###

Response to comments on “Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks” by Réjou-Méchain et al.

Dear Editor and reviewers,

We very much appreciated the careful assessment of our manuscript, and we were very pleased by the positive and constructive reviews. The reviewers made a number of useful suggestions to clarify and strengthen the manuscript, which we have addressed.

Below we provide detailed answers to all points raised in the reviews.

We hope that the corrected manuscript, in which all changes are highlighted in red, will now be suitable for publication in *Biogeosciences*.

Thank you for your time and consideration.

Sincerely yours,

On behalf of the authors, Maxime Réjou-Méchain

###

REFEREE 1

This paper poses and attempts to answer several important questions that are significant in the context of current efforts to infer large scale biomass maps from remote sensing and to make more general inferences on landscape scale biomass from a set of sample plots. In fact, the paper is not really about remote sensing per se, but about how accurately one can extrapolate measurements at one scale to a larger scale. In general, it illustrates that the sampling error when small plots are used to represent the average biomass of a larger area can lead to significant errors in the regression relation between the two. This is of special importance when training remote sensing data with plots that are significantly smaller than the resolution of the instrument. Though these conclusions seem fairly sound, the methodology could be improved, and there is some misleading text.

The following are the main scientific issues:

1. The wavelet approach is unhelpful for the purposes of this study. Given the autocorrelation structure of the data, it is relatively straightforward to calculate the variance associated with multiple samples. The wavelet analysis does not help for this and it is not at all clear why the authors have used this tangential approach rather than a less complicated and more informative autocorrelation analysis.

Response: In the new version of the manuscript, we have added empirical variograms for 20 x 20, 50 x 50 and 100 x 100 m subplots. These additional analyses consistently revealed a weak spatial autocorrelation at scales < 100 m. Because the wavelet analyses provide useful additional scale-wise information that will be of interest to some readers (including the other referee), we retained these in the main manuscript. We agree that the wavelet approach may be difficult for many readers to understand and to interpret, so we endeavored to better explain the usefulness of this approach and the meaning of the results.

2. On a related note: the statement about autocorrelation giving rise to a dependence of form s^(-gamma) is wrong, as is clear from an analysis based on autocorrelation.

Response: We did not mean to imply that spatial autocorrelation necessarily results in such relationships, and we agree that our wording here was misleading. We have modified the wording to clarify that a relationship of the form s^-0.5 is expected in the absence of autocorrelation, and that positive (negative) spatial autocorrelation will lead to a less (more) rapid decline in the CV with increasing sample size over relevant spatial scales.

3. In their discussion of dilution bias, the authors mix up two effects. The motivation in the text concerns errors in the ground measurements; this is not the same as accurate measurements of a variable quantity. The implications of this distinction need to be clarified in their analysis.

Response: The term "sampling error" is commonly used in the literature to refer to errors in estimating a true value for a population when measurements (however accurate) are done for only a sample. This term applies perfectly to our situation, where we are concerned with errors in estimating the true value for a larger area based on samples of a smaller area. We have revised the text to clarify this. That said, we agree that we used "sampling error" too broadly in the previous version, and have modified our text accordingly.

4. Why wasn't Deming regression used? This takes account of errors in both dependent and independent variables?

Response: The Deming regression is a special case of Reduced major axis (RMA) regression, which we used in our study. Both these approaches take into account both the error in x and the error in y. We chose the RMA approach because, unlike the Deming approach, RMA does not require prior knowledge of the ratio of the error variances in x and in y. This ratio is difficult to assess in practice, as it would require detailed knowledge not only of the true biomass density over the footprint of the remote sensing instrument and the sampling error of the ground plots but also of the errors in the remote sensing measurements. This is now mentioned in the discussion.

5. The authors have not properly understood the implications of negative autocorrelation in sampling to estimate a quantity. In particular, the second sentence of para from p.5727-p.5728 is not true. In fact, if there is negative correlation then averaging reduces the variance, so gives a better estimate; if there is no correlation it makes no difference what the spacing of the plots is.

Response: We agree that negative spatial autocorrelation in AGB would theoretically lead to a better estimate from a single large plot rather than multiple distant small ones. Given that our analyses of spatial structure have been modified with the addition of the variograms, the associated discussion has also been strongly modified. The new text is consistent with the reviewer's comment.

6. The authors allude to it only once, but an issue that is at least as serious as the topic of this paper, certainly in the tropics, is how representative the available set of plots is. This should be discussed somewhere, as it has effects very relevant to but well beyond the remote sensing problem and is important for REDD+.

Response: We agree. This is now mentioned in the discussion.

A weakness of the paper is its slipshod use of language, which may be because the first author is not a native English speaker (but many of the co-authors are!), but some of which is carelessness. These language issues are scientifically significant, as they change the meaning of many pieces of text. Examples of such (and related) issues include:

a. Heterogeneity is not the same as variability, and in most cases the authors mean the latter. This is fundamentally important for discussing statistical properties which rely on an underlying homogeneous population.

Response: We replaced heterogeneity by variability throughout the revised manuscript.

b. In related vein, what is meant by topographic heterogeneity, given the meaning of heterogeneity?

Response: We replaced "topographic heterogeneity" by "topographic variability".

c. The authors consistently talk about biomass when they really mean average biomass per unit area. This distinction is crucial as without it much of the paper is wrong. The initial text in Section 2.2 is therefore misleading.

Response: As suggested by referee 2, we now use AGBD for Aboveground Biomass Density (Mg. ha^{-1}) and AGB for Aboveground biomass (Mg). We thus modified section 2.2 accordingly.

d. They misuse "uncertainty"; in several cases they mean "error"

Response: "uncertainty" has been replaced by "error" in most places of the revised manuscript.

e. On p. 5719 there is an appeal to the Central Limit Theorem, but this is spurious: the result quoted is just a standard result on averages of independent samples. On the same page, what does ~ mean?

Response: We remove the reference to the Central Limit Theorem and rephrased the sentence including the ~ symbol.

f. The labelling of some of the Figs is misleading, e.g. Fig. 2a does not show sampling error; Fig.3a does not show spatial correlation, nor does 3b; it is wavelet variance.

Response: These labels have been changed.

Following are some more detailed comments on the text:

- On p.5717, l.5, it states that small ground samples will have large sampling errors if there is substantial local "heterogeneity". That is a tautology.

  Response: We agree that this sentence is rather a truism but given the questions investigated in our study, we believe that this statement must be made clear and unequivocal, even if it is a tautology.

- The end of the 1st para talks about the need to correct various errors, then fails to comment further on this.

Response: At the end of this paragraph, we state that there is a “need to quantify” the errors due to the spatial mismatches between sensors and field measurements. We address this issue by simulating circular footprints and square calibration plots (Fig. 6) and by investigating how the error associated with such spatial mismatch scales with both calibration plot and footprint areas (Fig. S10). Because subsequent comments highlighted a lack of clarity in these analyses, we improved the description of the methods.

- How meaningful are measurements at 5 m scale (p. 5719), given their dominance by edge effects?

  Response: We agree that measurement at 5-m scale are not relevant for remote sensing measurements. We included quantification of spatial variability at this scale to increase the range of scales over which we could investigate the decay of spatial variability with sample area. When more realistic simulations were done (e.g. for the dilution bias analysis), the smallest plot size was set to 0.04 ha (20x20 m), a sample area regularly used in remote sensing studies, even if large edge effects also occur at that scale.

- The use of the word “grain” instead of “scale” is unnecessary and confusing.

  Response: We replaced “grain” by “scale” throughout the revised manuscript.

- On p.5720 there is a reference to an area sˆ2, but s is an area.

  Response: This was indeed an error. This is now corrected.

- On p.5721 what does the phrase beginning “was perfectly perceptive : : :.” mean? That the remote sensing measurement is assumed to be correct??

  Response: Yes, this meant that the remote sensing measurement was assumed to infer the exact above ground value that would be measured in the field. We rephrased this sentence.

- On p. 5721 it implies that remote sensing fields of view are circles (or ellipsoids earlier); this may approximately be true for optical data but not for radar, where they are typically rectangular.

  Response: We simulated remote sensing footprints as circular to illustrate the general issue of mismatch between remote sensing field of view and ground measurements. We now make clear that this is merely a simple example. More realistic approaches would require sensor-specific 3D simulations. Radar products are indeed post-processed to represent rectangular areas. However, the original footprint do not precisely match the rectangular area as measured on the ground because radar is measuring the distance to features in slant-range rather than the true horizontal distance along the ground (i.e. Slant-range scale distortion occurs). We now address these issues both in the introduction and in the discussion.

- It is the root mean square error, not the mean error.

  Response: Please see next response.

- In (2) is it a condition that the field plot lies entirely within the circle? And why is the term ErrCV used;? This is misleading as it is not a CV and its connection to CV is not explained.

  Response: In this simulation, field plots were centered in circular remote-sensing footprints; thus, they were entirely within the circle when field plots were smaller than footprints or they sampled slightly different areas when field plots and circles were of similar size (e.g. the corners of the squares were not sampled by the circular footprint). As suggested by reviewer 2, we moved figure S2 to the main text in order to make this clear.

  The term ErrCV is used because it is the ratio of the RMSE to the mean AGB, which is thus analogous to a coefficient of variation. The formula was split into three equations to highlight the connection of ErrCV to the coefficient of variation.

- There are repeated statements in para 2 on p. 5722. What is meant by a "realistic reliability study"? Why and how is the ICC used? ICC is relevant to measurements made on units that are organized into groups. What are the groups here? The whole of this para following (4) is unclear.

  Response: We entirely rephrased this paragraph and provided more details on the ICC calculation.

- On p.5724, in para. 1, it seems strange not to mention at this point that the Asian sites show more elevation change, hence more AGB variation. This is not pointed out until several pages later.

  Response: We agree, this is now mentioned in the revised manuscript.

- It is unclear what the sentence about lower gamma values is meant to be saying.

  Response: Slopes greater than -1/2 indicate positive autocorrelation in AGB at the relevant scales, as illustrated in the simulation below.

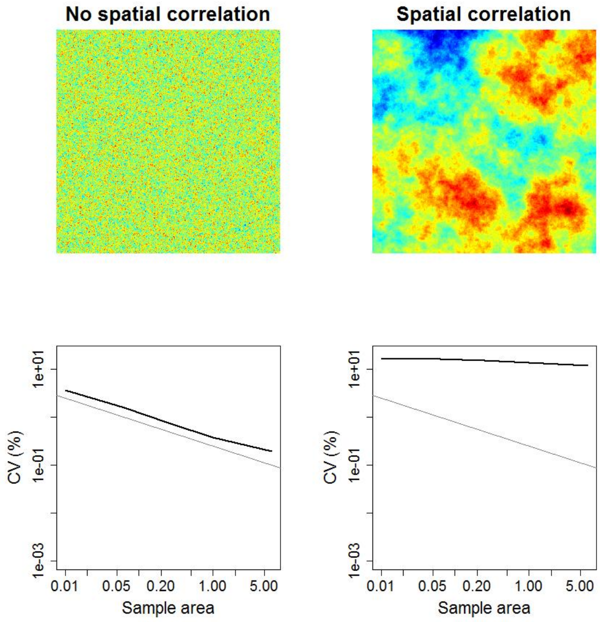

Figure: Simulation of the relationship between the coefficient of variation (CV) and the sample area under two different spatial autocorrelation schemes. Random fields with no spatial correlation (upper left) and with a positive spatial autocorrelation (upper right) were generated in a 500x500 grid with an exponential variogram model (sill of 0.025 and ranges of 1 and 100 respectively). As can be seen, with no spatial correlation (left panels), the logarithm of CV decreases linearly with the logarithm of the sample area with a slope of -0.5 (the -0.5 slope is illustrated in light grey). When positive spatial autocorrelation occurs, the slope is much shallower (right panels), with a slope of -0.05 in this particular simulation case.

- What does "expected" mean in Fig. 4? Is it being used in some statistical sense?

  Response: This corresponds to the slope that would have been obtained without bias. We have modified the text for clarity.

- p.5725. I could not see how the figure quoted tells us about shape effects, and the text does not explain this.

  Response: We have revised the legend of the figure S10 and associated manuscript for clarity.

- In para 2 what does "such models" refer to? This sentence is unhelpful overall. It should really say that "if the field measurements have large errors, etc. : : :.".

  Response: "Such models" referred to OLS-based models. We modified this text accordingly.

- As noted above, the authors are mixing up errors in the ground measurements with accurate measurements of a variable quantity.

  Response: see above.

- 1st para. in Section 4: "spatial" should be omitted. Where does 26% come from and what does it refer to?

Response: The word “spatial” has been removed. 26% is the average CV at the 0.25-ha scale. For clarity, we added a reference to table S2.

- p.5727. the first sentence confuses detection of change with estimation of biomass change.

  Response: This sentence has been removed.

- p.5730, Conclusions: there have been numerous studies of the errors in field sampling and their effects on carbon estimates. How do the authors suggest topographic variation be accounted for?

  Response: We now provide more details on how topographic variation might be explicitly considered in sampling designs in the discussion.

###
REFEREE 2

This paper deals with an important and critical issue: the scaling of plot estimates of carbon density to larger areas, in particular to remote sensing pixels. This is vital for REDD+, designing algorithms for predicting carbon stocks using data from new satellites addressing this issue such as BIOMASS (which will produce outputs at a coarse, 200 m pixel size, larger than most plots), and for developing a better understanding of forest dynamics.

Their findings in terms of the influence of topography are novel and interesting: topography has long been discussed as a source of error in comparing field data to remote sensing data, but its influence has not been quantified in this way before. Clearly from their findings sampling designs in hilly areas must be stratified by this topography. Also their findings in terms of the use of different sized subplots for developing regression equations are very important. It is known that using OLS regression for small plots would cause an underestimation slope, but it had been thought that RMA or Theil-Sen regression would correct for this: these results show that this is not the case. Finally this study shows that small forest plots (<0.25 ha) should not be directly compared to satellite remote sensing data for spatial sampling reasons, before even considering the known additional errors that result from small plots having a high edge-area ratio and a larger relative geolocation error.

I think the paper is excellent and should be published after minor corrections. I have a number of relatively minor comments below, with just one where I consider an additional analysis necessary. One of the key result of the paper is displayed in Figure 4, showing the expected errors that result from a calibration plot of a particular area compared to a remote sensing footprint of a particular size. This however considers only circular remote sensing footprints, which is not always appropriate. This is discussed in detail below, and should be addressed.

Response: We modeled the remote sensing footprint as circular for the sake of simplicity, as now stated and discussed in the revised version of the text. Please see the response to reviewer 1.

The other points I raise are not critical, and should be treated as suggestions.

- 5716 lines 14- 15 – need to explain why monitoring of forest carbon stocks is necessary for REDD+. REDD+ projects will it appears quantify changes in carbon stocks by stratifying areas, assigning carbon density values to each strata, and multiplying the two to get total stocks (and differencing these from year to year to calculate emissions). Thus lines 14-15 to not necessarily follow from your first paragraph. Need to justify space-based mapping of carbon stocks (as opposed to just landcover classes) in terms of degradation, regrowth and enhancement activities.

Response: We agree that this important point was missing and have modified this paragraph accordingly.

- 5717 line 16 – the issue raised here of uneven sampling of ground plots is raised here but not covered or discussed in the paper – if this is to be left in the introduction maybe it could also be raised in the discussion?

  Response: This is now discussed at the end of section 4.3.

- The bias introduced by non-random plot selection is hard to quantify, do the authors have any ideas on how to propagate this.

  Response: We agree that the bias introduced by non-random plot selection is hard to quantify; this is clearly an outstanding challenge, and we believe that this is beyond the scope of this paper. We now briefly mention this issue in the discussion.

- Page 5719 lines 1-2 – it is strange here to refer to a general paper about the proportion of biomass from lianas, as it is known that this can vary significantly from plot to plot. I think it is fine to exclude lianas, I can see how it simplifies the analysis and should not change the conclusions. But it would have been good to estimate the proportion of biomass from lianas within each plot (or maybe subplots too), and exclude those with a high (say >10 %) contribution of total biomass from lianas. This would prevent the possibility of lianas biasing the analysis.

  Response: Lianas were not measured in all plots so that we were not able to include them systematically. However, none of the plots had a high proportion of large lianas; thus, excluding lianas from the analyses is not expected to introduce bias in our results. We slightly modified the sentence to “We […] excluded lianas from our analyses in the few sites where these were censused. Lianas usually represent less than 5% of the total AGB (e.g. Schnitzer et al., 2012).”.

- Lines 3-5: information should be given here on how elevation data was collected for each plot. If this SRTM? Field survey measurements? The description here is too brief, it would not allow for this study to be replicated based only on this information. How is topographic heterogeneity defined? Moreover, given you don’t actually use topographic heterogeneity as a variable in the paper, but elevation range, why not in the rest of the paper call this variable what it is, i.e. ‘elevation range’, rather than heterogeneity?

  Response: Elevation data were either generated by field measurements with surveying equipment or by high-resolution airborne LiDAR. Topographic variability was defined as the mean standard deviation of elevation within 1-ha subplots (see legend of figure S1). This is now clearly stated in the manuscript.

  In the method section and in the results, we systematically used the term “topographic variability” in association with “elevation range” in order to avoid any ambiguity. We chose to keep “topographic variability” in the manuscript because the elevation range was an excellent proxy of it (Fig. S1) and because we believe that this term is more intuitive for readers.

- Page 5719 lines 8-10 and throughout: AGB is being used here to mean both Biomass density (Mg ha-1) and total tree biomass (AGB estimates for each stem, presumably in Mg or kg). This is often the case in the literature, but given the repeated use of AGB in this paper I found this confusing. I suggest the use of something like AGBD for Aboveground Biomass Density (Mg ha-1) and AGB for Aboveground biomass (Mg).

  Response: We agree and followed this recommendation throughout the revised paper.

- Page 5719 line 12 onwards: I like the use of CV(s) as a measure of divergence between subplot AGB values and the mean of the whole plot. However the comparisons are confused because the size of the large plots vary from 8 to 50 ha. This may not have much impact on the results, but does mean that the CV(1) values you use to test issues such as topographic heterogeneity are not all strictly comparable. This may be fine, but I would like you to test this impact by calculating CV values for 8 ha subsets of the larger plots, and seeing whether trends based on 8 ha subplots of all plots are significantly different from using a single CV value calculated per plot. In other words I would like you to test whether cutting all your largest plots into approximately the same size as your smallest plots, and then calculating CV values from comparable sized plots, influences the results. If not, then your current analysis is fine – it is better to have a single value per plot than the pseudo-replication of giving 6 values for a 50 ha plot, but you need to test whether the CV metric is sensitive to large plot size.

  Response: We investigated the consequence of calculating the CV for different plot areas. Because the largest possible standardized size was 4 ha (200x200 m), i.e. some plots had a side of “only” 200 m, we calculated a “standardized” CV for each site as the mean CV that would be obtained in 4-ha subplots chosen randomly (n=1000 random repetitions). The results areillustrated in the figure below and show that this “standardized” CV is highly correlated with the “unstandardized” CV for all spatial scales investigated. This is because most variation in AGBD occurs at small spatial scales. Using the standardized CV also gave a highly significant relationship with the elevation range (Spearman’s rho =0.59, p<0.001). Of course, the elevation range itself also depends on plot size, and ideally the standardized CV of AGBD within 4-ha plots would be correlated with the standardized CV of elevation within 4-ha plots – but the latter is not available for all plots. For the sake of simplicity, we retain the “unstandardized” values in the main text. We now provide the results from this “standardized” analysis in the new Fig. S4.

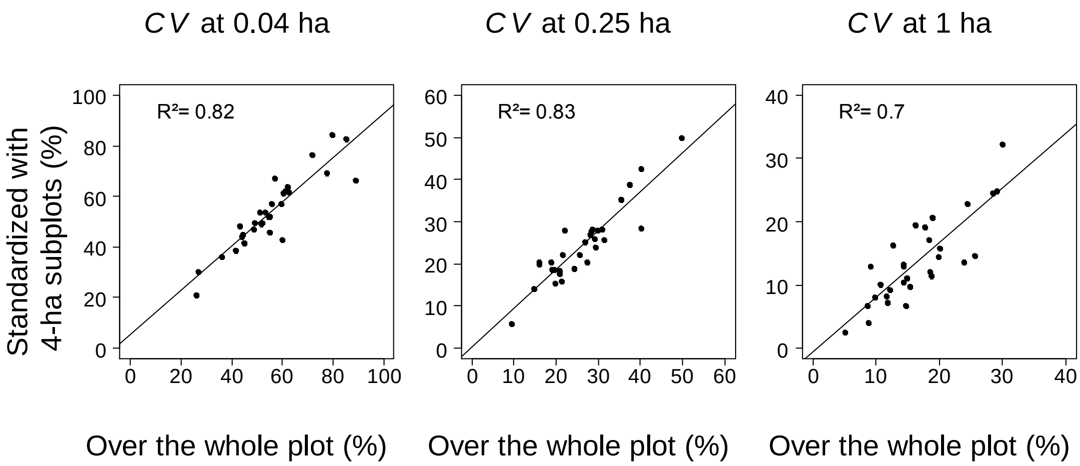

Figure: Relationship between the coefficient of variation in Above Ground Biomass calculated over the whole plot (x axis) and the mean CV calculated in 4-ha subplots (mean over 1000 randomly chosen subplots), relative to the 1:1 line. The R^2 of the linear model is reported within each panel.

- Page 5719 line 25: are you sure it is the Central Limit Theorem that implies this?

  Response: The reference to the Central Limit Theorem has been removed.

- Page 5721 line 1-: You need to justify decision to choose to represent remote sensing pixels as circles. This may be fair for most optical sensors, and potentially IceSAT GLAS mentioned, though it's not that simple as in both cases more information will pass to the sensor from the centre than edge of the pixel. But for radar and high resolution (airborne) LiDAR, looks/returns are aggregated into rectangular pixels. Radar in particular has no circular features, and the beams of a LiDAR sensor, while circular, from an aircraft are very small and aggregated into rectangular pixels. This needs to be discussed and addressed here and in the Discussion: It is okay to keep the analysis as is, but it must be made clear that the analysis as presented is relevant only to optical, and potentially spaceborne LiDAR, instruments. Alternatively, and perhaps more powerfully, the same analysis could also be performed with rectangular pixels of different sizes, allowing the effect of shape as well as size to be quantified.

  Response: As discussed above, we have improved the description of the methods explaining why we choose circular footprints. The main reason is to simulate a spatial mismatch between the footprint area and the calibration plot area, a mismatch which may occur for several reasons and take different forms. As now stated in the manuscript, a realistic approach would require a sensor-specific approach and 3D simulations, for example to explore the influence of the slant-range scale distortion in radar. We also note that when the calibration plot is fully inside the footprint, the shape of the footprint (circular or square) does not matter (whatever the shape of the footprint, circular or square, we found highly similar ErrCV values when the subplot was smaller than the footprint, not shown). The difference in shape mattered only when the calibration plot and the footprint have the same area, so that each encompasses area not encompassed by the other.

- The authors could also explain their choice of 'footprint' areas better. They have chosen 0.5, 1, 2 and 4 ha, but these do not match onto any remote sensing instrument I am aware of, and in particular start quite large. Surely more useful would have been to look at a range of pixels including Landsat size (30 m, i.e. 0.09 ha), ICESat 2 (50 m circular footprint, i.e. 0.2 ha) all the way up to MODIS size (250 m, i.e. 6.25 ha). If square pixels were being considered then the 4 ha pixel size makes some sense, as this is the probable size of pixels in the BIOMASS product, but a circular 4 ha pixel does not at all represent the viewpoint of BIOMASS, with this table probably overstating ErrCV in this case. In my view the analysis should be repeated with a wider, more realistic range of footprint sizes, and square as well as circular footprints.

Response: We chose these footprint scales to address those large enough that they are often calibrated via smaller plots, and those small enough that we can fit two of them within each of our study sites. The objective of our paper, after all, was to examine the implications of spatial variation in biomass density for calibration of large-footprint remote sensing with data for smaller plots. Thus, we restricted ourselves to simulations of fairly large footprints and small plots within them.

Many existing plots are larger than the footprints of Landsat and ICESat 2 (0.09 and 0.2 ha), and thus our concerns are by and large not relevant for these instruments. Further, as we showed, when calibration plots are small, most errors are due to the field sampling errors, almost independently from footprint size. Hence, with small resolution products such as Landsat or Icesat, sampling errors are likely to be very high if smaller plots are used or if spatial mismatches between the field and the sensor signal occur. We added a sentence in the text about that issue.

We would have liked to have simulated larger footprints such as 6.25 ha for MODIS, but even our large plots start to reach their limits as we increase the scale. In the current work, we simulate two nonoverlapping remote sensing footprints within each site. A 4-ha circular plot has a diameter of 226 m, while a 6.25 ha circular plot has a diameter of 282 m. In five of thirty sites the smaller plot dimension is 200, and thus these already get dropped by the 4-ha analysis, leaving us with 25 sites (and 50 simulated footprints). In 18 of 30 sites, the larger dimension is 500 m or less, leaving only 12 sites in which two overlapping 6.25 ha footprints could be simulated, and 13 in which one could be simulated, or potentially 37 simulated footprints total. Thus if we added larger footprints, the sites included would become ever more constrained. We would then either have different sets of sites for different simulated scales, making comparisons across scales difficult, or we would have a much reduced set of sites across all scales, which would fail to draw on what we see as the strengths of the CTFS-SIGEO network for addressing this question.

These constraints, and an interest in seeing how patterns change with the scale of the footprint, led us to restrict our footprints to the range 0.5-4 ha. We chose doubling sizes within this range (0.5, 1, 2, and 4 ha) because of our interest in seeing how patterns change with scale, and the power function nature of most such scaling relationships. We note that even though our results do not specifically encompass other scales, they provide a basis for estimating patterns at other scales through interpolation and extrapolation.

- Finally on this point, if space allows, it would be good to include Figure S2 in the main paper to illustrate this analysis. I do not think the text is terribly clear and this diagram is very helpful in explaining this analysis.

Response: We have improved the description of this approach and moved figure S2 to the main manuscript.

- Page 5721 line 17: In contrast the sampling error propagation analysis, including the assessment of regression dilution, assumes a square 4 ha plot (BIOMASS-like) – this is good and this is an excellent and sensible analysis to perform, but the inconsistency with the above is not discussed or explained. It makes the results from the two analyses difficult to compare.

  Response: As stated above, the effect of differing shape between calibration plot area and footprint area only occurs when these areas are of similar size, a case that was not investigated in the dilution bias analysis. Hence the two analyses are directly comparable as having a circular footprint or a square one would have produced extremely similar field sampling errors.

Discussion

The Discussion is excellently written and gets to the nub of the problem, explaining well all the significant findings of the study. Its implications for future research design are clearly stated. The explanation and interpretation of the wavelet analyses is well presented in my opinion. However, it ignores right until the end the role that high resolution remote sensing can play. High resolution data, especially aircraft lidar, can act as a stepping stone between small field plots and the relatively coarse resolution remote sensing methods discussed here (min considered pixel size in comparison is 0.5 ha). When discussing 'remote sensing' in this paper it appears the authors are discussing 'satellite remote sensing', whereas much biomass mapping effort uses airborne LiDAR sensors where individual trees can be distinguished and many (though not all) of the trends shown here would not apply. This is mentioned in the final two sentences, but it would be great for this to have its own paragraph highlighting studies that have succeeded in using high resolution data to effectively increase the size and number of plots that can be used to calibrate satellite remote sensing data.

Response: We agree that small-footprint LiDAR may offer a unique opportunity to act as a stepping stone between field plots and large remote sensing footprints. In the revised manuscript the discussion of this approach has been enlarged.

Fig 1: Forest cover map should be referenced as GLOBCOVER2009 as well as the Bontemps reference, and the different colours of forest should be included in a key or left out, i.e. with a single colour for all forests.

Response: The references have been included and the same colour has been applied to all forest types.

Additionally, though optionally, I think it would be useful to have lines showing the boundary between tropical, subtropical and temperate here, given these are used in later figures and analyses.

Response: Thermal zones from the IIASA classification have been added to the figure.