Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2020-36-SC1, 2020

Biogeosciences Discussions
Open Access
EGU

Interactive comment on "Improving maps of forest aboveground biomass: A combined approach using machine learning with a spatial statistical model" *by* Shaoqing Dai et al.

Wenli Huang

wenli.huang@whu.edu.cn

Received and published: 7 March 2020

This manuscript attempts to improve AGB maps by combining the machine learning model (SVM, RF, and RBF-ANN) and spatial statistical model (P-BSHADE). Overall the manuscript seems technically sound and, in most cases, is well written. The experiment is designed for one type of forest (Eucalyptus forest), thus I'm afraid the influence is limited. The results are reasonable, I question a number of aspects of the source data. Based on the comments below I suggest major revisions.

General questions: Q1: Your number of sample plots (N=30) is too small for machine learning models. How could the sample plots represent the region? The range of

Printer-friendly version

Discussion paper

biomass (1.02 – 135.79 Mg/ha). Did this range cover the whole range of your target species? The age of harvest trees ranged from 1 to 10 yr. What's the age range in the study area? What's the DBH range of your 90 harvested trees? Same questions as biomass and age.

Q2: How did you combined machine learning and spatial models? I got confused after reading your descriptions. Comparing to other methods, your description of the P-BSHADE model a little bit lengthy, suggest moving details into the supplement.

Q3: Your results suggest that plot-level biomass models need to be built per species and per ecoregion? The problem of not using allometric models is, how to quantify the AGB of not-so-common species?

Q4: Did you consider the influence of plot size? Say could your model build using a 20-m plot applied to 40-m or 100-m scale? This important when considering the need to apply models at larger geographical domains via the combined use of remote sensing datasets.

Q5: This study constructs local AGB allometric models, for a small Eucalyptus forest in Nanjing county. However, how should we apply your method in other places over a large geographical domain?

Q6: Did you compare your models and existing allometric models within the region? What's the influence of excluding small stems (living stem <8 cm) in your estimation of AGB?

Specific comments: L49: "the use of inadequate sampling data to construct the plot level prediction models" did you solve this issue?

L59: Selection of the allometric model could account for 20% uncertainty (Duncanson et al. 2017) Duncanson, L., Huang, W., Johnson, K., Swatantran, A., McRoberts, R., & Dubayah, R. (2017). Implications of allometric model selection for county-level biomass mapping. Carbon Balance and Management, 12

Printer-friendly version

Discussion paper

L82: Some recent studies integrated ground-based plot and remote sensing data for AGB mapping (Sun et al. 2011; Huang et al. 2019; Qi et al. 2019) Sun, G., Ranson, K.J., Guo, Z., Zhang, Z., Montesano, P., & Kimes, D. (2011). Forest biomass mapping from lidar and radar synergies. Remote Sensing of Environment, 115, 2906-2916

Huang, W., Dolan, K., Swatantran, A., Johnson, K., Tang, H., O'Neil-Dunne, J., Dubayah, R., & Hurtt, G. (2019). High-resolution mapping of aboveground biomass for forest carbon monitoring system in the Tri-State region of Maryland, Pennsylvania and Delaware, USA. Environmental Research Letters, 14, 095002

Qi, W., Saarela, S., Armston, J., Ståhl, G., & Dubayah, R. (2019). Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data. Remote Sensing of Environment, 232, 111283

L85: "First, existing studies that used machine learning methods have not considered the spatial heterogeneity of multiple environmental covariates (such as longitude, latitude, and forest structure)" This statement is too arbitrary. What does "structure" refer to? Shouldn't structure information came from lidar or radar?

L96: "multiple environmental covariates (such as longitude, latitude, and forest structure)" A duplicate statement, modify to be concise;

L140: Suggest add equations of the allometric models you used here.

L175: What software/package did you applied to construct your model?

L179: (reference series)?

Figure 3. (b) SVM and (c) RF are for classification, not regression;

L445: "we" should be "We".

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2020-36, 2020.