Supplementary Material

S1

1. Harvest of analytic-trees

1.1 Plot setup

The purpose of plot selection was to establish fixed and permanent plots representing regional *Eucalyptus* growing conditions and to provide analytic-tree data on the single-tree scale with adequate consideration of spatial heterogeneity. Patches were selected first and met the following six conditions: (1) patch records were available from FMPI data for 2009, (2) forest stand classification was classified as timber or commercial forest, (3) forest patches were disturbance free during the last seven years, including but not limited to logging, fire, and pests, (4) forest patches were not replanted, (5) patches contained closed-canopy forests, and (6) patches were not mixed stands, but monocultures. Based on these six conditions, 2980 *Eucalyptus* patches in the FMPI data were selected. Fixed and permanent plots were established. The 2980 selected patches were divided into ten groups based on forest age. Each stand group was planted at the same time. We calculated the mean basal area for each group as the basis for fixed plot selection, which was obtained from specified plot design and sampling procedures. In parallel, we considered site conditions, forest use, and forest origin (natural vs. man-made), and subsequently established 30 permanent square plots 20 m × 20 m in area. We recorded fixed-plot conditions by assigning a code to each fixed plot and recorded the environmental conditions, including the following direct and indirect attributes: age, community structure, canopy density, and understory shrub conditions. Finally, a full tree survey was done in each fixed plot. The following measures were obtained: DBH ≥ 8 cm in diameter for every tree, tree height, and other tree attributes.

1.2 Selection and cutting of standard wood

Standard wood was selected following a full tree survey. The following selection criteria were used: (1) Wood must be located within the plot; the stems were representative of the plot, with no disturbances (e.g., pests, fire, or anthropogenic activities) and the wood was healthy and not across plot border. (2) Based on full tree survey data, a tree sampling method was used to calculate average basal area, with three trees closest to the average values being selected (i.e., standard trees). These standard trees were cut down and the average biomass was calculated and multiplied by the stems per unit area to obtain the total *Eucalyptus* biomass per unit area.

1.3 Single-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.

For each tree, all roots with a diameter greater than 2 mm were collected. Coarse roots (>2 mm diameter) were gathered, starting at the stump and following the roots to their end. The tree density in the plots was high, and much of the root growth was shallow and lateral; therefore, it was difficult to distinguish the fine roots of one tree from those of adjacent trees. Consequently, large soil blocks (1 m^3 each) were excavated around each target tree. This block size was selected based on the assumption that any fine foreign roots collected within the block were balanced with intrinsic fine roots remaining in the soil. Fine roots were not collected outside of the blocks. All roots were extracted in the laboratory from each soil block by wet sieving, and 0.2–2 cm diameter roots were isolated by hand.

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. Therefore, stand biomass (mg/ha) was calculated based on total tree number.

2. Introduction to machine learning

2.1 Support vector machines for regression

A support vector machine (SVM) is a kind of categorized algorithm that improves the generalized machine learning ability by minimizing structural risks, so as to minimize the empirical risk and confidence intervals. In this way, it can achieve adequate statistical trends from a limited sample quantity. Compared with traditional machine learning methods, SVM adopts the principles to minimize structural risks. Along with minimizing sample point errors, SVM simultaneously narrows the upper bound of generalized error in the model so as to improve the generalization ability of the model and to solve the problems of excessive model learning, nonlinearity, and dimensionality (Ukil, 2002). The specific mathematical expression of a SVM is as follows:

The definition of a straight line is

$$y(x) = w^T x + b. \quad \text{(A.1)}$$

The distance from any point x to this straight line is

$$\frac{1}{||w||}(w^T x + b). \quad \text{(A.2)}$$

The information of *N*th training point (point coordinates, and class A or B) is recorded as (x_i, y_i)

The above expression is

$$\arg \quad \max_{w,b} \left\{ \frac{1}{||w||} \min_{n} [y_i (w^T x_i + b)] \right\}. \quad \text{(A.3)}$$

Taking longitude, DBH, tree height, and forest age as the selection characteristics, and using the biomass data of 30 plots as model training samples, the classification model of SVM is trained by using a C-classification method. The Gaussian inner product function serves as the kernel function.

2.2 Artificial neural networks radial basis function

The basic components of radial basis function artificial neural networks (RBF--ANNs) include an input layer, a hidden layer, and an output layer, which give the best approximation for nonlinear functions and the optimal global performance (Elanayar and Shin, 1994). The change from input layer space to hidden layer space is nonlinear, whereas the spatial transformation from hidden layer space to output layer is linear. The RBF network not only has good generalizability, but also requires less calculation. In general, its learning speed is faster than that of other algorithms, so it can usually avoid the lengthy process of iterative calculations found in the learning algorithm of BP neural networks and the possibility of falling into a local extremum. It has been widely used in many fields, such as meteorology (Nath et al., 2016), soil (Zakian, 2017), vegetation (Hilbert and Ostendorf, 2001), and engineering control (Sarimveis et al., 2004). Its methods are as follows:

$$R_j^u = exp\left[\frac{1}{2}\sum_{k=1}^{N}\left(\frac{x_k^u - C_{jk}}{\sigma_j^2}\right)^2\right], \quad \text{(A.4)}$$

$$V_j^u = \frac{R_j^u}{\sum_{j=1}^{L} R_j^u} \quad \text{(A.5)}$$

$$Q_j^u = \sum_{j=1}^{L} W_j^u V_j^u \quad \text{(A.6)}$$

X_k is the input of input layer $C_{jk} = (C_{j1}, C_{j2}, \ldots, C_{jn})$, T is the center of gravity for the basis function, σ_j is the width of the basis function, $R^u{}_j$ is the Gaussian-type radial basis function, V_j^u is the normalized radial basis function, L is the number of hidden layer units, X_k^u is the input mode, O_i is the output from the output layer, V_j is the output from the hidden layer, and W_j is the right of connection from the hidden layer units to the output layer units.

2.3 Random forest

Random forest (RF) is one of algorithm models and is a relatively new machine learning technique. Random forest is a data mining method developed by Breiman in 2001. It is a modern classification and regression technology and is also a combination of self-learning technologies (Breiman, 2001). The idea of combinatorial learning is to integrate several individual classifiers when classifying new instances and to decide the final classification of the instances by combining the classification results of multiple classifiers, so as to achieve better performance than that achieved by single classifiers. 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.

3. Introduction of P-BSHDE

3.1 Overview of P-BSHADE

P-BSHADE is an extension of the BSHADE method, which stands for the best linear unbiased estimation (BLUE) model for biased-spatial-location data (Hu et al., 2013). With the BSHADE model, the spatial correlation and heterogeneity of target data can be added into the model by using prior knowledge (such as temperature and forest AGB). In addition, through the rectification treatment of sample points, the BLUE model can estimate the target subjects. The strategy of the algorithm is to transform the problem into one of solving for the extremum of a multivariate function with constraint conditions, followed by using the Lagrange multiplier method and the overall estimate to acquire the corresponding parameters (Wang et al., 2011) (i.e., each sample in this method is given a certain weight, so that the variance between each sample and the true value is minimized to achieve the purpose of rectification).

Based on the BSHADE method, P-BSHADE is a BLUE-based interpolation method that considers both temporal and spatial heterogeneity. It can use biased samples to deduce the corresponding attributes of regions with missing samples. Therefore, the P-BSHADE model has the following two characteristics and assumptions: (1) the spatial distribution of target data (such as temperature and forest AGB) is heterogeneous, and (2) the information of correlation and difference among the target data between forests (or sites) is also included in the operation of the model (Xu et al., 2013). The effectiveness of the P-BSHADE method has been tested by using data of the annual average temperature in China from 1950 to 2000 (Xu, 2013).

3.2 P-BSHADE methods

a. Objective

The objective is to interpolate the AGB data of target sample plot (in modeling process) or of forest patch (in model application process) based on data acquired from other sample plots (in modeling process) or other forest patches (in model application process). Here we introduce, in the modeling process, how to interpolate the AGB data of target sample plot. A theoretical description is

$$\hat{y}_j = \Sigma_{i=1}^{n} w_{ij} y_i \ , \tag{A.7}$$

where $\hat{y}_j$ is the estimated AGB of the *j*th sample plot by P-BSHADE $(j = 1{\sim}30, n = 30)$, y_i is the reference AGB of the *i*th sample plot $(i = 1{\sim}30, n = 30)$, w_{ij} is the weight (contribution) of *i*th reference sample plot AGB to the *j*th sample plot AGB to be interpolated (when $j = 1, i = 2{\sim}30$; when $j = 1, i = 1,\ 3{\sim}30$). Appendix B gives a specific description of the P-BSHADE and the algorithm formulas. As expected, the two properties of the estimate of Eq.(7) are unbiased,

$$E(y_j) = E(\hat{y}_j) \ , \tag{A.8}$$

and minimum estimation variance,

$$min_w \left[\sigma_{\hat{y}_j}^2 = E(\hat{y}_j - y_i)^2\right] \ , \tag{A.9}$$

where E is the statistical expectation, $\hat{y}_j$ is the estimate AGB of the target plot.

b. Ratio between target sample plot and other sample plots data

The ratio between target sample plot and other sample plots data is one of the most important inputs for estimating target sample plot AGB and is an index of heterogeneity in the AGB spatial distribution. The relationship between target sample plot and other sample plots data can be expressed as

$$b_{ij} E y_j = E y_i \ , \tag{A.10}$$

In most cases, the AGB of two plots are not equal, and the relationship can be further expressed as the relative bias b_{ij} between the mathematical expectation of y_j and y_i. Considering Eq. (7), Eq. (10) can be written as

$$\Sigma_{i=1}^{n} w_{ij} b_{ij} = 1 \ , \tag{A.11}$$

This equation is generally valid for nonhomogeneous conditions. Clearly, the determination of b_{ij} requires the coefficients w_{ij} $(i = 1, \ldots, n, j = 1, \ldots, n)$ to be calculated, which is addressed in the following section.

c. Estimation of weight

The estimation problem is to find the weights w_{ij} that satisfy the unbiased condition and that minimize the estimation

variance:

$$\sigma_{\hat{y}_j}^2 = E\left(\hat{y}_j - y_i\right)^2 = C\left(\hat{y}_j\hat{y}_j\right) + C(y_i y_i) - 2C\left(\hat{y}_j y_i\right) \quad \text{(A.12)}$$

These weights can be calculated by minimizing the estimation variance and taking into account the unbiasedness,

$$\begin{bmatrix} C(y_1y_1) & \cdots & C(y_1y_n) & b_{1j} \\ \vdots & \ddots & \vdots & \vdots \\ C(y_ny_1) & \cdots & C(y_ny_n) & b_{nj} \\ b_{1j} & \dots & b_{nj} & 0 \end{bmatrix} \begin{bmatrix} w_{1j} \\ \vdots \\ w_{nj} \\ \mu \end{bmatrix} = \begin{bmatrix} C\left(y_1y_j\right) \\ \vdots \\ C\left(y_ny_j\right) \\ 1 \end{bmatrix} \quad \text{(A.13)}$$

μ is a Lagrange multiplier. The minimized variance in the estimation error can then be written as

$$\sigma_y^2 = \sigma_{y_i}^2 + \Sigma_{i=1}^n \Sigma_{k=1}^n C(y_i y_k) - 2\Sigma_{i=1}^n w_{ij} C\left(y_i y_j\right) + 2\mu\left(\Sigma_{i=1}^n w_{ij} b_{ij} - 1\right) \quad (14)$$

S2

Table B.1 Statistical description of forest patch data.

	Number	Minimum	Maximum	Mean	Standard deviation
Age(a)	2980	1	51	5.05	2.42
Stand density(stems/ha)	2980	135	3450	1377.63	241.10
DBH(cm)	2980	5.0	60.0	12.30	3.55
Tree height(m)	2980	1.5	48.50	13.40	3.99

Note: Of the 2980 forest patches, for which the maximum age is 51 years, there were only 24 forest patches over 10 years, all of which were identified as mature forest.

Table B.2 Tree structures for calculating biomass of 90 parse trees.

Age	DBH	Height	Individual biomass (kg)	Age	DBH	Height	Individual biomass (kg)
(yr)	(cm)	(m)	Aboveground	(yr)	(cm)	(m)	Aboveground
1	3.3	4.3	1.9376	6	15.0	20.8	82.2273
	3.0	4.0	2.2500		15.3	20.8	99.3969
	3.2	4.3	1.8514		15.0	21.1	102.5718
	2.1	3.3	1.1061		15.3	19.9	97.7377
	2.1	3.4	1.0697		15.0	21.2	93.3897
	2.4	3.3	1.3143		14.5	20.8	89.4676
	3.4	4.6	2.2976		14.6	19.4	81.7034
	3.3	4.7	2.3782		15.0	19.4	81.8693
	3.3	4.5	2.0494		14.6	20.1	87.1974
2	7.6	10.1	14.4861	7	18.0	20.4	119.9316
	8.0	8.5	14.7833		17.8	20.8	106.3167
	8.1	9.9	14.3030		18.0	20.4	143.0096
	7.2	10.5	12.1682		16.7	20.0	113.6738
	7.0	10.4	11.7154		16.6	20.9	99.6045
	7.0	10.8	11.1324		16.4	21.4	98.7499
	7.2	9.2	12.3033		16.9	19.8	102.7874

	7.2	9.5	11.0665		16.9	20.2	97.2996
	7.0	8.1	10.2483		15.6	20.3	89.5590
3	6.1	6.3	5.5350	8	14.3	21.1	89.6489
	7.0	6.9	8.8532		14.5	19.8	72.6971
	6.4	6.8	7.5987		14.0	19.2	90.9861
	6.2	7.6	6.3156		16.4	19.7	99.4468
	7.2	7.9	9.5706		16.4	20.1	97.8657
	7.2	7.7	9.7457		17.2	21.2	112.4650
	6.1	6.9	6.4039		14.0	17.7	63.5059
	6.2	9.4	9.2803		15.0	20.3	81.3824
	5.4	6.6	5.7853		14.9	19.3	84.1050
4	11.1	18.6	36.7169	9	16.9	25.5	110.3010
	12.1	17.3	50.7412		17.2	25.1	146.4738
	11.8	17.3	44.8078		17.5	24.5	130.5710
	8.9	11.7	16.5647		16.1	23.5	117.4427
	9.2	17.4	27.9658		15.8	22.9	106.7083
	8.8	15.2	24.5316		15.9	23.3	112.0993
	13.2	17.9	56.0009		18.4	26.6	168.4229
	13.1	18.2	58.7273		18.4	24.7	144.5210
	12.4	17.8	51.5655		18.3	26.0	167.0830
5	13.2	19.7	62.9911	10	18.2	27.0	136.6728
	13.9	16.5	68.7846		18.5	25.0	163.4031
	12.9	16.1	58.5322		18.2	26.2	150.9330
	13.4	19.3	81.9325		14.0	18.5	69.9841
	13.4	19.4	84.0987		13.9	22.1	76.9977
	13.1	18.9	73.2317		13.9	24.0	91.4171
	13.4	19.0	70.4283		17.6	23.8	118.4468
	12.9	17.1	70.5207		17.6	22.2	149.1616
	13.8	18.6	96.5537		17.6	25.6	138.2509

184 **Table B.3** How the optimal model were constructed.

185

Leave-one-out		Model 1	Model 2	…	Model 7
Validation data (Plot AGB)	Training data (Plot AGB and Predictor variables)	Simulated data1 (Simulated AGB 1)	Simulated data2 (Simulated AGB2)	Simulated data (Simulated AGB)	Simulated data7 (Simulated AGB7)
Plot ID	Plot ID	Plot ID	Plot ID	Plot ID	Plot ID
1	2-30	1 S1	1 S2	…	1 S7
2	1,3-30	2 S1	2 S2	…	2 S7
3	1-2,4-30	3 S1	3 S2	…	3 S7
…	…	…	…	…	…
29	1-28,30	29 S1	29 S2	…	29 S7
30	1-29	30 S1	30 S2	…	30 S7
AGB(group M)		AGB(group1)	AGB(group2)	…	AGB(group7)
		MAE1, MRE1 and RMSE1	MAE2, MRE2 and RMSE2	…	MAE7, MRE7 and RMSE7

186

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

188

Variables	Mean	Median	Standard deviation	variable coefficient	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	$0.13*10^{-5}$	117.45	117.5

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

189

190 **Table B.5** List of model accuracy indexes and their definitions.

191

Model accuracy index	Description	Interpretation
Mean Absolute Error (MAE)	Mean absolute error is the mean of the absolute deviations of all individual measurements from arithmetical mean values.	The mean of absolute deviations of the true biomass of the 90 parse trees from the average biomass of 90 parse trees obtained from a given simulation method. Because the deviations are expressed in absolute values, the mean absolute error will not be cancelled by positive and negative numbers. Therefore, the mean absolute error can better reflect the actual situation of the prediction error.
Mean Relative Error (MRE)	Mean relative error is the average value of the relative error, which is usually expressed as the absolute value, i.e., the absolute value of mean relative error. The relative error is the ratio of the absolute error to the measured value or the average of multiple measurements.	The average value of the ratio of the absolute error (the absolute value of the difference between the true value and the simulated value) for the biomass of each of the 90 parse trees to the simulated values. It is used to analyze the accuracy and precision of the results.
Root Mean Square Error (RMSE)	Square root of the ratio between the square of the deviation of the observed value from the true value and n, the number of observations. In actual measurement, the number of observations n is always limited, and the true value can only be substituted by the most reliable (best) value.	The average of the square root of the following value: for a real value and a simulated value of the biomass of each of 90 parse trees, the square of their difference divided by 90. Because the results are very sensitive to the extremely large or small errors in a set of measurements, it can better reflect the precision of the measurement.

Table B.6 Leave-on-out cross validation for machine learning (support vector machine, artificial neural network, and random forest), spatial statistical analysis (P-BSHDE), and results from the combination of the two.

Method	MAE	MRE	RMSE
SVM	11.167837	0.2478704	10.387622
ANN	12.148633	0.2669469	10.387868
RF	10.155275	0.2593000	9.429181
P-BSHADE	18.371450	0.3912975	14.077459
SVM-&P-BSHADE	6.882970	0.1246758	6.303799
ANN-&P-BSHADE	10.135638	0.2049183	9.633279
RF-&P-BSHADE	5.678865	0.1296616	5.299004

S3

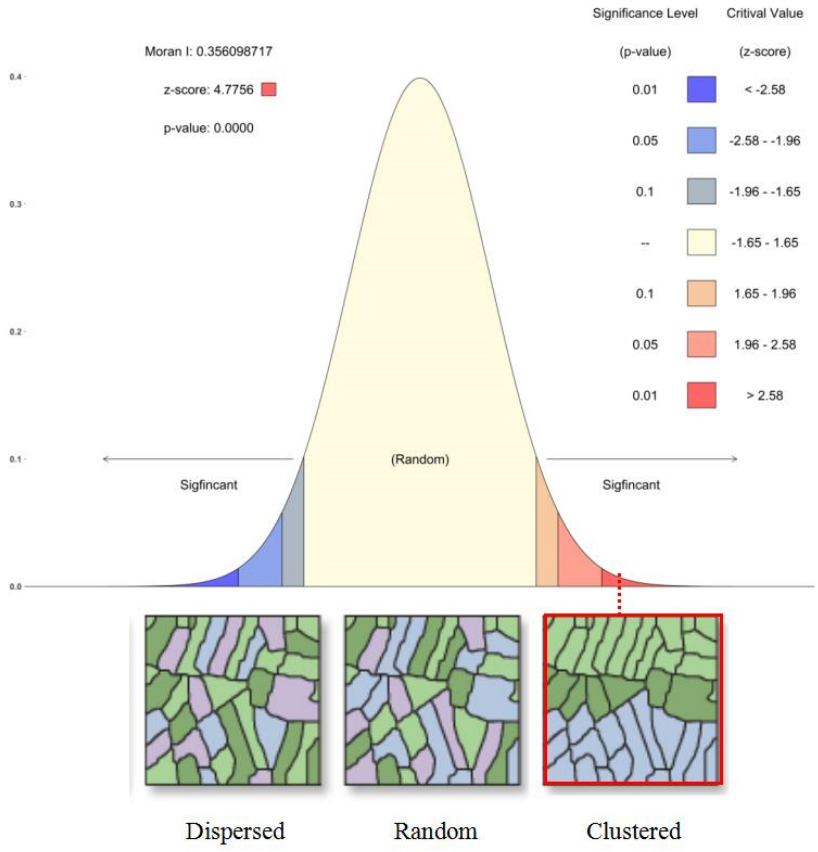

Figure C.1 Spatial autocorrelation report.

References

Breiman, L., 2001. Random forests. Machine Learning 45(1) 5-32.

Elanayar, V.T.S., Shin, Y.C., 1994. Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems. IEEE Transactions on Neural Networks 5(4) 594-603.

Hilbert, D.W., Ostendorf, B., 2001. The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates. Ecological Modelling 146(1-3) 311-327.

Hu, M.G., Wang, J.F., Zhao, Y., Jia, L., 2013. A B-SHADE based best linear unbiased estimation tool for biased samples. Environmental Modelling & Software 48(48) 93-97.

Nath, S., Kotal, S.D., Kundu, P.K., 2016. Seasonal prediction of tropical cyclone activity over the north Indian Ocean using three artificial neural networks. Meteorology and Atmospheric Physics 128(6) 751-762.

Sarimveis, H., Alexandridis, A., Mazarakis, S., Bafas, G., 2004. A new algorithm for developing dynamic radial basis function neural network models based on genetic algorithms ☆. Computers & Chemical Engineering 28(1–2) 209-217.

Ukil, A., 2002. Support vector machine. Computer Science 1(4) 1-28.

Wang, J.F., Reis, B.Y., Hu, M.G., George, C., Yang, W.Z., Qiao, S., Li, Z.J., Li, X.Z., Lai, S.J., Chen, H.Y., 2011. Area Disease Estimation Based on Sentinel Hospital Records. Plos One 6(8) e23428.

Xu, C.D., 2013. Modeling of uncertainty of temperature observation and anomaly stratification. University of Chinese Academy of Sciences.

Xu, C.D., Wang, J.F., Hu, M.G., Li, Q.X., 2013. Interpolation of missing temperature data at meteorological stations using P-BSHADE*. Journal of Climate 26(19) 7452-7463.

Zakian, P., 2017. An efficient stochastic dynamic analysis of soil media using radial basis function artificial neural network. Frontiers of Structural and Civil Engineering 11(4) 470-479.