

Article

Estimating Aboveground Biomass of Wetland Plant Communities from Hyperspectral Data Based on Fractional-Order Derivatives and Machine Learning

Huazhe Li ^{1,2,3,4}, Xiyang Tang ^{1,2,3,4}, Lijuan Cui ^{1,2,3,4}, Xiajie Zhai ^{1,2,3,4}, Junjie Wang ^{5,6} , Xinsheng Zhao ^{1,2,3,4}, Jing Li ^{1,2,3,4}, Yinru Lei ^{1,2,3,4}, Jinzhi Wang ^{1,2,3,4}, Rumiao Wang ^{1,2,3,4} and Wei Li ^{1,2,3,4,*} 

¹ Institute of Wetland Research, Chinese Academy of Forestry, Beijing 100091, China

² Beijing Key Laboratory of Wetland Services and Restoration, Beijing 100091, China

³ Institute of Ecological Conservation and Restoration, Chinese Academy of Forestry, Beijing 100091, China

⁴ Beijing Hanshiqiao National Wetland Ecosystem Research Station, Beijing 101399, China

⁵ College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, China

⁶ MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area, Guangdong Key Laboratory of Urban Informatics, Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China

* Correspondence: wetlands207@caf.ac.cn; Tel.: +86-010-6282-4186

Abstract: Wetlands, as a crucial component of terrestrial ecosystems, play a significant role in global ecological services. Aboveground biomass (AGB) is a key indicator of the productivity and carbon sequestration potential of wetland ecosystems. The current research methods for remote-sensing estimation of biomass either rely on traditional vegetation indices or merely perform integer-order differential transformations on the spectra, failing to fully leverage the information complexity of hyperspectral data. To identify an effective method for estimating AGB of mixed-wetland-plant communities, we conducted field surveys of AGB from three typical wetlands within the Crested Ibis National Nature Reserve in Hanzhong, Shaanxi, and concurrently acquired canopy hyperspectral data with a portable spectrometer. The spectral features were transformed by applying fractional-order differentiation (0.0 to 2.0) to extract optimal feature combinations. AGB prediction models were built using three machine learning models, XGBoost, Random Forest (RF), and CatBoost, and the accuracy of each model was evaluated. The combination of fractional-order differentiation, vegetation indices, and feature importance effectively yielded the optimal feature combinations, and integrating vegetation indices with feature bands enhanced the predictive accuracy of the models. Among the three machine-learning models, the RF model achieved superior accuracy using the 0.8-order differential transformation of vegetation indices and feature bands ($R^2 = 0.673$, RMSE = 23.196, RPD = 1.736). The optimal RF model was visually interpreted using Shapley Additive Explanations, which revealed that the contribution of each feature varied across individual sample predictions. Our study provides methodological and technical support for remote-sensing monitoring of wetland AGB.

Keywords: aboveground biomass; hyperspectral data; fractional-order derivative; machine learning; Shapley Additive Explanations



Citation: Li, H.; Tang, X.; Cui, L.; Zhai, X.; Wang, J.; Zhao, X.; Li, J.; Lei, Y.; Wang, J.; Wang, R.; et al. Estimating Aboveground Biomass of Wetland Plant Communities from Hyperspectral Data Based on Fractional-Order Derivatives and Machine Learning. *Remote Sens.* **2024**, *16*, 3011. <https://doi.org/10.3390/rs16163011>

Academic Editor: Melanie Vanderhoof

Received: 25 May 2024

Revised: 13 August 2024

Accepted: 13 August 2024

Published: 16 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the context of global warming, the enhancement of ecosystem carbon-sequestration capabilities has received increasing attention [1,2]. Wetlands, one of the three major terrestrial ecosystems, not only play a crucial role in water and soil conservation and regional climate amelioration, but also have significant carbon storage capacity, earning them the moniker “kidneys of the Earth” [3–5]. Although wetlands occupy only 5%–8% of the global terrestrial area, their carbon storage rate per unit area is higher than that of other terrestrial

ecosystems [6,7]. In the context of climate change, it is critical to pay attention to the carbon research of wetland ecosystems [8]. Wetland vegetation, a key component of wetland ecosystems, plays an indispensable role in carbon sequestration [9,10]. The biomass of wetland vegetation is a critical indicator for estimating wetland carbon stocks [11–13]. Therefore, conducting systematic surveys and long-term monitoring of wetland vegetation biomass are of great significance for the rational development and utilization of wetland resources [14]. This approach helps to better understand and assess the role of wetlands in the global carbon cycle and is also crucial for devising effective ecological protection and restoration strategies.

Conventional methods for obtaining aboveground biomass (AGB) rely primarily on field surveys. Although these methods are accurate and reliable for small-scale studies, they are not suitable for dynamic and large-area AGB estimations, and their applicability is greatly limited. The development of remote-sensing technology offers new possibilities for solving these problems. Models were devised that effectively estimate AGB by capturing spectral information that is closely related to ground objects and biomass [15]. Significant progress has been made in AGB estimation studies based on multispectral remote-sensing data [8], mainly focusing on predicting the AGB using single-band reflectance and calculated vegetation indices, namely the normalized differenced vegetation index (NDVI), enhanced vegetation index (EVI), and difference vegetation index (DVI) [12,16–18]. These studies have laid an important foundation for remote-sensing estimation of AGB. Nevertheless, multispectral remote sensing does not always capture key information in complex habitats because of the limited number of bands [19–21]. In contrast, hyperspectral remote-sensing technology, with its higher spectral resolution, yields more effective results in complex environments. Currently, hyperspectral remote-sensing technology is widely applied in the measurement and estimation of several environmental parameters, including biomass, chlorophyll content, functional traits, and soil nutrient content prediction [22–27].

Although hyperspectral remote-sensing technology provides rich spectral information, it also introduces issues with data dimensionality, making the selection of feature variables a key challenge in AGB estimation [28,29]. Data dimensionality can be reduced via two methods: feature transformation and feature selection. Feature transformation, such as vegetation indices [30–32], reduces data dimensions by transforming data, and feature selection reduces the dimensions of the dataset by selecting the most representative and informative features, for example, using correlation analysis and feature selection by importance [33,34]. Additionally, the selection of an estimation model is also a key step to improving the accuracy of AGB remote-sensing estimation. Machine learning technology, which exhibits strong computational capabilities for processing large numbers of remote-sensing data features, has been widely applied in vegetation biomass modeling [35,36]. Jacon et al. [28] explored the effects of five machine learning models (Classification and Regression Trees, Cubist, Partial Least-Squares Regression, Random Forest and Support Vector Machine) and four groups of indicators (reflectance, narrowband vegetation indices, absorption band parameters and their combination) in predicting AGB, and found that a Random Forest (RF) model based on vegetation indices provided the most stable AGB predictions. Guo et al. [37] proposed a method using Non-Negative Matrix Factorization (NMF) for the differential fusion of multi-source remote-sensing data and used the RF model for AGB estimation and accuracy assessment, with the result that the NMF-based differential fusion model had excellent inversion effects, with an R^2 of 0.6 and RMSE of 586.56 kg/ha. Huang et al. [29] compared the performance of Support Vector Machines, Least-Squares Boosting, and Gaussian Process Regression (GPR) models in AGB prediction, with the GPR model achieving the best results ($R^2 = 0.53$; RMSE = 19.837). The authors also confirmed that combining spectral features with the growing season could significantly improve the inversion accuracy of AGB. These results show that combining various variable-selection methods with different regression methods and incorporating environmental factors are important for developing models that can predict biomass more effectively.

The complex ecological conditions of wetlands can impact canopy spectra, providing challenges for extracting key spectral features related to wetland plant AGB. Previous studies have successfully extracted characteristic information closely related to wetland plant AGB through hyperspectral data and achieved AGB estimation [38,39], accumulating a wealth of practical experience and providing a reliable theoretical basis for estimating wetland plant biomass. Wang et al. [40] used an OSH hyperspectral sensor onboard the Zhuhai-1 satellite to apply an empirical linear method to invert reed biomass in the Hanzhong Wetland Nature Reserve in Beijing. Hemati et al. [41] developed a model for estimating AGB using AVIRIS-NG hyperspectral data and then scaled it up by combining Sentinel-1 and Sentinel-2 data to build a large-scale and multi-temporal AGB estimation model. Li et al. [42] assessed the accuracy of predicting reed biomass through univariate linear regression and partial least-squares regression using the original reflectance, first-order differential spectra, and trilateral parameters. The results demonstrated the optimal accuracy of the partial least-squares model in predicting fresh biomass ($R^2 = 0.89$, RMSE = 0.43). Dou et al. [43] selected the original spectral reflectance and vegetation indices and established biomass inversion models through linear and partial least-squares regression, confirming a greater degree of accuracy of the latter ($R^2 > 0.85$, RMSE% < 5.6%). Chen et al. [44] successfully implemented AGB estimation and mapping based on in situ hyperspectral features combined with Sentinel-2 data using a generative adversarial network with a constrained factor model (GAN-CF). Furthermore, to improve the inversion accuracy of wetland plant biomass, researchers have attempted to use radar data to fuse multi-source remote-sensing data [45,46]; this approach is an important development in hyperspectral inversion of wetland biomass. In summary, while existing research has primarily focused on homogeneous dominant-species communities, studies on complex mixed-species plant communities are lacking. Additionally, the use of hyperspectral data is currently limited to fixed narrow-band vegetation indices or enhancement of spectral features through integer-order differential transformations. The implementation of fractional-order differentiation provides an effective method for enhancing features in hyperspectral data [47], offering more detailed adjustments to the spectral curves, thus revealing subtle spectral features. How to fully leverage the complex information provided by hyperspectral data and more effectively highlight spectral features are pressing issues that warrant further research.

Therefore, we conducted field surveys in three typical wetland habitats within the Crested Ibis National Nature Reserve in Hanzhong, Shaanxi, China, to obtain AGB data for vegetation and, concurrently, acquire the canopy spectra of the communities. We analyzed the spectral features under different fractional-order differential transformations and the results of different feature-selection algorithms, and built AGB estimation models using XGBoost, RF, and CatBoost. The performance of the models was evaluated, and the output of the optimal estimation model was visualized through Shapley Additive Explanations (SHAP). The results of this study provide a new approach and a technical reference for using hyperspectral remote-sensing technology to determine the AGB of mixed-plant communities in wetlands, with the goal of large-scale wetland monitoring.

2. Materials and Methods

2.1. Study Area

The Crested Ibis National Nature Reserve (107°21'–107°44'E, 33°08'–33°35'N) is located in the southwestern part of Shanxi Province. It is bordered by the Qinling Mountains to the north and the Bashan Mountains to the south, and is situated on the edge of the Hanzhong Basin. The terrain features steep elevations in the northeast, gentle slopes in the south, and flat terrain in the middle. The reserve is located in a transitional zone from warm-temperate to northern-subtropical regions, characterized by a continental monsoon climate with distinct seasonal changes and simultaneous periods of heat and rain. The average annual temperature is 14.5 °C, with the hottest average temperature in July being 25.9 °C, and the coldest average temperature in January being 2.2 °C; the average annual precipitation is 839.7 mm. The reserve, with an area of approximately 37,549 hectares, hosts

approximately 321 species of woody plants from 152 genera across 72 families. The wetlands within the Crested Ibis National Nature Reserve comprise various types of wetland complexes, primarily consisting of wet meadows, with some areas exhibiting characteristics of shallow-water wetlands. The vegetation in these wetlands is predominantly herbaceous, with few woody plants, and the main vegetation types differ slightly across the locations. We selected three wet-meadow areas as our study sites, all of which are characterized by herbaceous vegetation and lack open water (Figure 1). Site 1 is at an average elevation of 680 m, with dominant species including *Echinochloa crus-galli*, *Oenanthe javanica*, and *Sagittaria trifolia*. The vegetation height ranges from 0.3 to 0.7 m. Site 2 is at an average elevation of 789 m, with dominant species including *Oenanthe javanica*, *Equisetum hyemale*, and *Juncus effusus*. The vegetation height ranges from 0.5 to 0.7 m. Site 3 is at an average elevation of about 585 m, with dominant species including *Echinochloa crus-galli*, *Equisetum hyemale*, and *Polygonum hydropiper*. The vegetation height ranges from 0.4 to 0.8 m.

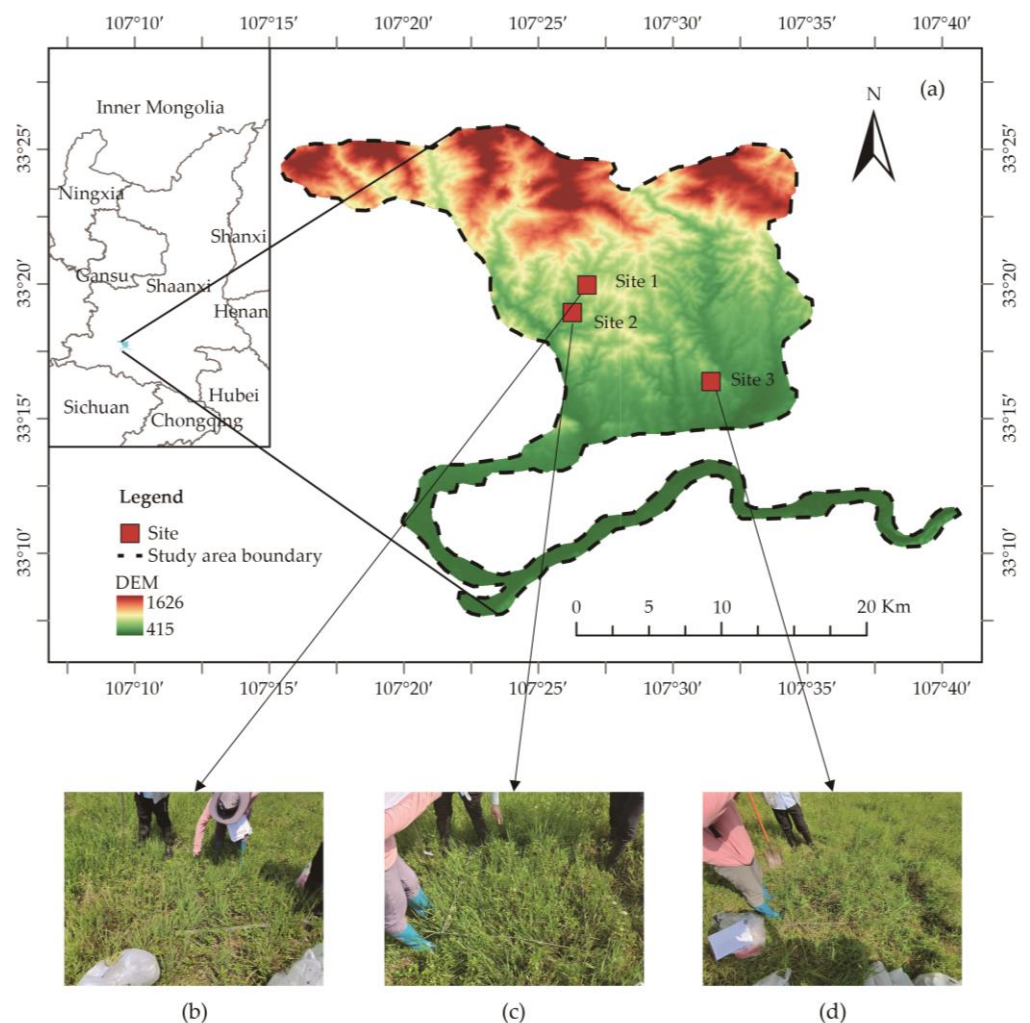


Figure 1. Study area; (a) Location of study area; (b) Survey Site 1; (c) Survey Site 2; (d) Survey Site 3.

2.2. Data Acquisition and Processing

The field data were collected between 21 July and 26 August 2022. The survey was based on the belt transect method and typical sample plots, totaling 102. The size of each plot was set to 1 m², with a minimum distance of 10 m between plots.

2.2.1. Hyperspectral Data Acquisition and Pre-Processing

Vegetation canopy spectral data were acquired using a FieldSpec-4 portable hyperspectral spectrometer (ASD, Analytical Spectral Devices, Inc., Boulder, CO, USA), covering

a spectral range of 350–2500 nm. Spectral data were collected under no-wind, clear, and slightly cloudy conditions, between 11:00 AM and 2:00 PM each day, to ensure a sufficient solar-elevation angle. During the measurement, the operator faced the direction of the incoming sunlight and positioned the sensor probe vertically, at approximately 1 m above the canopy, to avoid shadows, maintaining a constant field-of-view angle of 25°. Canopy spectral reflectance was collected ten times at each sample point, with the reflectance calibrated using a white reference panel every 15 min. After data collection, ten spectral curves from each sample point were averaged using ViewSpec Pro software (version 5.6), and the averaged spectra were exported as reflectance data, with a spectral sampling interval of 1 nm to represent the canopy spectral data for that point.

To avoid the effects of system noise and water vapor absorption on the canopy spectrum, the spectral range was set to 400–1350 nm for subsequent analyses. The Savitzky–Golay filtering algorithm was applied to smooth the vegetation-community-canopy spectrum and improve the signal-to-noise ratio using a second-order polynomial and a window size of seven [48].

2.2.2. AGB Data Acquisition

After collecting the hyperspectral data, destructive ground sampling was simultaneously conducted to obtain AGB data from the sample plots. This method involved cutting the AGB vegetation within the plots and drying it in a laboratory. The fresh weight of the vegetation was measured, and the material was wilted at 105 °C for 30 min and subsequently dried at 65 °C to a constant weight. Dry weight was measured using a high-precision balance (BSA2202s-CW, Sartorius AG, Göttingen, Germany). AGB is expressed as the dry weight of plants per unit area. The statistical results for the three locations are shown in Table 1. According to the ANOVA analysis results, the 102 collected AGB data points were arranged in ascending order, and every three samples were grouped. The middle data point from each group was selected as a validation sample, with the remaining data points used as training samples. Using this method, the AGB data were divided into 68 training samples and 34 validation samples.

Table 1. Aboveground Biomass Statistical Results (g/m²).

Site	Number of Samples	Maximum	Minimum	Average	Standard Deviation	Significant Difference
Site 1	20	219.06	108.56	157.26	32.84	a
Site 2	49	252.04	81.35	150.19	41.72	a
Site 3	33	280.02	96.71	167.87	41.32	a

Note: Significant differences were analyzed using one-way ANOVA. Sites with the same letter indicate no significant difference at the $p < 0.05$ level.

2.3. Fractional-Order Derivative

The fractional-order derivative, an extension of integer-order differentiation, was applied to remote-sensing data to sensitively capture changes in spectral reflectance details. This has been explored in numerous research fields [49,50], thereby proving its applicability. Currently, there is no unified formula to define fractional-order derivatives, and there are various definitions, including Riemann–Liouville, Caputo, and Grunwald–Letnikov (G-L). The most commonly used form is G-L [51], and our study employed the G-L form for differentiation. For hyperspectral data with a resolution of 1 nm, the differentiation formula is as follows:

$$\frac{d^u f(x)}{dx^u} \approx f(x) + (-u)f(x-1) + \frac{(-u)(-u+1)}{2}f(x-2) + \cdots + \frac{\Gamma(-u+1)}{m!\Gamma(-u+m+1)}f(x-m) \quad (1)$$

where $f(x)$ is the spectral reflectance; u represents the order of differentiation; and m is the length of hyperspectral data, ranging from 400 nm to 1350 nm. The fractional-order derivative was implemented using Matlab 2016a.

2.4. Selection of Vegetation Indices

We selected 15 vegetation indices that are widely used for AGB estimation (Table 2).

Table 2. Vegetation indices used in the study.

Vegetation Indices	Equation	Reference
VARI	$(R_{550} - R_{660}) / (R_{550} + R_{660} - R_{470})$	[32]
OSAVI	$1.16 \times (R_{800} - R_{670}) / (R_{800} + R_{670} + 0.16)$	[32]
MCARI	$((R_{700} - R_{670}) - 0.2 \times (R_{700} - R_{550})) \times (R_{700} / R_{670})$	[32]
TVI	$0.5 \times (520 \times (R_{750} - R_{550}) - 200 \times (R_{670} - R_{550}))$	[32]
EVI	$2.5 \times [(R_{800} - R_{270}) / (R_{800} + 6 \times R_{270} - 7.5 \times R_{475} + 1)]$	[52]
SAVI	$(1 + 0.5) \times (R_{800} - R_{670}) / (R_{800} + R_{670} + 0.5)$	[52]
RVI	R_{790} / R_{670}	[53]
CI green	$R_{780} / R_{550} - 1$	[54]
TCARI	$3 \times (R_{700} - R_{670}) - 0.2 \times (R_{700} - R_{550}) \times (R_{700} / R_{670})$	[54]
PRI	$(R_{531} - R_{570}) / (R_{531} + R_{570})$	[54]
CI red edge	$R_{800} / R_{740} - 1$	[54]
RSI	R_{825} / R_{735}	[54]
NDVI	$(R_{850} - R_{675}) / (R_{850} + R_{675})$	[55]
NDRE	$(R_{790} - R_{720}) / (R_{790} + R_{720})$	[55]
GNDVI	$(R_{750} - R_{550}) / (R_{750} + R_{550})$	[55]

Note: VARI, visible atmospherically resistant index; OSAVI, optimizing soil-adjusted vegetation index; MCARI, modified chlorophyll absorption ratio index; TVI, triangular vegetation index; EVI, enhanced vegetation index; SAVI, soil-adjusted vegetation index; RVI, ratio vegetation index; CI green, Green Chlorophyll Index; TCARI, transformed chlorophyll absorption reflectance index; PRI, photochemical reflectance index; CI red edge, red-edge chlorophyll index; RSI, ratio spectral index; NDVI, normalized difference vegetation index; NDRE, normalized difference red edge; GNDVI, green normalized-difference vegetation index.

2.5. Pearson Correlation Analysis

Correlation analysis is a method used to assess the statistical relationships between two or more variables, reflecting the degree of association through correlation coefficients. We used Python to calculate Pearson correlation coefficients to analyze the relationship between vegetation indices and AGB (Equation (2)). Vegetation indices that demonstrated strong correlations with AGB at a significance level of $p < 0.01$ were then selected as input variables for constructing the AGB prediction model.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

where x_i and y_i represent the reflectance values of the vegetation index and the AGB, respectively, and $\bar{x}$ and $\bar{y}$ represent the average values of the vegetation index and AGB for all samples, respectively. The correlation coefficient ranges from -1 to 1 . The larger the absolute value of the correlation coefficient, the stronger the correlation between the variables.

2.6. Optimal Feature Band Selection

Because of the large amount of information typically contained in hyperspectral data, which can lead to data redundancy and affect the model efficiency, it is necessary to reduce dimensionality by selecting the most important features [56]. We evaluated the importance of spectral band features under different fractional-order differentiation using two methods—Gini importance based on XGBoost and Mean Decrease Impurity (MDI) based on RF—and selected the top-20 most important bands as the feature bands.

In the RF model, the feature importance is usually assessed based on the MDI [57]. This method evaluates the importance of a feature by calculating the reduction in node impurities. The more impurity a feature reduces, the more important it is. Thus, the importance of each feature is calculated as an average across all trees in the RF model, yielding a final importance score for each feature.

In the XGBoost algorithm, feature importance can be assessed using various methods, including weight, gain, and cover [58]. By default, XGBoost uses gain as the criterion for assessing feature importance, which measures the average contribution of a feature when used as a split node to reduce model error. Through this method, the impact of a feature on the predictive performance is determined, which is crucial for feature selection and model optimization.

2.7. Regression-Model Construction and Evaluation

We employed three machine-learning regression techniques (RF, XGBoost, and CatBoost) to build regression models using the extracted feature data to estimate AGB.

RF is an ensemble learning method that builds and combines multiple decision trees. Each tree was built using random feature selection, making RF tolerant of outliers and noise in the data. This strategy not only enhances the robustness of the model, but also optimizes its generalizability, allowing RF to maintain a high predictive performance across various datasets [59,60].

XGBoost is a widely used machine learning algorithm for both regression and classification, and has demonstrated strong performance in numerous studies [61–63]. As an effective implementation of Gradient Boosting Decision Trees (GBDTs), XGBoost boosts model performance by building multiple decision trees, synthesizing their prediction results, and employing regularization strategies to prevent overfitting, thus enhancing the generalizability of the model [58].

CatBoost, introduced by Yandex in 2017, is an optimized gradient-boosting algorithm that refines the conventional GBDT. By sequentially integrating multiple base learners to ensure their dependency using “greedy” algorithms and ordered boosting techniques to optimize gradient transformation, and employing “oblivious trees” as predictors to reduce the risk of overfitting [64,65], CatBoost achieves excellent generalization capability and robustness, making it particularly suitable for regression tasks.

Machine learning models often contain multiple hyperparameters, the settings of which significantly affect model accuracy. Therefore, these hyperparameters must be optimized to determine the optimal model. In recent years, Bayesian optimization algorithms have gained attention and have been applied in various fields. This method treats the hyperparameter tuning process as a black-box function optimization and determines the next hyperparameter value based on previously obtained results, thereby avoiding many unnecessary evaluations [66]. Here, we employed the Bayesian optimization method to adjust the model hyperparameters. The specific hyperparameters and their value ranges are listed in Table 3. The model construction and Bayesian optimization in this study were implemented using Python 3.11.

Table 3. The hyperparameters of the model and their optimization ranges.

Model	Hyperparameter	Range
XGBoost	max_depth	(3, 10)
	learning_rate	(0.01, 0.1)
	n_estimators	(50, 500)
	gamma	(0, 0.5)
	min_child_weight	(1, 10)
	subsample	(0.6, 1.0)
	colsample_bytree	(0.6, 1.0)
RF	n_estimators	(50, 500)
	max_depth	(3, 10)
	min_samples_split	(0.01, 0.1)
	min_samples_leaf	(1, 10)
	max_features	(0.1, 1.0)

Table 3. Cont.

Model	Hyperparameter	Range
CatBoost	iterations	(10, 100)
	learning_rate	(0.01, 0.35)
	depth	(1, 11)
	l2_leaf_reg	(1, 11)

Lastly, to evaluate the predictive capability and inversion accuracy of each model, we selected several commonly used metrics to assess the performance of the machine learning models, including the coefficient of determination (R^2 ; Equation (3)), root mean square error (RMSE; Equation (4)), relative percent difference (RPD; Equation (5)), and index of agreement (IOA; Equation (7)). The closer R^2 and IOA are to 1, and the closer the RMSE is to 0, the better the predictive performance of the model [67]. In terms of RPD, a value less than 1.4 indicates that the model is not suitable for estimation; a value greater than 1.4 indicates good predictive ability; and greater than 2.0 indicates extremely high predictive accuracy [26].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (y_i - \bar{x})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \quad (4)$$

$$RPD = \frac{SD}{RMSE} \quad (5)$$

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{x})^2} \quad (6)$$

$$IOA = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (|y_i - \bar{y}| + |x_i - \bar{x}|)^2} \quad (7)$$

where n is the number of samples, y_i is the predicted AGB, x_i is the measured AGB, $\bar{y}$ is the average of the model-predicted AGB, and $\bar{x}$ is the average of the model-measured AGB.

2.8. Machine Learning Interpretability—Shapley Additive Explanations

The Shapley value was first proposed by Lloyd Shapley, a professor at the University of California, to address the issue of profit distribution in economic cooperation or game-theory contexts [68]. Specifically, the Shapley value is calculated by considering all possible combinations of participants, computing the incremental contribution (i.e., marginal contribution) to the total profit when each participant joins the combination, and then averaging these marginal contributions to determine each participant's Shapley value.

In machine learning models, the Shapley value is used as a method of model explanation, known as Shapley Additive Explanations (SHAP). It helps explain the contribution of each feature to the model's prediction outcome (i.e., the average marginal contribution of each feature), and is particularly suited for complex nonlinear models. The advantage of this method is that it not only assesses which variables are important, but also explains how variables influence the final prediction outcome [29]. This method was implemented using the “shap” package in Python, which includes various visualization tools, such as displaying feature importance or illustrating the contribution of features to the predicted values of samples.

3. Results

3.1. Spectral-Curve Characteristics of AGB

Figure 2 shows the spectral curves corresponding to the highest and lowest AGB at different sample sites. The high AGB values at all three sites correspond to spectral curves with high reflectance, featuring more distinct vegetation spectral characteristics. Conversely,

the low AGB values correspond to spectral curves with low reflectance, associated with sparser vegetation distributions influenced by aquatic backgrounds. Low AGB values correspond to lower vegetation coverage, resulting in greater exposure of the water. Water has a strong ability to absorb light, particularly in the near-infrared band, which significantly reduces the overall spectral reflectance. Furthermore, within the same site, differences in spectral curves corresponding to high and low AGB in the visible-light range are not significant, with only minor differences at the reflectance peak in the green-light band. However, in the near-infrared band, the spectral curves for high AGB are significantly higher than those for low AGB.

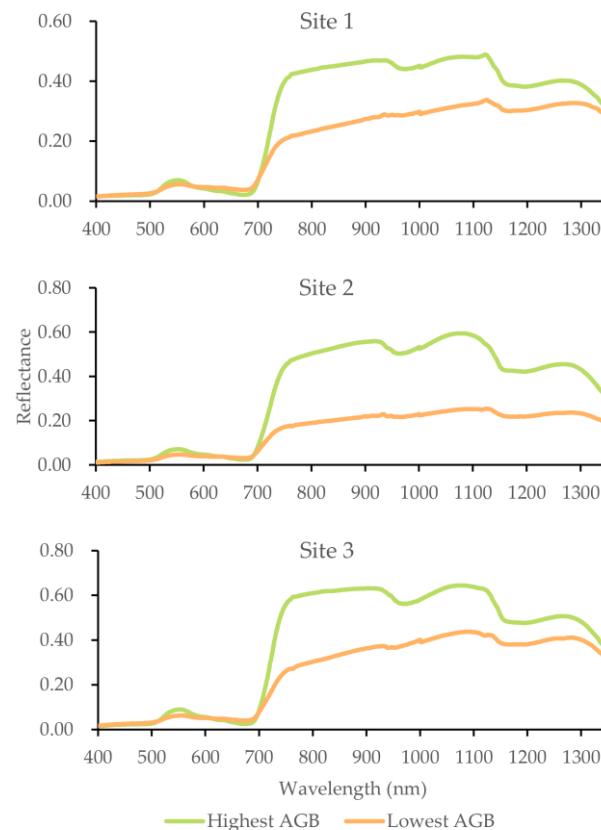


Figure 2. Aboveground biomass (AGB) spectral curves of different sample sites.

3.2. Spectral Characteristics after Fractional-Order Differentiation

In Figure 3 we present the spectral curves after fractional-order differentiation. At the 0.0 order, we observed a relatively flat region in the spectral range of 700–800 nm, which correlated with the absorption peaks of chlorophyll. As the order increased to 0.2 and 0.4, the feature peaks became more pronounced; however, the overall smoothness and noise levels of the curve remained relatively low. This suggests that smaller differentials can enhance key features in the spectral data. Under the differential transformations of 1.0 and 1.6 orders, the enhancement of feature peaks is more significant, with reflective peaks observable at specific wavelengths (for example, near 900 nm). However, from the 1.4-order differential onward, as the order of differentiation increases, noise also begins to increase and feature peaks start to diminish, possibly obscuring some weaker spectral features and affecting data interpretation. This is particularly true in the near-infrared region, where, although some sharp peaks are still visible, their utility is limited by high noise levels. This indicates that excessively high orders of differentiation may not be suitable for analyzing hyperspectral data, as they could lead to overprocessing and loss of information.

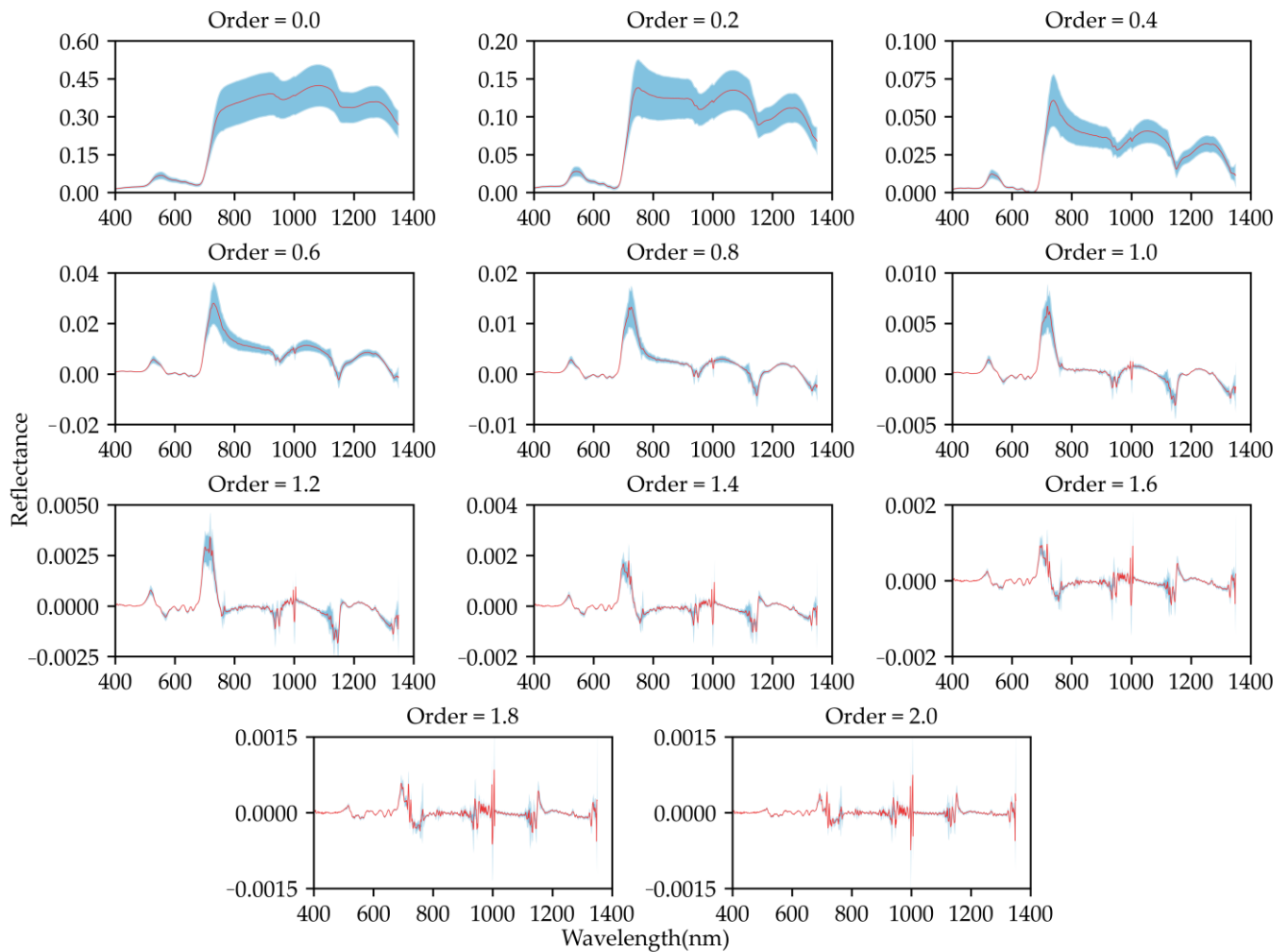


Figure 3. Results based on different FOD-preprocessed spectral curves. The blue areas represent the standard deviation interval of the spectra. The red line represents the average of the spectra curve.

3.3. Feature Extraction from Hyperspectral Data

3.3.1. Pearson Correlation Analysis between Classical Vegetation Indices and Biomass

From the 0.0 order to the 2.0 order, different orders of differentiation resulted in significant changes in the correlation with various vegetation indices (Figure 4). For example, NDVI, GNDVI, and RVI exhibited strong positive correlations at the 0.0 order of differentiation, whereas at higher orders this correlation weakened. PRI, VARI, and TCARI shifted from positive to negative correlations as the order of differentiation increased, especially at higher orders (1.8 and 2.0). Specifically, NDVI and GNDVI displayed strong positive correlations at lower orders of differentiation, indicating their effectiveness in capturing key vegetation features from the original spectral data. As the order of differentiation increased, this correlation weakened, especially at orders 1.8 and 2.0, where it became almost neutral or slightly negative. This may be due to an increase in noise from differentiation, which obscures the fundamental vegetation signals. PRI and VARI showed a slight positive or neutral correlation at lower orders of differentiation, but transformed into a distinct negative correlation at higher orders (particularly above 1.4). Indices designed to reduce soil background effects, namely SAVI and OSAVI, exhibited a shift from strong-positive to moderate-negative correlations as the order of differentiation increased. This might indicate that, as the order of differentiation increases, the spectral features of non-vegetative elements, such as soil, may be disproportionately emphasized. Subsequently, we selected vegetation indices that showed a significant correlation with AGB at a significance level of $p < 0.01$ for use in model construction, as detailed in Table 4.

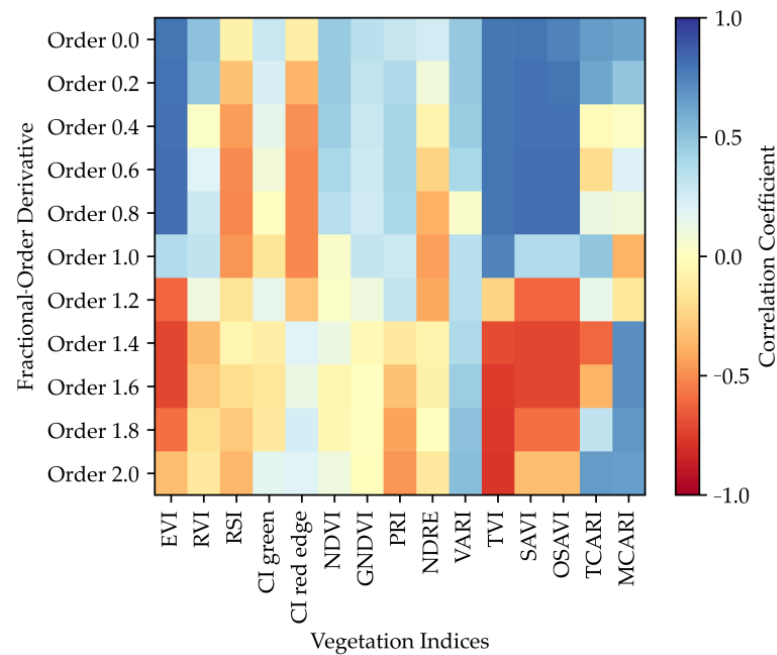


Figure 4. The correlation between vegetation indices from different fractional-order differential spectra and AGB.

Table 4. Significantly correlated vegetation indices.

Order	Feature VIs
0.0	EVI, RVI, NDVI, GNDVI, VARI, TVI, SAVI, OSAVI, TCARI, MCARI
0.2	EVI, RVI, RSI, CI red edge, NDVI, GNDVI, PRI, VARI, TVI, SAVI, OSAVI, TCARI, MCARI
0.4	EVI, RSI, CI red edge, NDVI, PRI, VARI, TVI, SAVI, OSAVI
0.6	EVI, RSI, CI red edge, NDVI, PRI, VARI, TVI, SAVI, OSAVI
0.8	EVI, RSI, CI red edge, NDVI, PRI, NDRE, TVI, SAVI, OSAVI
1.0	EVI, RVI, RSI, CI red edge, NDRE, VARI, TVI, SAVI, OSAVI, TCARI, MCARI
1.2	EVI, NDRE, VARI, SAVI, OSAVI
1.4	EVI, RVI, VARI, TVI, SAVI, OSAVI, TCARI, MCARI
1.6	EVI, PRI, VARI, TVI, SAVI, OSAVI, TCARI, MCARI
1.8	EVI, PRI, VARI, TVI, SAVI, OSAVI, TCARI, MCARI
2.0	EVI, RSI, PRI, VARI, TVI, SAVI, OSAVI, TCARI, MCARI

3.3.2. Feature Importance in RF and XGBoost

Figure S1 and Figure S2, respectively, present the feature-importance values for each band, calculated using the RF and XGBoost methods. Based on these results, we selected the top twenty bands with the highest importance values as the feature bands (Figure 5). This selection was consistently applied for each order of the fractional differentiation, from 0.0 to 2.0. For each fractional-order differentiation, the 20 most important spectral bands were selected, based on the importance scores of the model. Under lower orders of fractional-order differentiation (0.0–0.4), both XGBoost and RF models primarily selected spectral bands below 1100 nm, indicating that in the early stages of fractional-order differentiation, the key spectral features identified by the models are concentrated in lower-wavelength ranges. Starting from the 0.6 order, the XGBoost model began to recognize a broader range of important spectral bands, suggesting an increased sensitivity to a wider spectrum of features under higher orders of fractional-order differentiation. Conversely, the RF model continued to focus on the spectral bands within the 700–1200 nm range until before the 0.8 order of differentiation; from the 1.0 order of differentiation, as the order of differentiation increased, the distribution of the important bands identified by the RF model became more dispersed.

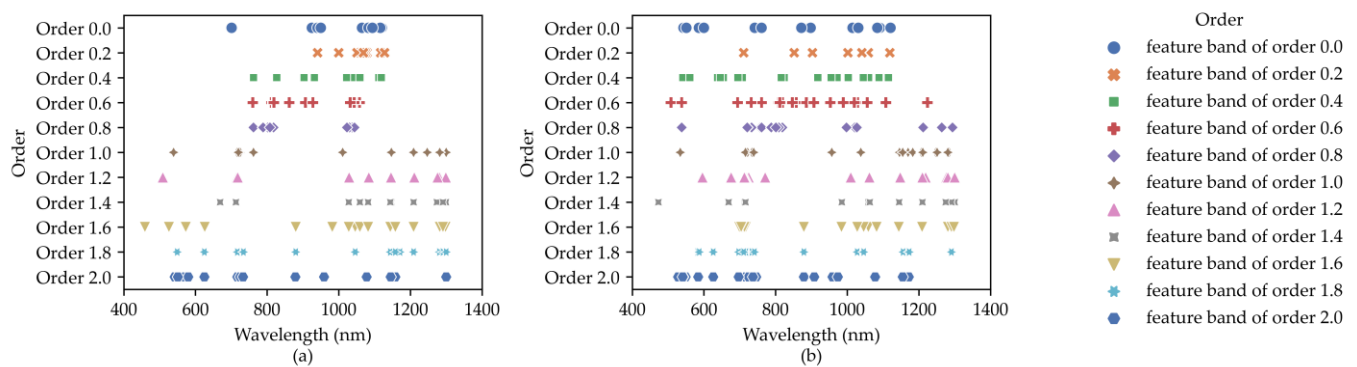


Figure 5. Features selected by feature importance from XGBoost and RF: (a) feature-importance selection results of XGBoost; (b) feature-importance selection results of RF.

3.4. Prediction Accuracy of Three Machine-Learning Models

3.4.1. Accuracy Evaluation Based on Vegetation Indices

In Table 5 we show the accuracy evaluation results of the AGB predictions using vegetation indices calculated from the spectral data under different orders of fractional-order differentiation. The results indicate that at lower orders of fractional-order differentiation, the models generally exhibit relatively high accuracy, with the R^2 values for the three models (XGBoost, RF, and CatBoost) ranging from 0.488 to 0.603. The RPD values of the RF model were all above 1.4, demonstrating a strong predictive potential. These results indicate that the models exhibit good adaptability and prediction accuracy when processing differentiated lower-order data. However, from orders 1.0 to 1.6, the model performance began to fluctuate, with a decline in the accuracy of models built using vegetation indices. Particularly at order 1.2, the R^2 values of the three models sharply dropped to approximately 0.2, and the RPD values were significantly lower than those at other orders. At order 1.8, the model performance recovered slightly, particularly at order 2.0, where the R^2 value for XGBoost reached 0.572 and the IOA value for XGBoost reached 0.849. However, the stability and prediction accuracy of the models under higher orders of differentiation are somewhat limited.

Table 5. Model-Accuracy Evaluation Based on Vegetation Indices.

Order	Model	R^2	IOA	RMSE (g/m ²)	RPD
0.0	XGBoost	0.567	0.858	27.613	1.458
	RF	0.557	0.803	27.604	1.459
	CatBoost	0.568	0.847	28.298	1.423
0.2	XGBoost	0.577	0.841	28.883	1.394
	RF	0.593	0.848	27.479	1.465
	CatBoost	0.585	0.867	27.260	1.477
0.4	XGBoost	0.533	0.821	30.817	1.307
	RF	0.577	0.854	27.078	1.487
	CatBoost	0.520	0.832	29.992	1.343
0.6	XGBoost	0.504	0.822	30.592	1.316
	RF	0.540	0.842	28.371	1.419
	CatBoost	0.488	0.817	29.604	1.360
0.8	XGBoost	0.562	0.854	27.707	1.453
	RF	0.603	0.859	25.472	1.581
	CatBoost	0.576	0.859	26.590	1.514
1.0	XGBoost	0.505	0.834	28.837	1.396
	RF	0.579	0.825	27.376	1.471
	CatBoost	0.564	0.830	27.380	1.471

Table 5. Cont.

Order	Model	R ²	IOA	RMSE (g/m ²)	RPD
1.2	XGBoost	0.197	0.584	35.726	1.127
	RF	0.234	0.599	35.067	1.148
	CatBoost	0.196	0.534	35.799	1.125
1.4	XGBoost	0.472	0.772	29.454	1.367
	RF	0.418	0.736	31.302	1.286
	CatBoost	0.435	0.765	30.591	1.316
1.6	XGBoost	0.328	0.709	35.116	1.147
	RF	0.408	0.737	32.102	1.254
	CatBoost	0.357	0.708	33.632	1.197
1.8	XGBoost	0.439	0.772	31.337	1.285
	RF	0.507	0.794	29.605	1.360
	CatBoost	0.439	0.772	31.805	1.266
2.0	XGBoost	0.572	0.840	28.064	1.435
	RF	0.527	0.813	28.262	1.425
	CatBoost	0.488	0.805	29.529	1.364

Note: The optimal accuracy is shown in bold.

Considering the performance of the individual models, XGBoost showed consistently similar performance across multiple orders, particularly at lower fractional orders of differentiation. In contrast, RF outperformed XGBoost and CatBoost in most orders of differentiation, especially at order 0.8, where its R², IOA, and RPD values reached their highest values of 0.603, 0.859 and 1.581, respectively, and had the lowest RMSE value (25.472 g/m²), indicating higher predictive accuracy. CatBoost performed similarly to XGBoost at certain orders (0.0 and 1.8), the accuracy of the two models being very similar; however, the accuracy of CatBoost fell short at higher orders of differentiation.

3.4.2. Model-Accuracy Evaluation Based on Feature Bands

We combined the feature bands extracted by the two methods under each order of fractional-order differentiation for use as inputs for the models. In Figure 6, we compare the performances of the three machine-learning models (XGBoost, RF, and CatBoost) across different fractional orders of differentiation (from 0.0 to 2.0) using four metrics (R², IOA, RMSE, and RPD). Overall, the R² values for all three models generally exceeded 0.5; across the various orders of differentiation, the IOA values were mostly above 0.8, the RMSE values mostly ranged between 25 g/m² and 30 g/m², and the RPD values were mostly above 1.4, peaking at 1.649. The accuracy of the models showed a trend of increasing and then decreasing with increasing orders of differentiation, specifically indicated by the initial increase and subsequent decrease in R², IOA, and RPD (Figure 6a,b,d), whereas RMSE initially decreased and then increased (Figure 6c). Additionally, there was a sharp decline in accuracy at order 2.0 for all three models, which may suggest that excessively high orders of differentiation introduce significant noise into the spectral data.

Looking at the performance of individual models, the RF model demonstrated the best performance at order 0.6 (R² = 0.650, IOA = 0.889, RMSE = 24.414 g/m², and RPD = 1.649), whereas XGBoost and CatBoost achieved their highest R² at order 1.0, with values of 0.604 and 0.633, respectively. At an order of 0.6, these two models exhibited optimal IOA, RMSE and RPD values (XGBoost IOA = 0.868, RMSE = 26.736 g/m², RPD = 1.506; CatBoost IOA = 0.886, RMSE = 24.942 g/m², RPD = 1.614). Combining all the metrics, the performance ranking of the three models was as follows: RF > CatBoost > XGBoost.

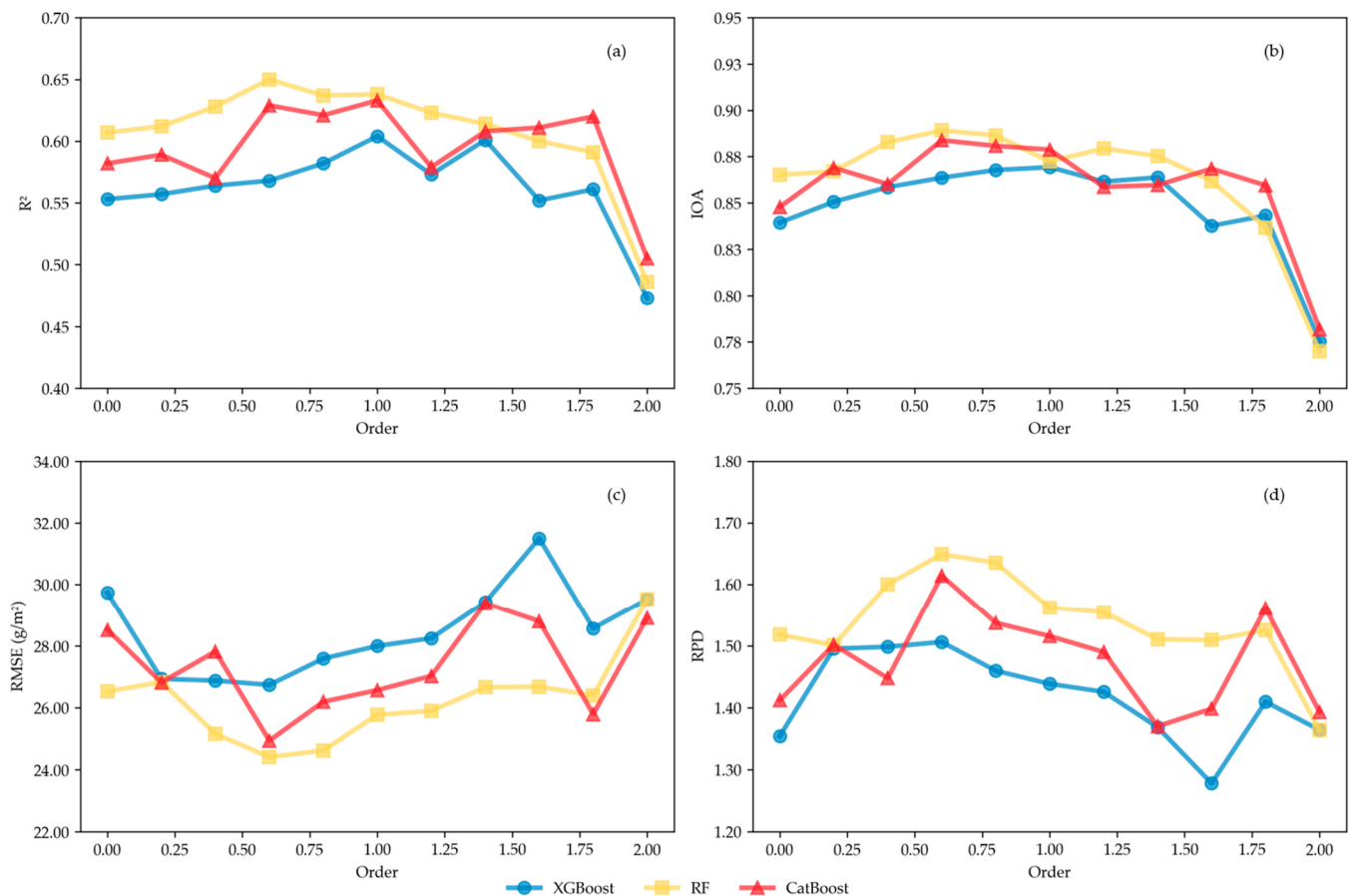


Figure 6. Model-accuracy evaluation based on feature bands. (a–d) represent R^2 , IOA, RMSE, and RPD, respectively.

3.4.3. Model-Accuracy Evaluation Integrating Feature Bands and Vegetation Indices

AGB prediction models were constructed by integrating the selected feature bands and vegetation indices (Figure 7). Compared with traditional strategies that use only vegetation indices or feature bands, the integrated approach demonstrated significant improvements in accuracy across most orders of differentiation, with R^2 generally exceeding 0.5, and most RPD values surpassing 1.4. This suggests that the combination of vegetation indices and feature bands effectively complemented each other, providing a richer set of features and thereby enhancing the predictive capability of the models.

However, we also observed a decline in model accuracy in certain instances, especially at the differentiation order of 2.0. Specifically, the accuracy of the three regression models (XGBoost, RF, and CatBoost) constructed using only vegetation indices was the highest among all strategies, with R^2 values of 0.572, 0.527, and 0.488, respectively. The models constructed using feature bands had lower R^2 values of 0.473, 0.486, and 0.505, respectively. After integrating these two types of data, the R^2 values of the models decreased to 0.514, 0.483, and 0.486, respectively, possibly because of increased spectral noise during the integration process.

As shown in Figure 7a–d, compared with the other two models, the RF model showed the lowest RMSE and the highest R^2 , IOA, and RPD values, indicating its superior predictive performance. Figure 7e–g show scatter plots of the optimal accuracy achieved by three models under transformations of fractional-order derivatives. At a differentiation order of 0.8, the XGBoost model demonstrated the best predictive accuracy, whereas the CatBoost model achieved optimal precision at the 0.6 differentiation order. In particular, at a differentiation order of 0.8, the RF model achieved the best accuracy among all the models, with an R^2 of 0.673, IOA of 0.898, RMSE of 23.196 g/m^2 , and RPD of 1.736.

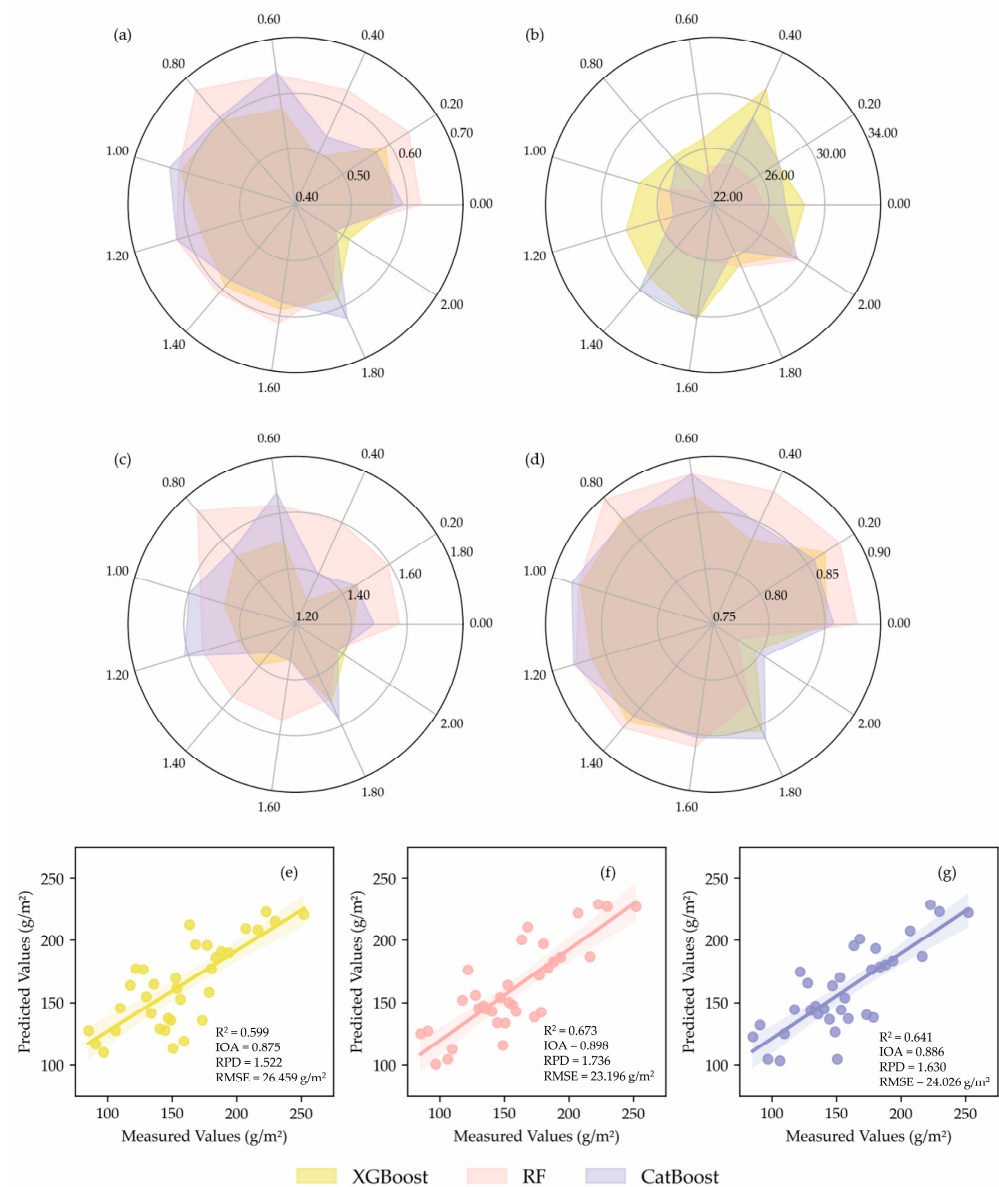


Figure 7. Model-accuracy evaluation based on feature bands and Vegetation Indices: (a–d) the R^2 , IOA, RMSE, and RPD of the three models, respectively; the inner scale represents values of four indices, and the outer scale represents order; (e) the relationship between the predicted AGB and the observed AGB using the optimal-accuracy XGBoost model (order = 0.8) based on the validation data; (f) the relationship between the predicted AGB and the observed AGB using the optimal-accuracy RF model (order = 0.8) based on the validation data; (g) the relationship between the predicted AGB and the observed AGB using the optimal-accuracy CatBoost model (order = 0.6) based on the validation data.

Notably, our results also revealed that R^2 and RMSE did not always display the same trends. In some cases, although a higher R^2 value indicates a better model fitting to the data, the corresponding RMSE may also be high, potentially owing to outliers in the data.

3.5. Interpretability of the Optimal Regression Model Using SHAP

We performed a SHAP analysis on the model with the highest accuracy (an RF model integrating vegetation indices and feature bands of the order of 0.6), as shown in Figure 8, in which the input variables with the greatest impact on the predicted AGB values are listed. The points in the figure represent the sample points, with each row containing 34 validation sample points. The depth of the color indicates the magnitude of the feature

value (vegetation index or feature-band reflectance), with deeper red indicating higher feature values and deeper blue indicating lower values. The x-axis represents the magnitude of the SHAP values, with the left side indicating an increase in the predicted value and the right a decrease. The y-axis represents the ranking of the variables based on their SHAP values.

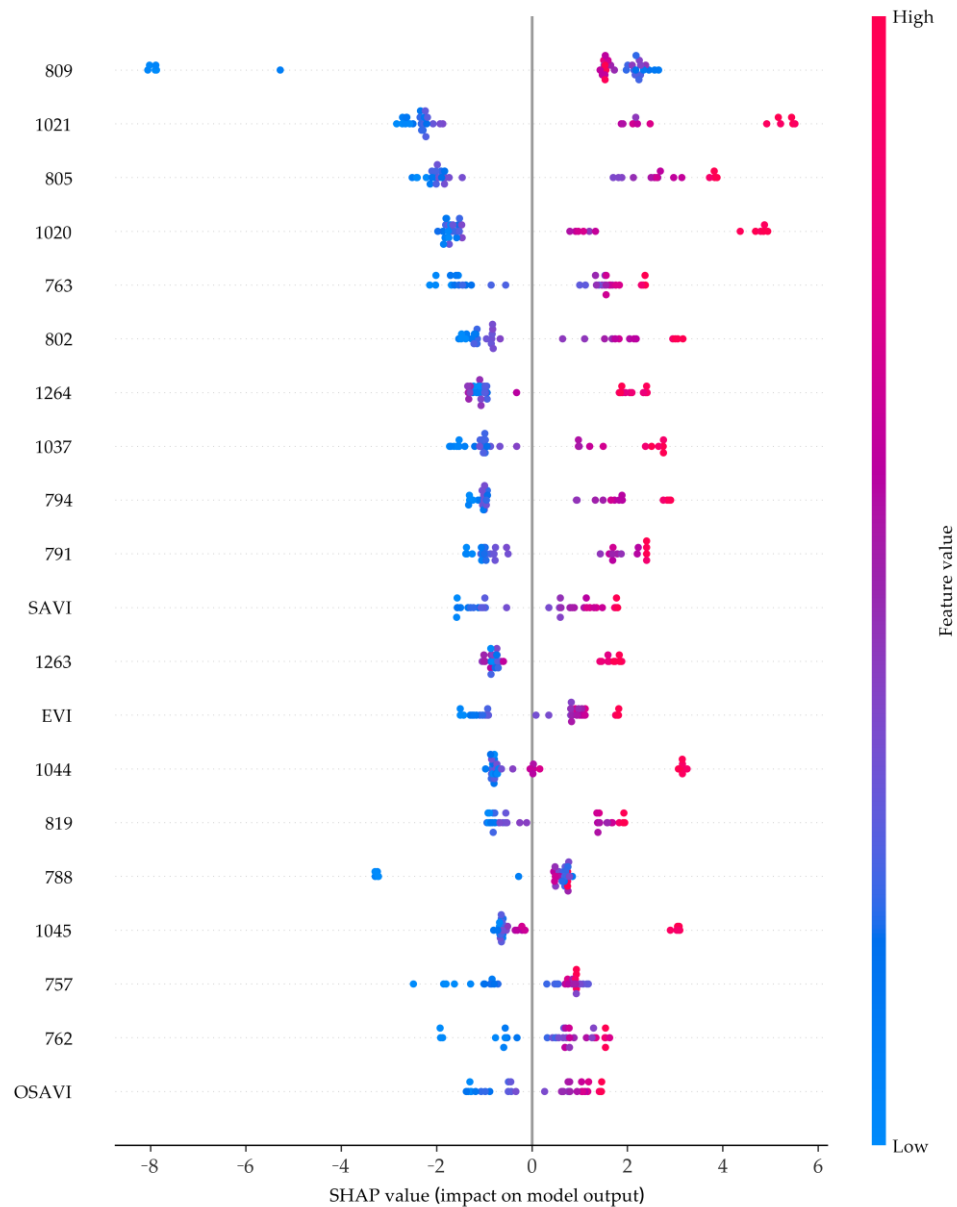


Figure 8. Visualization of feature contributions in the RF model achieving optimal accuracy under the strategy of integrating feature bands and vegetation indices.

As shown in the figure, the reflectance at 809 nm contributes most to the prediction, with the reflectance values of most samples increasing the size of the predicted value. Most other features also show that larger feature values increase the size of the predicted value; that is, the smaller the point, the larger the SHAP value, whereas smaller feature values decrease the size of the predicted value. Among the vegetation indices, SAVI was the most important, ranking 11th.

In Figure 9, the impact of each feature on AGB in the 10th and 24th samples is visualized through SHAP. Red indicates that the feature value increased the predicted value of AGB, and blue indicates that the feature value decreased the predicted value of AGB. The width of each feature represents the magnitude of the SHAP value. The two $f(X)$

values in Figure 8 (171.83 and 144.84, respectively) indicate that the predicted AGB values for these two samples are 171.83 g/cm^2 and 144.84 g/cm^2 , respectively, with the base value representing the average AGB of all data. Specifically, at 809 nm, the predicted value was increased in both samples, whereas at 1021 and 1020 nm, the predicted value of AGB decreased. Under the combined influence of all the features, the predicted AGB value for the 10th sample was much higher than the average, whereas the predicted AGB value for the 24th sample was much lower than the average.

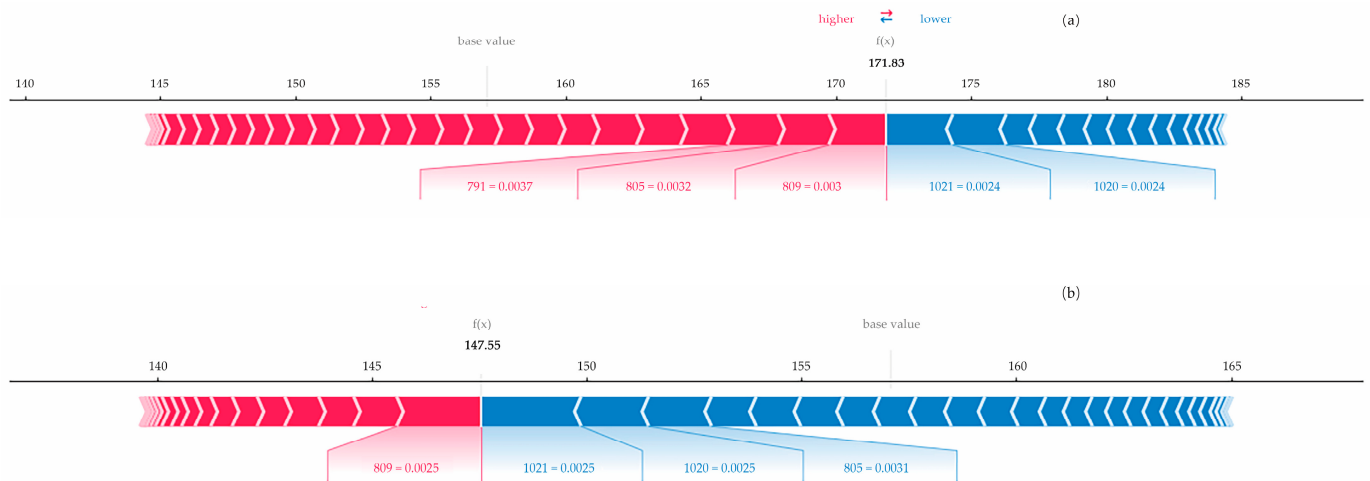


Figure 9. The feature effects on the estimated AGB values for the 10th and 24th samples: (a) represents the 10th sample, and (b) represents the 24th sample.

4. Discussion

Traditional AGB estimation methods primarily rely on classic vegetation indices to construct estimation models [69,70]. These methods either suffer from the limitations of spectral resolution and sensor performance, preventing the capture of fine vegetation features, or depend on a limited combination of bands, failing to fully harness hyperspectral information. This can result in significant limitations when dealing with complex mixed-species plant communities. Fully extracting spectral-feature information to improve model estimation accuracy is a pressing issue. In our study, we integrate fractional-order differentiation and spectral-feature extraction to explore how machine learning models can be used to predict AGB. This approach enhances the detection of key features related to AGB estimation while minimizing the risk of overfitting, overcoming some of the limitations of traditional methods, particularly in terms of resolution. However, whether the model can be directly applied to other wetland environments requires further training and optimization with samples from different habitats, which will be the focus of future research.

In the field of hyperspectral remote sensing, relying solely on raw spectra sometimes fails to satisfy the requirements for spectral-feature selection. Among the various methods for enhancing spectral features are differential transformation, logarithmic transformation, and standardized normal variate (SNV) [71,72]. However, in previous studies, the focus has been on the application of integer-order derivatives such as first and second derivatives [73,74]. The utility of this method has already been demonstrated in studies estimating chlorophyll [23,50]; however, its suitability for estimating AGB in mixed-species wetland-plant communities requires further exploration. Here, we analyzed the impact of fractional-order differentiation from 0.0 to 2.0 on the spectral features. The results showed that as the order of differentiation increased, some feature peaks became more pronounced, which may correspond to key biochemical parameters. However, when the order of differentiation reached 1.4, these feature peaks began to diminish, potentially obscuring important spectral information. Thus, the results demonstrate that appropriate fractional-order differentiation can effectively improve spectral features and reveal new characteristics. However, excessively high orders (such as above 1.4) may be detrimental to

highlighting features. Future studies should further evaluate this phenomenon through extensive experimental analyses.

Feature selection is a crucial step in the inversion of physiological and biochemical parameters of vegetation based on hyperspectral data. The results of a number of studies have confirmed that vegetation indices are widely considered effective means for estimating plant physiological and biochemical aspects [54,75,76]. In the present study, we analyzed the correlation of 15 commonly used vegetation indices and their fractional-order-differentiated transformations with total vegetation biomass (AGB). The results showed that the correlations between traditional vegetation indices and AGB—after being recalculated through fractional-order differentiation—were variously enhanced or weakened, and in some cases, shifted from positive to negative correlations, as was the case with TVI and OSAVI. This indicates that introducing new spectral features through fractional-order differentiation and recombination bands can optimize vegetation indices to a certain extent, revealing important biomass-related information that traditional indices cannot capture. However, the limitation of traditional vegetation indices lies in their dependency on fixed bands, which leads to the neglect of other key spectral features. In recent years, many scholars have explored new band combinations to determine the optimal vegetation indices for predicting the physiological and biochemical parameters [22,77]. By combining newly created vegetation indices with fractional-order differentiation, more significant spectral features are expected to be identified, thereby providing more effective variables for the accurate estimation of total vegetation biomass (AGB). This approach offers new research directions and potential technological breakthroughs in the field of vegetation remote-sensing monitoring.

In the context of high-dimensional features in hyperspectral data, the selection of important feature bands to alleviate the constraints of dimensionality was shown to be effective. Pearson correlation analysis is one of the most commonly used methods [78,79]; however, in recent years, some model-intrinsic feature-evaluation methods have been proven to effectively select feature bands, such as the VIP values in PLSR and feature importance in models such as XGBoost [22,80]. In the present study, we selected the top-20 most important bands through the XGBoost and RF importance ranking methods and combined them into feature bands for modeling. The selection of feature bands indicates that fractional-order differentiation may reveal new and relevant bands that contribute to model predictions. The selection results of the spectral bands under different orders of fractional-order differentiation indicated that lower orders of differentiation tended to select important features within a narrower wavelength range. As the order increased, both models exhibited a trend of incorporating a broader range of spectral features, indicating the potential for higher orders of differentiation to reveal new and informative spectral characteristics. The increasingly dispersed distribution of feature importance with higher orders of differentiation highlights the increasing complexity of spectral data interpretation, making the selection of key differentiation orders particularly important for evaluating feature importance.

We chose the XGBoost, RF, and CatBoost models to invert the AGB in wetland plant communities. Compared to previous studies on the inversion of wetland plant biomass (Table 6), our AGB range is from 81.35 g/m² to 280.02 g/m², which is lower than those reported in other studies. This can be attributed to differences in species composition. In our study area, the dominant vegetation consists of herbaceous plants such as *Echinochloa crus-galli*, *Equisetum hyemale*, and *Polygonum hydropiper*. Compared to the woody plants or dense monocultures like *Phragmites australis* found in other studies, the AGB values are indeed lower. This indicates that the model would need to be recalibrated when applied to targets with higher AGB. Although the accuracy achieved was not as high as that in the studies by Dou and Li [42,43], possibly due to the complex community composition and background-water reflections affecting the spectra, which weakened the correlation between the obtained spectra and AGB. Removing the influence of the water background will be an important direction for future research. Nonetheless, our

study still achieved considerable accuracy through more feature transformations and selection methods, as well as the exploration of new models, validating the feasibility of these methods. Our results also confirm the high accuracy and stability of the RF model, which has also been validated by the quantitative inversion of physiological and biochemical indicators in wetland plants [10,81,82]. The robustness of the RF model for feature selection and missing-data handling provides a solid foundation for complex ecological data analysis [83,84]. Although XGBoost can optimize computational speed and model performance, its performance did not meet expectations in our study, with R^2 only approximately 0.5, most RPDs not exceeding 1.4, and RMSE being relatively high. These results may stem from the complexity of the hyperparameters in the XGBoost algorithm [58], in which inappropriate configuration of these parameters could lead to model overfitting or insufficient learning of key data features. Not optimizing these hyperparameters may be the main reason why the performance of the model did not meet expectations. In future research, meticulous parameter tuning and cross validation could improve the predictive accuracy and stability of the XGBoost model.

Table 6. Comparison of results of the present study with those from previous research.

Dataset	Number of Total Samples	Dominant Wetland Vegetation	AGB Range (g/cm ²)	Regression Model	Accuracy Performance	Reference
ASD hyperspectral	180	<i>Phragmites australis</i>	5.50–9.00 kg/m ²	PLS	$R^2 = 0.87$	[42]
ASD hyperspectral	90	<i>Suaeda salsa</i>	0.12–0.81 kg/m ²	PLS	$R^2 = 0.95$	[43]
UAV hyperspectral and LiDAR	75	<i>Phragmites australis</i>	330.13–1351.78 g/m ²	$AGB = 207.098 \times \ln(H_{p99}) + 736.278 \times MSAVI + 765.635$	$R^2 = 0.648$	[45]
Multi-Temporal Images	182	/	/	RF	$R^2 = 0.65$	[41]
ASD hyperspectral	102	<i>Echinochloa crus-galli</i> , <i>Equisetum hyemale</i> , <i>Polygonum hydropiper</i>	81.35–280.02 g/m ²	XGBoost	$R^2 = 0.614$	Our Study
				RF	$R^2 = 0.673$	
				CatBoost	$R^2 = 0.635$	

When the model accuracy is deemed acceptable, further validation is required to determine whether the model is applicable to other wetland ecosystems using sample data from those wetlands. Significant differences in plant composition, soil conditions, and hydrological characteristics among different wetland types may imply that high-biomass samples are inadequately represented in our current dataset, thus impacting the model's transferability. Consequently, calibration may be necessary when applying the model to different wetlands. For instance, different wetlands may have distinct plant-species compositions, soil conditions, and hydrological features, all of which can affect spectral data [85]. Future research should focus on collecting and analyzing sample data from various types of wetlands to assess and validate the transferability of our model. By incorporating additional sample data, we aim to optimize the model to maintain high accuracy and stability across a wide range of ecosystems. Specific calibration methods for different wetland types also require further research and development to enhance the applicability and generalizability of the model.

Furthermore, with this study, we explored the impact of different feature-combination strategies on model accuracy. The results show that not all combinations of vegetation indices and specific bands improve the model accuracy. For instance, models built with data transformed at a 2.0 order of differentiation did not perform better than those using only vegetation indices, emphasizing the importance of selecting the best feature-input strategy to enhance model accuracy. Notably, we employed the Bayesian hyperparameter optimization method, which balances accuracy and efficiency. However, to ensure

that the best model parameter combinations are found in complex parameter spaces, the introduction of a grid-search algorithm may be a fruitful approach in future studies. A grid search systematically traverses all possible parameter combinations, and, although computationally costly, it provides more comprehensive search results, ensuring that the global optimum is determined [86].

Machine learning is often referred to as a black-box model, given the opaqueness of internal features, leading to problems in interpretability. In the present study, we constructed an interpretable model based on SHAP, which, unlike feature importance, not only evaluates the features input into the model but also assesses the effect of each feature on each group of samples, making the role of each feature in each group of samples transparent. This helps in understanding the logic of the model, thereby making the application of machine learning models more effective and enhancing their transparency and reliability [29,68].

Overall, hyperspectral remote-sensing technology demonstrates significant potential for predicting AGB, and integrating multi-source data to enhance the accuracy of predictive models is the current research focus. Future studies should concentrate on integrating ground-based hyperspectral, satellite, or airborne hyperspectral data, along with radar data, temperature, precipitation, and other environmental factors, starting from a multi-source data perspective, to enhance the accuracy of AGB prediction. This approach provides a comprehensive and accurate analytical framework for estimating vegetation biomass, marking an important direction for future research.

5. Conclusions

We extracted vegetation indices and feature bands through the fractional-order differentiation of hyperspectral data and constructed AGB regression models using XGBoost, RF, and CatBoost. The SHAP values were used to interpret the optimal model. We found that fractional-order differentiation can improve the detail and feature-enhancement capabilities of spectra to some extent, revealing features that cannot be expressed by raw spectral data, and enhancing the capture ability of vegetation indices and feature bands. However, higher fractional orders are not always better, and an optimal differentiation order exists. In our analysis, the range in differentiation order between 0.6 and 1.2 was found to be an optimal choice. Second, by comparing the estimation accuracy of the three strategies (vegetation indices, feature bands, and a combination of vegetation indices and feature bands) with the three regression models (XGBoost, RF, and CatBoost), we found that combining vegetation indices and feature bands improved inversion accuracy. The RF demonstrated superior accuracy and stability, with the best performance at the 0.8 order of differentiation, making it the optimal estimation model. Lastly, SHAP value analysis allowed us to visualize the top-20 features that most significantly affected AGB estimation, thus improving the interpretability and transparency of the model. Additionally, for each sample, the contributions of different features vary.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs16163011/s1>, Figure S1: Feature importance calculation results for each band using RF under fractional differentiation from 0.0 to 2.0; Figure S2: Feature importance calculation results for each band using XGBoost under fractional differentiation from 0.0 to 2.0.

Author Contributions: Conceptualization, H.L. and W.L.; methodology, H.L., X.T. and J.W. (Junjie Wang); data curation, H.L. and X.T.; funding acquisition, W.L.; project administration, Y.L. and J.W. (Jinzhi Wang); resources, W.L.; software, X.Z. (Xiajie Zhai) and X.Z. (Xinsheng Zhao); investigation, H.L. and X.T.; visualization, H.L. and W.L.; writing—original draft, H.L.; writing—review and editing, L.C., Y.L., J.W. (Jinzhi Wang), J.L., R.W. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by China's Special Fund for Basic Scientific Research Business of Central Public Research Institutes (CAFYBB2021MC006 and CAFYBB2021ZB003).

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

Acknowledgments: We would like to thank Jianwen Zeng and Jiaqi Yan of the research reserve for their assistance and support in facilitating our experiments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fu, C.; Xu, M. Achieving Carbon Neutrality Through Ecological Carbon Sinks: A Systems Perspective. *Green Carbon* **2023**, *1*, 43–46. [\[CrossRef\]](#)
2. Yu, G.; Zhu, J.; Xu, L.; He, N. Technological Approaches to Enhance Ecosystem Carbon Sink in China: Nature-Based Solutions. *Bull. Chin. Acad. Sci. (Chin. Version)* **2022**, *37*, 490–501.
3. Sharma, B.; Rasul, G.; Chettri, N. The Economic Value of Wetland Ecosystem Services: Evidence from the Koshi Tappu Wildlife Reserve, Nepal. *Ecosyst. Serv.* **2015**, *12*, 84–93. [\[CrossRef\]](#)
4. Na, X.; Zang, S.; Zhan, N.; Cui, J. Impact of Land Use and Land Cover Dynamics on Zhalong Wetland Reserve Ecosystem, Heilongjiang Province, China. *Int. J. Environ. Sci. Technol.* **2015**, *12*, 445–454. [\[CrossRef\]](#)
5. Bhatnagar, S.; Gill, L.; Regan, S.; Naughton, O.; Johnaton, P.; Waldren, S.; Ghosh, B. Mapping Vegetation Communities inside Wetlands Using Sentinel-2 Imagery in Ireland. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *88*, 102083. [\[CrossRef\]](#)
6. Rogers, K.; Kelleway, J.J.; Saintilan, N.; Megonigal, J.P.; Adams, J.B.; Holmquist, J.R.; Liu, M.; Schile-Beers, L.; Zawadzki, A.; Mazumder, D.; et al. Wetland Carbon Storage Controlled by Millennial-Scale Variation in Relative Sea-Level Rise. *Nature* **2019**, *567*, 91–95. [\[CrossRef\]](#)
7. Mendonça, R.; Müller, R.A.; Clow, D.; Verpoorter, C.; Raymond, P.; Tranvik, L.J.; Sobek, S. Organic Carbon Burial in Global Lakes and Reservoirs. *Nat. Commun.* **2017**, *8*, 1694. [\[CrossRef\]](#)
8. Naidoo, L.; Van Deventer, H.; Ramoelo, A.; Mathieu, R.; Nondlazi, B.; Gangat, R. Estimating above Ground Biomass as an Indicator of Carbon Storage in Vegetated Wetlands of the Grassland Biome of South Africa. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *78*, 118–129. [\[CrossRef\]](#)
9. Lu, X.; Jiang, M. Progress and Prospect of Wetland Research in China. *J. Geogr. Sci.* **2004**, *51*, 45–51.
10. Mutanga, O.; Adam, E.; Cho, M.A. High Density Biomass Estimation for Wetland Vegetation Using Worldview-2 Imagery and Random Forest Regression Algorithm. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 399–406. [\[CrossRef\]](#)
11. Shen, G.; Liao, J.; Guo, H.; Liu, J. Poyang Lake Wetland Vegetation Biomass Inversion Using Polarimetric RADARSAT-2 Synthetic Aperture Radar Data. *J. Appl. Remote Sens.* **2015**, *9*, 096077. [\[CrossRef\]](#)
12. Li, C.; Zhou, L.; Xu, W. Estimating Aboveground Biomass Using Sentinel-2 MSI Data and Ensemble Algorithms for Grassland in the Shengjin Lake Wetland, China. *Remote Sens.* **2021**, *13*, 1595. [\[CrossRef\]](#)
13. Pacini, N.; Hesslerová, P.; Pokorný, J.; Mwinami, T.; Morrison, E.H.J.; Cook, A.A.; Zhang, S.; Harper, D.M. Papyrus as an Ecohydrological Tool for Restoring Ecosystem Services in Afro-Tropical Wetlands. *Ecohydrol. Hydrobiol.* **2018**, *18*, 142–154. [\[CrossRef\]](#)
14. Wang, Y.; Yésou, H. Remote Sensing of Floodpath Lakes and Wetlands: A Challenging Frontier In the Monitoring of Changing Environments. *Remote Sens.* **2018**, *10*, 1955. [\[CrossRef\]](#)
15. Hestir, E.L.; Brando, V.E.; Bresciani, M.; Giardino, C.; Matta, E.; Villa, P.; Dekker, A.G. Measuring Freshwater Aquatic Ecosystems: The Need for a Hyperspectral Global Mapping Satellite Mission. *Remote Sens. Environ.* **2015**, *167*, 181–195. [\[CrossRef\]](#)
16. Zhao, Y.; Mao, D.; Zhang, D.; Wang, Z.; Du, B.; Yan, H.; Qiu, Z.; Feng, K.; Wang, J.; Jia, M. Mapping Phragmites australis Aboveground Biomass in the Momoge Wetland Ramsar Site Based on Sentinel-1/2 Images. *Remote Sens.* **2022**, *14*, 694. [\[CrossRef\]](#)
17. Zhou, R.; Yang, C.; Li, E.; Cai, X.; Wang, X. Aboveground Biomass Estimation of Wetland Vegetation at the Species Level Using Unoccupied Aerial Vehicle RGB Imagery. *Front. Plant Sci.* **2023**, *14*, 1181887. [\[CrossRef\]](#)
18. Doughty, C.L.; Ambrose, R.F.; Okin, G.S.; Cavanaugh, K.C. Characterizing Spatial Variability in Coastal Wetland Biomass across Multiple Scales Using UAV and Satellite Imagery. *Remote Sens. Ecol. Conserv.* **2021**, *7*, 411–429. [\[CrossRef\]](#)
19. Khaliq, A.; Comba, L.; Biglia, A.; Aimonino, D.R.; Chiaberge, M.; Gay, P. Comparison of Satellite and UAV-Based Multispectral Imagery for Vineyard Variability Assessment. *Remote Sens.* **2019**, *11*, 436. [\[CrossRef\]](#)
20. Transon, J.; d’Andrimont, R.; Maignard, A.; Defourny, P. Survey of Hyperspectral Earth Observation Applications from Space in the Sentinel-2 Context. *Remote Sens.* **2018**, *10*, 157. [\[CrossRef\]](#)
21. Baresel, J.P.; Rischbeck, P.; Hu, Y.C.; Kipp, S.; Hu, Y.C.; Barmer, G.; Mistele, B.; Schmidhalter, U. Use of a Digital Camera as Alternative Method for Non-Destructive Detection of the Leaf Chlorophyll Content and the Nitrogen Nutrition Status in Wheat. *Comput. Electron. Agric.* **2017**, *140*, 25–33. [\[CrossRef\]](#)
22. Tong, X.; Duan, L.; Liu, T.; Yang, Z.; Wang, Y.; Singh, V.P. Estimation of Grassland Aboveground Biomass Combining Optimal Derivative and Raw Reflectance Vegetation Indices at Peak Productive Growth Stage. *Geocarto Int.* **2023**, *38*, 2186497. [\[CrossRef\]](#)
23. Bhadra, S.; Sagan, V.; Maimaitijiang, M.; Maimaitiyiming, M.; Newcomb, M.; Shakoob, N.; Mockler, T.C. Quantifying Leaf Chlorophyll Concentration of Sorghum from Hyperspectral Data using Derivative Calculus and Machine Learning. *Remote Sens.* **2020**, *12*, 2082. [\[CrossRef\]](#)

24. Li, H.Z.; Cui, L.J.; Dou, Z.G.; Wang, J.J.; Zhai, X.J.; Li, J.; Zhao, X.S.; Lei, Y.R.; Wang, J.Z.; Li, W. Hyperspectral Analysis and Regression Modeling of SPAD Measurements in Leaves of Three Mangrove Species. *Forests* **2023**, *14*, 1566. [\[CrossRef\]](#)
25. Li, W.; Dou, Z.G.; Cui, L.J.; Wang, R.M.; Zhao, Z.J.; Cui, S.F.; Lei, Y.R.; Li, J.; Zhao, X.S.; Zhai, X.J. Suitability of Hyperspectral Data for Monitoring Nitrogen and Phosphorus Contents in Constructed Wetlands. *Remote Sens. Lett.* **2020**, *11*, 495–504. [\[CrossRef\]](#)
26. Wang, J.; Xu, Y.; Wu, G. The Integration of Species Information and Soil Properties for Hyperspectral Estimation of Leaf Biochemical Parameters in Mangrove Forest. *Ecol. Indic.* **2020**, *115*, 106467. [\[CrossRef\]](#)
27. Nie, L.; Dou, Z.; Cui, L.; Tang, X.; Zhai, X.; Zhao, X.; Lei, Y.; Li, J.; Wang, Z.; Li, W. Hyperspectral Inversion of Soil Carbon and Nutrient Contents in the Yellow River Delta Wetland. *Diversity* **2022**, *14*, 862. [\[CrossRef\]](#)
28. Jacon, A.D.; Galvão, L.S.; Dalagnol, R.; dos Santos, J.R. Aboveground Biomass Estimates over Brazilian Savannas Using Hyperspectral Metrics and Machine Learning Models: Experiences with Hyperion/EO-1. *GIScience Remote Sens.* **2021**, *58*, 1112–1129. [\[CrossRef\]](#)
29. Huang, W.; Li, W.; Xu, J.; Ma, X.; Li, C.; Liu, C. Hyperspectral Monitoring Driven by Machine Learning Methods for Grassland Above-Ground Biomass. *Remote Sens.* **2022**, *14*, 2086. [\[CrossRef\]](#)
30. Assiri, M.; Sartori, A.; Persichetti, A.; Miele, C.; Faelga, R.A.; Blount, T.; Silvestri, S. Leaf Area Index and Aboveground Biomass Estimation of an Alpine Peatland with a UAV Multi-Sensor Approach. *GIScience Remote Sens.* **2023**, *60*, 2270791. [\[CrossRef\]](#)
31. Brocks, S.; Bareth, G. Estimating Barley Biomass with Crop Surface Models from Oblique RGB Imagery. *Remote Sens.* **2018**, *10*, 268. [\[CrossRef\]](#)
32. Liu, Y.; Feng, H.; Yue, J.; Fan, Y.; Jin, X.; Zhao, Y.; Song, X.; Long, H.; Yang, G. Estimation of Potato Above-Ground Biomass Using UAV-Based Hyperspectral Images and Machine-Learning Regression. *Remote Sens.* **2022**, *14*, 5449. [\[CrossRef\]](#)
33. Liu, Y.; Feng, H.; Yue, J.; Jin, X.; Fan, Y.; Chen, R.; Bian, M.; Ma, Y.; Song, X.; Yang, G. Improved Potato AGB Estimates Based on UAV RGB and Hyperspectral Images. *Comput. Electron. Agric.* **2023**, *214*, 108260. [\[CrossRef\]](#)
34. Yang, H.; Li, F.; Wang, W.; Yu, K. Estimating Above-Ground Biomass of Potato Using Random Forest and Optimized Hyperspectral Indices. *Remote Sens.* **2021**, *13*, 2339. [\[CrossRef\]](#)
35. Yue, J.; Yang, G.; Tian, Q.; Feng, H.; Xu, K.; Zhou, C. Estimate of Winter-Wheat Above-Ground Biomass Based on UAV Ultrahigh-Ground-Resolution Image Textures and Vegetation Indices. *ISPRS J. Photogramm.* **2019**, *150*, 226–244. [\[CrossRef\]](#)
36. Jin, X.; Li, Z.; Feng, H.; Ren, Z.; Li, S. Deep Neural Network Algorithm for Estimating Maize Biomass Based on Simulated Sentinel 2A Vegetation Indices and Leaf Area Index. *Crop J.* **2020**, *8*, 87–97. [\[CrossRef\]](#)
37. Guo, R.; Gao, J.; Fu, S.; Xiu, Y.; Zhang, S.; Huang, X.; Feng, Q.; Liang, T. Estimating Aboveground Biomass of Alpine Grassland During the Wilting Period Using In Situ Hyperspectral, Sentinel-2, and Sentinel-1 Data. *IEEE Trans. Geosci. Electron.* **2024**, *62*, 1–16. [\[CrossRef\]](#)
38. Eon, R.S.; Goldsmith, S.; Bachmann, C.M.; Tyler, A.C.; Lapszynski, C.S.; Badura, G.P.; Osgood, D.T.; Bret, R. Retrieval of Salt Marsh Above-Ground Biomass from High-Spatial Resolution Hyperspectral Imagery Using Prosall. *Remote Sens.* **2019**, *11*, 1385. [\[CrossRef\]](#)
39. Pandey, P.C.; Anand, A.; Srivastava, P.K. Spatial Distribution of Mangrove Forest Species and Biomass Assessment Using Field Inventory and Earth Observation Hyperspectral Data. *Biodivers. Conserv.* **2019**, *28*, 2143–2162. [\[CrossRef\]](#)
40. Wang, Y.; Li, S.; Zheng, S.; Gao, W.; Zhang, Y.; Cao, B.; Cui, B.; Shao, D. Estimating Biomass and Carbon Sequestration Capacity of *Phragmites Australis* Using Remote Sensing and Growth Dynamics Modeling: A Case Study in Beijing Hanshiqiao Wetland Nature Reserve, China. *Sensors* **2022**, *22*, 3141. [\[CrossRef\]](#)
41. Hemati, M.; Mahdianpari, M.; Shiri, H.; Mohammadimanesh, F. Integrating SAR and Optical Data for Aboveground Biomass Estimation of Coastal Wetlands Using Machine Learning: Multi-Scale Approach. *Remote Sens.* **2024**, *16*, 831. [\[CrossRef\]](#)
42. Li, W.; Dou, Z.; Wang, Y.; Wu, G.; Zhang, M.; Lei, Y.; Ping, Y.; Wang, J.; Cui, L.; Ma, W. Estimation of Above-Ground Biomass of Reed (*Phragmites Communis*) Based on in Situ Hyperspectral Data in Beijing Hanshiqiao Wetland, China. *Wetl. Ecol. Manag.* **2019**, *27*, 87–102. [\[CrossRef\]](#)
43. Dou, Z.; Li, Y.; Cui, L.; Pan, X.; Ma, Q.; Huang, Y.; Lei, Y.; Li, J.; Zhao, X.; Li, W. Hyperspectral Inversion of Suaeda Salsa Biomass Under Different Types of Human Activity in Liaohe Estuary Wetland in North-Eastern China. *Mar. Freshw. Res.* **2020**, *71*, 482–492. [\[CrossRef\]](#)
44. Chen, C.; Ma, Y.; Ren, G.; Wang, J. Aboveground Biomass of Salt-Marsh Vegetation in Coastal Wetlands: Sample Expansion of in Situ Hyperspectral and Sentinel-2 Data Using a Generative Adversarial Network. *Remote Sens. Environ.* **2022**, *270*, 112885. [\[CrossRef\]](#)
45. Luo, S.; Wang, C.; Xi, X.; Pan, F.; Qian, M.; Peng, D.; Nie, S.; Qin, H.; Lin, Y. Retrieving Aboveground Biomass of Wetland *Phragmites Australis* (Common Reed) Using a Combination of Airborne Discrete-Return Lidar and Hyperspectral Data. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *58*, 107–117. [\[CrossRef\]](#)
46. Jensen, D.; Cavanaugh, K.C.; Simard, M.; Okin, G.S.; Castañeda-Moya, E.; McCall, A.; Twilley, R.R. Integrating Imaging Spectrometer and Synthetic Aperture Radar Data for Estimating Wetland Vegetation Aboveground Biomass in Coastal Louisiana. *Remote Sens.* **2019**, *11*, 2533. [\[CrossRef\]](#)
47. Chen, K.; Li, C.; Tang, R. Estimation of The Nitrogen Concentration of Rubber Tree Using Fractional Calculus Augmented NIR Spectra. *Ind. Crops Prod.* **2017**, *108*, 831–839. [\[CrossRef\]](#)
48. Wang, J.; Cui, L.; Gao, W.; Shi, T.; Chen, Y.; Gao, Y. Prediction of Low Heavy Metal Concentrations in Agricultural Soils Using Visible and Near-Infrared Reflectance Spectroscopy. *Geoderma* **2014**, *216*, 1–9. [\[CrossRef\]](#)

49. Ge, X.; Ding, J.; Jin, X.; Wang, J.; Chen, X.; Li, X.; Liu, J.; Xie, B. Estimating Agricultural Soil Moisture Content through UAV-Based Hyperspectral Images in the Arid Region. *Remote Sens.* **2021**, *13*, 1562. [\[CrossRef\]](#)
50. Zhang, A.; Yin, S.; Wang, J.; He, N.; Chai, S.; Pang, H. Grassland Chlorophyll Content Estimation from Drone Hyperspectral Images Combined with Fractional-Order Derivative. *Remote Sens.* **2023**, *15*, 5623. [\[CrossRef\]](#)
51. Wang, J.; Tiyyip, T.; Ding, J.; Zhang, D.; Liu, W.; Wang, F.; Tashpolat, N. Desert Soil Clay Content Estimation Using Reflectance Spectroscopy Preprocessed by Fractional Derivative. *PLoS ONE* **2017**, *12*, e0184836. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Liu, Y.; Liu, S.; Li, J.; Guo, X.; Wang, S.; Lu, J. Estimating Biomass of Winter Oilseed rape Using Vegetation Indices and Texture Metrics Derived from UAV Multispectral Images. *Comput. Electron. Agric.* **2019**, *166*, 105026–105036. [\[CrossRef\]](#)
53. Angela, K.; Heather, M.; David, L.; Mark, S.; Catherine, C. Assessment of RapidEye Vegetation Indices for Estimation of Leaf Area Index and Biomass in Corn and Soybean Crops. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 235–248.
54. Zhang, Y.; Xia, C.; Zhang, X.; Cheng, X.; Feng, G.; Wang, Y.; Gao, Q. Estimating the Maize Biomass by Crop Height and Narrowband Vegetation Indices Derived from UAV-Based Hyperspectral Images. *Ecol. Indic.* **2021**, *129*, 107985. [\[CrossRef\]](#)
55. Liu, Y.; Feng, H.; Fan, Y.; Yue, J.; Chen, R.; Ma, Y.; Bian, M.; Yang, G. Improving Potato Above Ground Biomass Estimation Combining Hyperspectral Data and Harmonic Decomposition Techniques. *Comput. Electron. Agric.* **2024**, *218*, 108699. [\[CrossRef\]](#)
56. Georganos, S.; Grippa, T.; Vanhuyse, S.; Lennert, M.; Shimoni, M.; Kalogirou, S.; Wolff, E. Less Is More: Optimizing Classification Performance Through Feature Selection in A Very-High-Resolution Remote Sensing Object-Based Urban Application. *GIScience Remote Sens.* **2018**, *55*, 221–242. [\[CrossRef\]](#)
57. Nicodemus, K.K. On the Stability and Ranking of Predictors from Random Forest Variable Importance Measures. *Briefings Bioinform.* **2011**, *12*, 369–373. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13 August 2016; pp. 785–794.
59. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5. [\[CrossRef\]](#)
60. Wan, R.; Wang, P.; Wang, X.; Yao, X.; Dai, X. Mapping Aboveground Biomass of Four Typical Vegetation Types in the Poyang Lake Wetlands Based on Random Forest Modelling and Landsat Images. *Front. Plant Sci.* **2019**, *10*, 1281. [\[CrossRef\]](#)
61. Jing, X.; Zou, Q.; Yan, J.; Dong, Y.; Li, B. Remote Sensing Monitoring of Winter Wheat Stripe Rust Based on mRMR-XGBoost Algorithm. *Remote Sens.* **2022**, *14*, 756. [\[CrossRef\]](#)
62. Bartold, M.; Kluczek, M. A Machine Learning Approach for Mapping Chlorophyll Fluorescence at Inland Wetlands. *Remote Sens.* **2023**, *15*, 2392. [\[CrossRef\]](#)
63. Fu, B.; Zuo, P.; Liu, M.; Lan, G.; He, H.; Lao, Z.; Zhang, Y.; Fan, D.; Gao, E. Classifying Vegetation Communities Karst Wetland Synergistic Use of Image Fusion and Object-Based Machine Learning Algorithm with Jilin-1 and UAV Multispectral Images. *Ecol. Indic.* **2022**, *140*, 108989. [\[CrossRef\]](#)
64. Samat, A.; Li, E.; Du, P.; Liu, S.; Xia, J. GPU-Accelerated CatBoost-Forest for Hyperspectral Image Classification Via Parallelized mRMR Ensemble Subspace Feature Selection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3200–3214. [\[CrossRef\]](#)
65. Zhang, Y.; Chang, Q.; Chen, Y.; Liu, Y.; Jiang, D.; Zhang, Z. Hyperspectral Estimation of Chlorophyll Content in Apple Tree Leaf Based on Feature Band Selection and the CatBoost Model. *Agronomy* **2023**, *13*, 2075. [\[CrossRef\]](#)
66. Song, G.; Wang, Q. Species Classification from Hyperspectral Leaf Information Using Machine Learning Approaches. *Ecol. Inform.* **2023**, *76*, 102141. [\[CrossRef\]](#)
67. Bannari, A.; Pacheco, A.; Staenz, K.; McNairn, H.; Omari, K. Estimating and Mapping Crop Residues Cover on Agricultural Lands Using Hyperspectral and IKONOS Data. *Remote Sens. Environ.* **2006**, *104*, 447–459. [\[CrossRef\]](#)
68. Sharma, P.; Mirzan, S.R.; Bhandari, A.; Pimpley, A.; Eswaran, A.; Srinivasan, S.; Shao, L.Q. Evaluating Tree Explanation Methods for Anomaly Reasoning: A Case Study of SHAP TreeExplainer and TreeInterpreter. In Proceedings of the 39th International Conference on Conceptual Modeling (ER), Vienna, Austria, 3–6 November 2020; pp. 35–45.
69. Li, F.; Zeng, Y.; Luo, J.; Ma, R.; Wu, B. Modeling Grassland Aboveground Bio-Mass Using A Pure Vegetation Index. *Ecol. Indic.* **2016**, *62*, 279–288. [\[CrossRef\]](#)
70. Wen, H.; Zhang, Y.; Wang, X.; Wang, R.; Wu, W.; Dong, J. Inversion Study of The Meadow Steppe Above-Ground Biomass Based on Ground and Airborne Hyperspectral Data. *Geocarto Int.* **2024**, *39*, 2370304. [\[CrossRef\]](#)
71. Yang, C.; Xu, J.; Feng, M.; Bai, J.; Sun, H.; Song, L.; Wang, C.; Yang, W.; Xiao, L.; Zhang, M.; et al. Evaluation of Hyperspectral Monitoring Model for Aboveground Dry Biomass of Winter Wheat by Using Multiple Factors. *Agronomy* **2023**, *13*, 983. [\[CrossRef\]](#)
72. Du, Y.; Wang, J.; Liu, Z.; Yu, H.; Li, Z.; Cheng, H. Evaluation on Spaceborne Multispectral Images, Airborne Hyperspectral, and LiDAR Data for Extracting Spatial Distribution and Estimating Aboveground Biomass of Wetland Vegetation *Suaeda salsa*. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *12*, 200–209. [\[CrossRef\]](#)
73. Gnyp, M.L.; Miao, Y.; Yuan, F.; Ustin, S.L.; Yu, K.; Yao, Y.; Huang, S.; Bareth, G. Hyperspectral Canopy Sensing of Paddy Rice Aboveground Biomass at different growth stages. *Field Crops Res.* **2014**, *155*, 42–55. [\[CrossRef\]](#)
74. Cui, L.; Dou, Z.; Liu, Z.; Zuo, X.; Lei, Y.; Li, J.; Zhao, X.; Zhai, X.; Pan, X.; Li, W. Hyperspectral Inversion of *Phragmites communis* Carbon, Nitrogen, and Phosphorus Stoichiometry Using Three Models. *Remote Sens.* **2020**, *12*, 1998. [\[CrossRef\]](#)
75. Cui, S.C.; Zhou, K.F. Acomparision of the Predictive Potential of Various Vegetation Indices for Leaf Chlorophyll Content. *Earth Sci. Inform.* **2017**, *10*, 169–181. [\[CrossRef\]](#)

76. Caturegli, L.; Gaetani, M.; Volterrani, M.; Magni, S.; Minelli, A.; Baldi, A.; Brandani, G.; Mancini, M.; Lenzi, A.; Orlandini, S.; et al. Normalized Difference Vegetation Index Versus Dark Green Colour Index to Estimate Nitrogen Status on Bermudagrass Hybrid and Tall Fescue. *Int. J. Remote Sens.* **2020**, *41*, 455–470. [[CrossRef](#)]
77. Jiang, X.; Zhen, J.; Miao, J.; Zhao, D.; Shen, Z.; Jiang, J.; Gao, C.; Wu, G.; Wang, J. Newly-Developed Three-Band Hyperspectral Vegetation Index for Estimating Leaf Relative Chlorophyll Content of Mangrove under Different Severities of Pest and Disease. *Ecol. Indic.* **2022**, *140*, 108978. [[CrossRef](#)]
78. Viljanen, N.; Honkavaara, E.; Näsi, R.; Hakala, T.; Niemeläinen, O.; Kaivosoja, J. A Novel Machine Learning Method for Estimating Biomass of Grass Swards Using A Photogrammetric Canopy Height Model, Images and Vegetation Indices Captured by A Drone. *Agriculture* **2018**, *8*, 70. [[CrossRef](#)]
79. Li, C.; Ma, C.; Pei, H.; Feng, H.; Shi, J.; Wang, Y.; Chen, W.; Li, Y.; Feng, X.; Shi, Y. Estimation of Potato Biomass and Yield Based on Machine Learning from Hyperspectral Remote Sensing Data. *J. Agric. Sci. Technol. B* **2020**, *10*, 195–213.
80. Li, W.; Zuo, X.; Liu, Z.; Nie, L.; Li, H.; Wang, J.; Dou, Z.; Cai, Y.; Zhai, X.; Cui, L. Predictions of *Spartina alterniflora* Leaf Functional Traits Based on Hyperspectral Data and Machine Learning Models. *Eur. J. Remote Sens.* **2024**, *57*, 2294951. [[CrossRef](#)]
81. Adam, E.; Mutanga, O.; Abdel-Rahman, E.M.; Ismail, R. Estimating Standing Biomass in Papyrus (*Cyperus papyrus* L.) Swamp: Exploratory of In Situ Hyperspectral Indices and Random Forest Regression. *Int. J. Remote Sens.* **2014**, *35*, 693–714. [[CrossRef](#)]
82. Tang, X.; Dou, Z.; Cui, L.; Liu, Z.; Gao, C.; Wang, J.; Li, J.; Lei, Y.; Zhao, X.; Zhai, X.; et al. Hyperspectral Prediction of Mangrove Leaf Stoichiometries in Different Restoration Areas Based on Machine Learning Models. *J. Appl. Remote Sens.* **2022**, *16*, 034525. [[CrossRef](#)]
83. Teluguntla, P.; Thenkabail, P.S.; Oliphant, A.; Xiong, J.; Gumma, M.K.; Congalton, R.G.; Yadav, K.; Huete, A. A 30-m Landsat-Derived Cropland Extent Product of Australia and China Using Random Forest Machine Learning Algorithm on Google Earth Engine Cloud Computing Platform. *ISPRS J. Photogramm. Remote Sens.* **2018**, *144*, 325–340. [[CrossRef](#)]
84. Gasela, M.; Kganyago, M.; De Jager, G. Using Resampled nSight-2 Hyperspectral Data and Various Machine Learning Classifiers for Discriminating Wetland Plant Species in A Ramsar Wetland Site, South Africa. *Appl. Geomat.* **2024**, *16*, 429–440. [[CrossRef](#)]
85. Byrd, K.B.; Ballanti, L.; Thomas, N.; Nguyen, D.; Holmquist, J.R.; Simard, M.; Windham-Myers, L. A Remote Sensing-Based Model of Tidal Marsh Aboveground Carbon Stocks for The Conterminous United States. *ISPRS J. Photogramm. Remote Sens.* **2018**, *139*, 255–271. [[CrossRef](#)]
86. Yang, L.; Shami, A. On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice. *Neurocomputing* **2020**, *415*, 295–316. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.