

## Estimation of nitrogen in cotton leaves using different hyperspectral region data

Qiang ZHANG<sup>1a</sup>, Lulu MA<sup>1b</sup>, Xiangyu CHEN<sup>1</sup>, Jiao LIN<sup>2</sup>, Caixia YIN<sup>1</sup>,  
Qiushuang YAO<sup>1</sup>, Xin LV<sup>1\*</sup>, Ze ZHANG<sup>1\*</sup>

<sup>1</sup>Shihezi University College of Agriculture/ The Key Laboratory of Oasis Eco-agriculture, Xinjiang Production and Construction Group, Shihezi 832003, China; [1510665156@qq.com](mailto:1510665156@qq.com); [malulu@stu.shzu.edu.cn](mailto:malulu@stu.shzu.edu.cn); [chenxinagyu@stu.shzu.edu.cn](mailto:chenxinagyu@stu.shzu.edu.cn); [yincaixia@stu.shzu.edu.cn](mailto:yincaixia@stu.shzu.edu.cn); [20182012108@stu.shzu.edu.cn](mailto:20182012108@stu.shzu.edu.cn); [luxin@shzu.edu.cn](mailto:luxin@shzu.edu.cn) (\*corresponding author); [zhangze1227@shzu.edu.cn](mailto:zhangze1227@shzu.edu.cn) (\*corresponding author)

<sup>2</sup>Tarim University College of Agriculture, Alar, 843300, China; [linjiao@taru.edu.cn](mailto:linjiao@taru.edu.cn)

<sup>ab</sup>These authors contributed equally to the work

### Abstract

As an important index of a plant's N nutrition, leaf nitrogen content (LNC) can be quickly monitored in real time with hyperspectral information, which is helpful to guide the precise application of N in cotton leaves. In this study, taking cotton dripping in Xinjiang, China, as the object of study, five N application treatments (0, 120, 240, 360, 480 kg ha<sup>-1</sup>) were set up, and the hyperspectral data and the N content of main stem functional leaves at the cotton flower and boll stage were collected. The results showed that (1) comparing the correlations of the three types of spectral data from the original spectra, first derivative spectra, and second derivative spectra with the LNC of cotton, the first derivative spectra increased the correlation between the reflectance in the peak and valley ranges of the spectral curves and the LNC of cotton; (2) in the three hyperspectral regions of VIS, NIR, and SWIR, all R<sup>2</sup> values of the estimation model for the LNC of cotton established based on the characteristic wavelengths of the original and the first derivative spectra were greater than 0.8, and the model accuracy was better than that of the second derivative spectra; and (3) the normalized root mean square error (n-RMSE) values of the validated model using MLR, PCR, and PLSR regression methods were all in the range of 10–20%, indicating that the established model could well estimate the nitrogen content of cotton leaves. The results of this study demonstrate the potential of the three hyperspectral domains of VIR, NIR, and SWIR to estimate the LNC of cotton and provide a new basis for hyperspectral data application in crop nutrient monitoring.

**Keywords:** characteristic wavelengths; leaf spectrum; near-infrared region; PCR; PLSR; short wave infrared region

### Introduction

Nitrogen is the “life element” of crop physiological activities. Nitrogen deficiency hinders the synthesis of crop protein, leads to a decline in the photosynthetic rate and leaf yellowing, affects the photosynthetic capacity and growth of crops, and thus affects the yield and quality of crops (Serrano *et al.*, 2000; Wang *et al.*, 2015). Hyperspectral imaging is considered a real-time and rapid tool for non-destructive monitoring of crop

Received: 13 Dec 2021. Received in revised form: 13 Mar 2022. Accepted: 15 Mar 2022. Published online: 16 Mar 2022.

From Volume 49, Issue 1, 2021, Notulae Botanicae Horti Agrobotanici Cluj-Napoca journal uses article numbers in place of the traditional method of continuous pagination through the volume. The journal will continue to appear quarterly, as before, with four annual numbers.

nitrogen. Be-cause of the correlation between nitrogen and other elements, characteristic wavelengths that are significant for nitrogen estimation are found in the vegetation spectrum (Curran, 1989; Homolová *et al.*, 2013).

In C3 plants, approximately 19% of leaf nitrogen is allocated to the light-harvesting complex, where it is mainly used to synthesize chlorophylls and chlorophyll-binding proteins in both photosystems (Chapin *et al.*, 1987). This relationship explains the strong correlation between nitrogen and chlorophyll suggested by many studies. In the early stage of crop growth, chlorophyll content and nitrogen content in leaves have a high correlation, and the characteristics of chlorophyll absorption in the visible spectrum can be directly used to estimate nitrogen content (le Maire *et al.*, 2008; Zhang *et al.*, 2020). Chlorophyll content in leaves of cotton was correlated with Vogelmann1 in the four growing periods (squaring stage, full budding stage, flowering and boll stage, and boll opening stage); the correlation coefficients were 0.944, 0.907, 0.895, and 0.930, respectively (Hong *et al.*, 2019). In the reproductive growth stage, due to the dynamic transfer of nitrogen, nitrogen is redistributed from vegetative organs to reproductive structures (Ohyama, 2010), and the chlorophyll content in leaves decreases, but the nitrogen content in plants will remain unchanged. The correlation between chlorophyll content in the canopy and nitrogen content in the aboveground parts will gradually weaken with the emergence or aging of re-productive organs (Berger *et al.*, 2020). Moreover, some studies suggest that the relationship between chlorophyll and nitrogen is highly dependent on the growing season of crops (Spaner *et al.*, 2005; Xu *et al.*, 2021). This relationship is unstable in different cultivars, so it is limited to estimate nitrogen by chlorophyll (Hosgood, 1994).

The different absorption peaks of protein (and nitrogen) are mainly in the range of the SWIR spectra in Curran's research work (1989). Furthermore, Kokaly (2009) suggested that the VIS and NIR could be used to directly evaluate the nitrogen status of crops. Zhang *et al.* (2010) constructed and evaluated the NIR spectral model of rice leaf nitrogen content by using partial least squares regression (PLS), principal component regression (PCR), and stepwise multiple regression (SMR) based on NIR spectroscopy. Lee *et al.* (2020) pointed out that the accuracy of the nitrogen concentration monitoring model for cotton leaves based on the ratio of red edge position and short wave near-infrared band was relatively high. Wang *et al.* (2021) believed that leaf structure parameters (leaf thickness) and dry matter content (protein, cellulose, and lignin) lead to high reflectivity and reflectivity fluctuations in the near-infrared and short-wave infrared regions and that 1500-1900 nm and 2000-2400 nm in the short-wave infrared region are the main wavelengths of nitrogen.

Using a field-portable spectrometer with spectral range of 350-2500 nm, researchers have established a spectral index based on the composition of the visible spectral domain and the near-infrared spectral domain to estimate the total nitrogen concentration of winter wheat leaves (Duan *et al.*, 2019). Wang *et al.* (2021) believed that the spectral signal in the short-wave infrared region is closely related to the nitrogen concentration in the leaves, but only the effects of chlorophyll (visible light region) and nitrogen concentration (short-wave infrared region) on the  $V_{max}$  of photosynthesis were studied. Shi *et al.* (Shi *et al.*, 2015) established the normalized difference spectral index (NDSI) and the three-band spectral index (TBSI) of five varieties using wavelengths in the spectral range of 400-2400 nm in rice, corn, tea, sesame, and soybean. The bands used involved three spectral domains of visible, near-infrared, and short-wave infrared, and the authors believed that species-specific differences could affect the hyperspectral estimation of crop leaf nitrogen concentration. Kennedy *et al.* (2013) studied the biochemical components of birch and eucalyptus through near-infrared (1100-2300 nm) spectral information. The results showed that the relative prediction standard deviation (RPD) value of the established nitrogen model was 3.90-4.78, showing excellent model performance ( $RPD \geq 3$ , indicating that the model had a favourable prediction effect).

The research conducted thus far is mainly based on the estimation and analysis of crop nutrient components in a few sensitive bands, such as considering only a single spectral region of visible light, near-infrared, and short-wave infrared, or using vegetation indices involving two spectral regions to construct models. Studies have not fully accounted for the interaction effects between the spectral domains, nor have they fully taken advantage of the detailed spectral information provided by hyperspectral equipment. In this

study, we measured the nitrogen content of main-stem functional leaves at the flower and boll stage of cotton in drip irrigation. The high spectral range of 350-2500 nm was divided into three spectral regions: visible (VIS), near-infrared (NIR), and short-wave infra-red (SWIR). By screening the characteristic wavelengths of the three spectral regions, the estimation accuracy of the high spectral regions of cotton leaves was analysed by multiple linear regression (MLR), principal component regression (PCR), and partial least square regression (PLSR) to explore and analyse the potential of the three hyperspectral regions in estimating nitrogen content in cotton leaves, and to provide a new idea for monitoring crop nutrition information.

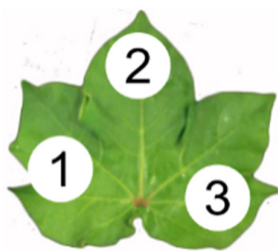
## Materials and Methods

### *Experimental design*

The test was conducted in the second company (85°59'41" E, 44°19'54" N) of the teaching experimental field of Shihezi University in Shihezi City, Xinjiang, China, from April to September 2020. The sunlight duration was 2721-2818 h. The active cumulus temperature of  $\geq 0$  °C was 4023-4118 °C, and that of  $\geq 10$  °C was 3570-3729 °C with a frost-free period of 168-171 d. The soil texture of the experiment was loam. The soil nutrients of the plough horizon (0-20 cm) were as follows: organic matter 19.06 g·kg<sup>-1</sup>, total nitrogen content 12.8 mg·kg<sup>-1</sup>, available potassium 165.1 mg·kg<sup>-1</sup>, available phosphorus 20.8 mg·kg<sup>-1</sup>. In order to obtain cotton leaves with different nitrogen content, five nitrogen levels-N0 (0 kg·ha<sup>-1</sup>), N120 (120 kg·ha<sup>-1</sup>), N240 (240 kg·ha<sup>-1</sup>), N360 (360 kg·ha<sup>-1</sup>), and N480 (480 kg·ha<sup>-1</sup>) - were set up in the experiment. All nitrogen fertilizer (urea, nitrogen content of 46%) was applied with water 8 times during the whole growth period of cotton. Potassium phosphate fertilizer (potassium dihydrogen phosphate, 150 kg·ha<sup>-1</sup>) was applied once before sowing cotton.

### *Acquisition of hyperspectral data of cotton leaves*

The main stem function leaf is an important source organ of cotton. Six cotton plants were selected for each treatment, and 226 samples were obtained. The hyperspectral instrument adopted was the sr-3500 portable ground object spectrometer of the Spectral Evolution Company of the United States, with the spectral range of 350-2500 nm. The spectrometer has a built-in light source and leaf clip. When obtaining hyperspectral data, we avoided the vein of cotton leaves. Three positions of each cotton leaf were measured, and each position was repeated three times. The average value was taken as the spectral data of the position, and the average reflectivity of the three positions was taken as the spectral data of the leaf (Figure 1). Before measuring different cotton leaves, the spectrometer was calibrated with the whiteboard.



**Figure 1.** Hyperspectral data acquisition location of cotton leaves (numbers 1, 2, and 3 represent the order and position of measurement)

*Determination of LNC*

The nitrogen content of cotton leaves was determined by the Kjeldahl method. The main stems and leaves of cotton were put into an envelope and dried in an oven at 105 °C for 30 min, and then dried at 85 °C until attaining a constant mass. The dried sample was crushed and then sieved through 100 mesh.

We took 0.1000 g of the sample with one-ten-thousandth balance and put it into the digesting tube. Then, it was digested with H<sub>2</sub>SO<sub>4</sub>-H<sub>2</sub>O<sub>2</sub>. The volume of digestive juice was fixed to 50 ml, and 10 ml was put into the Kjeldahl nitrogen analyzer for distillation, using sulfuric acid for titration. The nitrogen content of leaves was calculated according to the following formula:

$$LNC(g \cdot kg^{-1}) = \frac{c \times (v - v_0) \times 0.014 \times ts}{m \times 10^{-3}} \quad (1)$$

where  $c$  is the concentration of the dilute sulfuric acid solution (mol/L);  $v$  and  $v_0$  are the volumes of dilute sulfuric acid used in titrating sample solution and the volume of dilute sulfur used in titrating blank, respectively (ml); 0.014 is the molar mass of nitrogen (kg·mol<sup>-1</sup>);  $ts$  is the separation multiple, the ratio of constant volume to separated volume; 10<sup>-3</sup> is the conversion multiple between kilograms and grams; and  $m$  is the mass of the weighted sample (g).

*Selection of characteristic wavelength of the LNC in cotton*

According to previous studies, hyperspectral bands are divided into three spectral regions: visible region (VIS, 380-700 nm), near-infrared region (NIR, 700-1300 nm), and short-wave infrared region (SWIR, 1300-2500 nm). The VIS region contains 321 wavelengths, the NIR region contains 600 wavelengths, and the SWIR region contains 1200 wavelengths. Because each spectral region contains substantial redundant information and multicollinear data, to reduce the number of input variables and improve the modelling accuracy, we used the successive projections algorithm (SPA) to select the characteristic wavelength of the blade. The SPA can find the least redundant variables from the spectral information, effectively eliminate the collinearity between variables, minimize the collinearity between variables, and reduce the complexity of the model. The SPA of this study was completed in MATLAB R2014b.

*Establishment and evaluation of multiple linear regression model*

The model was constructed using multivariate linear regression (MLR), partial least square regression (PLSR), and principal component regression (PCR) in The Unscrambler X 10.4 software based on feature wavelength screening.

The correction set determination coefficient (R<sup>2</sup><sub>c</sub>), prediction set determination coefficient (R<sup>2</sup><sub>p</sub>), correction root mean square error (RMSEC), and prediction root mean square error (RMSEP) were used to quantify the accuracy and thereby evaluate the model. A favourable model should have higher R<sup>2</sup><sub>c</sub> and R<sup>2</sup><sub>p</sub> and lower RMSEC and RMSEP. The model was validated by the normalized root mean square error (n-RMSE), and then the 1:1 histogram of simulated and measured values was drawn to show the fitting and reliability of the model. When n-RMSE is less than 10%, the simulation performance of the model is excellent; when n-RMSE is (10%, 20%), the simulation performance of the model is good; when n-RMSE is (20%, 30%), the simulation performance of the model is average; when n-RMSE is more than 30%, the simulation performance is poor (Ma *et al.*, 2018; Duan *et al.*, 2019).

**Results***Statistical analysis of the LNC in cotton*

In this study, before the establishment of the hyperspectral inversion model of the LNC in cotton, the total samples were divided into the correction set and verification set by using the content gradient method. The content gradient method is a method of sample selection—that is, the samples are arranged in order of

composition content from large to small; the prediction set obtains every two data points, and the other data comprise the correction set (Yu *et al.*, 2018). There were 151 samples in the calibration set and 75 samples in the prediction set. The statistical parameters are shown in Table 1. The results show that the calibration set is consistent with the overall sample data range distribution, and contains the numerical range of the prediction set samples which can be used to establish and verify the model. The results show that the calibration set is consistent with the range of the total sample data, and contains the data range of the prediction set, which can be used for modelling and verification of the model.

**Table 1.** Statistical characteristics of samples

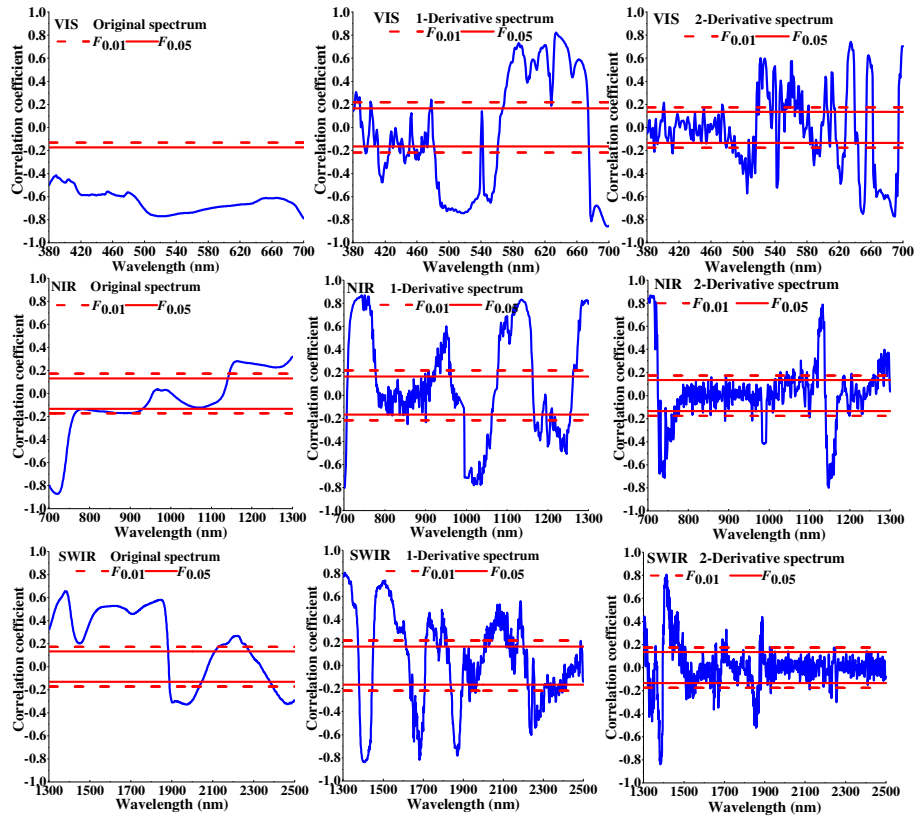
Treatment	Samples	Mean	Max	Min	SD	CV (%)
Total	226	24.18	35.88	7.66	7.69	31.81
Calibration set	151	24.20	35.88	7.66	7.71	31.85
Prediction set	75	24.13	35.59	8.03	7.65	31.74

*The correlation between spectral reflectance of different pretreatment and the LNC*

The derivative is a common spectral pre-processing method that can effectively eliminate the influence of linear background noise and baseline drift and improve the spectral resolution and sensitivity. The correlation between the original spectrum, the first derivative spectrum, and the second derivative spectrum and the LNC in cotton was studied (Figure 2). There was a negative correlation between the original spectral reflectance in the visible region and the nitrogen content in cotton leaves, and both reached an extremely significant level. The correlation at 507-541 nm and 697-700 nm was above 0.75 ( $F < 0.01$ ), indicating that the increase in the LNC would reduce the spectral reflectance of visible light, which is mainly because the increase in leaf nitrogen content would increase the chlorophyll content of leaves, and then increase the light absorption of leaves. After the first derivative transformation, there was a significant negative correlation between the spectral reflectance and the LNC at 567-673 nm, and the significance was more than 0.80 ( $F < 0.01$ ) at 633-636 nm, and more than 0.85 ( $F < 0.01$ ) at 697-700 nm (Figure 2, VIS-1-Derivative spectrum). After the second derivative treatment, although there were many bands with significant correlations with leaf nitrogen content, the correlation was greater 0.75 only at the four wavelengths of 650 nm in red valley and 688-690 nm in the red edge region, indicating that the correlation between leaf nitrogen concentration and spectral reflectance was different under different spectral pre-treatments. For the VIS, the first derivative spectrum can enhance the correlation between some spectral bands and leaf nitrogen content.

In the near-infrared spectral domain (NIR), there was a significant negative correlation between the original spectral reflectance and leaf nitrogen concentration at 701-763 nm, and a maximum negative correlation at 720 nm ( $r = -0.871$ ,  $F < 0.01$ ). This band range is the red edge region of leaf reflectance spectrum and the sensitive region of leaf structure and physicochemical parameter change response. The correlation between the first derivative spectrum and leaf nitrogen content changed rapidly from an extremely significant negative correlation to an extremely significant positive correlation in this band, and the correlation was more than 0.75 at 720-763 nm and reached the maximum positive correlation at 743 nm ( $r = -0.868$ ,  $F < 0.01$ ). Moreover, the second derivative spectra showed an extremely significant positive correlation with leaf nitrogen content in the range nm bands 924-961, 1077-1162, and 1262-1300, and an extremely significant negative correlation in the range of bands 701-706, 996-1062, 1167-1186, 1197-1205, and 1212-1255. These bands are due to the peaks and troughs of the high reflection platform in the near-infrared spectral region. It is considered that the first derivative spectrum will enhance the relationship between the reflectivity of the peak and trough range of the spectral curve and the nitrogen content of the blade. The correlation between the second derivative of the NIR spectrum and leaf nitrogen content reached the maximum positive correlation at 714 nm ( $r = -0.866$ ,  $F < 0.01$ ). There is an extremely significant positive correlation at 1110-1137 nm and an extremely significant negative correlation at 1141-1168 nm. This phenomenon is also caused by the peaks and troughs of the high reflection platform.

There are also several peaks and troughs in the SWIR due to the influence of water absorption and chemical bonds of proteins and other substances. At 1450, 1650, and 1900 nm, the inflection points of the SWIR infrared spectrum region were positive and had a negative correlation with leaf nitrogen content. From the correlation between the LNC and spectral reflectance in the SWIR in Figure 2, it can be seen that derivative can enhance the spectral correlation more obviously, and the correlation range between the original spectrum and leaf nitrogen content is  $-0.329$  to  $0.655$ . The correlation between the first derivative and leaf nitrogen content ranged from  $-0.817$  to  $0.838$ , with the maximum positive correlation at 1306 nm and the maximum negative correlation at 1682 nm. The correlation between the second derivative and leaf nitrogen content ranged from  $-0.839$  to  $0.807$ , with the maximum positive correlation at 1410 nm and the maximum negative correlation at 1383 nm.



**Figure 2.** The correlation between the reflectance of the three spectral regions and the LNC in cotton ( $F_{0.05}$ : significant correlation ( $P < 0.05$ );  $F_{0.01}$ : extremely significant correlation ( $P < 0.01$ ))

#### *Feature wavelength selection based on SPA*

SPA was used to select the characteristic wavelength of the original spectrum, the first derivative spectrum, and the second derivative spectrum. In VIS, there were only two characteristic wavelengths selected by the second derivative spectrum, and the two characteristic wavelengths were in the red region. Seven characteristic wavelengths were selected in the blue, green, and red bands of the original spectrum and the first derivative spectrum. Among the selected characteristic wavelengths, the wavelengths of 450 nm and 459 nm, 544 nm and 561 nm, and 674 nm and 676 nm were near the green reflection peak, blue reflection valley, and red reflection valley, respectively, indicating that the selected characteristic wavelengths conformed to the absorption characteristics of chlorophyll in visible light to a certain extent, and wavelengths of 693 nm, 699 nm, and 700 nm were at the red edge. The selected characteristic wavelengths significantly correlated with leaf nitrogen content.

In the near-infrared region, the higher reflection platform is formed mainly due to the influence of leaf structures such as mesophyll tissue. In this region, 19 characteristic wavelengths were selected. In the red edge position range of 700–760 nm, there are two characteristic wavelengths selected by the original spectrum and the first derivative spectrum, and one characteristic wavelength selected by the second derivative spectrum. The wavelengths of 812 nm, 842 nm, 845 nm, 880 nm, and 885 nm may be related to the fourth harmonic generation of N-H bonds. The wavelengths of 913 nm, 935 nm, 938 nm, 950 nm, 953 nm, and 957 nm may be related to the fourth harmonic generation of C-H bonds in leaves. The wavelengths of 975 nm, 976 nm, 977 nm, 995 nm, and 1000 nm may be related to the elongation of N-H bonds. SWIR is mainly the spectral region of water absorption, but the absorption of protein is also in this band. In this region, the second derivative spectrum had the most characteristic wavelengths, but only two wavelengths had a significant relationship with leaf nitrogen content. The characteristic wavelengths of 1501 nm, 1977 nm, 2145 nm, and 2308 nm were related to the first overtone and elongation of the N-H bond, and the characteristic wavelengths of 1712 nm, 1748 nm, 1723 nm, and 2405 nm were related to the first overtone and elongation of the C-H bond. The C-H bond and N-H bond are the main components of protein.

**Table 2.** Characteristic wavelengths of different spectrum regions

Spectral region	Pretreatment method	Numbers	Characteristic wavelength (nm)
VIS (380–700 nm)	original spectrum	7	459**, 475**, 508**, 544**, 674**, 693**, 700**
	first derivative spectrum	5	390, 450**, 561*, 676**, 699**
	second derivative spectrum	2	693**, 698**
NIR (701–1300 nm)	original spectrum	19	744**, 756**, 812*, 880*, 913*, 935, 938, 950, 953, 957, 965, 966, 968, 975, 976, 977, 995, 1201**, 1300**
	first derivative spectrum	6	706**, 730**, 885, 963*, 964, 975
	second derivative spectrum	5	707**, 842, 975, 1000, 1249
SWIR (1301–2500 nm)	original spectrum	4	1301**, 1374**, 1712**, 2405**
	first derivative spectrum	5	1397**, 1748**, 1945, 1977, 2496
	second derivative spectrum	7	1334**, 1380**, 1501, 1723, 2145, 2308, 2482

Notes: \*\* and \* represent correlations significant at the probability level of  $P < 0.01$  and  $P < 0.05$ , respectively.

#### *Establishment and validation of cotton leaf nitrogen content estimation model based on different hyperspectral regions*

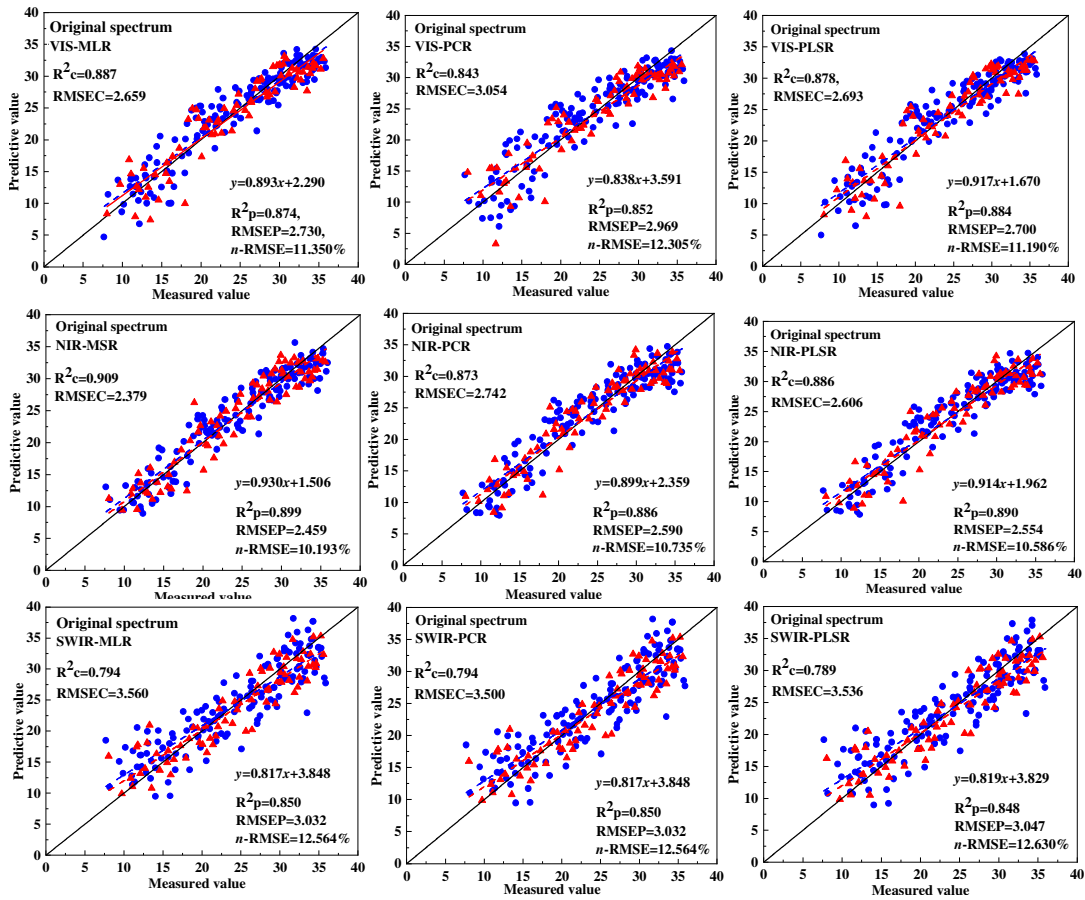
Using the characteristic wavelengths selected by the continuous projection algorithm (Table 2), the estimation model of cotton leaf nitrogen content was established based on simple multiple linear regression (MLR), principal component regression (PCR), and partial least squares regression (PLSR). The model parameters are shown in Figures 3-5. In the constructed model, the best estimation models of different regression models in different spectral domains were obtained based on the determination coefficient ( $R^2$ ) and the optimal principle of n-RMSE.

#### *Establishment and validation of the LNC in cotton estimation model based on original spectrum*

Based on the characteristic wavelengths screened by the original spectral reflectance, the LNC estimation model was built (Figure 3). By comprehensively comparing the modeling effects of the three spectral regions, the MLR model established in the NIR had the highest accuracy ( $R^2_c = 0.909$ ,  $R^2_p = 0.899$ ) due to more characteristic wavelengths), and the model constructed in the SWIR had the lowest accuracy ( $R^2_c = 0.789$ ,  $R^2_p = 0.848$ ). In the VIS and NIR domain, the accuracy of the PCR model was relatively low, and the  $R^2_c$  values of the model were 0.843 and 0.873, respectively. In the SWIR, the accuracy of the model based on the PLSR was slightly lower, and there were no differences in the model parameters of  $R^2$ , RMSE, and n-RMSE



based on the MLR and PCR. This may be due to the significant correlation between the selected characteristic wavelengths and leaf nitrogen content, and the less characteristic wavelengths. Using the characteristic wavelengths of three spectral regions, the n-RMSE of LNC in the cotton estimation model based on the original spectrum was 10.193-12.630%, indicating that the accuracy of the models was sufficient.

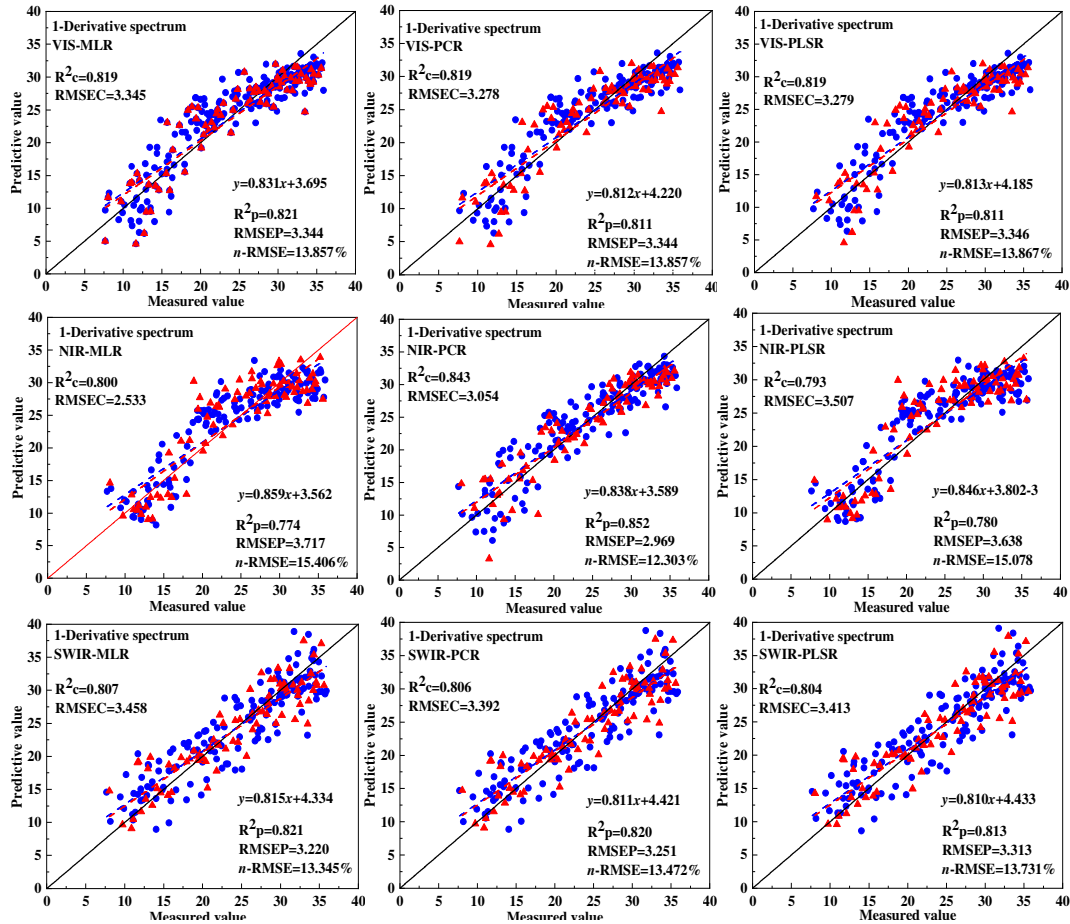


**Figure 3.** Regression model parameters of cotton LNC in different spectral regions based on original spectral characteristic wavelengths

*Establishment and verification of estimation model based on first derivative spectrum*

According to the estimation model established by analysing the characteristic wave-length of the first derivative spectrum (1-Derivative spectrum) (Figure 4), the model constructed by PCR in the NIR domain had the highest accuracy of  $R^2c$  and  $R^2p$ , which were 0.843 and 0.852, respectively. In the VIS and SWIR, the accuracy of  $R^2c$  and  $R^2p$  of the three methods was above 0.8. The accuracy of the model in the SWIR was increased.  $R^2c$  values were 0.807, 0.806, and 0.804, respectively; and  $R^2p$  values were 0.821, 0.820, and 0.813, respectively, indicating that the first derivative spectrum increases the accuracy of the model in the SWIR. For the validation model, the n-RMSE ranged from 12.303% to 15.406%. Compared with the validation model of the original spectrum, the n-RMSE value increased, and the model accuracy was low.



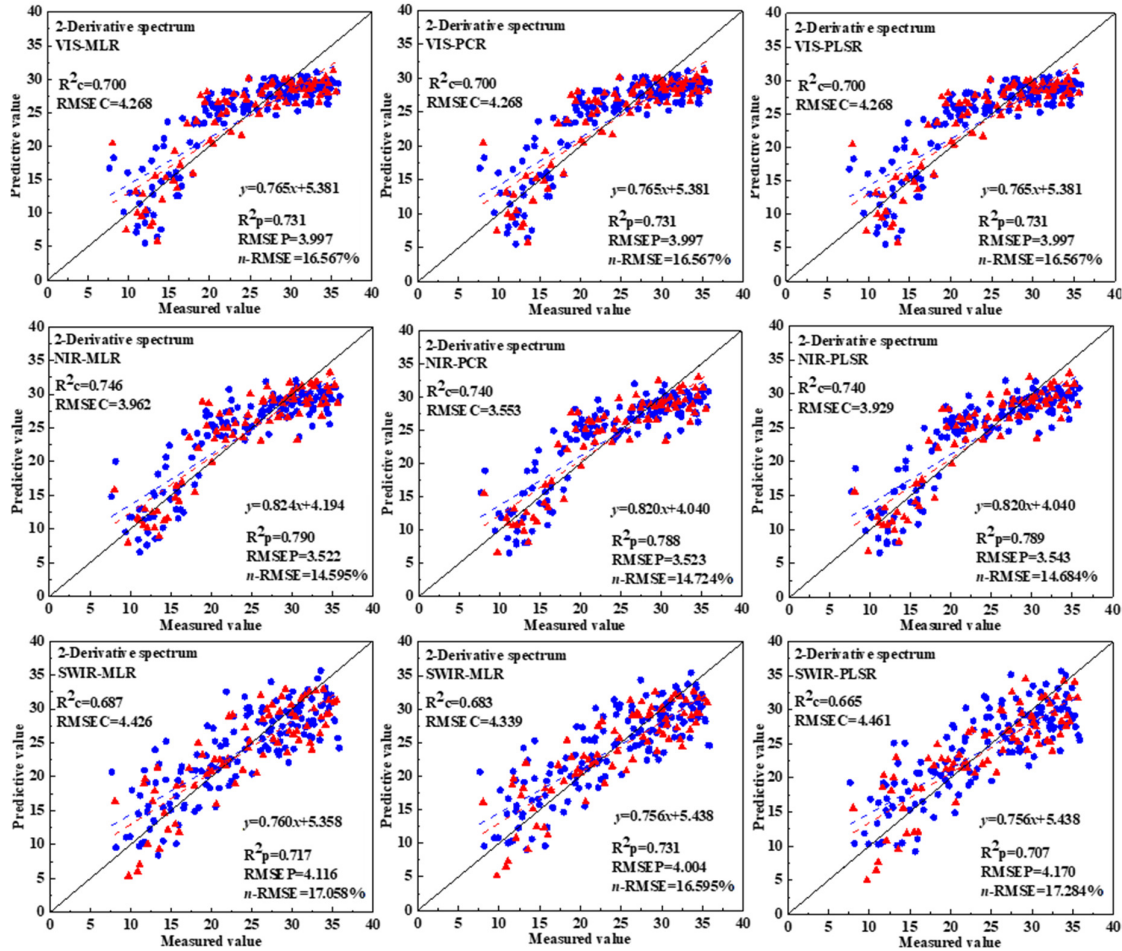


**Figure 4.** Regression model parameters of cotton LNC in different spectral regions based on first derivative spectrum characteristic wavelengths

*Establishment and verification of estimation model based on second derivative spectrum*

In the estimation model of the LNC using second derivative spectral reflectance (2-Derivative spectrum, Figure 5), the  $R^2_c$  and  $R^2_p$  values established by MLR, PCR, and regression methods in three spectral domains did not reach 0.8, and the values of  $R^2_p$  were greater than  $R^2_c$ . The accuracy of the MLR model was also the best in the NIR domain ( $R^2_c = 0.746$ ,  $R^2_p = 0.790$ ). Although the n-RMSE value of the validation model was 10-20%, it was the highest among the three types of data, indicating that the modelling effect using the second derivative spectra in the data processing method of three was worse than the other two types of data.

Through the analysis of the above results, when estimating the LNC by using different spectrum regions, the model based on the original spectrum in the VIS and NIR was superior, along with the effect of the model based on the first derivative spectrum in the SWIR. In contrast, the model established by using the second derivative spectrum in the three spectrum regions was poor. The optimal values of  $R^2_c$ ,  $R^2_p$ , and n-RMSE are the multiple linear regression models based on the original spectrum in the NIR. Perhaps due to the small number of the characteristic wavelengths selected, the accuracy of the PCR and PLSR models was not better than that of multiple linear regression models in the 27 estimation models of the LNC in cotton leaves.



**Figure 5.** Regression model parameters of cotton LNC in different spectral regions based on second derivative spectrum characteristic wavelengths

## Discussion

The application of spectrum regions to estimate the biochemical components of crops started at the earliest based on leaf spectra, mainly because the leaf spectra were not affected by crop canopy structure, soil background, or the monitoring environment, and the obtained spectral data were relatively ideal (Hasituya *et al.*, 2020; Lee *et al.*, 2020). In this study, the LNC of cotton was researched. Based on the VIS (380-700 nm), NIR (701-1300 nm), and SWIR (1301-2500 nm) of hyperspectral imaging, we studied the potential of three spectrum regions to estimate the LNC of crops, respectively.

The VIS is affected by the content of the chlorophyll, while the NIR band is mainly affected by the change in the spectral albedo caused by the structure of the leaves. In this study, the characteristic wavelength in the visible spectrum was mainly near the reflection peak or valley of red, green, and blue light and the range of the red edge position. The characteristic wavelength of the NIR was mainly related to the activities of N-H bonds and C-H bonds, which is consistent with previous research results, indicating that they have universal applicability as the key band for the study of plant spectral reflection (Cochrane, 2000; Liu *et al.*, 2011; Rubert-Nason *et al.*, 2013; Fan *et al.*, 2019).

In the VIS and NIR, it is considered that the accuracy of the model established by using the characteristic wavelength of the original spectrum and the first derivative spectrum is higher, which is consistent with

previous research results. In this study, the characteristic wavelengths in the visible spectrum were mainly near the reflection peak or valley of red, green, and blue light and the range of the red edge position. These results indicate that the information reflected by the leaves can be more retained by the uncomplicated data transformation in the estimation of biochemical components of crops by using the hyperspectral spectrum information of the leaves (Cochrane, 2000; Rubert-Nason *et al.*, 2013; Fan *et al.*, 2019; Zhang *et al.*, 2020). In the VIS and NIR, the accuracy of the model established by using the characteristic wavelengths of the original spectrum and the first derivative spectrum is higher, which is consistent with previous research results. These results indicate that the information reflected by the leaves can be more retained by the original spectrum or the uncomplicated data transformation in the estimation of bio-chemical components of crops by using the hyperspectral spectrum information of the leaves (Yi *et al.*, 2010; Fan *et al.*, 2019).

The SWIR is mainly the reflection curve related to the chemical bond (C-H, N-H, and O-H bonds) interactions and various biochemical nutrition characteristics of crops (Rubert-Nason *et al.*, 2013; Zhang *et al.*, 2018). For the SWIR, the characteristic wave-lengths selected based on the original spectrum have an extremely significant correlation with the LNC in cotton, which is consistent with the performance results of the VIS and NIR (Table 2). In terms of the characteristic wavelengths considering first derivative and second derivative screening, the characteristic wavelengths of 1301-1397 nm were related to the second overtone of the C-H and O-H bonds in the protein, the characteristic wave-lengths of 1712-1748 nm were mainly generated by the first overtone absorption of the C-H bond, and the characteristic wavelengths of 1977-2308 nm were related to the first overtone and extension of the N-H bond, which is consistent with the research results of Zhang *et al.* (2018) on the protein. Comparing and analysing the models established by the characteristic wavelength of three spectrum types, the  $R^2c$  and  $R^2p$  values of each regression model established by the first derivative were the highest.

By comparing and analysing the three modelling methods for the application of three spectrum domains, this study shows that PCR and PLSR have no advantage over MLR, which is different from the research results of Yi *et al.* (2010) and Zhang *et al.* (Zhang *et al.*, 2020). This may be due to the small number of characteristic wavelengths selected in this study, which makes PCR and PLSR unable to reduce and re-screen the characteristic wavelengths. The n-RMSE of the models established based on the three modelling methods were all in the range of 10-20%, indicating that the models by dividing the whole band into three spectrum domains and using various modelling methods were relatively stable and with suitable prediction accuracy, and could be applied to the estimation of LNC in cotton.

## Conclusions

In this study, we researched cotton leaf samples in the range of 7.66-38.88 g/kg. Through the correlations between the wavelength of three spectrum regions (VIS, NIR, and SWIR) and the LNC and the comparative analysis of the models constructed, the results showed that (1) the first derivative pre-treatment method can enhance the correlation between spectral band and leaf nitrogen content; (2) the multiple linear regression and principal component regression model based on the characteristic band screened by original spectrum and first derivative spectrum is more suitable for estimating leaf nitrogen content, with  $R^2c = 0.794 \sim 0.909$  and  $R^2p = 0.774 \sim 0.899$ ; (3) the model validation accuracy n-RMSE of the estimation models established by using VIS, NIR and SWIR is 10-20%, indicating that the three spectral regions have the potential to estimate the nitrogen content of cotton leaves, which provides a new idea for more comprehensive estimation of crop nutrient information by using hyperspectral data.

### Authors' Contributions

JZ and LLM were responsible for determination of some indexes and the writing of this manuscript. XYC and JL performed the experiments and collected all data sets. CXY and QSY were responsible for the mapping and typesetting. ZZ and XL, the corresponding author, were responsible for the revision and quality control of the paper. All authors read and approved the final manuscript.

### Ethical approval (for researches involving animals or humans)

Not applicable.

### Acknowledgements

This work was supported by the National Natural Science Foundation of China, China (Grant No. 42061058), Science and Technology Research Plan for Key Areas of XPCC, China (Grant No. 2020AB005) and Major scientific and technological projects of XPCC, China (Grant No. 2018AA00407).

### Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

### References

- Berger K, Verrelst J, Féret JB, Wang Z, Woche M, Strathmann M, ... Hank T (2020). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment* 242:111758. <https://doi.org/10.1016/j.rse.2020.111758>
- Chapin FS, Bloom AJ, Field CB, Waring RH (1987). Plant responses to multiple environmental factors. *Bioscience* 37(1):49-57. <https://doi.org/10.2307/1310177>
- Cochrane MA (2000). Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensing* 21(10):2075-2087. <https://doi.org/10.1080/01431160050021303>
- Curran PJ (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment* 30(3):271-278. [https://doi.org/10.1016/0034-4257\(89\)90069-2](https://doi.org/10.1016/0034-4257(89)90069-2)
- Duan DD, Zhao CJ, Li ZH, Yang GJ, Yang WD (2019). Estimating total leaf nitrogen concentration in winter wheat by canopy hyperspectral data and nitrogen vertical distribution. *Journal of Integrative Agriculture* 18(7):1562-1570. [https://doi.org/10.1016/S2095-3119\(19\)62686-9](https://doi.org/10.1016/S2095-3119(19)62686-9)
- Fan LL, Zhao JL, Xu XG, Liang D, Yang GJ, Feng HK, ... Wei PF (2019). Hyperspectral-based estimation of leaf nitrogen content in corn using optimal selection of multiple spectral variables. *Sensors* 19(13):2898. <https://doi.org/10.3390/s19132898>
- Hasituya, Li F, Elsayed S, Hu Y, Schmidhalter U (2020). Passive reflectance sensing using optimized two- and three-band spectral indices for quantifying the total nitrogen yield of maize. *Computers and Electronics in Agriculture* 173:105403. <https://doi.org/10.1016/j.compag.2020.105403>
- Homolová L, Malenovský Z, Clevers JGPW, García-Santos G, Schaepman ME (2013). Review of optical-based remote sensing for plant trait mapping. *Ecological Complexity* 15:1-16. <https://doi.org/10.1016/j.ecocom.2013.06.003>
- Hong S, Zhang Z, Zhang LF, Ma LL, Hai XY, Wang Z, ... Lv X (2019). Hyperspectral estimation model of chlorophyll content in cotton canopy leaves under drip irrigation at different growth stages. *Cotton Science* 31(2):138-146. <https://doi.org/10.11963/1002-7807.hslx.20190319>

- Hosgood B, Jacquemoud S, Andreoli G, Verdebout J, Pedrini G, Schmuck G (1994). Leaf optical properties experiment 93 (LOPEX93), in: Leaf Optical Properties Experiment 93 (LOPEX93). Institute for Remote Sensing Applications, European Commission-Joint Research Centre.
- Kokaly RF, Asner GP, Ollinger SV, Martin ME, Wessman CA (2009). Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sensing of Environment, Imaging Spectroscopy Special Issue* 113:S78-S91. <https://doi.org/10.1016/j.rse.2008.10.018>
- le Maire G, Francois C, Soudani K, Berveiller D, Pontauiller JY, Nathalie B, ... Dufrêne E (2008). Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sensing of Environment* 112(10):3846-3864. <https://doi.org/10.1016/j.rse.2008.06.005>
- Lee DH, Shin HS, Park JH (2020). Developing a p-NDVI map for highland kimchi cabbage using spectral information from UAVs and a field spectral radiometer. *Agronomy* 10(11):1798. <https://doi.org/10.3390/agronomy10111798>
- Liu F, Jin ZL, Naeem MS, Tian T, Zhang F, He Y, ... Zhou WJ (2011). Applying near-infrared spectroscopy and chemometrics to determine total amino acids in herbicide-stressed oilseed rape leaves. *Food Bioprocess Technology* 4(7):1314-1321. <https://doi.org/10.1007/s11947-010-0445-y>
- Ma LL, Lv X, Zhang Z, Ma GX, Hai XY (2018). Establishment of nitrogen nutrition diagnosis model for drip-irrigation cotton based on critical nitrogen concentration. *Transactions of the Chinese Society for Agricultural Machinery* 49(2):277-283. <https://doi.org/10.6041/j.issn.1000-1298.2018.02.035>
- Ohyama T (2010). Nitrogen as a major essential element of plants. In: *Nitrogen Assimilation in Plants*. Research Signpost, pp 1-18.
- Rubert-Nason KF, Holeski LM, Couture JJ, Gusse A, Undersander DJ, Lindroth RL (2013). Rapid phytochemical analysis of birch (*Betula*) and poplar (*Populus*) foliage by near-infrared reflectance spectroscopy. *Analytical and Bioanalytical Chemistry* 405(4):1333-1344. <https://doi.org/10.1007/s00216-012-6513-6>
- Serrano L, Filella I, Penuelas J (2000). Remote sensing of biomass and yield of winter wheat under different nitrogen supplies. *Crop Science* 40:723-731. <https://doi.org/10.2135/cropsci2000.403723x>
- Shi TZ, Wang JJ, Liu HZ, Wu GF (2015). Estimating leaf nitrogen concentration in heterogeneous crop plants from hyper-spectral reflectance. *International Journal of Remote Sensing* 36(18):4652-4667. <https://doi.org/10.1080/01431161.2015.1088676>
- Spaner D, Todd AG, Navabi A, McKenzie DB, Goonewardene, LA (2005). Can leaf chlorophyll measures at differing growth stages be used as an indicator of winter wheat and spring barley nitrogen requirements in Eastern Canada? *Journal of Agronomy and Crop Science* 191(5):393-399. <https://doi.org/10.1111/j.1439-037X.2005.00175.x>
- Tarpley L, Reddy KR, Sassenrath-Cole GF (2000). Reflectance indices with precision and accuracy in predicting cotton leaf nitrogen concentration. *Crop Science* 40(6):1814-1819. <https://doi.org/10.2135/cropsci2000.4061814x>
- Wang HG, Guo ZJ, Shi Y, Zhang YL, Yu ZW (2015). Impact of tillage practices on nitrogen accumulation and translocation in wheat and soil nitrate-nitrogen leaching in drylands. *Soil and Tillage Research* 153:20-27. <https://doi.org/10.1016/j.still.2015.03.006>
- Wang S, Guan KY, Wang ZH, Ainsworth EA, Zheng T, Townsend PA, ... Jiang CY (2021). Unique contributions of chlorophyll and nitrogen to predict crop photosynthetic capacity from leaf spectroscopy. *Journal of Experimental Botany* 72(2):341-354. <https://doi.org/10.1093/jxb/eraa432>
- Xu YJ, Liu X, Cao X, Huang CP, Liu E, Qian S, ... Qiu CW (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation* 2(4):100179. <https://doi.org/10.1016/j.xinn.2021.100179>
- Yi SH, Deng L, He SL, Zheng YQ, Zhang X (2010). Research on nitrogen content of leaf of Jincheng orange cultivar using visible near infrared spectroscopy model. *Journal of Fruit Science* 27(1):13-17. <https://doi.org/10.13925/j.cnki.gsx.2010.01.007>
- Yu L, Zhang T, Zhu YX, Zhou Y, Xia T, Nie Y (2018). Determination of soybean leaf SPAD value using characteristic wavelength variables preferably selected by IRIV algorithm. *Transactions of the Chinese Society of Agricultural Engineering* 34(16):148-154. <https://doi.org/10.11975/j.issn.1002-6819.2018.16.019>
- Zhang YK, Luo B, Song P, Pan DY, Lu WC, Zhou YN, ... Zhao CJ (2018). Rapid determination of soluble protein content for soybean leaves based on near infrared spectroscopy. *Transactions of the Chinese Society of Agricultural Engineering* 34(18):187-193. <https://doi.org/10.11975/j.issn.1002-6819.2018.18.023>

- Zhang Y, Wang L, Bai Y, Yang L, Lu Y, Sun Y, Zhang J, Li G (2020). Establishment and variation of soil inorganic nitrogen content estimation model based on maize leaf spectral characteristics. *Journal of Plant Nutrition and Fertilizers* 26:1206-1215. <https://doi.org/10.11674/zwyf.19466>
- Zhang YS, Yao X, Tian YC, Cao, WX, Zhu Y (2010). Estimating leaf nitrogen content with near infrared reflectance spectroscopy in rice. *Chinese Journal of Plant Ecology* 34:704-712. <https://doi.org/10.3773/j.issn.1005-264x.2010.06.010>
- Zhang Z, Ma LL, Hong S, Lin J, Zhang LF, Lv X (2020). Study on hyperspectral diagnosis of nitrogen nutrition index among different cotton varieties under drip irrigation. *Cotton Science* 32:392-403. <https://doi.org/10.11963/1002-7807.zzlx.20200716>



The journal offers free, immediate, and unrestricted access to peer-reviewed research and scholarly work. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose, without asking prior permission from the publisher or the author.



**License** - Articles published in *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* are Open-Access, distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) License.

© Articles by the authors; UASVM, Cluj-Napoca, Romania. The journal allows the author(s) to hold the copyright/to retain publishing rights without restriction.