Response to Anonymous Referee #1's interactive comment on "Multi-scale analysis of bias correction of soil moisture"

[R1] This is an excellent paper that makes a fundamental contribution to soil moisture time series analysis. In particular – it highlights the (temporal) scale dependence of relative multiplicative bias in modeled, in situ and remotely-sensed soil moisture data sets. This is a wholly new insight which has very important consequences for a number of important data assimilation and merging applications. I strongly recommend publication following minor revisions.

We thank the Referee #1 for the careful examination and positive endorsement of our manuscript. We take this opportunity to consider their comments (copied here and identified by [R1]) and improve the clarity of this work. Our responses are identified by [A] while extracts of the specific changes made in the revised manuscript are shown in quotations and in blue.

[R1] Below are some specific points to consider prior to publication. They are all minor suggestions except I did have problems following the motivation for Section 7 (see points 6C and 7 below). In addition, I believe there is a loose end involving additive biases which requires further clarification (point 2).

[A] Please see below for our specific responses.

[R1] 1) One advantage of applying time series analysis is that you can assume stationarity (at least in the weak sense that soil moisture expectations will no longer vary across the seasonal cycle). From this point of view, the seasonal cycle at a point is a deterministic feature that must be accounted for before random time series models can be applied to soil moisture time. For this reason, hydrologists often view seasonal dynamics as a unique time scale – and not simply just another time scale in a spectral range. Given that dealing with non-stationarity is a strength of wavelet analysis, the authors might want to discuss the implications of seasonality on their analysis. Instead of avoiding the issue by removing seasonality…it seems like the authors are addressing it head on (with a statistic tool that explicitly addresses seasonality). This seems like a step forward which might warrant a little more discussion.

[A] In our treatment, we use DWT to decompose a complex timeseries into components with variability at different time scales. All the individual components from different data sources can be compared to resolve for multiplicative bias and additive noise that account for their differences. For completeness, no individual component, e.g., seasonal cycle, has been or should be singled out for omission, but we can analyse individual components separately. Note also that the implicit assumption here is that different data sources observe the same seasonal cycle up to some multiplicative factor and noise. The same assumption applies to all other components. That is, from the viewpoint to formalise the inter-data comparisons, the signal component *f* can be taken as deterministic such that *f* is common to all data sources, but the overall data *p* is stochastic, and these extends to *f* and *p* representations at different time scales *j* [See also our response to Simon Zwieback's comments]. Lastly, because the detail timeseries at given scale *j* does not contain variations at scales >*j*, that timeseries can be better satisfy the weak-sense stationarity condition of matching mean and covariance. We revised the text to reflect these,

"…However in this work, we consider only variability across j and assume stationarity at each scale.
Pearson's linear correlation R and variance analyses (see Appendix A) are performed on
the Kyeamba's INS, AMS and MER SM (as p in Eq. 2) detail (p_j) and approximation ($p^{(a)}_j$) timeseries in Fig. 3. **The strength of MRA is that since the detail timeseries p_j at a given scale j does not contain variations at time scales > j, the weak-sense stationarity conditions can be better met.**

…

The interpretation of the discrepancies between X and Y can vary depending on the time period of the data and analysis, and the adopted signal/noise model. By using entire record of INS, AMS and MER data in MRA, the MS model does not observe time-varying additive bias (e.g., from using the moving-window approach (Su et al., 2014a)) and autocorrelated errors (from using lagged covariance (Zwieback et al., 2013)). **Rather, MRA and the MS model enable a description of the systematic differences based wholly in terms of multiplicative biases at individual time scales, and the random differences in terms of additive noise**. Specifically, this contrasts with the short time-window approach (≤ 32d), where multiplicative bias existing at coarse scales (e.g., $p^{(a)}_6$) will manifest as both time-varying additive and multiplicative biases."

[R1] 2) The other advantage of dealing with anomalies is that you can assume biases are wholly multiplicative and lack an additive component. How are additive biases accounted for in (2-3)? Are they just passed along to the approximation time series at the coarsest scale? This issue is addressed in Section 5 (lines 15-20 on page 9005) via the explicit removal of biases but it also seems relevant to results presented in Section 4.

[A] To discuss this point, we first distinguish between (1) an overall additive bias given by E(*X*)-E(*Y*), where the expectation value is taken over the entire timeseries, and (2) time-varying additive biases given by $E_t(X)$- $E_t(Y)$, where the expectation value is taken over some time-window located at some time *t*.

For (1) the overall additive bias, the reviewer is correct – the overall means E(*X*) and E(*Y*) are equivalent to the overall means of their coarse approximation timeseries $E(X_j^{(a)})$ and $E(Y_j^{(a)})$, respectively. This is now stated in text immediately after Eq. 1: $E(p_j^{(a)}) = E(p) = p_J^{(a)}(t) = \mu_p$

For (2) the time-varying additive biases, our model assumes that such biases arise from multiplicative bias present at coarser time scales. Therefore our model detects only for the presence of an overall additive bias, multiplicative biases at individual scales, and additive noise. This contrasts with the short moving time-window approach where the multiplicative biases existing at coarser scales will manifest as both time-varying additive and multiplicative biases. In other words, the perception and interpretation of the discrepancies between two data can vary depending on the time period of the analysis and the adopted signal/noise model.

It is also of note that the strength of wavelet analysis is decomposing a timeseries of no mean (additive) bias into multiple (with different frequency) timeseries with no mean additive bias, but only with multiplicative bias. At a given scale *j*, because the detail timeseries p_j does not contain variations of time scale >*j*, the weak-sense stationarity conditions for TC analysis with long timeseries can be better satisfied. To clarify these point, we revised the manuscript as follows,

"…multi-scale (MS) model that distinguishes the signal components of the two data *X* and *Y* **via an overall additive bias and a set of positive scaling coefficients $\alpha_{p,j}$, α'_p**, and assumes an additive and zero-mean independent but non-white noise model $\varepsilon_p(t)$. **Focusing on the zero-mean signal and noise components, the model reads**…

[Eq. 4 and 5]

…

The interpretation of the discrepancies between *X* and *Y* can vary depending on the time period of the data and analysis, and the adopted signal/noise model. By using entire record of INS, AMS and MER data, the MS linear model does not observe time-varying additive bias (from using the moving-window approach (Su et al., 2014a)) and autocorrelated errors (from using lagged covariance (Zwieback et al., 2013)). Rather, MRA and the MS model enable a description of the systematic differences based wholly in terms of multiplicative biases at individual time scales, and the random differences in terms of additive noise. **Specifically, this contrasts with the short time-window approach (≤ 32d), where multiplicative bias existing at coarse scales (e.g., $p^{(a)}_6$) will manifest as both time-varying additive and multiplicative biases."**

[R1] 3) The reference to "Fig. 2" right at the start of Section 3 does not seem consistent with the Figure 2 contained in the manuscript.

[A] The Kyeamba SM timeseries from INS, AMS and MER are shown by *p* (blue curves) in Figure 2. To clarify this, we revised the reference and the caption. In particular, the caption of Figure 2 now reads,

"Figure 2. MRA of INS, AMS and MER SM at Kyeamba. *p* denotes the original timeseries, p_j the detail timeseries, and $p_j^{(a)}$ the approximation timeseries. Grey shadings are > 5 day data gaps, red dots superimposed in $p_6^{(a)}$ are monthly means of *p*, and magenta lines are trend lines fitted to $p_8^{(a)}$.**"**

[R1] 4) Superscript "(a)" in (1) is not defined at first use.

[A] To clarify the use of the superscript to distinguish approximated representations from detail time series, we revised the text as follows. In particular, we added further clarification to the recursion chain and multi-resolution analysis.

"The 1-D orthogonal discrete wavelet transform (DWT) enables MRA of a timeseries *p(t)* of dyadic length $N=2^J$ and a regular sampling interval Δt by providing the mechanism to go from one resolution to another via a recursive function

$$p_{j-1}^{(a)}(t) = p_j^{(a)}(t) + p_j(t),$$

with an expectation value $E(p_{j-1}^{(a)}) = E(p) = p_J^{(a)}(t) = \mu_p$ and $E(p_j) = 0$, **where the superscript *(a)* labels approximated representations.** The integer *j=[1,J]* labels the scale of analysis with *j=1 (J)* denoting the finest (coarsest) scale, and serves to define a spectral range in a spectral analysis. The recursion therefore relates an approximation or coarse representation $p_j^{(a)}(t)$ of the signal at one resolution to that at a higher resolution $p_{j-1}^{(a)}(t)$ by adding some fine-scale detail denoted by p_j. **The end of the recursion chain leads to reconstruction of the original time series such that** $p_0^{(a)}(t) = p(t)$**, …"**

[R1] 5) Figure 3 – clarify difference between (a) and (b) in caption (hard to see small difference in super-script)

[A] We agree. The caption is revised to include "…(a) compares the correlation between their detail timeseries p_j, and (b) compares between their approximation timeseries $p_j^{(a)}$ …".

[R1] 6) Figure 5 contains a lot of information…a couple of things I struggled with when interpreting it:
A) In column (a) the "target" is the TC-based results (correct)…but in column (b) the target is unity? That change makes it a little difficult to read the figure horizontally.
Maybe break-out column (a) into another figure?

[A] We believe that the referee has misinterpreted the results (see also the next comment). To simplify our explanation, we focus on the AMS results. Figure 5a (top panel) shows the scaling factor $\alpha_{Y,j}$ of AMS (as *Y* as per model in Eqs. 4-6) with respect to INS (as *X*) BEFORE bias correction was applied to AMS to match INS. In choosing INS as the reference data *X*, we let $\alpha_{X,j}$ =1 (see text after Eq. 6). Figure 5a reveals that $\alpha_{Y,j} \neq \alpha_{X,j}$ (i.e., $\alpha_{Y,j} \neq 1$), indicating that there are multiplicative biases in *Y* (see also Eq. 6). In other words, $\alpha_{Y,j} \neq 1$ is a diagnosis for multiplicative bias. Within our MRA and MS linear model framework, our aim of the bias correction is to ensure $\alpha_{Y*,j}$ =1 AFTER applying a correction scheme, where *Y denotes the corrected data.**
Figure 5b-f show 5 different correction schemes, where the values of $\alpha_{Y*,j}$ after correction are being diagnosed, and most found not to produce $\alpha_{Y*,j}$ =1, except for the MS scheme.
To avoid causing similar confusion amongst readers, we have now revised Figure 5 and its caption. The text has also revised, please refer to the extracts in our next two responses.

[R1] B) "OLS" and "TC" can also be rescaling strategies: : :so it took me awhile to realize the color/symbols refer to strategies for calculating alpha AFTER various re-scaling strategies (listed horizontally along the top of the graphs) have been applied. Is that the correct interpretation of Figure 5? If so - is it really necessary to show the OLS results in each column? We already know they are biased by noise..you can already see that in column (a)?

[A] Only TC was used to perform bulk linear and A/S linear rescaling. The rescaling coefficients for their implementations are listed in the subfigures Figure 5b and d. For the diagnosis of multiplicative bias by looking at $\alpha_{Y,j}$ before and $\alpha_{Y*,j}$ after bias correction, we use TC for *j*>2 and OLS for *j*≤2. OLS was used because TC could not be conducted at those scales due to negative covariance.
We take on the Referee's advice and have revised Figure 5 to remove OLS plot-line, but noted in the caption that OLS was used only for diagnoses:
"

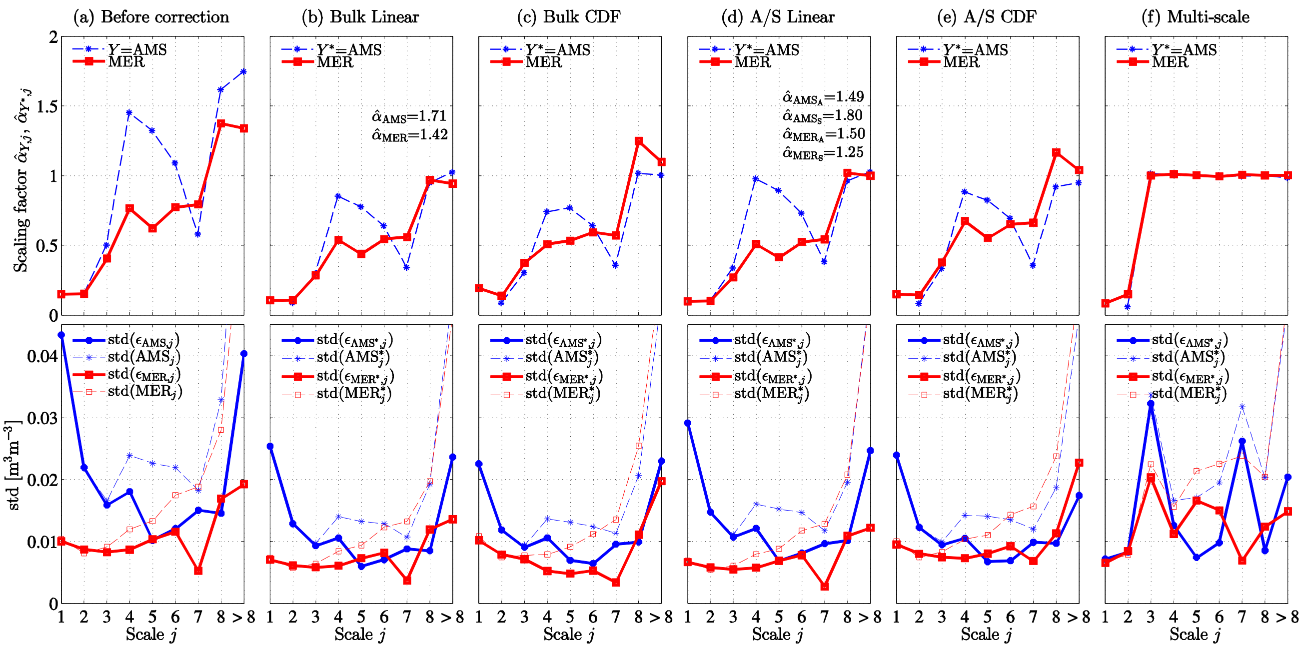

Figure 5. Bias correction of AMS and MER (as Y) with respect to INS (as X), showing the impact of 5 correction schemes on the scaling coefficients, noise and total std at individual scales. Estimated $\hat{\alpha}_{Y,j} \neq 1$ or $\hat{\alpha}_{Y*,j} \neq 1$ suggests multiplicative bias in Y_j or $Y*_j$ as per Eq. 6. (a) is the diagnosis of *Y* before correction, and (b–f) are that of *Y** after correction. The estimated $\hat{\alpha}_{Y,j}$ and $\hat{\alpha}_{Y*,j}$ for the diagnoses are derived using OLS (for *j* = 1,2) and TC (*j* > 2). The additional $\hat{\alpha}_{Y,j}$ values listed in (b, d) are the scaling coefficients used in the implementations of bulk and A/S linear rescaling. Scale j > 8 corresponds to $Y_8^{(a)}$."

[R1] C) Page 9009. Last paragraph. I don't follow where …but we also observed noise amplification in AMS at j=3,7..." Is shown. In Figure 5f (top row)? This seems like a key point but it could be tied better to the results in the figures. Does "alpha_Y,j < 1" refer to the "OLS" results in column (f)? If so, doesn't that just indicate the short-coming of OLS as an estimator of post-rescaling alpha's and NOT an indication that the alpha's have been poorly scaled? The author's should consider re-writing this paragraph for increase clarity.

[A] This comment by the referee is closely related to the last two. Following the referee's advice, we revised the text to make clear the following points,

- $\alpha_{Y,j}$ and $\alpha_{Y*,j}$ serve as a diagnostics for presence of multiplicative biases

- $\alpha_{Y,j}$ values in Figure 5a are used in the implementation of MS rescaling

-OLS is only used as a guide for estimate $\alpha_{Y,j}$ and $\alpha_{Y*,j}$ for the above diagnostic purposes when TC estimation could not be conducted.

"For illustrations, we correct the biases in AMS and MER SM with respect to INS SM at Kyeamba using the above five schemes. Using the above notations, AMS and MER are treated as *Y*, the corrected AMS* and MER* as *Y**, and INS as *X*. **MRA-TC was applied to observe their consequences in Fig. 5. In the upper panel, estimated $\hat{\alpha}_{Y,j}$ and $\hat{\alpha}_{Y*,j}$ values provide diagnostics for detecting the presence of multiplicative biases before and after application of the correction schemes. The lower panel plots the std of Y_j and Y*j and their associated errors $\varepsilon_{Y,j}$ and $\varepsilon_{Y*,j}$. The values of the scaling coefficients $\alpha_{Y,j}$ (before correction) and $\alpha_{Y*,j}$ (after), and the errors std($\varepsilon_{Y,j}$) and std($\varepsilon_{Y*,j}$) were estimated using TC. But where TC estimates could not be retrieved (for *j* = 1–2) due to negative correlation amongst the data triplet (e.g., resulting from significant noise and weak instrument), OLS-derived (under) estimates serve as a guide for the above diagnostic purposes.** Similarly the total std is a guide for error std in these cases.

…

By construction, **the MS rescaling uses the estimated $\hat{\alpha}_{Y,j}$ values from Fig. 5a to correct bias at all the scales. Fig. 5f shows the analysis of MS-corrected *Y**. The equivalence $\hat{\alpha}_{Y*,j}$ = 1 indicates that the multiplicative biases are eliminated at *j* > 2. At *j* = 1−2, as the scaling coefficients cannot be estimated by TC, CDF matching was applied to these scales such that the biases are still present at these scales.** Amid the reduction of biases, we also observed noise amplification in AMS at *j* = 3,7 and in MER at *j* = 3–7 **because of rescaling with less-than-unity $\hat{\alpha}_{Y,j}$ values in Eq. 10**. Indeed it is evident from Eq. (6) that it is possible to increase the noise variance and MSE when reducing the bias component of the MSE. This in turn leads to larger disagreement between INS and AMS in terms of RMSD and R, and the increased amplitudes of the noise observed in AMS in Fig. 6f."

In addition, we have also revised Figure 5 and its caption, see last comment.

[R1] 7) Page 9010. I don't quite follow the rationale for linking bias correction and noise correction here. I suspect that my problem is linked to something I missed in Figure 5 (see specific points above…especially point 6C). As a result, while Section 7 is interesting (and seems like a very nice extension of the MRA-based approach presented here), it does not seem tightly linked to the rescaling focus of the paper. However – as noted above – this might be due to my miss-interpretation of Figure 5. I'd recommend that the author's rewrite/re-clarify this connection for future readers of the paper.

[A] We have shown that we can correct multiplicative biases at every scale *j* in MS rescaling (or at two distinct scales of variation in A/S rescaling). However this can lead to amplification of the noise (and also the signal component) in *Y* when the rescaling coefficient(s) $\hat{\alpha}_{Y,j}$ is valued less than 1: $std(\varepsilon_{Y*,j}) > std(\varepsilon_{Y,j})$ where *Y* is the data before correction, and *Y is the data after correction. This contrasts with cases where $\hat{\alpha}_{Y,j}$ > 1, which leads to suppression of noise. This may be considered undesirable for an objective to produce more physically representative data with a simple error (or noise) structure on the whole. It is from this viewpoint that we argue that the task of bias correction**

cannot be separated from that of noise reduction. We have revised the text (see last comment) to better describe the problem of noise amplification.

"On the other hand, the A/S-based and MS methods can modify the original error profiles in the data across the scales, by amplifying (or suppressing) errors in individual components (either Y_j , Y_S, or Y_A) with less-than (greater-than) unity pre-correction α's. **This may be considered undesirable for an objective to produce more physically representative data with a simple error structure on the whole.** Therefore arguably, none of these methods is entirely satisfactory, in manners of not removing the multiplicative biases completely and/or changing error characteristics. **From this viewpoint,** the task of bias correction is seen as inseparable from that of noise reduction when considering MS (or A/S) bias correction, unless certain components in MRA were explicitly ignored.

…

The last example presents an impetus to consider noise removal prior to bias correction **and produce a simpler error structure in the bias corrected data $Y*$.**"

[R1] 8) Section 8. Regarding the potential impact of this work, I'd argue that the authors could be a little more assertive. For instance, it seems likely that the scale dependence of multiplicative biases explains the VERY poor (i.e. negative variance!) TC results that Draper et al. (2013) [RSE, "Estimating root-mean-square errors in remotely sensed soil Moisture…"] notes when applying TC to a raw (as opposed to climatological-anomaly) soil moisture time. Also, Yilmaz and Crow (2013) [already cited in paper] demonstrated the link between poor rescaling and errors in sequential data assimilation. Residual multiplicative bias (at any time scale) will cause filter innovations (i.e., back-ground minus observation) to contain residual signal (i.e., leaked signal). Leaked signal = auto-correlated innovations = sub-optimal filter performance. This is all admittedly a little bit bit-speculative but I would recommend that the author's be a bit more proactive about articulating the potential positive impact of this work. This is NOT a meaningless exercise in statistical estimation and it would be a shame if it was interpreted as such.

[A] The referee has highlighted two important observations. First, scale dependence of multiplicative bias (as observed in Figure 5a) can diminish correlation between data. We can observe in Figure 3b (the correlations between approximation timeseries) that as we include more components to the reconstruction, the correlation reduces significantly. On one hand, more noise is added to the reconstructed and noise suppresses correlation. On the other hand, this may be due to adding components with different multiplicative biases.

Second, we agree with the observation that the residual signal due to sub-optimal bias correction can impact filter performance in data assimilation, as illustrated by Yilmaz and Crow (2013). This also highlights the fact why we choose matching the statistics of the signal components in X and Y as the goal of a bias correction scheme.

We revised the manuscript with the following text,

"…on the other hand, stronger AMS-MER correlations at coarsest (temporal) scales and their mesoscale spatial resolutions would indicate lesser representativeness of in situ measurement at these spatio-temporal scales. **Furthermore, we observe that $R(p^{(a)}_j, q^{(a)}_j)$ reduces with decreasing j, as more components are added to the reconstruction of $p^{(a)}_j$ and $q^{(a)}_j$. The inclusion of noisy AMS_1 to the makeup of AMS leads to a drop in R(INS,AMS) and R(AMS,MER). Aside from including more noise to the approximation timeseries, adding components with different multiplicative biases (more later in Section. 6) can also diminish the correlations. The scale-dependence of multiplicative biases and added noise can contribute to the contrasting results of applying TC to raw versus anomaly SM timeseries in Draper et al. (2013). In particular, given the presence of noise in p_j for $j \geq 7$, error analysis of the anomaly SM (i.e., in p_j for $j \leq 6$) will under-estimate the total error in the raw data p.**

…

Here we define our optimality criterion based on the first criterion of matching the first two moments of the signal components in X and Y so that $Y*$ is suitable for bias-free data assimilation. **In particular, Yilmaz and Crow (2013) have shown that residual multiplicative biases due to sub-optimal bias correction scheme will cause filter innovations to contain residual signal and sub-optimal filter performance. Thus within the paradigm of the MS model, the goal of bias correction is to minimize the difference $|\alpha_{Y*j} - 1|$ for $\alpha_{X*j} = 1$, so that the multiplicative bias terms in Eq. 6 are eliminated.**"