Study On Spectral Characters-Chlorophyll Inversion Model of *Sabina Vulgaris* in Mu Us Sandy Land

Ning Wang
Inner Mongolia Agricultural University

Guang Yang (1150877126@qq.com)
Inner Mongolia Agricultural University

Xueying Han
Inner Mongolia Agricultural University

Guangpu Jia
Inner Mongolia Agricultural University

Feng Liu
Inner Mongolia Agricultural University

Xin Liu
Hulunbuir city development and Reform Commission

Haoyu Chen
Inner Mongolia Agricultural University

Tianqi Zhang
Inner Mongolia Agricultural University

Xinyu Guo
Inner Mongolia Agricultural University

Research Article

Keywords: Sabina vulgaris, Mu Us Sandy Land, Model prediction

Posted Date: December 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1130666/v1

Abstract

Sabina vulgaris is a group tree species in Mu Us Sandy Land. Understanding the growth status of *Sabina vulgaris* has guiding value for vegetation change monitoring. Chlorophyll is an important indicator to characterize the growth status of plants, and its content changes are important for analyzing the physiological growth status of plants and guiding the precise planting of plants. In this paper, the spectral reflectance and chlorophyll content of *Sabina vulgaris* were measured by SVC HR-1024 portable ground feature spectrometer and SPAD502 chlorophyll instrument, and the relationship between ground feature spectral characteristics and chlorophyll content of *Sabina vulgaris* was studied. The results show that there is a correlation between the vegetation index and chlorophyll, the effect of NDVI is the best, the bands with the highest correlation are the combined bands of 470nm-500nm, 610nm-680nm, and 740nm-840nm, and the wavelengths with the highest correlation are (660,790); Vegetation index, red-edge parameters, and chlorophyll have a certain correlation. The fitting effect of the model established by vegetation index is better than that established by red-edge parameters, and the highest R^2 is 0.97; Among the three modeling methods, the model fitting effect of partial least squares is the best, R^2 is > 0.91, and the disadvantage is that the processing process is complex; The processing method of the univariate linear regression model is the simplest, but the disadvantage is that the accuracy of the model is unstable, R^2 is between 0.1-0.9, so the multivariate linear regression model is the most suitable of the three methods(R^2>0.8).

Introduction

In recent decades, with climate change, population increase, unreasonable land use, and other factors, deserts and sandy land increased sharply. Due to the harsh sandy environment, which is not conducive to plant growth, there are fewer and fewer vegetation types that can survive in the sandy area for a long time. *Sabina vulgaris* is a desert plant in Mu Us sandy land that has survived long-term natural evolution. *Sabina vulgaris* is an evergreen shrub with fine branches, growing obliquely upward, and many smelly leaves and branches. It grows in clusters in fixed and semi-fixed sandy land[1]. Its growth characteristics determine that it has great canopy density in desert plants and adapts to the transpiration characteristics of sandy land, forming a unique community landscape in Mu Us sandy land.

Chlorophyll can absorb, reflect and transmit light energy at the same time[2]. The change of its content can characterize the growth status of vegetation and is directly related to plant aging and stress[3–4]. The visible light band of plants is a significant band for chlorophyll absorption[5], so chlorophyll and vegetation spectrum can be used to monitor plant growth.

In previous studies, the mathematical models of Vegetation Spectrum and chlorophyll are mostly used to estimate plant growth in a large area. In terms of crops, they are mainly used for crop monitoring, crop yield estimation, or pest control monitoring to assess crop growth, aging and stress in advance[6–8]. With

the development of remote sensing technology, hyperspectral remote sensing is applied to more and more fields.

In the 1960s, the US Department of Agriculture (USDA) pointed out that there were about 42 absorption bands related to vegetation leaf structure, pigment content, and nutrient elements in the range of 400~2400nm through detailed measurement and analysis of the spectral data of dried and mashed leaves of multiple vegetation types, This makes it possible for future generations to estimate the vegetation growth status and composition employing remote sensing. After that, researchers began to carry out spectral vegetation growth monitoring and continuously improve it[9].

In the 1980s, Miller et al. Defined the maximum value of the first-order differential spectrum in the range of 660-750nm as the "red-edge" position. Since then, scholars at home and abroad began to study the relationship between the content of chlorophyll and other pigments and the displacement of the "red-edge"[10].

The changes of red-edge position and red-edge slope are mainly related to the changes in plant chlorophyll content. It will generally cause spectral changes near the 700nm band of vegetation. When the leaf water shortage reaches the wilting coefficient, the position of the red-edge will shift red[11]. When the content of plant chlorophyll increases, the amplitude and peak area of the red-edge will also increase, The position of the red-edge tends to move towards the long wave direction[12]. The red-edge parameters of green vegetation can not only describe the phenological changes of observed vegetation and various physiological and biochemical parameters but also calculate the red-edge position, red-edge amplitude, and other parameters through the first-order differential value of its spectral reflectance in the range of 680-750nm. Therefore, it has great application value in vegetation spectral chlorophyll growth monitoring.

Then, with the in-depth study on the inversion of vegetation physical and chemical parameters, many scholars found that when using a single band for vegetation parameter inversion, there is often great uncertainty in the inversion results due to the influence of external environmental factors such as underlying surface. Therefore, the research on vegetation index has become a hot spot, and the establishment accuracy of the plant growth monitoring model has been improved[13]. After the establishment of the vegetation index was relatively perfect, scholars began to improve the prediction method of the model. From univariate linear analysis to multiple linear regression model, and then to principal component analysis, Mouazen compared principal component regression (PCR), partial least squares regression (PLSR), and backpropagation neural network (BPNN) Through the evaluation of the measurement accuracy of extractable forms such as selected soil properties by the three calibration methods, it is found that PLSR is superior to principal component analysis in model fitting accuracy, and the follow-up study of PLSR has become an ideal method for model establishment[14].

After that, spectral research was no longer limited to crops, and scholars began to turn their research direction to grassland and vegetation. It not only monitors grassland degradation and weed invasion, but

also quantitatively analyzes the spectral characteristics of degraded grassland species, effectively classifies weed species, and inverses the proportion of weeds, vegetation coverage, and plant height[15]; To solve the problem of mixed pixels in vegetation inversion, Li Zhe et al.[16] proposed an object-oriented multi endmember spectral mixing analysis method to study vegetation from the perspective of time scale and vegetation coverage level change, effectively reducing the amount of calculation and endmember change of mixed pixels, and the inversion value and accuracy are very high.

Although Chinese scholars have also begun to study the spectral characteristics of sandy plants in recent years, there are monitoring and mathematical models of sandy vegetation growth. Most studies focus on the analysis of spectral characteristics of different plant species in order to identify plant species by remote sensing monitoring and save manpower and material resources[17–18]; Or study the spectral characteristics of plants and environment in different succession stages in an ecosystem, or study the spectral characteristics of indicator species in grassland degradation, or establish a spectral model of forage quality. The above studies are to intervene in the pest control or stress of plant species in succession degradation stage in advance, curb water and soil loss and prevent the deterioration of the environment[19–21].

The important reason for sand damage in Mu Us sandy land is the low vegetation coverage. On the one hand, we can prevent and control the existing vegetation diseases and insect disasters and ensure their normal growth to protect the sandy ecological environment from deterioration. On the other hand, we can further plant according to the vegetation growth characteristics of local constructive plants, and carry out vegetation protection in advance during the critical period of the growth of *Sabina vulgaris*, It can further improve the harsh environment of Mu Us sandy land. Therefore, the study of spectral characteristics of *Sabina vulgaris* can quickly evaluate and monitor it and plays an important reference role in the evaluation of regional ecological and social value. This study starts with the spectral characteristics and chlorophyll of *Sabina vulgaris*, and uses the inversion model to monitor the growth characteristics of *Sabina vulgaris*, to provide an important reference for ecological environment investigation and comprehensive evaluation.

Study Area

The Mu Us Sandy Land is located at 37°27 '- 39°22 'N, 107°20' -111°50 'E[22]. The total area is about 4×104km^2. The Midwestern part of Mu Us Sandy land is a wavy plateau with the highest altitude of 1,600m[23]. Dunes of various types in The Mu Us Sandy Land account for 77% of the total area, and mobile dunes account for 47% of the total area[24]. The annual rainfall fluctuates greatly, increasing from 250mm to 440mm from northwest to southeast. Most of the precipitation is concentrated in summer, and the intensity of precipitation is the highest in August, which usually lasts for several days to more than ten days. Therefore, it is easy to cause drought and flood disasters, and drought is more frequent[25]. The annual average temperature is 6.0-9.0°C, the average temperature in January is -8.7-12°C and the average temperature in July is 20-24°C. The annual average temperature is 7.6°C, and the potential evaporation is

2,100-2,500mm. The main type of soil in Mu us sandy land is chestnut soil, which is alkaline and lacks organic matter and nutrients. Northwest desert steppe - steppe - southeast forest-steppe. Grasslands account for 90% of the total area. Most of the vegetation grows on sandy girders. Vegetation types mainly include sandy vegetation, meadow vegetation, halophytic vegetation, and marsh vegetation, among which the sandy vegetation accompanying all kinds of sandy land has the largest area[26].

Methods

Acquisition of spectral data. Spectral data was measured by SVC HR-1024 portable ground object spectrometer produced by Spectra Vista, USA. The spectral range was 350-2,500nm, the number of channels was 1024, and the spectral resolution was: 3.5≤nm within 350-1,000nm, ≤8.5nm within 1,000-1,850nm, ≤6.5nm within 1,850-2,500nm, minimum integration time 1ms, signal acquisition method is Bluetooth transmission[27].

The field measurement of plant spectrum is greatly affected by the solar altitude Angle, so clear and cloudless weather should be selected for measurement, and the measurement time is from 10:00 to 14:00[28]. Spectral measurement of ground objects in the study area was carried out in early May and mid-July 2019, and dark current collection and whiteboard calibration were required before measurement. In case of obvious weather changes such as strong wind halfway, it is necessary to calibrate again with a whiteboard, and it is necessary to calibrate again with a whiteboard every time the location is changed. To ensure the accuracy of the test results, the samples were randomly sampled. A total of 70 groups of Samples were measured with 5 replicates per group 350 times. Among them, 55 groups were used for the model establishment and 15 groups were used for model verification. During measurement, the measurement time of each spectral data was set as 5s, and the measurement height was 50cm from the probe to the ground object, The top view of spectral data acquisition is shown in figure.1a, Monitoring range of spectrometer probe figure.1b. The identification of *Sabina vulgaris* refer to flora of China and “Chinese Virtual Herbarium”, the number of *Sabina vulgaris* is LZD0000136.

Determination of chlorophyll content. Chlorophyll was measured by SPAD (Soil Plant Analysis Development Unit) 502 chlorophyll meter produced by KONICA MINLTA, Japan. SPAD502 chlorophyll meter will not cause any damage to plants.The selected plants were measured in the area with a diameter of 22cm around the center of *Sabina vulgaris* during spectral measurement. 560 measurements in total, Chlorophyll was measured repeatedly for 8 times in each treatment plot, and its average value was taken as the value of chlorophyll SPAD at this point, Corresponding to spectral data.

Fusion of spectral data. Firstly, the software SVC HR-1024 of the ground object spectrometer is used to eliminate the bands with large variations in the spectral curve data of the study area. The normal spectral curve should have no intersection and be smooth. Then, the SIG File Overlap/Matching function of the software is used to match the data. Finally, the SIG File Merge function is used to Merge the data and output the data to Excel format.

Smoothing of spectral data. Due to the difference in the response of the band to the energy during the measurement of spectral data, the spectral curve always has noise, so the data curve needs to be smoothed to eliminate the small amount of noise contained in the signal. Commonly used smoothing method has moving average method, static average method, moving average method includes three-point smoothing method, five-point smoothing method, nine-point smoothing method, etc... The five-point smoothing method was used in this paper [27], The comparison before and after smoothing is shown in Figure 2.

Resampling of spectral data. The resampling of spectral data is to ensure the accuracy of the spectral data prediction model established in the later stage[29]. ENVI5.1 is used to resample spectral data. The resampling interval is 10nm.

First-order differential processing. After first-order differential processing of spectral data can eliminate the systematic error of different data, better eliminate the impact of noise on data, more prominent spectral characteristics of vegetation. The data after first-order differential processing is shown in Figure 3, the first-order differential formula is shown in Table 1.

$$R\left(\lambda_i\right) = \frac{\left[R\left(\lambda_{i+1}\right) - R\left(\lambda_i\right)\right]}{\Delta\lambda_i} \times R$$

1

Where R (λ_i) is the first-order differential spectrum of wavelengths λ_{i+1} andλ_i. R(λ_i) and R(λ_{i+1}) are the original spectral reflectance at *i* and *i*+1. Δ(λ_i) is the wavelength difference between λ_{i+1} and λ_i.

Table.1 Red-edge Parameter

Red-edge Parameter	definition
The red-edge area(S_r)	Sum of first-order differentials in the range 680~760nm
The red-edge position(D_r)	Wavelength corresponding to the maximum first-order differential in the range of 680~760nm
The red keep out appearance value(R_p)	A maximum first-order differential in the range of 680~760nm

Statistical analysis and modeling. *Selection of vegetation index*. The correlation between spectral data and chlorophyll of cypress in sandy land was analyzed by MATLAB R2012a software. The selected vegetation index and its calculation formula are shown in Table 2:

Table.2 Vegetation index calculation method[29]

Vegetation index	A formula to calculate
RVI	R_{800}/R_{680}
DVI	R_{800}-R_{680}
NDVI	$(R_{800}-R_{680})/(R_{800}+R_{680})$
mSR	$(R_{750}-R_{445})/(R_{705}+R_{445})$
mNDVI	$(R_{750}-R_{705})/(R_{750}+R_{705}-2R_{445})$

Extraction of optimal spectral index wavelength combination. The original spectral data of 350-2500nm band were selected, and the reflectance of any composition was analyzed by normalized vegetation index (NDVI), ratio vegetation index (RVI) and difference vegetation index (DVI), and the correlation between chlorophyll and reflectance was analyzed by MATLAB R2012a. The formula is as follows [30]:

$$N_I\left(R_{\lambda_1}, R_{\lambda_2}\right) = \left|R_{\lambda_1} - R_{\lambda_2}\right| / \left|R_{\lambda_1} + R_{\lambda_2}\right|$$

2

$$R_I\left(R_{\lambda_1}, R_{\lambda_2}\right) = R_{\lambda_1} / R_{\lambda_2}$$

3

$$D_I\left(R_{\lambda_1}, R_{\lambda_2}\right) = R_{\lambda_1} - R_{\lambda_2}$$

4

Mathematical modeling. The prediction model of chlorophyll content was established by unary linear regression, multiple stepwise linear regression, and the partial least square method. The unitary linear regression model has small input and output, and the calculation method is convenient and simple, which is completed by Origin2017[30]. The multiple stepwise linear regression model selects one of the most important independent variables into the regression equation according to the weight of each data involved in modeling. It is widely used in the spectral prediction model and completed by SPSS26[31]. The partial least square method is also widely used in the field of spectroscopy. Through the analysis of multiple independent variables and dependent variables, the correlation between them can be maximized in the modeling process, which is completed by MATLAB R2012a. To verify the accuracy of the model and make the model more accurate, 70 groups of data were divided into 55 groups of prediction model data and 15 groups of validation model data.

The accuracy evaluation of the hyperspectral prediction model is mainly based on the determination coefficient and total root means square difference. R^2 is related to the stability of the model, that is, if the

value of R^2 is large, the stability of the model is better; Moreover, R_{MSE} is related to the prediction ability of the model: the smaller the value of R_{MSE}, the higher the accuracy of the established model and the stronger the prediction ability [32].

Extracting optimal vegetation index from structural equation model. The structural equation model is developed from linear equation. Its principle is to use the Bayesian estimation method to hypothetically guess the data and various indicators through previous empirical theory, and extract the relationship between unmeasurable information from the measured information. From the previous retrieval, the structural equation is mainly used to estimate and test the relationship between some non directly measurable variables in psychology, management, and economics[33].

In this study, the structural equation model is mainly used to screen vegetation index. Different vegetation indices have different representations of vegetation chlorophyll and spectrum. For example, normalized vegetation index overcomes the disadvantage of ratio vegetation index, limits the value to [-1,1], and can eliminate most of the changes in irradiance conditions related to solar angle, terrain, cloud shadow, and atmospheric conditions. However, in areas with dense vegetation, The normalized vegetation index will tend to supersaturation early, and can not timely reflect the growth process from yellow to dry in the process of vegetation growth[34]. Therefore, the vegetation index selected in the final modeling has a great impact on the accuracy of modeling.

In this study, according to the five vegetation indexes for establishing linear regression, three indexes and red-edge parameters are randomly selected to establish the structural joint equation model, and finally, the vegetation index most suitable for establishing the mathematical model is selected.

Result

Analysis of spectral characteristics of Sabina vulgaris. The spectral curve of Cypress on sandy land conforms to the spectral characteristics of general green plants (Fig. 4). Green plants are mainly affected by various pigments (chlorophyll, lutein, carotenoid, etc.) contained in leaves in the visible light band, among which chlorophyll plays the most important role[35]. Due to the strong absorption of pigments to electromagnetic waves and other radiation in this band, the reflection and transmission of leaves are very low. In the 420nm-450nm blue waveband and the 620nm-780nm red waveband, chlorophyll strongly absorbs radiation waves and easily forms absorption valleys. The reflection between these two absorption valleys is relatively reduced and forms reflection peaks, which makes plants appear green. If the normal growth of plants is inhibited in some form in the visible band, the decrease of chlorophyll content will increase the reflection of plants in the blue-green band and reduce the absorption.

The curve obviously shows the characteristics of "five grains and four peaks" of green plants. The main characteristics of vegetation spectrum are "Red Valley" and "green peak" in a visible light band; The red-edge appears between 680-760 nm, which is a diagnostic spectral feature of vegetation, and the Red

Valley forms high reflection in this band; There is a small reflection peak near the wavelength of 800 nm, namely "green peak". With the increase of chlorophyll, the spectral curve will shift to the right.

In the near-infrared band, the main influencing factor of green plants is the cell structure inside the leaves. In this band, the absorption energy of leaves is low, but the reflection and transmission are similar. High reflection is formed in the 680-1300nm spectrum.

In the infrared band, the transmission of plants is very small, and the absorption and incidence are similar. The main influencing factor is the water content in plant cells. Generally, two main water absorption bands are formed in the band centered at 1400nm and 1900nm.

Extraction of optimal spectral index wavelength combination. In the correlation analysis between the ratio of original spectral reflectance of the two bands, the vegetation index, the improved RVI, NDVI, mNDVI with chlorophyll, the RVI performed best (Fig. 5), and blue to yellow indicated a high negative correlation to high positive correlation. In the ratio vegetation index, the highest correlation is in the combined band of 610-680nm and 700-940nm, and the highest correlation is in the combined band of 350-430nm and 650-690nm in the improved red-edge ratio vegetation index. In NDVI, the highest correlation was found in the combined bands of 470 nm-500 nm, 610 nm-680 nm, and 740 nm-840 nm. The correlation between NDVI and chlorophyll was 0.95. Through comparative analysis, it was found that the band with the highest correlation between NDVI and chlorophyll was (660,790), the best ratio vegetation index was (630,720), and the best improved red-edge ratio vegetation index was (360,450). Therefore, in the follow-up monitoring, we can focus on the band with better performance to monitor the growth of *Sabina vulgaris*.

Selection of optimal index. In this study, according to the five vegetation indexes for establishing linear regression, three indexes and red-edge parameters are randomly selected to establish the structural joint equation model, and finally, the vegetation index most suitable for establishing the mathematical model is selected. Among the five vegetation indices selected in this study, NDVI has a close relationship with chlorophyll and has a significant characterization of chlorophyll, which is consistent with the previous research results. Similarly, the characterization of chlorophyll by RVI and mNDVI is also significant. In previous studies, the disadvantage of NDVI is that it is not suitable for areas with large vegetation coverage, while the disadvantage of RVI is that it is not suitable for measurement areas with too sparse vegetation and soil impact. The *Sabina vulgaris* in this study belongs to tree species with unsaturated vegetation coverage. At the same time, the selected measurement area and experimental design during measurement also minimize the error of ground soil on this measurement, Therefore, in this study, their characterization of chlorophyll is more obvious(figure. 6b, figure. 6c), and they are the preferred indexes in the follow-up modeling.

However, the characterization of mSR is relatively unstable. When DVI and mNDVI establish equations, the characterization of the relationship between mSR and chlorophyll is significant(figure. 6a), but it has a negative correlation with the characterization of chlorophyll when RVI and NDVI establish equations(figure. 6c). This is because the influence ways of the structural equation model are divided

into direct influence and indirect influence. In all indicators of establishing a structural equation model, The relationship between indicators and indicators is interactive. It can be seen from the figure that among the equations established by mSR, RVI, and NDVI, RVI and D_r have an obvious characterization effect, which will indirectly affect the direct characterization effect of mSR on chlorophyll. Therefore, the characterization of chlorophyll is unstable, but the above mNDVI, RVI, and NDVI are also indirectly affected by other factors, However, the performance is still obvious and stable, so the three are the best indexes for modeling(figure. 6).

Mathematical model. *Univariate linear regression model*. The univariate linear regression analysis was conducted on the spectral data and chlorophyll (Fig. 7). According to the fitting results, the best fitting effect is the normalized vegetation index, with R^2 above 0.9. According to the conclusions of previous researchers, NDVI itself is used to monitor the vegetation growth status. At the same time, NDVI can also eliminate the error impact of atmospheric radiation, so it is more suitable for this study, The results are also satisfactory; The worst effect is the difference vegetation index, R^2 is only 0.16;

However, except for the difference vegetation index, the correlation between other vegetation indexes and chlorophyll is about 0.6, in which the correlation coefficient R^2 of mSR and RVI is > 0.7, because the applicable condition of RVI itself is "the ratio of scattering of green leaves in the near-infrared band to chlorophyll absorption in the red band"[37]. It can be seen that RVI itself is more suitable for studying the spectrum of green plants, Therefore, the effect is ideal when it is used for the correlation between green plant spectrum and chlorophyll; Similarly, mSR corrects the specular reflection efficiency of leaves. The original vegetation index RVI based on mSR is more suitable for this study, so the fitting effect of mSR in this study is also ideal. Although mNDVI is an improved value of NDVI, it is only very sensitive to small changes in leaf canopy, gap segments and senescence[38]. In previous studies, it is also mostly used for fine agriculture, vegetation monitoring, and vegetation stress detection. mNDVI is selected because it is suitable for vegetation monitoring. However, according to the linear fitting results, the research used for *Sabina vulgaris* is not ideal. The subsequent mNDVI should be more used for spectral monitoring of broad-leaved tree species or monitoring with high vegetation coverage.

Multiple stepwise regression model analysis. Advantages of multiple stepwise regression analysis: the regression equation includes all independent variables that have a significant impact on the dependent variable, and does not include the regression equation of independent variables that have no significant impact on the dependent variable. Stepwise regression analysis is a regression analysis method based on this principle. Its essence is to derive an algorithm skill for studying and establishing the optimal multiple linear regression equation based on multiple linear regression analysis. It mainly uses the principle of regression analysis, adopts the double test principle, and gradually introduces and eliminates independent variables to establish the optimal regression equation[39].

In this study, vegetation index and red-edge parameters were used as independent variables and chlorophyll content as dependent variables. Two multivariate linear stepwise regression models were established by SPSS26. The regression equation is constructed as shown in Table 3. According to Table

3, the fitting degree of the model constructed by vegetation index is much higher than that constructed by red-edge parameters, and its RMSE is also relatively small, so the accuracy of its prediction model is high. But generally speaking, the accuracy of multivariate stepwise regression model is higher than that of univariate linear regression model.

Table 3
Multiple stepwise regression model equation

parameter	Model	Model accuracy		Test model accuracy	
		R^2	R_{mse}	R^2	R_{mse}
vegetation index	$y=-0.542+5.063_{NDVI}+7.373_{RVI}$	0.938	0.194	0.934	0.124
Red-edge parameter	y=-5.962+0.19Rp+23.114Dr	0.885	0.183	0.820	0.194

Partial least squares regression model. The partial least squares regression models of different leaf coverage areas were established by programming with MATLAB R2012a and taking vegetation index and red-edge characteristic parameters as inputs respectively. For the accuracy of the results, three indexes mNDVI, RVI, and NDVI with the best characterization of chlorophyll were selected from 3.3 to establish the partial least squares regression model, and the three vegetation indexes and red-edge parameters were used as inputs respectively. The results are shown in Table 4. Compared with multiple stepwise regression analysis, the partial least squares regression model has higher accuracy than the multiple stepwise regression model, the correlation coefficients of vegetation index and red-edge parameter model have increased, while R_{mse} has decreased and is less than 0.1. The fitting effect of the model with vegetation index as input is higher than that of the prediction model, As a model, the effect is better.

Based on the above three modeling methods, the model accuracy established by partial least square method is better than univariate linear regression model and multivariate linear regression model, and the fitting effect is the best.

Table.4 partial least squares regression equation

parameter	Model accuracy		Test model accuracy	
	R^2	R_{mse}	R^2	R_{mse}
vegetation index	0.971	0.094	0.982	0.037
Red-edge parameter	0.914	0.091	0.938	0.043

Discussion

As important sand fixing vegetation in fixed and semi-fixed sandy land, *Sabina vulgaris* is of great significance in preventing wind and fixing sand and improving the ecological environment. Taking Mu Us *Sabina vulgaris* as the research object, the spectral data were measured in the study area, and the

chlorophyll content of *Sabina vulgaris* was measured. During measurement, compared with ordinary broad-leaved trees, the chlorophyll content of Sabina vulgaris is generally low. Firstly, the leaf type of *Sabina vulgaris* is coniferous[40], and secondly, the measurement principle of the chlorophyll meter is more suitable for broad-leaved trees[41]. In the measurement process, even if the leaves are covered with small holes, there will be some gaps in them, or the leaves will be artificially integrated into a cluster, At this time, the instrument will also have measurement error and data error because the leaves are too thick. Therefore, the spectrophotometer method will continue to be used to measure chlorophyll content in subsequent experiments.

In this paper, three modeling methods are used to predict the chlorophyll content of *Sabina vulgaris*, and the model verification results are ideal. The modeling results show that the three modeling methods are feasible to predict the chlorophyll content of *Sabina vulgaris* and monitor its growth by using spectral characteristic parameters. The univariate linear fitting effect is unstable, which mainly depends on the selection of vegetation index, and different plants have different sensitivity to vegetation index[42], so univariate linear fitting is not recommended; The fitting effects of multiple stepwise regression analysis and partial least squares method are good, but the modeling process of partial least squares method is more complex[43], so it is more appropriate to finally choose multiple stepwise regression analysis for modeling.

Among the three methods, the red-edge parameters are the data extracted by the first-order differential of the original spectrum. Although the data processed by the first-order differential can better reduce human interference and strengthen its absorption effect on water[44–45], this study focuses more on the modeling of visible light band and chlorophyll, Therefore, the accuracy of the model established by vegetation index is higher than that of red-edge parameters, which is more stable and universal for the model established with chlorophyll.

In most of the past articles on plant hyperspectral characteristics and chlorophyll, in the models established by the spectral characteristics of all plants and different chlorophyll contents, the value of the determination coefficient has reached more than 0.8, which shows that a good correlation can be established between the spectral characteristics of color plants and chlorophyll through the analysis of vegetation index[46–48]. The significance of this correlation lies in: A model for inversion of green plant growth can be established through the change of plant chlorophyll content. Through the analysis of model data, any change of green plants in the growth process can be determined, achieving the purpose of timely prevention and control of diseases and pests and human regulation.

Due to the differences in a growth environment and internal factors, the spectral reflectance curves of different vegetation types are not the same, but they are in the form of "five grains and four peaks". Therefore, this study is not only applicable to *Sabina vulgaris*, but also applicable to other vegetation. The extraction of vegetation index from the original spectral reflectance of vegetation is helpful to highlight the spectral characteristics of vegetation and establish the model. However, the preferred vegetation index in this paper is common, but this method has certain limitations and does not have universality.

With the change of operation mode and measurement conditions, the wavelength of the preferred vegetation index will also change. Therefore, this paper chooses to use the structural program model to optimize the best vegetation index for subsequent modeling.

In fact, in previous studies, the structural equation model is mostly used to study the deep-seated correlation between dependent variables and independent variables, and can also be used in the mathematical modeling part of this paper. However, the author believes that although the structural equation model can mine the deep-seated correlation, the structural equation model is more complex in modeling and analysis, The purpose of this study is to find a model with a good fitting effect between spectrum and chlorophyll and relatively good operation. Therefore, structural equation model has not been applied to the model establishment of this study. However, based on the advantages of the structural equation model, the selection of the optimal index for mining this paper is very appropriate. At the same time, the author compares the mNDVI, RVI, and NDVI indexes screened out by the structural equation model with the unselected mNDVI and mSR (adding the NDVI with high accuracy in each modeling at the same time). The results are indeed the screened mNDVI The higher accuracy of RVI and NDVI shows that it is feasible to screen indexes with higher accuracy by structural equation model.

At the same time, as sandy vegetation, the spectral characteristics of *Sabina vulgaris* are the same as those of general green plants, so the results of this paper are also applicable to other desert plants.

Conclusion

In this paper, the spectral reflectance and chlorophyll content of Sabina vulgaris were measured by SVC HR-1024 portable ground object spectrometer and SPAD502 chlorophyll meter. The data were post-processed by MATLAB R2012a, SPSS26, and other software, and the ground object spectral characteristics, characteristics of chlorophyll content, and the relationship between them were studied. The results show that:

(1) The spectral curve of *Sabina vulgaris* has a reflection peak at 570nm in the visible band, an absorption Valley at 680nm, and an obvious reflection platform in the near-infrared band, which is in line with the spectral characteristics of green plants;

(2) There is a correlation between the vegetation index and chlorophyll in the original spectral composition of *Sabina vulgaris*. The effect of the normalized index is better than that of the specific value vegetation index and improved red-edge ratio vegetation index. The combined bands of 470-500nm, 610-680nm, and 740-840nm have the highest correlation, and the wavelength with the highest correlation is (660,790);

(3) There is a certain correlation between vegetation index, red-edge parameters, and chlorophyll. The fitting effect of the model established by vegetation index is better than that established by red-edge parameters. The model fitting degree R^2 of normalized vegetation index is as high as 0.93 and the red-edge parameter R^2 is 0.78;

(4) Among the three modeling methods, partial least squares have the highest accuracy, $R^2 > 0.9$, and are the most stable. The disadvantage is that the processing process is complex; The processing method of a univariate linear regression model is the simplest, but the disadvantage is that the accuracy of the model is not high, which has a great relationship with the selection of vegetation index. Sometimes the results are not objective, in which the lowest R^2 is 0.16 and the maximum R^2 is 0.99; The R^2 of multiple linear regression remains between 0.88-0.94, so the multiple linear regression model is the most suitable among the three.

Declarations

Acknowledgments

This research was supported by Inner Mongolia Autonomous Region's major scientific and technological project "Research and Industrialization Demonstration of Near-Natural Restoration Technology of Sandy Ecosystem" (Grant No. 2019ZD003).

Author contributions

N.W. wrote the main manuscript text and X.H. prepared Figs. 1–7. In addition, G.Y. reviewed and

revised the manuscript. G.J ., F.L., H.C., X.L., T.Z., and X.G Carried out field survey in Mu us Sandy land. All authors discussed the results and commented on the manuscript. The identification of *Sabina vulgaris* refer to flora of China. The final determiner of the plant is the corresponding author G.Y. No samples were collected in this experiment, and the *Sabina vulgar* is in the study area was not damaged.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.W. or Y.G..

References

1. DONG, X. J., CHEN, Z. X., ALATENG, B., LIU, Z. M., & SIDENG, D. B. (1999). A preliminary study on the water regimes of Sabina Vulgaris in Maowusu sandland, China. *Chinese Journal of Plant Ecology*, *23*(4), 311. https://www.plant-ecology.com/EN/Y1999/V23/I4/311 (1999).
2. Fassnacht F E, Stenzel S, Gitelson A A. Non-destructive estimation of foliar carotenoid content of tree species using merged vegetation indices. *Journal of Plant Physiology*,176: 210-217. https://doi.org/10.1016/j.jplph.2014.11.003 (2015).
3. Yu, K., Lenz-Wiedemann, V., Chen, X., & Bareth, G. Estimating leaf chlorophyll of barley at different growth stages using spectral indices to reduce soil background and canopy structure effects. *ISPRS*

Journal of Photogrammetry and Remote Sensing, *97*, 58-77. https://doi.org/10.1016/j.isprsjprs.2014.08.005 (2014).

4. Jay, S., Maupas, F., Bendoula, R., & Gorretta, N. Retrieving LAI, chlorophyll and nitrogen contents in sugar beet crops from multi-angular optical remote sensing: Comparison of vegetation indices and PROSAIL inversion for field phenotyping. *Field Crops Research*, *210*, 33-46. https://doi.org/10.1016/j.fcr.2017.05.005 (2017).
5. Féret, J. B., Gitelson, A. A., Noble, S. D., & Jacquemoud, S. PROSPECT-D: towards modeling leaf optical properties through a complete lifecycle. *Remote Sensing of Environment*, *193*, 204-215. https://doi.org/10.1016/j.rse.2017.03.004 (2017).
6. Liu S, Yu H, et al. Study on Inversion Model of Chlorophyll Content in Soybean Leaf Based on Optimal Spectral Indices. *Spectroscopy and Spectral Analysis*,41(06):1912-1919. (2021).
7. Liang L., et al. Estimating crop chlorophyll content with hyperspectral vegetation indices and the hybrid inversion method. *International Journal of Remote Sensing*,, 37(13): 2923-2949. https://doi.org/10.1080/01431161.2016.1186850 (2016).
8. Bao, Y., et al. Vis-SWIR spectral prediction model for soil organic matter with different grouping strategies. *Catena*, *195*, 104703. https://doi.org/10.1016/j.catena.2020.104703 (2020).
9. Wang X. Study on Chlorophyll and Nitrogen Content of WalnutForest Crown Based on Hyperspectral Information. *Chinese Academy of Forestry Sciences*.(2020).
10. Miller, J. R., et al. W. Seasonal patterns in leaf reflectance red-edge characteristics. *International Journal of Remote Sensing*, *12*(7), 1509-1523. https://doi.org/10.1080/01431169108955186 (1990).
11. Danson, F. M., & Plummer, S. E. Red-edge response to forest leaf area index. *Remote Sensing*, *16*(1), 183-188. https://doi.org/10.1080/01431169508954387 (1995).
12. Horler D. H. N, Dockray M., Barber J. The Red-edge of Plant Leaf Reflectance. I*nternational Journal of Remote Sensing*, 4:273-288. https://doi.org/10.1080/01431168308948546 (1983)
13. Kokaly, R. F. Investigating a physical basis for spectroscopic estimates of leaf nitrogen concentration. *Remote Sensing of Environment*, *75*(2), 153-161. https://doi.org/10.1016/S0034-4257(00)00163-2 (2001).
14. Cheng, Y., et al. On Change of Soil Moisture Distribution With Vegetation Reconstruction in Mu Us Sandy Land of China, With Newly Designed Lysimeter. *Frontiers in Plant Science*, *12*, 182. https://doi.org/10.3389/fpls.2021.609529 (2021).
15. Xing, F., et al. Mapping the occurrence and spatial distribution of noxious weed species with multisource data in degraded grasslands in the Three-River Headwaters Region, China. *Science of The Total Environment*, *801*, 149714. https://doi.org/10.1016/j.scitotenv.2021.149714 (2021).
16. Zhang, S., et al. Fractional vegetation cover estimation of different vegetation types in the Qaidam Basin. *Sustainability*, *11*(3), 864. https://doi.org/10.3390/su11030864 (2019).
17. Pan, Y., et al. Effect of Seasonal Grazing on Trade-off Among Plant Functional Traits in Root, Stem and Leaf of Leymus chinensis in the Temperate Grassland of Inner Mongolia, China. *Chinese Bulletin of Botany*, *52*(3), 307. https://www.chinbullbotany.com/EN/Y2017/V52/I3/307 (2017).

18. Lyu, X., et al. A new method for grassland degradation monitoring by vegetation species composition using hyperspectral remote sensing. *Ecological Indicators*, *114*, 106310. https://doi.org/10.1016/j.ecolind.2020.106310 (2020).
19. Möckel, T., et al. Classification of grassland successional stages using airborne hyperspectral imagery. *Remote sensing*, *6*(8), 7732-7761. https://doi.org/10.3390/rs6087732 (2014).
20. Wijesingha, J., Astor, T., Schulze-Brüninghoff, D., Wengert, M., & Wachendorf, M. Predicting forage quality of grasslands using UAV-borne imaging spectroscopy. *Remote Sensing*, *12*(1), 126. https://doi.org/10.3390/rs12010126 (2020).
21. Liu, X., Tian, Z., Zhang, A., Zhao, A., & Liu, H. Impacts of climate on spatiotemporal variations in vegetation NDVI from 1982–2015 in Inner Mongolia, China. *Sustainability*, 11(3), 768. https://doi.org/10.3390/su11030768 (2019).
22. Guo, Z. C., et al. Biomass and vegetation coverage survey in the Mu Us sandy land-based on unmanned aerial vehicle RGB images. *International Journal of Applied Earth Observation and Geoinformation*, *94*, 102239. https://doi.org/10.1016/j.jag.2020.102239 (2021).
23. Karnieli, A., Qin, Z., Wu, B., Panov, N., & Yan, F. Spatio-temporal dynamics of land-use and land-cover in the Mu Us sandy land, China, using the change vector analysis technique. *Remote Sensing*, *6*(10), 9316-9339. https://doi.org/10.3390/rs6109316 (2014).
24. Liang, P., & Yang, X. Landscape spatial patterns in the Mu Us Sandy Land, northern China and their impact factors. *Catena*, *145*, 321-333. https://doi.org/10.1016/j.catena.2016.06.023 (2016).
25. Zhou, H., et al. Network analysis reveals the strengthening of microbial interaction in biological soil crust development in the Mu Us Sandy Land, northwestern China. *Soil Biology and Biochemistry*, *144*, 107782. https://doi.org/10.1016/j.soilbio.2020.107782 (2020).
26. Han, X., et al. Spatiotemporal dynamic evolution and driving factors of desertification in the Mu Us Sandy Land in 30 years. *Scientific Reports*, *10*(1), 1-15. https://doi.org/10.1038/s41598-020-78665-9 (2020).
27. Yang G, Li Q. Spectral characteristics of artificial biological crust of desert alga in the east edge of Kubuqi Desert. *Science of Soil and Water Conservation*, 12(01):90. (2014).
28. Niu Y, Liu Y,. Analysis of Hyperspectral Characteristics and Extraction of Remote Sensing. *Journal of Ecology and Rural Environment*, 33(07):632-644. (2017).
29. NIU Y, et al.Interpreting Parameters of Five Types of Typical Psammo-Vegetation of the Horqin Sandy Land. *Journal of Ecology and Rural Environment*. *33*(7), 632-644. (2017).
30. Chuvieco, E. *Fundamentals of satellite remote sensing: An environmental approach*. CRC press. https://doi.org/10.1201/9780429506482 (2020).
31. Liu X, Yang G. Spectral Characteristics of Plant-soil Mixture with Spectral Reflection Measurement. *Journal of Northeast Forestry University*, 48(02):54-60. (2020).
32. Sun B. Study on hyperspectral characteristics and estimating models about physiological parameters of rape. *Northwest A&F University*. (2017).

33. Hekmati, M., Hasanirad, S., Khaledi, A., & Esmaeili, D. Green synthesis of silver nanoparticles using extracts of Allium rotundum l, Falcaria vulgaris Bernh, and Ferulago angulate Boiss, and their antimicrobial effects in vitro. *Gene Reports*, *19*, 100589. https://doi.org/10.1016/j.genrep.2020.100589 (2020).
34. Sharafatmandrad, M., & Mashizi, A. K. Visual value of rangeland landscapes: A study based on structural equation modeling. *Ecological Engineering*, *146*, 105742. https://doi.org/10.1016/j.ecoleng.2020.105742 (2020).
35. Jiang, Z., et al. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote sensing of environment*, *101*(3), 366-378. https://doi.org/10.1016/j.rse.2006.01.003 (2006).
36. Tang, Y. L., Huang, J. F., Wang, X. Z., Wang, R. C., & Wang, F. M. Comparison of the characteristics of hyper spectra and the red edge in rice, corn and cotton. *Scientia Agricultura Sinica*, *37*(1), 29-35. (2004).
37. Tulokhonov, A. K., Tsydypov, B. Z., Voloshin, A. L., Batueva, D. Z., & Chimeddorj, T. Spatio-temporal characteristics of vegetation cover of arid and semiarid climatic zones in mongolia on the basis of vegetation index NDVI. *Arid Ecosystems*, *4*(2), 61-68. https://doi.org/10.1134/S2079096114020115 (2014).
38. Malingreau, J. P. The vegetation index and the study of vegetation dynamics. In *Applications of remote sensing to agrometeorology* (pp. 285-303). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-2235-8_12 (1989).
39. Wang S H, Sun W, Li S W, et al. Interannual variation of the growing season maximum normalized difference vegetation index, mNDVI, and its relationship with climatic factors on the Tibetan Plateau. *Polish Journal of Ecology*, 63(3): 424-439. https://doi.org/10.3161/15052249PJE2015.63.3.012 (2015)
40. Mouazen, A. M., Kuang, B., De Baerdemaeker, J., & Ramon, H. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma*, *158*(1-2), 23-31. https://doi.org/10.1016/j.geoderma.2010.03.001 (2010).
41. Ning, L., Liu, C. X., He, W. M., & Yu, F. H. Interactions of the indigenous evergreen shrub Sabina vulgaris with coexisting species in the Mu Us sandland. *Journal of Plant Ecology*, *6*(1), 48-56. https://doi.org/10.1093/jpe/rts019 (2013).
42. Ates, F., & Kaya, O. The Relationship Between Iron and Nitrogen Concentrations Based On Kjeldahl Method and SPAD-502 Readings in Grapevine (Vitis vinifera L. cv.'Sultana Seedless'). *Erwerbs-Obstbau*, *63*(1), 53-59. https://doi.org/10.1007/s10341-021-00580-8 (2021).
43. Gao L, et al. Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159: 364-377. https://doi.org/10.1016/j.isprsjprs.2019.11.018 (2020).

44. Hair J F, Ringle C M, Sarstedt M. Partial least squares: the better approach to structural equation modeling?. *Long Range Planning*, 45(5-6): 312-319. DOI: 10.1016/j.lrp.2012.09.011 (2012).
45. Zhu, J., et al. Spectral reflectance characteristics and chlorophyll content estimation model of Quercus aquifolioides leaves at different altitudes in Sejila Mountain. *Applied Sciences*, *10*(10), 3636. https://doi.org/10.3390/app10103636 (2020).
46. Sun, J., et al. Using Spectral Reflectance to Estimate the Leaf Chlorophyll Content of Maize Inoculated With Arbuscular Mycorrhizal Fungi Under Water Stress. *Frontiers in Plant Science*, *12*, 1053. https://doi.org/10.3389/fpls.2021.646173 (2021).
47. Dzikiti, S., et al. Determining the water status of Satsuma mandarin trees [Citrus Unshiu Marcovitch] using spectral indices and by combining hyperspectral and physiological data. *Agricultural and Forest Meteorology*, *150*(3), 369-379. https://doi.org/10.1016/j.agrformet.2009.12.005 (2010).
48. Xu, X., et al. Predicting leaf chlorophyll content and its nonuniform vertical distribution of summer maize by using a radiation transfer model. *Journal of applied remote sensing*, *13*(3), 034505. https://doi.org/10.1117/1.JRS.13.034505 (2019).

Figures

Figure 1

Top view of measured spectral data (a), schematic diagram of spectral data, and chlorophyll content measurement. In Fig. B, position a is the probe position of the spectrometer, and elliptical area B is the area that can be monitored by the probe. The probe is 50cm away from the canopy of Sabina vulgaris. When measuring chlorophyll, it must be within the range that can be monitored by the probe. Therefore, the radius of the chlorophyll monitoring area is 28.86cm.

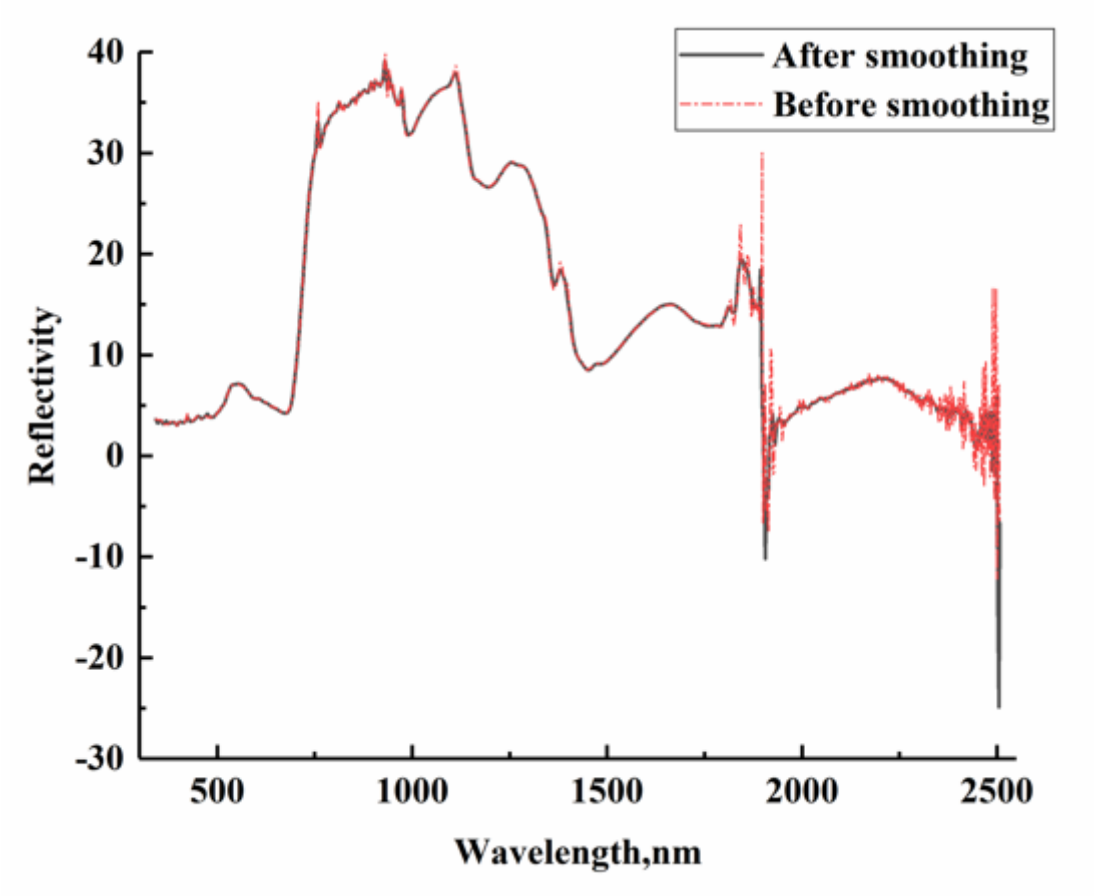

Figure 2

Comparison of spectral smoothing effect. The red dotted line is before smoothing, and the black solid line is after smoothing. The impact of noise in the 1800-2500 nm band is obvious. After smoothing, the noise error is reduced and the accuracy of data is enhanced.

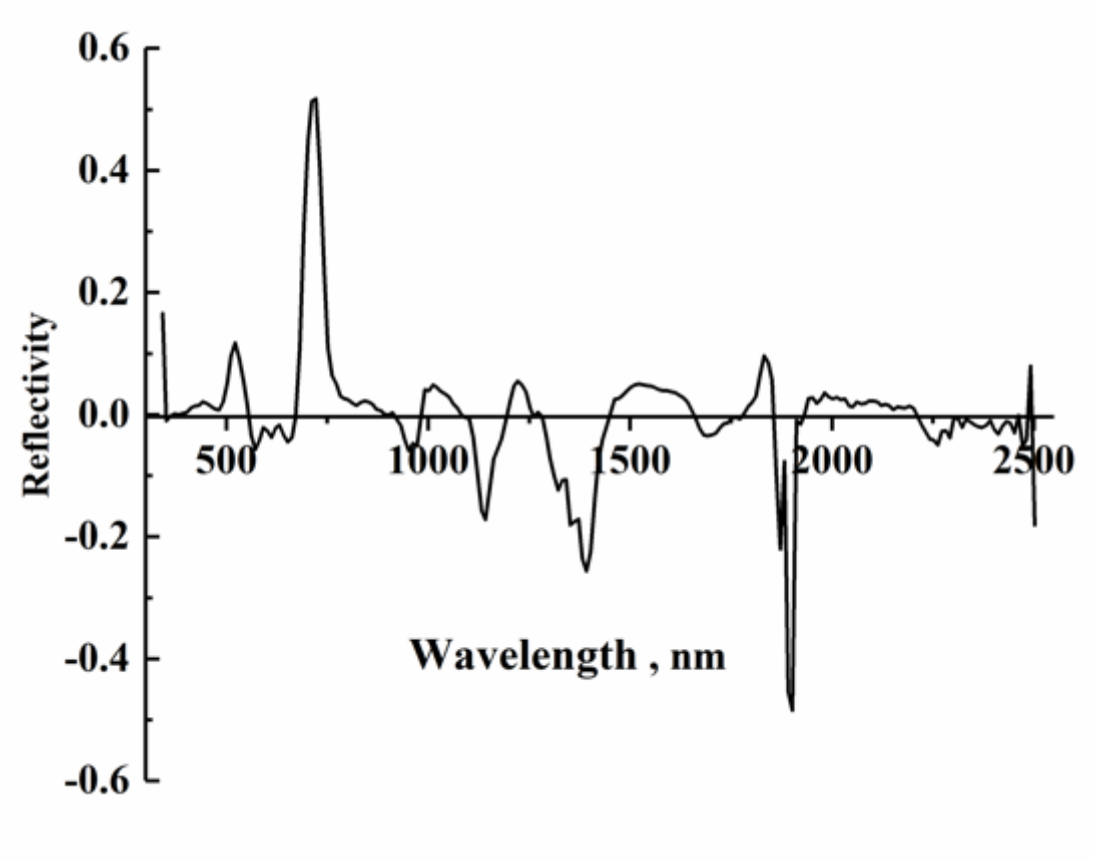

Figure 3

Spectral data after first-order differential processing strengthens the change of absorption valley or reflection peak at each place.

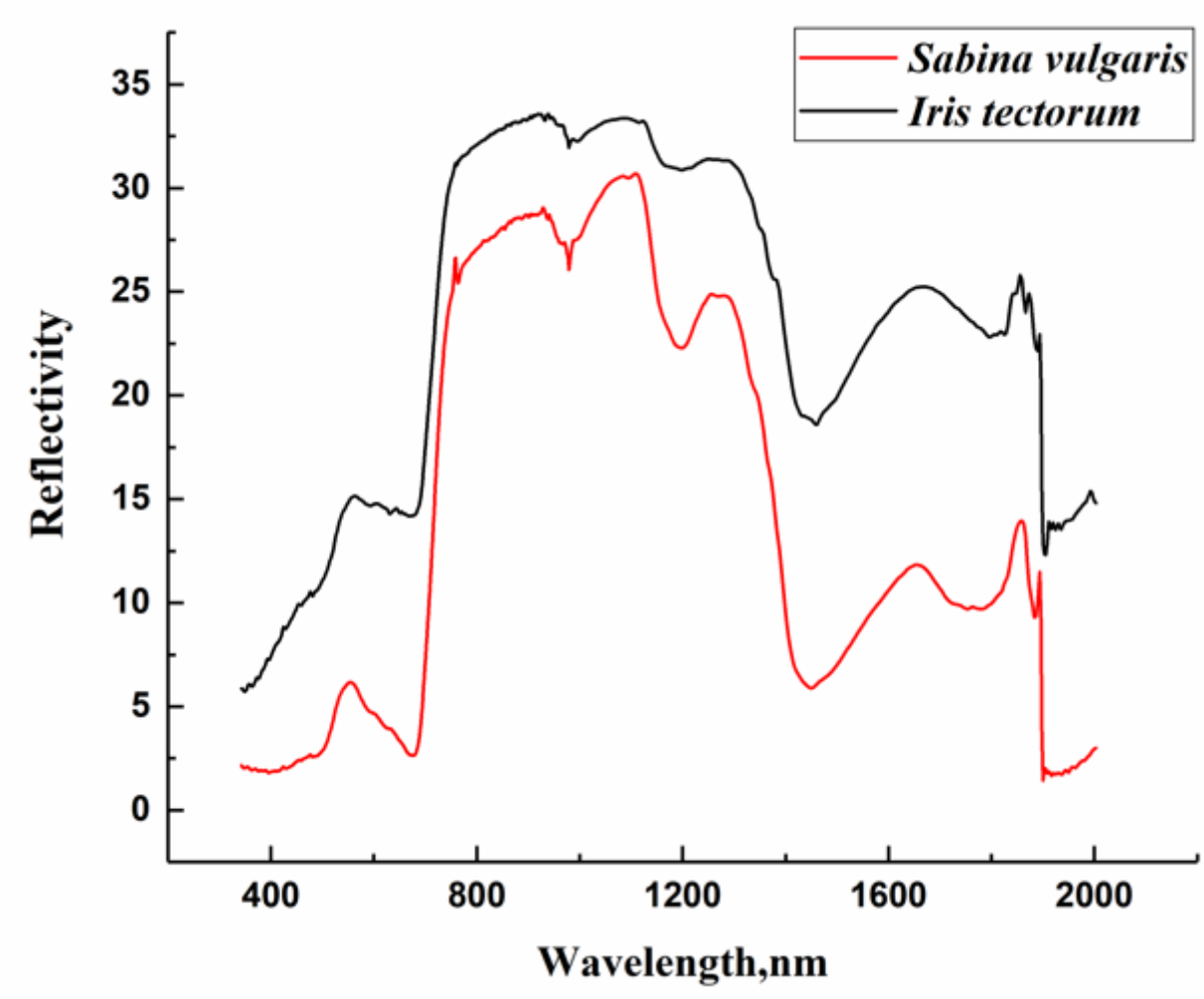

Figure 4

The spectral curves of *Sabina vulgaris* and other green plants (taking *Iris tetorum* as an example) show the law of "five grains and four peaks". The difference is the wavelength positions of "absorption Valley" and "reflection peak". In the figure, it is obvious that the reflectivity of *Iris tetorum* is higher than *Sabina vulgaris* at each band.

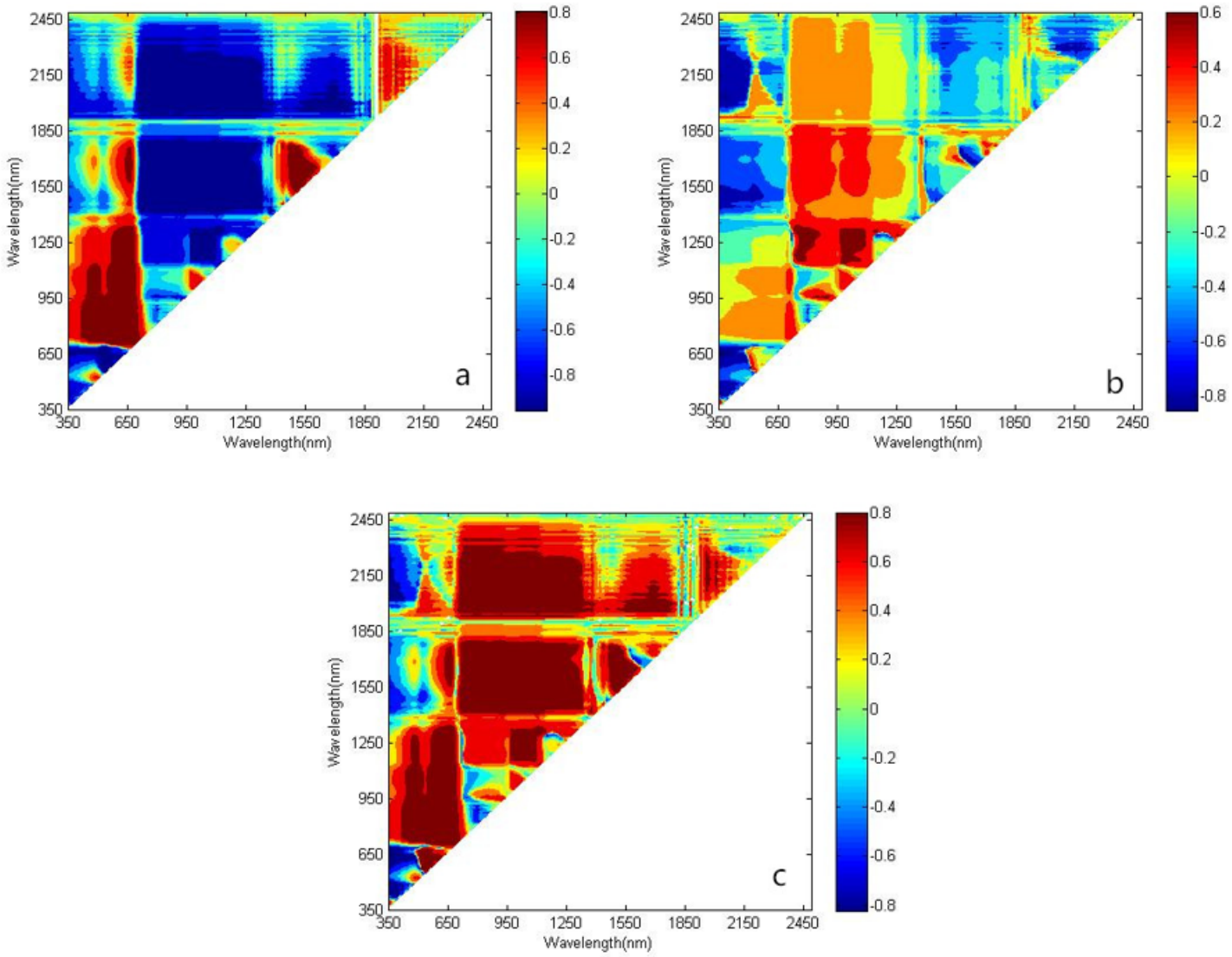

Figure 5

Correlation analysis between original band spectrum and chlorophyll: RVI (a), DVI (b), NDVI (c), and blue to red indicates high negative correlation to high positive correlation.

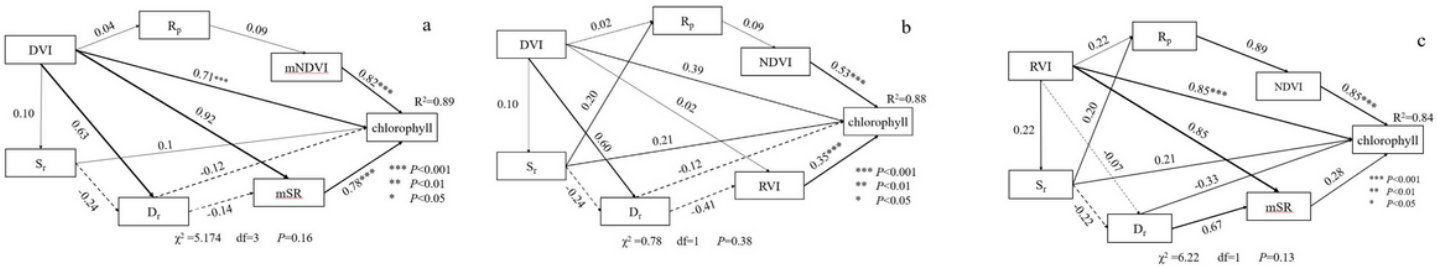

Figure 6

Analyzes the relationship between chlorophyll and NDVI、RVI、DVI、mSR、mNDVI, red-edge amplitude, red-edge area, and red-edge position by using structural equation model. The solid line and dotted line represent positive coefficient and negative coefficient respectively; The thickness of the arrow indicates the size of the standardized path coefficient; The R^2 value represents the proportion of variance interpretation of each endogenous variable.

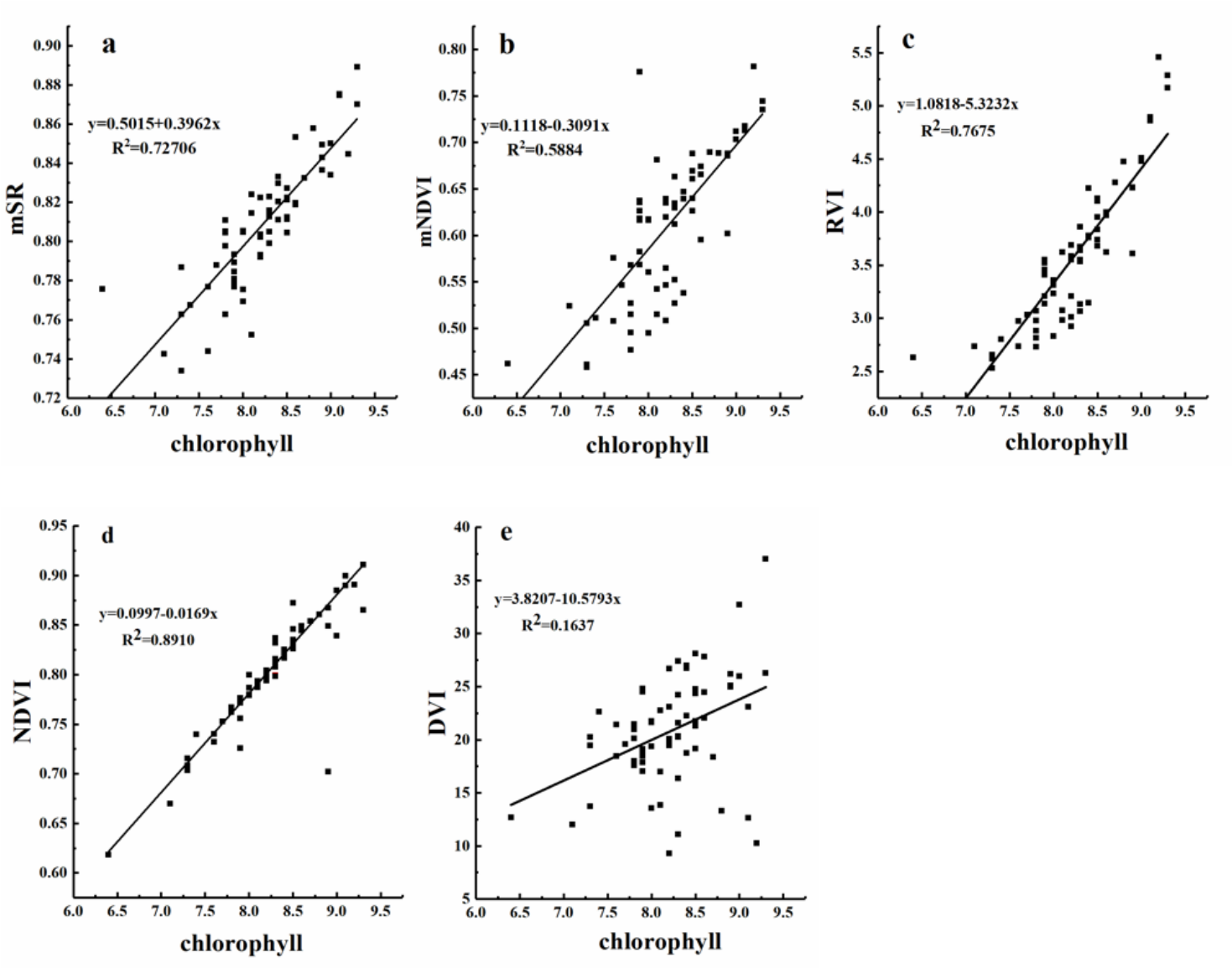

Figure 7

Univariate linear regression analysis, fitting of MSR and chlorophyll content (a), fitting of mNDVI and chlorophyll content (b), fitting of RVI and chlorophyll content (c), fitting of NDVI and chlorophyll content (d), fitting of DVI and chlorophyll content (e).