

Institute of Urban Environment,
Chinese Academy of Sciences,
CN 361021, China



Dr. Yin Ren
Email: yren@iue.ac.cn
5th Jun 2020

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. (*Response to Anonymous Referee #2*)

Shaoqing Dai et al.

yren@iue.ac.cn

Dear Editor,

Thank you for your email and the reviewer’s comments concerning our manuscript entitled “Improving maps of forest aboveground biomass: A combined approach using machine learning with a spatial statistical model” (ID: bg-2020-36). We thank the reviewer for the very helpful comments.

Detailed reply to reviewer's comments:

Comment 1: 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 community would benefit strongly from an approach capable of predicting accurate the AGB at scale.

Response to comment 1: We very much appreciate the positive comments on our manuscript from Anonymous Reviewer #2. The reviewer's efforts have helped us greatly improve this paper.

General comments:

Comment 2: 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.

Response to comment 2: Thank you for your valuable comments, which have significantly improved our manuscript. Your opinions on the ML methods and their presentation in the paper are very useful and important. Thanks for pointing out the mistakes we made about machine

learning in the previous version. We have carefully and thoroughly reviewed and modified this article, especially the presentation of machine learning methods (see specific responses below).

The definitions of parametric and nonparametric models are sourced from Statistics and may be confused with some concepts in Computer Sciences. Here, we have carefully followed the definitions of ‘parametric models’ and ‘nonparametric models’ from a classical textbook on artificial intelligence titled ‘Artificial Intelligence: A Modern Approach (3rd edition)’. Page 737 of this book reads, ‘A learning model that summarizes data with a set of parameters of fixed size (independent of the number of training examples) is called a parametric model.’ Page 737 reads, ‘A nonparametric model is one that cannot be characterized by a bound set of parameters. For example, suppose that each hypothesis we generate simply retains within itself all of the training examples and uses all of them to predict the next example. Such a hypothesis family would be nonparametric because the effective number of parameters is unbounded — it grows with the number of examples. This approach is called instance-based learning or memory-based learning.’ kNN, SVM, and RF all belong to this instance-based learning approach.’ Page 758 reads, ‘Nonparametric models use all the data to make each prediction, rather than trying to summarize the data first with a few parameters. Examples include nearest neighbors and locally weighted regression.’

Nonparametric ML methods are called ‘nonparametric’ not because there are no parameters in the model; indeed, nonparametric ML models typically have one or more hyperparameters (external parameters of the model whose values cannot be estimated from the data) plus a number of common parameters (internal parameters whose values can be estimated using the data). Examples of nonparametric ML methods include k-nearest neighbors, decision trees, and nonlinear support vector machines.

Whether ANN is a parametric or nonparametric method is debatable. As the number of ANN parameters grows with the number of neurons/layers it is therefore considered as nonparametric in some references (e.g., Orr 1996, which is a basic and classical paper for radial basis function networks and has been cited more than 1000 times). This work follows Orr (1996) and considers RBF-ANN as a nonparametric model. We agree with you that we also need to describe the advantages of these nonparametric models in dealing with nonlinear fitting, especially in our forest ecology nonlinear systems.

References

Orr, M. J. L.: Introduction to radial basis function networks. University of Edinburgh, 1996.

Russell, S. J, & Norvig, P.: Artificial intelligence: a modern approach, 3rd edition, Pearson Publisher, 2016.

List of changes to machine learning:

1. The third paragraph of the Introduction; see the box immediately below ‘List of changes to machine learning’ and the response to comments 8–13.
2. The Methods section ‘(1) Machine learning’ part of Section 2.4.3 (‘Model training’); see the box immediately below ‘List of changes to machine learning’ and the response to comments 17–19.
3. We deleted the inappropriate statement about machine learning in Section 4.2 (‘Model comparisons’) of the Discussion, and modified references; see the box immediately below ‘List of changes to machine learning’ and the response to comments 25–27.
4. We have revised Figure 3; see the response to comment 28.

For your convenience, the revised text is reproduced below:

1. The third paragraph of the Introduction:

With advances in hardware and the further requirement of forest biomass modeling accuracy, machine learning methods have become prevalent in the field of forest biomass estimation. Traditional parametric methods, which summarize data with a fixed number of parameters regardless of sample size (e.g., logistic regression and perceptron) (Liu et al., 2020; Russell & Russell, 2016), usually require prior knowledge on a system to be modelled and assume a specific function for the system. Then the training data is used to estimate parameters of the presumed function. By comparison, nonparametric machine learning algorithms, which do not need sufficient understanding on the systems to be modelled, but seek the best fitting of the training data. They cannot be characterized by a fixed number of parameters (the number of parameters depends on the number of training examples, e.g., k-nearest neighbor, nonlinear support vector machine, and random forest). As a result, they are more elastic, can fit different function forms without prior knowledge, and achieve excellent fit performance (Russell and Norvig, 2016). Moreover, these nonlinear nonparametric models have advantages in dealing with nonlinear fitting and have great application potential in nonlinear systems such as forest ecology. Therefore, nonparametric machine learning algorithms may offer higher prediction

accuracy in forest AGB estimation (Frey et al., 2019; Gleason and Im, 2012).

Frey, U. J., Klein, M., and Deissenroth, M.: Modelling complex investment decisions in Germany for renewables with different machine learning algorithms, *Environmental Modelling & Software*, 118, 61-75, <https://doi.org/10.1016/j.envsoft.2019.03.006>, 2019.

Gleason, C. J., and Im, J.: Forest biomass estimation from airborne LiDAR data using machine learning approaches, *Remote Sensing of Environment*, 125, 80-91, 2012.

Liu, B., Gao, L., Li, B., Marcos-Martinez, R., and Bryan, B. A.: Nonparametric machine-learning as an effective tool for mapping forest cover and exploring influential factors, *Landscape Ecology*, <https://doi.org/10.1007/s10980-020-01046-0>, 2020.

Russell, S. J., & Norvig, P.: *Artificial intelligence: a modern approach*, 3rd edition, Pearson Publisher, 2016.

2. The Methods section '(1) Machine learning', part of Section 2.4.3 ('Model training'):

SVM is a method of supervised learning in machine learning which is often used to solve classification problems and can also be used to solve regression problems. The basic principle of SVM for classification is to find a hyperplane in the feature space and separate the positive and negative samples with the minimum misclassification rate (Hearst et al., 1998). The principle of SVM for regression is very similar to that of SVM for classification, with only slight differences (Sayad, 2020). The SVM for regression keeps the main idea of minimizing error, individualizing the hyperplane which maximizes the margin, and part of the error is acceptable (SaedSayad, 2020). RBF-ANN is a three-layer neural network model which includes an input layer, a hidden layer (a Gaussian RBF is used here), and an output layer. The transformation from input space to hidden space is linear or nonlinear, whereas the transformation from hidden space to output space is linear. The function of the hidden layer is to map the vector from the indivisible low-dimensional linear state to the separable high-dimensional linear state so as to greatly accelerate the learning and convergence speed and avoid getting stuck in a local optimum (Elanayar and Shin, 1994). RF is a relatively new machine learning technique and data mining method which combines self-learning technologies and was developed by Breiman in 2001. RF is a combination of tree predictors such that each tree depends on the values of a random vector that is sampled independently and with the same distribution for all trees in the forest. RF is effective in prediction (Breiman, 2001).

Breiman, L.: Random forests, *Machine Learning*, 45, 5-32, 2001.

Elanayar, V. T. S., and Shin, Y. C.: Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems, *IEEE Transactions on Neural Networks*, 5, 594-603, 1994.

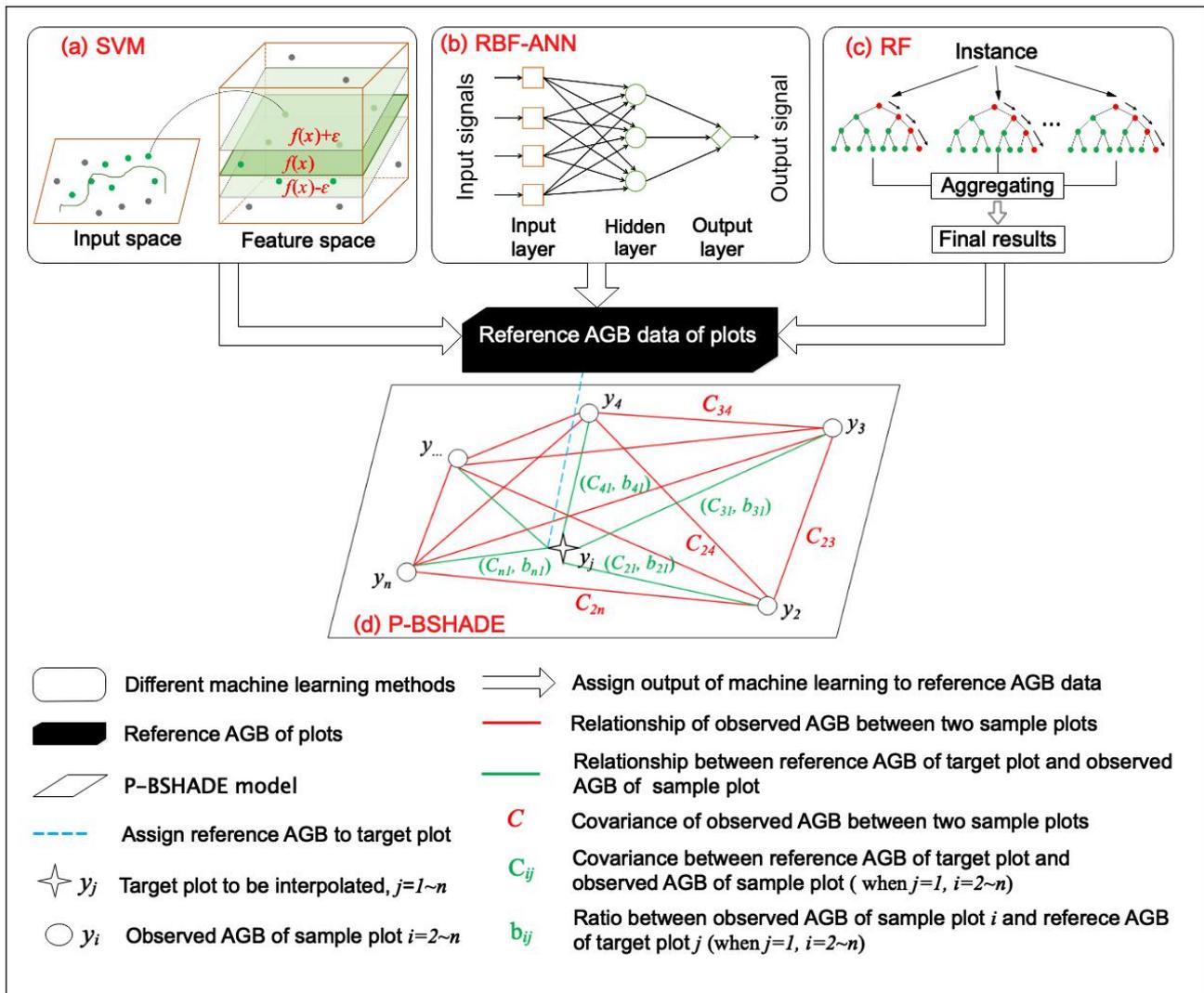
Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., & Scholkopf, B.: Support vector machines, *IEEE Intelligent Systems*, 13, 18-28, 1998.

Sayad S.: https://www.saedsayad.com/support_vector_machine_reg.htm, last access: 02 June 2020.

3. Modified reference:

Probst, P., Wright, M. N., Boulesteix, A-L.: Hyperparameters and tuning strategies for random forest. *WIREs Data Mining Knowledge Discovery*. e1301, <https://doi.org/10.1002/widm.1301>, 2019.

4. Revised Figure 3:



Comment 3: 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 model 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).

Response to comment 3: Many thanks for your valuable recommendations to improve this work. We agree that ML models can also detect and use spatial relationships. However, it is not the focus of this work to improve the performance through optimizing the structures of ML models or tuning hyper-parameters. As forest ecology modelers, we are more interested in the performance of combining machine learning with a spatial statistical method in forest AGB estimation as previous literature has shown that such a combined method can further improve the predictive performance in the spatial interpolation of environmental variables (Li et al., 2011). Therefore, the scientific question this work attempts to answer is: Given the limited expertise/knowledge of non-ML-community practitioners to further improve ML performance, can we further improve the predictive performance by combining machine learning with a spatial statistical method?

In the previous version, we did not include latitude as a predictor because it is controversial whether longitude and latitude should be used as predictors, and some research efforts found that including them could lead to considerable overfitting (e.g., Meyer et al., 2019). Moreover, as suggested, we have tested the performance of models by including both longitude and latitude as predictors, and compared the models with those that only include longitude or neither longitude nor latitude (see **response to comment 24** for detailed results). The results showed that including longitude and latitude as predictors did not improve the performance of RF, SVM, or RBF-ANN.

However, we acknowledge that your recommendations are valuable. We have discussed your recommendations in the revised manuscript, including them as future directions/limitations of this work. We have also read and cited the reference you kindly provided.

For your convenience, the revised text is reproduced below:

In this paper, we use a correlation analysis to determine which predictors are included to train a model, without considering whether the interaction among predictors has an impact on the forest AGB. Although we found that the interaction of longitude and latitude did have an impact on the AGB, our experiments also found that including them as predictors simultaneously did not improve the performance of the machine learning models. This may be due to two reasons: 1) we have not found the specific interaction relationship between longitude and latitude. In the future, we will test whether the different interaction patterns between these two predictors can improve the performance of machine learning in forest AGB estimation; and 2) in this work, we manually adjusted the hyper-parameters and did not fully explore the potential of machine learning methods. In future work, we will improve the performance of model

estimation through optimizing the structures of machine learning models or tuning hyper-parameters (Probest et al., 2019).

Li, J., Heap, A. D., Potter, A., and Daniell, J. J.: Application of machine learning methods to spatial interpolation of environmental variables, *Environmental Modelling & Software*, 26, 1647-1659, <https://doi.org/10.1016/j.envsoft.2011.07.004>, 2011.

Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T.: Importance of spatial predictor variable selection in machine learning applications - Moving from data reproduction to spatial prediction. *Ecological Modelling*, 411, <https://doi.org/10.1016/j.ecolmodel.2019.108815>, 2019.

Probst, P., Wright, M. N., Boulesteix, A-L.: Hyperparameters and tuning strategies for random forest. *WIREs Data Mining Knowledge Discovery*. e1301, <https://doi.org/10.1002/widm.1301>, 2019.

Comment 4: 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 y_0 requires calculation of coefficients w_{ij} (. . .), which is addressed in the following section'. . . in the MS: 'This equation is generally valid for nonhomogeneous conditions. Clearly, the determination of b_{ij} requires calculating the coefficients w_{ij} (. . .), which is addressed in the following section.'

Response to comment 4: We are deeply sorry for that. Lines 241–280 described the specific equations for the P-BSHADE. In order to make this new method understandable in our forest AGB estimate, we only changed the descriptions of specific symbols in the equations but did not modify other important descriptions, such as the sentences you listed above, in order to avoid misunderstanding. We apologize for this and accept your kind advice in comment 22. Therefore, we have summarized the P-BSHADE method in the manuscript and have given detailed information in the Supplementary Material.

Specific comments:

Comment 5: L28: I suggest that you re-position the following sentence in abstract. ‘The study was conducted’ should come after the introduction of our methods

Response to comment 5: Thank you for your professional advice. We have put this sentence after the introduction of the methods.

Comment 6: L39: Sentence is redundant

Response to comment 6: Thank you for your kind advice. We have deleted the sentence.

Comment 7: L52: ‘the present study. . .’

Response to comment 7: Thank you for your kind advice.

Comment 8: 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).

Response to comment 8: Thanks for your kind advice. We have revised this sentence as follows:

With advances in hardware and the further requirement of forest biomass modeling accuracy, machine learning methods have become prevalent in the field of forest biomass estimation.

Comment 9: 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 $p \gg n$ 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.

Response to comment 9: Thanks for this comment.

Here, we adopted the definition of ‘parametric models’ in a classical textbook on artificial intelligence titled ‘Artificial Intelligence: A Modern Approach (3rd edition).’ Page 737 reads, ‘A learning model that summarizes data with a set of parameters of fixed size (independent of the number of training examples) is called a parametric model.’ In the revised version, we have corrected the statement as ‘...which summarize data with a fixed number of parameters regardless of sample size.’

Whether ANN is a parametric or nonparametric method is debatable. As the number of ANN parameters grows with the number of neurons/layers, ANN is considered as nonparametric in some references (e.g., Orr 1996, which is a basic and classical paper for radial basis function networks and has been cited more than 1000 times). This work follows Orr (1996) and considers RBF-ANN as a nonparametric model.

We agree with you that we also need to describe the advantages of these nonparametric models in dealing with nonlinear fitting, especially in our forest ecology nonlinear systems. Therefore, this sentence (L65) has been revised. The revised content in our manuscript now reads:

Traditional parametric methods, which summarize data with a fixed number of parameters regardless of sample size (e.g., logistic regression and perceptron) (Liu et al., 2020; Russell & Russell, 2016), usually require prior knowledge on a system to be modelled and assume a specific function for the system. Then the training data is used to estimate parameters of the presumed function. By comparison, nonparametric machine learning algorithms, which do not need sufficient understanding on the systems to be modelled, but seek the best fitting of the training data. They cannot be characterized by a fixed number of parameters (the number of parameters depends on the number of training examples, e.g., k-nearest neighbor, nonlinear support vector machine, and random forest). As a result, they are more elastic, can fit different function forms without prior knowledge, and achieve excellent fit performance (Russell and Norvig, 2016). Moreover, these nonlinear nonparametric models have advantages in dealing with nonlinear fitting and have great application potential in nonlinear systems such as forest ecology. Therefore, nonparametric machine learning algorithms may offer higher prediction accuracy in forest AGB estimation (Frey et al., 2019; Gleason and Im, 2012).

References

Frey, U. J., Klein, M., and Deissenroth, M.: Modelling complex investment decisions in Germany for renewables with different machine learning algorithms, *Environmental Modelling & Software*, 118, 61-75, <https://doi.org/10.1016/j.envsoft.2019.03.006>, 2019.

Gleason, C. J., and Im, J.: Forest biomass estimation from airborne LiDAR data using machine learning approaches, *Remote Sensing of Environment*, 125, 80-91, 2012.

Liu, B., Gao, L., Li, B., Marcos-Martinez, R., and Bryan, B. A.: Nonparametric machine-learning as an effective tool for mapping forest cover and exploring influential factors, *Landscape Ecology*, <https://doi.org/10.1007/s10980-020-01046-0>, 2020.

Orr, M. J. L.: Introduction to radial basis function networks. University of Edinburgh, 1996.

Russell, S. J, & Norvig, P.: Artificial intelligence: a modern approach, 3rd edition, Pearson Publisher, 2016.

Comment 10: L68: remove 'nonparametric' and use non-linear

Response to comment 10: Thank you for reminding us of this point once again. We have revised it.

Comment 11: 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?

Response to comment 11: Thank you for this comment. The mistake we made in L65 has been revised as presented in the response to comment 9. The definitions of parametric and nonparametric models are sourced from Statistics and may be confused with some concepts in Computer Sciences. Here, we have carefully followed the definitions in classical textbook on artificial intelligence titled 'Artificial Intelligence: A Modern Approach (3rd edition)'. Page 737 reads, 'A nonparametric model is one that cannot be characterized by a bound set of parameters. For example, suppose that each hypothesis we generate simply retains within itself all of the training examples and uses all of them to predict the next example. Such a hypothesis family would be nonparametric because the effective number of parameters is unbounded — it grows with the number of examples. This approach is called instance-based learning or memory-based learning.' kNN, SVM, and RF all belong to this instance-based learning approach. Page 758 reads, 'Nonparametric models use all the data to make each prediction, rather than trying to summarize the data first with a few parameters. Examples include nearest neighbors and locally weighted regression.'

References

Russell, S. J, & Norvig, P.: Artificial intelligence: a modern approach, 3rd edition, Pearson Publisher, 2016.

Comment 12: 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

Response to comment 12: Thank you for correcting our mistake. We have revised it. The revised content in our manuscript now reads as shown in our **response to comment 9**.

Comment 13: 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.

Response to comment 13: Thank you for your kind reminder. We have revised it. The revised content in our manuscript now reads as shown in our **response to comment 9**.

Comment 14: L90: Why? Or at least provide a reference

Response to comment 14: Sorry, we made a mistake here. Thanks for bringing this to our attention. Our original intention was as follows: because of the spatial autocorrelation and spatial heterogeneity of forest AGB, its data distribution often does not conform to the assumption of the independent identical distribution of the traditional spatial statistical model. We have rewritten it as follows:

Second, the data distribution of forest AGB is not consistent with the assumption of independent identical distribution required by the traditional spatial statistical model when it has spatial autocorrelation and spatial heterogeneity.

Comment 15: L97: Please provide example references

Response to comment 15: We agree with you that example references were necessary. We

have now cited some important references into our revised manuscript. The corresponding references are:

Qiu, Q. Y., Yun, G. L., Zuo, S. D., Yan, J., Hua, L. Z., Ren, Y., Tang, J. F., Li, Y. Y., Chen, Q.: Variations in the biomass of Eucalyptus plantations at a regional scale in southern China, *Journal of Forestry Research*, 29, 1263-1276, <https://doi.org/10.1007/s11676-017-0534-0>, 2018.

Kobler, J., Zehetgruber, B., Dirnböck, T., Jandl, R., Mirtl, M., & Schindlbacher, A.: Effects of aspect and altitude on carbon cycling processes in a temperate mountain forest catchment, *Landscape Ecology*, 34, 325-340, <https://doi.org/10.1007/s10980-019-00769-z>, 2019

Comment 16: L187: ‘Each model was trained on 30 datasets. . .’ But within the CV, right? So it should be ‘trained on 29 datasets’.

Response to comment 16: Thank you for your correction. Since we have accepted one reviewer's suggestion on the cross-validation strategy, converted to adopt spatial block cross-validation strategy, we will rewrite Section 2.4.2 (‘split datasets’). The original sentence of L187 will also be revised.

Comment 17: L199-200: Not exactly true, the activation function makes the transformation linear or non-linear. The fundamental matrix multiplication is a linear function.

Response to comment 17: Thank you for your kind reminder. We have revised it. The revised content in our manuscript now reads:

The transformation from input space to hidden space is either linear or nonlinear, whereas the transformation from hidden space to output space is linear.

Comment 18: 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.

Response to comment 18: We used RBF-ANN because it has been reported to have excellent performance in environmental applications, such as the spatial interpolation of soil moisture and nutrients (Xu, 2012), the prediction of river flow time series (Ghorbani et al., 2016), and the prediction of the energetic efficiency of a roughened solar air heater (Ghritlahre et al., 2018).

As suggested, we have further explained RBF-ANN in detail. Details of how the RBF function is used by RBF-ANN can be found at the following websites: (1) https://www.saedsayad.com/artificial_neural_network_rbf.htm, (2) <https://pythonmachinelearning.pro/using-neural-networks-for-regression-radial-basis-function-networks/>). We will add references to these websites in our manuscript and the Supplementary Material.

RBF-ANN is a three-layer neural network model which includes an input layer, a hidden layer, and an output layer. A Gaussian RBF is used inside the hidden-layer neurons of RBF-ANN. This differentiates an RBF-ANN from a regular neural network. First, each input vector is input into each basis. Then, the approximated function value is obtained by a simple weighted sum. Like any neural network, RNF-ANN is trained by back-propagation. Finally, the RBF network is implemented in a class and used to approximate a simple function. The transformation from input space to hidden space is either linear or nonlinear, whereas the transformation from hidden space to output space is linear. The function of the hidden layer is to map the vector from the indivisible low-dimensional linear state to the separable high-dimensional linear state so as to greatly accelerate the learning and convergence speed and avoid getting stuck in a local optimum (Elanayar and Shin, 1994; Xia and Xiu, 2007).

Elanayar, V. T. S., and Shin, Y. C.: Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems, *IEEE Transactions on Neural Networks*, 5, 594-603, 1994.

Ghorbani, M. A., Zadeh, H. A., Isazadeh, M., & Terzi, O.: A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction, *Environmental earth sciences*, 75(6), 476.1-476.14, <https://doi.org/10.1007/s12665-015-5096-x>, 2016.

Ghritlahre, H. K., & Prasad, R. K.: Exergetic performance prediction of solar air heater using MLP, GRNN and RBF models of artificial neural network technique, *Journal of Environmental Management*, 223(OCT.1), 566-575, <https://doi.org/10.1016/j.jenvman.2018.06.033>, 2018.

Xia, C. L., Xiu, J.: RBF ANN Nonlinear Prediction Model Based Adaptive PID Control of Switched Reluctance Motor, Proceedings of the CSEE, 27, 626-635, https://doi.org/10.1007/11893295_69, 2007.

Xu, Y.: Application of RBF artificial neural network with block effect to spatial interpolation of soil properties, Shuikexue Jinzhan/advances in Water Science, 23(1), 67-73, <https://doi.org/10.14042/j.cnki.32.1309.2012.01.005>, 2012.

Comment 19: 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.

Response to comment 19: Thank you for your professional guidance in machine learning. The incorrect description has been deleted.

Comment 20: L227: What are these “obvious advantages”?

Response to comment 20: The detailed advantages of P-BSHADE compared with other traditional spatial statistics are described in the following paragraph. We have modified it to be more coherent and logical. The revised content in our manuscript now reads:

P-BSHADE is markedly different from the Kriging and Inverse Distance Weighting (IDW) algorithms. Compared with Kriging and IDW, the application of P-BSHADE to forest AGB interpolation has the following obvious advantages: **(1)** The spatial distribution of forest AGB is also characterized by spatial autocorrelation and heterogeneity, which have been taken into account in the P-BSHADE model. Taking into account spatial heterogeneity can effectively solve the difference in forest AGB distribution caused by different terrain or geographical location. However, Kriging and IDW only consider the spatial correlation between plots. **(2)** Additionally, P-BSHADE considers strongly correlated sample plots as neighboring plots, whereas the Kriging and IDW algorithms consider sites that are spatially close.

Comment 21: L248: ‘when $j=1, i=2,3,\dots,30$: when $j=1, i=1,3,5,\dots,30$)’ What do you mean?

Response to comment 21: We made a mistake in L248. It should be ‘when $j=1, i=2,3,\dots,30$: when $j=2, i=1,3,5,\dots,30$ ’. The purpose of this passage was to illustrate the ‘leave-one-out

cross-validation' situation. Given that we have changed the cross-validation strategy to spatial block cross-validation, we will revise this part.

Comment 22: 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.

Response to comment 22: Thank you for your kind advice. We have summarized the P-BSHADE method in the manuscript and have included detailed information in the Supplementary Material.

Comment 23: L312: 'A detailed description of the combined models. . .' is missing in the Supplementary Material

Response to comment 23: Thank you for your kind correction. Considering that the combination of the two models is the innovation of our manuscript, we included an introduction of how to combine the two methods in the main manuscript and deleted this sentence.

Comment 24: 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.

Response to comment 24: Thank you for your kind advice. Using GeoDetector (<http://www.geodetector.cn>), we found that the interaction of longitude and latitude had an effect on AGB (see Table 1 below).

Table 1. Results of factor detector and interaction detector from GeoDetector.

	Longitude	Latitude	Interaction of longitude and latitude
q statistic	0.54	0.09	0.75
P value	0.004	0.57	Interaction result: enhanced, nonlinear

We also tested the performance of models when longitude and latitude were both used as a predictor, and compared the results with those of models which only include longitude or which include neither longitude nor latitude (see Table 2 below). The results showed that using both longitude and latitude as predictors did not improve the performance of RF, SVM, or RBF-ANN. However, this does not mean that machine learning cannot capture the spatial relationships. We also show that the combination of three MLs and P-BSHADE can all improve the prediction accuracy compared to the single ML in all of the situations. We think that this further illustrates the advantages of the combined models.

Table 2. Performance of models in different predictor situations.

Model	MAE	MRE	RMSE	nRMSE
Situation 1: includes longitude and latitude				
SVM	18.260	2.260	25.929	0.548
RBF-ANN	14.665	0.324	21.429	0.453
RF	16.435	1.292	27.419	0.579
P-BSHADE	12.069	0.251	19.490	0.412
SVM & P-BSHADE	5.075	0.108	6.495	0.137
RBF-ANN & P-BSHADE	12.164	0.244	17.546	0.371
RF & P-BSHADE	6.344	0.153	12.279	0.259
Allometric Model	16.412	0.577	25.080	0.530
Situation 2: includes longitude but excludes latitude				
SVM	13.659	1.382	20.251	0.428
RBF-ANN	14.884	0.324	20.719	0.438
RF	14.919	1.107	21.371	0.451
P-BSHADE	12.069	0.251	19.490	0.412
SVM & P-BSHADE	6.879	0.128	10.757	0.227
RBF-ANN & P-BSHADE	12.597	0.260	17.819	0.376
RF & P-BSHADE	5.816	0.137	9.520	0.201
Allometric Model	16.412	0.577	25.080	0.530

Situation 3: excludes longitude and latitude				
SVM	15.824	2.034	21.609	0.456
RBF-ANN	14.175	0.373	20.354	0.430
RF	15.201	1.092	21.330	0.451
P-BSHADE	12.069	0.251	19.490	0.412
SVM & P-BSHADE	9.679	0.173	16.055	0.339
RBF-ANN & P-BSHADE	12.227	0.232	18.124	0.383
RF & P-BSHADE	6.104	0.136	9.911	0.209
Allometric Model	16.412	0.577	25.080	0.530

Note: MAE: mean absolute error; MRE: mean relative error; RMSE: root mean square error; nRMSE: normalized root mean square error

Comment 25: L479: 'Machine learning models appear adept at tackling high-dimensional problems.' Yes, they do, but you do not have a 'high-dimensional problem'.

Response to comment 25: Thank you for your kind advice. We have deleted this inappropriate sentence.

Comment 26: 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?

Response to comment 26: Thank you for your kind suggestion. We have deleted this sentence.

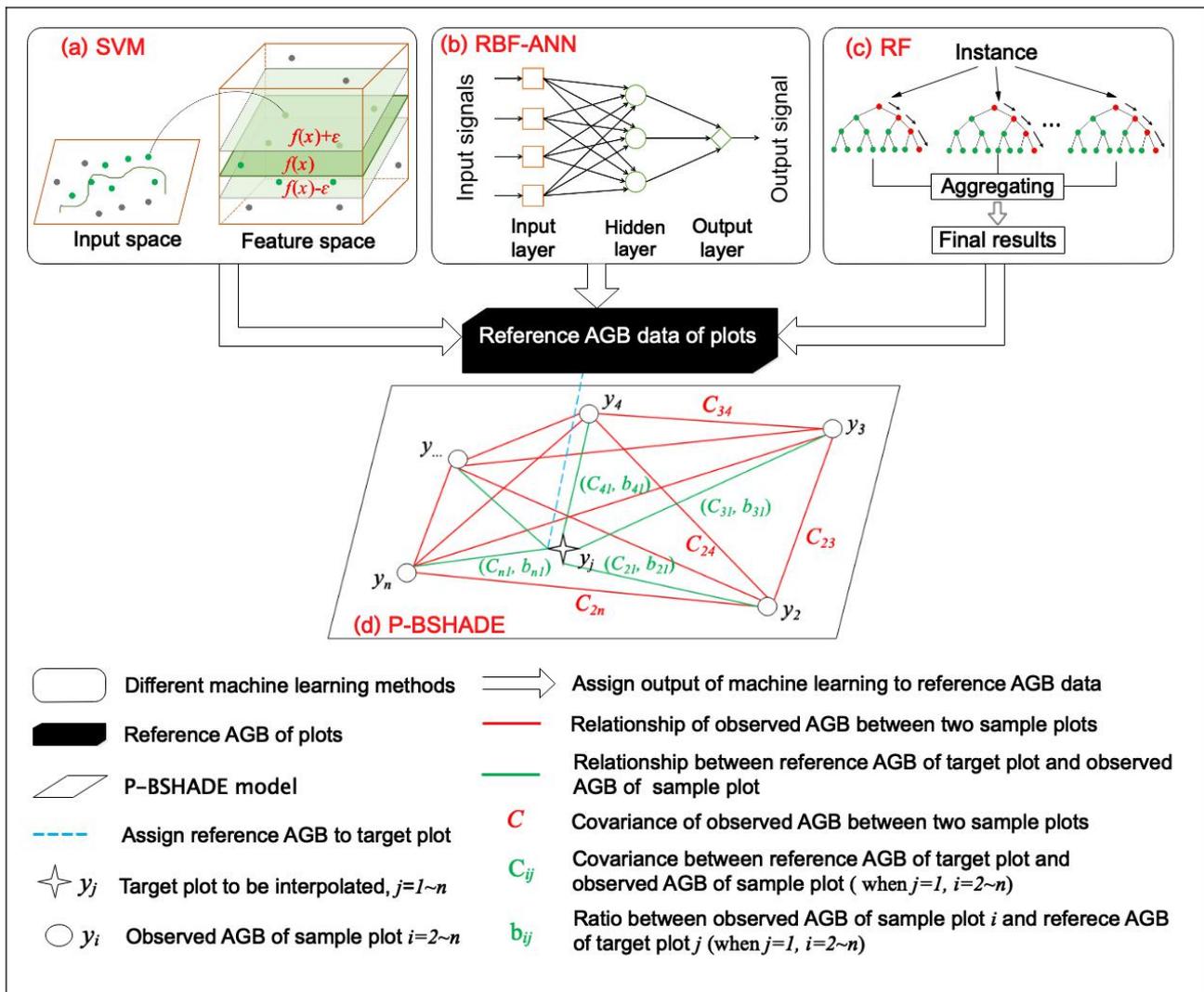
Comment 27: L568: RF and SVM are also sensitive to hyper-parameters. It is myth that RF does not need hyper-parameter tuning (see <https://doi.org/10.1002/widm.1301>).

Response to comment 27: We are sorry for our arbitrary description of machine learning. It has been deleted. We have also read and cited the reference you kindly provided.

Figures:

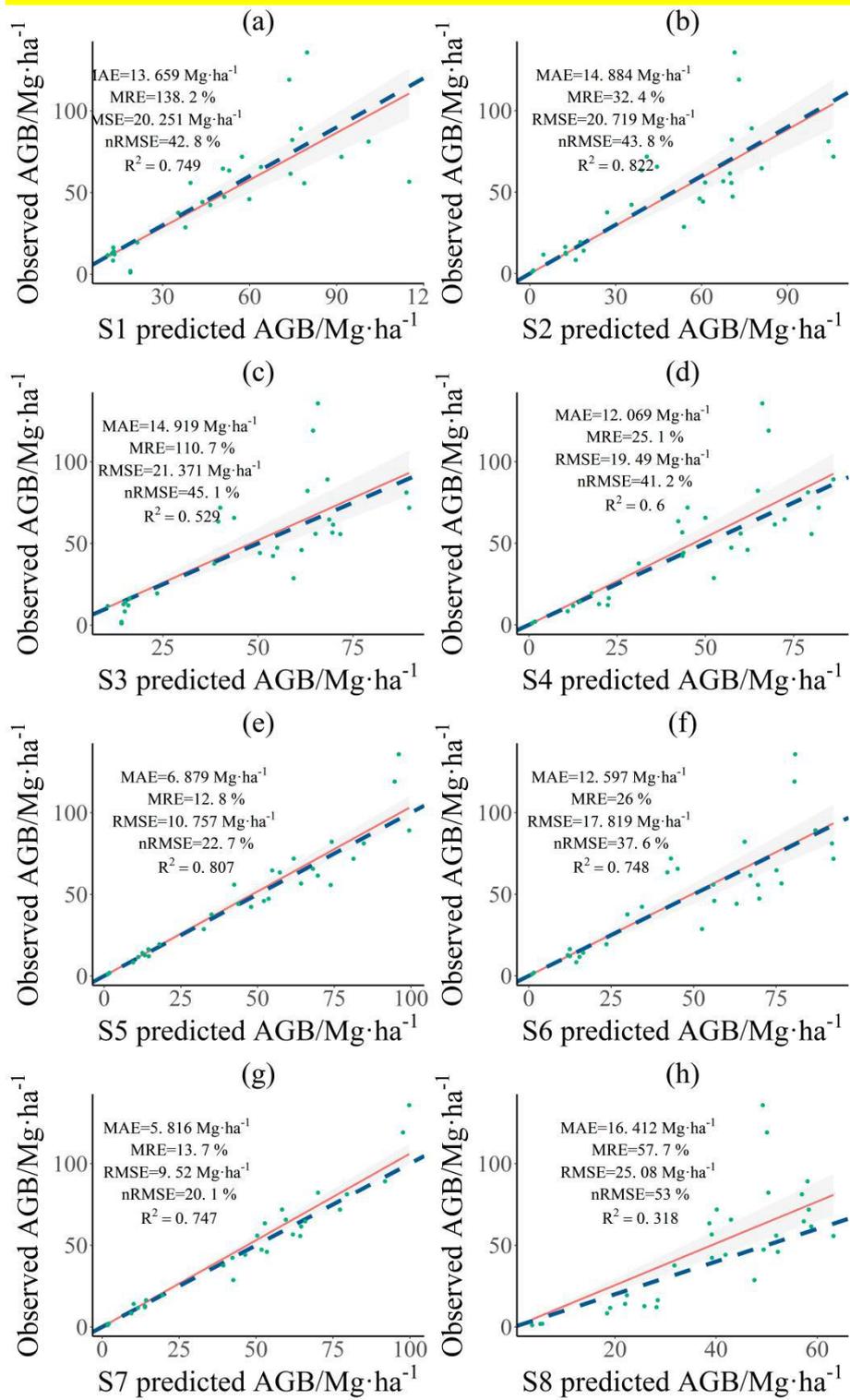
Comment 28: Fig 3: The sematic figures of the ML do not fit to the way you used them in your work: RBF-ANN, RF, and SVM are illustrated as classifier with 4, 3, and 2 response classes, however, you use them as regression models (e.g. one output node for the RBF ANN, no majority voting for the RF etc.)

Response to comment 28: Thank you for your kind advice. We have modified Figure 3 as follows:



Comment 29: Fig 6: Revise your color choice, e.g. it is difficult to distinguish between the yellow and blue line.

Response to comment 29: Thank you for the kind advice. As suggested, we have revised Figure 6 in the revised manuscript.



We thank the reviewer and remain at your disposal for any further questions.

Yours sincerely,

Yin Ren