BIO Web of Conferences **230**, 01003 (2026) https://doi.org/10.1051/bioconf/202623001003
ICAA A 2026

Applications of Derivative Spectroscopy on determination of nitrogen status of tea plant (Camellia sinensis)

Junhong Pan[1, 2], *Wei* Wu[1, 3], *Ronghui* Lu[4], *Wensen* Yu[1, 2], *Tzong-Jer* Chen*[1, 2], *Rengui* Chen[1, 2]

[1]School of Mathematics & Computer Science, Wuyi University Wuyishan, Fujian, 354300, China
[2]Fujian Key Laboratory of Big Data Application and Intellectualization for Tea Industry (Wuyi University), China
[3]College of Computer Science and Technology, Zhejiang University of Technology, Zhejiang, China.
[4]Information Technology and Laboratory Management Center, Wuyi University, Wuyishan, Fujian, 354300, China

Abstract. Nitrogen (N) fertilizer plays an important role in tea plantation with significant impacts on the formation of vital compounds, pigments and flavor-related substances. Excess and deficiency of N can reduce crop yield and quality yet impact environmental concerns. Therefore, it is crucial to tailor fertilization levels based on the N status of tea plants to avoid both inadequate and excessive N application. This study has evaluated if it is possible to improve the effectiveness of spectral indices to estimate the nitrogen concentration using derivative spectroscopy and dimension reduction techniques to process the spectral signatures. The results demonstrated that this study involved spectrum smoothing, derivative implementation, and measurements of Normalized Difference Spectral Index (NDSI) and Ratio Spectral Index (RSI) indices. Analysis identified the top 25 Pearson correlation coefficient (PCC) for both indices and employed Principal Component Analysis (PCA) to determine ten principal component scores. A ten-component linear regression model was developed, and its effectiveness was assessed. This model was then used to predict N levels in tea gardens. Derivative spectroscopy techniques can improve the effectiveness of indices to estimate the nitrogen concentration as compared to non-derivative indices. The statistics values are always higher for derivative indices. This model was then used to predict N levels in tea gardens. To increase Signal to Noise Ratio (SNR), a scheme added three spectra and corresponding N values together then find an average. This procession made very good results on statistics for multiple linear regression. The results are promising, suggesting that further research in this field is needed.

* Corresponding author: s838502@oz.nthu.edu.tw

Keyword: Tea plant; Nitrogen; Multiple Linear Regression; Derivative Spectroscopy; Principal Component Analysis.

1 Introduction

The tea plant (Camellia sinensis) is a major economic crop in China, valued for its health benefits and mental well-being effects, which have led to a rise in cultivation [1]. In southern China, tea is consumed in large quantities, and the plant is widely grown as an important economic resource.

Fertilization is a key aspect of tea plantation management, essential for providing the diverse nutrients necessary for optimal tea plant growth. Among these nutrients, nitrogen (N) is particularly important due to its high demand. It plays a critical role in the formation of vital compounds, including pigments and flavor-related substances [1].

The demand for N by tea crops throughout their development is well-documented. The application of N fertilizer ensures a consistent supply necessary for optimal plant growth. However, both excessive and insufficient nitrogen can negatively impact production, operating costs, and environmental conservation, leading to reduced crop yield and quality [2]. Continuous N fertilizer application affects not only plant N absorption but also raises environmental concerns. Therefore, it is crucial to tailor fertilization levels based on the N status of tea plants to avoid both inadequate and excessive N application [1].

Leaf N content can be used to estimate the overall N level in the whole plant, which can be determined through chemical analysis in a laboratory or by remote sensing methods [1]. Manually methods are time-consuming, destructive to samples, and require specialized laboratory equipment. Consequently, there is a need for rapid, non-destructive, and highly accurate methods to assess N levels. Researchers have developed predictive models by analyzing nitrogen-sensitive spectral bands and constructing vegetation indices strongly correlated with nitrogen levels. Recent advancements have also been made in reconstructing N content in crops using remote sensing data.

Remote sensing monitoring of vegetation nitrogen content primarily relies on hyperspectral or multispectral data. Algorithms for analyzing N levels in crops include methods based on spectral indices, regression analysis, and radiative transfer models [3]. Reflectance measurements were once the most widely used to obtain spectral indices as indicators of chlorophyll concentration (i.e. chlorophyll concentration in crops is closely related to N levels.). However, the effectiveness of these indices for estimating N concentration in plants is limited.

Significant progress has been made in recent years in researching N inversion for major crops such as wheat, corn, and rice using hyperspectral technology. By analyzing nitrogen-sensitive spectral bands and developing vegetation indices strongly correlated with nitrogen levels, researchers have established predictive models that have yielded promising results [3].

Spectral indices related to plant N have become a common approach for monitoring N concentration. Currently, there are over 30 major spectral indices available [3]. These indices are generally categorized into those based on N absorption wavelengths (such as the Normalized Difference Nitrogen Index, NDNI) and those related to chlorophyll content (e.g., the Canopy Chlorophyll Content Index, CCCI; Modified Chlorophyll Absorption in Reflectance Index, MCARI; and Transformed Chlorophyll Absorption in Reflectance Index, TCARI) [4].

Due to the significant correlation between chlorophyll concentration and N content, chlorophyll content indices derived from chlorophyll absorption wavelengths are often used to estimate N levels indirectly. Notable examples include the improved MTARI index proposed by Meng Qing-ye *et al.* [5]. Among these indices, the CCCI is currently widely used for assessing vegetation nitrogen content [3]. Tilling *et al.* were the first to apply this index to airborne hyperspectral data for nitrogen inversion in wheat [6].

In general, spectral indices for vegetation nitrogen content in remote sensing inversion are derived from the vegetation index framework proposed by Rouse [7]. These indices typically take two forms: the normalized index form and the ratio form. The normalized index form is represented by the Normalized Difference Spectral Index (NDSI), which is calculated as:

$$NDSI = \frac{x_2 - x_1}{x_2 + x_1} \quad (1)$$

The ratio form is represented by the Ratio Spectral Index (RSI), calculated as:

$$RSI = \frac{x_2}{x_1} \quad (2)$$

where x_2, x_1 can refer to any specific wavelengths of original reflectance or the relative reflectance after a series of transformations [3].

Linear regression is a commonly used inversion method in remote sensing for assessing vegetation nitrogen content [3]. This approach was adopted in earlier studies to estimate nitrogen levels in crops [2]. Wu *et al.* developed advanced regression models using linear, exponential, logarithmic, and parabolic expressions to relate leaf area index (LAI) to nitrogen content. Among these, the exponential model, which incorporates the vegetation index variable "VI4" (red side area/yellow side area), demonstrated the best modeling performance [8]. In addition to linear regression, other methods such as support vector machines (SVM) and neural networks have been used to estimate nitrogen content in vegetation. Axelsson *et al.* found that SVM regression provided the best results [9]. Huang *et al.* reported that neural network regression offered a better fit compared to stepwise multiple regression, suggesting a potential nonlinear relationship between leaf biochemical content and canopy reflectance [10]. Wang *et al.* showed that the backpropagation neural network (BPN) model for inverting nitrogen content in rape outperformed stepwise linear regression [11]. Despite the emergence of new empirical inversion methods, the simplicity and intuitiveness of linear regression keep it a valuable approach in this field [2].

In spectral analysis, hyperspectral data often require preprocessing before modeling. Common preprocessing schemes include:

1. **Smoothing**: This step reduces noise in the spectral data. The Savitzky-Golay filter is one of the most widely used methods for smoothing in spectral processing.
2. **Spectrum Derivative**: This technique enhances the signal-to-noise ratio (SNR) by reducing noise. First-order differential transformations of spectra can effectively improve data SNR [1].
3. **Principal Component Analysis (PCA)**: PCA is used for dimensionality reduction, data compression, and feature extraction. The first few principal components usually account for a significant portion of the variance in the original data. PCA loadings, which are the weighting coefficients for each wavelength, indicate the relative importance of each wavelength in contributing to the principal components [1-3].

Derivative analysis is particularly promising for use with remote sensing data. Derivatives not only emphasize subtle spectral details; they also minimize illumination and atmospheric effects [12]. A first-order derivative represents the rate of change of reflectance with respect to wavelength. For example, consider a Gaussian curve (Figure 1(a)) used to simulate the original signal. The first-order derivative of this Gaussian curve starts and ends at zero, as shown in Figure 1(b), and passes through zero at the same wavelength as the peak of the Gaussian curve.

In contrast, even-order derivatives are characterized by a strong negative or positive band with a minimum or maximum at the same wavelength as the band's peak, as illustrated in

Figures 1(c) and (e) [12]. For even-order derivatives, there is a peak maximum or minimum at the same wavelength as the Gaussian curve, while for odd-order derivatives, this wavelength corresponds to a zero-crossing point [12].

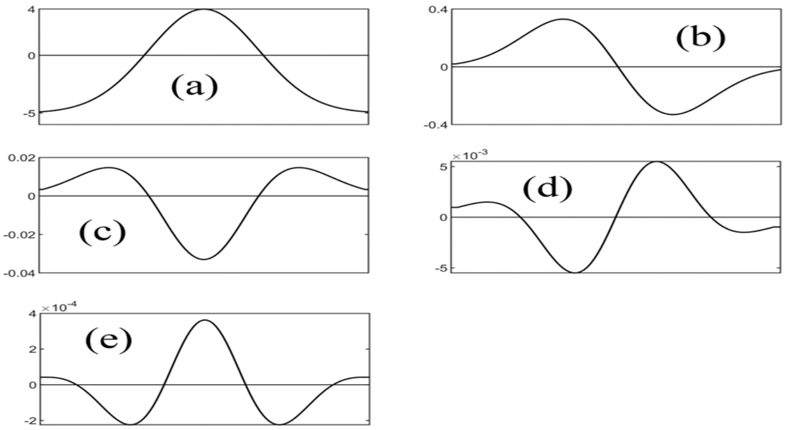

Figure 1. An original Gaussian curve with various version of derivatives, (a)original, (b)first derivative, (c) second derivative, (d) third derivative, (e) fourth derivative.

Studies have shown that the Pearson correlation coefficient (PCC) between N content and the first-order derivative of crop reflectance spectra is higher than that between N content and the original reflectance spectra [13]. Since even-order derivatives of the spectrum exhibit negative or positive bands with minima or maxima at the same wavelengths as the original spectrum, it can be hypothesized that the PCC between N content and crop reflectance spectra may be better for even-order derivatives compared to the first-order derivatives and the original reflectance spectra.

This work has evaluated if it is possible to improve the effectiveness of indices to estimate the N concentration using the specificity of derivative spectroscopy techniques to process the spectral signatures. The canopy reflectance spectra of tea leaf were taken first in Wuyishan and the N concentration of leaf was measured by chemical experiments in laboratory. Following, the NDSI and RSI were measured for both the original spectra and several derivative spectra. Then the PCC between N contents with the NDSI, RSI were calculated respectively. The top PCCs of NDSI and RSI with leaf N contents were collected. These NDSI and RSI with top PCCs correspond to N were collected and then send to make PCA transformation and ten PCA scores were measured. A ten-component linear regression model was developed, and its effectiveness was assessed. This model was then used to predict N levels in tea gardens. A flow chart for brief presents the process of this work can be found in Figure 2.

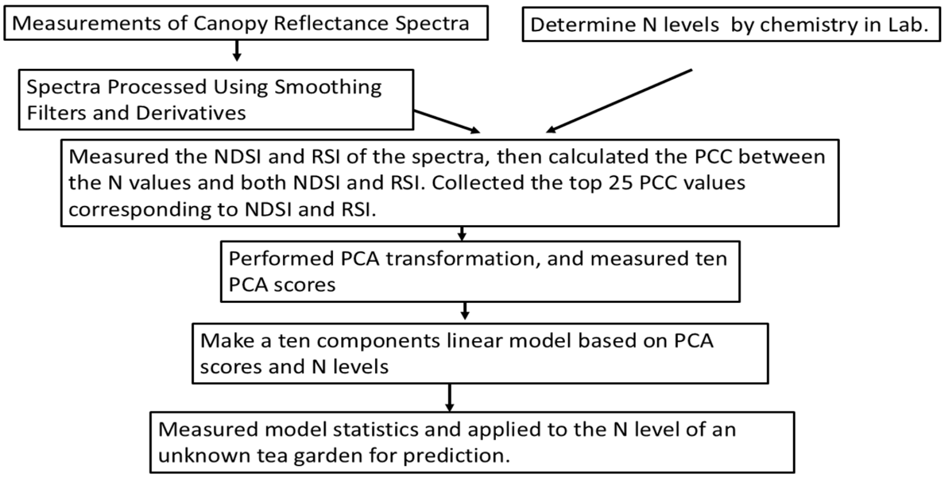

Figure 2. A flow chart for brief presents the process of this work.

2 Methods and Materials

2.1 Tea plant area

The Wuyishan City is the area of data collection, located in the northwest of Fujian Province, China, positioned around 118° E, 28° N. The altitude of sampling location is about 300 meters on average, with mean annual temperature of 18-20°C, 80-85% RH, rainfall around 2300 mm annually. Wuyishan stands as a globally renowned natural and cultural heritage site, serving as the birthplace of Chinese tea culture and art. Wuyi "rock tea", renowned for its superior quality among Oolong tea, holds a distinguished place among China's most celebrated tea [14]. The soil type is acid latosol, PH around 4.5-6.5, rich in organic matter, suitable for the growth and development of tea crop.

2.2 Experimental Setup

The tea leaves start to flourish around April, and Wuyi rock tea is harvested annually during this time. By July, the regrowing canopy may reach its luxuriant growth. The spectral data of experiment in this work were taken on July, 2022 and April, 2023.
Experiments were conducted in four different tea plantations ("YZX16," "YP30," "JGK42," and "HS54"), all within a 10 km radius of Wuyi University. The number following each abbreviation represents the number of sampling points in the respective plantations. The area of the tea plantations where the samples were taken ranges from approximately 500 to 1500 square meters. Due to the varying sizes of the plantations, the number of sampling points differs in each plantation.
All the tea plantations are Wuyi rock tea basically may with a little variety. The tea trees planted are of the single variety named: Cinnamon, growing well with no pests or diseases. The sampling points were primarily selected based on appearance and location before experiment. Each sampling point was separated by about 10 m. Experiments were performed between 10:00 to 14:00 on a cloudless day.

AvaField-1 portable hyperspectral spectrometer (Avantes, Netherlands) was used for spectrum collection. The performances of AvaField-1 spectrometer are range: 200-1100 nm; resolution: 2.4 nm; sampling interval: 0.6 nm; accuracy: ±0.3 nm; field of view: 25 degrees. The detectors used in this system is a 2048X14 pixel array CCD detector, with TE cooling. The detector was fixed on a tripod and was set above the tea canopy about 0.5 m. The experimental assistant wore dark clothes. A standard whiteboard with 99% reflection rate was used for calibration priori every sampling point.

2.3 Data Collection Methodology

A notebook was connected to the AvaField-1 for data collection. The measurement of spectral data in each sampling point were five times consecutive collections. The final spectra for one sampling point were obtained by averaging five individual spectra, with representative tea tree spectra shown in Figure 3.

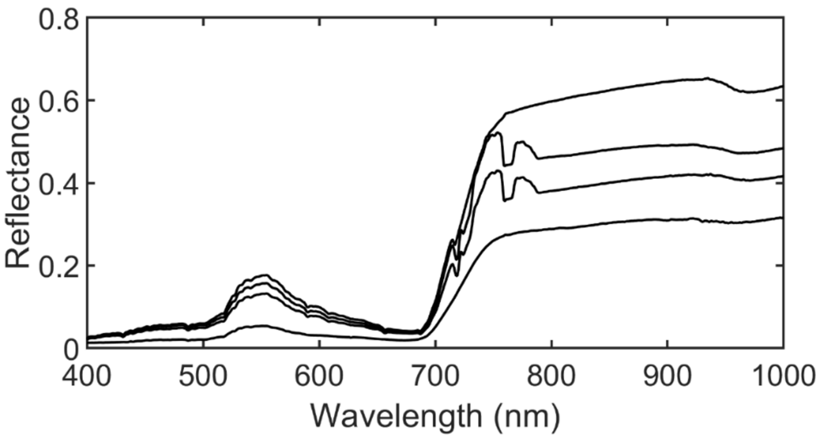

Figure 3. Four representative tea tree spectra.

Following the spectral data were collected, 15 canopy tea leaves were plucked at each sampling point randomly. These leaves put into the numbered bag, and send to the laboratory for N determination.

2.4 N measurement

The N of leaves was measured by artificial destructive method. The nitrogen content was determined by using the Kjeldahl N analyzer method. The standard material used was GBW07603(GSV-2), the standard N concentration was 15±0.3g/Kg, and the final nitrogen content (g/kg) can be obtained.

2.5 Spectral data processing

Since the measurement of reflectance spectra in real fields is influenced by various factors, such as environmental conditions and the structure of the tea canopy, accuracy, reliability, and resolution of the data can be affected by light source fluctuations and instrument drift.

Therefore, preprocessing of the spectrum is necessary before it can be used for content inversion in the biomass model.
Initially, all spectral data were smoothed using the Savitzky-Golay method to reduce random noise and improve the signal-to-noise ratio. To ensure continuity and ease of calculation, the original spectrum, which has 1092 bands at 0.6 nm intervals, was interpolated using a spline method to 600 bands at 1 nm intervals (ranging from 400 to 1000 nm). Both interpolation and smoothing operations were performed using Matlab R2022b. The strength of the linear relationship between leaf nitrogen content and spectra for one tea plantation was measured using Pearson's Correlation Coefficient (PCC).
The processed spectra were then processed advance using the first, second, and fourth derivatives. The PCC for representative bands in the original, first, second, and fourth derivative spectra is shown in Figure 4. This figure demonstrates that the PCC is consistently higher for those derivative spectra compared to the original spectra.

2.6 feature transformation

NDSI and RSI were calculated for all spectra, including the original and all derivative versions, using equations (1)

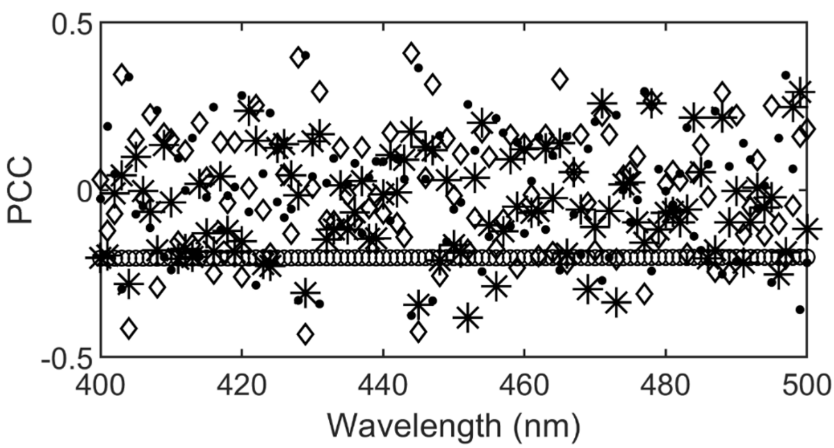

Figure 4. The PCC for representative bands (400-500 nm) in the original, first, second, and fourth derivative spectra. original: **o**, First derivative: *, Second derivative: ●, Fourth derivative, ◇

and (2). This process transforms the spectra into both feature space and index forms. The linear correlation of these indices with leaf nitrogen content was evaluated using Pearson's Correlation Coefficient (PCC). For each sampling garden, the top 25 PCC values for both NDSI and RSI indices, in relation to leaf nitrogen content, were recorded. For example, in a tea garden named HS54, there are 54 sampling points with corresponding nitrogen values and spectra, and a PCC value was computed for the NDSI index relative to these 54 N values and spectra. Other gardens are the same.

2.7 Principal Component Analysis (PCA)

PCA identifies a new coordinate system—linear combinations of the original base—that best captures the data with reduced dimensionality. PCA returns as many components as there are inputs. The top 25 PCC values of both NDSI and RSI indices. In this study, the first 10 components were selected to explain at least 95% of the total variance, while other components were excluded to achieve dimensional reduction [2]. The score as map to these components were collected.

2.8 Multiple Linear Regression

The score values of PCA for NDSI and RSI features were analyzed using a multiple linear regression routine in an attempt to determine they correlated with N values. Multiple linear regression fits an observed dependent data set (e.g., concentration of nitrogen) using a linear combination of independent variables (e.g. 10 scores). It has built a ten components linear regression model for each of the different sets of scores generated (PCA) with nitrogen concentration. The goodness of fit of each model was evaluated by calculating the correlation coefficient squared (R^2), the adjusted squared correlation coefficient (adjR^2), the square root of the mean square error (RMSE) and the statistical significance (p-value) of the model (analysis of the variance) [2, 15]. The linear equation determined by the regression has the form as shown in Eq. (3):

$$Nitrogen = a_0 + \sum_{i=1}^{10} a_i X_i \quad (3)$$

3 Results

3.1 PCC for indices with N

For each sampling garden, the top 25th PCC values for both NDSI and RSI indices in relation to leaf nitrogen content were recorded. Results for NDSI are shown in Table I, while results for RSI are in Table II. Both Tables indicate that the top 25th PCC values are higher for those spectra indices taken derivative. The most of fourth derivative yields the highest top 25th PCC for NDSI, as shown in Table I. The effect is consistently observed for those PCC by RSI, the PCC values are always higher for derivative indices compared to non-derivative indices. It is obviously that the NDSI got higher PCC than RSI do, as noted in Tables.

Table I. The top 25th PCC of NDSI with N for four tea gardens.

NDSI

Tea Gardens	YZX16	YP30	JGK42	HS54
Original	0.56	0.43	0.35	0.67
1st Derivative	0.67	0.63	0.50	**0.71**
2nd Derivative	0.70	0.64	**0.54**	0.69
4th Derivative	**0.71**	**0.69**	0.52	0.68

Table II. The top 25th PCC of RSI with N for four tea gardens.

RSI

Tea Gardens	YZX16	YP30	JGK42	HS54

Original	0.56	0.43	0.31	0.65
1st Derivative	0.60	0.52	0.42	0.69
2nd Derivative	0.60	**0.52**	**0.42**	**0.69**
4th Derivative	**0.63**	0.47	0.42	0.67

3.2 Statistics for Multiple Linear Regression

Statistics for a ten-component linear regression model applied to various sets of scores generated by NDSI and PCA with nitrogen concentration are presented in Table III for different tea plantations (III-1 for YZX16, III-2 for YP30, III-3 for JGK42, and III-4 for HS54). This table illustrates how well a ten-component linear regression model of each garden fits the N values, respectively. A lower p-value, higher R^2, and adjusted R^2, along with a lower RMSE, indicate a better model fit. A same statistic of RSI is indicated in Table IV (IV-1 for YZX16, IV-2 for YP30, IV-3 for JGK42, and IV-4 for HS54).
The higher the order of derivative got better statistics for NDSI but some variation for YP30, as illustrated in Table III. These effects are the same in Table IV for RSI. The best statistics shown as a bold form. These Tables shown that derivative spectroscopy techniques can improve the effectiveness of indices to estimate the nitrogen concentration as compared to non-derivative indices. The statistics values are always higher for derivative indices. It is obviously that the NDSI got better statistics than RSI do, as noted in Tables.

4 Discussions

4.1 Testing the Predictive Ability of Regression Equations

Regression equations were tested for the ability to predict N for new data sets. To accomplish this the coefficients in the regression equations of garden JGK42 was used as a model to create linear equation. The ten-component linear regression model of JGK42 (NDSI) applied to predict the N values in other gardens (i.e. same as an un-known garden), prediction results are indicated in Table V. These spectra of the other gardens were treated as before. The best statistics shown as a bold form. The result in this Table shows that derivative spectroscopy techniques can get the best predictions even higher to $R^2 = 0.77$ and Adj $R^2 = 0.73$ (garden YZX16, 4th Derivative). The other good predictions both of 1st derivative for garden YP30 and HS54.

Table III. Statistics for a ten-component linear regression model applied to various sets of scores generated by NDSI and PCA with nitrogen concentration.

(a)YZX16

	p-value	R^2	Adj R^2	RMSE
Original	0.42	0.72	0.15	4.69
1st Derivative	<0.01	0.96	0.87	1.83
2nd Derivative	**<0.01**	**0.98**	**0.95**	**1.12**
4th Derivative	<0.01	0.95	0.87	1.83

(b)YP30

	p-value	R^2	Adj R^2	RMSE
Original	0.04	0.57	0.35	4.05
1st Derivative	**<0.01**	**0.86**	**0.78**	**2.35**
2nd Derivative	<0.01	0.83	0.74	2.54
4th Derivative	<0.01	0.69	0.53	3.45

(c)JGK42

	p-value	R2	Adj R2	RMSE
Original	<0.01	0.56	0.42	3.78
1st Derivative	<0.01	0.66	0.55	3.3
2nd Derivative	<0.01	0.69	0.59	3.16
4th Derivative	**<0.01**	**0.76**	**0.69**	**2.77**

(d)HS54

	p-value	R2	Adj R2	RMSE
Original	<0.01	0.61	0.52	3.49
1st Derivative	**<0.01**	**0.69**	**0.61**	**3.15**
2nd Derivative	<0.01	0.66	0.59	3.23
4th Derivative	<0.01	0.64	0.56	3.36

4.2 Effects of SNR reduction

The higher the derivative may decline the SNR of signal [12]. The smooth filter applied on a derivation signal can improve the SNR and recommended Savitzky-Golay filter applied in a derivative spectrum [12].
Sampling garden named HS54 and JGK42 were selected for testing because there are the most spectra and N values in these data set. Both data were treated as above and followed an additional smoothing using Savitzky-Golay filter on three derivative versions of spectra. The SNR is improved but this additional procession made any better results on PCC, as noted in Table VI (a). Table VI (a) showed comparisons of top 25th PCC with N for smoothed vs. non-smoothed version for both JGK42 and HS54 tea gardens.
Based on this, a scheme originated in signal procession area named "average filter" was tried to apply for increasing SNR. This scheme added three spectra together point by point to made smooth and find an average, and N values too. Results are shown in Table VI (b), this additional procession obvious made better results on PCC.
Statistics for a ten-component linear regression model applied to sets of smoothing using average filter are presented in Table VII for HS54 and JGK42. The additional average filter procession obvious made better results on model
statistics of multiple linear regression. These effects indicated that the scheme of average filter should be considered in signal procession process in future. statistics for multiple linear regression.

4.3 Adaptability of the data and algorithms

Due to its simplicity and intuitiveness, which are unlikely to affect the effectiveness of the preprocessing scheme, the commonly used inversion method, Multiple Linear Regression, is adopted in this study to assess vegetation nitrogen content [2, 3]. The data and algorithms demonstrate good adaptability, highlighting the effectiveness of

Table IV Statistics for a ten-component linear regression model applied to various sets of scores generated by RSI and PCA with nitrogen concentration.

(a)YZX16

	p-value	R^2	Adj R^2	RMSE
Original	0.51	0.68	0.03	5.02
1st Derivative	0.08	0.88	0.64	3.07
2nd Derivative	0.11	0.86	0.59	3.25
4th Derivative	**0.04**	**0.92**	**0.75**	**2.56**

(b)YP30

	p-value	R^2	Adj R^2	RMSE
Original	**0.03**	**0.59**	**0.37**	**3.98**
1st Derivative	0.28	0.41	0.10	4.75
2nd Derivative	0.04	0.56	0.33	4.09
4th Derivative	**0.04**	**0.58**	**0.35**	**4.03**

(c)JGK42

	p-value	R^2	Adj R^2	RMSE
Original	<0.01	0.47	0.31	4.10
1st Derivative	**<0.01**	**0.49**	**0.33**	**4.06**
2nd Derivative	<0.01	0.48	0.32	4.09
4th Derivative	0.12	0.36	0.15	4.55

(d)HS54

	p-value	R^2	Adj R^2	RMSE
Original	<0.01	0.56	0.46	3.71
1st Derivative	<0.01	0.61	0.52	3.51
2nd Derivative	**<0.01**	**0.66**	**0.58**	**3.29**
4th Derivative	<0.01	0.63	0.54	3.44

4.4 Adaptability of the data and algorithms

Due to its simplicity and intuitiveness, which are unlikely to affect the effectiveness of the preprocessing scheme, the commonly used inversion method, Multiple Linear Regression, is adopted in this study to assess vegetation nitrogen content [2, 4]. The data and algorithms demonstrate good adaptability, highlighting the effectiveness of derivative spectroscopy techniques. The model achieves a reasonable level of accuracy, and the data aligns well with this approach. However, adaptability could be further enhanced by exploring alternative algorithms.

In addition to linear regression, nonlinear methods have been shown to outperform linear regression in N reconstruction studies [2, 9-11]. The SVM and PLSR regressions have been used to estimate N content in vegetation [2, 9], demonstrating superior performance.

Recently, Liang et al. estimated N content in wheat leaves and suggested that the Random Forest (RF) regression algorithm performs better than SVM regression [15]. To further enhance the effectiveness of the preprocessing scheme, the PLSR, and RF regression algorithms were applied to estimate N content for data of garden JGK42.

Figure 5 presents statistics for two regression models applied to scores generated by NDSI and PCA with nitrogen concentration for the JGK42 tea plantation. The figure clearly demonstrates that both regression algorithms achieve higher R2 and adjusted R2, as well as lower MAE and RMSE, depending on the spectral derivatives.

Table V Statistics for predicting nitrogen concentration using a ten-component linear regression model by JGK42, applied to various tea gardens.

(a) Statistics for predict YZX16

	MAE	MSE	RMSE	R^2	Adj R^2
Original	3.08	15.61	3.95	0.36	-0.9
1st Derivative	2.59	9.93	3.15	0.59	-0.23
2nd Derivative	2.40	7.77	2.79	0.67	0.04
4th Derivative	**1.87**	**5.55**	**2.36**	**0.77**	**0.31**

(b)Statistics for predict YP30

	MAE	MSE	RMSE	R^2	Adj R^2
Original	3.61	18.51	4.30	0.24	-0.16
1st Derivative	**2.51**	**10.74**	**3.30**	**0.56**	**0.33**
2nd Derivative	2.87	13.36	3.66	0.45	0.16
4th Derivative	3.15	15.67	3.96	0.36	0.02

(c)Statistics for predict HS54

	MAE	MSE	RMSE	R^2	Adj R^2
Original	2.45	9.39	3.07	0.30	0.13
1st Derivative	**2.45**	**9.39**	**3.07**	**0.63**	**0.54**
2nd Derivative	2.63	11.57	3.40	0.54	0.43
4th Derivative	2.68	11.86	3.44	0.53	0.42

MAE = Mean Absolute Error; MSE = Mean Squared Error; RMSE = Root Mean Squared Error

Table VI: The top 25th PCC of NDSI with N for JGK42 and HS54.

(a)Smooth vs. non-smooth

Tea Gardens	JGK42SM	JGK42	HS54SM	HS54
Original	0.35	0.35	0.67	0.67
1st Derivative	0.51	0.50	0.71	0.69
2nd Derivative	0.57	0.54	0.69	0.69
4th Derivative	0.56	0.52	0.68	0.67

SM=smooth version

(b)Average filter vs. non

Tea Gardens	JGK42Ave	JGK42	HS54Ave	HS54
Original	**0.62**	0.35	**0.89**	0.67
1st Derivative	**0.80**	0.50	**0.95**	0.69
2nd Derivative	**0.79**	0.54	**0.92**	0.69
4th Derivative	**0.79**	0.52	**0.92**	0.67

Ave= average filter

5 Conclusion

In this study, derivative spectroscopy techniques were evaluated for correlating spectra with N content using a linear regression model. The derivative approach enhances the accuracy of these indices for estimating N concentration by analyzing spectral signatures. The results demonstrated that the study involved spectrum smoothing, derivative

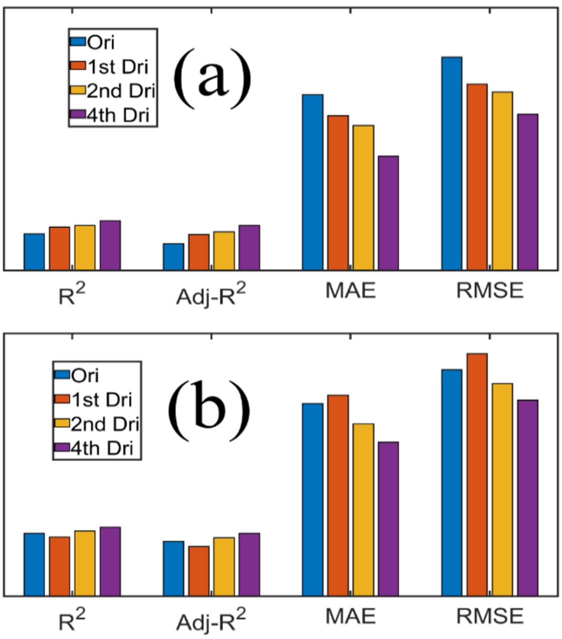

Figure 5. Statistics for (a)PLSR, (b)RF regression model applied to JGK42 of scores generated by NDSI and PCA with nitrogen concentration. (Ori=Original, 1^{st} , 2^{nd} 4^{th} derivation, MAE = Mean Absolute Error; RMSE = Root Mean Squared Error)

Table VII: Statistics for a ten-component linear regression model applied to HS54 and JGK42 both spectra (original and various derivative) were treated using Average filter.

(a)HS54 treated using Average filter

	p-value	R^2	Adj R^2	RMSE
Original	<0.01	0.96	0.90	1.58
1st Derivative	**<0.01**	**0.98**	**0.95**	**1.11**
2nd Derivative	<0.01	0.92	0.81	2.19
4th Derivative	<0.01	0.96	0.90	1.6

(b)JGK42 treated using Average filter

	p-value	R^2	Adj R^2	RMSE

Original	0.38	0.84	0.32	4.21
1st Derivative	0.03	0.98	0.90	1.64
2nd Derivative	<**0.01**	**0.99**	**0.98**	**0.71**
4th Derivative	0.04	0.97	0.90	1.80

implementation, and measurements of NDSI and RSI indices. Analysis identified the top 25 PCC for both indices and employed PCA to determine ten principal component scores. A ten-component linear regression model was developed, and its effectiveness was assessed. This model was then used to predict N levels in tea gardens. The results are promising, suggesting that further research in this field is warranted.

Acknowledgment

This work was supported in part by: Project of Fuxiaquan innovation platform, Fujian Province, (#2022-p-029). A research grant from Wuyi University, (#YJ202215). Two Natural Science Foundation projects of Fujian Province (#2022J011204) and (#2022J011206). Natural Science Foundation of Nanping City, Fujian Province (N2021J008-2). Education and Scientific Research Grants for Young and Middle-aged Teacher, Fujian Province (JAT190780). Nanping City Resource Chemical Industry Science and Technology Innovation joint funded project (N2020Z016), "Sky, Earth and Air" integration of highway ecological environment monitoring key technology research and application. China.

References

1. Y. Wang, X. Hu, Z. Hou, J. Ning, Z. Zhang, Discrimination of nitrogen fertilizer levels of tea plants (Camellia sinensis) based on hyperspectral imaging. Journal of Science Food and Agriculture. 98, 4659–4664 (2018)
2. F. Rodriguez-Moreno, and F. Llera-Cid, PCA versus ICA for Dimension Reduction of Spectral Signatures in the Search for an Index of Nitrogen Concentration in Plants. Spanish Journal of Agricultural Research. **9**, 1168-1175 (2011)
3. Y.-Z. Chen, B.-J. Fu, X.-M. Feng, Overview and Outlook on Remote Sensing Inversion of Vegetation Nitrogen Content. Acta Ecologica Sinica. **37**, 6240-6252 (2017)
4. L. Wang, Y. Wei, Revised Normalized Difference Nitrogen Index (NDNI) for Estimating Canopy Nitrogen Concentration in Wetlands. Optik - International Journal for Light and Electron Optics. **127,** 7676–7688 (2016)
5. 5.Q.-Y. Meng, H. Dong, Q.-M. Qin, J.-L. Wang, J.-H. Zhao, MTCARI: A Vegetation Index for Monitoring Leaf Chlorophyll Content Based on Hyperspectral Remote Sensing. Spectroscopy and Spectral Analysis. **32**, 2218-2222 (2012)
6. A.-K. Tilling, G.-J. O'Leary, J.-G. Ferwerda, S.-D. Jones, G.-J. Fitzgerald, D. R.-B. Rodriguez, Remote Sensing of Nitrogen and Water Stress in Wheat. Field Crops Research. **104**, 77-85. (2007)
7. J.-W. Rouse, R.-H. Haas, J.-A. Scheel, D.-W. Deering, Monitoring Vegetation Systems in the Great Plains with ERTS. Proceedings of the 3rd Earth Resource Technology Satellite (ERTS) Symposium, 48-62. (1974), vol. 1
8. W.-B. Wu, J.-Y. Li, Z.-B. Zhang, C.-J. Ling, X.-K. Lin, X.-L. Chang, Estimation model of LAI and nitrogen content in tea tree based on hyperspectral image.

Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), **34**, 195201 (2018)

9. C. Axelsson, A.-K. Skidmore, M. Schlerf, A. Fauzi, W. Verhoef, Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression. International Journal of Remote Sensing. **34**, 1724-1743 (2013).
10. Z. Huang, B.-J. Turner, S.-J. Dury, I.-R. Wallis, W.-J. Foley, Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. Remote Sensing of Environment. 93, 18-29 (2004)
11. Y. Wang, F.-M. Wang, J.-F. Huang, X.-Z. Wang, Z.-Y. Liu, Validation of artificial neural network techniques in the estimation of nitrogen concentration in rape using canopy hyperspectral reflectance data. International Journal of Remote Sensing. **30**, 4493-4505 (2009).
12. F. Tsai, W. Philpot, Derivative Analysis of Hyperspectral Data, Remote Sensing of Environment. 66.1, 41-51(1998)
13. Z. Li, S.-L.Yang, Y.-E. Liu, Y. Wang, C. Zhang, Y.-X. Xiao, H.-F. Xu, Hyperspectral total nitrogen inversion of wetland plants based on differential transformation. Research of Environmental Sciences. **35**, 1268-1276 (2022)
14. J.-P. Xue, Z.-P Zhuo, Y. LIU, Z. ZHOU, H.-J. Huang, R.-N. Xu, Z.-L. Guo, L. Sun, Z.-C. Chen, H. Liao, Variation of Wuyi Rock Tea in Mineral Nutrition by Region and Its Affecting Factors. Acta Pedologica Sinica. **59**, 528–535. 2022
15. R.-F. Kokaly, R.-N. Clark, Spectroscopic Determination of Leaf Biochemistry Using Band-Depth Analysis of Absorption Features and Stepwise Multiple Linear Regression. Remote Sensing of Environment. **67,** 267–287(1999)