Response to reviewers

Line numbers mentioned in this reply refer to our clean version of the revised manuscript.

Reviewer #2

Comments:

This work extended the C-band data set to the previously missing period by rescaling the QSCAT Ku-band dataset during 2001-2007. Data-rescaling was used to unify the backscatter values from different sensors and then the machine learning method was used to address the monthly values differences. This is a quite useful dataset for further detecting forest structure and resilience dynamics. I have some minor comments as below:

Response: Thank you greatly for reviewing our manuscript. We apologize for the late reply due to the covid situation in China. We have carefully considered the comments and revised the manuscript accordingly. The suggestion of considering the overfitting issue is especially useful. Thank you!

To compare the linear regression, CDF and new data rescaling method, the author should compare their performances at global scale, i.e. a map showing the pearson r and RMSE pixel by pixel.

Response: Before replying to this comment, we would like to mention that, thanks to a comment of Referee #3, we now avoid calling our data rescaling method a "new method", as similar approaches have been used by previous research (Brocca et al. 2010 & 2013).

Thank you for the suggestion of mapping the Pearson r and RMSE pixel by pixel. We very much appreciate this suggestion but refrain us from showing such a map. Seen from Fig. R2 below, the Pearson r, RMSE, and rRMSE by CDF and linear correction could be even higher than that obtained with the “new scaling method”. However, the scaled signals by CDF and linear correction are obviously less satisfying. Thus, we believe Pearson r and RMSE can be misleading here.

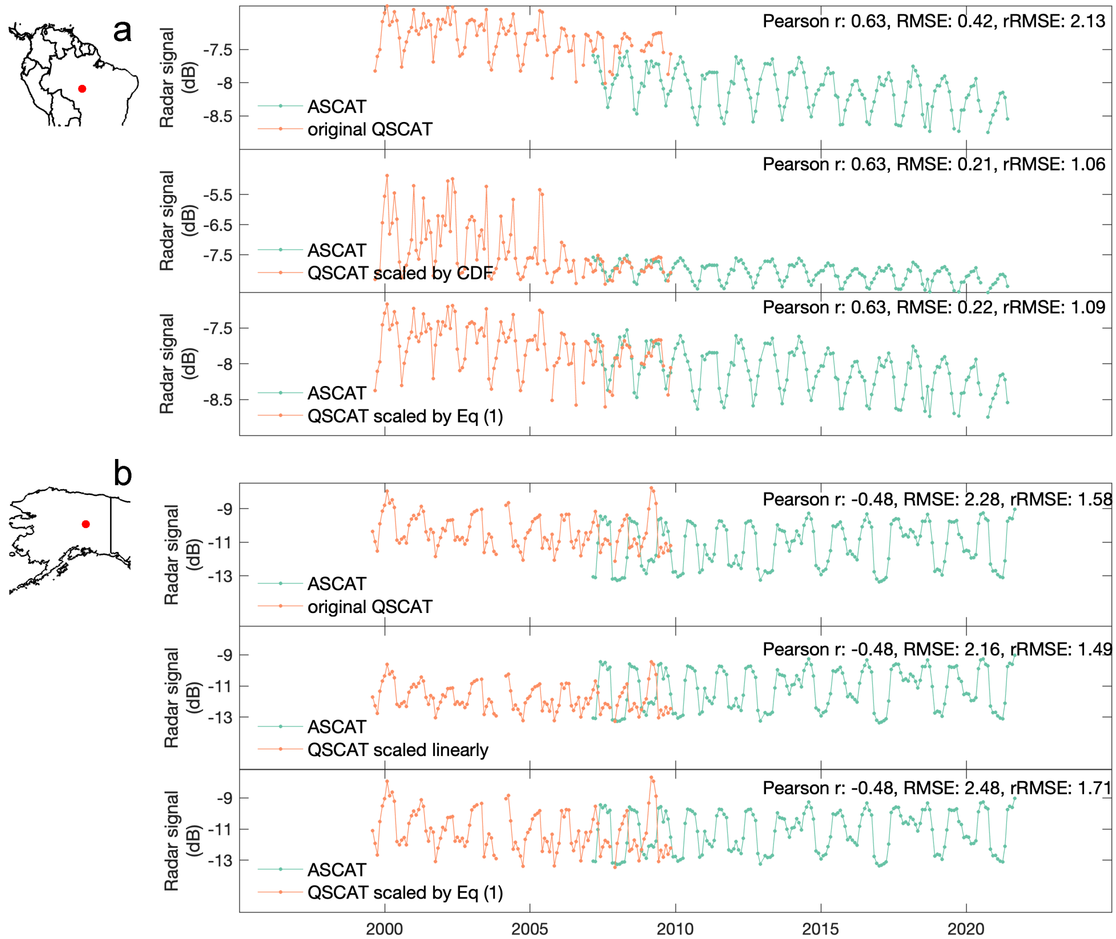

Fig. R2. Same as Fig. 3 in the main text but showing the Pearson r, RMSE, and rRMSE between ASCAT and QSCAT signals in the overlapping period.

In fact, during our calculation, we visually inspected the rescaling results of the three methods for a large number of pixels (every 100 of all the pixels). We found that at the global scale, the three methods performed almost equally well in most pixels (Fig. R1 in responses to Referee #1, also shown as Fig. S1 in the revised manuscript), but linear regression and CDF created very unnatural results for pixels with a strong signal trend or sudden changes in signal. We therefore focused on these two particular kinds of pixels in Fig. 3. This point has been made clearer in the legend of Fig. 3 (lines 716-720).

Similar for Fig 4, the author can show the spatial map of the performance of scaled Ku-band and corrected Ku-band pixel by pixel.

Response: The maps suggested by Referee #2 have actually been shown as Figs. 6 (Pearson r-based assessment), 7 (RMSE-based assessment) & S2 (rRMSE-based assessment). Per your suggestion of checking the overfitting issue below, all these figures have been updated.

It seems that such rescaling method can also apply to other merging tasks. Can you discuss a bit of its potential usage to benefit the big data environmental science field?

Response: As above-mentioned, we now avoid calling our rescaling method a totally new method, because Referee #3 has pointed out that similar approaches have been used by previous research (Brocca et al. 2010 & 2013). However, we followed your suggestion to discuss more the potential usage of the data rescaling method in earth science studies (lines 379-384).

The author could include a table mentioning the specific information of available microwave dataset, i.e. their time and spatial coverage, time and spatial resolution, etc, to prove the uniqueness of constructing the time series over non-overlapped period with QSCAT Ku-band data.

Response: Indeed! Thank you for this very useful suggestion. We now added such a table (Table 1).

As you used the decision tree regression, have you checked whether the over-fit issue exist or not?

Response: Thank you for reminding us of this very important issue. Previously we used the 'fitrtree' function in Matlab without tuning the parameters (i.e., default value of 1 for 'MinLeafSize'). A small value for 'MinLeafSize' means a deep tree, and vice versa. Thus, overfitting could indeed occur due to a small value of 'MinLeafSize'.

Per your suggestion, we now use cross-validation to find the best 'MinLeafSize' value. Cross-validation is a suggested approach by Matlab to overcome the overfitting issue (https://ww2.mathworks.cn/help/stats/improving-classification-trees-and-regression-trees.html), and has been used by previous research (Sankaran et al., 2005; Pekel 2020). We used five-fold cross validation as there are only ~60 overlapping monthly observations between QSCAT and ERS/ASCAT, but we verified that the results were not changed if 10-fold was used. This point has been stated in lines 258-260. Meanwhile, we also followed a suggestion of Referee #3 to include all three climatic variables into one regression tree (instead of building the regression trees separately).

We tuned the parameter pixel by pixel. Taking one pixel in the Tibetan Plateau as an example (Fig. R3 below), the cross-validated errors decrease initially with the increase of 'MinLeafSize', reach its minima when 'MinLeafSize' is around 6, then increase sharply. Previously the depth of the regression tree is 6, but now after cross-validation, the depth becomes 4 (Fig. R4). Fig. R5 further shows the C-band and Ku-band signals before and after signal correction: the Ku-band signals corrected by the optimal tree showed highly similar dynamics with the C-band signals, with a r value of 0.9 (Fig. R5c), and this accuracy is only slightly lower than that created by the "default tree" (0.93, Fig. R5b).

After addressing this comment, all results have been updated, and all related figures (Figs 4-9, S2, S3) have been redrawn. Encouragingly, the new results are highly similar to the old ones, suggesting that the over-fitting issue is not severe in the previous results. We thank Referee #2 once again for this important suggestion!

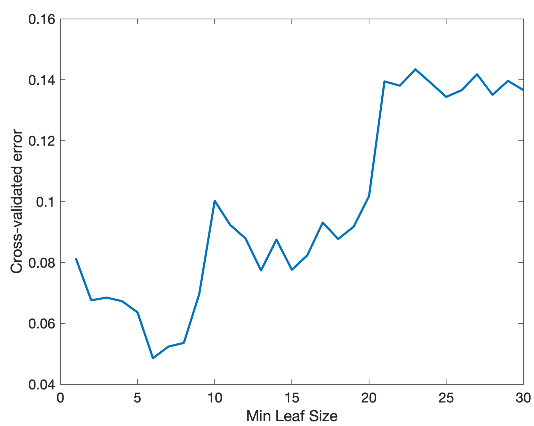

Fig. R3. 'MinLeafSize' parameter *vs* cross-validated errors.

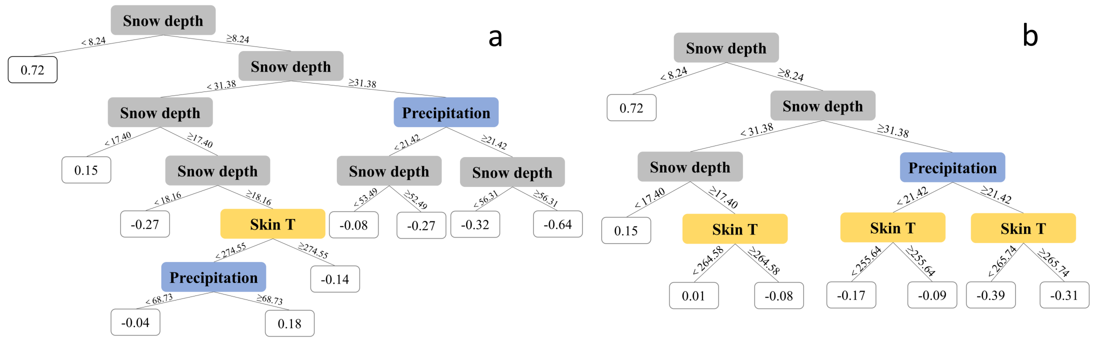

Fig. R4. Comparison between regression tree with (a) default parameters and (b) with optimal 'MinLeafSize' parameter.

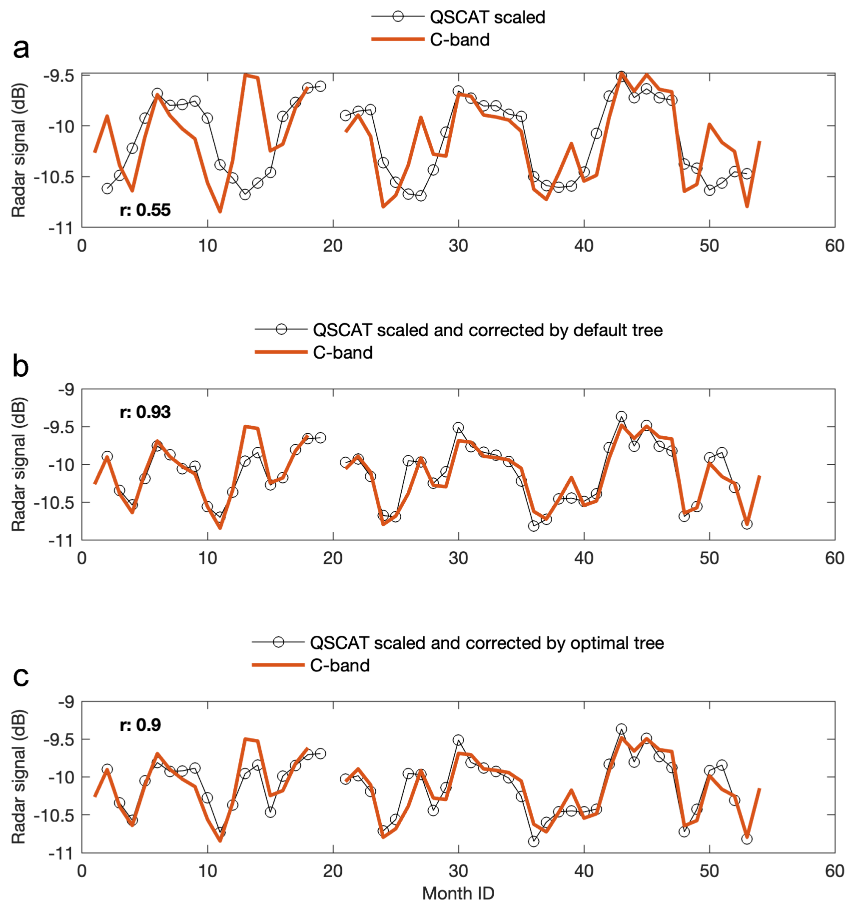

Fig. R5. Performances of the decision tree modelling for correcting the signal differences between C- and Ku- band signals. (a) shows the C-band signal and the scaled Ku-band signal before correction. (b) shows the C-band and Ku-band signal corrected by the decision tree with default parameters. (c) shows the C-band and Ku-band signals corrected by the decision tree with optimal "MinLeafSize" parameter.

For Fig 8, there is large overlap between type 1 and type 2 pixels. If the author just compared the corresponding pearson r values between corrected Ku-band and C-band values to find the appropriate regressor, the type of each pixel can be determined. Why are some pixels assigned by two type?

Response: We apology for this misleading figure. This figure was drawn in GIS and each pixel was shown as a point. Points in GIS have a size; thus their locations appear "overlapping" but actually they do not. We have redrawn this figure (Fig. 8) into a raster map to avoid this misunderstanding.

We hope the revision has addressed all your concerns! We thank you once again for the very helpful comments!

References

Pekel E. Estimation of soil moisture using decision tree regression[J]. Theoretical and Applied Climatology, 2020, 139(3): 1111-1119.

Brocca L, Melone F, Moramarco T, et al. Scaling and filtering approaches for the use of satellite soil moisture observations[J]. Remote Sensing of Energy Fluxes and Soil Moisture Content, 2013, 411: 426.

Brocca L, Melone F, Moramarco T, et al. Improving runoff prediction through the assimilation of the ASCAT soil moisture product[J]. Hydrology and Earth System Sciences, 2010, 14(10): 1881-1893.

Sankaran, M., Hanan, N. P., Scholes, R. J., Ratnam, J., Augustine, D. J., Cade, B. S., Gignoux, J., Higgins, S. I., Le Roux, X., and Ludwig, F.: Determinants of woody cover in African savannas, Nature, 438, 846-849, 2005.