Response to comments

Paper #: essd-2023-220

Title: A coarse pixel scale ground "truth" dataset based on the global in situ site measurements to support validation and bias correction of satellite surface albedo products

Journal: Earth System Science Data

We really appreciate the rigorous attitude of the reviewer for providing so many valuable suggestions. We revised the paper carefully and tried to give satisfactory answers to the reviewers' questions. The corresponding modifications are highlighted in red font in the revised paper.

First, we have added the diagram showing the upscaling and evaluation process.

Second, we have emphasized the intention and objective of generating such a pixel scale ground albedo "truth" dataset, and further explained the uncertainty of the upscaling model.

Third, we have plotted the distribution of the accuracy indicators for different networks and rephrased the sentence about the performance of BSRN network.

Fourth, we have explained that only the median values of the boxplots were focused due to the even sample sizes for each level of spatial heterogeneity.

Fifth, the reason why the results (percentages) in Figure 11 is quite different in the revised manuscript than those in the previous manuscript was given in this manuscript: the addition of new in situ sites and the the different data sources of the reference data.

Reviewer #1

I would still suggest adding the diagram showing the upscaling and evaluation process but avoiding plagiarism.

Re: We sincerely appreciate your rigorous science attitude. As you suggested, we have added and improved the diagram by adding the procedure of evaluation process of the upscaling model and the generation of the pixel scale reference albedo dataset based on the upscaling model and *in situ* site measurements in the revised manuscript. The revised diagram is as follows:

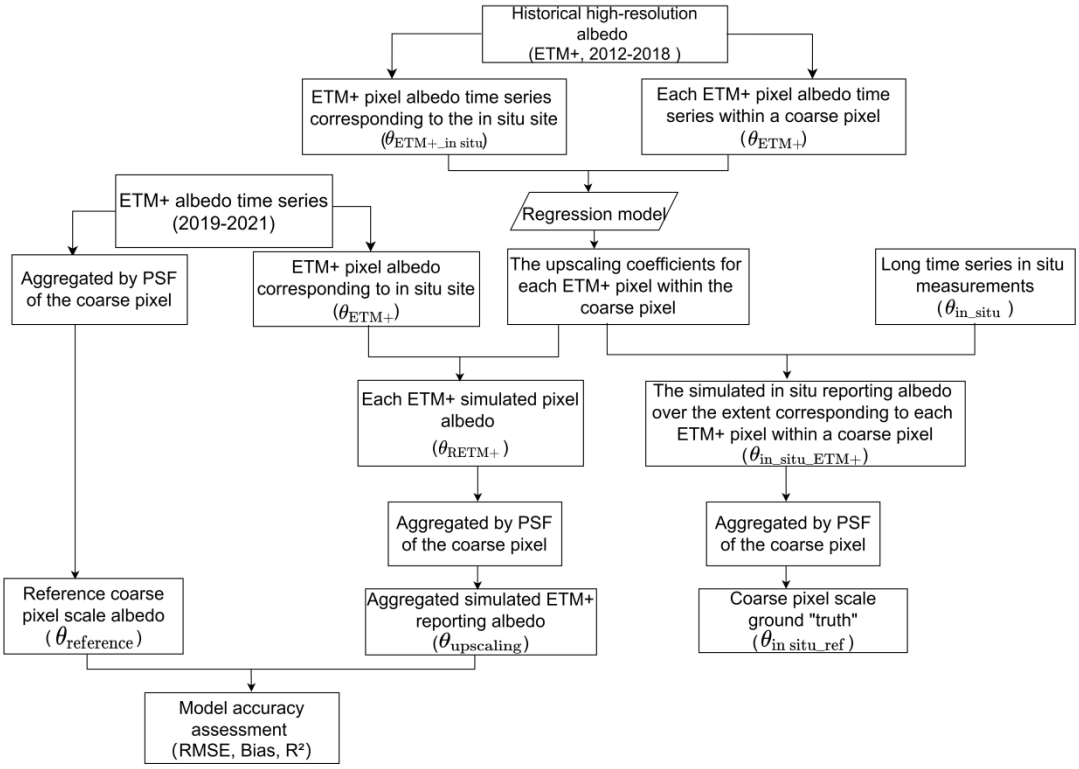

Figure R1: The flowchart of generating coarse pixel scale ground "truth" based on upscaling model.

The method part should emphasize the improvements compared with the published methods on developing a global pixel scale ground "truth" dataset.

Re: In fact, this paper is the continuation and deepening of our previous work. The upscaling method was proposed in our previous research, but the effectiveness of this upscaling method on the global scale was still unknown. Furthermore, this method has not been utilized for individual *in situ* site measurements from sparsely globally distributed observation networks (e.g., SURFRAD, BSRN, and Fluxnet). Under this background, this study aims to first comprehensively evaluate the effectiveness of upscaling methods on the global scale, and then apply this upscaling method to the 416 in situ sites over the globe. Finally, a pixel scale ground "truth" dataset was provided for validation, bias correction, and other applications that need the linkage between *in situ* measurements and satellite pixels.

Besides, I recommend the author delve deeper into the previous comments 'the uncertainty of the upscaled maps as similar to any other product and it is questionable its utility as a reference ground-truth'.

Re: Thank you for your insightful comments. It's important to note that the accuracy (between 0.03 and 0.05) of current coarse-resolution surface albedo satellite products was generally assessed over relatively homogeneous land surfaces. But their accuracy over heterogeneous are still unknown, because the effect of scale mismatch between *in situ* measurements and satellite pixel cannot be ignored but not resolved. The scale mismatch is still the key challenge over heterogeneous surface. And this is original intention of generating such a pixel scale ground truth dataset. Moreover, we would like to point out that the accuracy of the pixel scale ground truth cannot be determined through the comparison with other products since the products themselves contain errors. In fact, the advantage of the pixel scale ground truth was proved

through the comparison with in situ single site measurements in terms of their agreements with a coarse pixel scale albedo value. Therefore, although the pixel scale ground truth dataset is not the absolute truth due to its own uncertainty, it shows an advantage over single *in situ* sites when matched with satellite pixel.

Reviewer #2

2nd Review of "A coarse pixel scale ground "truth" dataset based on the global in situ site measurements to support validation and bias correction of satellite.

2nd Review of "A coarse pixel scale ground "truth" dataset based on the global in situ site measurements to support validation and bias correction of satellite surface albedo products" by Fei Pan et al.

The new manuscript has been revised greatly. Authors should answer the following comments before the publication.

Major comments:

1. Lines 287-288: It's difficult to see "BSRN network generally exhibits higher accuracy and satisfies the precision benchmarks" in Figure 5. Please show more analysis (Figures or tables as you like) to support your point.

Re: We are so sorry for not explaining this clearly. To illustrate this point, we have plotted the distribution of the accuracy indicators for different networks (Figure R2). It can be seen that the advantage of BSRN is not significant compared to other networks given that their RMSE and R^2 are comparable. To clarify this point, the sentence has been rephrased as "*Both GCOS and CEOS LPV albedo best practice protocols (Wang et al., 2019) indicate the better performance of BSRN than other networks. However, this phenomenon does not occur with this upscaling model given the comparable RMSE and R^2 among different networks*" in the revised manuscript.

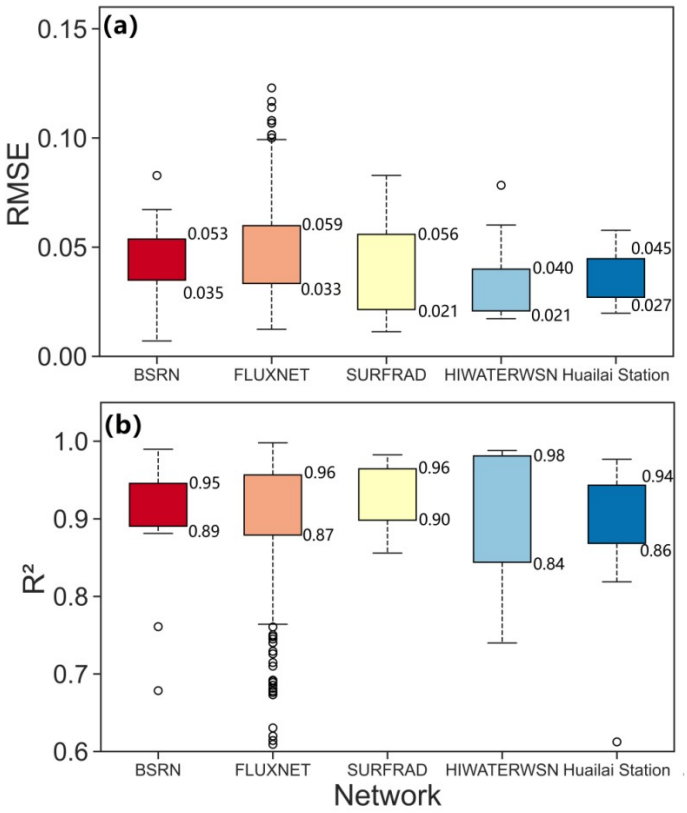

Figure R2. Distribution of RMSE (a), and R^2 (b) of the five Networks used in the study.

2. Lines 297-298: Authors highlighted that "It is worth noting that when the spatial heterogeneity exceeds 0.1, the model's stability fluctuates considerably, indicated by the larger height of the boxplots of RMSE and R2." However, the height of the boxplots of RMSE with spatial heterogeneity < 0.1 in Figure 6a is much larger than that of the other two obviously, while the outliers with spatial heterogeneity < 0.1 in both panels of Figure 6 are much more than those of the other two significantly. Therefore, how can the authors highlight the above result?

Re: We appreciate the reviewer's careful observation and comment. In fact, the data points beyond the upper and lower edges of the boxplots were identified as the outliers, and this is just the unique advantage of boxplots. These outliers should be excluded from the analysis. Generally, the median value as well as the interquartile range should be the measures of performance. However, since the sample sizes are not equal under various spatial heterogeneity conditions, only the median values were focused in this study because it is less influenced by sample size.

The sentence "*It is worth noting that when the spatial heterogeneity exceeds 0.1, the model's stability fluctuates considerably, indicated by the larger height of the boxplots of RMSE and R^2.*" was not enough rigorous and thus was revised as "*It is worth noting that when the spatial heterogeneity exceeds 0.1, the R^2 of the model fluctuates considerably, indicated by the larger height of the boxplots*" in the revised manuscript.

3. Figure 11: Please explain why the results (percentages) in Figure 11 is quite different in the revised manuscript than those in the previous manuscript (version 1)? In previous version, the RRMSEs are quite

lower (around 30%) than those (around 100%) in the revised version.

Re: The difference in the RRMSEs between this version and last version was caused by two reasons. First, the addition of new in situ sites. In the revised version, we have incorporated a significant amount of new site data into our analysis. The introduction of these new sites led to differences between the overall accuracy and performance of the current version and the previous version. Second, the different data sources of the reference data. In the previous version, the MODIS albedo product was used as the reference. However, in the current version, the aggregated Landsat ETM+ albedo on the 500 m pixel scale was used as the reference.

For comparison purpose, we have plotted the RRMSE based on the MODIS albedo product (Figure R3) and the aggregated HJ albedo (Figure R4) over the 416 *in situ* sites , respectively. It can be seen that although the RRMSE present large difference when different data was used as the reference, the RRMSE of pixel scale ground "truth" were always smaller than the single *in situ* site measurements, demonstrating the advantage of pixel scale ground truth over single *in situ* site measurements. It is important to note that the absolute accuracy of the pixel scale ground truth cannot be determined through comparison with other products (e.g., MODIS albedo or ETM+ albedo) since the products themselves contain errors. Instead, the advantage of the pixel scale ground truth was proved through the comparison with *in situ* single site measurements in terms of their relative accuracy relative to a coarse pixel scale albedo value (e.g., MODIS albedo or ETM+ albedo). In other words, the value of RRMSE was not the focus, but the difference of RRMSEs between the pixel scale ground truth and single *in situ* site measurements was the key. In order to clarify this point, we have added the sentence "*Although the errors of the pixel scale ground "truth" are not negligibly small, it is important to note that this kind of error cannot reveal the absolute accuracy of pixel scale ground "truth" given that the reference data itself contain errors. In fact, the focus of this evaluation is not the value of RRMSEs but the difference of RRMSEs between the pixel scale ground "truth" and single in situ site measurements. It can be seen that the accuracy of the pixel scale ground "truth"……*" in **Section 4.2**. Since the previous studies (Peng et al., 2015; Wen et al., 2022, Wu et al., 2016) generally used the aggregated high-resolution albedo as the reference on the coarse pixel scale, we employed the aggregated Landsat ETM+ albedo in this version to be consistent with previous studies.

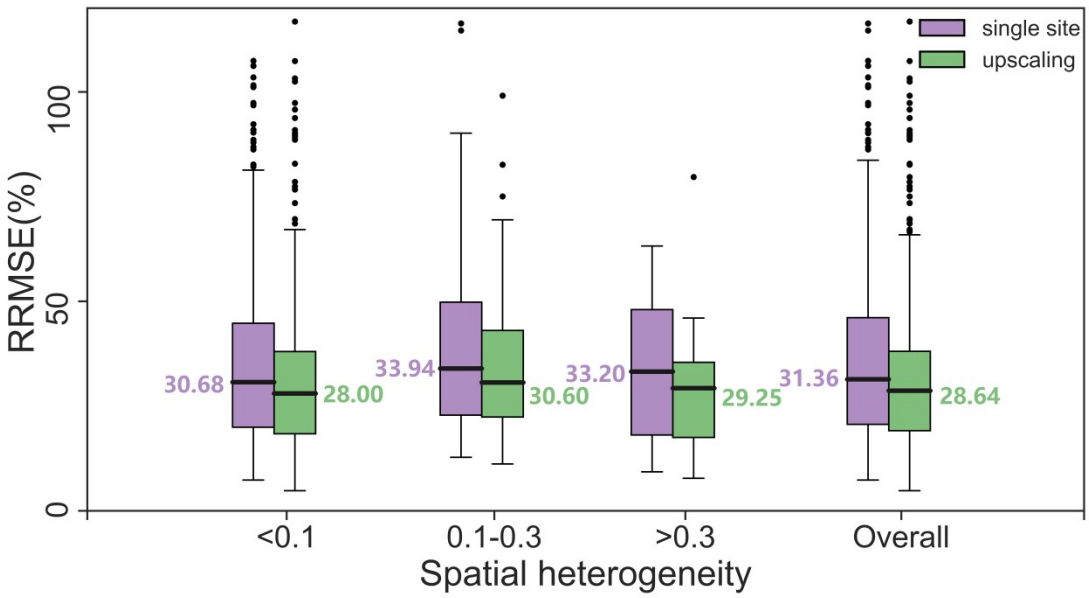

Figure R3: The boxplots of RRMSE of pixel scale ground "truth" and single site measurements. The reference data was MODIS albedo product.

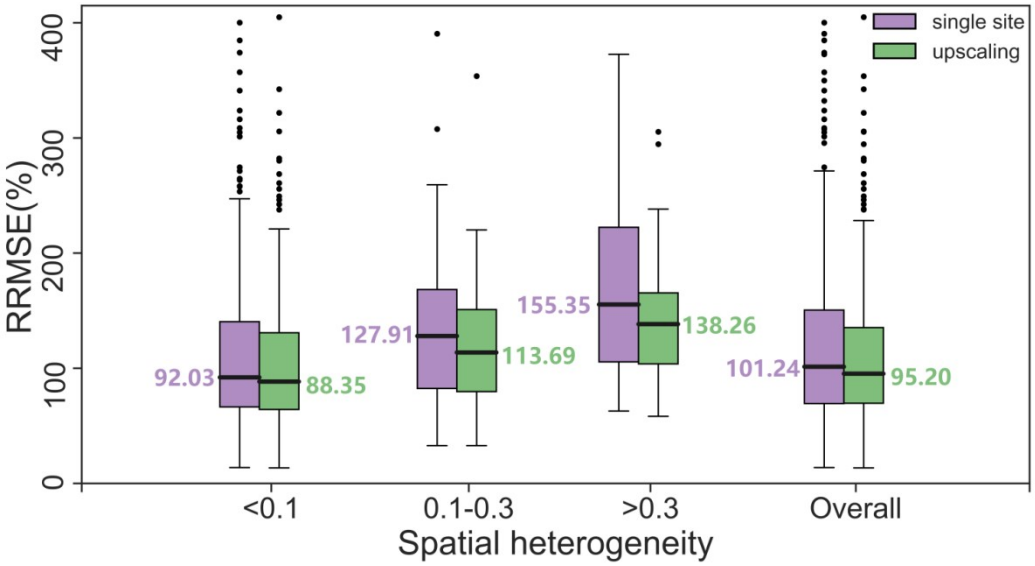

Figure R4: The boxplots of RRMSE of pixel scale ground "truth" and single site measurements. The reference data was aggregated ETM+ albedo.

Minor comments:

1. Section 2.2: the description of ETM+ data should be described in detail. I cannot find the resolution of the data here which should not be mentioned in section 3.2. Please modify.

Re: The resolution of ETM+ imagery bands has been mentioned in Section 3.2.1 as "……*a critical component of the upscaling approach involves the acquisition of upscaling coefficients derived from 30-meter ETM+ albedo covering the period from 2012 to 2018*".

2. Please check the caption of Figure 7. Duplicated [200-500].

Re: We have corrected these errors in the revised manuscript.

3. The contents of functions are overlapping in the PDF version. Please double-check the typing of all functions.

Re: We are sorry for this mistake. We have carefully reviewed the revised manuscript and have made sure that the formatting and typing of all functions are correct.