remote sensing

Article

Optimization and Validation of Hyperspectral Estimation Capability of Cotton Leaf Nitrogen Based on SPA and RF

Xiangyu Chen [1], Xin Lv [1], Lulu Ma [1], Aiqun Chen [2], Qiang Zhang [1] and Ze Zhang [1,*]

[1] Xinjiang Production and Construction Group Oasis Eco-Agriculture Key Laboratory, College of Agriculture, Shihezi University, Shihezi 832003, China
[2] Agricultural Technology Extension Station, Xinjiang Production and Construction Corps Fourth Division, Kokodara 835213, China
* Correspondence: zhangze1227@shzu.edu.cn

Abstract: By studying the spectral information of cotton leaf nitrogen content, sensitive feature bands and spectral indices for leaf nitrogen content were screened, and different methods were used to model the screened feature bands and indices to find a method with higher accuracy and stability of the inversion model, which provides a theoretical basis and technical support for remote sensing estimation of cotton nitrogen content in Xinjiang. The experiment was conducted in 2019–2020 at the Second Company of Shihezi University Teaching Experimental Farm in Xinjiang, China, with six fertilization treatments (0, 120, 240, 360, 480 kg/hm pure N), sampled at five key fertility stages of cotton (squaring stage, full budding stage, flowering, boll stage, and boll opening stage), and the obtained data were used in two modeling approaches (eigenbands and spectral indices) to establish a cotton nitrogen estimation model and estimate the cotton leaf N content. The results showed that the nonlinear model using SVR was validated with an R^2 of 0.71 and an RMSE of 3.91. The linear models of MLR and PLS were developed for the feature bands screened by SPA and RF, respectively, and the best modeling result was achieved by SPA-PLS with a validated R^2 of 0.722 and an RMSE of 3.83. The existing spectral indices were optimized by screening the central wavelength and the simple linear regression model was constructed. The inversion effect of the SVR model with the characteristic spectral modeling was better than the index results. However, compared with the direct use of the characteristic wavelengths and the SVR way of modeling, the accuracy of leaf N content estimation by the model built by optimizing the spectral indices was reduced but the stability was greatly improved, and it can be used as a hyperspectral model for leaf N content at full fertility. The hyperspectral estimation of leaf N content in cotton can be used as a hyperspectral estimation method for the whole fertility period.

Keywords: machine learning; drip irrigation cotton; canopy spectra; spectral index; nitrogen concentration

Citation: Chen, X.; Lv, X.; Ma, L.; Chen, A.; Zhang, Q.; Zhang, Z. Optimization and Validation of Hyperspectral Estimation Capability of Cotton Leaf Nitrogen Based on SPA and RF. *Remote Sens.* **2022**, *14*, 5201. https://doi.org/10.3390/rs14205201

Academic Editor: Xiuliang Jin

Received: 17 August 2022
Accepted: 7 October 2022
Published: 17 October 2022

1. Introduction

Crop nitrogen monitoring is an important means to evaluate crop growth and rational application of nitrogen fertilizer [1–3]. At present, the modern production of cotton in Xinjiang urgently needs a timely and reliable technology to monitor the nitrogen nutrition status of cotton so as to guide accurate fertilization. As a rapid non-destructive monitoring technology, remote sensing can monitor the growth and nutritional status of a variety of crops (wheat, corn, cotton, etc.) without damaging the plant tissue structure [4–7]. The research shows that the estimation model of crop nitrogen content based on a single sensitive band has poor accuracy. The spectral index based on multiple sensitive bands and the combination of sensitive bands can effectively reduce the impact of soil environment and atmospheric conditions on canopy spectrum, and then improve the prediction accuracy of the model [8]. For example, Wright [9] established a quadratic polynomial estimation model of leaf nitrogen content based on normalized vegetation index (NDVI) and normalized green band difference vegetation index (GNDVI); Feng Wei [10] and Song [11] showed that the red edge position parameters and leaf nitrogen content can reach a very significant level, and the leaf nitrogen content can be monitored by reprie, SDr-SDb, and fd729. At present, there is still the problem that the spectral inversion models established for different regions and crop varieties are not common [12]. Therefore, how to make full use of spectral information to further optimize the nitrogen quantitative analysis model is a current research hotspot. For example, the algorithm combining the competitive adaptive reweighting algorithm and correlation coefficient method proposed by Yang Baohua [13] selects 30 sensitive bands from 2151 original bands and establishes a more accurate nonlinear regression model for predicting wheat nitrogen content, but the model is not only complex, and the correlation between the screened variables is high. Liu Haijun [14] and others used spectral reflectance data or spectral differential data to select nitrogen characteristic bands and construct appropriate spectral vegetation index, which can significantly enhance the sensitivity of vegetation to nitrogen. The prediction effect of vegetation index composed of different bands on nitrogen is not the same; Lee [15] and others analyzed the relationship between the nitrogen concentration of cotton leaves and 190 spectral ratio indexes and found that the ratio of the red edge position to the short wave infrared band was more accurate and accurate in predicting the nitrogen concentration of cotton leaves.

Continuous projection algorithm (SPA) [16] is a forward variable selection algorithm that minimizes the collinearity of vector space. It can select the spectral variables with the lowest information redundancy from the full band spectral information to ensure the minimum collinearity between the selected characteristic bands so as to reduce the fitting complexity in the process of model establishment and speed up the fitting operation speed. At the same time, the sensitive characteristic bands screened by the algorithm have the order of importance and can directly reflect the quantitative relationship between the screened variables and dependent variables. Therefore, it has attracted extensive attention in the research of Hyperspectral Information; Random forest (RF) [17] is a typical supervised learning algorithm, an integrated learning algorithm obtained on the basis of decision trees. Randomization is performed on the use of variables (columns) and the use of data (rows) to generate many classification trees, and then the results of the classification trees are aggregated. Random forest improves the prediction accuracy without a significant increase in computing power. Random forest is insensitive to multivariate covariance, and the results are more robust to missing data and unbalanced data and can predict well the role of up to several thousand explanatory variables. Support vector machine regression (SVM) [18] has significant advantages in solving nonlinear problems. In order to adapt to the nonlinearity of sample set, the traditional fitting method usually adds a higher-order term after the linear equation. Although this method is effective, the increased adjustable parameters also increase the risk of overfitting. Support vector machine uses kernel function to solve this contradiction. It uses kernel function to replace the linear term in linear equation. The commonly used kernel functions are:

Polynomial kernel function

$$\mathrm{K(x, y) = (xy + 1)^d \ (d = 1, 2, \ldots, n)} \tag{1}$$

Radial basis kernel function

$$\mathrm{K(x, y)} = \exp(-\gamma \,||x - y||^2) \tag{2}$$

Sigmoid kernel function

$$\mathrm{K(x, y) = tanh(v(xy)} + c) \tag{3}$$

The kernel function can map the input space to the high-dimensional space and construct the linear decision function in the high-dimensional space to realize the linear regression. Therefore, SVR algorithm can better control the overfitting phenomenon. At the same time, it also has global optimization and better generalization ability.

The purpose of this study is to compare the full band, characteristic band, and optimize the model accuracy of spectral modeling by using the spectral information of cotton canopy in Xinjiang and the method of machine learning. Excavate the influence of canopy spectrum on leaf nitrogen content in key growth period of cotton under different data scales, find the inversion method of leaf nitrogen content that can better represent the characteristic planting mode of Xinjiang, and estimate the leaf nitrogen content in each growth period of cotton, in order to provide theoretical basis and technical support for remote sensing estimation of nitrogen content in key growth period of cotton in Xinjiang, and provide scientific guiding suggestions for field management of its later growth. Although there is some basis for using spectral technology to detect crop nitrogen nutrition [19–22], the law of spectral reflectance changes due to different crops, varieties, regions, and planting methods [23]. In particular, there are great differences between Xinjiang and coastal and inland areas in cultivation methods. Therefore, the research on cotton nutrition remote sensing monitoring technology of different varieties under Xinjiang's unique cultivation methods needs to be further deepened in order to establish a universal model for remote sensing monitoring of cotton nitrogen nutrition.

Therefore, in order to improve the accuracy of cotton nitrogen nutrient monitoring models, this study used different methods to process the spectra, and then screened out the sensitive bands combined with the vegetation indices established using hyperspectral data after different prediction processing methods. Nitrogen monitoring models were established, and the inverse accuracy and stability of each model were compared to provide a basis for accurate management of cotton fields in Xinjiang.

2. Materials and Methods

2.1. Study Area and Experimental Design

To further investigate the application of RF and SPA in cotton leaf nitrogen nutrient monitoring, we conducted field nitrogen fertilization experiments. The experiment was carried out in the second company of the teaching and experimental field of Xinjiang Shihezi University from 2019 to 2020 (86°01′E, 44°26.5′N). The test area is shown in Figure 1 (left). The altitude is 429 m, the annual sunshine hours are 2721~2818 h, the frost-free period is 168~171 D, and the active accumulated temperature $\geq$10 °C is 3570~3729 °C. It belongs to a typical temperate continental climate, with a long and severe winter and short and hot summer. The soil texture of the test area is loam, and the basic physical and chemical properties of the soil are: organic matter 19.90 g/kg, alkali hydrolyzable nitrogen 60.88 mg/kg, available phosphorus 17.95 mg/kg, available potassium 134 mg/kg, and the previous crop is cotton.

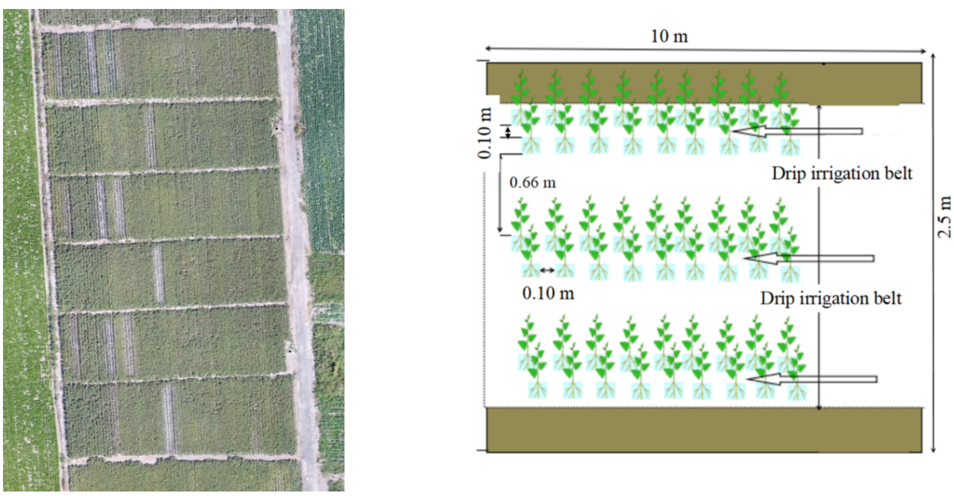

Figure 1. The top view of cotton test area (**left**) and internal row spacing configuration diagram of cotton field experimental community (**right**).

N, P, and K fertilizers were applied as water drops throughout the cotton growth. Among them, urea (N, 46%) 276 kg·hm^{-2}, monoammonium phosphate (P_2O_5, 61%) 174 kg·hm^{-2}, and potassium sulfate (K_2O, 50%) 225 kg·hm^{-2}, were applied as the main fertilizers. A total of nine drip applications were made throughout the growing period, and the application ratios of the three fertilizers (N, P, and K) are shown in Table 1. For each application, the proportionally weighed three fertilizers were poured into the corresponding pressure differential application tank and dissolved, and then the plot was drip irrigated.

Table 1. Fertilization program.

Date	Fertilizer Application Ratio		
	N	P	K
7 June	2.5	7.7	0
15 June	7.5	11.7	6.7
24 June	7.5	11.7	6.7
2 July	12.5	19.2	20
18 July	20	19.2	20
26 July	25	15.4	13.3
5 August	15	15.4	13.3
15 August	10	0	13.3
25 August	0	0	6.7

The cotton variety is "Xinlu Early 53" with sub-membrane drip irrigation technology. It was sown in mid-April each year. A plot trial was conducted. The film width was 205 cm, planting density 180,000 plants/ha, and the spacing configuration was 66 cm + 10 cm, for a total of 1 film, 3 tubes, and 6 rows (Figure 1 right). There were six types of N application, 0, 120, 240, 348, 360, and 480 kg/hm^2 of pure N, denoted as N0, N1, N2, NC, N3, and N4, respectively, where N0 was the control. The experimental plots were planted in continuous cotton with an overall East–West orientation, and the protection rows were set at the East and West ends. The overall design of the experiment was a completely randomized grouping design with three replications and a total of 18 plots, each with an area of 25 m^2, with isolation zones between treatments.

2.2. Hyperspectral Data Acquisition

Ground versus spectral data were collected on days 63, 75, 86, 95, and 120 after seedling emergence. Hyperspectral data of cotton canopy were obtained by a hyperspectral radiometer (SR3500 full-spectrum portable ground object spectrometer from Spectrum Evolution, Southlake, TX, USA). This spectrometer has a full wavelength range (350–2500 nm) with a sampling interval of 1 nm and a total of 2150 data sets. Cotton canopy spectral reflectance for each growth period should be measured in clear weather, without cirrus or dense clouds and with low wind speed, and the measurement time range should be controlled from 11:00 to 15:00. Measurements are made with the sensor probe vertically downward at a height of 1 m from the cotton canopy. Three points per cell were collected and each point was repeated five times. Standard whiteboard correction should be performed before and after each group of sample observation to reduce the influence of cloud cover and solar height variation on spectral reflectance and to improve test accuracy.

2.3. Collection of Nitrogen Content in Cotton Canopy Leaves

After collecting the cotton canopy spectral data, the different organs of the samples (leaves, stems, buds, etc.) were dried separately, the dry weight was measured, then crushed, sieved through a 100-mesh sieve, and the cotton leaf samples were digested with H_2O_2-H_2SO_4 in 50 mL of digestion solution. Then, a 10-mL aliquot was placed in a Kjeldahl nitrogen tester (Haineng-K9840 automatic Kjeldahl nitrogen tester) for distillation and the distilled solution was titrated with (1/2) H_2SO_4 to determine the nitrogen concentration (LNC) of the cotton leaves.

2.4. Data Processing

All data, consisting of 96 observations from all experiments, were pooled in a calculation spreadsheet. The dataset was randomly separated into two databases: 75% for the calibration set and 25% for the validation set. To address the influences of the "dilution effect", growth stage, cultivar, site, and year on the performances of spectral indices and PLSR method in deriving the canopy N content of cotton, we organized the datasets into three dataset formations with different cultivars, sites, and years, in addition to organizing data combinations into calibration and validation datasets.

PLRS is a method that specifies a linear relationship between a set of independent and response variables. In this study, PLSR was used to model the correlation between canopy reflectance spectra (predictor variables) and canopy N content (response variable). The PLSR modelling was performed using software developed by Viscarra Rossel. All calibration spectral data used for building the PLSR models were corrected for light scattering using Standard Normal Variate Transformation (SNV) techniques. The performance of the model was estimated by comparing the differences in prediction abilities using the coefficient of determination (R^2) and the root mean square error of cross-validation/prediction (RMSE). The higher the R^2 and the lower the RMSE, the higher the precision and accuracy of the model to predict the canopy N content.

3. Results

3.1. Variation in Canopy N Content

To investigate the corresponding relationship between the canopy spectral data and the nitrogen content of cotton leaves, we investigated the correlation between the spectral data and the nitrogen content of cotton leaves. The results showed that the nitrogen concentration of cotton leaves in each nitrogen treatment reached the maximum before and after the budding stage. After the full flowering stage, reproductive growth began, nitrogen and other elements transferred to reproductive organs, and the nitrogen concentration in leaves began to decrease gradually. The change of canopy nitrogen content is shown in Table 2. The results showed that the leaf nitrogen concentration of three leaf layers decreased with the delay of growth period, and the leaf nitrogen concentration of five nitrogen application levels was higher than that of each leaf layer under no nitrogen treatment (N0). Taking NC (conventional fertilization level) as an example, in the upper layer of cotton plant, the leaf nitrogen concentration in each growth stage decreased by 8.19% (full flowering stage), 7.41% (flowering and boll stage), 14.78% (full boll stage), and 43.61% (boll opening stage) respectively. In the middle layer of cotton plant, compared with the budding stage, it decreased by 13.59% (full flowering stage), 12.79% (flowering and Bolling stage), 32.08% (full Bolling stage), and 46.29% (Bolling stage). In the lower layer of cotton plant, compared with the budding stage, it decreased by 10.28% (full flowering stage), 19.75% (flowering and Bolling stage), 30.60% (full Bolling stage), and 46.82% (Bolling stage). The change trend of leaf nitrogen content in each layer of cotton is basically the same as the growth period, showing that the leaf nitrogen content in each layer decreases with the growth period, and the reduction rate increases with the growth period. However, there are some differences between different leaf layers. There is a significant gap between the middle and lower layers and the upper layer at full boll stage. It is mainly because the nutrients required by cotton in full boll stage increase sharply, and the absorption and transportation law of nutrients in cotton plants is from bottom to top, which leads to the rapid decline of nitrogen content in middle and lower leaves. When the growth process reaches the boll opening stage, the nutrients supplied by the middle and lower layers are not enough to support the normal growth and development of cotton bolls, and the nutrients continue to be supplied to the upper leaves.

Table 2. Nitrogen concentrations.

	Budding Period	Full Bloom	Blooming Period	Boll Stage	Spitting Period
	41.9 ± 3.4 a	36.0 ± 5.3 a	33.9 ± 4.0 ab	29.8 ± 4.3 b	19.9 ± 3.5 c
N0	41.2 ± 4.2 a	36.6 ± 4.6 a	34.9 ± 3.6 ab	31.6 ± 3.5 b	19.9 ± 5.5 c
	33.7 ± 4.8 a	34.7 ± 3.6 a	32.4 ± 3.2 ab	31.4 ± 3.9 b	21.5 ± 5.8 c
	43.6 ± 5.0 a	37.7 ± 4.5 a	38.4 ± 4.2 a	35.0 ± 4.2 ab	23.3 ± 4.5 c
N1	41.5 ± 5.2 a	38.5 ± 3.2 a	39.0 ± 4.8 a	33.4 ± 4.7 b	22.2 ± 3.9 c
	29.8 ± 4.1 a	38.9 ± 3.9 a	36.2 ± 4.8 ab	29.7 ± 3.7 bc	22.5 ± 2.3 c
	41.1 ± 4.5 a	41.1 ± 5.0 a	39.0 ± 3.7 a	30.0 ± 4.9 b	17.8 ± 4.6 c
N2	41.9 ± 3.9 a	41.9 ± 5.4 a	39.4 ± 4.4 a	33.2 ± 3.6 b	20.3 ± 3.4 c
	40.9 ± 4.4 a	40.9 ± 5.6 a	35.5 ± 4.3 ab	30.3 ± 3.6 b	20.9 ± 2.5 c
	40.8 ± 3.6 a	40.2 ± 3.8 a	40.6 ± 3.1 a	32.2 ± 4.7 b	22.9 ± 2.3 c
N3	42.6 ± 3.2 a	41.3 ± 4.4 a	39.9 ± 4.6 a	32.6 ± 3.6 b	22.7 ± 3.3 c
	35.6 ± 5.8 a	38.3 ± 3.4 a	37.6 ± 3.8 a	30.6 ± 4.7 b	22.9 ± 3.7 c
	46.6 ± 4.6 a	38.4 ± 5.7 a	38.8 ± 3.2 a	28.7 ± 4.7 b	19.9 ± 5.3 c
N4	43.8 ± 5.3 a	39.0 ± 4.7 a	40.0 ± 3.8 a	35.7 ± 4.9 b	20.4 ± 2.2 c
	39.5 ± 3.2 a	37.4 ± 3.6 a	38.4 ± 4.3 a	32.1 ± 3.9 b	22.2 ± 4.9 c
	45.3 ± 5.1 a	40.4 ± 4.0 a	39.0 ± 5.1 a	35.9 ± 4.7 b	23.7 ± 4.8 c
NC	42.1 ± 4.8 a	41.3 ± 4.2 a	38.2 ± 3.1 a	29.7 ± 4.1 b	23.5 ± 3.9 c
	35.3 ± 5.3 a	38.3 ± 3.2 a	34.6 ± 3.3 ab	30.7 ± 3.5 b	23.5 ± 5.7 c

Note: a, b, and c are used to indicate that in the data analysis, statistical differences exist. The same letter means the difference between two data groups is not significant, and different letters mean that the difference between data is significant.

3.2. Relationship between Canopy N Content and Spectral Data

There are many soil hyperspectral band data, and there are serious multicollinearity among spectral data variables, which affects the complex structure of the model. The research shows that the redundant information variables can be effectively removed, and the spectral characteristic information can be amplified through the spectral index [24]. According to the spectral characteristics of cotton canopy leaves and previous inductive studies, 14 indexes composed of bands sensitive to nitrogen content were selected for analysis. The specific calculation formula is shown in Table 3.

Table 3. Spectral index calculation formula.

Serial Number	Abbreviation	Calculation Formula
1	SRPI	R430/R680
2	mSR705	(R750 − R445)/(R705 − R445)
3	mNDVI 705	(R750 − R705)/(R750 + 2R445)
4	NPCI	(R680 − R430)/(R680 + R430)
5	RENDVI	(R750 − R705)/(R750 + R705)
6	RI-1dB	R735/R720
7	VOG	R740/R720
8	DCNI	(R720 − R700)/(R700 − R670) /(R720 − R670 + 0.03)
9	PRI	(R531 − R570/(R531 + R570)
10	RVI	R800/R670
11	NDVI	(R800 − R670)/(R800 + R670)
12	VOG3	(R734 − R747)/(R715 + R720)
13	ND705	(R750 − R705)/(R750 + R705)
14	NRI	(R570 − R670)/(R570 + R670)

Note: R is the original spectral reflectance. For example, R430 represents the spectral reflectance at 430 nm, and the rest are similar.

To evaluate the stability of spectral indices in deriving canopy N content, we established the relationships between representatively published spectral indices and canopy N content with seven dataset formations using calibration datasets. As illustrated in Figure 2, most spectral indices had only weak relationships with canopy N content. With the exception NDVI and SRPI, none of the spectral indices showed a consistent performance in estimating the canopy N content across seven calibration dataset formations. In addition, all spectral indices showed a poor predictive ability for the calibration datasets during the period before flowering. This may be due to the influence of variation of the above-ground biomass and canopy structure of cotton.

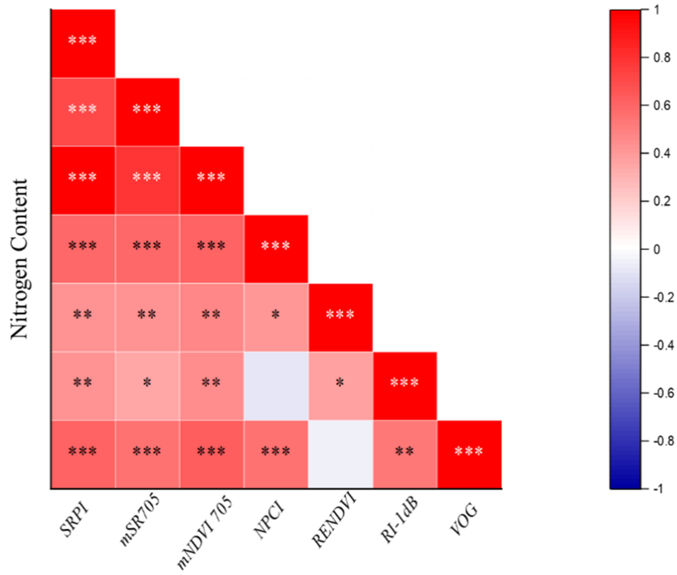

Figure 2. Vegetation index and nitrogen correlation analysis (* represents that the sig value is less than 0.05, ** represents that the sig value is less than 0.01, *** represents that the sig value is less than 0.001).

3.3. Screening Spectral Data

We used SPA to extract the sensitive characteristic band of leaf nitrogen content from resampling and smoothing preprocessed hyperspectral data in MATLAB. After repeated sampling and inspection, when RMSE reached the lowest value of 0.03, 39 characteristic bands sensitive to leaf nitrogen content in canopy spectrum were optimized, and the number of spectral bands decreased by 90.9%. The selected characteristic bands are respectively shown in Table 4 according to the order of importance. The reflectance of each selected sensitive characteristic band has a significant correlation with the nitrogen content of leaves, and all have passed the significance test at the level of 0.01. Among them, the absolute value of the correlation coefficient at 687 nm is the highest, the absolute value of the correlation coefficient at 1540 nm is the lowest, and the absolute value of the correlation coefficient is 0.852.

Table 4. Screening results of spectral characteristic bands.

Math	Number	Result
SPA	39	1987, 648, 2103, 700, 1706, 545, 694, 759, 561, 474, 740, 1806, 540, 1893, 783, 1058, 580, 628, 481, 602, 755, 1950, 677, 454, 1806, 671, 773, 424, 761, 393, 1991, 410, 714, 433, 351, 387, 438, 730, 687
RF	16	1847, 1851, 734, 1058, 1955, 687, 686, 1907, 783, 688, 1818, 541, 773, 1987, 513, 1901

In RF, this paper sets the threshold of selection possibility as 0.5, and finally selects a total of 16 sensitive bands (the 16 sensitive bands are arranged in descending order of the selected probability), in which the highest selection probability is 0.7994 and the lowest is 0.506.

3.4. Evaluation of Optimised Spectral Indices

SPA and RF were used to screen sensitive spectral bands and optimize the central wavelength of five spectral indexes with extremely significant correlation. The optimized results are shown in Table 5. The results show that SPA-DCNI has the longest wavelength selection migration distance, and the three bands are shifted by 11 nm in total. The central wavelength migration distance of RF-ND705 is the shortest, and the total migration distance of the two bands is 2 nm. A simple linear regression model was established between the optimized spectral index and the nitrogen content in cotton leaves (the ratio of modeling set to verification set was 2:1), and the estimation ability of the optimized spectral index on nitrogen content in cotton leaves was analyzed. It was found that the optimized spectral index could be used to invert the nitrogen content in cotton leaves better, and most of the R^2 models were above 0.5. Rf-DCNI had the best estimation ability, R^2 = 0.791, RMSE = 3.74. Rf-RVI had poor estimation ability, R^2 = 0.684, RMSE = 4.76.

Table 5. Optimize the spectral index fitting results.

Abbreviation	Optimal Center Wavelength			R^2	RMSE
SPA-mSR705	755	445	708	0.764	3.97
SPA-RI-1 dB	738	719		0.702	4.69
SPA-DCNI	725	706	677	0.832	4.01
SPA-RVI	789	677		0.710	4.56
SPA-ND705	755	708		0.709	4.09
RF-mSR705	750	448	707	0.780	4.48
RF-RI-1 dB	736	723		0.702	4.21
RF-DCNI	723	705	673	0.791	3.74
RF-RVI	788	673		0.684	4.76
RF-ND705	750	707		0.694	3.92

The MSR model established by the above optimized spectral index was verified, and the verification results were shown in Table 6. The results showed that the model accuracy of the cotton leaf nitrogen estimation model established by the optimized spectral index was improved in different degrees, with R^2 above 0.5 and RMSE below 5.00. SPA-DCIN has the highest accuracy, R^2 = 0.762, RMSE = 3.06. Rf-msr705 showed the greatest improvement, with R^2 increased from 0.402 to 0.659, an increase of 63.93%. The results showed that the optimized spectral index performed well in retrieving the nitrogen content of cotton leaves, but there were still some gaps compared with the nitrogen estimation model established by spectral characteristic bands.

Table 6. Uses SPA and RF to optimize the spectral index's leaf nitrogen content prediction model.

Index	Equation	Before R^2	Before RMSE	After R^2	After RMSE
SPA-mSR705	Y = 8.72X − 10.351	0.659	4.53	0.678	3.73
SPA-RI-1 dB	Y = 32.016X − 13.283	0.601	4.87	0.613	4.57
SPA-DCNI	Y = 1.868X + 18.515	0.673	4.46	0.762	3.06
SPA-RVI	Y = 46.58X − 2.238	0.521	5.02	0.591	4.92
SPA-ND705	Y = 70.772X − 5.8457	0.616	4.55	0.638	4.45
RF-mSR705	Y = 7.631X − 9.176	0.402	4.59	0.659	4.89
RF-RI-1 dB	Y = 34.13X − 14.752	0.542	4.57	0.642	4.77
RF-DCNI	Y = 2.87X + 9.331	0.607	3.86	0.714	4.66
RF-RVI	Y = 57.9X − 12.437	0.584	4.97	0.589	4.87
RF-ND705	Y = 76.532X − 7.788	0.538	4.51	0.597	4.11

3.5. Testing of the Estimation Ability of the Model

To verify the application effect of the model, SPA-DCNI, RF-DCNI, SPA-PLS, and RF-PLS models with high inversion accuracy in the above different modeling methods were selected, and the spectral data of the four key growth periods were brought into the model for cross-validation. All data were divided into two SVR modeling sets and validation sets (modeling sets accounted for 70% and validation sets accounted for 30%), and the MODEL LNC estimation model was established using SVR. The results are shown in Table 7. The modeling methods with the highest accuracy were SVR, RF-PLS, RF-PLS and SPARF-DCNI, R^2 0.673, 0.783, 0.774, and 0.634, respectively, and RMSE of 4.62, 3.12, 3.28, and 3.52, respectively. By comparing the accuracy of models in the whole growth period, as shown in Figure 3, the SVR, SPA-PLS, and RF-PLS models established based on characteristic wavelength have excellent performance in some periods, they are not good on the whole. Spa-DCNI and RF-DCNI models established based on optimization center wavelength have very stable overall realization, and their R^2 values are all above 0.59. It has good inversion accuracy and stability.

Table 7. Each fertility period model validates.

Model		Budding Period	Full Bloom	Blooming Period	Boll Stage
SVR	R^2	0.673	0.772	0.641	0.379
	RMSE	4.62	3.37	4.88	6.21
SPA-PLS	R^2	0.523	0.667	0.648	0.529
	RMSE	5.33	4.61	4.79	5.15
RF-PLS	R^2	0.577	0.783	0.774	0.434
	RMSE	4.96	3.12	3.28	5.43
SPA-DCNI	R^2	0.620	0.635	0.672	0.590
	RMSE	3.76	3.82	3.76	4.68
RF-DCNI	R^2	0.593	0.601	0.623	0.634
	RMSE	4.72	4.11	3.97	3.52

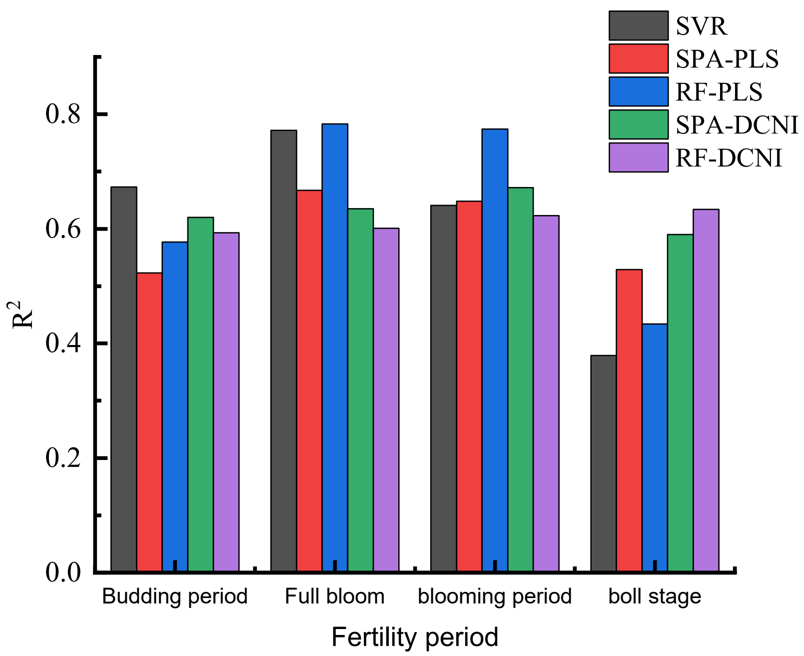

Figure 3. The modeling results.

4. Discussion

The comparative analysis shows that when using SVM to model the regression model, as a nonlinear model, it has strong learning ability, and the accuracy of the model is higher than that of the linear model. However, the final fitting result of the model depends too much on the selection of kernel function [25,26]. In the selection process, the appropriate kernel function parameters can reduce the modeling accuracy of the model from 0.98 to 0.76 and improve the verification accuracy from 0.55 to 0.71, which not only solves the overfitting phenomenon in the construction of the model, but also ensures the stability of the model. The biggest advantage of spa method is to eliminate the collinearity between wavelength variables and ensure the contribution of the screened variables to the estimated variables [27]. However, in the process of screening variables by this method, some effective information in the spectral information may be eliminated, or some wavelength variables with low correlation with the estimated variables may be introduced [28]. Although the correlation between some sensitive characteristic wavelengths and leaf nitrogen content selected in this study passed the significance test at the level of 0.01, it cannot be brought into any spectral index for optimization. Although the selected sensitive characteristic bands of leaf nitrogen content have high correlation with leaf nitrogen content, its physical significance remains to be discussed. In MLR, the choice of factors and the expression of this factor are just speculation, which affects the diversity of selected factors and the immeasurability of some factors, so that regression analysis is limited in some cases [29]. The accuracy of the model established in this study is slightly lower than that of the leaf nitrogen estimation model established by Yincai Xia and Cummings [30,31] using canopy spectrum. Compared with the above research, the prediction accuracy of the model is low, which may be mainly due to the different data acquisition environment. The data used in this study were collected in the field. There are many uncontrollable interference factors, such as weather and man-made, in the process of data collection, which reduces the accuracy of spectral data and affects the accuracy of model detection [32].

The optimal central wavelength for retrieving leaf nitrogen content varies with vegetation index and crop species. The optimal central wavelength of spectral index selected in this experiment is different from the defined central bands such as red light and red edge. The reason is that when Sims et al. [33] proposed the optimal central wavelength, the research object is the chlorophyll of trees in California, USA. Therefore, for the inversion of leaf nitrogen content of cotton in Xinjiang, the exponential optimal central wavelength also shifted. At the same time, it is also similar to the wheat nitrogen inversion studied by Zhang Xiaoyuan [34]. The optimal central wavelength selected for msr705 is basically the same, indicating that the selection of central wavelength has certain commonality in nitrogen inversion. It should be pointed out that although the estimation accuracy of cotton leaf nitrogen content by using optimized spectral indexes spa-dcni and rf-dcni is slightly lower than that of the estimation model spa-pls established by using sensitive band, it has achieved better estimation results than spa-pls and other characteristic band models. The stability in the estimation of the whole growth period is much higher than that of other modeling methods, and the application potential in the estimation of vegetation physical and chemical parameters is worthy of further discussion. Therefore, when using remote sensing technology to monitor crop biochemical parameters, the accuracy of information parameters obtained from spectral images, the applicability of vegetation index and the selection of modeling methods need to be further studied. The high correlation between spectral reflectance and leaf nitrogen content does not mean that the reflectance at this wavelength must have an indication of nitrogen content. It is necessary to comprehensively consider the population characteristic leaf area index effects of internal structure of leaves, vegetation coverage, and soil background [35]. In addition, based on the canopy spectral information of cotton Xinluzao 53, this study carried out the estimation of leaf nitrogen content, and achieved high estimation accuracy, but the estimation ability of this method on leaf nitrogen content of other varieties needs to be tested.

5. Conclusions

Using a variety of hyperspectral parameters of canopy hyperspectral reflectance and evolution under different experimental conditions, this study compared the quantitative relationship between cotton leaf nitrogen content and canopy reflectance spectrum established by MLR and PLS, and preliminarily put forward the sensitive hyperspectral parameters and prediction equation of cotton leaf nitrogen status. By comparing the accuracy relationship of the model constructed among characteristic band modeling, optimized spectral index modeling, and SVM modeling, the following conclusions are drawn:

(1) Among the three modeling methods, the inversion of the model constructed using eigenbands is the best at each fertility period. However, there is a problem of non-uniformity in the characteristic bands among the fertility periods. The accuracy of the model built using spectral indices decreased to some extent compared with the eigenbands, but the estimation was the most stable throughout the growth period, and it could effectively estimate the nitrogen content of cotton leaves.
(2) In estimating the nitrogen content of leaves for a specific growth period, higher accuracy can be obtained with models built using characteristic spectral bands. However, with inversion of leaf nitrogen content at full growth period, the model built using spectral indices can invert cotton leaf nitrogen content better and more consistently. Combining the two, the optimized spectral index using the characteristic waveform has better correlation with the nitrogen content of cotton leaves, and the inversion effect is more stable, which is a good idea to optimize the accuracy of the model.
(3) In the next step, we will continue to study the relationship between spectral incidence and nitrogen content of cotton leaves in depth, adopt more advanced algorithms, and consider the information differences brought by different cotton varieties and different growing regions on this basis, with a view to establishing a general model for cotton nitrogen nutrition estimation applicable to all cases.

Author Contributions: Conceptualization, X.L. and Z.Z.; Data curation, X.C. and L.M.; Formal analysis, A.C. and Q.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by National Natural Science Foundation of China, China (Grant No. 42061058), Science and Technology Research Plan for Key Areas of XPCC, China (Grant No. 2020AB005).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Berger, K.; Verrelst, J.; Féret, J.B.; Wang, Z.; Wocher, M.; Strathmann, M.; Danner, M.; Mauser, W.; Hank, T. Crop nitrogen monitoring, Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sens. Environ.* **2020**, *242*, 111758. [CrossRef]
2. Zha, H.; Miao, Y.; Wang, T.; Li, Y.; Zhang, J.; Sun, W.; Feng, Z.; Kusnierek, K. Improving Unmanned Aerial Vehicle Remote Sensing-Based Rice Nitrogen Nutrition Index Prediction with Machine Learning. *Remote Sens.* **2020**, *12*, 215. [CrossRef]
3. Mubarak, I.; Janat, M. Response of new cotton variety (Rassafa) to nitrogen fertilizer and partial rootzone drying drip irrigation. *Open Agric. J.* **2018**, *12*, 1–10. [CrossRef]
4. Tuvdendorj, B.; Wu, B.; Zeng, H.; Batdelger, G.; Nanzad, L. Determination of appropriate remote sensing indices for spring wheat yield estimation in Mongolia. *Remote Sens.* **2019**, *11*, 2568. [CrossRef]
5. Diao, C. Remote sensing phenological monitoring framework to characterize corn and soybean physiological growing stages. *Remote Sens. Environ.* **2020**, *248*, 111960. [CrossRef]
6. Xu, W.; Chen, P.; Zhan, Y.; Chen, S.; Zhang, L.; Lan, Y. Cotton yield estimation model based on machine learning using time series UAV remote sensing data. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102511. [CrossRef]
7. Yadav, P.K.; Thomasson, J.A.; Hardin, R.; Searcy, S.W.; Braga-Neto, U.; Popescu, S.C.; Martin, D.E.; Rodriguez, R.; Meza, K.; Enciso, J.; et al. Detecting Volunteer Cotton Plants in a Corn Field with Deep Learning on UAV Remote-Sensing Imagery. *arXiv* **2022**, arXiv:2207.06673.
8. Ma, Y.; Zhang, Q.; Yi, X.; Ma, L.; Zhang, L.; Huang, C.; Zhang, Z.; Lv, X. Estimation of Cotton Leaf Area Index (LAI) Based on Spectral Transformation and Vegetation Index. *Remote Sens.* **2022**, *14*, 136. [CrossRef]
9. Wright, D.L.; Rasmussen, V.P.; Ramsey, R.D.; Baker, D.J.; Ellsworth, J.W. Canopy reflectance estimation of wheat nitrogen content for grain protein management. *GIScience Remote Sens.* **2004**, *41*, 287–300. [CrossRef]
10. Wei, F.; Yan, Z.; Yongchao, T.; Weixing, C.; Xia, Y.; Yingxue, L. Monitoring leaf nitrogen accumulation in wheat with hyper-spectral remote sensing. *Acta Ecol. Sin.* **2008**, *28*, 23–32. [CrossRef]

11. Song, X.; Xu, D.; He, L.; Feng, W.; Wang, Y.; Wang, Z.; Coburn, C.A.; Guo, T. Using multi-angle hyperspectral data to monitor canopy leaf nitrogen content of wheat. *Precis. Agric.* **2016**, *17*, 721–736. [CrossRef]
12. Huete, A.R.; Jackson, R.D.; Post, D.F. Spectral response of a plant canopy with different soil backgrounds. *Remote Sens. Environ.* **1985**, *17*, 37–53. [CrossRef]
13. Yang, B.; Wang, M.; Sha, Z.; Wang, B.; Chen, J.; Yao, X.; Cheng, T.; Cao, W.; Zhu, Y. Evaluation of aboveground nitrogen content of winter wheat using digital imagery of unmanned aerial vehicles. *Sensors* **2019**, *19*, 4416. [CrossRef] [PubMed]
14. Haijun, Q.; Xiu, J.; Liu, Z.; Maxime, D.I.; Shaowen, L. Predicting sandy soil moisture content with hyperspectral imaging. *Int. J. Agric. Biol. Eng.* **2017**, *10*, 175–183. [CrossRef]
15. Tarpley, L.; Reddy, K.R.; Sassenrath-Cole, G.F. Reflectance indices with precision and accuracy in predicting cotton leaf nitrogen concentration. *Crop Sci.* **2000**, *40*, 1814–1819. [CrossRef]
16. Yong YI, N.; Guanghui, W. Hyperspectral characteristic wavelength selection method for moldy maize based on continuous projection algorithm fusion information entropy. *J. Nucl. Agric. Sci.* **2020**, *34*, 356.
17. Chan, J.C.W.; Paelinckx, D. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sens. Environ.* **2008**, *112*, 2999–3011. [CrossRef]
18. Devos, O.; Ruckebusch, C.; Durand, A.; Duponchel, L.; Huvenne, J.P. Support vector machines (SVM) in near infrared (NIR) spectroscopy, Focus on parameters optimization and model interpretation. *Chemom. Intell. Lab. Syst.* **2009**, *96*, 27–33. [CrossRef]
19. Daughtry, C.S.; Walthall, C.L.; Kim, M.S.; De Colstoun, E.B.; McMurtrey Iii, J.E. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* **2000**, *74*, 229–239. [CrossRef]
20. Prado Osco, L.; Marques Ramos, A.P.; Roberto Pereira, D.; Akemi Saito Moriya, É.; Nobuhiro Imai, N.; Takashi Matsubara, E.; Estrabis, N.; de Souza, M.; Marcato Junior, J.; Gonçalves, W.N.; et al. Predicting canopy nitrogen content in citrus-trees using random forest algorithm associated to spectral vegetation indices from UAV-imagery. *Remote Sens.* **2019**, *11*, 2925. [CrossRef]
21. Li, Z.; Jin, X.; Yang, G.; Drummond, J.; Yang, H.; Clark, B.; Li, Z.; Zhao, C. Remote Sensing of Leaf and Canopy Nitrogen Status in Winter Wheat (*Triticum aestivum* L.) Based on N-PROSAIL Model. *Remote Sens.* **2018**, *10*, 1463. [CrossRef]
22. Chauhan, S.; Darvishzadeh, R.; Boschetti, M.; Nelson, A. Estimation of crop angle of inclinationfor lodged wheat using multi-sensor SAR data. *Remote Sens. Environ.* **2020**, *236*, 111488. [CrossRef]
23. Fu, Y.; Yang, G.; Li, Z.; Li, H.; Li, Z.; Xu, X.; Song, X.; Zhang, Y.; Duan, D.; Zhao, C.; et al. Progress of hyperspectral data processing and modelling forcereal crop. *Comput. Electron. Agric.* **2020**, *172*, 105321. [CrossRef]
24. Kobayashi, N.; Tani, H.; Wang, X.; Sonobe, R. Crop classification using spectral indices derived from Sentinel-2A imagery. *J. Inf. Telecommun.* **2020**, *4*, 67–90. [CrossRef]
25. Deiss, L.; Margenot, A.J.; Culman, S.W.; Demyan, M.S. Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. *Geoderma* **2020**, *365*, 114227. [CrossRef]
26. Golhani, K.; Balasundram, S.K.; Vadamalai, G.; Pradhan, B. A review of neural networks in plant disease detection using hyperspectral data. *Inf. Processing Agric.* **2018**, *5*, 354–371. [CrossRef]
27. Wang, D.; Li, G.; Wu, S.; Lin, L. Apply a new characteristic wavelength screening method to improving the spectral quantitative analysis precision of multiple immune cell types. *Infrared Phys. Technol.* **2022**, *123*, 104140. [CrossRef]
28. Li, Z.; Tang, X.; Shen, Z.; Yang, K.; Zhao, L.; Li, Y. Comprehensive comparison of multiple quantitative near-infrared spectroscopy models for Aspergillus flavus contamination detection in peanut. *J. Sci. Food Agric.* **2019**, *99*, 5671–5679. [CrossRef]
29. Weih, M.; Hamnér, K.; Pourazari, F. Analyzing plant nutrient uptake and utilization efficiencies, comparison between crops and approaches. *Plant Soil* **2018**, *430*, 7–21. [CrossRef]
30. Cummings, C.R. In-Season Corn Nitrogen Status Diagnosis and Precision Management with Proximal and Remote Sensing. Doctoral Dissertation, University of Minnesota, Minneapolis, MN, USA, 2020.
31. Cai, Y.; Guan, K.; Nafziger, E.; Chowdhary, G.; Peng, B.; Jin, Z.; Wang, S.; Wang, S. Detecting in-season crop nitrogen stress of corn for field trials using UAV-and CubeSat-based multispectral sensing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 5153–5166. [CrossRef]
32. Bera, S.; Shrivastava, V.K. Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification. *Int. J. Remote Sens.* **2020**, *41*, 2664–2683. [CrossRef]
33. Sims, D.A.; Gamon, J.A. Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance, a comparison of indices based on liquid water and chlorophyll absorption features. *Remote Sens. Environ.* **2003**, *84*, 526–537. [CrossRef]
34. Zhang, X.; Zhang, L.; Zhang, X.; Wang, S.; Tian, J.; Zhai, Y. Sensitivity of different spectral vegetation index for estimating winter wheat leaf nitrogen. *Sci. Agric. Sin.* **2017**, *50*, 474–485.
35. Xue, J.; Su, B. Significant remote sensing vegetation indices, A review of developments and applications. *J. Sens.* **2017**, *2017*, 1353691. [CrossRef]