**Responses to Referees Comments**

**We are very grateful for the recommendations and comments we have received from the reviewers. We have addressed these comments in the revised manuscript to the best of our ability. We believe that the manuscript is now significantly improved and we look forward to your decision. We would also like to thank the two anonymous reviewers for their constructive comments, which have been very helpful for improving this manuscript.**

**Anonymous Referee #1**

*The study addresses the problem that maps of forest aboveground biomass (AGB) are often based on sample plots which may not be representative for the entire area of interest. In a case study for Eucalyptus plantations in southern China, the authors compare AGB predictions produced with different approaches. These include three machine learning methods and one spatial statistical method as well as combinations of them. Based on a set of goodness-of-fit statistics, the authors conclude that a combination of random forest and the spatial statistical approach P-BSHADE produce the best AGB predictions and that this approach could reduce uncertainty due to non-representative sampling. Non-representative sample distribution is indeed a big problem in mapping of AGB and other forest attributes. However, the manuscript does not clearly state how the presented methodology can improve this situation. While research question (1) can partly be answered by the given results (e.g., Fig. 5), a clear answer to research question (2) is not possible based on the conducted analyses. Several sections need considerable improvements for the manuscript to become a consistent story.*

**Response: Thanks for your constructive advice. We agree with many of the problems that you have pointed out and your comments helped us improve our article. We agree with the reviewers that our current combined approach to spatial statistics and machine learning is not sufficient to address the uncertainty of regional AGB mapping due to non-representative plots. The most fundamental reason for this is the lack of non-destructive measurements for every tree covering the entire region. Regarding the first research question, we reviewed several papers on the application of machine learning to geosciences and added some discussion to section 4.2 (Lines 415-431). The second research question is correlated with the validation problem. To address this problem, we would have needed systematic sampling plots (forest inventory data) and corresponding non-destructive measurements for each tree in each plot in the entire region to validate. However, we don’t have the systematic sampling plots yet. In future studies, we will attempt to collect this data to validate the regional scale results.**

**Based on this limitation, we revised the main research question in the current study to: “Can the integration of spatial statistical and machine learning methods improve the accuracy of AGB models at the plot level?” (Lines 103-104) The plot level is a bridge that connects single tree AGB measurements to AGB estimates at a regional scale. Accurate AGB mapping at the plot scale provides a basis for future upscaling to the regional scale. Our proposed combination of spatial statistical analysis and machine learning methods can significantly improve the accuracy of plot-scale AGB mapping. We therefore restructured our paper based on this idea. To this end, we collected non-destructive measurements for an additional 22 independent plots and repeated the workflow of the optimal model to test the combined models (Lines 249-255). The section on regional scale AGB mapping has been shortened (Lines 256-261) and moved to the Supplementary Material (Lines 172-189).**

**In addition, we defined some general concepts used in this article.**

**Allometric model at tree level: we fitted an allometric model at the tree level using 90 harvested trees.**

**Allometric model at plot level: we fitted an allometric model at the plot level using 30 sample plots.**

**True AGB of sample plots: we applied an allometric model at the tree level and calculated the sum of AGBs for all trees in each sample plot.**

**Reference AGB of sample plots: this is the predicted value of the fitted allometric model at the sample plot level.**

*Abstract:*

*The Abstract currently focuses mostly on the problem statement. It should be more specific about the used methods and most important results.*

**Response: Thank you for pointing out this issue. Based on your advice, we have rewritten the abstract as follows (Lines 19-38):**

**Abstract：**Aboveground biomass (AGB) estimates in plot level is a bridge that connects single tree AGB measurements to AGB estimates at a regional scale. Accurate AGB mapping at the plot scale provides a basis for future upscaling to the regional scale. However, the uncertainty and error propagation inherent in different prediction models make this process challenging. Allometric models are most commonly used in building plot-level AGB models, but they cannot fully capture the complex and spatially heterogeneous landscapes where multiple environmental covariates (such as longitude, latitude, and forest structure) affect the spatial distribution of AGB. To address this challenge, we tested in this study, an approach that combines machine learning with spatial statistics for constructing a more accurate AGB plot-level model. The study was conducted in a Eucalyptus plantation in Nanjing, China. We develop, evaluate, and compare three different machine learning models (support vector machine (SVM), random forest (RF), and the radial basis function-artificial neural network (RBF-ANN)), one spatial statistics model (P-BSHADE), and three their combinations (SVM & P-BSHADE, RF & P-BSHADE, ANN & P-BSHADE), based on data from 30 sample plots and their corresponding environmental covariates. We found that the performance indices (RMSE, MAE, and MRE) of all combined models were substantially smaller than those of any individual models, with the RF & P-BSHADE combination method having the least values. Thus our results clearly demonstrate that combined models, especially the RF & P-BSHADE, can improve the accuracy of AGB plot-level models and reduce uncertainty in plot-level or even large forested landscape AGB estimates. The research results would be important by virtue of reducing the uncertainty in regional carbon balance estimates.

*Introduction:*

*In general it should be made clear that this study does not* *use remote sensing data for AGB mapping. As many studies do use remote sensing and the authors cite several such studies, it is not clear to the reader that the forest structure used in this study is entirely ground based. This should be clear from the beginning.*

**Response:** **Thank you, we agree with your comment. We did not use remote sensing data for AGB mapping, so the remote sensing articles we cited may confuse some readers. In response to this comment, we have changed some cited articles (Lines 50-56) to avoid misunderstanding and revised the last paragraph of the Introduction (as follows, lines 95-97). In stating the objective of this study, we mentioned that the data used in this paper is based on ground surveys (Lines 95-97). In addition, we present the data used in this paper in detail in the Materials & Methods section to preclude the misunderstanding** **(Lines 122-135).**

**“The objective of this study is to develop and evaluate a combined machine learning and spatial statistics method for improving the prediction accuracy of AGB spatial mapping at the plot level by ground based samples.”**

*L. 53: Please explain what is meant by back-end processing. The difference to front-end processing is not clear.*

**Response: Thank you for pointing out this problem. Our explanation of the difference between back-end and front-end processing is presented as follows. However, our manuscript does not prove that the methodology solved the problem of non-representative samples, so we have revised our research questions (Lines 101-104) and relevant parts of the manuscripts accordingly (Lines 19-38, 46-56, 89-104, 235-255, 333-365, 415-431), and also deleted this section.**

**“Compared with front-end processing, which focuses on the processing of model input data, while back-end processing aims to the advanced model and algorithms which have the weak model parameter assumption and the good prediction performance based on small samples. These back-end processing approaches may substantially increase the accuracy of AGB maps.”**

*L. 86: Explain what is meant with stability of the second steps.*

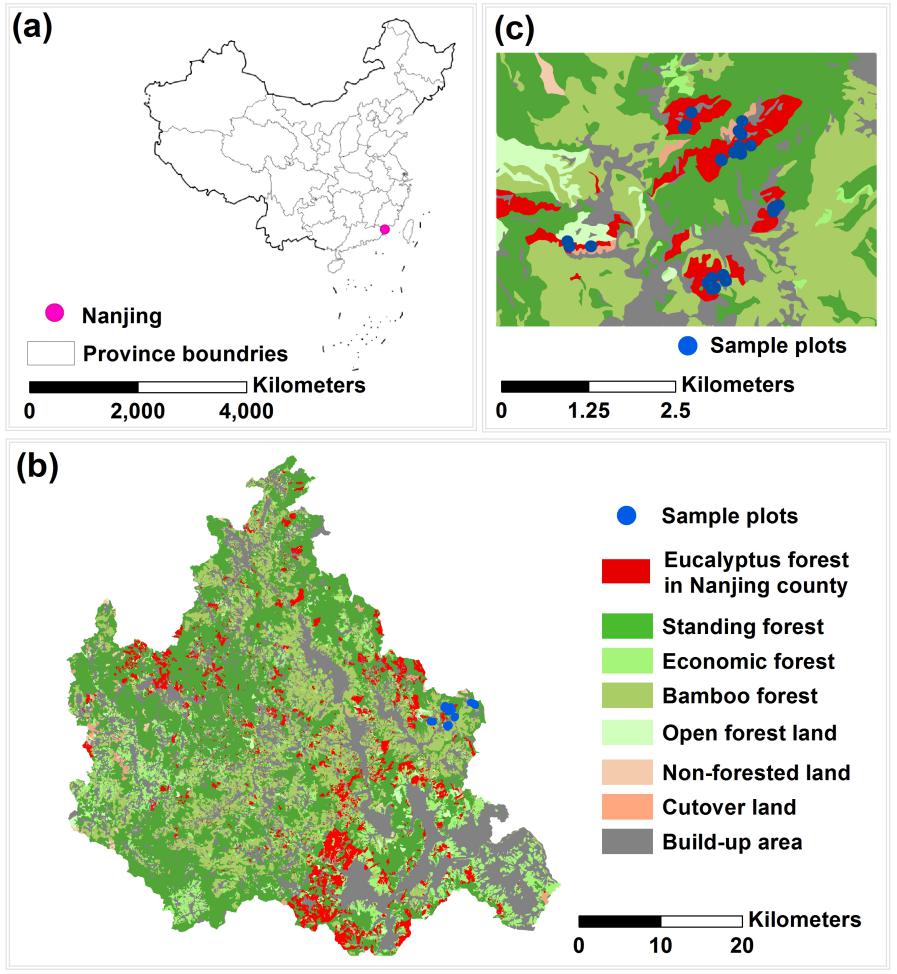
**Response: Our statement was incorrect, so we have corrected it** **(Lines 86-88). We meant “spatial stratified heterogeneity,” not “stability of the second steps.” Spatial stratified heterogeneity is a ubiquitous ecological phenomenon in which the within-strata variance is less than the between-strata variance, such as occurs in ecological zones with many ecological variables.** **Spatial stratified heterogeneity reflects the essence of nature, implies potentially distinct mechanisms by strata, suggests possible determinants of observed processes, allows the representativeness of observations of the earth, and enforces the applicability of statistical inferences. We have modified this section as follows (Lines 86-88):**

**“Second, the assumptions of the spatial statistical method (e.g., spatial autocorrelation and spatial stratified heterogeneity), which may not always be valid in forest AGB.”**

*Materials & Methods:*

*L. 111: According to the numbers given in L. 103 and 110* *about 10% of the forest in Nanjiing county is Eucalyptus forest. So why isn’t ~10% of Fig. 1b red? Also mind the typo “Bamboo foreat” in legend.*

**Response: Thank you. We have corrected these errors in Fig. 1b , as shown below (Line 118).**



*L. 119: Can you describe the FMPI data a bit* *further? Are these 2980* *patches polygons of* *irregular shape representing stands of same structure? If yes, what are typical sizes? Or are they grid cells?*

**Response: Thank you for your kind advice. The 2,980 polygon patches are irregular, they are not grid cells. The structure in each** **patch is homogeneous, but the structure of the different patches is not necessarily the same. The approximate sizes of the patches ranged from 0.15 to 30 ha. We have added a more detailed description of the FMPI data** **as follows, but this section is now located in the Supplementary Material section** **(Lines 152-163) due to the major revisions in paper.**

**“The FMPI data was consisted of many irregular polygons divided from forest according to their forest structured characteristics. Each polygon was homogeneous structured. In this study, we selected the FMPI data of Eucalyptus plantation forest (including 2980 patches).”**

*L. 126: Please* *remove information irrelevant in this study. I think, soil and* *topographical variables are not further used in the analysis.*

**Response:** **Thank you.** **We have removed all information that may be irrelevant to this study, including soil and topographical variables. Some of the material that is marginally relevant has been moved to the Supplementary Material (Lines 158-160).**

*L. 129-131: It is unclear what is meant by systematic* *and stratified sampling in this context.*

**Response: Thank you. Our statement was incorrect. The establishment of FMPI data involves** **accuracy verification and adopts the method of systematic sampling by setting up fixed sample plots of 0.0667 hm2 with equal distance (1 km). We have corrected our statement about systematic and stratified sampling, as follows, and moved it to the Supplementary Material section (Lines 161-163) after major revision of the paper.**

**“The accuracy of forest patch variables was tested using systematic sampling. A 95% sampling precision was required.”**

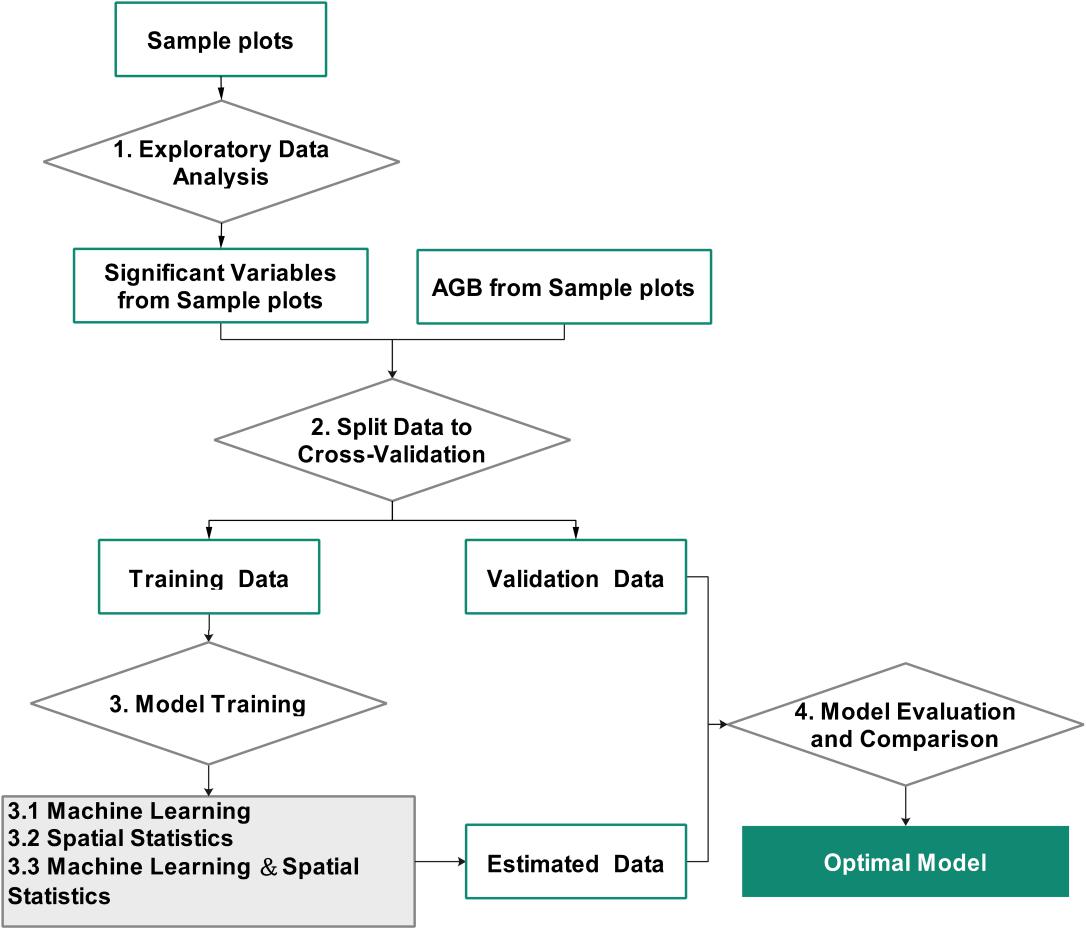
*L. 142:* *Foliage and* *roots should not be mentioned if they are not relevant for the analysis.*

**Response:** **Thank you for your kind advice. We have removed the description of roots (Lines 132-133). In this study, AGB included tree foliage, stems, and branches. In our calculations of AGB, we weighed the stem, branches, and foliage of each harvested tree separately, then added them together to obtain AGB. To avoid misunderstanding, we have rewritten the sentence (Lines 132-133) as follows:**

**“We then weighed the biomass of each organ (foliage, stems, and branches) for each tree to obtain AGB of each harvest-tree.”**

*L. 156: Typo in* *Fig. 2 “Model Trainning”.*

**Response: Thank you. We have corrected this typo as shown below (Line 145):**



*L. 167: Please explain briefly the concept of spatial stratified heterogeneity and how the geographic detector method works.*

**Response:** **Thank you for your kind advice. We have added a brief explanation on the concept of spatial stratified heterogeneity and how the geographic detector method works (Lines 155-162), as follows:**

**“****The spatial stratified heterogeneity (****referring to the characteristic that the within-strata variance less than the between strata-variance, it is ubiquitous in ecological phenomena, such as AGB)** **of the true AGB of sample plots was evaluated using a q-statistic value by the GeogDetector model, which is a software tool based on spatial variation analysis of the geographical strata of variables, as proposed by Wang et al. (2016). First, we used the K-means algorithm to obtain the strata of true AGB for preprocessing of GeogDetector. Then, we regarded true AGB as Y, regarded the strata of true AGB as X and put them into the GeogDetector model to obtain the q-statistics value (Wang et al., 2010;Wang et al., 2016).”**

*L. 178: At several places in the manuscript the word “simulate” is used, where I would use* *“predict”. Simulation is usually associated with process-based model simulations, while statistical model output is more commonly referred to as prediction. Consider changing the wording to avoid confusion.*

**Response: Thank you. We have corrected this usage throughout the manuscript (Lines 173, 183, 197, 213).**

*L. 182: What do you mean with “localization biomass model”? To my understanding the P-BSHADE interpolates the AGB values of the neighboring plots.*

**Response: Thank you for pointing out this issue. “Localization biomass model” refers to the allometric model. The allometric model, which is based on stand age, was developed by our team using the same data sets used in the rest of this study. Detailed information on the model can be found in “Variations in the biomass of Eucalyptus plantations at a regional scale in Southern China” (doi: https://doi.org/10.1007/s11676-017-0534-0). After further consideration, we think the** **term “localization biomass model” may not be appropriate here, therefore it has been changed to “allometric model” in this revision (Line 176).**

**In essence, the P-BSHADE model is a data fusion approach which combines observed samples with a reference series (related variables). The P-BSHADE model regards strongly correlated plots as neighboring plots.**

**We have rewritten the method for the P-BSHADE model as follows (Lines 203-214):**

**“The P-BSHADE model is an interpolation approach based on the assumption of spatial autocorrelation and spatial heterogeneity. It is markedly different from the Kriging and Inverse Distance Weighting (IDW) algorithms, because the latter two algorithms only regard spatial autocorrelation as an assumption condition. Besides, P-BSHADE regards the strongly correlated sample plots as the neighboring plots. The core of the model is minimizing the variances of predicted error and unbiased estimation. The model is also a data fusion approach which could combine the observed samples with and reference series (related variable). In brief, the P-BSHADE includes two steps. First, obtain reference AGB of all sample plots according to the allometric model. Second, use the reference AGB of target sample plot and true of other sample plots to obtain the weights relationship between target sample plot and other sample plots, and put true AGB of other sample plots and weights into fomula (2) to obtain the predicted AGB of the sample plots. The specific mathematical expression of a P-BSHADE is as follows (Hu et al., 2013;Xu et al., 2013):”**

*L. 192-208: The descriptions of all machine learning methods are hard to understand. I suggest to simplify and only explain the principles of each of them. Technical details are confusing if they are not explained at length. Since all are well known machine learning methods, I would refer to the original literature for details.*

**Response: Thank you for your kind advice. We have simplified the explanation of all machine learning methods as follows (Lines 190-194):**

**“The SVM is a categorized algorithm that improves the generalized machine learning ability by minimizing structural risks in order to minimize the empirical risk and confidence intervals (Drucker et al., 1996). The RBF-ANN provides the best approximation for nonlinear functions and optimal global performance (Elanayar and Shin, 1994). The RF is a relatively new machine learning technique that combines self-learning technologies (Breiman, 2001).”**

*L. 213: Is it mean DBH and mean tree height?*

**Response: Yes. Thanks for pointing this out. We added the word “mean” to avoid confusion (Line 199).**

*L. 215-227: The explanation of P-BSHADE is not sufficient to understand the method. Since this is not a standard method it disserves more explanation. Again, for technical details I would refer to the original literature, but the principle ideas should be clearly explained.*

**Response: Thank you for your kind advice. We have rewritten the method for the P-BSHADE model as follows (Lines 203-214):**

**“The P-BSHADE model is an interpolation approach based on the assumption of spatial autocorrelation and spatial heterogeneity. It is markedly different from the Kriging and Inverse Distance Weighting (IDW) algorithms, because the latter two algorithms only regard spatial autocorrelation as an assumption condition. Besides, P-BSHADE regards the strongly correlated sample plots as the neighboring plots. The core of the model is minimizing the variances of predicted error and unbiased estimation. The model is also a data fusion approach which could combine the observed samples with and reference series (related variable). In brief, the P-BSHADE includes two steps. First, obtain reference AGB of all sample plots according to the allometric model. Second, use the reference AGB of target sample plot and true of other sample plots to obtain the weights relationship between target sample plot and other sample plots, and put true AGB of other sample plots and weights into fomula (2) to obtain the predicted AGB of the sample plots. The specific mathematical expression of a P-BSHADE is as follows (Hu et al., 2013;Xu et al., 2013):”**

*Also, how is it related to existing methods like kriging and inverse distance weighted interpolation? What are the differences to them? Does distance between plots play a role and how many neighboring plots are considered for a prediction?*

**Response: Thank you. The P-BSHADE model is an interpolation approach based on an assumption of spatial autocorrelation and spatial heterogeneity. It is quite different from the Kriging and Inverse Distance Weighting (IDW) algorithms, which assume only spatial autocorrelation. In addition, P-BSHADE considers strongly correlated sample plots as neighboring plots, whereas the Kriging and Inverse Distance Weighting algorithms consider sites that are close in proximity. We have rewritten the method for the P-BSHADE model as follows** **(Lines 203-214):**

**“The P-BSHADE model is an interpolation approach based on the assumption of spatial autocorrelation and spatial heterogeneity. It is markedly different from the Kriging and Inverse Distance Weighting (IDW) algorithms, because the latter two algorithms only regard spatial autocorrelation as an assumption condition. Besides, P-BSHADE regards the strongly correlated sample plots as the neighboring plots. The core of the model is minimizing the variances of predicted error and unbiased estimation. The model is also a data fusion approach which could combine the observed samples with and reference series (related variable). In brief, the P-BSHADE includes two steps. First, obtain reference AGB of all sample plots according to the allometric model. Second, use the reference AGB of target sample plot and true of other sample plots to obtain the weights relationship between target sample plot and other sample plots, and put true AGB of other sample plots and weights into fomula (2) to obtain the predicted AGB of the sample plots. The specific mathematical expression of a P-BSHADE is as follows (Hu et al., 2013;Xu et al., 2013):”**

*L. 252: The number of patches (2980) should be mentioned here to make clearer that you are not speaking of the 30 plots here.*

**Response: Thank you for your kind advice. We added the number of patches (2,980) to avoid confusion (Line 257).**

*L. 258: D is probably not breast height, but diameter at breast height.*

**Response: Thank you for pointing out our mistake. We have corrected it and moved it to the Supplementary Material section** **(Lines 176-179) .**

*L. 256-261: This is not a validation! This is a comparison of AGB predicted with the optimal model to AGB predicted via a simple allometry. A validation would imply that the data used for comparison is (a close approximation of) the truth.* *Then it could be tested how close the predictions can get to the truth. But in the case here, the authors are not stating that they consider the simple allometric AGB to be the truth.*

**Response:** **Thank you for your comment, we agree on many points. To address this issue, we would have needed systematic sampling plots (forest inventory data) and corresponding non-destructive measurements for each tree in each plot in the entire region to validate. However, we don’t have the systematic sampling plots. In future studies, we will attempt to collect this data to validate the regional scale results. Based on this limitation, we revised the main research question in the current study (Lines 103-104) to, “Can the integration of spatial statistical and machine learning methods improve the accuracy of AGB models at the plot level,” and restructured our paper. To this end, we collected non-destructive measurements for an additional 22 independent plots and repeated the workflow of the optimal model to test the combined models (Lines 250-255). The section on regional scale AGB mapping has been shortened (Lines 257-260) and moved to the Supplementary Material (Lines 128-186).**

*It is not clear why they make this comparison and also not the role of the CIW.*

**Response: As we do not have full systematic sampling data for validation, the comparison and CIW in this article are not necessary, so we deleted these sections.**

*Results:*

*L. 264: Change Mg ha-1 plot-1 to Mg ha-1.*

**Response:** **Thank you, we have made this correction (Lines 264-265).**

*L. 280: Please use clear variable names in Fig. 4. E.g., “vol” and “volume” are ambiguous.*

**Response: Thank you, we corrected Fig. 4 based on your recommendation (Line 277).**

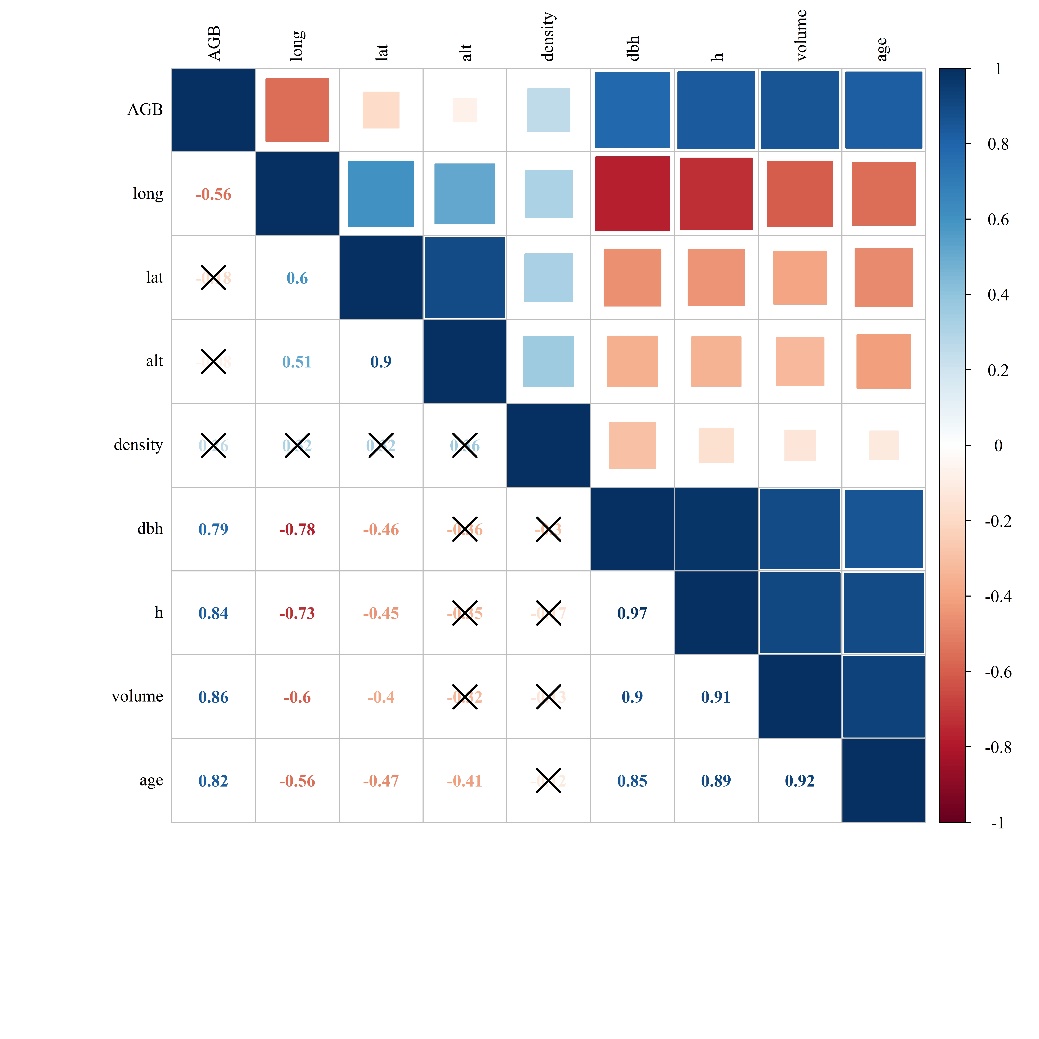
****

Figure 4. Pearson’s correlation coefficients between AGB and other variables represented by numbers and squares. Negative numbers represent that the corresponding variables are negatively correlated and are colored in red, while positive blue numbers represent positive correlations. Larger absolute numbers, darker colors, and larger squares all indicate stronger correlation, while “” indicates the correlation are insignificant.

*L. 296-300: To understand this a more detailed explanation of the method is required (see* *comment on L. 167). The k-means clustering was not mentioned before and a reader unfamiliar with the method does not know how to interpret q-values.*

**Response: Thanks for pointing this out. We added a brief explanation about the concept of spatial stratified heterogeneity and how the Geographical Detector method works in the Materials ＆ Methods section (Lines 155-162).**

**“The spatial stratified heterogeneity (referring to the characteristic that the within-strata variance less than the between strata-variance, it is ubiquitous in ecological phenomena, such as AGB) of the true AGB of sample plots was evaluated using a q-statistic value by the GeogDetector model, which is a software tool based on spatial variation analysis of the geographical strata of variables, as proposed by Wang et al. (2016). First, we used the K-means algorithm to obtain the strata of true AGB for preprocessing of GeogDetector. Then, we regarded true AGB as Y, regarded the strata of true AGB as X and put them into the GeogDetector model to obtain the q-statistics value (Wang et al., 2010;Wang et al., 2016).”**

*L. 308: This graphic nicely summarizes the main results.*

**Response: Thank you for your nice comment.**

*However, it raises two crucial questions, which are not answered in the text:*

*1) If S4 are interpolations based on the* *actual AGB values of the other plots and S5 to S7 are interpolations based on ML-predicted AGB of the other plots, how can S4 be worse than S5 to S7? The ML-predictions should introduce additional uncertainty compared to S4 which uses the actual AGBs. This has to be explained.*

**Response:** **Thank you for your constructive advice. As we explained in our response to the comment on L. 182, 215-227, we have added a detailed description of the P-BSHADE model (Lines 203-214). Taking S4 as an example, it includes three steps, as follows:**

1. **For each leave-one-out cross-validation of S4, we first obtained the reference AGB of the target sample plot using the plot-level allometric model (reference series).**
2. **We then used the reference AGB of the target sample plot and the true AGB of the other sample plots to calculate the weight relationship between the target sample plot and the other sample plots.**
3. **We input the true AGB and the weights of the other sample plots into a specific formula to obtain the predicted AGB of the target sample plots.**

**As for the combined methods, the reference AGB of the target sample plot was obtained using corresponding machine learning methods. The differences of S4-S7 methods were only the reference series, the more accurate reference AGB, the more accurate predicted AGB. The reason may be that the allometric model at the plot level cannot express complex non-linear changes, so the accuracy of the machine learning method with good non-linear fitting ability may be better than the allometric model and the combined algorithm (S5-S7) would be better than P-BSHADE alone (S4).**

1. *If we accept the fact questioned in my question 1), it remains the question why the* *combined approach can strongly improve the SVM and RF approaches, but not the* *ANN approach (S6 is hardly better than S2). This should be discussed.*

**Response: Thank you for your comment. We added this to the discussion in section 4.2 (Lines 421-431). The reasons for this may be: (1) the RF and SVM are easier to use and optimize than the RBF-ANN (Raczko and Zagajewski, 2017). In addition,** **RBF-ANN is sensitive to hyper-parameters and requires optimized parameters to obtain better fitting results. In this study, we did not use any optimized algorithms, such as the genetic algorithm, to obtain parameters in the machine learning model. Furthermore, the number of training samples determines the number of nodes in the hidden layer of RBF-ANN and the number of nodes significantly affects prediction accuracy. Only 30 training samples were used in this study, which may have affected the ability of the combined approach to improve the ANN approach. (2) RBF-ANN is more suitable for nonlinear stochastic dynamic systems (Elanayar and Shin, 1994). The relationship between AGB and environmental covariates is likely a monotonically increasing function.**

*It would be informative to also show* *1:1-scatterplots (AGB predictions vs. observations) for the different methods. And R2 values should be provided, because the given measures of error (RMSE, MRE, MAE) are uninformative with regard to whether there is any trend between predictions and observations at all. I strongly advise to provide R2 values of predictions vs. observations for methods S1-S7.*

**Response: Thanks for your kind advice. We have added a figure (Fig. 6) which shows 1:1-scatterplots (AGB predictions vs. observations) and R2 values for the different methods (Lines 333-349).**

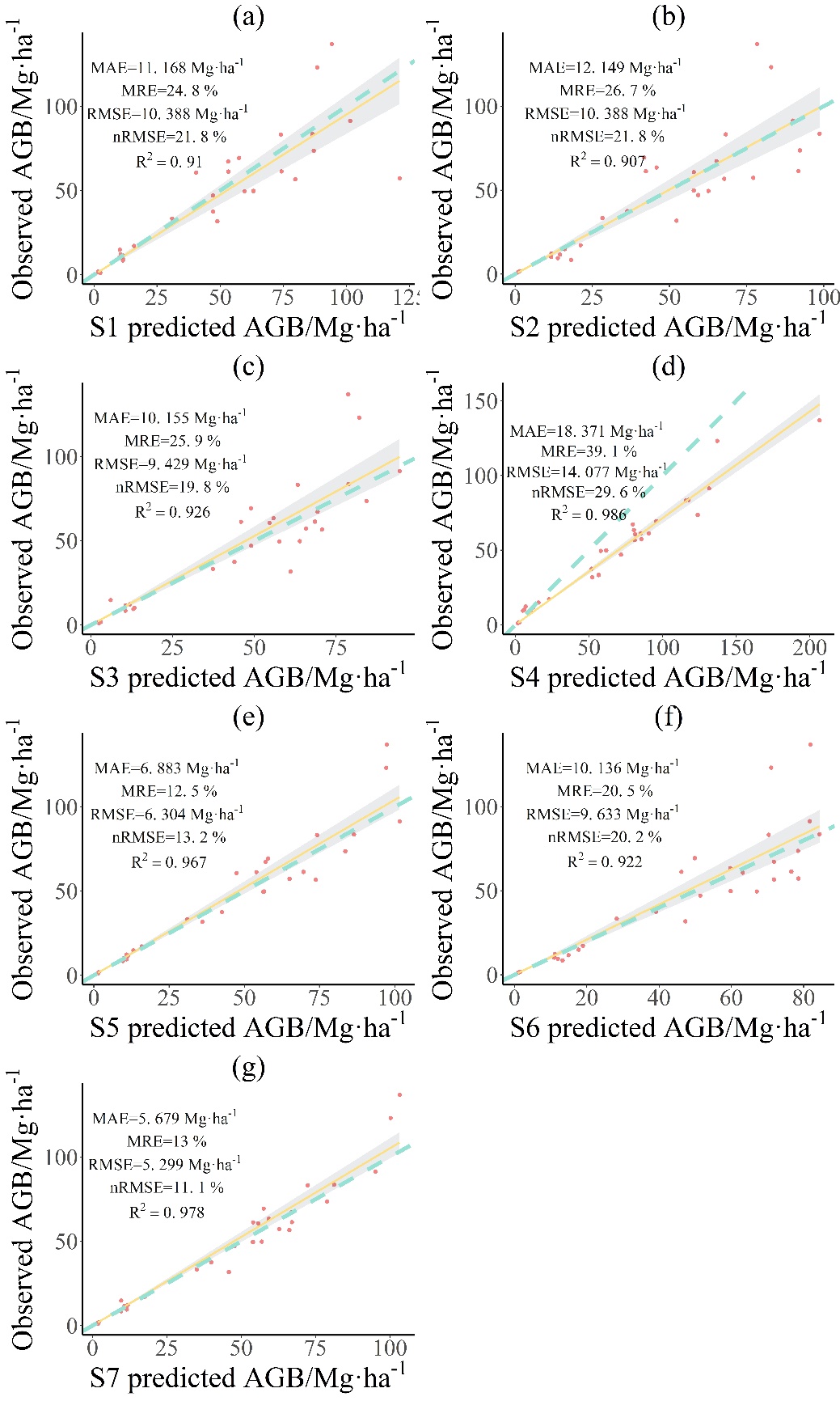
****

Figure 6 Comparisons of predicted and observed AGB for the accuracy assessment, (a), (b), (c), (d), (e), (f), (g) represent the SVM, RBF-ANN, RF, P-BSHADE, SVM&P-BSHADE, RBF-ANN&P-BSHADE, RF&P-BSHADE respectively, green dashed lines correspond to 1:1 relationship, each dot represents a sample plot individual, yellow solid lines indicate trend lines of dots.

*L. 343: The text in Fig. 6 is too small to read. Also the maps are too small for meaningful interpretation. I suggest to show only map (a) in large and to put map (b) to the appendix.* *The caption text “green outside of study area” is confusing. Green is part of the DTM color palette.*

**Response: Thanks for your kind advice. We have revised** **Fig. 6 based on your recommendation, as shown below. And it has been moved to Supplementary Material (Lines 218-219).**

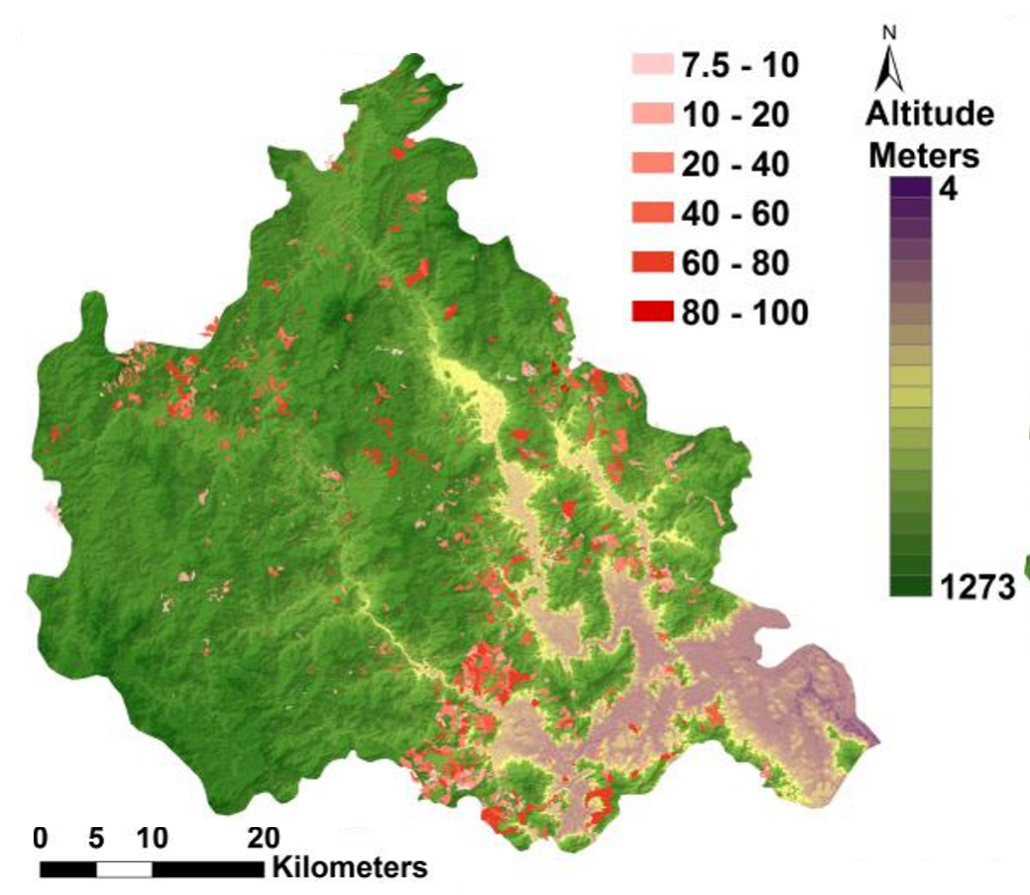


Figure C.3. Upscaling map of AGB using RF&P-BSHADE.

*L. 348-350: The total difference between the county-wide predictions of the optimal model and predictions using* *the simple allometric approach is only 0.17%. Does this mean that the methodology presented in this study is not needed?*

**Response: Thanks for pointing out this issue. Indeed, total AGB** **obtained** **from our approach is similar to that from the simple allometric** **approach. However, total AGB does not reflect** **the** **spatial distribution of AGB, so differences between different approaches can be concealed by total AGB. Although overall estimates based on regional scale carbon sequestration maps are similar, substantial differences exist in spatial distribution in maps of local areas and spatial distribution of AGB is critical to mapping accurate AGB. Based on your questions and suggestions, we added a description of the spatial differences between the results of the two approaches in the results section (Lines 363-365), as follows:**

**“The relative percent difference in total AGB between the two methods is 0.17%. Meanwhile, the MRE of AGB between the two methods ranged from 0.04% to 99.8% with an average of 19.93%.”**

*Discussion:*

*L. 418: This section does not answer the questions raised in the comment on L. 308.*

**Response: Thanks for pointing out this problem. Please refer to our previous responses to the comments on L.182 and L.308. We have revised our research question (Lines 103-104), as follows:**

**“Can the integration of spatial statistical and machine learning methods improve the accuracy of AGB models at the plot level.”**

*L. 435-469: Comparing absolute RMSE values from different studies representing very different forest types is not meaningful, because the intrinsic AGB variability is very different between, e.g., a tropical rainforest (large) and a Eucalyptus plantation (small). If anything, normalized* *nRMSE (i.e., RMSE divided by the mean) should be compared.*

**Response: Yes, you’re right, we agree. We have revised our comparison of RMSE and nRMSE (Lines 458-469), as follows:**

**“Based on calculated nRMSE indexes, the AGB prediction accuracy of the combined RF & P-BSHADE method (11.13%) was higher than that obtained by Babcock et al. (2015) (33.91%) in Colorado, USA. In that study, the authors used a combination of airborne LiDAR, a forest inventory database, and a Bayesian spatial hierarchical framework model and introduced spatial random effects to compensate for the residual spatial dependence and non-stationarity of model covariates. The AGB prediction accuracy of the method developed in the current work was also higher than that obtained by Ioki et al. (2014) (nRMSE=26%) in northern Borneo using a stepwise linear regression model with airborne LiDAR and a ground survey. Furthermore, it was higher than the accuracy obtained by Hansen et al. (2015) in the tropical submontane rain forest (34.4%) using fusion maps of multi-source databases combined with multiple regression analysis. Our prediction performance is close to that obtained by Kim et al. (2016) (9.2%) who studied the Intact Tropical Rain Forest using a voxel-based method based on airborne LiDAR in conjunction with field monitoring in Brunei.”**

**References:**

**Babcock, C., Finley, A. O., Bradford, J. B., Kolka, R., Birdsey, R., and Ryan, M. G.: LiDAR based prediction of forest biomass using hierarchical models with spatially varying coefficients, Remote Sensing of Environment, 169, 113-127, 2015**

**Ioki, K., Tsuyuki, S., Hirata, Y., Phua, M.-H., Wong, W. V. C., Ling, Z.-Y., Saito, H., and Takao, G.: Estimating above-ground biomass of tropical rainforest of different degradation levels in Northern Borneo using airborne LiDAR, Forest Ecology and Management, 328, 335-341, https://doi.org/10.1016/j.foreco.2014.06.003, 2014.**

**Hansen, H. E., Gobakken, T., Bollandsås, M. O., Zahabu, E., and Næsset, E.: Modeling Aboveground Biomass in Dense Tropical Submontane Rainforest Using Airborne Laser Scanner Data, Remote Sensing, 7, 10.3390/rs70100788, 2015.**

**Kim, E., Lee, W.-K., Yoon, M., Lee, J.-Y., Son, Y., and Abu Salim, K.: Estimation of Voxel-Based Above-Ground Biomass Using Airborne LiDAR Data in an Intact Tropical Rain Forest, Brunei, Forests, 7, 10.3390/f7110259, 2016.**

*L. 477: How were the MRE values 0.04% to 99.8% and the average 19.93% calculated? This should be reported in the Materials & Methods and in the Results sections. It could also be visualized in a 1:1-plot.*

**Response: Thanks for pointing this out. Values 0.04% to 99.8% are relative errors and 19.93% is the mean of relative error (MRE). We have revised this sentence (Lines 491-192), as follows:**

**“Meanwhile, the relative error (RE) of AGB between the two methods ranged from 0.04% to 99.8% with an MRE of 19.93%.”**

**In view of our inability to address the uncertainty of regional AGB mapping due to non-representative plots, we revised the article, focusing on the combined model to improve the accuracy of AGB fitting. Therefore, the comparison of AGB predicted by the optimal model to AGB predicted by simple allometry has been moved to** **S1.6 of the Supplementary Material (Lines 167-186). We also added the formula used to calculate RE to this section (Lines 173-175).**

**“6 Model application and upscaling mapping of AGB**

**We applied the chosen optimal model to each Eucalyptus forest patch (2,980 patches) and estimated the total AGB for all patches in the study area. We regarded the irregular polygon forest patches of Forest Management and Planning Inventory (FMPI) as a homogenous sample plot and applied the optimal plot-level model to the FMPI for the upscaling of forest AGB. We compared this upscaling of forest AGB with the AGB map obtained by an allometric model, relative error (RE) (see A.9) of AGB between the two methods was calculated. The allometric model was expressed as the formula , where DBH is the diameter at breast height (m), H is the tree height (m), and a and b are constants. This model is acknowledged as a fast, simple, and basic method to calculate regional AGB. In our study, we used the AGB, mean H, and mean DBH of 30 sample plots to constitute the plot-level allometric model.**

**(A.9)**

**where is the predictive AGB values of each irregular polygon forest patch by the optimal model, is the predictive AGB values of each irregular polygon forest patch by allometric model.”**

**And we added Figure C.4 to present the 1:1 plot in the Supplementary Material (Lines 220-223), as shown:**

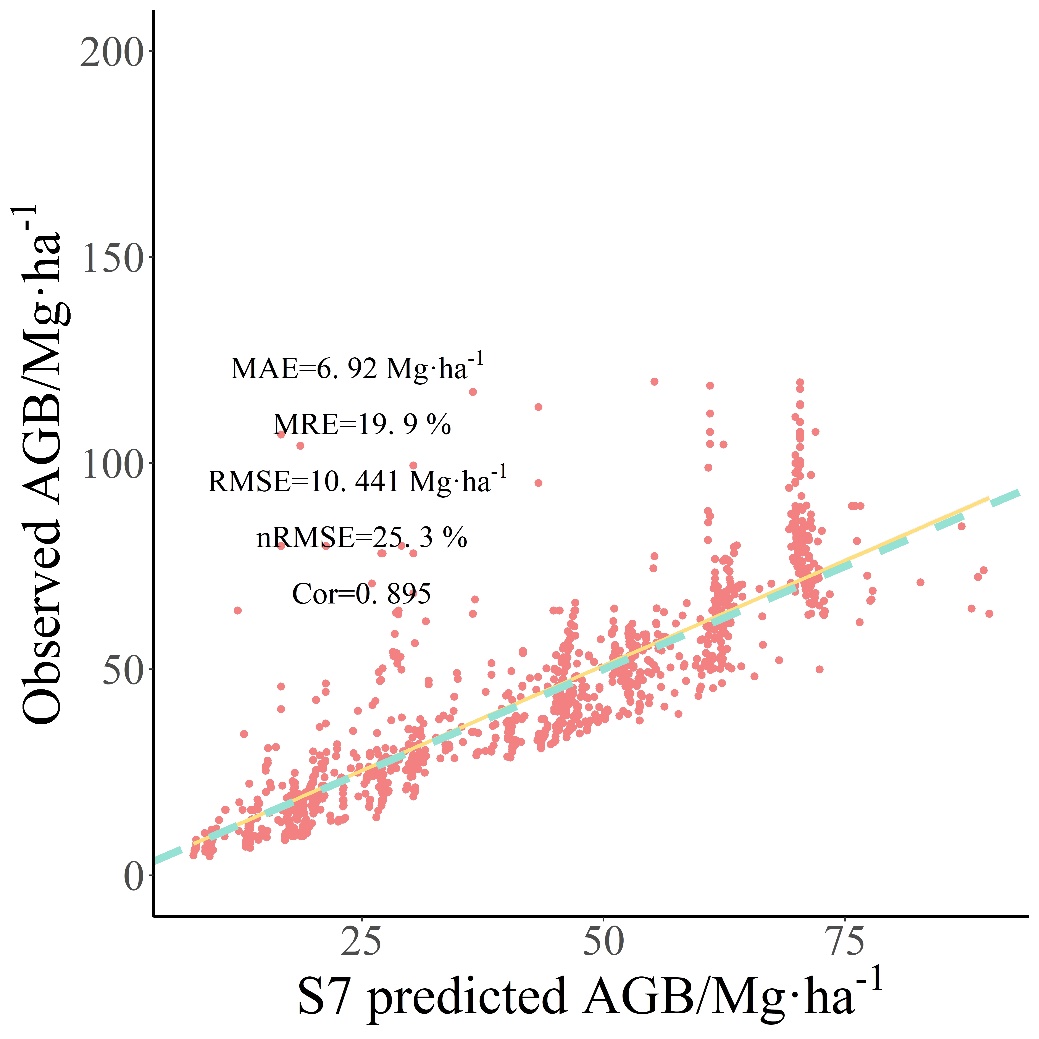
****

Figure C.4 Comparisons between upscaling of RF&P-BSHADE with the allometric model, green dashed line corresponds to a 1:1 relationship, each dot represents a forest patch individual, yellow solid line indicates trend line of dots.

*A fundamental question arising from this comparison is, how can we know whether RF&P-BSHADE or the allometric approach is closer to the truth? This would require independent ground-truth plots in other parts of the study area. In fact, an answer to research question (2) raised in the Introduction would require independent ground-truth plots. Without them I see no justification for the claim that the presented methodology can improve the accuracy of AGB mapping in regions where only non-representative sample units are available (stated in L. 27).*

**Response: Yes, you’re right, we agree. To address this issue, we would have needed systematic sampling plots (forest inventory data) and corresponding non-destructive measurements for each tree in each plot in the entire region to validate. However, we don’t have the systematic sampling plots. In future studies, we will attempt to collect this data to validate the regional scale results.**

**Based on this limitation, we revised the main research question in the current study (Lines 103-104) to: “Can the integration of spatial statistical and machine learning methods improve the accuracy of AGB models at the plot level.” We therefore restructured our paper based on this idea. To this end, we collected non-destructive measurements for an additional 22 independent plots and repeated the workflow of the optimal model to test the combined models (Lines 250-255). The results are shown in Fig. 7 (Lines 350-356) and section 3.3 (Lines 345-349), as follows. The section on regional scale AGB mapping was shortened** **(Lines 257-260) and moved to the Supplementary Material (Lines 128-186).**

**“We compared three machine learning methods with three corresponding combined machine learning and spatial statistical methods by the change of MAE, MRE, RMSE, nRMSE in two periods, 2012 and 2019 (Figure 7). The result showed that the combined machine learning and spatial statistical methods could both improve the accuracy of single machine learning method when based on different sample plots from two periods (2012 and 2019). The combined methods are robust and optimal.**

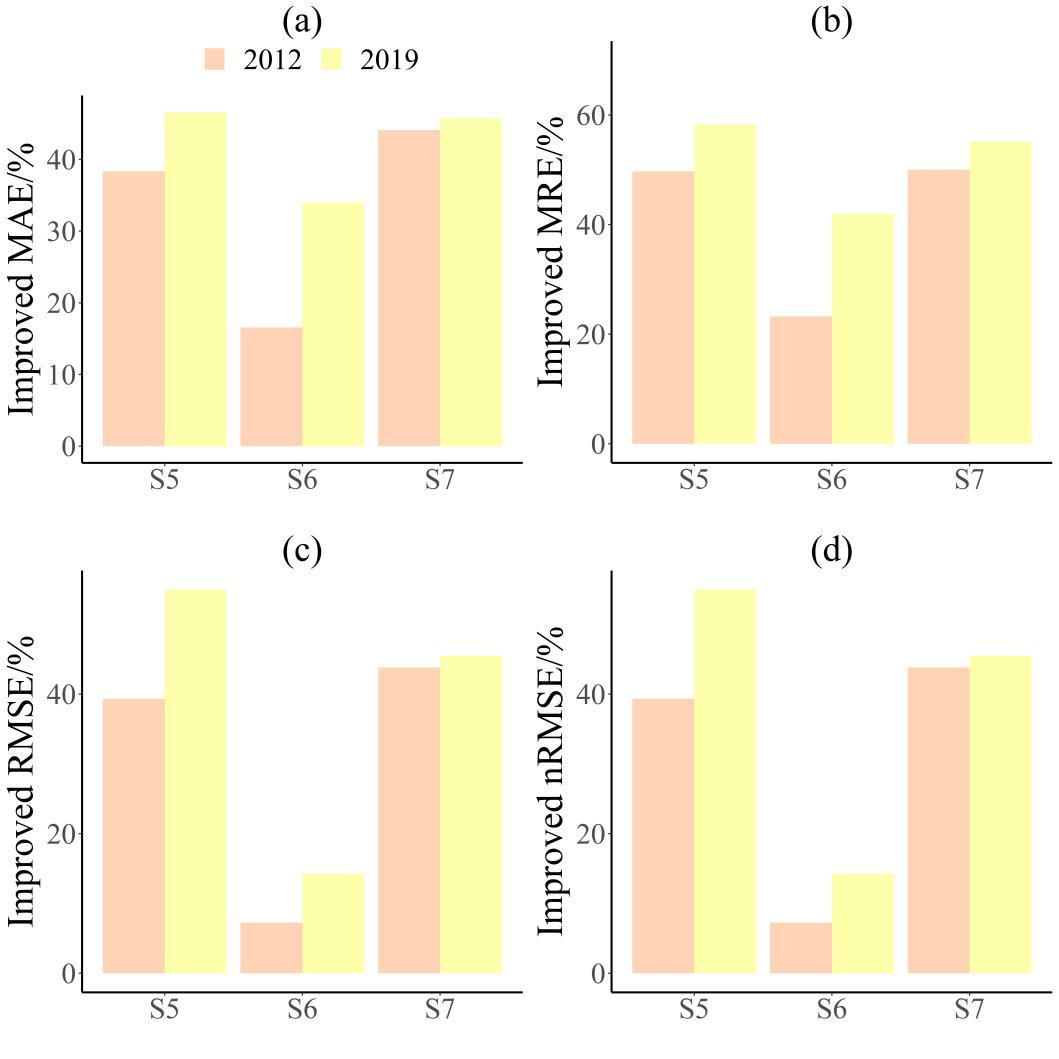


Figure 7 The improvement in accuracy assessment indexes of three combined machine learning and spatial statistical methods by comparison with three corresponding machine learning methods, (a), (b), (c), and (d) represent the MAE, MRE, RMSE, and nRMSE respectively; S1-S5 represents RMSE comparison of S5 with S1, S2-S6 represents RMSE comparison of S6 with S2, S3-S7 represents RMSE comparison of S7 with S3 (S1=SVM, S2=RBF-ANN , S3=RF, S4=P-BSHDE , S5=SVM & P-BSHDE, S6=RBF-ANN & P-BSHDE, S7=RF & P-BSHDE).”

*The Discussion and* *Conclusion are in general very broad and should be* *more specific with regard to the results in this study.*

**Response: Thanks for pointing out this issue. The Discussion and Conclusions are very broad because we believe the method presented here can be applied to different research fields and different regions, even though we used a local case to verify our method. But to also address your point, we have added additional discussion of key results to these sections.**

**Changes to the Discussion section (Lines 415-431):**

**“The three combined machine learning and spatial statistical methods produced higher AGB prediction accuracy than any method individually. The accuracy of the RF & P-BSHADE and SVM & P-BSHADE methods were significantly higher than the individual methods, but the RBF-ANN & P-BSHADE method was only slightly higher. The accuracies of the combined methods depended on the accuracy of the reference series (machine learning predicted result) (Xu et al., 2013). In other words, the higher the accuracy of the predicted machine learning results, the higher the accuracy of the combined method. Therefore, the following probably reasons about different improvements of three combined methods maybe: (1) the RF and SVM models are easier to use and optimize than RBF-ANN (Raczko and Zagajewski, 2017). RBF-ANN is sensitive to hyper-parameters and usually requires optimized parameters to obtain better fitting results. However, in this study, we did not use any optimized algorithms, such as the genetic algorithm method, to obtain parameters in the machine learning model. Furthermore, the number of training samples determines the number of nodes in the hidden layer of the RBF-ANN model, and the number of nodes significantly affects the prediction accuracy. With only 30 training samples used in this study, the combined approach may have been unable to strongly improve prediction accuracy. (2) RBF-ANN is more suitable for nonlinear stochastic dynamic systems (Elanayar and Shin, 1994), whereas the relationship between AGB and environmental covariates in this study is likely a monotonically increasing function.”**

**Changes to the Conclusions section (Lines 510-516):**

**“****In this study, we proposed a method to integrate the advantages of machine learning and spatial statistics, different datasets, and multiple environmental covariates to improve the accuracy of plot-level AGB estimation models. Using the most accurate data for harvested trees and sample plots, we explored the prediction performance of different methods in AGB modeling. Although data from the sample plots and harvested trees were collected only from Eucalyptus forests located in the Nanjing region of China, the proposed method and the findings can provide references for AGB mapping in other countries and in different types of tropical forests.”**

*Supplements:*

*L. 55-69: Please shorten strongly by removing all information not relevant for the study.*

**Response: Thank you, we revised this section by removing non-relevant information. The remaining information is relevant to the study (Lines 64-70), as follows:**

**“1.3 harvest-tree measurements**

**Aboveground biomass was divided into three tissue types: stems, branches, and foliage. Four to six branches were systematically sampled from each tree at regular intervals over the entire crown length. Foliage was collected from each of the sampled branches. The stem was sectioned into meter-long pieces by using a chainsaw.**

**The fresh weight of three tissue types was obtained in the field, and 500 g of each sample (i.e., stems, branches, and foliage) were placed in plastic bags. The samples were maintained under refrigeration during transportation to the laboratory. Fresh samples were oven dried at 85 °C to determine the constant dry weight.”**

*L. 70-115: Please shorten strongly (see comment on L. 192-208 in main text). Please don’t repeat text from the Materials & Methods section. Equations are confusing if they are not explained in full detail. E.g., what are w and T in (A.1) etc. In L. 103 a T is mentioned which does not appear in the equations above. In conclusion, I recommend avoiding all equations and explain the mechanics of the different ML methods with words.*

**Response: Thanks for pointing this out. We have revised this section (Lines 72-98), as follows:**

**“2.1 Support vector machines for regression**

**A support vector machine (SVM) is a type of categorized algorithm that improves generalized machine learning ability by minimizing structural risks in order to minimize empirical risk and confidence intervals. In this way, it achieves adequate statistical trends from a limited number of samples. Compared with traditional machine learning methods, SVM adopts the principle of minimizing structural risks. Along with minimizing sample point errors, SVM simultaneously narrows the upper bound of generalized error in the model to improve the generalization ability of the model and to solve the problems of excessive model learning, nonlinearity, and dimensionality (Ukil, 2002).**

**The SVM classification model was trained using a C-classification method, with longitude, DBH, tree height, and forest age as the selection characteristics and the biomass data from the 30 plots as model training samples. The Gaussian inner product function served as the kernel function.**

**2.2 Radial basis function artificial neural networks**

**The basic components of radial basis function artificial neural networks (RBF-ANNs) include an input layer, a hidden layer, and an output layer, which are able to provide the best approximation for nonlinear functions and optimal global performance (Elanayar and Shin, 1994). The change from the input layer space to the hidden layer space is nonlinear, whereas the spatial transformation from the hidden layer to the output layer space is linear. The RBF-ANN has good generalizability, requires fewer calculations, and has a faster learning speed than other machine learning algorithms. Therefore, the RBF-ANN avoids lengthy iterative calculations, such as those found in the learning algorithms of back propagation neural networks, and the possibility of falling into a local extremum. RBF-ANN is widely used in many fields, including meteorology (Nath et al., 2016), soil (Zakian, 2017), vegetation (Hilbert and Ostendorf, 2001), and engineering control (Sarimveis et al., 2004).**

**2.3 Random forest**

**The random forest (RF) algorithm model is a relatively new machine learning technique and data mining method developed by Breiman in 2001. It is a modern classification and regression technology that combines self-learning technologies (Breiman, 2001). In order to achieve a better performance than individual classifiers, combinatorial learning approaches integrate several individual classifiers to determine the final classification of a case. If a single classifier is considered as a decision maker, the method of combinatorial learning is equivalent to a decision-making process involving multiple decision makers.”**

*L. 116-171: The explanation of* *P-BSHADE is very technical and non-intuitive. The overview should contain a simple explanation of what it does and how it compares to kriging or IDW. The spatial aspect is not clear at all. What role do plot positions and distances play?*

**Response: Thanks for pointing this out. We revised our introductory overview about P-BSHADE (see Lines 100-114 and response to *L*. 215-227). P-BSHADE uses the covariance of the value of observed samples and target samples (reference series) instead of distance to identify neighboring plots for prediction. Therefore, positions and distances between plots do not apply here.**

*L. 120: What does temperature have to do with it?*

**Response: We’re very sorry for the confusion. Temperature is not relevant to this study, so we removed the word from this section** **(Line 103).**

*L. 188: Table B.4: What is variable coefficient? Do you mean* *coefficient of variation (CV)? Was longitude used with two decimal precision only? Given the narrow range 117.45 to 117.5, the precision should probably be more than two decimals.*

**Response: We apologize for this mistake, and thank you for your kind advice. Yes, we meant coefficient of variation (CV). We have revised this in Table B.4 (Lines 202-204), as follows:**

Table B.4 Statistical description of AGB and selected variables for sample plots.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variables | Mean | Median | Standard deviation | **Coefficient of variation** | Minimum | Maximum |
| Aboveground biomass, AGB (t/ha) | 47.34 | 46.64 | 34.46 | 0.73 | 1.02 | 135.79 |
| Longitude, long | 117.48 | 117.47 | 0.02 | 0.13\*10-5 | **117.446** | **117.503** |
| Diameter at breast height, DBH (cm) | 12.29 | 13.19 | 4.48 | 0.36 | 2.19 | 17.99 |
| Tree height, h (m) | 12.98 | 14.42 | 4.72 | 0.36 | 2.83 | 18.23 |
| Age (a) | 5.5 | 5.5 | 2.92 | 0.53 | 1 | 10 |

*L. 197: Fig. C.1: Please increase font size. Labels are missing at the x- and y-axis.*

**Response: Thank you for noticing these errors. We have revised Fig. C.1 in S3 of the Supplementary Material (Lines 214-215).**

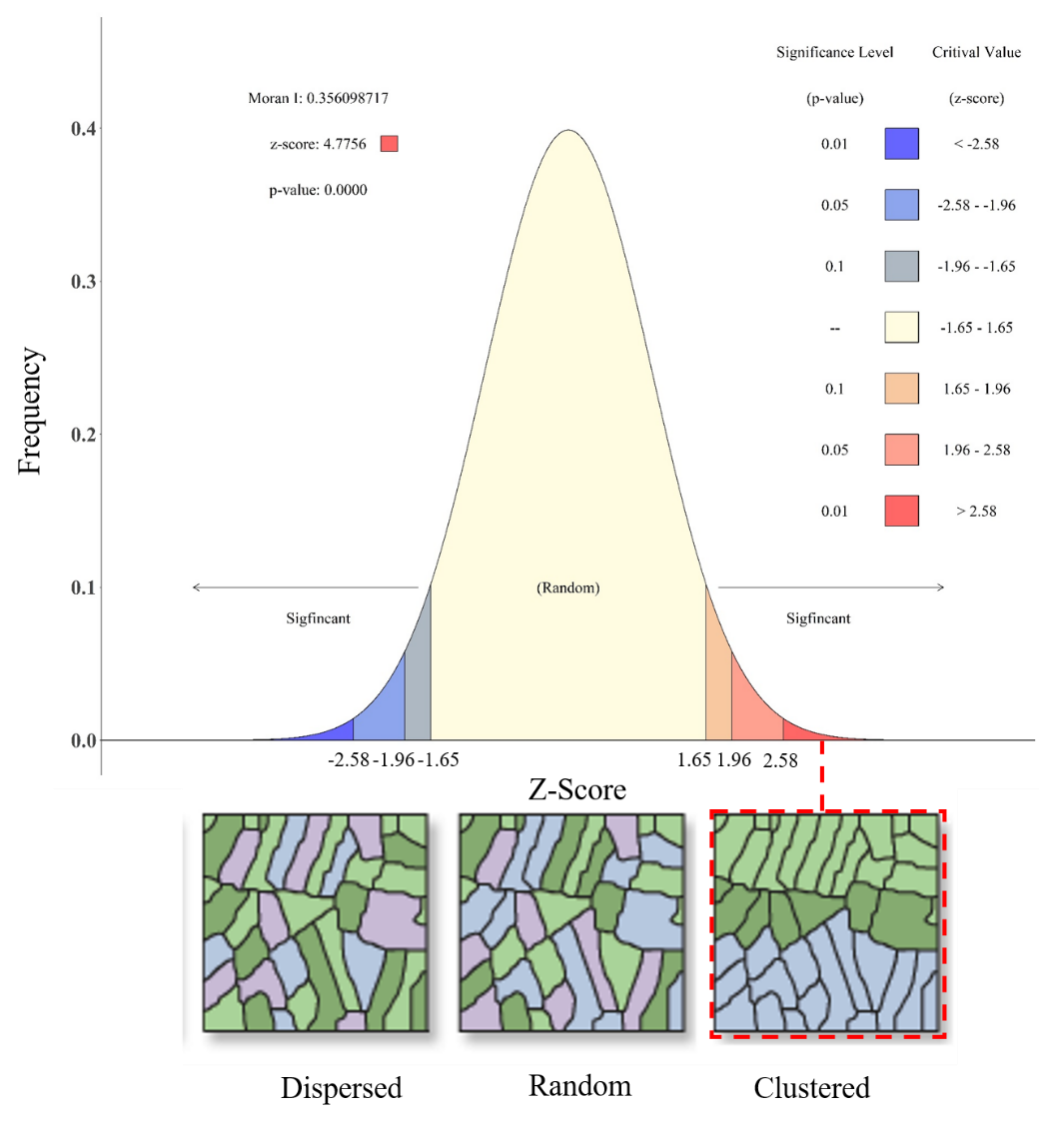


Figure C.1 Spatial autocorrelation report.