

Article

Comparative Analysis of Feature Importance Algorithms for Grassland Aboveground Biomass and Nutrient Prediction Using Hyperspectral Data

Yue Zhao ¹, Dawei Xu ¹, Shuzhen Li ¹, Kai Tang ², Hongliang Yu ¹, Ruirui Yan ¹ , Zhenwang Li ^{3,*} , Xu Wang ^{1,*} and Xiaoping Xin ¹

- ¹ State Key Laboratory of Efficient Utilization of Arid and Semi-Arid Arable Land in Northern China, Key Laboratory of Grassland Resource Monitoring Evaluation and Innovative Utilization, Ministry of Agriculture and Rural Affairs, Hulunber Grassland Ecosystem National Observation and Research Station, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China
 - ² State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
 - ³ Jiangsu Key Laboratory of Crop Genetics and Physiology/Jiangsu Key Laboratory of Crop Cultivation and Physiology, Agricultural College, Yangzhou University, Yangzhou 225009, China
- * Correspondence: zwli@yzu.edu.cn (Z.L.); wangxu01@caas.cn (X.W.)

Abstract: Estimating forage yield and nutrient composition using hyperspectral remote sensing is a major challenge. However, there is still a lack of comprehensive research on the optimal wavelength for the analysis of various nutrients in pasture. In this research, conducted in Hailar District, Hulunber City, Inner Mongolia Autonomous Region, China, 126 sets of hyperspectral data were collected, covering a spectral range of 350 to 1800 nanometers. The primary objective was to identify key spectral bands for estimating forage dry matter yield (DMY), nitrogen content (NC), neutral detergent fiber (NDF), and acid detergent fiber (ADF) using principal component analysis (PCA), random forests (RF), and SHapley Additive exPlanations (SHAP) analysis methods, and then the RF and Extra-Trees algorithm (ERT) model was used to predict aboveground biomass (AGB) and nutrient parameters using the optimized spectral bands and vegetation indices. Our approach effectively minimizes redundancy in hyperspectral data by selectively employing crucial spectral bands, thus improving the accuracy of forage nutrient estimation. PCA identified the most variable bands at 400 nm, 520–550 nm, 670–720 nm, and 930–950 nm, reflecting their general spectral significance rather than a link to specific forage nutrients. Further analysis using RF feature importance pinpointed influential bands, predominantly within 930–940 nm and 700–730 nm. SHAP analysis confirmed critical bands for DMY (965 nm, 712 nm, and 1652 nm), NC (1390 nm and 713 nm), ADF (1390 nm and 715–725 nm), and NDF (400 nm, 983 nm, 1350 nm, and 1800 nm). The fitting accuracy for ADF estimated using RF was lower ($R^2 = 0.58$), while the fitting accuracy for other indicators was higher ($R^2 \geq 0.59$). The performance and prediction accuracy of ERT ($R^2 = 0.63$) were noticeably superior to those of RF. In conclusion, our method effectively identifies influential bands, optimizing forage yield and quality estimation.

Keywords: remote sensing data analysis; forage quality assessment; spectral feature selection; agricultural resource management; environmental sensing technologies



Citation: Zhao, Y.; Xu, D.; Li, S.; Tang, K.; Yu, H.; Yan, R.; Li, Z.; Wang, X.; Xin, X. Comparative Analysis of Feature Importance Algorithms for Grassland Aboveground Biomass and Nutrient Prediction Using Hyperspectral Data. *Agriculture* **2024**, *14*, 389. <https://doi.org/10.3390/agriculture14030389>

Academic Editor: Fabio Sciubba

Received: 8 January 2024

Revised: 7 February 2024

Accepted: 26 February 2024

Published: 28 February 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

Grasslands, an integral ecosystem type, serve pivotal functions in climate regulation, biodiversity preservation, and water resource maintenance [1–3]. Globally, grassland ecosystems encompass approximately 52.5 million square kilometers, which constitutes around 26% of the Earth's total surface area [4]. In China specifically, grasslands extend

over an estimated 4 million square kilometers, accounting for nearly 41.7% of the nation's landmass [5]. In recent years, multiple factors have adversely impacted grassland ecosystems, particularly in China, where overgrazing, land-use shifts, and climate change have contributed to the deterioration of grassland resources and the attenuation of ecological functions [6–8]. In this context, identifying an efficacious approach to obtain precise information on grassland yield and quality is of paramount importance for monitoring the entire grassland ecosystem.

Dry matter yield (DMY), nitrogen content (NC), neutral detergent fiber (NDF), and acid detergent fiber (ADF) are important indicators to represent the forage's nutritional value, digestibility, and energy density. Numerous scholars employ various chemical analysis techniques to ascertain the constituents associated with forage yield and quality [9]. DMY is of vital importance for evaluating forage production potential and determining livestock farming scales [10]. NC constitutes one of the pivotal indicators in the appraisal of forage nutritional quality, as it delineates the available nitrogen element content within the forage [11]. ADF and NDF are integral parameters for examining forage fiber composition and digestibility [12]. ADF predominantly measures indigestible constituents in forage, such as cellulose, lignin, and silicates, whereas NDF quantifies total fiber content, encompassing cellulose, hemicellulose, and lignin. Ordinarily, ADF and NDF contents exhibit an inverse correlation with forage digestibility and energy density. Consequently, lower ADF and NDF values are indicative of superior forage nutritional value and digestibility, ultimately contributing to enhanced livestock production efficiency.

Multispectral remote sensing technology demonstrates potential in predicting forage yield and quality indicators. However, due to its specific and limited band configurations, its accuracy is often suboptimal [13]. In contrast, hyperspectral data, with its continuous spectral bands, offer superior capabilities in representing material components, showing significant advantages in estimating forage yield and quality-associated chemical constituents [14]. Yet, the strong intercorrelation among hyperspectral bands introduces redundant information, leading to challenges such as increased computational complexity and multicollinearity. Consequently, there is an urgent need to analyze the most significant bands for estimating different indicators, aiming to alleviate the redundancy in hyperspectral data and achieve accurate estimations using select bands.

Many studies have estimated the biomass and quality parameters of grasslands using hyperspectral data. For instance, Oliveira and Wijesingha, 2020, compared the performance of various machine learning models based on hyperspectral data for estimating grassland biomass [14,15]. Gitelson, 2003, employed a method that actively selected new training samples based on spectral diversity and prediction uncertainty to estimate DMY, nitrogen content, and nitrogen absorption [16]. Li et al., 2017, evaluated the random forests (RF) method for predicting grassland LAI using ground measurements and remote sensing data. Parameter optimization and variable reduction were conducted before model prediction. Two variable reduction methods were examined: the variable importance value method and the principal component analysis (PCA) method. Finally, the sensitivity of RF to highly correlated variables was tested [17]. However, previous studies have lacked research on estimating ADF and NDF. Additionally, few machine learning models used in estimation have analyzed the importance of features, and traditional importance analysis results cannot explain whether the relationship between features and estimation results is positive or negative. A substantial number of scholars employ principal component analysis (PCA) as a means to determine the significance of hyperspectral bands [18]. By reducing the dimensionality of high-dimensional spectral data and extracting principal components, PCA emphasizes crucial bands. Although this method aids in streamlining models, it is constrained by its dependence on linear assumptions, resulting in a limited capacity to analyze chemical constituents with weak linear relationships to reflectance. In the study by Nishikawa et al., 2023, it was found that during the process of estimating DMY and NC using visible and near-infrared (V-NIR) spectroscopy, PCA was not the optimal method for feature extraction due to its lack of interpretability [19]. Furthermore, it does not ex-

implicitly reflect the impact of individual features on target variables, leading to suboptimal interpretability. RF feature importance analysis proficiently identifies the significance of hyperspectral bands [20]. This machine learning algorithm is capable of managing complex non-linear relationships, thereby uncovering key bands. Numerous algorithms designed for the interpretation of machine learning models, such as the least absolute shrinkage and selection operator, assist in overcoming the black box effect during the estimation process [21]. However, these algorithms are hindered by a lack of local interpretability, rendering it challenging to provide comprehensive feature contribution explanations for individual samples. More critically, traditional methods encounter limitations in discerning the positive and negative contributions of features to prediction outcomes, complicating the optimization of models based on the results of interpretation algorithms. SHapley Additive exPlanations (SHAP) analysis is a game theory-based explanation framework that seeks to quantify the contributions of features in machine learning models [22]. By generating detailed explanations for individual samples, SHAP analysis enables a more profound exploration of the roles specific bands play in estimating chemical constituents. Zhong et al., 2021, explored the modeling potential of deep convolutional neural networks (DCNNs) for soil properties based on a large soil spectral library [23]. The SHAP method was used to interpret the outputs of the DCNN model (LucasResNet-16), demonstrating the potential of deep learning in utilizing hyperspectral soil data for modeling [23]. In this study, it is particularly significant that the approach can identify the positive and negative contributions of features to the prediction outcomes, thereby enabling a more comprehensive understanding of the relationship between spectral bands and chemical compositions.

The primary objective of this research is to develop a high-accuracy forage chemical composition prediction method based on a reduced number of spectral bands, thereby promoting the development of cost-effective multispectral sensing devices for the prediction of forage yield and quality, specifically, to explore and assess machine learning approaches employing hyperspectral remote sensing data for the estimation of forage DMY, NC, NDF, and ADF content, and to provide an interpretation of feature importance within the estimators and to identify the most indicative bands corresponding to the various chemical constituents of forage.

2. Materials and Methods

2.1. Study Site

The study site is situated in Hailar City, Hulunber, Inner Mongolia Autonomous Region (49°23'13" N, 120°02'47" E), at an average altitude of 631 m, and is characterized by a mid-temperate semi-arid continental climate. In 2022, the region experienced an average annual temperature of approximately 3 °C. The annual average precipitation was approximately 350–370 mm, predominantly occurring between July and September. The plant community within the experimental area primarily consists of *Leymus chinensis*, accompanied by prevalent species such as *Vicia amoena*, *Thalictrum squarrosum*, and *Pulsatilla turczaninowii*. The soil type is classified as dark chestnut calcareous soil. Starting from 2013, long-term annual mowing treatments have been carried out regularly, and the area has been encircled and safeguarded. A gradient treatment of nitrogen fertilizer concentrations was employed in the area. Each plot encompasses an area of 60 m², with a 2 m buffer strip established between subplots. In total, seven different treatments were applied to each group. Fertilization treatments were implemented using two approaches: chemical and organic fertilizers. Since 2014, our study has conducted fertilization treatments on experimental fields at the beginning of June each year, employing a broadcasting method for granular fertilizers. Taking into consideration the average application rate of organic fertilizers in local farmlands, findings from relevant literature, and baseline soil survey data, our experiment established 4 distinct fertilization levels based on nitrogen content. Urea was used as the chemical fertilizer, with application rates set at 0 kg/ha², 75 kg/ha², 150 kg/ha², and 225 kg/ha². The organic fertilizer, composed of granular N + P₂O₅ + K₂O (dry basis) ≥ 7% and organic matter ≥ 45%, had effective nutrient application rates of

0 kg/ha², 63 kg/ha², 127 kg/ha², and 190 kg/ha². The NC remained consistent between the two fertilization techniques at each level (Figure 1).

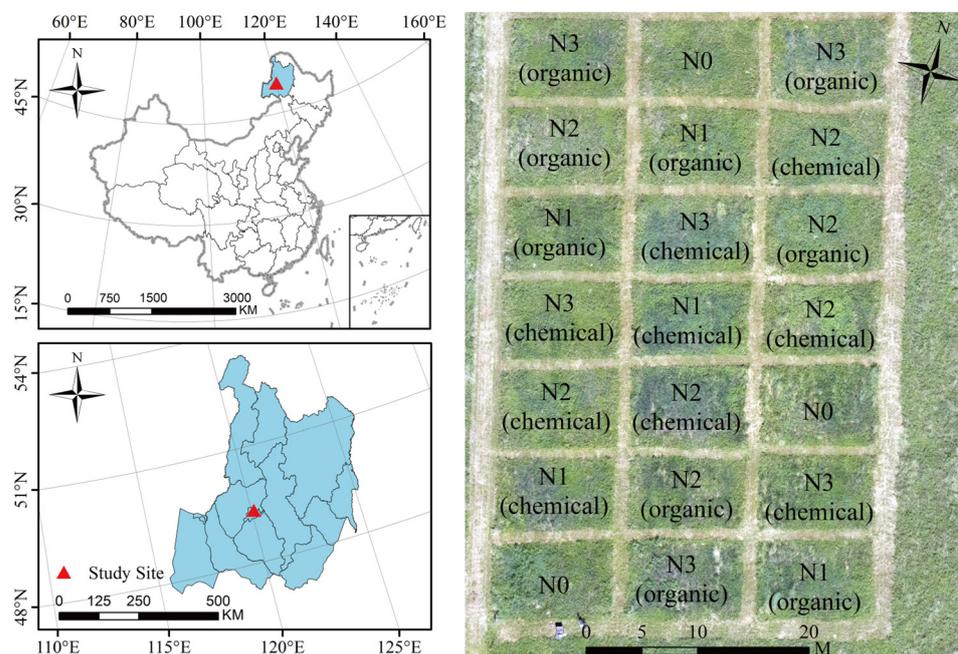


Figure 1. The location of the Hulunber research area in the Inner Mongolia Autonomous Region. The orthophoto for the experimental area with a gradient of nitrogen fertilizer applications is displayed on the right. “N0” represents the untreated grass, while “N3” represents the grass treated with the highest concentration of nitrogen fertilizer. “Chemical” and “organic” indicate that the nitrogen fertilizer types are chemical fertilizer and organic fertilizer, respectively.

2.2. Ground Measurements

Field investigations were conducted during the peak growing season in July 2022. Six evenly distributed 1 m × 1 m quadrats were established within each plot, leading to a total of 126 quadrats. All plants within each quadrat were harvested, and forage samples were subjected to an initial drying process at 105 °C for 30 min. Subsequently, the temperature was reduced to 60 °C and maintained until the samples were completely dry, yielding the forage’s DM. The samples were then ground, and their nutritional chemical components were analyzed. Plant NC was determined using the Kjeldahl method after sulfuric acid and hydrogen peroxide digestion, while the van Soest method was employed to measure plant NDF and ADF content [24,25].

2.3. Hyperspectral Data Sets

In this study, hyperspectral remote sensing technology was used for predicting the nutritional chemical components of forage. To gather spectral data, we used the RS8800 hyperspectral instrument (Spectral Evolution, Haverhill, MA, USA). The RS8800 has a broad wavelength range (350–2500 nm), with a spectral resolution of 3 nm between 350 and 1000 nm and 10 nm between 1001 and 2500 nm. Spectral sampling intervals are 1.5 nm in the 350–1000 nm range and 2 nm in the 1001–2500 nm range. The instrument features a full field of view angle of 28° and a spectral response time of 10 milliseconds, making it suitable for studying various vegetation types. During the in situ measurements, the RS8800 hyperspectral instrument was positioned one meter above the ground to ensure the stability and accuracy of the spectral data. To circumvent issues arising from low solar angles, measurements were conducted between 10 a.m. and 2 p.m. The operator faced the sun to avoid any obstructions. Throughout the measurement process, we adhered to the recommendations and guidelines specified in the instrument’s user manual, calibrated the device using a whiteboard, and maintained stable environmental conditions.

Moreover, we collected spectral data from six sample points within each plot to increase data representativeness.

We employed the isolation forest algorithm on the hyperspectral data to remove outliers, thereby enhancing the accuracy of forage nutrient component predictions. This approach mitigates the impact of noise on the RF regression model, consequently improving the prediction accuracy for DMY, NC, ADF, and NDF.

2.4. Experiments Process

This study employs a series of analytical techniques to evaluate and estimate the nutritional value of forage samples. Initially, forage samples were collected, and comprehensive spectral information was obtained through hyperspectral data. The isolation forest algorithm was used to detect outliers in the hyperspectral data, ensuring the quality and accuracy of the data. Subsequently, reflectance data were utilized to predict the samples' DMY, NC, ADF, and NDF. To process and analyze the high-dimensional hyperspectral data, three different methods were adopted for dimensionality reduction and for extracting the most informative features, with a comparative analysis conducted. On this basis, the study not only used the RF algorithm but also employed the ERT algorithm for predictive modeling, thereby ensuring the accuracy of the results. By comparing these two models, the study aims to identify the most suitable estimation method for hyperspectral data in the context of forage nutritional value assessment (Figure 2).

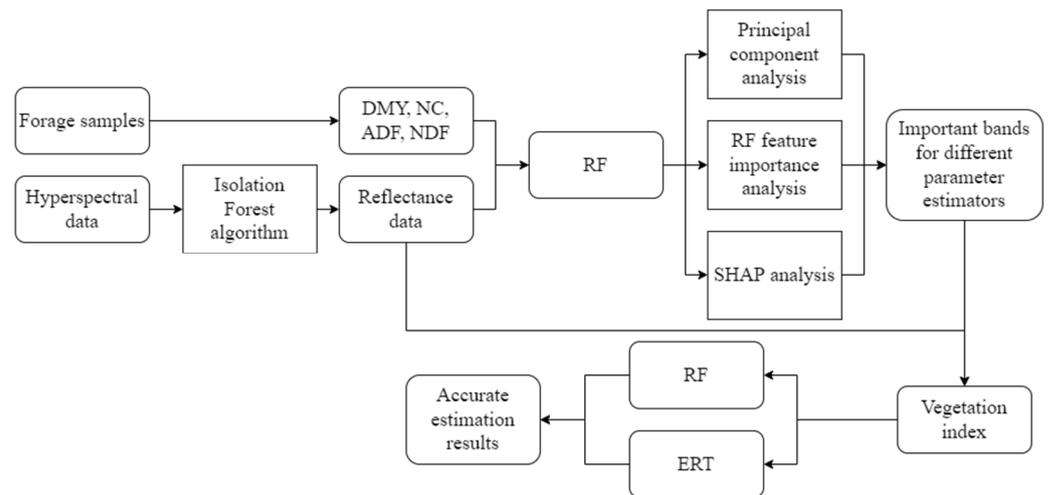


Figure 2. Flowchart for comparing algorithms of feature importance in estimating nutritional value of forage.

2.5. Band Importance Analysis Algorithm

Band importance analysis was conducted using three methods: PCA, RF’s feature importance analysis, and SHAP. The PCA analysis allowed us to identify the most important features in our data set, while the RF modeling provided us with a way to measure the relative importance of each feature. We used the SHAP analysis to identify the most significant contributions of each feature to the prediction results, allowing us to gain deeper insights into the complex relationships between the spectral features and the grassland nutrient properties.

2.5.1. Principal Component Analysis

In our application of principal component analysis (PCA), we focused on x-loadings to determine the wavelengths showing the highest variance in the dataset. These x-loadings, indicative of each wavelength’s contribution to the principal components, highlighted bands representing the most significant general spectral variability. This method allowed

us to identify those bands that are most pronounced across all spectra, signifying the greatest variance, rather than correlating them with specific forage nutrient characteristics.

2.5.2. RF Feature Importance Analysis

In our study, we utilized the random forest (RF) algorithm’s inherent capability to calculate the importance of each wavelength (feature). This is based on how much each feature contributes to the decision-making process in the model. We then normalized these importance scores to a scale of 0 to 1, allowing for a straightforward comparison of each wavelength’s relevance in predicting forage chemical parameters.

2.5.3. SHAP Analysis

SHAP (SHapley Additive exPlanations) analysis, a cutting-edge interpretability technique, was employed to assess each feature’s (wavelength’s) individual contribution to our model’s predictions. Unlike the broad-spectrum approach of PCA and the information gain assessment in RF, SHAP provides a detailed, feature-specific analysis. It allows us to distinguish both the positive and negative impacts of each wavelength, offering a comprehensive understanding of their roles in predicting forage chemical parameters. This nuanced analysis, grounded in cooperative game theory, enhances our model’s transparency and elucidates the “black-box” nature of complex algorithms, setting it apart from conventional methods. The method is based on the SHAP values from game theory, which are used to quantify the contribution of each feature to the model’s prediction outcome. For a given feature i , its SHAP value can be calculated using the following formula:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)] \tag{1}$$

Herein, N is the feature set, S is an arbitrary subset of features that do not contain feature i , and $v(S)$ is the contribution value of feature set S to the prediction result, which is usually predicted by the model. $|S|$ is the number of elements in set S . ϕ_i is the SHAP value of feature i .

2.6. Vegetation Indices

In order to reduce data redundancy, we calculated the corresponding vegetation index from the selected optimal band to predict the parameters. Five vegetation indices (VIs) were selected as features for the RF and ET model (Table 1). These include the Normalized Difference Vegetation Index (NDVI), the Enhanced Vegetation Index (EVI), the Normalized Difference Red Edge Index (NDRE), the Normalized Difference Water Index (NDWI), and the Blue-light Normalized Difference Vegetation Index (BNDVI). These vegetation indices encompass spectral information from multiple bands, such as blue, red, red-edge, near-infrared, and shortwave-infrared.

Table 1. Vegetation indices for yield and quality estimation used in this study.

Index Acronym	Formula	References
NDVI	$\frac{NIR - RED}{NIR + RED}$	[26]
EVI	$G \times \frac{NIR - RED}{NIR + C_1 \times RED - C_2 \times BLUE + L}$	[27]
NDRE	$\frac{NIR - REEDGE}{NIR + REEDGE}$	[28]
NDWI	$\frac{NIR - SWIR}{NIR + SWIR}$	[29]
BNDVI	$\frac{NIR - BLUE}{NIR + BLUE}$	[30]

2.7. Estimation Process and Evaluation Metric

The estimation process used in this study is based on the methods proposed by Breiman, 2001, and Geurts et al., 2006, and random forest (RF) and extreme randomized trees (ERT) algorithms were used as estimators and the results were compared [31,32].

The random forest (RF) algorithm is widely adopted for regression tasks. This technique enhances the model's generalization abilities by incorporating multiple decision tree models and implementing bootstrap sampling along with random feature selection strategies [33]. The primary advantages of the RF algorithm lie in its capacity to handle various data types, its resilience against overfitting, and its high predictive accuracy and robustness [34]. In this study, the Python 3.10 environment was employed, utilizing the scikit-learn package to implement the RF regression algorithm.

The Extra-Trees algorithm (ERT) was utilized for estimating the yield and quality indices of forage. This ensemble method constructs multiple decision trees with random splits on features, enhancing predictive accuracy for multi-dimensional agricultural data [35]. Extra-Trees introduces greater randomness in the selection of split nodes, potentially reducing computational time and improving model performance in certain scenarios.

Initially, vegetation indices containing important bands were used as input features, and grassland DMY, NC, ADF, and NDF were designated as target variables. Subsequently, we applied K-fold cross-validation (K = 5) to dividing the dataset into training and testing sets. In this method, the entire dataset is randomly divided into five equal parts. In each iteration, one of the parts is used as the test set, while the remaining four parts are combined into the training set. This process is repeated five times, with each part having one opportunity to serve as the test set. Cross-validation enables more reliable estimation of the model's generalization capability, reduces the risk of overfitting, and obviates the need for an additional validation set, thereby streamlining the model training and evaluation process.

The performance of an RF and ERT regression model for predicting four plant nutrient indicators was evaluated using three evaluation metrics: R^2 , mean absolute error (MAE), and root-mean-square error (RMSE).

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (2)$$

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

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (5)$$

Herein, y_i represents the observed value, \hat{y}_i denotes the predicted value, \bar{y} signifies the mean of the observed values, and n refers to the number of samples. Smaller values of MAE and RMSE indicate smaller errors. R^2 expresses the degree of fit in the regression analysis.

3. Results

3.1. The Grass Yield and Quality

In the analysis of grassland yield and quality parameters, the DMY exhibited a range from 302.83 g/m² to 773.14 g/m², with an average value of 485.54 g/m² and a standard deviation of 107.12 g/m². The NC varied between 1.45 g/100 g and 2.60 g/100 g, averaging at 2.05 g/100 g with a standard deviation of 0.26 g/100 g. The ADF percentages spanned from 54.51% to 68.97%, with a mean value of 61.88% and a standard deviation of 3.27%. Lastly, the NDF percentages ranged from 30.71% to 38.16%, with an average of 35.11% and a standard deviation of 1.29%. These statistics reflect the inherent variability and central tendencies observed in the field measurements of grassland yield and quality parameters. The results of in situ measurements are given in Table 2.

Table 2. Forage sample measurements: Min: minimum; Max: maximum; Mean: average; Std: standard deviation of the attribute; DMY: dry matter yield (g/m²); NC: nitrogen content (g/100 g); ADF: acid detergent fiber (%); NDF: neutral detergent fiber (%). N0 represents the untreated grass, while N3 represents the grass treated with the highest concentration of nitrogen fertilizer. All represents the data of all samples.

	DMY (g/m ²)				NC (g/100 g)			
	Min	Max	Mean	Std	Min	Max	Mean	Std
All	302.83	773.14	493.87	107.12	1.45	2.60	2.04	0.26
N0	302.83	542.52	404.89	60.33	1.45	2.34	1.88	0.18
N1	304.26	708.28	484.54	125.54	1.51	2.49	2.06	0.24
N2	345.57	755.58	521.77	83.27	1.53	2.60	2.00	0.32
N3	415.58	773.14	571.12	96.26	1.92	2.53	2.24	0.17

	ADF (%)				NDF (%)			
	Min	Max	Mean	Std	Min	Max	Mean	Std
All	54.51	68.97	61.99	3.27	30.71	38.16	34.90	1.29
N0	54.51	65.03	59.59	2.61	30.71	36.95	34.45	1.29
N1	55.22	67.97	61.43	2.95	32.52	37.64	34.90	1.43
N2	57.06	68.63	63.11	2.93	31.56	38.16	35.02	1.26
N3	58.26	68.97	64.29	2.82	31.94	36.63	35.23	1.05

The interrelationship among four key grassland parameters was elucidated through a Pearson correlation matrix. It is evident that DMY and NC exhibit a strong positive correlation ($r = 0.61, n = 126, p < 0.05$), suggesting that an increase in DMY might be associated with heightened NC. This can be attributed to the fact that higher NC often leads to enhanced plant growth, contributing to increased biomass. On the contrary, the ADF and NC show a moderate negative correlation ($r = -0.38, n = 126, p < 0.05$). This inverse relationship implies that as the ADF content in the grassland increases, there tends to be a decrease in its NC. ADF, being an indigestible part of plant material, might deter the accumulation of nitrogen-rich compounds. Additionally, the slight negative correlation between DMY and NDF ($r = -0.25, n = 126, p < 0.05$) indicates that an increase in DMY might slightly reduce the NDF content, although this relationship is not as strong as the aforementioned pairs. The Pearson correlation coefficients of yield and quality parameters of forage are displayed in Figure 3.

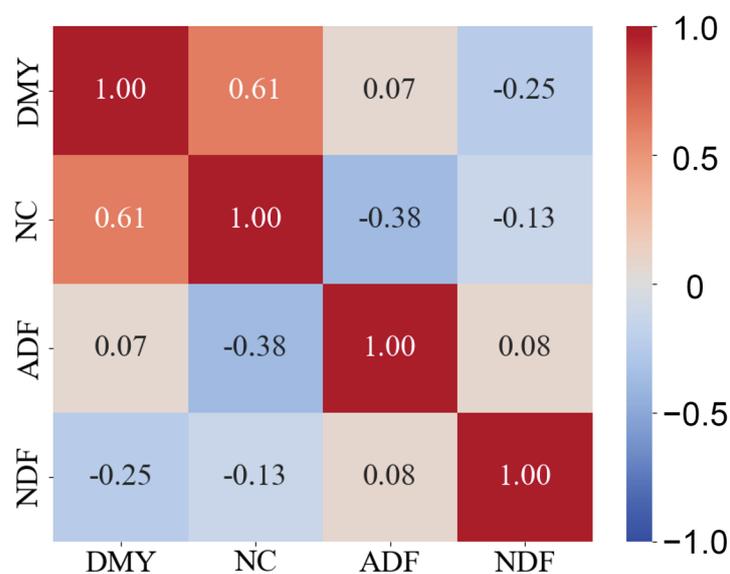


Figure 3. Pearson correlation analysis of yield and quality parameters of forage.

3.2. Hyperspectral Band Analysis

All hyperspectral measurement data after removing outliers are displayed in Figure 4. “N0” to “N3” represent different treatments of grass with varying concentrations of nitrogen fertilizer.

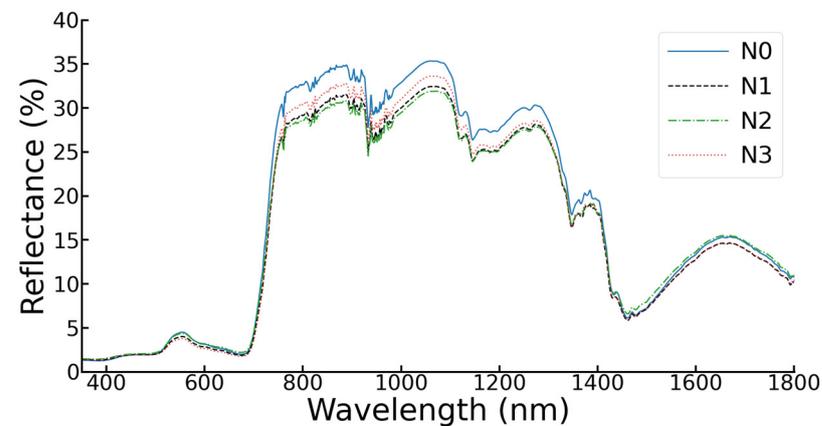


Figure 4. After removing outliers from the hyperspectral data using the isolation forest algorithm, the wavelength is plotted on the x -axis and reflectance on the y -axis.

Correlation analysis was conducted between the reflectance values across all wavelengths and the four indices: DMY, NC, NDF, and ADF, as illustrated in Figure 5. The results indicate that for DMY, NC, and NDF, there is a pronounced positive correlation in the 500–800 nm range and a negative correlation around 1300 nm. Among them, the correlation between NC and reflectance is superior to the others, followed by DMY, and then NDF. In contrast, ADF did not exhibit a significant correlation with reflectance as observed with the other three indices.

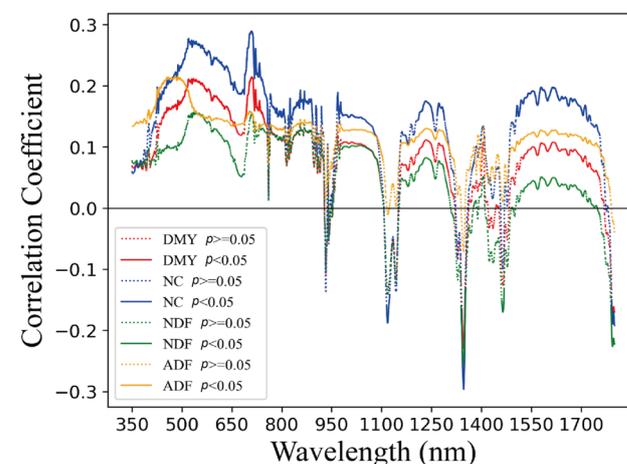


Figure 5. The reflectance of all wavelengths in hyperspectral data shows correlation coefficients with DMY, NC, ADF, and NDF, where dashed lines indicate non-significance ($p \geq 0.05$) and solid lines indicate significance ($p < 0.05$).

3.3. Feature Importance Analysis

3.3.1. The Feature Importance Based on PCA

The cumulative variance ratios plot was utilized to determine the optimal number of principal components to be retained in the analysis, and the “elbow” method was employed to identify the balance point between the explained variance and the number of principal components.

As shown in Figure 6, the cumulative variance ratios increased rapidly with the increase in the number of principal components and then approached a plateau after a

certain number of components. The “elbow” point, which represents the optimal number of principal components, was determined to be 4. This finding indicates that the first 4 principal components contain the majority of the information in the original data and are sufficient to explain the variation in the grassland nutritional components.

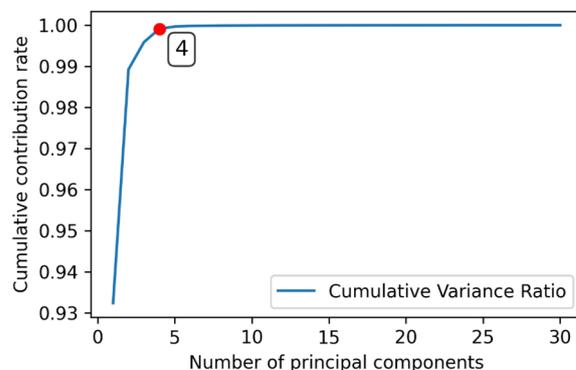


Figure 6. Principal component analysis (PCA) of hyperspectral data for predicting grassland nutrient concentrations. The plot shows the cumulative variance ratios explained by each principal component (PC) as a function of the number of PCs included in the analysis. The elbow point is marked by a red dot at PC4. The X-axis represents the number of PCs, and the Y-axis represents the cumulative variance ratio.

Figure 7 illustrates the top four X-loadings ranked principal components (PCs) obtained from PCA analysis. The X-loading peaks are primarily concentrated in the 400 nm, 520–550 nm, 670–720 nm, and 930–950 nm intervals, corresponding to the blue, green, red edge, and near-infrared (NIR) bands, respectively, which are known to capture rich information on vegetation nutritional content. Of particular interest is PC4, which exhibits a higher contribution to the PCs in the 400 nm and 520–550 nm range. In the red edge band (670–720 nm), PC2, PC3, and PC4 all exhibit some peaks. The 930–950 nm range is located in the NIR region of the spectrum. PC4 shows some peaks in this range, as do several other PCs.

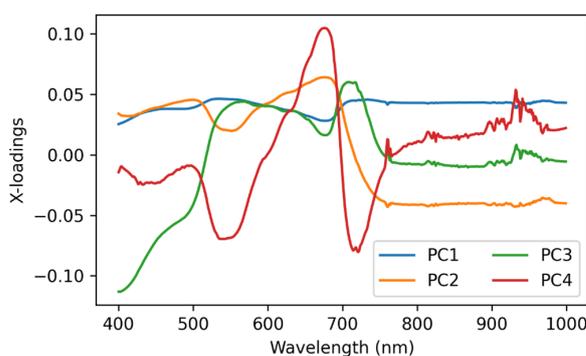


Figure 7. X-loadings analysis of principal component analysis (PCA). The x-axis represents wavelength (nm), and the y-axis represents the x-loadings values of the principal components.

3.3.2. The Feature Importance Based on RF

An RF regression model was used to identify the important wavelengths for predicting grassland nutrient content through evaluation and ranking of feature importance. Figure 8 shows the evaluation and ranking of the top 10 important wavelengths for four nutrient evaluation indices (DYM, NC, ADF, and NDF) using hyperspectral reflectance as features in the RF regression model. To facilitate the comparison of the differences in the importance of each wavelength in the model, we standardized the importance values. The results showed that 930–940 nm and 700–730 nm were important features in the model for analyzing the four nutrient components.

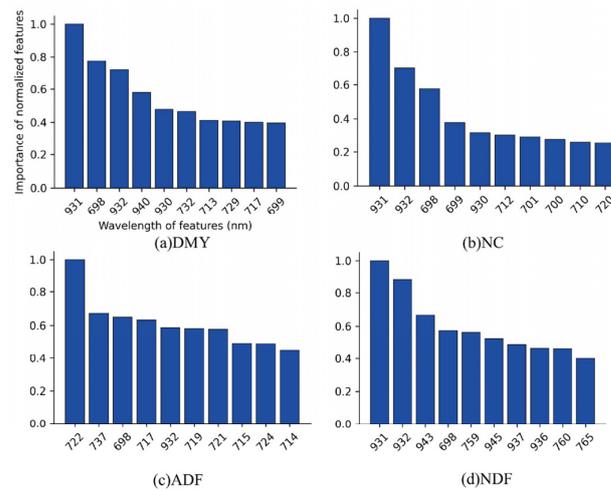


Figure 8. Feature importance of the random forests (RF) model for estimating forage yield and nutritional indices. The *x*-axis represents feature wavelengths in nanometers (nm), and the *y*-axis denotes the importance of each feature: (a) DMY, (b) NC, (c) ADF, and (d) NDF.

3.3.3. The Feature Importance Based on SHAP

SHAP analysis was employed to explain the predictions of the ERT model. The top 20 features in terms of contribution values for each model prediction were calculated and their results for the four nutrient components and yield are presented separately, as shown in Figure 9.

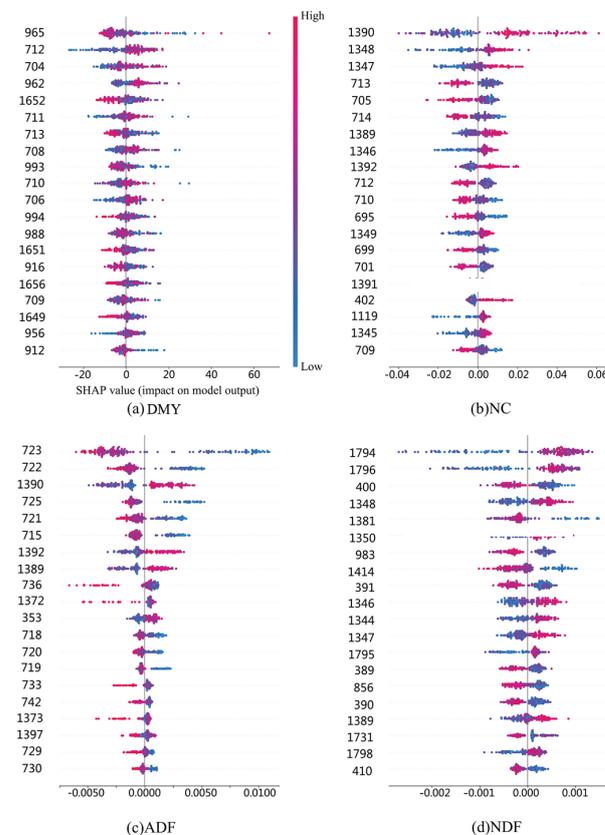


Figure 9. SHapley Additive exPlanations (SHAP) analysis of the Extra-Trees algorithm (ERT) model employed for estimating forage yield and nutritional indices. The *x*-axis denotes

SHAP values, while the y -axis represents feature wavelengths in nanometers (nm). The color scheme corresponds to the reflectance values associated with each feature: DMY, NC, ADF, and NDF.

In evaluating the DMY estimator, the wavelengths 965 nm, 712 nm, and 1652 nm, spanning the near-infrared, red-edge, and shortwave-infrared bands, exhibit pronounced reflectance contributions. For the NC estimator, significant reflectance is observed at 1390 nm in the shortwave-infrared and 713 nm in the red-edge domain. The SHAP analysis for ADF indicates subdued correlations across most bands, with a notable negative correlation between the 1390 nm, 715–725 nm range, and ADF content, especially when juxtaposed with other forage nutritional metrics. Regarding the NDF estimator, the key wavelengths are 400 nm, 983 nm, 1350 nm, and 1800 nm. While 400 nm and 983 nm manifest a negative correlation, the 1350 nm and 1800 nm wavelengths in the shortwave-infrared correlate positively with NDF.

3.4. Estimation of Grassland Yield and Quality

The reflectance of significant bands and computed vegetation indices were used as inputs for both RF and ERT regression analyses. Table 3 shows that for RF, the R^2 values for all plant nutritional indicators are above 0.57, with the highest for DMY (0.80). DMY has the highest MAE (49.78) and RMSE (57.17). For ERT, the model performance improved, with R^2 values increasing notably for NC (0.67) and NDF (0.72). ERT also showed lower MAE and RMSE across all indicators, suggesting enhanced prediction accuracy compared to RF, particularly in estimating NC and NDF.

Table 3. The performance evaluation metrics of both the RF and ERT regression models for four plant nutrient indicators.

		DMY (g/m ²)	NC (g/100 g)	ADF (%)	NDF (%)
RF	R^2	0.80	0.57	0.58	0.62
	MAE	49.78	0.14	0.78	1.84
	RMSE	57.17	0.18	1.01	2.17
ERT	R^2	0.82	0.67	0.63	0.72
	MAE	44.16	0.12	0.82	1.47
	RMSE	53.05	0.16	1.02	1.85

Figure 10 shows the predicted values and measured values of each plant nutrient index for the best-performing runs of the 5-fold cross-validation. Overall, the predicted values and measured values are in good agreement, with a relatively concentrated distribution near the 1:1 line. Consistent with the results presented in Table 3, most of the model's predictions are reliable. However, for samples with relatively small or large measured values, the RF model tends to produce higher or lower predicted values, respectively.

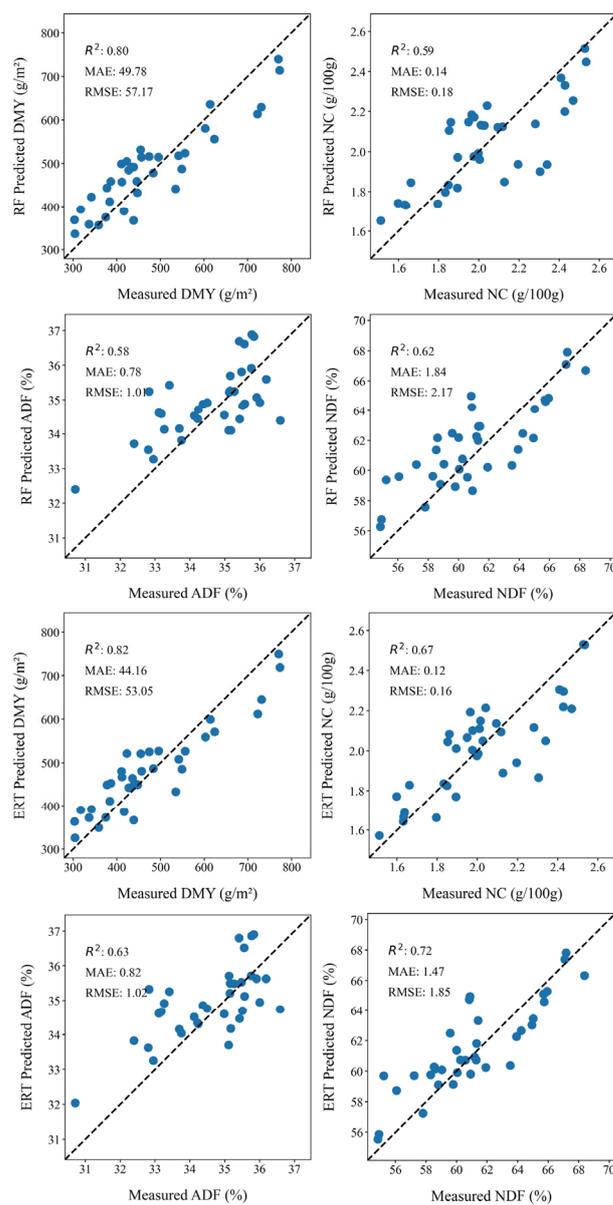


Figure 10. Scatterplots of RF and ERT predictions versus observed values for four nutrient parameters: DMY, NC, ADF, and NDF. The x-axis represents the observed values, while the y-axis represents the RF or ERT predictions. The black dashed line in each scatterplot represents the line of perfect prediction.

4. Discussion

In PCA, the significance of the spectra for all indicators as a whole is reflected through the X-loadings values of the principal components. Four principal components highlighted the significance of multiple wavelengths. Specifically, the 400 nm and 520–550 nm wavelengths fall within the blue and green light bands, respectively. This range is pivotal for estimating crude protein content [36] and is closely associated with chlorophyll and protein concentrations. An increase in chlorophyll content results in a decrease in reflectance in the visible light region, especially in the blue and green areas of the spectrum [37]. The red edge band (670–720 nm), which is associated with chlorophyll and pigment content such as chlorophyll a and b, plays a crucial role in vegetation analysis. Several studies have demonstrated the importance of the red edge band for estimating canopy chlorophyll content [38,39]. Vegetation indices constructed using the red edge band show greater potential for predicting chlorophyll content than traditional indices, such as NDRE [28]. In the 930–950 nm range, vegetation indices derived from this range of reflectance values

have been found to provide robust estimates of lipid content and aboveground biomass in grasslands [40].

The RF feature importance analysis can provide band importance results for individual indicators. Chen et al., 2018, present a method for automatically determining the optimal feature subset using RF-RFE, validated across two completely distinct molecular biology datasets [41]. Additionally, the SHAP analysis offers insights into the positive and negative correlations between various wavelengths and indicators. The results show that 965 nm, 712 nm, and 1652 nm are the most important bands for the DYM estimator. It is common for the reflectance in proximity to these bands to be linked with the photosynthetic activity in forage grasses, specifically the chlorophyll content [42]. The reflectance at 700 nm is associated with the leaf area index (LAI), biomass, and internal leaf structure [10], while the reflectance around 930 nm is related to the moisture content of the biomass [43,44]. As a rule, enhanced photosynthetic activity correlates with increased reflectance within these wavelength ranges. When encountering low chlorophyll content, substantial fluctuations are observed in dry matter content, whereas, with elevated chlorophyll content, a positive correlation manifests between dry matter content and chlorophyll content. This phenomenon can be explicated through the impact of nitrogen fertilization on forage growth and photosynthesis. Adequate nitrogen fertilization facilitates plant growth, augments chlorophyll content, and elevates photosynthetic efficiency. Nonetheless, both excessive and insufficient applications of nitrogen fertilizer may culminate in hindered plant growth. In scenarios with diminished chlorophyll content, one plausible explanation is a deficit of nitrogen fertilizer. Nitrogen constraint in plants results in reduced chlorophyll content and biomass. Under such circumstances, a positive correlation between chlorophyll content and DYM may arise, with lower chlorophyll content corresponding to lower DYM, exemplified by the distribution of blue points on the left side of Figure 10. Alternatively, an overabundance of nitrogen fertilizer may be the cause. Excessive uptake of nitrogen by plants could engender growth inhibition or other physiological perturbations, such as nitrogen toxicity [45,46]. In this situation, chlorophyll content may be diminished, while the high or low dry matter content is contingent upon the plant's response to nitrogen toxicity. Consequently, the relationship between chlorophyll content and dry matter content may be ambiguous in this context. Conversely, in instances of elevated chlorophyll content, plants might be experiencing an optimal supply of nitrogen fertilizer, leading to high chlorophyll content and photosynthetic efficiency, ultimately resulting in increased dry matter content. Under these conditions, a positive correlation between chlorophyll content and dry matter content is probable.

In the NC estimator, the most important bands are concentrated around 1390 nm and 713 nm. When juxtaposed with the DYM model's findings, the NC estimator accentuates the impact of the shortwave-infrared band on the prognostic outcomes to a greater extent. Primarily, the positive association between the shortwave-infrared band reflectance and NC is corroborated across all samples in the 1390 nm SHAP analysis, whereby higher red-edge band reflectance typically corresponds to elevated nitrogen and chlorophyll levels [47]. Furthermore, although the near-infrared band exhibits a substantial advantage in terms of model contribution, its performance in correlation is not particularly prominent. The distribution of SHAP values across varying reflectance values appears rather dispersed, with a clear demarcation in the red-edge band samples surrounding a SHAP value of 0. This phenomenon may be ascribed to the extensive range of nitrogen application rates since, in addition to NC sensitivity, the near-infrared band is also highly responsive to biomass, as corroborated by a plethora of research [48–50]. Elevated nitrogen application rates bolster forage nitrogen concentration; however, they do not perpetuate the augmentation of biomass. Consequently, model uncertainty is manifested in the near-infrared band reflectance. Contrasting the prominent contribution displayed by the near-infrared band, the shortwave-infrared band seemingly offers a superior representation of the correlation with NC, thus serving as a more robust spectral band for the estimation of forage NC.

In the ADF estimator, the results show that the reflectance around 1390 nm and 715 nm–725 nm is the most important spectral information, and there is a significant negative correlation between them. On the one hand, the absorption capacity of chlorophyll within this spectral domain is notably strong, and on the other hand, cellulose and hemicellulose in forage grass also exert some influence in this region [51]. As a result, with the increasing ADF content in the samples, it is plausible that the chlorophyll content experiences a concomitant reduction [15]. Subsequently, the reflectance values in the vicinity of the 715–725 nm band demonstrate an increment, exhibiting an inverse correlation. However, the reflectance within this band also pertains to the water content and structure of the leaves, wherein forage leaves with elevated fiber content are likely to possess more substantial or compact structures, consequently influencing the spectral reflectance properties within the 1390 nm [52]. From this inference, it can be deduced that leaves with higher water content and fiber content may exhibit elevated reflectance values in this particular wavelength band, and the reflectance within this band can serve as a significant feature in the ADF estimator.

Through the NDF estimator analysis results, several significant wavelengths can be identified. Notably, in the NDF estimator, the importance of reflectance in the shortwave-infrared band is particularly prominent. Samples from this band in the SHAP analysis are distinctly distributed on both sides of a SHAP value of zero. This phenomenon might be attributed to the absorption effect of moisture content in the forage at this specific wavelength. For forage, spectral bands near the shortwave-infrared are primarily associated with water content [53], possessing strong absorption characteristics. This is due to the spectral properties of water at wavelengths 1350 nm and 1800 nm, which are mainly caused by the vibrations of hydrogen bonds present in water [54]. NDF primarily consists of lignin, cellulose, and hemicellulose, which are the main structural components of plant cell walls. The structure of the plant cell wall directly impacts the water content of plant tissues. For instance, a higher NDF content in plant tissues might imply thicker plant cell walls, potentially leading to an increase in the moisture content of plant tissues. Consequently, this results in a pronounced positive correlation for 1350 nm and 1800 nm in the NDF estimation model. Simultaneously, in the RF feature importance analysis, the spectral region within the 930–940 nm range is also considered significant. This range, situated within the near-infrared spectral domain, is closely associated with leaf structure and pigments [55]. The absorption characteristics of C–H, C–N, N–H, and O–H bonds present in cellulose, protein, and starch are manifested within this range [56].

In the inversion results of machine learning models, due to the uneven distribution of data and the presence of extremes caused by nitrogen gradients, there is a bias in the predicted values for samples with relatively large or small measurements. This effect is particularly evident in the RF model [57]. Specifically, the nitrogen concentration gradient caused a large overall range of values for the plant nutrient indices in the comparison treatment and introduced extreme values that deviated from the central tendency. Similar situations have been observed in other studies. Dhal et al., 2022, optimized nutrient levels for plant growth in hydroponic irrigation systems, to overcome the impact of excessively high data dimensions in the sample, machine learning techniques specifically designed for small datasets were adopted [58]. This included using feature selection methods to train the three most important features identified in the results and then deploying them to a machine learning model on an Android application [58]. As a result, the RF and ERT algorithm, which relies on decision trees, may have overemphasized the contribution of these extreme values in constructing the trees, which in turn skewed the predictions for samples with smaller and larger observed values [30,59]. Additionally, the model's sensitivity to these extreme values may have led to overfitting [31], which further contributed to the bias in the predicted values. Therefore, when applying regression analysis based on decision tree structure models in control experiments, considering the data distribution and the presence of extremes is crucial.

5. Conclusions

In summary, this study successfully identified key spectral bands within hyperspectral data, enhancing the precision of above-ground biomass and nutrient estimation in grasslands. Specifically, for the DMY estimator, the bands at 965 nm, 712 nm, and 1652 nm were deemed most significant; for the NC estimator, the reflectance at 1390 nm and 713 nm bands held the highest importance; for the ADF estimator, the bands around 1390 nm and within the range of 715 nm to 725 nm were of higher importance; and for the NDF estimator, the reflectance at 400 nm, 983 nm, 1350 nm, and 1800 nm bands was the most critical features for the model.

By employing advanced machine learning techniques, including RF and ERT regression, we demonstrated the potential of these methods in agricultural remote sensing. The models exhibited a high degree of fit for estimating grassland biomass and nutrient levels ($0.57 < R^2 < 0.82$). These findings contribute to more effective grassland quality assessment and can inform the development of cost-effective multispectral sensors, underscoring the significance of targeted spectral analysis in precision agriculture.

Author Contributions: Conceptualization, Y.Z., X.W. and X.X.; methodology, Y.Z. and Z.L.; software, Y.Z., Z.L. and D.X.; validation, Y.Z., S.L. and K.T.; formal analysis, Y.Z. and Z.L.; investigation, Y.Z. and K.T.; resources, X.W., R.Y. and X.X.; data curation, Y.Z. and H.Y.; writing—original draft preparation, Y.Z. and D.X.; writing—review and editing, Y.Z., Z.L. and K.T.; visualization, Y.Z.; supervision, X.W. and X.X.; project administration, X.W. and X.X. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Key Research and Development Program of China (2021YFD1300502), the National Natural Science Foundation of China (42301365), the Natural Science Foundation of Jiangsu Province (No. BK20230566) and the Lv Yang Jin Feng Plan of Yangzhou city (YZLYJFJH2022YXBS137).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We are grateful to Jiangwen Wang and Chunling Zhao with the Hulunber Grassland Ecosystem Observation and Research Station, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences (CAAS) for their work during the ground data collection.

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

References

1. Tilman, D.; Downing, J.A. Biodiversity and Stability in Grasslands. *Nature* **1994**, *367*, 363–365. [[CrossRef](#)]
2. Liu, W.; Zhang, Z.; Wan, S. Predominant Role of Water in Regulating Soil and Microbial Respiration and Their Responses to Climate Change in a Semiarid Grassland. *Glob. Chang. Biol.* **2009**, *15*, 184–195. [[CrossRef](#)]
3. Hector, A.; Dobson, K.; Minns, A.; Bazeley-White, E.; Hartley Lawton, J. Community Diversity and Invasion Resistance: An Experimental Test in a Grassland Ecosystem and a Review of Comparable Studies: Community Diversity and Invasion. *Ecol. Res.* **2001**, *16*, 819–831. [[CrossRef](#)]
4. Török, P.; Brudvig, L.A.; Kollmann, J.; Price, J.N.; Tóthmérész, B. The Present and Future of Grassland Restoration. *Restor. Ecol.* **2021**, *29*, e13378. [[CrossRef](#)]
5. Liu, Y.; Wang, Q.; Zhang, Z.; Tong, L.; Wang, Z.; Li, J. Grassland Dynamics in Responses to Climate Variation and Human Activities in China from 2000 to 2013. *Sci. Total Environ.* **2019**, *690*, 27–39. [[CrossRef](#)] [[PubMed](#)]
6. Conant, R.T.; Paustian, K. Potential Soil Carbon Sequestration in Overgrazed Grassland Ecosystems: Potential C Sequestration in Overgrazed Grasslands. *Glob. Biogeochem. Cycles* **2002**, *16*, 90–1–90–9. [[CrossRef](#)]
7. Qi, A.; Holland, R.A.; Taylor, G.; Richter, G.M. Grassland Futures in Great Britain—Productivity Assessment and Scenarios for Land Use Change Opportunities. *Sci. Total Environ.* **2018**, *634*, 1108–1118. [[CrossRef](#)]
8. Gang, C.; Zhou, W.; Chen, Y.; Wang, Z.; Sun, Z.; Li, J.; Qi, J.; Odeh, I. Quantitative Assessment of the Contributions of Climate Change and Human Activities on Global Grassland Degradation. *Environ. Earth Sci* **2014**, *72*, 4273–4282. [[CrossRef](#)]
9. Hejcman, M.; Češková, M.; Schellberg, J.; Pätzold, S. The Rengen Grassland Experiment: Effect of Soil Chemical Properties on Biomass Production, Plant Species Composition and Species Richness. *Folia Geobot.* **2010**, *45*, 125–142. [[CrossRef](#)]
10. Chen, J.; Gu, S.; Shen, M.; Tang, Y.; Matsushita, B. Estimating Aboveground Biomass of Grassland Having a High Canopy Cover: An Exploratory Analysis of in Situ Hyperspectral Data. *Int. J. Remote Sens.* **2009**, *30*, 6497–6517. [[CrossRef](#)]

11. Gao, J.; Liang, T.; Yin, J.; Ge, J.; Feng, Q.; Wu, C.; Hou, M.; Liu, J.; Xie, H. Estimation of Alpine Grassland Forage Nitrogen Coupled with Hyperspectral Characteristics during Different Growth Periods on the Tibetan Plateau. *Remote Sens.* **2019**, *11*, 2085. [[CrossRef](#)]
12. Fu, G.; Wang, J.; Li, S. Response of Forage Nutritional Quality to Climate Change and Human Activities in Alpine Grasslands. *Sci. Total Environ.* **2022**, *845*, 157552. [[CrossRef](#)] [[PubMed](#)]
13. Sharma, P.; Leigh, L.; Chang, J.; Maimaitijiang, M.; Caffé, M. Above-Ground Biomass Estimation in Oats Using UAV Remote Sensing and Machine Learning. *Sensors* **2022**, *22*, 601. [[CrossRef](#)] [[PubMed](#)]
14. Oliveira, R.A.; Näsi, R.; Niemeläinen, O.; Nyholm, L.; Alhonoja, K.; Kaivosoja, J.; Jauhiainen, L.; Viljanen, N.; Nezami, S.; Markelin, L.; et al. Machine Learning Estimators for the Quantity and Quality of Grass Swards Used for Silage Production Using Drone-Based Imaging Spectrometry and Photogrammetry. *Remote Sens. Environ.* **2020**, *246*, 111830. [[CrossRef](#)]
15. Wijesingha, J.; Astor, T.; Schulze-Brüninghoff, D.; Wengert, M.; Wachendorf, M. Predicting Forage Quality of Grasslands Using UAV-Borne Imaging Spectroscopy. *Remote Sens.* **2020**, *12*, 126. [[CrossRef](#)]
16. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between Leaf Chlorophyll Content and Spectral Reflectance and Algorithms for Non-Destructive Chlorophyll Assessment in Higher Plant Leaves. *J. Plant Physiol.* **2003**, *160*, 271–282. [[CrossRef](#)] [[PubMed](#)]
17. Li, Z.; Xin, X.; Tang, H.; Yang, F.; Chen, B.; Zhang, B. Estimating Grassland LAI Using the Random Forests Approach and Landsat Imagery in the Meadow Steppe of Hulunber, China. *J. Integr. Agric.* **2017**, *16*, 286–297. [[CrossRef](#)]
18. Lyu, X.; Li, X.; Dang, D.; Dou, H.; Xuan, X.; Liu, S.; Li, M.; Gong, J. A New Method for Grassland Degradation Monitoring by Vegetation Species Composition Using Hyperspectral Remote Sensing. *Ecol. Indic.* **2020**, *114*, 106310. [[CrossRef](#)]
19. Nishikawa, H.; Oenema, J.; Sijbrandij, F.; Jindo, K.; Noij, G.-J.; Hollewand, F.; Meurs, B.; Hoving, I.; van der Vlugt, P.; Bouten, M.; et al. Dry Matter Yield and Nitrogen Content Estimation in Grassland Using Hyperspectral Sensor. *Remote Sens.* **2023**, *15*, 419. [[CrossRef](#)]
20. Zhao, Y.; Zhu, W.; Wei, P.; Fang, P.; Zhang, X.; Yan, N.; Liu, W.; Zhao, H.; Wu, Q. Classification of Zambian Grasslands Using Random Forest Feature Importance Selection during the Optimal Phenological Period. *Ecol. Indic.* **2022**, *135*, 108529. [[CrossRef](#)]
21. Ristok, C.; Poeschl, Y.; Dudenhöffer, J.; Ebeling, A.; Eisenhauer, N.; Vergara, F.; Wagg, C.; Van Dam, N.M.; Weinhold, A. Plant Species Richness Elicits Changes in the Metabolome of Grassland Species via Soil Biotic Legacy. *J. Ecol.* **2019**, *107*, 2240–2254. [[CrossRef](#)]
22. den Broeck, G.V.; Lykov, A.; Schleich, M.; Suci, D. On the Tractability of SHAP Explanations. *J. Artif. Intell. Res.* **2022**, *74*, 851–886. [[CrossRef](#)]
23. Zhong, L.; Guo, X.; Xu, Z.; Ding, M. Soil Properties: Their Prediction and Feature Extraction from the LUCAS Spectral Library Using Deep Convolutional Neural Networks. *Geoderma* **2021**, *402*, 115366. [[CrossRef](#)]
24. Kirk, P.L. Kjeldahl Method for Total Nitrogen. *Anal. Chem.* **1950**, *22*, 354–358. [[CrossRef](#)]
25. Van Soest, P.J.; Robertson, J.B.; Lewis, B.A. Methods for Dietary Fiber, Neutral Detergent Fiber, and Nonstarch Polysaccharides in Relation to Animal Nutrition. *J. Dairy Sci.* **1991**, *74*, 3583–3597. [[CrossRef](#)]
26. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [[CrossRef](#)]
27. Liu, H.Q.; Huete, A. A Feedback Based Modification of the NDVI to Minimize Canopy Background and Atmospheric Noise. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 457–465. [[CrossRef](#)]
28. Lin, S.; Li, J.; Liu, Q.; Li, L.; Zhao, J.; Yu, W. Evaluating the Effectiveness of Using Vegetation Indices Based on Red-Edge Reflectance from Sentinel-2 to Estimate Gross Primary Productivity. *Remote Sens.* **2019**, *11*, 1303. [[CrossRef](#)]
29. Xu, H. Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [[CrossRef](#)]
30. Lukas, V.; Huňady, I.; Kintl, A.; Mezera, J.; Hammerschmiedt, T.; Sobotková, J.; Brtnický, M.; Elbl, J. Using UAV to Identify the Optimal Vegetation Index for Yield Prediction of Oil Seed Rape (*Brassica Napus* L.) at the Flowering Stage. *Remote Sens.* **2022**, *14*, 4953. [[CrossRef](#)]
31. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
32. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely Randomized Trees. *Mach Learn* **2006**, *63*, 3–42. [[CrossRef](#)]
33. Noi, P.T.; Degener, J.; Kappas, M. Comparison of Multiple Linear Regression, Cubist Regression, and Random Forest Algorithms to Estimate Daily Air Surface Temperature from Dynamic Combinations of MODIS LST Data. *Remote Sens.* **2017**, *9*, 398. [[CrossRef](#)]
34. Alnahit, A.O.; Mishra, A.K.; Khan, A.A. Stream Water Quality Prediction Using Boosted Regression Tree and Random Forest Models. *Stoch Environ. Res. Risk Assess* **2022**, *36*, 2661–2680. [[CrossRef](#)]
35. Dhal, S.B.; Mahanta, S.; Gumero, J.; O'Sullivan, N.; Soetan, M.; Louis, J.; Gadepally, K.C.; Mahanta, S.; Lusher, J.; Kalafatis, S. An IoT-Based Data-Driven Real-Time Monitoring System for Control of Heavy Metals to Ensure Optimal Lettuce Growth in Hydroponic Set-Ups. *Sensors* **2023**, *23*, 451. [[CrossRef](#)] [[PubMed](#)]
36. Capolupo, A.; Kooistra, L.; Berendonk, C.; Boccia, L.; Suomalainen, J. Estimating Plant Traits of Grasslands from UAV-Acquired Hyperspectral Images: A Comparison of Statistical Approaches. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 2792–2820. [[CrossRef](#)]
37. Zhou, Z.; Morel, J.; Parsons, D.; Kucheryavskiy, S.V.; Gustavsson, A.-M. Estimation of Yield and Quality of Legume and Grass Mixtures Using Partial Least Squares and Support Vector Machine Analysis of Spectral Data. *Comput. Electron. Agric.* **2019**, *162*, 246–253. [[CrossRef](#)]

38. le Maire, G.; François, C.; Dufrêne, E. Towards Universal Broad Leaf Chlorophyll Indices Using PROSPECT Simulated Database and Hyperspectral Reflectance Measurements. *Remote Sens. Environ.* **2004**, *89*, 1–28. [[CrossRef](#)]
39. Inoue, Y.; Guérif, M.; Baret, F.; Skidmore, A.; Gitelson, A.; Schlerf, M.; Darvishzadeh, R.; Olioso, A. Simple and Robust Methods for Remote Sensing of Canopy Chlorophyll Content: A Comparative Analysis of Hyperspectral Data for Different Types of Vegetation: Simple Sensing of Canopy Chlorophyll Content. *Plant Cell Environ.* **2016**, *39*, 2609–2623. [[CrossRef](#)] [[PubMed](#)]
40. Obermeier, W.A.; Lehnert, L.W.; Pohl, M.J.; Makowski Gianonni, S.; Silva, B.; Seibert, R.; Laser, H.; Moser, G.; Müller, C.; Luterbacher, J.; et al. Grassland Ecosystem Services in a Changing Environment: The Potential of Hyperspectral Monitoring. *Remote Sens. Environ.* **2019**, *232*, 111273. [[CrossRef](#)]
41. Chen, Q.; Meng, Z.; Liu, X.; Jin, Q.; Su, R. Decision Variants for the Automatic Determination of Optimal Feature Subset in RF-RFE. *Genes* **2018**, *9*, 301. [[CrossRef](#)]
42. Sims, D.A.; Gamon, J.A. Relationships between Leaf Pigment Content and Spectral Reflectance across a Wide Range of Species, Leaf Structures and Developmental Stages. *Remote Sens. Environ.* **2002**, *81*, 337–354. [[CrossRef](#)]
43. Ollinger, S.V. Sources of Variability in Canopy Reflectance and the Convergent Properties of Plants. *New Phytol.* **2011**, *189*, 375–394. [[CrossRef](#)]
44. Riano, D.; Vaughan, P.; Chuvieco, E.; Zarco-Tejada, P.J.; Ustin, S.L. Estimation of Fuel Moisture Content by Inversion of Radiative Transfer Models to Simulate Equivalent Water Thickness and Dry Matter Content: Analysis at Leaf and Canopy Level. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 819–826. [[CrossRef](#)]
45. Naicker, R.; Mutanga, O.; Peerbhay, K.; Agjee, N. The Detection of Nitrogen Saturation for Real-Time Fertilization Management within a Grassland Ecosystem. *Appl. Sci.* **2023**, *13*, 4252. [[CrossRef](#)]
46. Cui, H.; Sun, W.; Delgado-Baquerizo, M.; Song, W.; Ma, J.-Y.; Wang, K.; Ling, X. Phosphorus Addition Regulates the Responses of Soil Multifunctionality to Nitrogen Over-Fertilization in a Temperate Grassland. *Plant Soil* **2022**, *473*, 73–87. [[CrossRef](#)]
47. Gao, J.; Liang, T.; Liu, J.; Yin, J.; Ge, J.; Hou, M.; Feng, Q.; Wu, C.; Xie, H. Potential of Hyperspectral Data and Machine Learning Algorithms to Estimate the Forage Carbon-Nitrogen Ratio in an Alpine Grassland Ecosystem of the Tibetan Plateau. *ISPRS J. Photogramm. Remote Sens.* **2020**, *163*, 362–374. [[CrossRef](#)]
48. Bazzo, C.O.G.; Kamali, B.; Hütt, C.; Bareth, G.; Gaiser, T. A Review of Estimation Methods for Aboveground Biomass in Grasslands Using UAV. *Remote Sens.* **2023**, *15*, 639. [[CrossRef](#)]
49. Guerini Filho, M.; Kuplich, T.M.; Quadros, F.L.F.D. Estimating Natural Grassland Biomass by Vegetation Indices Using Sentinel 2 Remote Sensing Data. *Int. J. Remote Sens.* **2020**, *41*, 2861–2876. [[CrossRef](#)]
50. Xu, D.; Wang, C.; Chen, J.; Shen, M.; Shen, B.; Yan, R.; Li, Z.; Karnieli, A.; Chen, J.; Yan, Y.; et al. The Superiority of the Normalized Difference Phenology Index (NDPI) for Estimating Grassland Aboveground Fresh Biomass. *Remote Sens. Environ.* **2021**, *264*, 112578. [[CrossRef](#)]
51. Barnetson, J.; Phinn, S.; Scarth, P. Estimating Plant Pasture Biomass and Quality from UAV Imaging across Queensland's Rangelands. *AgriEngineering* **2020**, *2*, 523–543. [[CrossRef](#)]
52. Lu, X.; Zhang, S.; Tian, Y.; Li, Y.; Wen, R.; Tsou, J.; Zhang, Y. Monitoring Suaeda Salsa Spectral Response to Salt Conditions in Coastal Wetlands: A Case Study in Dafeng Elk National Nature Reserve, China. *Remote Sens.* **2020**, *12*, 2700. [[CrossRef](#)]
53. Zhao, X.; Wu, B.; Xue, J.; Shi, Y.; Zhao, M.; Geng, X.; Yan, Z.; Shen, H.; Fang, J. Mapping Forage Biomass and Quality of the Inner Mongolia Grasslands by Combining Field Measurements and Sentinel-2 Observations. *Remote Sens.* **2023**, *15*, 1973. [[CrossRef](#)]
54. Ceccato, P.; Flasse, S.; Tarantola, S.; Jacquemoud, S.; Grégoire, J.-M. Detecting Vegetation Leaf Water Content Using Reflectance in the Optical Domain. *Remote Sens. Environ.* **2001**, *77*, 22–33. [[CrossRef](#)]
55. Wang, Z.; Skidmore, A.K.; Darvishzadeh, R.; Heiden, U.; Heurich, M.; Wang, T. Leaf Nitrogen Content Indirectly Estimated by Leaf Traits Derived from the PROSPECT Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3172–3182. [[CrossRef](#)]
56. Fernández-Habas, J.; Carriere Cañada, M.; García Moreno, A.M.; Leal-Murillo, J.R.; González-Dugo, M.P.; Abellanas Oar, B.; Gómez-Giráldez, P.J.; Fernández-Rebollo, P. Estimating Pasture Quality of Mediterranean Grasslands Using Hyperspectral Narrow Bands from Field Spectroscopy by Random Forest and PLS Regressions. *Comput. Electron. Agric.* **2022**, *192*, 106614. [[CrossRef](#)]
57. Ma, Y.; Hou, P.; Zhang, L.; Cao, G.; Sun, L.; Pang, S.; Bai, J. High-Resolution Quantitative Retrieval of Soil Moisture Based on Multisource Data Fusion with Random Forests: A Case Study in the Zoige Region of the Tibetan Plateau. *Remote Sens.* **2023**, *15*, 1531. [[CrossRef](#)]
58. Dhal, S.B.; Bagavathiannan, M.; Braga-Neto, U.; Kalafatis, S. Nutrient Optimization for Plant Growth in Aquaponic Irrigation Using Machine Learning for Small Training Datasets. *Artif. Intell. Agric.* **2022**, *6*, 68–76. [[CrossRef](#)]
59. Lanjewar, M.G.; Parab, J.S.; Shaikh, A.Y.; Sequeira, M. CNN with Machine Learning Approaches Using ExtraTreesClassifier and MRMR Feature Selection Techniques to Detect Liver Diseases on Cloud. *Clust. Comput* **2023**, *26*, 3657–3672. [[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.