Hyperspectral Analysis and Regression Modeling of SPAD Measurements in Leaves of Three Mangrove Species

Huazhe Li 1,2,3, Lijuan Cui 1,2,3, Zhiguo Dou 1,2,3, Junjie Wang 4,5, Xiajie Zhai 1,2,3, Jing Li 1,2,3, Xinshe Zhao 1,2,3, Yinru Lei 1,2,3, Jinzhi Wang 1,2,3 and Wei Li 1,2,3,*

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 College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, China
5 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.: +86-10-6282-4186

Abstract: Mangroves have important roles in regulating climate change, and in reducing the impact of wind and waves. Analysis of the chlorophyll content of mangroves is important for monitoring their health, and their conservation and management. Thus, this study aimed to apply four regression models, eXtreme Gradient Boosting (XGBoost), Random Forest (RF), Partial Least Squares (PLS) and Adaptive Boosting (AdaBoost), to study the inversion of Soil Plant Analysis Development (SPAD) values obtained from near-ground hyperspectral data of three dominant species, Bruguiera sexangula (Lour.) Poir. (B. sexangula), Ceriops tagal (Perr.) C. B. Rob. (C. tagal) and Rhizophora apiculata Blume (R. apiculata) in Qinglan Port Mangrove Nature Reserve. The accuracy of the model was evaluated using $R^2$, RMSE, and MAE. The mean SPAD values of $R. apiculata$ (SPAD$_{avg} = 66.57$), with a smaller dispersion (coefficient of variation of 6.59%), were higher than those of $C. tagal$ (SPAD$_{avg} = 61.56$) and $B. sexangula$ (SPAD$_{avg} = 58.60$). The first-order differential transformation of the spectral data improved the accuracy of the prediction model; $R^2$ was mostly distributed in the interval of 0.4 to 0.8. The accuracy of the XGBoost model was less affected by species differences with the best stability, with RMSE at approximately 3.5 and MAE at approximately 2.85. This study provides a technical reference for large-scale detection and management of mangroves.

Keywords: hyperspectral data; successive projections algorithm; mangroves; SPAD measurements

1. Introduction

Mangrove forests are a special type of coastal ecosystem in the transition zone between terrestrial and marine ecosystems [1,2]. They are a complex-structured salt-tolerant wetland ecological community composed of woody plants, such as evergreen shrubs and trees, with mangrove plants as the mainstay. These plants can tolerate extreme environments such as high salinity, high temperature, flooding, and poor soils, and are important for shoreline protection. They play a role in slowing coastal erosion, provide resiliency to natural disasters such as storms and help protect agricultural land [3,4]. Mangrove plants provide habitats and breeding sites for many marine animals and are an important component in maintaining ecological balance. Mangroves have high socioeconomic value [5], but over-exploitation and human activities have threatened growth in mangrove forests and there has been a significant reduction in mangrove area [6–8]. In recent years, there has been an increasing awareness of the need to protect mangrove ecosystems, and many mangrove conservation and restoration efforts have been made [9,10] to ensure the sustainable development of mangrove ecosystems.
Chlorophyll is an important chemical in plants that is involved in both photosynthesis and growth and development [11]. Chlorophyll content is a key indicator that enables effective monitoring of plant health and overall growth. The traditional method for measuring chlorophyll content is chemical analysis in the laboratory, which is time-consuming and costly, although the accuracy of the measurement is high [12,13]. Many studies have shown a strong correlation between chlorophyll content and Soil and Plant Analysis Development (SPAD) values, which are typically measured with a convenient handheld meter that does not destroy plant tissue [14]. Brown et al., (2022) demonstrated that SPAD measurements can be used to represent leaf chlorophyll content in forests under calibrated conditions [15]. In mangrove studies conducted as early as 1997, Connelly reported a large correlation ($R^2 > 0.6$) between the chlorophyll content of mangroves and SPAD-502 readings [16]. Neres et al., (2020) also demonstrated that chlorophyll concentration indices of SPAD-502 and those of mangroves can be inter-converted [17]. Therefore, SPAD can be used to represent the leaf chlorophyll content of mangrove plants. The determination of the SPAD values of mangrove plant leaves might be expected to be effective in indicating the health status of mangrove plants and provide a basis for decision making in the monitoring and management of mangroves.

Accurate leaf SPAD values can be obtained through field measurements; however, the complex growth environment of mangroves makes field experiments less accessible, and only the SPAD measurements of plant leaves can be obtained, making it difficult to reflect the changes in the SPAD measurements of the whole plant or even the plant community. Remote sensing technology has attracted the attention of many ecologists because of its ability to quickly acquire target feature information without contact, enabling monitoring over a large area. While multispectral data have few bands and low resolution and can only monitor certain indicators under specific conditions, hyperspectral data have the advantages of more bands and higher spectral resolution and in many aspects contain more information to achieve more accurate monitoring [18,19]. Most studies on mangrove forests that have used remote sensing have been carried out to classify mangroves and invert aboveground biomass [20,21]. For example, Yang et al., (2022) proposed an Enhanced Mangrove Vegetation Index (EMVI) based on hyperspectral images and achieved a classification of mangroves [22]. Pandey et al., (2019) classified ten species of mangrove plants and mapped their biomass distributions based on field collection and hyperspectral data [23]. These studies are important for the analysis of mangrove landscape patterns and growth conditions, but only a few studies have focused on the SPAD measurements of mangroves.

In recent years, studies on the prediction of mangrove chlorophyll content using hyperspectral data have progressed considerably, with some studies focusing on canopy chlorophyll content and others on leaf SPAD measurements. For example, Dou et al., (2018) estimated the SPAD measurements of mangroves at different stages of restoration in coastal wetlands in Quanzhou using a full-band spectral model and a red-edge position regression model with real-world hyperspectral data and SPAD values, and found that the full-band model had better inversion accuracy [24]. Wang et al., (2020) conducted a prediction study of the SPAD of mangrove plant leaves from near-ground hyperspectral data using continuous wavelet transform and Random Forest (RF) regression models and explored the effect of soil factors on the accuracy of the inversion [25]. Zhen et al., (2021) used Sentinel-2 imagery and measured SPAD data to develop linear regression and Kernel Ridge regression models for SPAD inversion with five newly developed vegetation indices and implemented SPAD mapping for the entire region [26]. Jiang et al., (2022) newly developed several new three-band vegetation indices based on hyperspectral remote sensing data to assess the chlorophyll content of mangroves under stress from pests and found that these vegetation indices responded to changes in the chlorophyll content of mangrove leaves and could contribute to pest and disease monitoring in mangroves [27]. Fu et al., (2022) used Zhuhai-1 hyperspectral satellite (OHS) imagery and Sentinel-2A multispectral data to calculate traditional vegetation indices and combined vegetation indices and constructed a
remote sensing inversion model of mangrove Canopy Chlorophyll Content (CCC) based on a single linear regression algorithm, a machine learning regression algorithm, and a stack integrated learning regression algorithm to explore the best remote sensing inversion model for CCC of mangroves in the Beibu Gulf [28]. However, there have been relatively few studies on the application of hyperspectral data and integrated learning algorithms to the inversion of the SPAD measurements of the leaves of different species of mangrove plants using first-order differential processing.

Thus, to this end, we measured leaf hyperspectral data and SPAD values and tested four regression models: eXtreme Gradient Boosting (XGBoost), Random Forest (RF), Partial Least Squares (PLS), and Adaptive Boosting (AdaBoost). We predicted the SPAD values of leaves from three mangrove species in the Hainan Qinglan Port Mangrove Nature Reserve and explored the applicability of the four regression models to predict SPAD values at the leaf scale.

2. Materials and Methods

2.1. Study Area

Qinglan Port Mangrove Nature Reserve (19°22’ N–19°35’ N, 110°40’ E–110°48’ E) is located in the northeast of Hainan Island within the boundary of Wenchang City, and has a total area of 2931.20 ha. It is one of the largest mangrove areas in China, and has the largest distribution of natural mangrove vegetation. According to a survey, 24 mangrove species exist in the Qinglan Port Mangrove Nature Reserve, accounting for 85.71% of those present within the whole of China [1], and the dominant mangrove species in the typical vegetation of the reserve are Bruguiera sexangula (Lour.) Poir. (B. sexangula), Ceriops tagal (Perr.) C. B. Rob. (C. tagal) and Rhizophora apiculata Blume (R. apiculata). The reserve is located at the northern edge of the tropics and belongs to the tropical monsoon maritime climate zone, with a warm climate, little seasonal variation, an average annual temperature of 23.9 °C, sufficient sunshine, abundant rainfall, and an average annual precipitation of 1799.4 mm [29]. Qinglan Port is a lagoon harbor and has many rivers such as the Wenchang River and the Wenjiao River flowing into it. It forms a typical lagoon-estuarine wetland habitat with deep coastal silt and weak winds and waves, which provides good survival conditions for the growth of mangroves [30].

2.2. Data Acquisition

In April 2019, we sampled the leaves of three dominant mangrove species, B. sexangula, C. tagal and R. apiculata, along the coastline in the Bamen Bay area of the reserve, setting 60 sample points for each species for a total of 180 sample points. Each sample point was set on a single area of species within a 10 m square, ensuring as far as possible that the distance between sample points was greater than 50 m, and random sampling was conducted in the survey area. We selected the third leaf from the top of different branches of the same mangrove tree for collection at each sampling site as a way of ensuring that the harvested leaves were mature. Ten leaves from each mangrove tree were collected as duplicate samples, and the collected leaves were placed in sealed bags and quickly transported to a nearby laboratory where their spectra and SPAD values were measured.

2.2.1. Hyperspectral Data Acquisition

We used an ASD FieldSpec 4 (Analytical Spectral Devices, Inc., Longmont, CO, USA) portable ground object spectrometer to measure the collected leaf spectra of mangrove plants in the band range of 350–2500 nm with a resolution of 1.4 nm in the range of 350–1000 nm and 1.1 nm in the range of 1001–2500 nm. For indoor hyperspectral measurements of leaves, we used an embedded light source (a built-in halogen lamp) to directly measure the leaves. Measurements were performed using the ASD FieldSpec 4 with a leaf clip and a spectrometer for leaf spectral data acquisition, placing the leaf clip in the middle of the leaf and avoiding the main leaf veins. Ten spectral data points were measured for each leaf, and the average of 100 measurements per 10 leaves was used as the final result.
The spectrometer was recalibrated every 15 min during the acquisition period using a white calibration plate with a known reflectance of 99% [31].

2.2.2. SPAD Value Measurement

The SPAD values were measured immediately after measuring the leaf spectra, taking care to avoid measuring the main leaf veins, at equal intervals from the tip to the end of the leaf. We used a chlorophyll meter SPAD-502 Plus (Konica Minolta, Inc., Tokyo, Japan) to measure five parts of each leaf: tip, upper middle, middle, lower middle, and end of the leaf [32], and the results of the five measurements were averaged as the SPAD readings for that leaf. SPAD values range from 0 to 100 and are dimensionless. We measured and recorded the SPAD values for 180 samples from leaves of three mangrove species.

2.3. Data Pre-Processing

First, the leaf hyperspectral data were preprocessed using ViewSpec Pro 6.0. The 100 spectral data points collected from 10 leaves of each sample were averaged, and the results were exported as the leaf reflectance spectral data of this mangrove sample. Because the leaf spectra have systematic noise at 350–399 nm and 2451–2500 nm, we first removed the noise from the leaf reflection spectra of 180 samples and retained only the band from 400–2450 nm for analysis [33]. We then smoothed the 400–2450 nm band with a second-order polynomial fit and a Savitzky-Golay (SG) smoothing filter with a 7-data-point window size [34], which ensured that the shape and width of the signal were unchanged while filtering out noise.

This differential transform can eliminate systematic errors between spectral data, weakens the effects of background noise, such as atmospheric radiation, scattering, and absorption, and enhances subtle changes in the slope of the spectral profile [35]. For vegetation, this variation is related to the biochemical absorption characteristics of the vegetation, which facilitates the extraction of the spectral features of the detected features and is one of the most common spectral transformation methods. In this study, a first-order differential transform was used to process the original spectral data. The transform equation is as follows:

$$FD(\lambda_i) = \frac{R(\lambda_{i+1}) - R(\lambda_{i-1})}{\Delta \lambda}$$

where $\lambda_i$ is the wavelength of each band, $FD(\lambda_i)$ is the first-order differential spectral value at wavelength $\lambda_i$, and $\Delta \lambda$ is the wavelength value between band $i$ and band $i + 1$.

The successive projection algorithm (SPA) is a forward iterative search method that starts with one wavelength and then adds a new variable in each iteration until the number of selected variables reaches a set value N. The aim of the SPA is to select the wavelength with the least redundant spectral information to solve the covariance problem [36,37]. We used the original spectral reflectance (OR) and first-order derivative reflectance (FD) as independent variables and the SPAD values of the three mangrove plants as dependent variables. We implemented the SPA using MATLAB 2016a to use the calculated bands as the characteristic bands for predicting SPAD.

2.4. Model Building and Validation

The 60 samples from each mangrove plant were divided into 42 training sets and 18 validation sets at a ratio of 7:3. The training sets were used for model building, and the validation sets were used to validate the model prediction accuracy. The training and validation of the four regression models were implemented using the Sklearn library in Python 3.9.

The accuracy of the model prediction results was evaluated using the $R^2$, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). The larger the $R^2$, the smaller
the RMSE and MAE, indicating a higher accuracy and stability of the model and better prediction [38,39]. The calculation formulae are as follows:

\[ R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - x_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \]  

(2)

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2} \]  

(3)

\[ \text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - x_i| \]  

(4)

where \( x_i \) and \( y_i \) are the measured and predicted values of sample \( i \), \( \bar{x} \) and \( \bar{y} \) are the means of the measured and predicted values, respectively; and \( n \) is the number of samples.

3. Results

3.1. SPAD Statistical Analysis

We statistically analyzed the SPAD measurements of the leaves of the three mangrove species and performed a one-way ANOVA based on SPSS for the SPAD measurements of the leaves of the three mangrove plants, and the results are shown in Table 1. The standard deviations and coefficients of variation of the SPAD measurements of the leaves of the three mangrove plants showed the following trends: \( R. \) apiculata < \( C. \) tagal < \( B. \) sexangula, and their coefficients of variation were all approximately 10%, with the smallest being 6.59% for \( R. \) apiculata. In addition, by performing one-way ANOVA, we found that the leaf SPAD measurements of the three mangrove plants were significantly different at the level of \( p < 0.05 \); therefore, we independently analyzed the hyperspectral inversions of the leaf SPAD values of the three mangrove plants.

<table>
<thead>
<tr>
<th>Species</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Average</th>
<th>Standard Deviation</th>
<th>Coefficient of Variation</th>
<th>Significant Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bruguiera sexangula</td>
<td>47.62</td>
<td>75.80</td>
<td>58.60</td>
<td>6.07</td>
<td>10.36%</td>
<td>a</td>
</tr>
<tr>
<td>(B. sexangula)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ceriops tagal (C. tagal)</td>
<td>49.94</td>
<td>69.74</td>
<td>61.56</td>
<td>5.16</td>
<td>8.39%</td>
<td>b</td>
</tr>
<tr>
<td>Rhizophora apiculata</td>
<td>57.63</td>
<td>75.71</td>
<td>66.57</td>
<td>4.39</td>
<td>6.59%</td>
<td>c</td>
</tr>
<tr>
<td>(R. apiculata)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The letters a, b, and c represent significant differences in leaf SPAD measurements of different mangrove species.

3.2. Comparison of Spectral Curves

We averaged the spectral data at 400–2450 nm for each mangrove plant leaf after smoothing the reflectance data with an SG filter and removing the noise band to obtain the average reflectance spectral curves of the leaves from the three mangrove species, as shown in Figure 1. Overall, all three spectral curves were consistent with the spectral trend characteristics of green vegetation. Two absorption valleys appeared in the blue light band around 490 nm and the red band around 670 nm, which were formed by the strong absorption by chlorophyll in plants in the visible light band, whereas a small reflection peak was observed in the green light band around 550 nm. In addition, high reflectance is observed in the 800–1300 nm band because of the strong reflectance of near-infrared light by the cellular structure and pigments of the vegetation, resulting in a sharp increase in the reflectance curve between 670 nm and 750 nm. Absorption valleys are also formed at 1450 nm and 1900 nm of the spectral curve due to the strong absorption by water in the plants.
In addition, the reflectance of *C. tagal* was significantly higher than that of *B. sexangula* and *R. apiculata* in the visible band and in the 800–1300 nm interval, whereas in the 1500–1800 nm and 2100–2300 nm intervals, the reflectance of *B. sexangula* was higher and the reflectance of *R. apiculata* in each band was relatively smaller than that of the other two mangrove species.

### 3.3. Successive Projection Algorithm for Filtering Feature Bands

The continuous projection method can maximize the screening of the characteristic wavelengths that contain the key information of the spectral reflection curve, and the covariance between variables is small. Application of the method achieves dimensionality reduction of the hyperspectral data, effectively reduces the input variables in the model operation, and improves the efficiency of the model operation. The results of the feature bands screened using the continuous projection algorithm are shown in Figure 2 and Table 2. For both the OR and FD data, the feature bands screened by the SPA were more distributed in the parts with evident changes in characteristics, such as wave peaks or troughs, which reduced the model input variables while retaining the feature information of the spectrum as much as possible.

**Figure 1.** Average spectral curves of three mangrove species.

**Figure 2.** Results of SPA feature band screening. (a–c) Screening results of the origin spectral data of *B. sexangula*, *C. tagal*, and *R. apiculata*, respectively; (d–f) screening results of the first-order differential spectral data of *B. sexangula*, *C. tagal*, and *R. apiculata*, respectively.
Table 2. Feature band statistics.

<table>
<thead>
<tr>
<th>Species</th>
<th>Processing Method</th>
<th>Number of Bands</th>
<th>Feature Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. sexangula</td>
<td>Original reflectance (OR)</td>
<td>16</td>
<td>404, 418, 420, 444, 535, 650, 699, 736, 1389, 1663, 1893, 2186, 2347, 2435, 2445</td>
</tr>
<tr>
<td></td>
<td>First-order differential reflectance (FD)</td>
<td>6</td>
<td>674, 703, 1666, 1803, 1805, 1842</td>
</tr>
<tr>
<td>C. tagal</td>
<td>Original reflectance (OR)</td>
<td>14</td>
<td>401, 407, 422, 424, 427, 504, 1421, 1694, 1864, 1997, 2145, 2399, 2420, 2436</td>
</tr>
<tr>
<td></td>
<td>First-order differential reflectance (FD)</td>
<td>16</td>
<td>407, 415, 430, 432, 439, 467, 718, 1064, 1073, 1169, 1925, 1931, 2205, 2226, 2250, 2412</td>
</tr>
<tr>
<td>R. apiculata</td>
<td>Original reflectance (OR)</td>
<td>11</td>
<td>407, 413, 422, 435, 462, 637, 725, 775, 998, 1125, 2446</td>
</tr>
<tr>
<td></td>
<td>First-order differential reflectance (FD)</td>
<td>14</td>
<td>405, 417, 423, 433, 464, 481, 887, 1656, 1923, 1931, 1934, 1941, 2220, 2223</td>
</tr>
</tbody>
</table>

We extracted more than ten feature bands from the different spectral treatments of each mangrove species using SPA. Except for the first-order differential transformation spectra of B. sexangula, six feature bands were extracted. Finally, we combined the original spectra and first-order differentially transformed spectra as the feature set for this species, with 22 feature bands extracted for B. sexangula, 30 for C. tagal, and 25 for R. apiculata.

3.4. Model Building

Based on the feature bands screened by SPA and the measured SPAD values, XGBoost, RF, PLS, and AdaBoost regression models were developed for the 42 training samples from the three mangrove species.

3.4.1. SPAD Regression Model of B. sexangula

The eight regression models built using the original spectral data and the first-order differential spectral data of B. sexangula are shown in Figure 3. For the OR data, the AdaBoost model had the highest modeling accuracy, with an $R^2$ of 0.98 and the smallest RMSE of 0.79. For the FD data, both the XGBoost and AdaBoost models had high accuracy, with an $R^2$ of 0.99; however, the RMSE of the XGBoost model was smaller at 0.61. For both the OR and FD data, the accuracy of the PLS regression model was significantly lower than that of the other three models.

3.4.2. SPAD Regression Model of C. tagal

The modeling results based on the spectral data and SPAD values of C. tagal are shown in Figure 4. As with the modeling results of B. sexangula, the modeling accuracy of the PLS
model for both spectral data was also the lowest, and the distribution of its data points was more discrete, especially for the modeling of the original spectral data; the $R^2$ was only 0.71 and the RMSE was 2.60, which was significantly higher than that of the other models. In addition, the four regression models based on FD data have slightly improved accuracy compared with the regression models based on FD data, as shown by the different degrees of improvement in $R^2$ and reduction in RMSE; the best accuracy is the AdaBoost model based on FD data, with an $R^2$ of 0.99 and an RMSE of 0.49.

Figure 4. Modeling results of C. tagal. (a–d) Scatter plots of XGBoost, RF, PLS and AdaBoost models built using the first-order differential spectral data, respectively; (e–h) scatter plots of XGBoost, RF, PLS and AdaBoost models built using the original spectral data, respectively.

3.4.3. SPAD Regression Model of R. apiculata

The modeling results of R. apiculata are presented in Figure 5, showing that the PLS model, based on OR data, had the worst accuracy of all regression models built for the three mangrove plants, with an $R^2$ of only 0.64 and an RMSE of 2.68. However, the accuracy of the AdaBoost model based on the FD data is very high, with $R^2$ reaching 0.99 and only 0.11 for the RMSE value. The data points are densely distributed on a 1:1 diagonal, with only a few sample points deviating from the 1:1 diagonal. In terms of the overall modeling accuracy, AdaBoost > XGBoost > RF > PLS, and it is evident that a first-order differential transformation of the OR data can improve the modeling accuracy.

Figure 5. Modeling results of R. apiculata. (a–d) Scatter plots of XGBoost, RF, PLS and AdaBoost models built using the original spectral data, respectively; (e–h) scatter plots of XGBoost, RF, PLS and AdaBoost models built using the first-order differential spectral data, respectively.
3.5. Evaluation of Model Validation Accuracy

We validated the accuracy of several regression models using the remaining 18 validation samples for each mangrove plant using three metrics, $R^2$, RMSE, and MAE, to evaluate the stability and accuracy of different spectral treatments and different regression models for predicting the SPAD values of mangrove plant leaves. The results are presented in Table 3 and Figure 6.

Table 3. Validation of regression model accuracy.

<table>
<thead>
<tr>
<th>Model</th>
<th>$B. \text{ sexangula}$</th>
<th>$C. \text{ tagal}$</th>
<th>$R. \text{ apiculata}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>RMSE</td>
<td>MAE</td>
</tr>
<tr>
<td>XGBoost OR</td>
<td>0.72</td>
<td>3.23</td>
<td>2.79</td>
</tr>
<tr>
<td>XGBoost FD</td>
<td>0.65</td>
<td>3.48</td>
<td>2.87</td>
</tr>
<tr>
<td>Random Forest OR</td>
<td>0.60</td>
<td>3.70</td>
<td>2.94</td>
</tr>
<tr>
<td>Random Forest FD</td>
<td>0.78</td>
<td>3.27</td>
<td>2.52</td>
</tr>
<tr>
<td>Partial Least Squares OR</td>
<td>0.74</td>
<td>3.87</td>
<td>3.03</td>
</tr>
<tr>
<td>Partial Least Squares FD</td>
<td>0.73</td>
<td>3.99</td>
<td>2.80</td>
</tr>
<tr>
<td>AdaBoost OR</td>
<td>0.51</td>
<td>4.14</td>
<td>3.23</td>
</tr>
<tr>
<td>AdaBoost FD</td>
<td>0.67</td>
<td>3.49</td>
<td>2.76</td>
</tr>
</tbody>
</table>

Figure 6. Radar plot for model validation. (a) Radar plot of $R^2$, (b) radar plot of RMSE, (c) radar plot of MAE.

Among the SPAD prediction models for $B. \text{ sexangula}$, the validation accuracy of the RF model based on the FD data was the highest ($R^2 = 0.78$, RMSE = 3.27, MAE = 2.52), whereas the AdaBoost model exhibited extremely high modeling accuracy, and its validation accuracy was the worst among the four models. For $C. \text{ tagal}$, the best validation accuracy was for the PLS model based on the FD data ($R^2 = 0.77$, RMSE = 3.70, MAE = 3.21), which was a significant improvement compared with the PLS model based on the OR data ($R^2 = 0.37$, RMSE = 4.96, MAE = 3.95). The validation accuracy of the prediction models of $R. \text{ apiculata}$ is poor, and the accuracy decreases significantly compared with that of the modeling, and its $R^2$ is mostly distributed in the interval of 0.2–0.4. Only the validation accuracy of the AdaBoost model based on the FD data is a little better, but its $R^2$ is only 0.44, and the RMSE and MAE are 3.38 and 2.81, respectively.

4. Discussion

The chlorophyll content of leaves is an important indicator of plant growth and development, and many researchers have conducted related studies on remote sensing estimation of chlorophyll content in leaves [40–42]. However, relatively few studies have been conducted on the rapid prediction of chlorophyll content in the leaves of multiple
mangrove species. The rapid prediction of leaf chlorophyll content in a variety of mangrove plants based on hyperspectral techniques is beneficial for the accurate monitoring of the management and growth status of mangrove plants on a large scale and can also provide a basis for species classification studies based on differences in chlorophyll content. This study compared the accuracy and stability of four regression models (XGBoost, RF, PLS, and AdaBoost) in SPAD inversion based on hyperspectral data measured by an ASD geospectrometer, which can provide a methodological reference for leaf-scale SPAD prediction.

The volume of the hyperspectral data was large, and the correlation between the bands was strong. Before using hyperspectral data for data analysis, extraction of feature bands using feature selection methods and the elimination of bands that are not correlated with chlorophyll markedly reduces the complexity of the model and is an important part of realizing a reduced and simplified model for hyperspectral data [43,44]. Using SPA, we reduced the number of bands from 2051 to approximately 10–20 for both the OR and FD spectral data, which greatly reduced the dimensionality of the data. However, when the feature bands were extracted, sometimes the number of bands given by SPA was not the number of bands corresponding to the smallest RMSE; therefore, we assumed that the optimal number of feature bands should be shuffled by combining the effects of the number of bands and the RMSE value on the model. In addition, the feature bands of both the OR and FD data were mainly concentrated in the parts of the wave peaks and troughs with more evident changes in reflectance (Figure 2), indicating that the feature bands extracted by the SPA retained well the characteristic information of the spectra. The feature bands of the leaf spectra of the three mangrove species were extracted using the SPA.

We used four regression models to explore the inversion of SPAD. The RF and PLS models have been widely used for the prediction of the elemental content of vegetation, confirming the universality of these two models, whereas the XGBoost and AdaBoost models are still relatively rarely used for the prediction of plant physiological and biochemical indicators. The RF model employs a modified CART decision tree as a learner, which significantly enhances its ability to handle nonlinear multiple-compound regression problems. As a decision tree is split and combined based on a tree structure, it can adapt to variable nonlinear data distributions, facilitating the handling of various complex data structures and improving the efficiency and accuracy of the algorithm [45,46]. PLS models have the advantage of removing multicollinearity and handling high-dimensional data, which can simplify the model and improve regression accuracy by compressing the dimensionality of independent variables [47]. AdaBoost is an integrated learning method that can be used to improve the accuracy of a classifier gradually by continuously iterating to adjust the weights of the samples. It uses multiple weak classifiers combined with a single strong classifier to achieve higher accuracy than a single weak classifier [48]. XGBoost is an integrated learning algorithm based on a tree model that is optimized and improved using a Gradient Boosting Machine algorithm. It is widely used in practical machine learning applications owing to its advantages of high accuracy and robustness, its capacity for handling missing values and high-dimensional features, it prevents overfitting, and is highly interpretable [49,50].

In this study, the PLS model was worse than the other four models, which may be related to the small number of samples in the training dataset, whereas an excessively small amount of data may lead to the inability of PLS to find effective patterns to complete the modeling, thus leading to poor modeling accuracy. In the validation of the models, the validation accuracy of all models decreased significantly compared to the modeling accuracy, which may be due to the use of the default parameters of the models and the relatively small number of validation samples, resulting in underfitting or overfitting of the models. Further in-depth research is needed to optimize the parameters of each regression model for the SPAD inversion of mangrove leaves. In addition, as shown in Table 3, regardless of the regression model, the validation accuracy showed a trend of $B. \text{sexangula} > C. \text{tagal} > R. \text{apiculate}$, with large differences between the species. The factors
that lead to the large variation in the accuracy of the model between species are not yet known, which again requires in depth studies to explore the intrinsic reasons.

Previous studies have laid the foundation for studying chlorophyll inversion in mangrove plants. Table 4 shows some of the results of researchers who have achieved good inversion results by focusing more on narrow-band vegetation indices to improve the accuracy of inversion. We used several regression models to invert chlorophyll content and found that XGBoost was relatively stable. It was also interesting to observe the effects of combining the vegetation index with the XGBoost regression model. Additionally, many researchers are studying the use of stacking models for the inversion of metrics, which can combine the advantages of multiple algorithms and be more generalizable [51]. For example, Liu et al., (2023) constructed a high-precision stacked inversion model for apple leaf chlorophyll content based on boosting and stacking strategies; the $R^2$ value of its validation set reached 0.9644, which was better than that of the traditional integrated learning model [52]. Lin et al., (2022) constructed a stacked AdaBoost model to invert the heavy metal content in soil; the stacked model was more stable and had better prediction accuracy than the traditional single machine learning model [53]. The performance of the stacking model during the inversion of leaf SPAD in mangrove plants will be the focus of future research.

### Table 4. Comparison of other research results with this paper.

<table>
<thead>
<tr>
<th>Index</th>
<th>Specie</th>
<th>Method</th>
<th>Accuracy</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPAD</td>
<td><em>Aegiceras corniculatum</em></td>
<td>Newly developed three-band Vis: VI1</td>
<td>$R^2 = 0.658$ &lt;br&gt; RPD = 1.724</td>
<td>[26]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Newly developed three-band Vis: VI4</td>
<td>$R^2 = 0.671$ &lt;br&gt; RPD = 1.753</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Aegiceras corniculatum</em>, <em>Bruguiera gymnorrhiza</em>, <em>Kandelia obovata</em>, <em>Rhizophora stylosa</em>, and <em>Avicennia marina</em></td>
<td>Newly developed RSSI (B8, B2, B3)</td>
<td>$R^2_{adj} = 0.801$&lt;br&gt;RMSE$_V$ = 3.46&lt;br&gt;RE$_V$ = 4.63</td>
<td>[25]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KRR model (9 bands without red-edge bands of Sentinel-2)</td>
<td>$R^2_{adj} = 0.776$&lt;br&gt;RMSE$_V$ = 3.68&lt;br&gt;RE$_V$ = 4.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Aegiceras corniculatum</em></td>
<td>RF</td>
<td>$R^2_{Val} = 0.671$&lt;br&gt;RPD = 1.552</td>
<td>[54]</td>
</tr>
<tr>
<td></td>
<td>Long-term restoration of <em>Kandelia candel</em></td>
<td>Red edge position model</td>
<td>$R^2 = 0.8392$&lt;br&gt;RMSE = 3.98</td>
<td>[23]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Full-wave band spectrum model</td>
<td>$R^2 = 0.9353$&lt;br&gt;RMSE = 2.26</td>
<td></td>
</tr>
<tr>
<td>Chlorophyll</td>
<td><em>Rhizophora apiculata</em> and <em>Bruguiera gymnorrhiza</em></td>
<td>Simple ratio (559 nm/885 nm)</td>
<td>$R^2 = 0.75$&lt;br&gt;RMSE = 0.60</td>
<td>[55]</td>
</tr>
<tr>
<td></td>
<td><em>Kandelia candel</em>, <em>Avicennia marina</em>, <em>Aegiceras corniculatum</em>, and <em>Sonneratia apetala</em></td>
<td>IMDATT = (R$<em>{527} - R</em>{746}$)/(R$<em>{527} - R</em>{747}$)</td>
<td>$r = -0.88$&lt;br&gt;RMSE = 8.70</td>
<td>[56]</td>
</tr>
</tbody>
</table>

The aim of this study was to investigate the accuracy and stability of SPAD inversion of mangrove plant leaves using hyperspectral data to provide theoretical support for the monitoring of mangrove growth. In addition, the research in this study was mainly based on inversion studies conducted using machine learning models. Deep learning models can effectively deal with nonlinear relationships, do not require manual extraction of features, and can learn the best feature representation adaptively, thus improving the generalization ability of the model and having higher prediction advantages [57]. They are now also gradually being applied to the inversion studies of some plant physiological and biochemical indicators [58,59]. Furthermore, the UAV technology is becoming more mature and is a better choice for remote sensing data acquisition. Compared to other remote sensing platforms, UAVs have high temporal and spatial resolution imaging capabilities, which are cost-effective and highly adaptable [60]. Many researchers have confirmed that
combining UAV hyperspectral data with LiDAR data can effectively improve inversion accuracy [61,62], which will also be a new trend in studying SPAD inversion.

5. Conclusions

In this study, we used SPA to screen hyperspectral data for feature bands, built and validated four regression models based on the feature bands, and determined the SPAD values. The following conclusions were drawn.

1. The leaves of *B. sexangula* had lower SPAD measurements than those of *C. tagal* and *R. apiculata*, and plants with higher leaf SPAD measurements had less variation in SPAD measurements.
2. The accuracy and stability of the model established using first-order differentially transformed spectral data were slightly higher than those established using the original spectral data. First-order differential transformation of the original spectral data can improve the model accuracy.
3. The PLS model was found to be the most suitable model for the SPAD inversion of *B. sexangula* leaves, whereas the most suitable model for SPAD inversion of *C. tagal* and *R. apiculata* leaves was found to be the XGBoost model.
4. The XGBoost model had better adaptability to the leaf SPAD inversion of different species, and its stability is the best.

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

Funding: This research was funded by The Special Fund of Chinese Central Government for Basic Scientific Research Operations in Commonweal Research Institutes (CAFYBB2021MC006).

Data Availability Statement: All data included in this study are available upon request by contact with the corresponding author.

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

References


35. Yang, H.Y.; Du, J.M. Classification of desert steppe species based on unmanned aerial vehicle hyperspectral remote sensing and continuum removal vegetation indices. Optik 2021, 247, 167877. [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.