Visualization of Chlorophyll Content Distribution in Apple Leaves Based on Hyperspectral Imaging Technology

Xin Wen¹, Xicun Zhu¹,²*, Ruiyang Yu¹, Jingling Xiong¹, Dongsheng Gao³, Yuanmao Jiang³, Guijun Yang⁴

¹College of Resources and Environment, Shandong Agricultural University, Tai’an, China
²Key Laboratory of Agricultural Ecology and Environment, Shandong Agricultural University, Tai’an, China
³College of Horticulture Science and Engineering, Shandong Agricultural University, National Apple Engineering and Technology Research Center, Tai’an, China
⁴National Engineering Research Center for Information Technology in Agriculture, Beijing, China

E-mail: *zx.c@sdau.edu.cn

Abstract

We took distribution visualization of chlorophyll content in apple leaves to estimate the nutrient content and growth levels of apple leaves. 130 mature and non-destructive apple leaves were collected, and imaging spectroscopy data were collected by SOC710VP hyperspectral imager. The chlorophyll content of the leaves was determined on the spectral information of the leaves. After pre-processing, we took linear wavelength stepwise regression method to choose the sensitive wavelength of chlorophyll content. And then we established partial least squares, principal component analysis and stepwise regression model. Finally, the chlorophyll content distribution visualization was realized. The results showed that the sensitive wavelengths of the chlorophyll content were 712.50 nm, 509.95 nm, 561.22 nm, 840.62 nm, 696.67 nm and 987.91 nm. The $R^2$, RMSE, RE of the optical chlorophyll content estimation model, and principal component analysis regression model, were 0.800, 0.319 and 26.4%. The chlorophyll content of each pixel on the hyperspectral image of apple leaves was calculated by the best estimation model and we completed the visualization distribution of chlorophyll content, which provided a technical support for the rapid detection of nutrient distribution.

Keywords

Apple Leaves, Chlorophyll Content, Hyperspectral, Visualization

1. Introduction

Chlorophyll in plant leaves plays an important role in plant metabolism and
Chlorophyll in apple leaves plays an important role in photosynthesis [1] [2] [3]. The use of hyperspectral imaging technology to estimate the chlorophyll estimation and distribution of apple leaves is of great significance for the nutrient distribution and precise fertilization of apple leaves [4]. According to the characteristics of chlorophyll reflection and absorption of specific wavelength spectra, chlorophyll spectroscopy diagnosis of a large number of crops has been carried out at home and abroad. Curran [5] studied between the original spectrum and chlorophyll content and the spectral and chlorophyll content after first-order differential treatment. In contrast, it was found that first-order differentiation of the spectrum can well eliminate the influence of the background environment or illumination on the spectral reflectance. Song Kaishan et al. [6] took the correlation analysis of the hyperspectral reflectance and chlorophyll content of soybean canopy, and screened the sensitive band model with large correlation coefficient, and carried out the inversion estimation of chlorophyll content. Shi Jiyong [7] studied the rapid and non-destructive detection of chlorophyll content in cucumber leaves with cucumber as the research object. The multi-linear regression model was established by using hyperspectral image information, and the chlorophyll content distribution map was drawn, indicating that non-destructive testing was performed by hyperspectral image technology. Leaf chlorophyll content and distribution are feasible. Yu Keqiang [8] used hyperspectral imaging technology to visualize the nitrogen distribution of pepper leaves, and obtained the inversion map of nitrogen distribution in pepper leaves. The results show that hyperspectral imaging technology can be used for nondestructive detection of plant nutrients. Spectral information can only reflect the concentration of nutrients, and can not reflect the spatial distribution characteristics. Hyperspectral imaging technology can be used to obtain spectral information and image information, and chlorophyll content distribution inversion [9] [10] [11] [12].

Hyperspectral imaging technology combines the advantages of both spectroscopy and image. It has the characteristics of high resolution, multi-band, and map integration. It can detect the appearance characteristics and internal components of objects, and can utilize the multi-band spectrum to the content of plant nutrients. Quantitative analysis can be used to visualize the spatial distribution differences using images [13]-[20]. Therefore, this study carried out the estimation and visualization of chlorophyll content in apple leaves, and used hyperspectral imaging technology to obtain the changes of leaf nutrient status during apple growth, which could provide technical support for precise fertilization management.

2. Experimental Part

2.1. Sample Collection

The study area is an apple orchard in Qixia County, Yantai City, Shandong Province. Study object Red Fuji apple tree leaves. The study sample was apple
leaves collected in September 2016. At the same time, the experimental results are representative for the distribution of the apple orchard. 130 healthy leaf samples were collected by random sampling method. Each sample has 3 leaves. After collecting the samples, they were quickly loaded into the fresh box, sealed, numbered and brought back to the laboratory for determination.

2.2. Collecting Hyperspectral Data from Apple Leaves

The experiment used SOC710VP hyperspectral imager to collect the imaging spectrum data, which consisted of an imaging spectrometer, two 15 W halogen lamps, a national standard gray board, a black box and a computer equipped with acquisition software. Assemble the instruments, place the national standard gray board inside the dark box; install the halogen lamp in the dark box, install one on each side, adjust the position of the halogen lamp to make the halogen light source cross-illuminate the center of the gray board; fix it in the center of the gray board An imaging spectrometer that aligns the imaging spectrometer lens with the center of the gray plate and adjusts the lens aperture to F5.6; connects the computer to the imaging spectrometer. After the assembly of the hyperspectral imaging data system platform is completed, the system is de-bugged. Place the blade in the center of the gray plate, connect the halogen lamp power supply, and open the image acquisition software. When the software displays the sample image, adjust the lens focal length of the imaging spectrometer to make the sample image clearly visible. Close the black box and ensure that the halogen lamp is the only light source. Try to collect the imaging spectrum information and make sure the data acquisition system platform is correct. As shown in Figure 1.

In order to reduce the uneven distribution of the intensity of the light source during the acquisition process and the noise generated by the dark current in the lens, it is necessary to perform black and white correction of the image before the image is acquired. The formula is as follows:

\[ R = \frac{I - B}{W - B} \]  

Among them: \( I \) is the original acquisition image, \( B \) represents the all black calibration image, \( W \) represents the all white reflection calibration image, and black and white correction is performed according to the formula to obtain the corrected image \( R \).

2.3. Determination of Chlorophyll Content

Weigh about 0.20 g of freshly cut leaves, transfer the leaves to a mortar, add a small amount of quartz sand, \( \text{CaCO}_3 \), and then add a little 95% ethanol solution, grind the leaves to a green liquid, put the filter paper into the funnel, and use alcohol It is moistened, and the funnel is placed in a 25 ml brown volumetric flask, drained with a glass tube, the green liquid is completely transferred to a brown volumetric flask, the mortar is washed several times with alcohol, and the green
liquid is completely transferred into the volumetric flask; When there is no liq-
uid, the funnel is taken out and the volume is adjusted with ethanol solution; the
chlorophyll solution is introduced into the cuvette, and the absorbance at 665
nm, 649 nm, and 470 nm is measured by a spectrophotometer, and the con-
centration of chlorophyll in the solution is calculated. The formula for calculating
chlorophyll is as follows:

\[
Chla = 13.95 \times A_{665} - 6.88 \times A_{649} \\
Chlb = 24.96 \times A_{649} - 7.32 \times A_{665} \\
Chl(a+b) = Chla + Chlb
\]

Among them: \( Chla \) represents the content of chlorophyll \( a \), \( Chlb \) represents
the content of chlorophyll \( b \), and \( Chl(a+b) \) is the total content of chlorophyll.

2.4. Pretreatment of Spectral Data of Apple Leaves

Due to the influence of external environment, light source, instrument accuracy
and other factors, the spectral information obtained by hyperspectral technology
will appear noise, distortion, information redundancy, etc. Therefore, the spec-
trum needs to be preprocessed to eliminate the noise and information redu-
dancy of the spectral data. After all, spectral data preprocessing is the basic
processing of hyperspectral images, and subsequent research is based on these
processes. At the same time, spectral data preprocessing is also an important
method to effectively improve spectral accuracy and screen effective spectral
information. This study used multivariate scatter correction for pretreatment.

Multivariate scatter correction is a relatively common treatment method for
multiple wavelengths. The spectral data is processed by a multivariate scatter
correction method to reduce the effects of light scattering and enhance spectral
absorption information related to the component content.

2.5. Sensitive Wavelength Screening

The amount of information contained in hyperspectral data is very large. If all
the data is calculated, it will not only cost manpower and time, but also
over-fitting the model, which requires spectral feature extraction. In this way, both spectral data dimensionality reduction and irrelevant information can be removed, and errors can be reduced.

In order to reduce information redundancy and speed up the calculation, this study used SPSS software to screen sensitive wavelengths using stepwise regression analysis.

2.6. Model Establishment Method

Partial least squares (PLS) regression analysis is a statistical method for finding a linear regression model by projecting predictors and observation variables into a new space. Both data X and Y are projected into the new space, so the PLS series of methods are called bilinear factor models. Compared with the traditional multiple linear regression model, it can be performed under the condition of multiple correlations of independent variables, allowing the number of samples to be less than the number of variables, and the regression coefficients of each independent variable are easier to explain.

Principal Component Regression Analysis is a regression analysis based on the principal component as an independent variable. It is a method for analyzing multivariate collinearity problems. It generally consists of two steps: one is to re-linearly combine the independent variables, and the other is Delete the difference in the new component variable, leaving the main component. After the multi-collinearity in the regression model is eliminated by principal component analysis, the principal component variables are used as independent variables for regression analysis, and then the original variables are substituted back to obtain a new model according to the score coefficient matrix. Principal component regression models have good stability.

The basic idea of stepwise regression is to introduce variables into the model one by one. After each introduction of an explanatory variable, an F-test is performed, and the explanatory variables that have been selected are subjected to t-test one by one. When the originally introduced explanatory variables are changed due to the introduction of later explanatory variables If it is no longer significant, delete it. To ensure that only the significant variables are included in the regression equation before each new variable is introduced. This is an iterative process until neither a significant explanatory variable is selected into the regression equation nor an insignificant explanatory variable is removed from the regression equation. In order to ensure that the final set of explanatory variables is optimal.

2.7. Model Accuracy Test Method

In order to evaluate the practical performance and predictive ability of the model, the accuracy of the model is tested, and the accuracy of the model is tested by using the decision coefficient ($R^2$), the root mean square error (RMSE), and the relative error ($RE$). The fitting determination coefficient ($R^2$) represents the
closeness between the measured value and the estimated value. The relative error (RE) and the root mean square error (RMSE) represent the degree of dispersion between the measured value and the estimated value. In general, the larger R² is, the smaller the RE and RMSE are, indicating that the estimation accuracy of the model is higher. Calculated as follows:

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

(5)

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\]

(6)

\[
RE = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i - \hat{y}_i}{y_i}
\]

(7)

In the above formula: \( y_i \) is the model to estimate the content of nutrient elements, \( \hat{y}_i \) is the measured value of nutrient elements, \( \bar{y} \) is the average value of the measured values, and \( n \) is the number of samples.

3. Results and Discussion

3.1. Spectral Curve Characteristics of Apple Leaves

Figure 2 is a correlation between the original hyperspectral reflectance of apple leaves and leaf chlorophyll content and the correlation between spectral and chlorophyll content after multi-scatter correction. It can be seen that the chlorophyll content of apple leaves and the original spectrum are at a wavelength of 530 - 570 nm. It is significantly correlated with 710 - 740 nm, and the absolute value of the correlation coefficient is high. When the wavelength is greater than 600 nm, the absolute value of the correlation coefficient begins to decrease. At the wavelengths of 400 - 500 nm and 690 nm, the correlation coefficient is close to 0, and then the wavelength increases. Large, the correlation coefficient is small. It can be seen that the wavelength band after the wavelength greater than 780 nm hardly reflects the chlorophyll content information, and the absolute

---

**Figure 2.** Comparison of correlation curves of pretreatment of apple leaf chlorophyll spectrum.
value of the correlation coefficient reaches the maximum value of 0.65 at the wavelength of 720 nm. The chlorophyll content of apple leaves and the multi-scattering corrected spectra showed a significant negative correlation at wavelengths of 540 - 580 nm and 720 - 750 nm, and significant positive correlations at wavelengths of 460 - 520 nm and 680 - 710 nm, with wavelengths of 420 - 520 nm and 540. The correlation coefficients are higher than 0.65 at 580 nm, 670 - 690 nm and 720 - 750 nm, reaching a maximum of 0.79 at a wavelength of 710 nm, and a small correlation coefficient at wavelengths of 500 - 530 nm and 630 - 650 nm, and wavelengths greater than 780 nm. The chlorophyll content information is hardly reflected, and it can be seen that the multivariate scatter correction spectral value is more correlated with the chlorophyll content than the original spectral value.

3.2. Characteristic Wavelength Selection

A stepwise regression analysis method was used. According to the screening results, five sensitive wavelengths were selected, which were 712.5 nm, 454.09 nm, 561.22 nm, 530.4 nm and 987.91 nm. After the multi-scatter correction process, six sensitive wavelengths were selected by stepwise linear regression method, which were 712.5 nm, 509.95 nm, 561.22 nm, 840.62 nm, 696.67 nm and 987.91 nm, respectively.

3.3. Model Establishment and Inspection

A total of 130 samples, of which 98 were used as prediction set samples to establish partial least squares model, principal component regression model, stepwise regression model, and the modeled results were fitted with predicted values and measured values. The results are shown in Figures 3-5.

A partial least squares model, a principal component regression model, a stepwise regression model were established using 32 samples as test sets, and the predicted and measured values were fitted to the test results. The results are shown in Figures 6-8.
3.4. Comparison of Chlorophyll Content Prediction Models in Apple Leaves

The comparison between the original spectrum and the chlorophyll content prediction model after multivariate scatter correction treatment is shown in Table 1 and Table 2, respectively.

Figure 4. (a) PCA calibration by original spectral data; (b) PCA calibration of spectral data by MSC processing.

Figure 5. (a) Step by step calibration by original spectral data; (b) Step by step calibration of spectral data by MSC processing.

Figure 6. (a) PLS prediction by original spectral data; (b) PLS prediction of spectral data by MSC processing.
**Figure 7.** (a) PCA prediction by original spectral data; (b) PCA prediction of spectral data by MSC processing.

**Figure 8.** (a) Step by step prediction by original spectral data; (b) Step by step prediction of spectral data by MSC processing.

**Table 1.** Comparison of modeling set and detection set of chlorophyll estimation model before pretreatment.

<table>
<thead>
<tr>
<th>Model</th>
<th>Modeling R2C</th>
<th>RMSEC</th>
<th>REC</th>
<th>Prediction R2P</th>
<th>RMSEP</th>
<th>REP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS</td>
<td>0.3733</td>
<td>0.6235</td>
<td>16.81%</td>
<td>0.2225</td>
<td>0.6428</td>
<td>18.04%</td>
</tr>
<tr>
<td>PCR</td>
<td>0.7418</td>
<td>0.4004</td>
<td>10.56%</td>
<td>0.7320</td>
<td>0.3827</td>
<td>9.77%</td>
</tr>
<tr>
<td>Step by step</td>
<td>0.8046</td>
<td>0.3482</td>
<td>8.26%</td>
<td>0.5926</td>
<td>0.5439</td>
<td>14.74%</td>
</tr>
</tbody>
</table>

**Table 2.** Comparison of modeling set and test set of chlorophyll estimation model after multivariate scatter correction.

<table>
<thead>
<tr>
<th>Model</th>
<th>Modeling R2C</th>
<th>RMSEC</th>
<th>REC</th>
<th>Prediction R2P</th>
<th>RMSEP</th>
<th>REP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS</td>
<td>0.6791</td>
<td>0.4308</td>
<td>11.08%</td>
<td>0.6526</td>
<td>0.4354</td>
<td>11.09%</td>
</tr>
<tr>
<td>PCR</td>
<td>0.8121</td>
<td>0.3485</td>
<td>9.42%</td>
<td>0.8004</td>
<td>0.3188</td>
<td>26.4%</td>
</tr>
<tr>
<td>Step by step</td>
<td>0.8402</td>
<td>0.3149</td>
<td>7.19%</td>
<td>0.6899</td>
<td>0.4627</td>
<td>12.81%</td>
</tr>
</tbody>
</table>
It can be concluded from Table 1 and Table 2 that the fitting result of the modeling set of the principal component regression model in the predicted model of apple leaf chlorophyll determines the coefficient $R^2$ to be 0.8121 and the root mean square error RMSE to be 0.3485. For the 9.42%, the fitting result of the test set determines that the coefficient $R^2$ is 0.8004, the root mean square error RMSE is 0.3188, and the relative error RE is 26.4%. The principal component regression model after multivariate scatter correction processing is the optimal model.

3.5. Visualization of Chlorophyll Content Distribution in Apple Leaves

According to the established chlorophyll prediction model, the MSC-PCA model predicts the best chlorophyll content, and the regression equation is

$$Y_1 = 7.51 - 5.30 \times R_{714} + 43.70 \times R_{512} - 24.19 \times R_{563} - 9.09 \times R_{843} + 15.59 \times R_{698} + 2.25 \times R_{991},$$

it can be used to calculate the chlorophyll of each pixel on the blade.

In the ENVI environment, the mask background tool was used to remove the background, and the pure leaf hyperspectral image was extracted. The MSC-PCA model was used to perform the band calculation on the hyperspectral image of the apple leaf, that is, the pixel was gradually solved to obtain the chlorophyll of the apple leaf. Visualizing the distribution map, the value of each pixel in Figure 9 is the chlorophyll value of that point on the blade.

According to the figure, the distribution of chlorophyll on the leaves can be visually observed. Chlorophyll is more evenly distributed on both sides of the veins, and the chlorophyll content in the veins is lower than that in the mesophyll. The color of the tip of the blade is generally darker than the end, and the chlorophyll content of the first segment is higher than the end. Because the resolution of the hyperspectral image itself and the surface layer of the leaf contain a waxy layer, the image can only distinguish the distribution of the main veins. In addition, the edge part of the leaf indicates not only the estimation of chlorophyll content, but also the fluctuation of the edge of the blade. The resulting light is unevenly reflected. In summary, according to the MSC-PCA model, the distribution of chlorophyll in leaves can be estimated more accurately.

4. Conclusions

In this paper, the response characteristics of chlorophyll content in apple leaves were studied. The spectral characteristics of each band were analyzed. The spectral data were processed to model and predict the chlorophyll content of apple leaves. The best prediction model was selected to realize the distribution visualization.

1) The variation law of chlorophyll hyperspectral characteristics was analyzed, and the chlorophyll spectrum was processed in various ways. The correlation coefficient of chlorophyll corresponding spectrum after multi-scattering correction
Figure 9. Visual distribution of chlorophyll content.

spectra pretreatment was higher. The correlation coefficient reached 0.79 at the wavelength of 712.5 nm.

2) The prediction model of chlorophyll based on hyperspectral apple leaves was established, and the optimal model was selected. The fitting result of the modeling set of the principal component regression model determines the coefficient R² is 0.8121, the root mean square error RMSE is 0.3485, the relative error RE is 9.42%, and the fitting result of the test set determines the coefficient R² is 0.8004. The ratio is 0.3188 and the relative error RE is 26.4%.

3) A visual distribution map of chlorophyll in apple leaves was obtained. Using the established chlorophyll optimal linear regression model equation, the chlorophyll content of each pixel on the leaf was calculated, and the visual distribution map of chlorophyll in apple leaves was obtained, which provided intuitive and efficient technical support and more intuitive information expression for fruit tree nutrition monitoring. The results show that the visualization of chlorophyll cloth in apple leaves can be realized by hyperspectral imaging technology. The distribution of nutrients in leaves can further analyze the nutritional information of plants, and provide a scientific way to effectively detect the growth of plants and the distribution of nutrients.

Acknowledgements

This paper was supported by Shandong Major Scientific and Technological Innovation Project (2018CXGC0209), the National Natural Science Foundation of China (41671346), the Taishan Scholar Assistance Program from Shandong Provincial Government, Funds of Shandong “Double Tops” Program (SYL2017XTTD02), National Key Research and Development Program of China (2017YFE0122500).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

X. Wen et al.


