International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.8, No.12 (2015), pp.49-58
http://dx.doi.org/10.14257/ijsip.2015.8.12.06

Detection of Starch Content in Potato Based on Hyperspectral Imaging Technique

Wei Jiang[12], Junlong Fang[1*], Shuwen Wang[1] and Yongcun Fan[1]

[1]*School of Electrical Engineering and Information, Northeast Agricultural University, Harbin, 150030, China*
[2]*Department of Computin, Harbin Finance University, Harbin, 150030, China*

Abstract

Detection of starch content in potato is studied applying hyperspectral imaging technique in the paper. The original and preprocessing spectra were processed with partial least square(PLS) method to build prediction model of starch content. The original spectra between 400 and 1000nm was preprocessed with smoothing, second derivation, and multiplicative scatter correction (MSC). Prediction model was built with preprocessing spectra by applying principal component analysis (PCA). Known from the result, the model based on the preprocessing spectra preprocessed with smoothing and PCA is the best of all prediction models built in research. The determination coefficient (R^2)of calibration set and prediction set was 0.8234 and 0.9031 respectively. The root mean square error of calibration set (RMSEC) and root mean square error of validation set(RMSEV) was 0.5633 and 0.5025, respectively. It indicated that this method could be applied in detection of starch content in potato. The study could offer theoretical and practical reference for further study in the future.

Keywords: *Hyperspectral imaging, Potato, Starch, Partial least square*

1. Introduction

Potato has high economic value and is one of the most grain crops in the world. Potato riches in starch and starch content is generally 15%-18% in China at present, while starch content could reach 30% in some countries. The level of potato starch content is one of the main standard to measure the quality of potato [1]. It is important significance for potato breeding and deep processing to accurately and rapidly determine the content of potato starch. At present, it is wide to determine starch content by the traditional methods, but it is difficult to promote in the analysis and test for the large qualities of samples because of the long test time, high cost, not easy to master and inaccuracy result of determination [2].

Hyperspectral imaging technology is the combination of image technology and spectrum technology, It not only can be used to characterize image features of measured sample in the spatial distribution, but also to obtain its spectral properties from a pixel or groups of pixels [3-4]. Diffuse reflection is reflected from the incident surface after light coming into the potato inner interacts with the component of internal organization, and internal quality information of potato could be carried by the reflected light [5-7]. Hyperspectral imaging technology has been gradually applied to the quality detection field of fruit and vegetable due to the advantage acquiring spatial spectrum and image information of fruits and vegetables [8]. It is applied at the detection of internal quality for the moisture content of potatoes, total sugar, moisture content and hardness of bananas, sugar content and moisture content of Xuehua pear [9]. In this study, potato was used as experimental samples and hyperspectral imaging technology was utilized to obtain

* Corresponding Author

ISSN: 2005-4254 IJSIP

spectral curve information acquisition of potato, and the spectral information were extracted by the principal component analysis (PCA) method and then hyperspectral mathematical model about starch content was constructed according to PLS method, and finally this model was validated by a validation set. This model can provide efficient and accurate analysis for detection of potato starch.

2. Material and Method

2.1. Sample and Equipment

Taking potato as the research object, which were selected different varieties from around city of Harbin in Heilongjiang province. There are no surface defects can be used as experimental samples and the samples were cleaned by water for further experiment. The images of 118 samples were acquired, and the 100 images samples were randomly selected to construct model, and the other were taken as predict samples.

The image were acquired by hyperspectral image acquisition system made in HeadWall Company, USA, shown as Figure 1. The system consists of three parts, image acquisition unit, a light source and a sample conveying platform. The image acquisition unit includes an image spectrometer, CCD camera and lens. The light source is a fiber halogen lamp with 150W adjustable power. The slit width of hyperspectral image spectrometer is 25μm and the spectral range is from 400 to 1000nm. The spectral resolution is 1.29nm, and spacing of image acquisition band is 3nm, and the spatial resolution is 0.15mm.

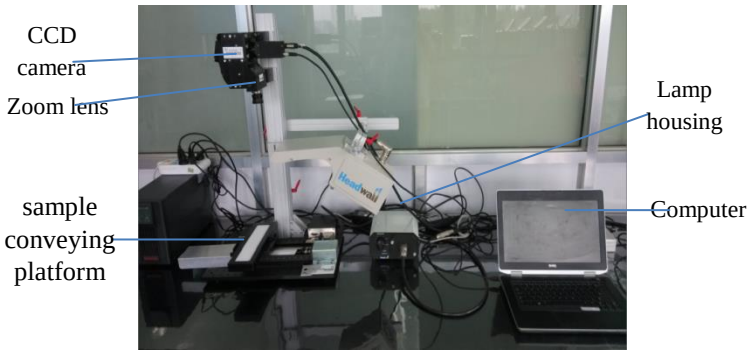


Figure 1. Hyperspectral Image System

2.2. Chemical Analysis Method of Starch Content

Measurement of starch content: homogenate precipitation method [10].

The principle of this method is that potato starch is easy to extract. According to the starch grain diameter is mostly between 20 and 60μm, starch is swashed through the sieve with 100μm aperture sieve, which sediment starch below the sieve, while the cellulose is above the sieve and protein, inorganic salt, sugar, soluble material are left in the water. The starch is filtrated, dried by air and weighed, and then starch content of potato tuber could be calculated.

The specific preparation process is: weighing potato, treating 20min at 5000 r/min by homogenate machine, screening with 120 mesh nylon sieve and swashing, screening with 180 meshes and swashing again, and finally filtering by using weight filter paper, drying by air and weighing. The starch content for the samples could be obtained as eqn. (1).

$$starch\ \ content\ \ \% = \frac{the\ \ total\ \ weight\ \ of\ \ filter\ \ paper\ \ and\ \ starch - filter\ \ paper\ \ weight}{initial\ \ sample\ \ weight} \quad (1)$$

2.3. Hyperspectral Analysis Method

2.3.1. Hyperspectral Image Acquisition: Before acquiring the hyperspectral image, according to the intensity of the light source the exposure time of spectral camera was set to ensure to get clear images, and the speed of conveying device was adjusted to avoid the distortion of image size and spatial resolution. After several debugging, the related parameters for ultimately determination is 26cm of object distance, 350ms of exposure time, 70W of the power for the light source. During each experiment process, a selected potato is placed on a white background plate, and then put on the objective stage. When acquiring image, linear detector is horizontal scan in the vertical direction of the optical focal plane. As the movement of objective platform, linear detector scans the image of the entire plane so as to complete the sample image acquisition.

2.3.2. Hyperspectral Image Correction: In different wavelengths, it is resulted in too large noise contained by weak band of light source distribution because of the influence of non-uniform distribution of light source intensity and the existence of dark current in the camera. So the hyperspectral image must be corrected. In order to reduce the interference of temperature change of light source on the image, when a total of acquisition image arrived at 20 each time, it must be gathered a whole white calibration image and a whole black calibration image, and the hyperspectral image was obtained and calibrated according to eqn. (2) [11-15].

$$R = \frac{R_s - R_d}{R_w - R_d} \tag{2}$$

Which, R is a calibrated image; Rs is the original sample image; Rw is the reference image acquired from the white spectral on calibration panel; and Rd is the dark current image obtained with a cap covering the camera lens.

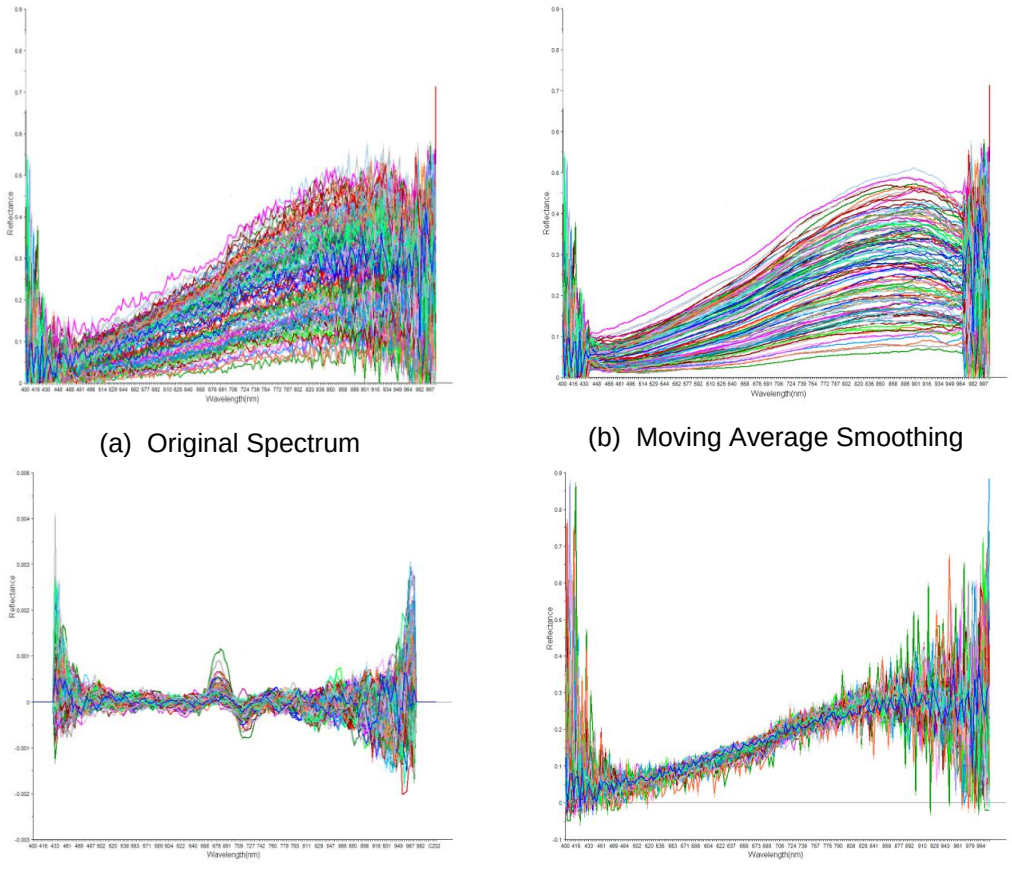
2.3.3. Data Processing of Hyperspectral Image

In order to weaken or eliminate the effect of baseline drift, scattering and other kinds of non-objective factors on the spectra, it is necessary to pretreat the spectra acquired hyperspectral imaging spectrometer [16]. In this paper, the original spectra firstly were processed by moving average smoothing (13 of average window size), the second derivative (11 of smoothing window size) and multiplicative scatter correction (MSC) and so on, and then the PCA method was used to extract the image feature and PLS was employed to construct the model, and finally the quantitative analysis model of the potato starch content was determined. Figure 2 shows the effect of different pretreatment on potato spectra. Which Figure 2a is the original spectral image, and Figure 2b is a spectra by smooth 13 points processing, and Figure 2c is a spectra by second derivative 11 points processing, and Figure 2d is a spectra by multiplicative scatter correction (MSC).

It can be seen from Figure 2 that there are more peaks and troughs by second derivative processing, these peaks and troughs are sharper by multiplicative scatter correction (MSC) processing and influence the effect of modeling. It is also compared to the influences between the original spectra and pretreatment spectra (moving average smoothing, second derivative and multiplicative scatter correction) for the calibration models, and according to the principles of the largest decision coefficient and minimum root mean square error, the pretreatment is determined.

2.3.4. Principal Component Analysis: Principal component analysis(PCA) is a projection along with the direction of maximum covariance from high dimensional data space to low dimensional data space. Because each component is independent each other,

it can be achieved for dimensionality reduction and elimination of redundant information. Principal component analysis principle is as follows [17]:



(a) Original Spectrum (b) Moving Average Smoothing

(c) Second Derivative (d) Multiplicative Scatter Correction

Figure 2. Preprocessing of Original Spectrum

A matrix $X_{n\times p}$ is a spectral matrix of samples, n is the number of samples, p is the number of variables (band). x_i (i=1,2,... The i sample, n) express the sample i, X_k (k=1,2,... , p) express the band k. V is the covariance matrix of sample.

Eigenvalue $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ is obtained by calculating the characteristic matrix V, the corresponding unit eigenvector is $u_1, u_2, \ldots, u_p$, and the principal component of sample k is shown as eqn. (3):

$$y_k = u_k^I X, k = 1,2,\ldots, p \tag{3}$$

Which, y_1 is the first principal component, and y_2 is the second principal component, … , by analogy.

For PCA, the proportion of the principal components k (y_k) to the total variance is shown as eqn. (4):

$$\lambda_k \Big/ \sum_{i=1}^{p} \lambda_i \quad (k = 1,2,\ldots, p) \tag{4}$$

Eqn. (4) is called the contribution rate of principal component y_k. The contribution rate of first principal component y_1 is maximum, and it is indicated that y_1 has the strongest

ability to explains the original variable x_k (k=1,2,…, p), while the explain ability of y_2, ..., y_p decrease successively.

The purpose using principal component analysis is to reduce the number of raw data variables, so the total number of principal components will not use in general and it will not result in great impact to the total variance to ignore some small variance of the principal components. The contribution rate for the first q principal components is called cumulative contribution rate. It can be used to represent the capability of original variable x_k (k=1,2,…, p) explained by $y_1, y_2, \ldots, y_q$. In generally, the dimension q of the principal components is usually smaller, so that the percentage of the cumulative contribution rate reached a higher value (usually above 85%). In order to, the principal component $y_1, y_2, \ldots, y_q$ can be used to instead of the original variables and achieve the purpose of reducing the dimension of the variables.

2.3.5. The Establishment of Regression Model: Partial least squares algorithm (PLS) is a common mathematical optimization techniques and it can get the best function of a set of data through the minimization of error square. Partial least squares regression (PLSR) is the most commonly modeling for the spectral analysis[18] and it can solve the problem which sample number is smaller than variable number. This method combines factor analysis with regression analysis, and decomposes the spectral matrix X and the concentration matrix Y by the principal component and establishes the liner regression model of matrix X and Y by relating the information of matrix X and Y, and then predicts the the unknown sample concentration. The principle is [19]:

First of all, the matrix X and Y are decomposed at the same time, following as eqn. (5) and (6):

$$X = T \bullet P + E \tag{5}$$

$$Y = U \bullet Q + F \tag{6}$$

Which, T and U are the score of matrix X and Y, respectively. P and Q are load matrix X and Y. E and F are the error matrix.

Then, T and U are established a linear relationship.

$$U = T \bullet B \tag{7}$$

Where B is the regression coefficient matrix.

Finally, the score T_{un} of an unknown sample X_{un} can be obtain by calibration set matrix P, and then Y_{un} can be calculated during the forecast process.

$$Y_{un} = T_{un} BQ + F \tag{8}$$

In this research, PLS models were established for whole band of original spectral and whole band of the spectral processed by smoothing, second derivative and multiplicative scatter correction respectively. And the other PLS model were constructed for the spectral which was firstly smoothed and then extracted features by the principal component analysis.

Three spraying quality indexes are mutually associated and constrained. More intensive spraying patterns provide low drift and bad uniformity of droplet distribution. Although uniformity of droplet distribution will be good spraying with small size droplets, more drift will come on. Uniformity of droplet distribution could be improved by increasing vertical distance between nozzle and target, but more drift would come on inevitable. So it

is very hard to get the most optimal parameters for all spraying quality indexes at the same time [13-15].

3. Result and Analysis

3.1. Chemical Values of the Starch Content

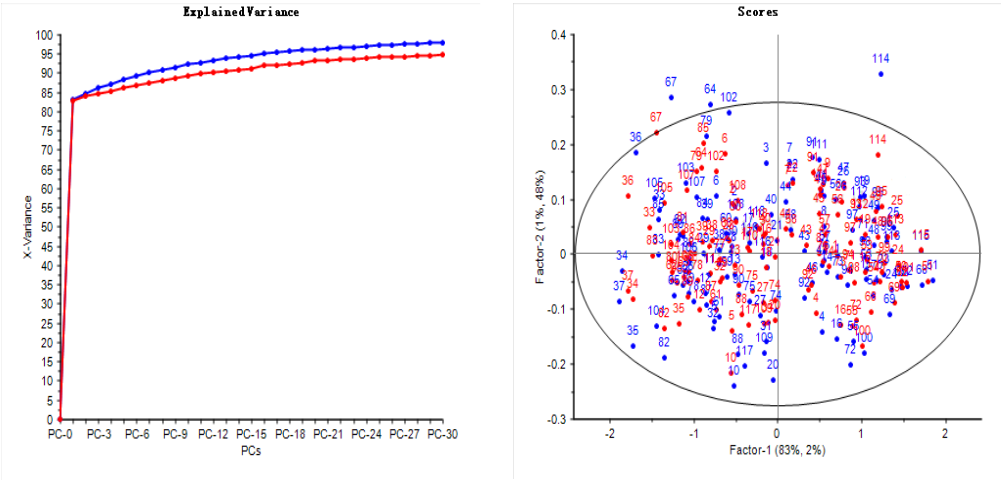
The statistical distribution of potato starch obtained by the homogenate precipitation method is listed in Table 1.

Table 1. The Statistical Distribution of Starch Content of Potato Samples

Sample classification	Number	Min(%)	Max(%)	Average(%)
Total	118	11.84	18.15	14.93
Calibration set	100	11.84	18.06	14.89
Validation set	18	12.16	18.15	15.18

3.2. Selection of Pretreatment Method

The Analysis for the mathematic model was realized by Matlab programming. Firstly, the different pretreatment methods was employed to analyze the whole spectra and then PCA treatment was used. The optimal number of principal components was determined and chosen by cross validation method, and finally the PLS model was established. The spectral range of hyperspectral image for potato test was set in 400~1000nm because the speed of processing data would be reduced for too large amount of data. In addition, a large amount of redundant information contained in spectral image data was resulted in the heavy correlation of the image of adjacent bands.



(a) Explained Variance (b) PCA Score

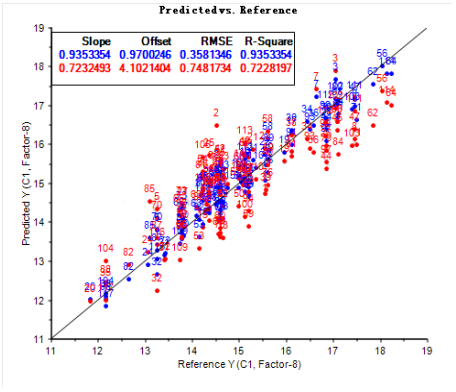
Figure 3. Spectral Smoothing -PCA Diagram

PCA was used to deal with the images of whole band original spectral and the spectral by preprocessed for potato and several principal components were extracted. PLS regression modeling was established and finally the optimal principal components of starch content was determined for potato. The spectral smoothing-PCA diagram is shown as Figure 3. Figure 3a expresses the variance explanation, and Figure 3b is the score matrix diagram. It can be seen from Figure 3a that the first principal component contribution rate was 82.4% based on principal component extraction principle, and the cumulative contribution rate was over 85% for the selected principal components. The number of principal component could be from 6 to 30. So PLS regression model could be

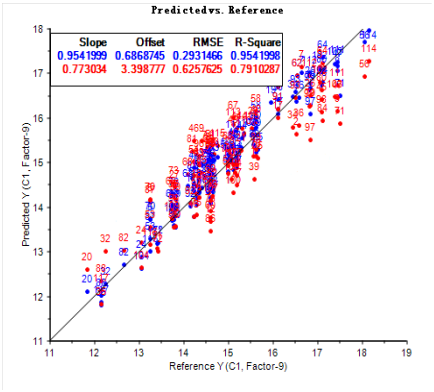
established by the different number of principal components instead of whole band data. It is the optimal model that the number of principal components is 15. According to Figure 3 (b), abnormal samples can be seen and eliminated, but abnormal samples were not removal in this research [20].

3.3. PLSR Modeling Analysis

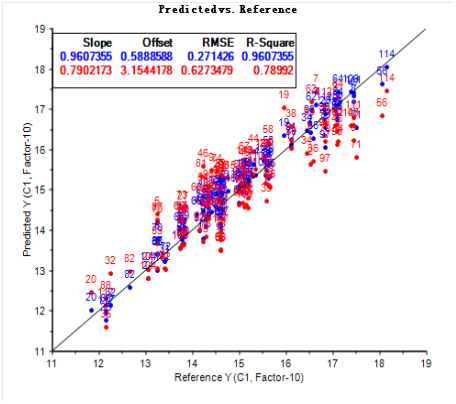
At first, the different spectral pretreatment methods were used to treat the original whole band, and then PCA method was employed. The optimal number of principal components was determined and chosen by the cross validation method, and PLS model was established. The prediction of PLS model for the set of validation samples was shown in Figure 4. Figure 4a was the original spectral modeling, Figure 4b was the spectral modeling by smoothing processing, Figure 4c was the spectral modeling by second derivative, and Figure 4d was the spectral modeling by multiplicative scatter correction(MSC), Figure 4e was the spectral modeling by smooth and PCA processing in which blue was reference value and red was the predicted value.



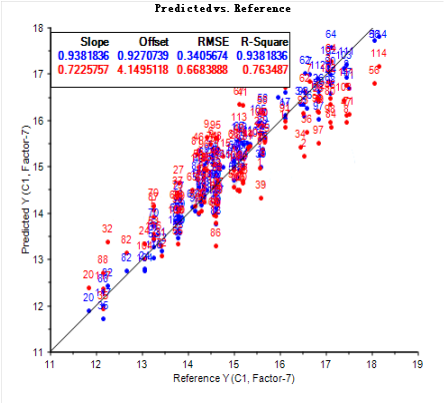
(a) Original spectrum



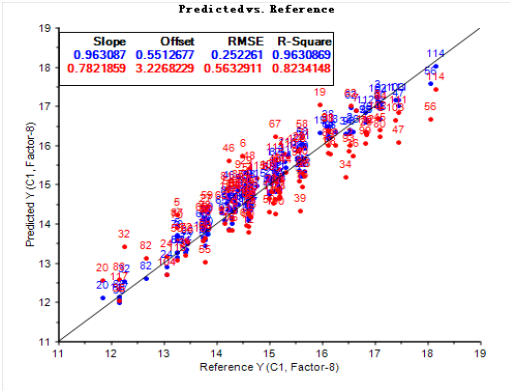
(b) Smooth correction



(c) Second derivative



(d) Multiplicative scatter correction



(e) Smooth -PCA-PLS

Figure 4. PLS Model of Calibration Set

Table 2. Performance of PLSR Model

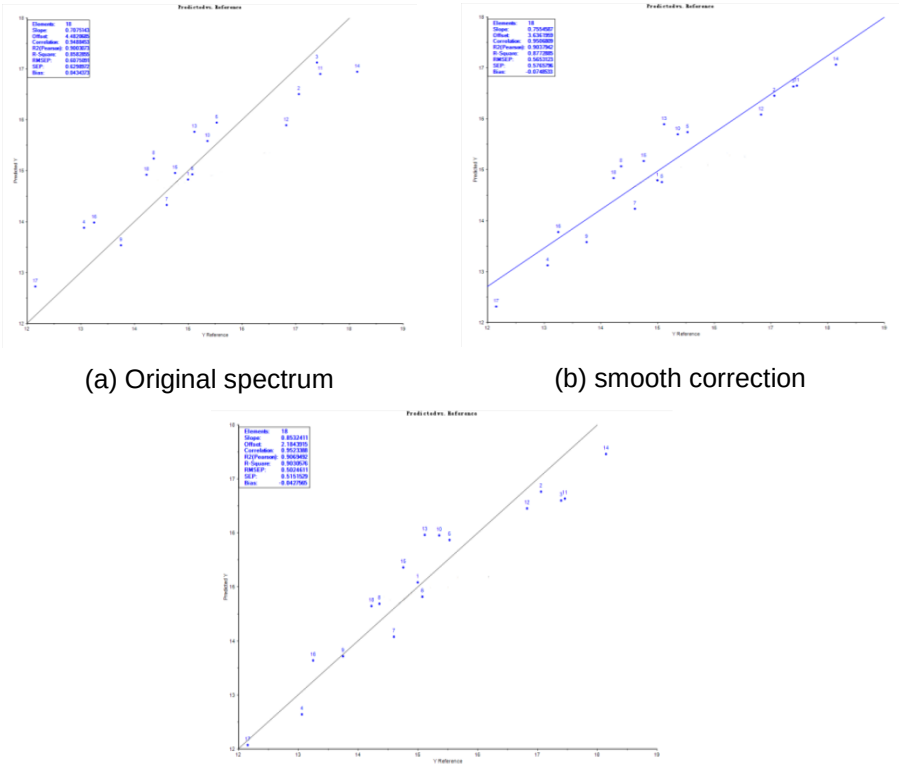
Method	Spectral points	Factors No.	RMSE	R^2
Original spectrum-PLS	203	8	0.7482	0.7228
Smooth-PLS	203	9	0.6258	0.7910
Smooth-PCA-PLS	15	8	0.56329	0.82341
Second derivative -PLS	203	10	0.62735	0.7899
MSC-PLS	203	7	0.6684	0.7635

According to Figure 4, it can be known that the prediction values and reference values of every model of corrected set are near the regression line, and the nearer distance between the points and line, the better prediction results are. It is seen from Figure 4a-4d that the prediction results of spectral model treated by the smoothing are the better for original spectrum of whole band and different pretreatment spectrum. better effect. And the predicted value and reference value were convergence in Figure 4e, and it is suggested that the prediction model of spectral data is the best treated by smoothing and extraction of principal component.

The effect of different methods of data pretreatment for the spectra of potato on the calibration model can be seen in Table 2. The determination coefficient R2 for the set of modeling and cross validation are high and root mean square error (RMSE) is low, so the models are accurate. The analysis model of spectrum is the best established by the combination with the treatment of smooth-moving and PCA. The determination coefficient R2 of calibration set is 0.9031 and RMSE is 0.5025. Therefore, according to the pretreatment method, the model is not only accurate, but also the spectral dimension can be reduced. It is advantage for rapid detection.

3.4. Validation of Regression Model

The accuracy and reliability of prediction for the hyperspectral validation model was corrected by 18 samples of the validation set. The correlation between chemical analysis values and prediction value of model for potato starch content was obtained by PLS regression. The scatter distribution of verification set of sample predicted by model is shown in Figure 5. The abscissa expresses the chemical values of starch content, and the ordinate shows the forecast value of the starch content. The sample is uniformly distributed around the regression line, and it can be seen that the slope of regression trend line of PCA is close to 1 and the correlation coefficient is 0.9031. It is suggested that higher significance level, it is the better prediction for the model [21].



(a) Original spectrum (b) smooth correction

(c) Smooth-PCA-PLS

Figure 5. Scattering Distribution of PLSR Model for Validation Set

4. Discussion

1) Potato spectral data were obtained by hyperspectral imaging technology. In the whole band range from 400 to 1000nm, comparing and analyzing the effect of models obtained by the different pretreatment such as the moving average smoothing, second derivative, and multiplicative scatter correction(MSC), it can be known that modeling effect and predicting accuracy is better treated by moving average smoothing method.

2) The hyperspectral image of potato by smoothed treatment was dealt with PCA. PLS regression model was constructed for the principal component of different number of reducing dimension. The starch content of potato can be accurately predicted by the first 15 principal component according to comparison and analysis. The determination coefficient R^2 of the prediction set reached 0.9031, and the root mean square error RMSE is 0.5025. Therefore, there is the higher accuracy of prediction for validation model established by hyperspectral image.

In a word, hyperspectral method has the characteristics of convenient, fast, accurate, non-destructive, non-pollution, and it is feasible to be used to measure the starch content of potato. It provide a rapid and non-destructive detection method for detecting the qualities of potato by the hyperspectral image technology, and it lays the foundation for the widely application of hyperspectral technology.

Acknowledgments

The authors would like to acknowledge the financial support from the National Science Technology Pillar Program during the Twelfth Five-year Plan Period "Research and demonstration for modern agriculture and machinery equipment" project

(No.2014BAD06B03); Special fund for system construction of a modem agricultural industry technology(No.CARS-10-P22).

References

[1] W. Chunying, C. Yili and S. Ying, "Research progress of Influence the quality for Fried potato", Chineese Potato, vol. 3, **(2003)**, pp. 23-24.
[2] Y. Chen, J. Yue and Y. Ding, "Determination of Starch Content in Potato with Acid-hydrolysis", Seed, vol. 28, no. 9, **(2009)**, pp. 109-120.
[3] M. Benxue, Y. Yibin and R. Xiuqin, "Advance in nondestructive detection of fruit internal quality based on hyperspectral imaging", Spectroscopy and Spectral Analysis, vol. 29, no. 6, **(2009)**, pp. 1611-1615.
[4] Z. Zhu, L. Xiaoyu, T. Hailong and G. Hailong, "Detection of potato external defects based on hyperspectral imaging technology", Transactions of the Chinese Sociely of Agricultural Engineering, vol. 28, no. 21, **(2012)**, pp. 221-229.
[5] Y. Peng and R. Lu, "Transaction of the ASAE", vol. 49, no. 1, **(2006)**, pp. 259.
[6] Y. Peng and R. Lu, "Journal of Food Engineering", vol. 82, **(2006)**, pp. 142.
[7] L. Renfu and P. Yankun, Biosystems Engineering, vol. 93, no. 2, **(2006)**, pp. 161.
[8] R Zhang, Y. Ying and X. Rao, "Quality and safety assessment of food and agricultural products by hyperspectral fluorescence imaging", Journal of the Science of Food and Agriculture, vol. 92, no. 12, **(2012)**, pp. 2397-2408.
[9] R. Zhang, "Detection of soluble solids content and firmness of tomato using hyperspectral imaging", Zhejiang University, **(2014)**, pp. 63-75.
[10] T. Feng, L. Yuzhen and M. Zhanlan, "The comparison of detection method for Starch Content in Potato", Science and Technology of Qinghai Agriculture and Forestry, **(1997)**, pp. 1.
[11] Z. Xiaoyan, L. Wei, X. Li, Z. Fengmin, Y. Yanchen and Y. Bingnan, "An Near-infrared Prediction Model for Quality Indexes of Potato Processing", Infrared, vol. 33, no. 12, **(2012)**, pp. 33-38.
[12] H. Min, Z. Xiao, Z. Qibing and F. Chaoli, "Based on corn seeds of hyperspectral image feature extraction and recognition", ACTA PHOTONICA SINICA, vol. 41, no. 7, **(2012)**, pp. 868-872.
[13] L. Dan, H. Jianguo, L. Guishan, H. Xiaoguang, W. Songlei and W. Longguo, "Non-destructive detection of moisture content in gherkin using Hyperspectral imaging", Infrared and Laser Engineering, vol. 43, no. 7, **(2014)**, pp. 2393-2397.
[14] Z. Zhu, L. Xiaoyu, G. Hailong, T. Hailong, L. Peng and W. Dongdong, "Comparison of Different Variable Selection Methods on Potato Dry Matter Detection by Hyperspectral Imaging Technology", Transactions of the Chinese Society for Agricultural Machinery, vol. 43, no. 2, **(2012)**, pp. 128-133.
[15] G. ElMasry, N. Wang and C. Vigneault,, "Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks", Postharvest Biol., Technol., vol. 52, no. 1, **(2009)**, pp. 1-8.
[16] L. Wanzhen, Y. Hongfu and X. Guangtong, "Modern analysis technique of near infrared spectroscopy [M]", Beijing: China Petrochemical Press.
[17] L. Bonian and W. Libin, "Data Analysis Method using MATLAB", Beijing: China Machine Press, pp. 106-116.
[18] P. Geladi and B R. Kowalski, Analytica Chimica Acta, vol. 185, **(1986)**, pp. 1.
[19] Z. Fengle, Z. Hailiang, S. Yongni and H. Yong, "Visualization of the Chilling Storage Time for Turbot Flesh Based on Hyperspectral Imaging Technique", vol. 34, no. 7, **(2014)**, pp. 1938-1942.
[20] M. Shungeng, L. Ning and Z. Mingxiang, "Outlier Diagnosis and Calibration Model Optimization for Near Infrared Spectroscopy Analysis", Spectroscopy and Spectral Analysis, **(2004)**, vol. 24, no. 10, pp. 1205-1209.
[21] H. Tiansheng, J. Jun, W. Ning, M. O.N gadi, Z. Zuoxi and L. Zhen, "Non-destructive in section of chinese pear quality based on hyperspectral imaging technique", Transaction of the CSAE, vol. 23, no. 2, **(2007)**, pp. 151-155.

Authors

Wei Jiang, born in 1980, in Zhaodong, China. PhD, Lecturer. Research Interests: Electrical Engineering, Science and Technology of Computer. E-mail: jwhancg@126.com

Corresponding author: Junlong Fang, PhD, professor, Research Interests: Electrical Engineering. Email:58525638@qq.com,
Mailing address: 59 Mucai Street, Harbin, 150030, China,
Tel: 13945697595.

