Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2020-36-RC2, 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.

Anonymous Referee #2

Received and published: 16 May 2020

Dear editors, dear authors,

The manuscript 'improving maps of forest aboveground biomass: A combined approach using machine learning with a spatial statistical model' by Dai et al. presents a new approach to predict more accurately the Aboveground biomass.

They do so by combining a statistical approach, the P-BSHADE model, with machine learning models. They claim that the joint model approach is superior to machine learning models and the P-BSHADE model alone.

I found the general ideas to use machine learning for such a predictive task and to combine machine learning models with statistical models very appealing and I think that the

Printer-friendly version

Discussion paper

community would benefit strongly from an approach capable of predicting accurate the AGB at scale.

General comments:

As I am not part of the remote sensing community I cannot say much about the novelty of this approach, but I have a good experience with machine learning methods and I will focus on the methodological part of this manuscript.

a) I must say that the paper is sloppy in multiple respects, many statements about machine learning are inaccurate and partly wrong. Especially in the third paragraph of their introduction (L60-L73), many of their claims about machine learning are only partially true or are confusing (see specific comments): they say 'nonparametric machine learning algorithms, in which the number of parameters depends on the number of training examples', however, if they are nonparametric how can their number of parameters depend on anything? Also, the authors use RF, ANN, and SVM in their work as regression models but why do they explain and illustrate them as classifier? (see method section and Fig. 3). In summary, the authors should carefully revise all their statements about ML and explain correctly their used ML models.

b) They claim that the joint model combines the advantages of ML and the P-BSHADE model, the predictive non-linearity advantage of ML and the ability of the P-BSHADE to capture spatial relationships. However, if they are given the chance I think that ML models are also capable of detecting and using spatial relationships, that means, you have to provide them not only longitude but also latitude as predictor! Based on correlation with AGB, the authors selected only longitude, however, I would assume that an interaction of longitude and latitude would be a good predictor of spatial relationships (two variables of an interaction can show by themselves low correlation). Moreover, ML models such as RF are outstanding in detecting interactions and higher-order interactions (if they are given the chance). Also, hyper-parameter tuning is important in ML to improve predictive performance, even for RF! (e.g. see Probest et al., 2019

https://doi.org/10.1002/widm.1301). I recommend that the authors re-evaluate the performance of the ML models with hyper-parameter tuning, nested cross-validation, and additional predictors (at least latitude).

c) I found it very distressing that section about the P-BSHADE model (L241-280) was taken almost literally from a previous work by one of the co-authors (Xu et al., 2013)! An illustration: Xu et al. 2013 (https://doi.org/10.1175/JCLI-D-12-00633.1): 'This equation is generally valid for a nonhomogeneous condition. Clearly, determination of yˆ0 requires calculation of coefficients wij (. . .), which is addressed in the following section'. . . in the MS: 'This equation is generally valid for nonhomogeneuous conditions. Clearly, the determination of bij requires calculating the coefficients wij (. . .), which is addressed in the following section.'

Specific:

L 28: I suggest that you re-position the following sentence in abstract. 'The study was conducted' should come after the introduction of our methods

L39: Sentence is redundant

L52: 'the present study. . .'

L63-L65: not the development of computer-science techniques but the advances in hardware are responsible for the popularity of ML. Most of the ML techniques are quite old (e.g. Artificial neural networks, even CNNs).

L65: "which summarize data with a fixed number of parameters based on sample size". This statement is inaccurate or even wrong (it is difficult to understand the authors' intention). A) "summarize data" is wrong, or what do you mean? B) fixed number of parameters based on sample size; I think this is wrong because fitting 'parametric' models with p » n is a common task/problem with well-known solutions (e.g. regularization/elastic net). Moreover deep neural networks are highly parametrized models and are not 'non-parametric'. Here, you should focus on the distinction between linear

and non-linear models.

L68: remove ‘nonparametric’ and use non-linear

L68-69: What do you mean with that the number of parameters on the number of training examples? This statement conflicts with your previous statement L65 and why or how does the number of parameters depend on n in kNN, SVM, and RF?

L70: restrict variable types? A linear regression is not restricted to specific data types. Actually, kNN and SVM require the same contrasts as a linear regression and only RF is able to handle non-contrasted categorical predictors

L71: What do you mean with the distribution of predictor variables? I think that kNN and SVM are indeed affected by the distribution of the predictors because they use distance measurements.

L90: Why? Or at least provide a reference

L97: Please provide example references

L187: ‘Each model was trained on 30 datasets. . .’ But within the CV, right? So it should be ‘trained on 29 datasets’.

L199-200: Not exactly true, the activation function makes the transformation linear or non-linear. The fundamental matrix multiplication is a linear function.

L197-199: Is there a reason you use an RFB ANN? You could also use a normal DNN with relu activation functions and several hidden layers. Could you also explain the RBF-ANN in detail? I think that most users do not know how the RBF function is used by RBF-ANN.

L205: Wrong, RF can overfit, I thin with the law of large numbers you refer to the number of trees which is true that increasing the number of tree does not increase the generalization error. But if we assume that we have a sufficient number of trees, RF can overfit.

L227: What are these “obvious advantages”?

L248: ‘when j = 1, i = 2, 3, . . .30: when j = 1, i = 1, 3, 5, . . ., 30)’ What do you mean?

L241-280: The description of the P-BSHADE model is too close to the original work Xu et al. 2013! Either you rewrite it completely in your own words, or which I suggest is that you try to summarize the method and move a detailed description to the Appendix. The P-BSHADE model is not the focus of your work.

L312: ‘A detailed description of the combined models. . .’ is missing in the Supplementary Material

L347: Because of a low correlation you did not choose latitude, however, I hypothesize that the interaction between longitude and latitude has an effect on AGB, which you would not see in the correlation table unless to test explicitly the correlation between the interaction and the AGB.

L479: ‘Machine learning models appear adept at tackling high-dimensional problems.’ Yes, they do, but you do not have a ‘high-dimensional problem’.

L511: ‘. . . regression trees . . .’ this applies only for RF and I wonder if the RF really suffer from a skewed response distribution. Do you have a reference?

L568: RF and SVM are also sensitive to hyper-parameters. It is myth that RF does not need hyper-parameter tuning (see https://doi.org/10.1002/widm.1301).

Figures:

Fig 3: The sematic figures of the ML do not fit to the way you used them in your work: RBF-ANN, RF, and SVM are illustrated as classifier with 4, 3, and 2 response classes, however, you use them as regression models (e.g. one output node for the RBF ANN, no majority voting for the RF etc.)

Fig 6: Revise your color choice, e.g. it is difficult to distinguish between the yellow and blue line.

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

Printer-friendly version

Discussion paper