Application of machine learning algorithms in estimating aboveground biomass of Tamarix shrub dunes (postprint)

Authors: Dong Yaqing, Song Shaoteng, Guo Xiaoqian, Zhao Yuanjie

Date: 2026-01-30T16:53:16+00:00

Abstract

Tamarix chinensis, as an important shrub species in arid-region shrublands, has strong stress resistance and growth adaptability, can stabilize sand dunes, slow down wind-blown sand movement, and promote ecological restoration. The aboveground biomass (AGB) of Tamarix shrublands is a key indicator for assessing vegetation status and desertification control in arid regions. In this study, Tamarix shrubland dunes in the lower reaches of the Tarim River were taken as the research object. Based on Landsat 8 remote sensing imagery, 92 band indices, vegetation indices, and texture indices were extracted as feature variables. Stepwise regression, least absolute shrinkage and selection operator (LASSO), and extreme gradient boosting (XGBoost) algorithms were used for variable selection, and random forest (RF), support vector regression (SVR), and back-propagation neural network regression (BPNN) models were constructed to estimate the AGB of Tamarix shrubland dunes. The adaptability between different feature selection methods and models and their influence on AGB estimation accuracy were assessed, and the application of multivariate algorithms in AGB estimation of Tamarix shrubland dunes was explored. The results showed that: (1) The variable sets selected by stepwise regression, LASSO, and XGBoost algorithms were significantly correlated with the AGB of Tamarix shrubland dunes, and the collinearity among variables was low (VIF<5), verifying the effectiveness of the algorithms. (2) The models constructed based on LASSO and XGBoost algorithms had significantly higher accuracy than those based on stepwise regression, among which the RF model constructed using the LASSO algorithm performed best (R2=0.73, RMSE=447.63 g·m-2); multivariate combinations significantly improved the predictive capability of the models. (3) Using the RF model based on the LASSO algorithm, the mean AGB of Tamarix shrubland dunes in the study area in the lower reaches of the Tarim River was estimated to be 1733.63 g·m-2, with a total AGB of 1.71$\times$108 kg, indicating that the model has high reliability and regional applicability. The findings can provide a

reference for the selection and accuracy improvement of remote sensing inversion methods for aboveground biomass of shrub vegetation in desert regions.

Full Text

Application of Machine Learning Algorithms for Estimating Aboveground Biomass of Tamarix Nebkha

DONG Yaqing[1,2,3], **SONG Shaoteng**[1,2,3], **GUO Xiaoqian**[1,2,3], **ZHAO Yuanjie**[1,2,3]

[1]School of Geographical Sciences, Hebei Normal University, Shijiazhuang 050024, Hebei, China
[2]Hebei Technology Innovation Center for Remote Sensing Identification of Environmental Change, Shijiazhuang 050024, Hebei, China
[3]Hebei Key Laboratory of Environmental Change and Ecological Construction, Shijiazhuang 050024, Hebei, China

Abstract: *Tamarix chinensis*, a dominant shrub species in arid ecosystems, exhibits strong stress resistance and adaptive growth capacity, playing a crucial role in stabilizing sand dunes, mitigating wind erosion, and promoting ecological restoration. The aboveground biomass (AGB) of Tamarix nebkha serves as a key indicator for assessing vegetation status and desertification control in arid regions. This study focuses on Tamarix nebkha in the lower reaches of the Tarim River. Based on Landsat 8 imagery, 92 feature variables were extracted, including spectral bands, vegetation indices, and texture indices. Variable selection was performed using stepwise regression, least absolute shrinkage and selection operator (LASSO), and extreme gradient boosting (XGBoost) algorithms. Random forest (RF), support vector regression (SVR), and backpropagation neural network (BPNN) models were constructed to estimate Tamarix nebkha AGB. The study evaluated the adaptability of different feature selection methods and models, their impact on estimation accuracy, and explored the application potential of multivariate algorithms for Tamarix nebkha AGB estimation. Results demonstrate that: (1) All three variable selection algorithms identified variable sets significantly correlated with Tamarix nebkha AGB, with low multicollinearity among variables (VIF < 5), confirming algorithm effectiveness. (2) Models built with LASSO and XGBoost algorithms achieved significantly higher accuracy than those using the stepwise method. The RF model based on LASSO performed best ($R^2 = 0.73$, RMSE $= 447.63\ g \cdot m^{-2}$), with multivariate combinations substantially enhancing predictive capability. (3) The LASSO-based RF model estimated mean AGB of Tamarix nebkha in the lower Tarim River region at $1733.63\ g \cdot m^{-2}$, with total biomass of 1.71×10^8 kg, demonstrating high reliability and regional applicability. These findings provide valuable references for method selection and accuracy improvement in remote sensing inversion of shrub AGB in desert ecosystems.

Keywords: Tamarix nebkha; aboveground biomass; LASSO algorithm; ma-

chine learning; remote sensing estimation

1 Materials and Methods

1.1 Study Area

The lower reaches of the Tarim River are located in the Bayingolin Mongol Autonomous Prefecture of Xinjiang, representing an extremely arid region in China. The area experiences a temperate continental climate characterized by perennial aridity, low precipitation, and large temperature variations. The plant community structure is simple, dominated by *Populus euphratica*, *Tamarix chinensis*, and *Phragmites australis*. The study area is situated along the Qara-to-Taitema Lake section of the lower Tarim River [Figure 1: see original paper]. Based on land use data from the Chinese Academy of Sciences' Resource and Environmental Science Data Center, shrubland areas were extracted and interpreted using high-resolution remote sensing imagery. Field surveys conducted in July 2022 identified Tamarix nebkha as the dominant vegetation type, covering an area of 100.80 km^2.

1.2 Data Sources and Preprocessing

1.2.1 Field Biomass Data Sample collection was conducted in July 2022. Landsat 8 image raster data were converted to vector polygons for precise spatial matching. During field sampling, 30 sample plots of 30 m × 30 m were established based on pixel spatial locations. Aboveground biomass of Tamarix shrubs was surveyed within each plot. AGB refers to the biomass of above-ground parts (branches and leaves) on the sand dune surface. For each shrub, three representative standard branches were randomly selected to measure canopy projection area. Branches were cut at the base, oven-dried to constant weight, and weighed. The ratio between canopy area and biomass was established using standard branches, and total shrub AGB was extrapolated using the average biomass per unit canopy area derived from high-resolution imagery.

1.2.3 Feature Variables Landsat 8 imagery (spatial resolution: 30 m) was obtained from Google Earth Engine (https://earthengine.google.com/). Preprocessing included radiometric calibration, atmospheric correction, and clipping. Three categories of remote sensing variables were extracted: original spectral band reflectance, vegetation indices, and texture indices, totaling 92 features. The seven spectral bands included coastal, blue, green, red, near-infrared, and two shortwave infrared bands. Twelve vegetation indices suitable for sparse vegetation in arid regions were selected . Texture information was derived using the gray-level co-occurrence matrix (GLCM) from bands B1–B7, calculating eight texture features: mean, variance, contrast, correlation, dissimilarity, homogeneity, second moment, and entropy. All preprocessing and feature extraction were performed using ENVI 5.3.

1.3 Modeling Approaches

1.3.1 Variable Selection Methods Model accuracy in biomass estimation depends heavily on feature variable selection. This study employed three methods: stepwise regression, LASSO, and XGBoost. Stepwise regression iteratively selects variables with significant contributions while eliminating redundant or non-significant variables ($P > 0.05$), reducing complexity while maintaining accuracy. LASSO achieves sparse linear model estimation through coefficient shrinkage and regularization, driving some coefficients to zero for feature selection. XGBoost combines gradient boosting with regularization, assessing feature importance through contributions to the loss function during tree construction, offering strong overfitting prevention for complex tasks.

1.3.2 Model Construction Three machine learning models were developed: Random Forest (RF), Support Vector Regression (SVR), and Backpropagation Neural Network (BPNN). RF generates multiple bootstrap samples to build independent decision trees, aggregating predictions through voting, making it suitable for high-dimensional data and effective at reducing overfitting. SVR uses kernel functions to map data into high-dimensional space, fitting a hyperplane to minimize error, excelling at capturing non-linear relationships. BPNN iteratively adjusts weights through backpropagation to minimize output-target error, demonstrating strong adaptability and fault tolerance for complex non-linear problems. Model parameters are summarized in .

1.3.3 Model Accuracy Assessment Ten-fold cross-validation was employed to systematically evaluate model stability and predictive reliability. Samples were randomly divided into ten subsets; nine subsets trained the model while one subset tested it, repeating ten times. Each sample participated in testing once. Model accuracy was assessed using coefficient of determination (R^2) and root mean square error (RMSE). Final performance metrics represent averages across all test sets:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{y})^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{y}_i)^2}{n}}$$

where Y_i is the measured AGB ($g \cdot m^{-2}$), $\hat{y}_i$ is the estimated AGB ($g \cdot m^{-2}$), $\bar{y}$ is the mean measured AGB, and n is the sample size.

2 Results and Analysis

2.1 Variable Set Selection

Normality testing confirmed that measured Tamarix nebkha AGB data approximated a normal distribution (mean = 1369.99 $g \cdot m^{-2}$, SD = 996.04 $g \cdot m^{-2}$), satisfying parametric modeling assumptions [Figure 3: see original paper].

Stepwise regression selected a variable set comprising MSAVI, Correlation_B4, Correlation_B5, and Correlation_B6 . All variables showed VIF < 5, indicating no multicollinearity. The LASSO algorithm' s regularization path showed coefficient shrinkage as λ increased [Figure 4: see original paper]. Variables retained at one standard error included Correlation_B5, Correlation_B6, and NDVI_{563}. XGBoost ranked feature importance, with Correlation_B6 showing the highest importance score (>0.15), followed by other texture correlation indices [Figure 5: see original paper]. Seven variable sets (L1-L7) were constructed by incrementally adding the most important features . Comparative modeling revealed that the L7 variable set (containing Correlation_B6, Correlation_B3, Homogeneity_B5, and NDVI_{563}) achieved optimal performance across RF, SVR, and BPNN models, with highest R^2 and lowest RMSE [Figure 6: see original paper]. Therefore, the XGBoost-selected L7 variable set was adopted for final model construction.

2.2 Model Accuracy Comparison

Pearson correlation analysis validated significant relationships between selected variables and AGB [Figure 7: see original paper]. All three selection methods retained texture correlation indices (e.g., Correlation_B6) and vegetation indices calculated from red and near-infrared bands (e.g., MSAVI, NDVI_{563}), with correlations significant at $P < 0.01$. Variable sets from LASSO and XGBoost showed lower inter-variable correlations ($|r| < 0.65$) compared to stepwise regression, confirming effective variable selection.

Model performance varied significantly across variable selection methods and algorithms . Stepwise-based models showed R^2 ranging 0.64–0.68 and RMSE of 508.92-528.74 $g \cdot m^{-2}$. LASSO-based models improved to R^2 = 0.68-0.73 and RMSE = 447.63-487.21 $g \cdot m^{-2}$. XGBoost-based models achieved intermediate performance (R^2 = 0.66-0.70, RMSE = 483.67-511.04 $g \cdot m^{-2}$). The LASSO-based RF model performed best overall (R^2 = 0.73, RMSE = 447.63 $g \cdot m^{-2}$), demonstrating that optimal combinations of variable selection and machine learning algorithms substantially improve estimation accuracy.

Visualization of RF model predictions across variable sets confirmed that LASSO-selected variables produced the strongest alignment between predicted and measured values in both training and test sets [Figure 8: see original paper], indicating superior systematic variable selection and stable results.

2.3 AGB Estimation

The LASSO-based RF model was applied to estimate Tamarix nebkha AGB across the study area [Figure 9: see original paper]. Results showed a mean AGB of 1733.63 $g \cdot m^{-2}$, ranging from 2629.17 $g \cdot m^{-2}$ to 996.04 $g \cdot m^{-2}$, with total biomass of 1.71×10^8 kg. Spatial distribution exhibited clear heterogeneity, with high AGB concentrated near tributaries of the Tarim River, indicating significant hydrological influence. Areas distant from river channels showed reduced AGB due to severe drought stress.

---

3 Discussion

3.1 Importance of Variable Selection for AGB Estimation

Remote sensing-based vegetation biomass estimation often relies on multi-source features (spectral, vegetation, and texture indices), but high inter-feature correlation can induce multicollinearity, compromising model stability and prediction accuracy. Appropriate feature selection is therefore critical for reliable AGB estimation. This study employed three distinct methods to ensure robust variable selection.

All three methods consistently prioritized vegetation indices constructed from red and near-infrared bands (e.g., MSAVI, NDVI_{563}), confirming their importance for Tamarix nebkha AGB estimation. Spectrally, red bands capture chlorophyll absorption sensitive to photosynthetic activity, while near-infrared bands respond to canopy structure. The resulting vegetation indices integrate absorption and scattering information, simultaneously reflecting physiological status and structural characteristics, making them particularly sensitive to biomass changes in arid desert backgrounds. This aligns with Adame Campos et al.' s findings on mixed pine-oak forests in central Mexico, demonstrating consistent spectral response patterns across ecosystems.

Machine learning-based feature selection (LASSO, XGBoost) outperformed traditional stepwise regression, effectively identifying the most contributive environmental factors under complex, high-dimensional, and multicollinear conditions. This advantage has been widely recognized in recent research. For instance, Huang et al. demonstrated that LASSO-based feature selection significantly improved grassland AGB model accuracy and stability. Similarly, Li et al. showed that machine learning algorithms could identify optimal variable sets from high-dimensional hyperspectral data, substantially enhancing forest AGB estimation. These results underscore the superiority of machine learning-based variable selection for feature optimization and parameter identification in complex ecosystems.

3.2 Regional Applicability of Tamarix Nebkha AGB Models

All three models demonstrated satisfactory fitting capability for Tamarix nebkha AGB estimation, though accuracy varied significantly. The RF model consistently outperformed SVR and BPNN across all variable selection methods, with higher average R^2 and lower RMSE, highlighting its advantages in handling non-linear relationships and multi-dimensional features. This conclusion aligns with numerous studies on AGB estimation in arid oases, forests, and grasslands, confirming RF's superior ability to extract non-linear vegetation structural features and its enhanced stability and regional adaptability.

From a feature selection perspective, LASSO and XGBoost improved prediction accuracy across all models, with the most pronounced enhancement in RF models. In the optimal LASSO-based RF model, vegetation indices (NDVI_{563}, MSAVI) served as primary predictors, while texture features (Correlation_B5, Correlation_B6) provided complementary information on branch density and distribution patterns, mitigating saturation effects inherent to spectral features alone. This demonstrates that rational variable selection effectively eliminates redundancy, enhances model stability, and maximizes the predictive potential of machine learning algorithms.

However, the unique characteristics of the study area—sparse vegetation and strong wind-sand environments—may influence model performance. Future research should integrate multi-source remote sensing data (e.g., high-resolution UAV imagery, radar data) to leverage complementary advantages for optimized feature extraction and model construction. Expanding the study area and sample size to validate model applicability across diverse environmental conditions would further enhance practical utility. Overall, this study provides a scientific foundation for remote sensing inversion methods in desert ecosystems, confirming the feasibility and effectiveness of multi-feature fusion and machine learning algorithms in arid region vegetation research.

---

4 Conclusions

This study investigated Tamarix nebkha AGB estimation using Landsat 8 imagery, three variable selection methods (stepwise regression, LASSO, XGBoost), and three machine learning models (RF, SVR, BPNN). Key conclusions are:

1) All three variable selection algorithms identified variable sets significantly correlated with Tamarix nebkha AGB, including texture indices (e.g., Correlation_B6) and vegetation indices from red and near-infrared bands (e.g., MSAVI, NDVI_{563}). Low multicollinearity ($VIF < 5$) and significant correlations ($P < 0.01$) confirmed algorithm accuracy and effectiveness.

2) Models based on LASSO and XGBoost achieved significantly higher accuracy than stepwise regression. The LASSO-based RF model performed

best ($R^2 = 0.73$, RMSE = 447.63 $g \cdot m^{-2}$). Multivariate combinations substantially enhanced predictive capability, demonstrating that optimal variable selection is crucial for improving Tamarix nebkha AGB estimation accuracy.

3) Across variable selection methods, RF models showed superior fitting performance, particularly for independent variable fitting. The LASSO-based RF model estimated mean Tamarix nebkha AGB in the lower Tarim River region at 1733.63 $g \cdot m^{-2}$, with total biomass of 1.71×10^8 kg. The model objectively reflects AGB content with high reliability and regional applicability.

These results provide theoretical and practical references for remote sensing estimation of shrub AGB in desert ecosystems.

References

[1] Tang X L, Zhao X, Bai Y F, et al. Carbon pools in China' s terrestrial ecosystems: New estimates based on an intensive field survey[J]. Proceedings of the National Academy of Sciences, 2018, 115(16): 4021-4026.

[2] Maestre F T, Eldridge D J, Soliveres S, et al. Structure and functioning of dryland ecosystems in a changing world[J]. Annual Review of Ecology Evolution and Systematics, 2016, 47(1): 215-237.

[3] Mahdianpari M, Salehi B, Mohammadimanesh F, et al. Fisher linear discriminant analysis of coherency matrix for wetland classification using PolSAR imagery[J]. Remote Sensing of Environment, 2018, 206: 300-317.

[4] Liang B, Di L, Zhao C Y, et al. Altitude distribution of aboveground biomass of typical shrubs in the Tianlaochi watershed of Qilian Mountains[J]. Acta Agrestia Sinica, 2013, 21(4): 664-669.

[5] Li D R, Wang C W, Hu Y M, et al. General review on remote sensing based biomass estimation[J]. Geomatics and Information Science of Wuhan University, 2012, 37(6): 631-635.

[6] Shi Y L, Wang Z H, Zhang G J, et al. Evaluating the research status of the remote sensing mediated monitoring of forest biomass: A bibliometric analysis of WOS[J]. Forests, 2024, 15(3): 524.

[7] Fernández Alonso J M, Llorens R, Sobrino J A, et al. Exploring the potential of Lidar and Sentinel-2 data to model the structural characteristics of gorse shrublands in NW Spain[J]. Remote Sensing, 2022, 14(23): 6063.

[8] Anees S A, Mehmood K, Khan W R, et al. Integration of machine learning and remote sensing for above ground biomass estimation through Landsat-9 and field data in temperate forests of the Himalayan region[J]. Ecological Informatics, 2024, 82: 102732.

[9] Xu Z P, Shu L L, Wu W G, et al. Classification of Shangri-La grasslands based on Google Earth Engine and remote sensing estimation of their biomass[J]. Pratacultural Science, 2024, 41(10): 2250-2262.

[10] Liu W J, Xu C, Zhang Z M, et al. Machine learning based grassland above-ground biomass estimation and its response to climate variation in Southwest China[J]. Frontiers in Ecology and Evolution, 2023, 11: 1146850.

[11] Chen M G, Yin C X, Lin T, et al. Integration of unmanned aerial vehicle spectral and textural features for accurate above ground biomass estimation in cotton[J]. Agronomy, 2024, 14(6): 1313.

[12] Zhang J L, Zhang L, Zhang X H, et al. Inversion of aboveground biomass of desert shrub vegetation in Junggar Basin[J]. Journal of Xinjiang Agricultural University, 2019, 42(3): 202-209.

[13] Gao M L, Gong Z N, Zhao W J, et al. The study of *Vitex negundo* shrubs canopy biomass inversion in Beijing Jundu mountainous area based on vegetation indices[J]. Acta Ecologica Sinica, 2014, 34(5): 1178-1188.

[14] Wang C L, Guo P, Wang B, et al. Construction of shrub biomass estimation model based on hyper spectrum in the eastern Qilian Mountains[J]. Grassland and Turf, 2020, 40(6): 39-45.

[15] Yan X G, Li J, Smith A R, et al. Evaluation of machine learning methods and multi-source remote sensing data combinations to construct forest above-ground biomass models[J]. International Journal of Digital Earth, 2023, 16(2): 4471-4491.

[16] Luo M, Wang Y F, Xie Y H, et al. Combination of feature selection and CatBoost for prediction: The first application to the estimation of aboveground biomass[J]. Forests, 2021, 12(2): 216.

[17] Zhang X F, Li X Y, Sharma R P, et al. Forest above-ground biomass estimation based on strongly collinear variables derived from airborne laser scanning data[J]. Ecological Indicators, 2024, 166: 112517.

[18] Cui L H, Zheng S, Xu M. Remote sensing inversion of forest and grassland aboveground biomass in Inner Mongolia, China[J]. Geographical Science, 2024, 44(12): 2215-2224.

[19] Li Y C, Li C, Li M Y, et al. Influence of variable selection and forest type on forest aboveground biomass estimation using machine learning algorithms[J]. Forests, 2019, 10(12): 1073.

[20] Zandler H, Brenning A, Samimi C. Quantifying dwarf shrub biomass in an arid environment: Comparing empirical methods in a high dimensional setting[J]. Remote Sensing of Environment, 2015, 158: 140-155.

[21] Li Z Z, Bi S D, Hao S, et al. Aboveground biomass estimation in forests with random forest and monte carlo based uncertainty analysis[J]. Ecological Indicators, 2022, 142: 109246.

[22] Huang T B, Ou G L, Xu H, et al. Comparing algorithms for estimation of aboveground biomass in *Pinus yunnanensis*[J]. Forests, 2023, 14(9): 3550.

[23] Xie D B, Huang H C, Feng L Y, et al. Aboveground biomass prediction of arid shrub dominated community based on airborne LiDAR through parametric and nonparametric methods[J]. Remote Sensing, 2023, 15(13): 3344.

[24] Tamiminia H, Salehi B, Mahdianpari M, et al. Comparison of machine and deep learning methods to estimate shrub willow biomass from UAS imagery[J]. Canadian Journal of Remote Sensing, 2021, 47(2): 209-227.

[25] Wang X M, Yang X F, Zhao F, et al. Estimation of aboveground biomass in the arid oasis based on the machine learning algorithm[J]. Ecology and Environmental Sciences, 2023, 32(6): 1007-1015.

[26] Ding J Q, Huang W L, Liu Y C, et al. Estimation of forest aboveground biomass in Northwest Hunan Province based on machine learning and multi-source data[J]. Scientia Silvae Sinicae, 2021, 57(10): 36-48.

[27] Ge J, Hou M J, Liang T G, et al. Spatiotemporal dynamics of grassland aboveground biomass and its driving factors in North China over the past 20 years[J]. Science of the Total Environment, 2022, 826: 154226.

[28] Jiang R G, Liang J C, Zhao Y, et al. Assessment of vegetation growth and drought conditions using satellite based vegetation health indices in Jing-Ji region of China[J]. Scientific Reports, 2021, 11(1): 13775.

[29] Cai T Y, Chang C C, Zhao Y B, et al. Within season estimates of 10 m aboveground biomass based on Landsat, Sentinel-2 and PlanetScope data[J]. Scientific Data, 2024, 11(1): 1276.

[30] Chen L, Hao W Q, Gao D L. The latest applications of optical image texture in forestry[J]. Journal of Beijing Forestry University, 2015, 37(3): 1-12.

[31] Li Y C, Li M Y, Li C, et al. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms[J]. Scientific Reports, 2020, 10(1): 9952.

[32] Sun L K, Liu W Q, Chen T, et al. Review on mechanism of habitat adaptability and resource value of *Tamarix* species[J]. Journal of Desert Research, 2016, 36(2): 349-356.

[33] Huang S, Tang L N, Hupy J P, et al. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing[J]. Journal of Forestry Research, 2020, 32(1): 1-6.

[34] Wang Z J, Xu H, He J L, et al. Additive dynamic biomass models for *Salix psammophila* in the arid sandy area in Ningxia[J]. Journal of Northeast Forestry University, 2019, 47(9): 52-57.

[35] Bai Y F, Chen C Q, Xu H L, et al. Relationship between spatial distribution of aboveground biomass of desert vegetation and groundwater depth in the lower

reaches of Tarim River, Xinjiang, China[J]. Scientia Silvae Sinicae, 2016, 52(11): 1-10.

[36] Hu R Y, Chang J X, Guo A J, et al. Ecosystem variation and ecological benefits analysis of the mainstream of Tarim River[J]. Arid Land Geography, 2024, 47(4): 622-633.

[37] Adam E, Mutanga O, Abdel Rahman E M, et al. Estimating standing biomass in papyrus (*Cyperus papyrus* L.) swamp: Exploratory of in situ hyperspectral indices and random forest regression[J]. International Journal of Remote Sensing, 2014, 35(2): 693-714.

[38] Lin W K, Lu Y G, Li G Y, et al. A comparative analysis of modeling approaches and canopy height based data sources for mapping forest growing stock volume in a northern subtropical ecosystem of China[J]. GIScience & Remote Sensing, 2022, 59(1): 568-589.

[39] Luo M, Anees S A, Huang Q Y, et al. Improving forest aboveground biomass estimation by integrating individual machine learning models[J]. Forests, 2024, 15(6): 975.

[40] Hengl T, Mendes de Jesus J, Heuvelink G B, et al. SoilGrids 250 m: Global gridded soil information based on machine learning[J]. Plos One, 2017, 12(2): e0169748.

[41] Singh A, Kushwaha S K P, Nandy S, et al. Aboveground forest biomass estimation by the integration of TLS and ALOS PALSAR data using machine learning[J]. Remote Sensing, 2023, 15(4): 1143.

[42] Sharifi A, Amini J, Tateishi R. Estimation of forest biomass using multivariate relevance vector regression[J]. Photogrammetric Engineering & Remote Sensing, 2016, 82(1): 41-49.

[43] Zhang D H, Ni H. Inversion of forest biomass based on multi-source remote sensing images[J]. Sensors, 2023, 23(23): 9313.

[44] Willmott C J, Ackleson S G, Davis R E, et al. Statistics for the evaluation and comparison of models[J]. Journal of Geophysical Research: Oceans, 2012, 90(C5): 8995-9005.

[45] Yu X H, Ge H L, Lu D S, et al. Comparative study on variable selection approaches in establishment of remote sensing model for forest biomass estimation[J]. Remote Sensing, 2019, 11(12): 1437.

[46] Fernández Manso O, Fernández Manso A, Quintano C. Estimation of aboveground biomass in Mediterranean forests by statistical modelling of ASTER fraction images[J]. International Journal of Applied Earth Observation and Geoinformation, 2014, 31: 45-56.

[47] Adame Campos R L, Ghilardi A, Gao Y, et al. Variables selection for aboveground biomass estimations using satellite data: A comparison between relative

importance approach and stepwise Akaike information criterion[J]. ISPRS International Journal of Geo-Information, 2019, 8(6): 245.

[48] Zhou D J, Chahal R, Gotlib I H, et al. Comparison of lasso and stepwise regression in psychological data[J]. Methodology, 2024, 20(2): 121-143.

[49] Wang E, Huang T B, Liu Z, et al. Improving forest above ground biomass estimation accuracy using multi-source remote sensing and optimized least absolute shrinkage and selection operator variable selection method[J]. Remote Sensing, 2024, 16(23): 4497.

[50] Huang W Y, Li W L, Xu J, et al. Hyperspectral monitoring driven by machine learning methods for grassland above ground biomass[J]. Remote Sensing, 2022, 14(9): 2086.

[51] Song K X, Jiang F G, Hu Z D, et al. Remote sensing inversion of aboveground biomass of grassland in the Xizang Autonomous Region[J]. Acta Ecologica Sinica, 2023, 43(13): 5600-5613.

[52] Liu W L, Fa H J, Ba Y J, et al. Estimation of the aboveground biomass of mangrove forest in Zhenzhu Bay, Guangxi[J]. Marine Geology Frontiers, 2024, 40(11): 35-45.

[53] Zhou J, Zan M, Zhai L L, et al. Remote sensing estimation of aboveground biomass of different forest types in Xinjiang based on machine learning[J]. Sustainable Horizons, 2025, 15(1): 6187.

Note: Figure translations are in progress. See original paper for figures.