In a) and b) the first line shows the full model, whereas κ is set to zero in the inference in the second, and λ , μ and κ were set to zero in the third. In these cases, the RMSE is equal to b , and these values are not shown in a).

products with the posterior $[\cdot]$ means inferred from each simulation run $[\cdot]_r$

$$\text{RMSE} = \left(\frac{1}{N_\pi R} \sum_{n=1}^{N_\pi} \sum_{r=1}^R (\hat{\pi}_{n,r} - \pi_n)^2 \right)^{\frac{1}{2}} \quad (6)$$

where $\hat{\pi}_{n,r}$ is the posterior mean of parameter π estimated for product n (out of N_π) in run r . To quantify systematic deviations, we computed an average absolute bias (in the root mean square sense) between the truth and posterior

5 mean for each parameter:

$$b = \left(\frac{1}{N_\pi R} \sum_{n=1}^{N_\pi} \left(\sum_{r=1}^R (\hat{\pi}_{n,r} - \pi_n)^2 \right) \right)^{\frac{1}{2}}, \quad (7)$$

and we subsequently averaged the magnitudes over all products (referred to as mean absolute bias). To explore the impact of the number of observations on the accuracy, we did this for 100, 250 and 500 observations, which bound the number of observations that were available in the SMAP study.

10 The error parameters could be estimated with sufficient fidelity in the simulation study (Fig. 2; full model). The accuracy of the estimated time-variable additive bias parameter μ was found better than $0.01 \text{ m}^3 \text{ m}^{-3}$, and thus likely sufficient to detect relevant non-constant biases. The sensitivity coefficients λ are retrieved with comparable precision: the RMSE of 0.05 corresponds to a differential bias between dry and wet conditions of around $0.01 \text{ m}^3 \text{ m}^{-3}$ (assuming a soil moisture dynamic range $\sim 0.25 \text{ m}^3 \text{ m}^{-3}$). Again, this should be sufficient for many applications. Finally, the time-constant noise magnitude σ

15 could be estimated accurately ($\ll 0.005 \text{ m}^3 \text{ m}^{-3}$).

Bayesian triple collocation yields a distribution of the parameters and thus naturally provides uncertainty estimates. [.]The posterior standard deviation s_p compares favourably to the RMSE errors: Fig. 2b shows that the posterior standard deviations are within 10 to 20 % of the RMSE errors. Even though these two quantities represent different kinds of uncertainty, they are expected to be comparable for large samples (MacKay, 2003). The posterior distribution hence provides a useful summary of the estimation uncertainty.

To test the sensitivity of the estimates to model assumptions, we extended the simulation study. The most critical aspect turned out to be the specification of the bias terms: neglecting variable bias terms can impair the overall estimation quality. Neglecting the complex error structure leads to an overestimation of the error magnitude (Fig. 2). In our case, setting μ , λ and κ to 0 induced an increase in the RMSE of σ by a factor of two. This increase was mainly due to a large bias, as the varying offsets were wrongly attributed to quasi-random noise. The posterior uncertainty estimates also became inaccurate. Conversely, neglecting only the variability of the error magnitude (i.e. setting $\kappa = 0$) had limited impact on the retrieval of the other parameters.

To test for additional model assumptions, we modified the model and the forward simulations. The impact on the estimation accuracy was typically limited (see the supplement), so we only provide a short summary. First, the model for the noise term ϵ had a moderate influence on the estimation quality. A mismatch between assumed and simulated ϵ distributions increased the RMSE of the error variance by a small amount ($<0.001 \text{ m}^3 \text{ m}^{-3}$ for σ) for a range of error distributions and strongly autocorrelated errors. The other crucial assumption in the model is the probability distribution for the soil moisture. Also here, the changes are typically small (up to 10% improvement in the RMSE, but a decrease in bias) when replacing the standard time-invariant model by a seasonally variable model or by a different time-invariant model. The improvement suggests that the model-internal soil moisture distribution can have an impact on the estimated biases, in particular when the actual soil moisture is correlated with the explanatory variable, as it was in the simulated data. We would hence expect that for most applications it is the seasonal and sub-seasonal time scales that the soil moisture model should be able to capture. For comparison, autocorrelation on the inter-storm time scale that is not captured by our model but present in the simulated data did not seem to introduce major limitations (sufficient fidelity for the full model, Fig. 2). The prior distributions had an even smaller impact. Making the prior distributions twice as wide or the tails less heavy changed the estimates by only a few percent.

4 SMAP case study

4.1 Materials and methods

4.1.1 Data

[.]To estimate the biases of the SMAP soil moisture product, we used $N = 3$ soil moisture data sets in the probabilistic inference: apart from the SMAP product, these were in-situ data from dense networks or sparse sites, and MERRA-2 reanalysis data. The SMAP data set we analysed was the SMAP enhanced Level-2 soil moisture product, which is

Table 2. Network sites from north to south, including their Koeppen-Geiger climate regime.

Site	Location	Climate	Crop cover [%]	Samples
Kenaston	Canada (Saskatchewan)	Cold	90	351
Carman	Canada (Manitoba)	Cold	80	352
South Fork	USA (Iowa)	Cold	90	323
REMEDHUS	Spain	Temperate	80	475
Fort Cobb	USA (Oklahoma)	Temperate	60	413
Bell Ville	Argentina	Arid	90	399
Monte Buey	Argentina	Arid	90	400

disseminated on the 9 km EASE-Grid 2.0 at a resolution of 33 km (Chan et al., 2017; O' Neill et al., 2017). It contains a variety of estimates of the top (5 cm) soil moisture, of which we chose the standard product (baseline V single channel algorithm, 9 am morning passes). The single channel retrievals rely on an a priori τ derived from a MODIS climatology; these τ values are included in the disseminated product. We studied all available data since the beginning of the record in April 2015 until August 2017, i.e. up to three annual growing seasons. After removing flagged retrievals (Colliander et al., 2017), the number of available measurements is on the order of 300-500, which is not ideal but should be sufficient according to Fig. 2a.

The analyses focus on seven locations in North America, South America and Europe with significant crop cover, due to the availability of high-quality dense in-situ networks (Tab. 2). At these SMAP core or candidate sites, continuous calibrated in-situ measurements at 5 cm depth are collected at multiple locations within a SMAP grid cell (Colliander, 2017; Colliander et al., 2017).

To provide a better overview of the spatial patterns, we also used data from > 200 in-situ sites in the contiguous United States (SCAN and USCRN networks). These sparse sites consist of a single station per satellite pixel, and their representativeness is hence not comparable to that of the network sites. The USDA's SCAN network has been in continuous operation since 1999 and provides soil moisture data at 2 in (5 cm) depth (Schaefer et al., 2007). The USCRN network consists of 114 sites whose location was chosen to be maximally representative of its surroundings; we used the 5 cm soil moisture observations (Bell et al., 2013; Diamond et al., 2013; Palecki et al., 2017). We assigned these sites a dominant land cover based on the MODIS MCD12C1 land cover product (Friedl et al., 2017).

For the third soil moisture product we used the MERRA2 reanalysis (M2T1NXLND.5.12.4) (Gelaro et al., 2017; Global Modeling and Assimilation Office, 2017). It is the most recent reanalysis product of NASA's Global Modeling Office, available at a resolution of 55 km. For sensitivity analyses, we also used GLDAS-2 (Noah model, GLDAS_NOAH025_3H.2.1), a popular land assimilation data set (Roddell and Beaudoin, 2017).

To quantify the error structure as a function of $\Delta\tau$, we estimated $\Delta\tau$ as the difference between an external estimate of τ and the SMAP input τ . The external estimate was based on independent L-band microwave observations by the SMOS satellite (Level 3 operational algorithm). The multi-angular observations and the temporal aggregation of multiple overpasses are conducive to providing robust, if noisy, measurements of the vegetation optical depth relevant to SMAP (Al Bitar et al.,

2017). To reduce the impact of high-frequency noise, the $\Delta\tau$ data were smoothed to a temporal resolution of 16 days (LOWESS filter). The explanatory variable used in the Bayesian inference is the normalized version

$$w_{\Delta\tau}(t) = \frac{1}{\text{std}(\Delta\tau)} (\Delta\tau(t) - \text{mean}(\Delta\tau)). \quad (8)$$

4.1.2 Error estimation

5 In our estimation we specified the error structure based on our hypothesized impact of a vegetation misspecification. The τ - ω model (Kurum et al., 2011) predicts that a vegetation misspecification $\Delta\tau = \tau_{\text{inv}} - \tau_{\text{true}}$ will induce both a sensitivity L and an offset M which scale approximately linearly with $\Delta\tau$ (Fig. 3a). There is thus only one explanatory variable ($P_{\lambda,n} = P_{\mu,n} = 1$), namely the normalized $w_{\Delta\tau}$ of Eq. 8. By slightly simplifying the notation, the error model reads

$$y_{\text{SMAP}}(t) = \underbrace{(l + \lambda w_{\Delta\tau}(t))}_{L(t)} (\theta(t) - \theta_0) + \underbrace{(m + \mu w_{\Delta\tau}(t))}_{M(t)} + \theta_0 + \epsilon(t). \quad (9)$$

Here θ_0 was taken to be the mean value of the in-situ product. We quickly recall the interpretation of the error parameters. λ and μ quantify the temporal changes in the sensitivity and offset, respectively, that are associated with changes in $\Delta\tau$. Their magnitudes correspond to the temporal standard deviation of the sensitivity and offset, respectively (Fig. 3b). Their sign expresses the direction of the dependence on $\Delta\tau$. It is predicted to be positive: as $\Delta\tau$ grows, the inversion increasingly
 15 overcompensates the vegetation-induced loss of sensitivity to soil moisture and increase in brightness temperature, which in turn inflates both L and M , respectively.

[..]

To compute the predicted biases in Fig. 3a), we assumed the τ - ω model applied and was correctly specified (temperature, dielectric mixing model [Dobson; silt loam], single-scattering albedo $\omega = 0.05$, etc.). For a given value of τ_{true} , we
 20 simulated the V-polarized brightness temperatures for dry and wet soil moisture conditions. These brightness temperatures were in turn the basis for estimating soil moisture by inverting the τ - ω model using the wrong τ_{inv} as a function of $\Delta\tau$. For both dry and wet soil moisture conditions, the deviation was an estimate of the retrieval bias: their mean was taken to be an estimate of the offset M , whereas their difference allowed us to estimate L . When plotted against $\Delta\tau$, M and L increase nearly linearly and only show a weak dependence on τ_{true} . The slope of this relation is thus well
 25 but not perfectly defined. We refer to the slopes as μ^* (for M) and λ^* (for L), respectively. To account for the spread due to the slight curvature and dependence on τ , we estimated the likely range of values by computing the slopes from the differences in L or M between five equally spaced values of $\Delta\tau$ (between -0.1 and 0.1), repeated for equally many values of τ_{true} (between 0.1 and 0.6). The range of these values was $\lambda_{\text{pred}}^* \in [2.0, 3.8]$ and $\mu_{\text{pred}}^* \in [0.33, 0.65] \text{ m}^3 \text{ m}^{-3}$. These ranges will later be compared to data-driven estimates, thus providing a first-order assessment of the agreement
 30 between predictions and observations, despite the simplified setup and neglect of other retrieval errors.

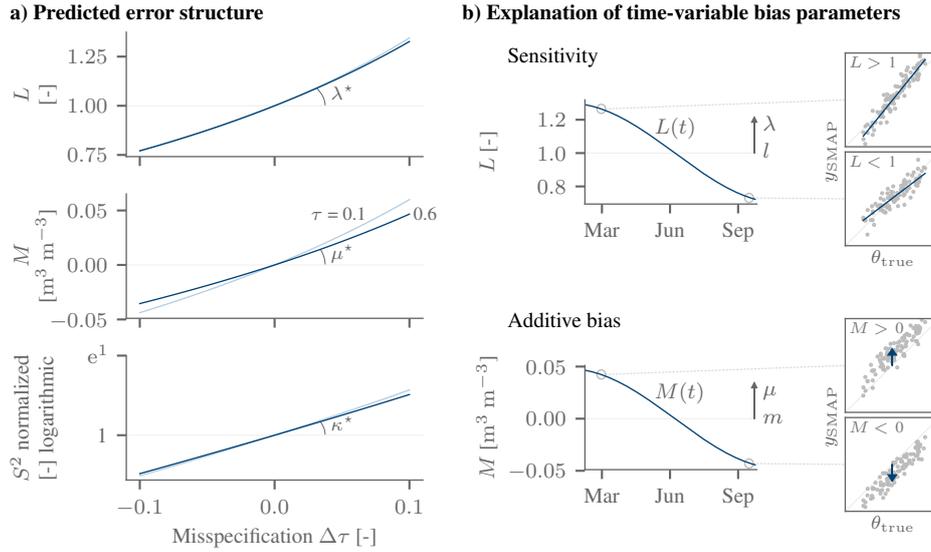


Figure 3. a) The expected error structure due to a misspecified vegetation in the SMAP retrieval. Additive bias M , sensitivity L and variable noise level S^2 induced by a misspecified vegetation optical depth τ , as predicted by the τ - ω model for the SMAP satellite and single-scattering albedo $\omega = 0.05$. Two different values of the true τ were assumed (0.06 in dark blue, 0.01 in light blue). The slopes of these approximately linear relationships will later be compared to independent data-driven estimates. b) General explanation of the bias terms, illustrated for a hypothetical time-changing sensitivity $L(t)$ and offset $M(t)$ (underlying change in $\Delta\tau$ not shown). A varying sensitivity changes the response of the SMAP retrieval to a unit change in the true soil moisture. When it is larger than one, the SMAP data have a larger dynamic range than the true soil moisture (illustrated by the slope > 1 in the inset). The time-average value of L is l , and the temporal standard deviation of L is given by $|\lambda|$ (length of arrow). A variable M induces non-constant offsets, and the magnitude of its temporal variability is given by $|m|$. $M > 0$ corresponds to a positive offset (shown in the inset).

The same overcompensation that increases the sensitivity also increases the noise level, so that we also made the variance of $\epsilon(t)$, $S^2(t)$, time dependent. We used two explanatory variables, $\exp\Delta\tau$ and the external τ (both normalized, $P_{\kappa,n} = 2$). The reason for also including τ was that the noise level S^2 is predicted to depend on τ even if $\Delta\tau$ is constant, in contrast to the bias terms. By transforming $\Delta\tau$ to $\exp\Delta\tau$, the predicted dependence of S^2 shown in Fig. 3 could be reproduced accurately.

- 5 The dependence on $\exp\Delta\tau$ is positive, i.e. an increase in $\Delta\tau$ is associated with an increase in noise level. We would thus expect $\kappa_{\exp\Delta\tau} > 0$.

The same error structure was assumed for the re-analysis data[. . .]. The inclusion of a $\Delta\tau$ -dependent bias for the reanalysis product is not driven by physical reasoning, but for statistical reasons. By controlling for the same explanatory variables for both products, the impact of potential confounders - e.g. a seasonal bias that is correlated with $\Delta\tau$ - on the bias estimates of the remotely sensed product can be reduced. If this were not done, the model would try to partially adjust the time-variable bias term of the remote sensing product to minimize the systematic differences to the re-analysis product, thus distorting the bias estimates. The in-situ observations were taken as reference product [. . .] y_0 : $L = 1$, $M = 0$, $\kappa = 0$). All

other model components – the soil moisture and error distributions, the priors – [..]were identical to the simulation study (Tab. 1).

[..]

To compare the estimated biases with the model predictions, we re-expressed λ and μ in absolute terms by reversing the rescaling from $\Delta\tau$ to $w_{\Delta\tau}$. Thus, [..]

$$\lambda^* = \frac{\lambda}{\text{std}(\Delta\tau)} \quad \text{and} \quad \mu^* = \frac{\mu}{\text{std}(\Delta\tau)} \quad (10)$$

are, respectively, the slope of L and M vs $\Delta\tau$. In other words, [..]the estimated λ^* describes the inferred change in L for a unit change in $\Delta\tau$, and similarly for μ^* . They can be directly compared to the model predictions [..]described above. Note that the division only reverses the scaling but not the offset inherent in the normalization from $\Delta\tau$ to $w_{\Delta\tau}$. Estimating the derivative at the mean value of $\Delta\tau$ (due to the offset) rather than at $\Delta\tau = 0$ has, however, no impact for a purely linear relation.

To test the robustness of the estimates, we varied the input data and the model configuration. Instead of using the SMOS τ as reference, we also derived $\Delta\tau$ from the SMAP dual channel (LOWESS smoothed) retrievals (O’ Neill et al., 2017) and from contemporaneous MODIS NDVI data (Didan, 2017). We converted the MODIS NDVI to τ using the same equations as in the generation of the SMAP input climatology. We refer to these model runs as SMAP DC τ and MODIS τ , respectively. We also modified the external soil moisture products: instead of MERRA-2 we used GLDAS-2 (GLDAS θ), and we also dropped the reanalysis data set altogether (no reanalysis). This is possible in a Bayesian setting; to improve the identification of the errors, we assigned a narrower prior distribution to the in-situ noise magnitude ($0.02 \text{ m}^3 \text{ m}^{-3}$). [..]To account for a potential confounding of τ itself, which may also have an effect on the biases, we included the smoothed SMOS τ as second explanatory variable for L and M , referred to as τ control. Finally, we also modified the model configuration. We made the probability model for θ vary seasonally [..](θ spline) to account for the non-negligible impact of this detail of the model specification on the estimates observed in the simulation (Sec. 2.1.2[. .]).

4.1.3 Estimates of vegetation-soil moisture coupling

To explore the relation between time-variable vegetation biases and estimates of vegetation-water coupling, we analysed the coefficient of determination R^2 between τ and soil moisture anomalies. These were derived from the SMOS τ time series and both SMAP and in-situ soil moisture, respectively. The associated anomalies τ' and θ' were obtained by subtracting a seasonal climatology that in turn was the smoothed (30 day) multi-year average of the input time series. To compare the SMAP and the in-situ soil moisture data, we then computed the difference in the coefficient of determination ΔR^2

$$\Delta R^2 = R^2(\tau'_{\text{SMOS}}, \theta'_{\text{SMAP}}) - R^2(\tau'_{\text{SMOS}}, \theta'_{\text{in-situ}}) \quad (11)$$

If the SMAP soil moisture were only contaminated by random noise relative to the in-situ data, ΔR^2 would tend to be negative. Time-variable vegetation biases, on the other hand, can induce positive values. Time-average biases cancel out. Of course, there are also other error sources such as representativeness errors, especially for the sparse networks. Further, the time series are

comparatively short, but ΔR^2 can provide a first tentative assessment of the reliability of coupling metrics such as R^2 in the presence of time-variable biases.

4.2 Results

4.2.1 Network sites

5 SMAP biases that track the vegetation misspecification $\Delta\tau$ are common at the core validation sites. Especially changes in the sensitivity can be large. [Reanalysis bias parameters were estimated as well, but they are considerably smaller in magnitude than those of the SMAP product.](#)

The varying sensitivity of SMAP is illustrated for the South Fork site, Iowa (USA), in Fig. 4a). In July, when the predominantly cultivated corn is heading and flowering (Tomer et al., 2008), the magnitude of the SMAP response to rainfall events
10 matches that of the in-situ data. Conversely, in autumn (senescence, post harvest) the sensitivity of SMAP to soil moisture variations is diminished. The Bayesian inference reproduces this visually inferred pattern, as the sensitivity L drops by more than 0.5 (or 50%). By design, the temporal changes in L are governed by changes in $\Delta\tau$, which drops by about 0.1 from July to September. Over the entire time series, its temporal variability is $|\lambda| = 0.3$. Note that even when $\Delta\tau \approx 0$ (August), the sensitivity is too low. As $\Delta\tau$ is essentially zero on average over the entire time series ([\[.\] -0.01](#)), the sensitivity at $\Delta\tau = 0$
15 corresponds to the time-average value l . Its posterior median is 0.64 and thus less than the expected value of 1. We will return to the time-average biases later.

Pronounced changes in the sensitivity are found for all network sites but one (Fig. 4b). For the most part their magnitude is large, as the $|\lambda|$ values correspond to a temporal variability of around 10 to 50%. The inferred relation to $\Delta\tau$ is consistently positive ($\lambda > 0$), i.e. as $\Delta\tau$ increases over time, so does the SMAP sensitivity $L(t)$. The inferred direction hence matches
20 the model prediction (Fig. 3a) [of a positive \$\lambda\$](#) . The model does not constrain the magnitude of the λ parameter because the latter is an internally normalized quantity. When converted into absolute quantities (λ^*), the inferred dependence of L on $\Delta\tau$ matches the model predictions reasonably well (Fig. 4b). [In other words, the data-derived, completely independent estimate is broadly consistent with the predicted impact of a \$\tau\$ misspecification in the retrieval, despite limitations in the estimates \(e.g. issues with the reference \$\tau\$ \) and the model predictions \(e.g. assumed knowledge of the land surface temperature\) of](#)
25 λ^* . [There is no clear apparent dependence of \$\lambda^*\$ on location or land cover properties; for instance, Monte Buey and Bell Ville are within < 100 km of one another, and despite the similarity in planted crops the latter's \$\lambda^*\$ is considerably larger.](#)

Also the additive biases track changes in $\Delta\tau$ at the network sites, but to a lesser extent (Fig. 4c). Over time, the biases vary by up to $0.02 \text{ m}^3 \text{ m}^{-3}$ in magnitude. The inferred direction, $\mu > 0$, matches the predictions, as an increase in $\Delta\tau$ corresponds to a larger offset. However, the magnitude of the inferred change of the offset with $\Delta\tau$, μ^* , is generally smaller than predicted.
30 A very small value of μ^* is found at the South Fork site: it is a factor of ten smaller than its predicted range (Fig. 4c). It is also small enough to be practically irrelevant, as the additive bias at South Fork is inferred to be essentially constant (Fig. 4a).

The time-variable biases are complemented by the time-average biases, which are quite large at several network sites. The time-average sensitivity l deviates from its nominal value of one by more than 30% at three out of 8 sites (Fig. 5). It tends to

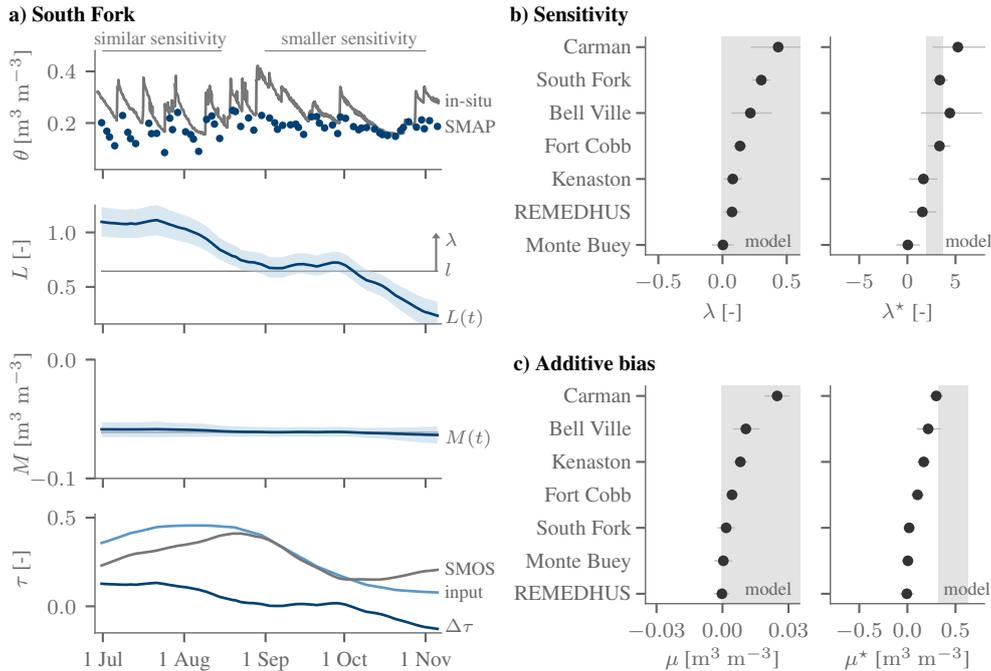


Figure 4. Time-variable biases over the network sites. a) The SMAP product’s sensitivity decreases from summer to late autumn 2015 at the South Fork site (42.4N, 93.4W), which is reflected by a decrease in the inferred sensitivity L that tracks a decrease in $\Delta\tau$: in early summer $L(t) \approx 1$, but it subsequently drops well below the time-average value of $l = 0.7$. Its temporal standard deviation is given by $\lambda = 0.3$ (arrow). $M(t)$ is approximately constant ($M(t) \approx m$) because μ is small (not shown). b) Over the network sites, the association of changes in the sensitivity with changes in $\Delta\tau$ is predominantly positive ([..] marker: posterior median, lines: 5–95% uncertainty), as predicted by the model (positive λ ; grey background). The magnitude of the dependence for a unit change in $\Delta\tau$, λ^* of Eq. 10, is broadly consistent with predictions by the τ - ω model of Sec. 4.1.2. c) The additive biases are of the right direction (positive), but the unnormalized quantities μ^* smaller than [..] predicted by the model.

be larger than one, as may be expected considering the surface soil moisture SMAP is sensitive to has a larger dynamic range than the moisture at the depth of the probes. The South Fork site of (Fig. 4a) with $l < 1$ is thus somewhat unusual. However, it is typical in that its time-average additive bias m is negative. Conversely, the noise level σ inferred using our approach tends to be small, usually on the order of $0.03 \text{ m}^3 \text{ m}^{-3}$ (Fig. 5). These values are not directly comparable to standard RMSE estimates because our approach disentangles the quasi-random noise from a sensitivity that deviates from 1, temporally variable biases associated with $\Delta\tau$ and in-situ errors. Also, the quasi-random errors tend to increase with $\Delta\tau$ (Fig. 5), as predicted by the τ - ω model (Fig. 3a). However, there is considerable variability in the magnitude across the study sites.

4.2.2 Sensitivity analysis

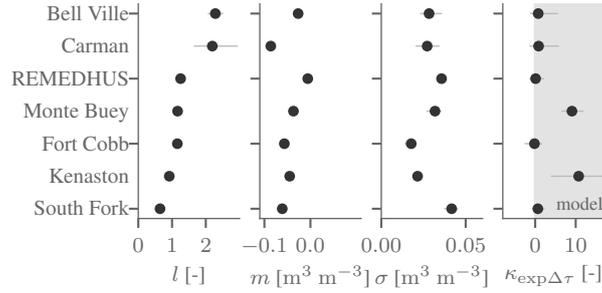


Figure 5. The time-average error properties l (sensitivity), m (additive bias) and σ (noise standard deviation), and the κ parameter (sensitivity of SMAP's noise level to $\Delta\tau$) as inferred for all network sites. Meaning of markers and lines as in Fig. 4.

[..] Our sensitivity analyses focus on the reference τ product. When the SMAP dual channel result is used as the reference τ product, the bias parameters change little for the vast majority of sites (Fig. 6[.], column 1). When the posterior uncertainties are taken into account, the λ and μ values tend to overlap with those obtained using the SMOS τ product. This is despite potential disadvantages of the SMOS product (different resolution and footprint than SMAP, different retrieval model and use of NDVI-based regularization in the retrieval (Al Bitar et al., 2017)) or of the SMAP DC product (e.g. no multi-angular information, measurement noise-induced error correlations with soil moisture estimate). Taken together, the similarity over the network sites indicates that the results are not particularly sensitive to whether the SMOS or the SMAP DC τ product serves as the reference for $\Delta\tau$.

By contrast, [..] the estimates can change substantially when τ is derived from contemporaneous NDVI data, [..] and predominantly they are smaller in magnitude (Fig. 6, column 2). If the problem with the use of the NDVI climatology in the retrieval were the use of a climatology alone, we would expect similar estimates. Conversely, we would expect the estimates to be smaller in magnitude if it was the link between NDVI and τ [..] that led to an inaccurate vegetation correction, because the NDVI-based $\Delta\tau$ would provide little information on the biases. The smaller estimates that were actually observed may thus indicate that the use of a climatology [..] is not a dominant error source in the SMAP vegetation input data. [..]

The estimates of the time-variable biases are reasonably robust to other aspects of the model specification. The impact of replacing the MERRA2 with the GLDAS2 soil moisture or dropping it altogether is also small (Fig. 6, columns 3 and 4). [..] By additionally including the smoothed SMOS τ as an explanatory variable, potential confounding on $\Delta\tau$ could be assessed: the estimates changed very little for all but one station (Fig. 6, column 5). When allowing the soil moisture parameters A and B to vary seasonally (Eq. 5), the parameter estimates do not change substantially for all sites but one (Fig. 6, [..] column 6). The standard setup of the probabilistic model hence seems adequate for quantifying the SMAP error structure.

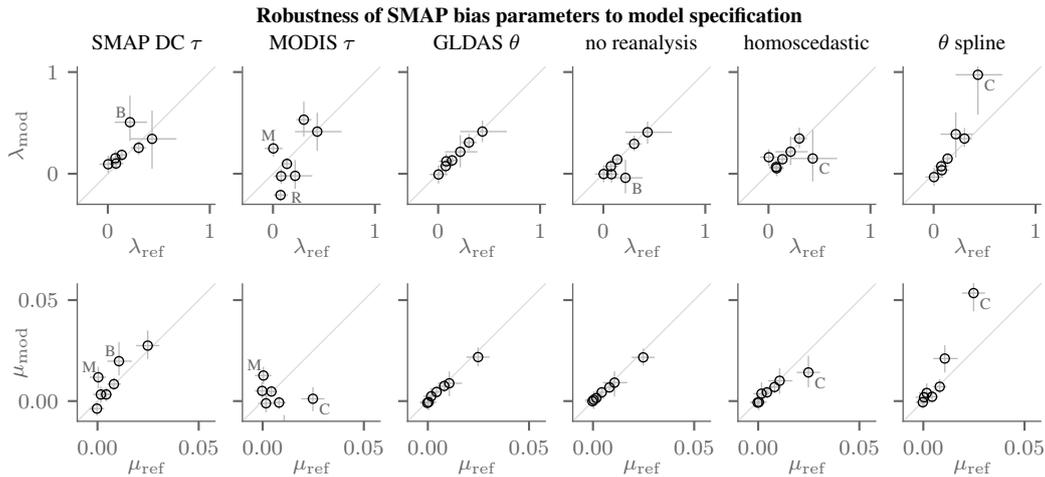


Figure 6. Robustness quantified by changes in the estimated λ (top row, [-]) and μ (bottom row, [$\text{m}^3 \text{m}^{-3}$]) parameters. Each panel compares the reference estimates on the horizontal axis (median: marker, lines: 10–90% posterior probability interval) to those obtained with the modified model on the vertical axis (cf. Sec. 4.1.2). The most prominent outliers are annotated – B: Bell Ville, C: Carman, M: Monte Buey, R: REMEDHUS.

4.2.3 Sparse sites

Across the sparse sites within the contiguous US we commonly find pronounced time-variable biases (Fig. 7a-b). While the results over the sparse sites are deemed much less reliable than those over the network sites, they are similar in areas of overlap. The largest changes in sensitivity of $\lambda \approx 0.3$ are found over croplands (e.g. Midwest, Central Valley in California). As predicted by the model, the λ are predominantly positive but notable exceptions occur in the Mississippi Delta. Large and positive λ are also common over pastures and grasslands. The additive bias parameters μ tend to show a similar spatial pattern, in that they are largest over crop- and grasslands (Fig. 7b). [..]

Large biases over crop- and grasslands are also found when $\Delta\tau$ is computed from the SMAP DC product (Fig. 8a,b). The spatial patterns are very similar. Again, there is a close correspondence between the sparse and the network sites. Thus, irrespective of the $\Delta\tau$ product, the analysis of the sparse sites reveals sizeable biases over grasslands in addition to croplands.

To analyse the estimated noise level for all three products, we computed a normalized version σl^{-1} , where the division by l accounts for the different dynamic ranges of the three products by scaling the noise level with respect to the in-situ data (Fig. 9; SMOS-based $\Delta\tau$). SMAP achieves a median value of $0.045 \text{ m}^3 \text{ m}^{-3}$, a higher value than that of the in-situ data or MERRA-2 (0.029 and $0.040 \text{ m}^3 \text{ m}^{-3}$, respectively). For all three products, the corresponding values over the network sites are smaller by $\sim 50\%$. For the in-situ data, the larger noise level at the sparse sites is not surprising, owing to their limited representativeness. However, direct comparisons could be misleading. For instance, the larger noise

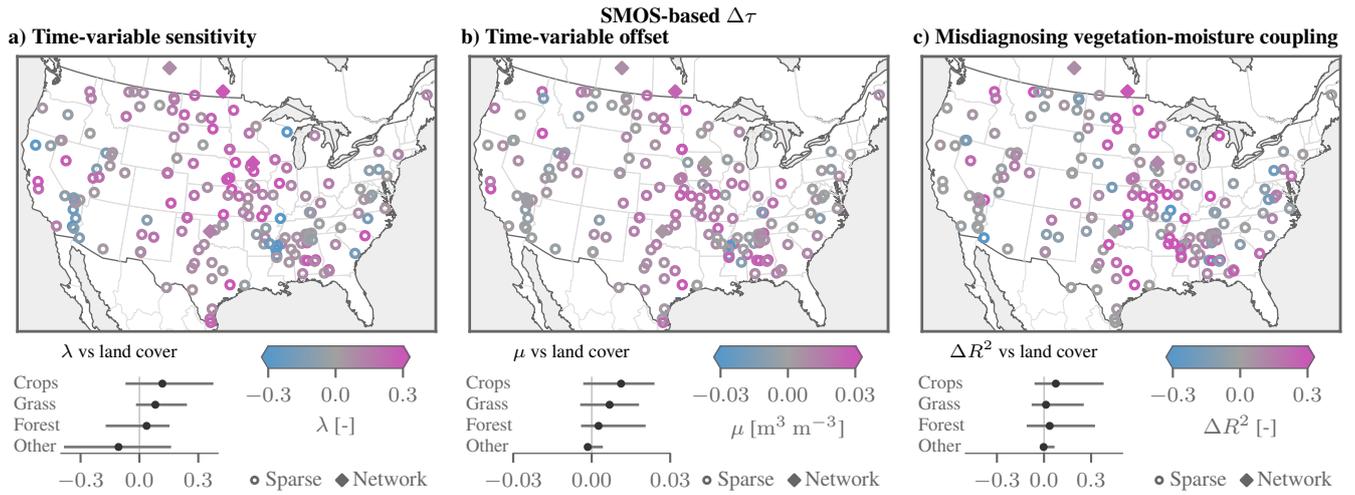


Figure 7. Time-variable biases and ΔR^2 coupling metric across the contiguous US obtained using the SMOS τ product, shown for the network sites and the sparse sites (only those with >250 valid SMAP observations). a) The time-dependent sensitivity parameters λ are predominantly positive, and the largest magnitudes are found over crop- and grasslands. b) The additive biases exhibit a similar spatial pattern. c) The degree of association R^2 between anomalies of vegetation τ and soil moisture is commonly higher for SMAP than for in-situ soil moisture ($\Delta R^2 > 0$).

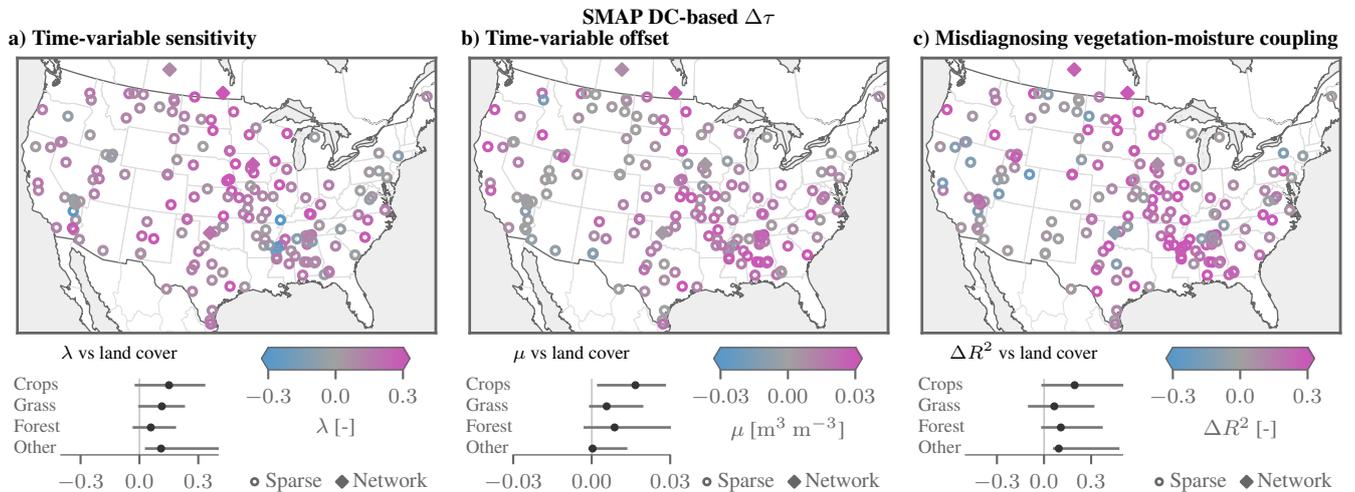


Figure 8. Time-variable biases and ΔR^2 coupling metric across the contiguous US, obtained using the SMAP DC τ product. Cf. Fig. 17 for results based on the SMOS τ product.

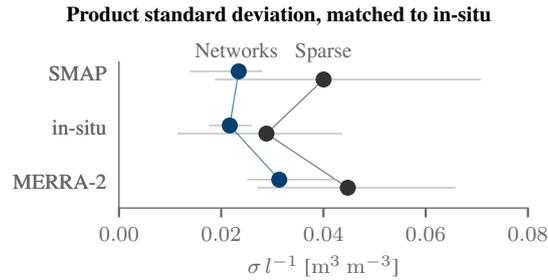


Figure 9. Estimated noise level normalized to the in-situ dynamic range. For both the network and the sparse in-situ sites, the distribution of the posterior median values of σl^{-1} is summarized by the median (marker) and the interquartile range (horizontal bar).

level estimates (and greater spread of these estimates) may be partially accounted for by the small number of available networks and by the heterogeneous land cover and vegetation conditions across the sparse sites in the contiguous US.

4.2.4 Vegetation-soil moisture coupling

The observed coupling between vegetation and soil moisture anomalies is larger when using SMAP than when using in-situ soil moisture data (Fig. 7c). Positive values of ΔR^2 are particularly pronounced over croplands (0.12 on average), with the spatial pattern largely conforming to that of the time-variable biases. Conversely, if the time-variable errors in the remotely sensed soil moisture were purely random, the degree of association would decrease, i.e. $\Delta R^2 < 0$. Even though spatial representative errors are likely large for the sparse networks, the ΔR^2 values are comparable at proximal network sites. When the SMAP DC τ product is used instead, the results are similar in terms of both magnitudes and spatial patterns (Fig. 8c).

10 5 [..]

[..] While the spatial patterns largely match those of the time-variable [..] biases, the link between them is not clear and not necessarily uniform across all sites. The computation of anomalies largely removes seasonal offsets, which constitute a major fraction of the estimated additive biases. However, it does not remove higher-frequency variations or inter-annual differences, although the record is too short to reliably study those. Neither can it account for the changes in sensitivity, which are particularly large over croplands. Finally, the in-situ soil moisture anomalies, predominantly derived from single probes, are subject to major uncertainties. All these factors likely contribute to the elevated associations between the τ and the SMAP soil moisture anomalies ($\Delta R^2 > 0$), but the precise impact of time-variable [..] biases on our ability to diagnose such interactions remains an open question.

[..]

5 Discussion

5.1 Detected time-variable errors in the SMAP product

By applying Bayesian triple collocation to the SMAP soil moisture product, we detect time-variable biases. These time-variable biases track the misspecification of the vegetation optical depth $\Delta\tau$ during the soil moisture retrieval. They are both additive and multiplicative, i.e. not only the offset but also the sensitivity changes over time. Especially the changes in sensitivity can be large over croplands, as seasonal variations in $[\dots]L$ on the order of 30% ($\lambda \approx 0.3$) deserve attention in future studies (Fig. 4, 7). The spatial patterns of the time-variable biases largely match those of the temporal autocorrelation estimated by Dong et al. (2018), which were also largest over croplands. These and our findings suggest that the NDVI-based vegetation correction in the SMAP retrieval introduces particularly large errors in agricultural regions.

A mechanistic interpretation of the inferred biases is complicated by a number of poorly understood factors. First, the time-variable biases are relative to the in-situ data. The results over the sparse sites should hence be interpreted with caution due to representativeness error, even if they are similar to those at the dense high-quality network sites $[\dots]$ (Fig. 7). Also at the network sites, residual time-dependent biases $[\dots]$ in the in-situ data cannot be ruled out completely.

Another major uncertainty are errors in the satellite-derived τ products, which are not accounted for in the estimation. One reason why these errors are difficult to quantify is that in the context of soil moisture retrieval τ can be considered as essentially a model-internal effective quantity (Parrens et al., 2017). As such, an observation-based estimate of τ reflects not only the vegetation conditions but also inaccuracies of the tau-omega model itself, the way it is parameterized and other environmental conditions. An instance for the latter are roughness changes associated with harvest in croplands (Patton and Hornbuckle, 2013), which likely contribute to the autumnal increase in SMOS τ in Fig. 4. To a good degree of approximation, roughness changes will be captured by the effective τ that the SMOS or SMAP DC algorithms retrieve from the brightness temperatures (Parrens et al., 2017). Nevertheless, the estimates used in this study will still be affected by errors with respect to this effective quantity that can be random and systematic (e.g. due to the different incidence angles for SMOS and SMAP). The impact of such errors on the estimated biases is unknown, but analogies to simple regression models suggest that they can distort these estimates in either direction (Frost and Thompson, 2002).

While it is premature to attribute the inferred biases completely to an imperfect vegetation correction, there are two lines of reasoning that suggest that the inferred biases are not spurious. First, they are fairly consistent across croplands, and also between sites with sparse and dense in-situ $[\dots]$ networks (Fig. 7). Also, they tend to be large both in absolute terms (e.g. $\lambda > 0.1$) and compared to the posterior uncertainties. Further, they are also robust to the specification of the input τ product (SMAP DC instead of SMOS) and to several model modifications (Sec. 4.2.2). However, these results are purely descriptive in that they only quantify associations, rather than establishing a causal link. A first step towards such a mechanistic interpretation is the comparison of the time-variable biases with predictions by the τ - ω model. This second line of reasoning suggests that the estimated multiplicative biases λ^* are largely consistent with theoretical expectations (Fig. 4). However, this analysis is contingent on i) the tau-omega model being appropriate and correctly specified (e.g. $[\dots]$

]known ω), ii) there being no confounding biases such as seasonal inundation, and iii) the sufficient accuracy of the input τ product. It is difficult to dispel these concerns, and indeed the deviations from the predictions (for μ^*) may indicate that unconsidered phenomena also contribute to the time-varying biases in addition to those resulting from the vegetation correction.

5 One further caveat is that also time-average biases are present (additive bias: $m \neq 0$, sensitivity: $l \neq 1$). For instance, the SMAP retrievals at the South Fork site have too low a sensitivity ($l < 1$) and are too dry (negative m) on average (Fig. 4a). Our analysis has focused on the time-dependent biases, partly because the time-independent biases are better known and more commonly compensated for in analyses such as data assimilation studies (Yilmaz and Crow, 2013; Kornelsen and Coulibaly, 2015; Colliander et al., 2017). Also, there are many potential sources for these time-invariant biases between the retrievals and
10 the in-situ data, e.g. the calibration of the in-situ probes, the dielectric mixing model in the retrieval, or an offset in the mean input vegetation optical depth (i.e. $\text{mean}(\Delta\tau) \neq 0$).

We conclude that our key finding is the presence of sizeable time-variable biases in the SMAP product. They are associated with, but likely not entirely caused by, deviations of the a priori τ used in the soil moisture retrieval from τ estimated from contemporaneous microwave data.

15 5.2 Implications for hydrological applications

The time-varying biases can have a negative impact in many applications. The changing sensitivity impedes the seasonal comparison of soil moisture dynamics, as the same SMAP-observed change corresponds to a wide range of actual soil moisture changes depending on the season [..](e.g. Fig. 4a). Variable sensitivities are particularly problematic for characterizing droughts, as extreme conditions may not be apparent in the SMAP data when the sensitivity happens to be small [..](as would
20 happen for reduced vegetation water content). These issues also carry over to inter-annual comparisons. Inter-annual differences in the vegetation water content may, owing to the use of a climatology for τ in the retrieval, induce a spurious vegetation signal in the soil moisture [..]retrievals.

The spurious vegetation signal in the soil moisture data may distort estimates of water-vegetation coupling. We find inflated values of R^2 between the SMOS vegetation optical depth and SMAP soil moisture, whereas purely random noise
25 would decrease the R^2 (Fig. 7c). [..]While the spatial patterns largely match those of the estimated biases, this does not imply a causal link between the two. However, the inflated R^2 values hint at potential pitfalls in [..]using remotely sensed soil moisture to study global hydrology.

[..]

5.3 Estimating complex error structures

30 Soil moisture products can be subject to complex, time-variable errors, as revealed by our novel method for estimating such complex error structures from data. Other estimation procedures are conceivable, especially if high-quality in-situ data are available, and should be explored in the future. Our Bayesian triple collocation approach is widely applicable because it yields consistent estimates of error magnitudes and biases even when no error-free reference data set is

available. It does, however, have to be assumed to be free of systematic error. The method is flexible, so that the error structure parameterization can be adapted to the problem at hand. We hope that this will enable the community to better characterize the uncertainties of remotely sensed soil moisture products. The knowledge of time-variable structural errors is key to improving the products, and it also helps to inform the application of these data sets in practice.

- 5 The presented approach could be applied to a wide range of variables besides soil moisture, such as wind speed, land surface fluxes and leaf area index. The issue of non-constant error sources, be they associated with environmental conditions or varying observational parameters, likely pertains to many such variables. By shedding light onto residual biases, our approach could in the future contribute to the development of improved retrieval approaches.

6 Conclusions

- 10 We developed a probabilistic approach for estimating complex error structures to study time-variable biases in the SMAP soil moisture product. We hypothesized that temporal changes in the error structure arise due to an inaccurate vegetation correction in the retrieval, so that the biases relative to the in-situ data track the misspecification in the vegetation optical depth $\Delta\tau$. Our conclusions are as follows

1. Sizeable temporal changes in the offset and the sensitivity were detected, and they were particularly large over croplands (e.g. change in sensitivity $\sim 30\%$).
2. While the estimated time-variable biases track the $\Delta\tau$ (with respect to contemporaneous SMOS or SMAP dual channel estimates), they are not necessarily entirely caused by an inappropriate vegetation correction. The attribution to this source is complicated by the potential presence of confounding variables or errors in the reference τ . However, the time-variable multiplicative biases match the expected biases, as predicted by the τ - ω model, in direction and magnitude, suggesting that the vegetation correction is indeed an issue. Conversely, the additive biases only match in direction.
3. The time-variable biases impede the seasonal comparison of remotely sensed soil moisture values. In particular, extreme conditions like droughts may not be apparent in the SMAP data when the sensitivity happens to be small.
4. The presented estimation approach is widely applicable because it yields consistent estimates of error magnitude and biases even when no error-free reference data set is available. Further, it is flexible in that a wide range of different kinds of error structures can be estimated purely from observations.

More generally, our findings illustrate the importance of recognizing time-variable biases in [..] soil moisture products. The uncertainty analyses, including the choice of metrics or error parameters, must account for them. The widely used RMSE cannot distinguish between such systematic errors and white noise. Neglecting that distinction can easily give rise to misleading interpretations, for instance into how water availability regulates [..] vegetation processes. Remotely sensed soil moisture products provide unique insight into [..] such vegetation-hydrology couplings, but only up to a point as phenology-dependent

biases of unknown size are potentially present in all remotely sensed products. Neglecting the associated seasonal and inter-annual biases will distort observation-based estimates of the couplings. Robust estimates of such complex error structures can help to mitigate these biases, and thus to exploit the full potential of observational data sets.

7 Acknowledgements

5 The authors are grateful to Kaighin McColl, Dara Entekhabi and his [[. \]group](#), and the SMAP team for comments and suggestions. [They thank Alexandra Konings and Wade Crow for insightful reviews](#). S.Z. acknowledges support from the Swiss National Science Foundation (P2EZP2_168789). The support of the Canadian Space Agency, Environment and Climate Change Canada and the Natural Sciences and Engineering Research Council of Canada is acknowledge for support of the Kenaston network. A contribution to this work was made at the Jet Propulsion Laboratory, California Institute of Technology, under a
10 contract with the National Aeronautics and Space Administration. The University of Salamanca team's involvement in this study was supported by the Spanish Ministry of Economy and Competitiveness with the project PROMISES: ESP2015-67549-C3 and the European Regional Development Fund (ERDF). The remotely sensed and reanalysis data are available at the URLs provided in the references (registration generally required). The sparse in-situ data can be found at *ismn.geo.tuwien.ac.at*. The SMAP network in-situ data are available from <https://nsidc.org/data/nsidc-0712> until October 2016; the 2017 data are expected
15 to be uploaded in early 2018; until then, they are available from the authors upon request. The Python code for Bayesian triple collocation has been made available at <https://github.com/szwieback/BayesianTripleCollocation>.

References

- Adegoke, J. O. and Carleton, A. M.: Relations between Soil Moisture and Satellite Vegetation Indices in the U.S. Corn Belt, *Journal of Hydrometeorology*, 3, 395–405, [https://doi.org/10.1175/1525-7541\(2002\)003<0395:RBSMAS>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0395:RBSMAS>2.0.CO;2), 2002.
- Al Bitar, A., Mialon, A., Kerr, Y. H., Cabot, F., Richaume, P., Jacquette, E., Quesney, A., Mahmoodi, A., Tarot, S., Parrens, M., Al-Yaari, A., Pellarin, T., Rodriguez-Fernandez, N., and Wigneron, J.-P.: The global SMOS Level 3 daily soil moisture and brightness temperature maps, *Earth System Science Data*, 9, 293–315, <https://doi.org/10.5194/essd-9-293-2017>, 2017.
- Bell, J. E., Palecki, M. A., Baker, C. B., Collins, W. G., Lawrimore, J. H., Leeper, R. D., Hall, M. E., Kochendorfer, J., Meyers, T. P., Wilson, T., and Diamond, H. J.: U.S. Climate Reference Network Soil Moisture and Temperature Observations, *Journal of Hydrometeorology*, 14, 977–988, <https://doi.org/10.1175/JHM-D-12-0146.1>, 2013.
- Brooks, S. P. and Gelman, A.: General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, 7, 434–455, <https://doi.org/10.1080/10618600.1998.10474787>, 1998.
- Chan, S., Bindlish, R., O’Neill, P., Jackson, T., Njoku, E., Dunbar, S., Chaubell, J., Piepmeier, J., Yueh, S., Entekhabi, D., Colliander, A., Chen, F., Cosh, M., Caldwell, T., Walker, J., Berg, A., McNairn, H., Thibeault, M., Martínez-Fernández, J., Uldall, F., Seyfried, M., Bosch, D., Starks, P., Collins, C. H., Prueger, J., van der Velde, R., Asanuma, J., Palecki, M., Small, E., Zreda, M., Calvet, J., Crow, W., and Kerr, Y.: Development and assessment of the SMAP enhanced passive soil moisture product, *Remote Sensing of Environment*, pp. –, <https://doi.org/10.1016/j.rse.2017.08.025>, 2017.
- Chan, S. K., Bindlish, R., O’Neill, P. E., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar, S., Piepmeier, J., Yueh, S., Entekhabi, D., Cosh, M. H., Caldwell, T., Walker, J., Wu, X., Berg, A., Rowlandson, T., Pacheco, A., McNairn, H., Thibeault, M., Martínez-Fernández, J., González-Zamora, ., Seyfried, M., Bosch, D., Starks, P., Goodrich, D., Prueger, J., Palecki, M., Small, E. E., Zreda, M., Calvet, J. C., Crow, W. T., and Kerr, Y.: Assessment of the SMAP Passive Soil Moisture Product, *IEEE Transactions on Geoscience and Remote Sensing*, 54, 4994–5007, <https://doi.org/10.1109/TGRS.2016.2561938>, 2016.
- Colliander, A.: SMAP/In Situ Core Validation Site Land Surface Parameters Match-Up Data, Version 1, NASA National Snow and Ice Data Center Distributed Active Archive Center, <https://doi.org/10.5067/DXAVIXLY18KM>, 2017.
- Colliander, A., Jackson, T., Bindlish, R., Chan, S., Das, N., Kim, S., Cosh, M., Dunbar, R., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., al Jassar, H., Lopez-Baeza, E., Martínez-Fernández, J., González-Zamora, A., Livingston, S., McNairn, H., Pacheco, A., Moghaddam, M., Montzka, C., Notarnicola, C., Niedrist, G., Pellarin, T., Prueger, J., Pulliainen, J., Rautiainen, K., Ramos, J., Seyfried, M., Starks, P., Su, Z., Zeng, Y., van der Velde, R., Thibeault, M., Dorigo, W., Vreugdenhil, M., Walker, J., Wu, X., Moneris, A., O’Neill, P., Entekhabi, D., Njoku, E., and Yueh, S.: Validation of SMAP surface soil moisture products with core validation sites, *Remote Sensing of Environment*, 191, 215 – 231, <https://doi.org/http://dx.doi.org/10.1016/j.rse.2017.01.021>, 2017.
- Diamond, H. J., Karl, T. R., Palecki, M. A., Baker, C. B., Bell, J. E., Leeper, R. D., Easterling, D. R., Lawrimore, J. H., Meyers, T. P., Helfert, M. R., Goodge, G., and Thorne, P. W.: U.S. Climate Reference Network after One Decade of Operations: Status and Assessment, *Bulletin of the American Meteorological Society*, 94, 485–498, <https://doi.org/10.1175/BAMS-D-12-00170.1>, 2013.
- Didan, K.: MYD13A2: MODIS/Aqua Vegetation Indices 16-Day L3 Global 1 km SIN Grid V006, <https://doi.org/10.5067/MODIS/MYD13A2.006>, <https://lpdaac.usgs.gov/>, accessed 30 September 2017, 2017.
- Doherty, J. and Welter, D.: A short exploration of structural noise, *Water Resources Research*, 46, n/a–n/a, <https://doi.org/10.1029/2009WR008377>, w05525, 2010.

- Dong, J., Crow, W. T., and Bindlish, R.: The Error Structure of the SMAP Single and Dual Channel Soil Moisture Retrievals, *Geophysical Research Letters*, 45, 758–765, <https://doi.org/10.1002/2017GL075656>, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL075656>, 2018.
- Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions, *Remote Sensing of Environment*, <https://doi.org/10.1016/j.rse.2017.07.001>, 2017.
- Entekhabi, D., Reichle, R., Koster, R., and Crow, W.: Performance Metrics for Soil Moisture Retrievals and Application Requirements, *J. Hydrometeor.*, 11, 832–840, 2010.
- Friedl, m., Menashe-Sulla, D., and MOSAPS SIPS: MCD12C1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG, NASA LP DAAC, <https://doi.org/10.5067/MODIS/MCD12C1.006>, accessed: 10 October 2017, 2017.
- Frost, C. and Thompson, S. G.: Correcting for regression dilution bias: comparison of methods for a single predictor variable, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 163, 173–189, <https://doi.org/10.1111/1467-985X.00164>, <https://rssl.onlinelibrary.wiley.com/doi/abs/10.1111/1467-985X.00164>, 2002.
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), *Journal of Climate*, 30, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>, 2017.
- Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S.: A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models, *The Annals of Applied Statistics*, 2, 1360–1383, 2008.
- Global Modeling and Assimilation Office: MERRA-2 tavg1_2d_lnd_nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Land Surface Diagnostics V5.12.4, Goddard Earth Sciences Data and Information Services Center, <https://doi.org/10.5067/RKPHT8KC1Y1T>, https://disc.gsfc.nasa.gov/datasets/M2T1NXLND_V5.12.4/, accessed on 14 September 2017, 2017.
- Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., and Wagner, W.: Recent advances in (soil moisture) triple collocation analysis, *International Journal of Applied Earth Observation and Geoinformation*, 45, 200 – 211, <https://doi.org/10.1016/j.jag.2015.09.002>, 2016.
- Harvey, A. C.: Estimating regression models with multiplicative heteroscedasticity, *Econometrica*, 44, 461–465, 1976.
- Hoffman, M. and Gelman, A.: The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo, *Journal of Machine Learning Research*, 15, 1351–1381, 2014.
- Konings, A. G., Piles, M., Das, N., and Entekhabi, D.: L-band vegetation optical depth and effective scattering albedo estimation from SMAP, *Remote Sensing of Environment*, 198, 460 – 470, <https://doi.org/http://dx.doi.org/10.1016/j.rse.2017.06.037>, 2017.
- Kornelsen, K. and Coulibaly, P.: Reducing multiplicative bias of satellite soil moisture retrievals, *Remote Sensing of Environment*, 165, 109 – 122, <https://doi.org/10.1016/j.rse.2015.04.031>, 2015.
- Kurum, M., Lang, R. H., O'Neill, P. E., Joseph, A. T., Jackson, T. J., and Cosh, M. H.: A First-Order Radiative Transfer Model for Microwave Radiometry of Forest Canopies at L-Band, *IEEE Transactions on Geoscience and Remote Sensing*, 49, 3167–3179, <https://doi.org/10.1109/TGRS.2010.2091139>, 2011.

- Lawrence, H., Wigneron, J.-P., Richaume, P., Novello, N., Grant, J., Mialon, A., Bitar, A. A., Merlin, O., Guyon, D., Leroux, D., Bircher, S., and Kerr, Y.: Comparison between SMOS Vegetation Optical Depth products and MODIS vegetation indices over crop zones of the USA, *Remote Sensing of Environment*, 140, 396 – 406, <https://doi.org/http://dx.doi.org/10.1016/j.rse.2013.07.021>, 2014.
- Loew, A. and Schlenz, F.: A dynamic approach for evaluating coarse scale satellite soil moisture products, *Hydrology and Earth System Sciences*, 15, 75–90, <https://doi.org/10.5194/hess-15-75-2011>, 2011.
- MacKay, D. J. C.: *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.
- Momen, M., Wood, J. D., Novick, K. A., Pangle, R., Pockman, W. T., McDowell, N. G., and Konings, A. G.: Interacting Effects of Leaf Water Potential and Biomass on Vegetation Optical Depth, *Journal of Geophysical Research: Biogeosciences*, 122, 3031–3046, <https://doi.org/10.1002/2017JG004145>, <http://dx.doi.org/10.1002/2017JG004145>, 2017JG004145, 2017.
- 10 O’Neill, P., Chan, S., Njoku, E., Jackson, T., and Bindlish, R.: SMAP Enhanced L2 Radiometer Half-Orbit 9 km EASE-Grid Soil Moisture, Version 1., NASA National Snow and Ice Data Center Distributed Active Archive Center, <https://doi.org/10.5067/CE0K6JS5WQMM>, https://nsidc.org/data/SPL2SMP_E, accessed on 25 August 2017, 2017.
- Palecki, M., Lawrimore, J., Leeper, R., Bell, J., Embler, S., and Casey, N.: U.S. Climate Reference Network Processed Data from USCRN Database Version 2, National Centers for Environmental Information, NESDIS, NOAA, U.S. Department of Commerce, 15 <https://doi.org/doi:10.7289/V5MS3QR9>, 2017.
- Parrens, M., Wigneron, J.-P., Richaume, P., Al Bitar, A., Mialon, A., Fernandez-Moran, R., Al-Yaari, A., O’Neill, P., and Kerr, Y.: Considering combined or separated roughness and vegetation effects in soil moisture retrievals, *International Journal of Applied Earth Observation and Geoinformation*, 55, 73 – 86, <https://doi.org/https://doi.org/10.1016/j.jag.2016.11.001>, <http://www.sciencedirect.com/science/article/pii/S0303243416301866>, 2017.
- 20 Patton, J. and Hornbuckle, B.: Initial Validation of SMOS Vegetation Optical Thickness in Iowa, *IEEE Geoscience and Remote Sensing Letters*, 10, 647–651, <https://doi.org/10.1109/LGRS.2012.2216498>, 2013.
- Roddell, M. and Beaudoin, H.: GLDAS Noah Land Surface Model L4 3 hourly 0.25 x 0.25 degree V2.1, NASA GESDISC DATA ARCHIVE, <https://doi.org/10.5067/E7TYRXPJKWOQ>, https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/GLDAS_NOAH025_3H.2.1/, accessed 30 September 2017, 2017.
- 25 Salvatier, J., Wiecki, T. V., and Fonnesbeck, C.: Probabilistic programming in Python using PyMC3, *PeerJ Computer Science*, 2, e55, <https://doi.org/10.7717/peerj-cs.55>, 2016.
- Schaefer, G. L., Cosh, M. H., and Jackson, T. J.: The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN), *Journal of Atmospheric and Oceanic Technology*, 24, 2073–2077, <https://doi.org/10.1175/2007JTECHA930.1>, <https://doi.org/10.1175/2007JTECHA930.1>, 2007.
- 30 Su, C.-H. and Ryu, D.: Multi-scale analysis of bias correction of soil moisture, *Hydrology and Earth System Sciences*, 19, 17–31, <https://doi.org/10.5194/hess-19-17-2015>, 2015.
- Su, C.-H., Ryu, D., Dorigo, W., Zwieback, S., Gruber, A., Albergel, C., Reichle, R. H., and Wagner, W.: Homogeneity of a global multi-satellite soil moisture climate data record, *Geophysical Research Letters*, 43, 11,245–11,252, <https://doi.org/10.1002/2016GL070458>, 2016GL070458, 2016.
- 35 Tomer, M., Moorman, T., and Rossi, C.: Assessment of the Iowa River’s South Fork watershed: Part 1. Water quality, *Journal of Soil and Water Conservation*, 63, 360–370, 2008.

- Xu, T., Valocchi, A. J., Ye, M., and Liang, F.: Quantifying model structural error: Efficient Bayesian calibration of a regional groundwater flow model using surrogates and a data-driven error model, *Water Resources Research*, 53, 4084–4105, <https://doi.org/10.1002/2016WR019831>, 2017.
- Yilmaz, M. T. and Crow, W. T.: The Optimality of Potential Rescaling Approaches in Land Data Assimilation, *Journal of Hydrometeorology*, 5 14, 650–660, <https://doi.org/10.1175/JHM-D-12-052.1>, 2013.
- Zwieback, S., Scipal, K., Dorigo, W., and Wagner, W.: Structural and statistical properties of the collocation technique for error characterization, *Nonlin. Processes Geophys.*, 19, 69–80, 2012.
- Zwieback, S., Su, C.-H., Gruber, A., Dorigo, W. A., and Wagner, W.: The Impact of Quadratic Nonlinear Relations between Soil Moisture Products on Uncertainty Estimates from Triple Collocation Analysis and Two Quadratic Extensions, *Journal of Hydrometeorology*, 17, 10 1725–1743, <https://doi.org/10.1175/JHM-D-15-0213.1>, 2016.