

Maturity stage distinction of pear based on visible/near infrared spectroscopy technology

Weiwei Xu ¹, Min Lin ^{1,*}, Yongmei Huang ¹, Huijun Liu ¹ and Xinqi Zhou ²

¹College of Metrology and Measurement Engineering, China Jiliang University
Hangzhou, China

²Focused Photonics(Hangzhou),Inc.,Hangzhou 310052,China

*Corresponding author, e-mail address: linm@cjljlu.edu.cn

Abstract. This paper selected pear as the research object, established a model for distinguishing different maturity stages of pear which bases on visible/near-infrared spectroscopy, discussed the influence of four types of spectral pretreatment, such as the Savitzky-Golay (SG) smoothing, first derivative correction, second derivative correction, and multiplicative scatter correction(MSC) on the model, and verified that competitive adaptive reweighted sampling(CARS) algorithm can effectively extract the effective information from the spectrum. Maturity stage distinction model of pears was established based on Visible/Near Infrared Spectroscopy Technology by using partial least squares discriminant (PLSDA). It analysed four different common pretreatment methods to improve the model effect. The result shows that the first derivative correction carries the best predicted accuracy, and the recognition rate is up to 86.11%.The information from the spectrum was filtrated from the pretreated spectrum by adopting CARS algorithm, to optimize the maturity stage distinction model of pears which was established before. The results demonstrates that the CARS algorithm can actually extract effective information, and the recognition rate is up to 97.22%.

1. Introduction

The pear which is used as the research object is mature in mid-august every year. If the picking is premature, the maturity is not enough, the fruits' quality will be poor, on the contrary, it will make against the storage and transportation. Both of them will result in the unsatisfactory economic income, so timely picking can maximize the economic benefit. Accurately judging the maturity stage of the fruit can not only be used to arrange the optimal time of picking, but also make for cultivating good seeds, storage, transportation and further processing[1].

Traditional fruit maturity differentiation method is roughly divided into two kinds [2]. One is that the practitioner judge based on the experience. The other is the physical and chemical index detection instrument. The former has the advantages of quick and intuitive, while accuracy is the biggest drawback. The latter has the advantages of the former, also with high accuracy, small volume, light weight, easy to carry, but still not perfect. Currently, hardness tester and the sugar meter are generally used to discriminate the fruit maturity stage, both of the two kinds of physical and chemical indicators testing instruments will damage fruit samples. Therefore, it can only take random sampling method, and the result has the occasionality, rather than universal because of the destructive.

This paper puts forward a pear maturity near-infrared detection method based on CARS algorithm, which is a fast and intuitive, nondestructive, high accuracy and universal assessment method of fruit maturity stage [3]. It not only includes the advantages of the traditional assessment method of maturity, but also overcomes the disadvantages.



2. Theory

2.1. Introduction of Spectral Pretreatment

This paper adopts four pretreatment methods, which is first derivative correction, second derivative correction, the Savitzky-Golay smoothing, and multiplicative scatter correction. Signal smoothing is a common method to eliminate noise, which can improve the signal-to-noise ratio (SNR) of high frequency noise, and the selection of filter width is very important [4]. First derivative correction and second derivative correction both can effectively eliminate interference from other background, distinguish overlapping peaks, and improve resolution and sensitivity [5]. Multiplicative scatter correction is used to eliminate the influence of scattering on the spectrum.

2.2. Competitive Adaptive Reweighted Sampling

Near infrared spectrum usually consists of a large number of data points. Wavelength points is far more than sample number, and the collinearity of the spectrum is very serious. The effective information is very weak, and the noise level is not consistent. Therefore, wavelength variable screening is a necessary pretreatment method. CARS algorithm is a new method of variable selection which is based on Darwin's survival of the fittest theory. The method can pick out the maximum absolute value of regression coefficient through each cycle constructed model by using the exponential decay function, and get the subset which has the minimum root-mean-square error of cross-validation (RMSECV) among N subsets by using cross validation. It is defined as optimal variable subset. This method can not only remove the non-information variables in the original spectral data, but also compress and remove the collinear variables, which can effectively be used in the selection of high spectral data variables [6].

2.3. Partial Least Squares Discrimination Analysis

PLSDA is a multivariate statistical analysis method for discriminant analysis. Discriminant analysis is a common statistical analysis method based on observation or measurement to determine how to classify the object. PLSDA can reduce the influence of multiple collinearity among the variables.

3. Experiment

3.1. Technological Process

The research process of this paper has these following steps: preparing the pear samples, obtaining the physical and chemical indicators, collecting spectra, selecting the best pretreatment method, extracting the characteristic wavelength of the pretreated spectrum by using CARS algorithm, and establishing the calibration model. That is the way to identify the maturity stages of the unknown pear samples.

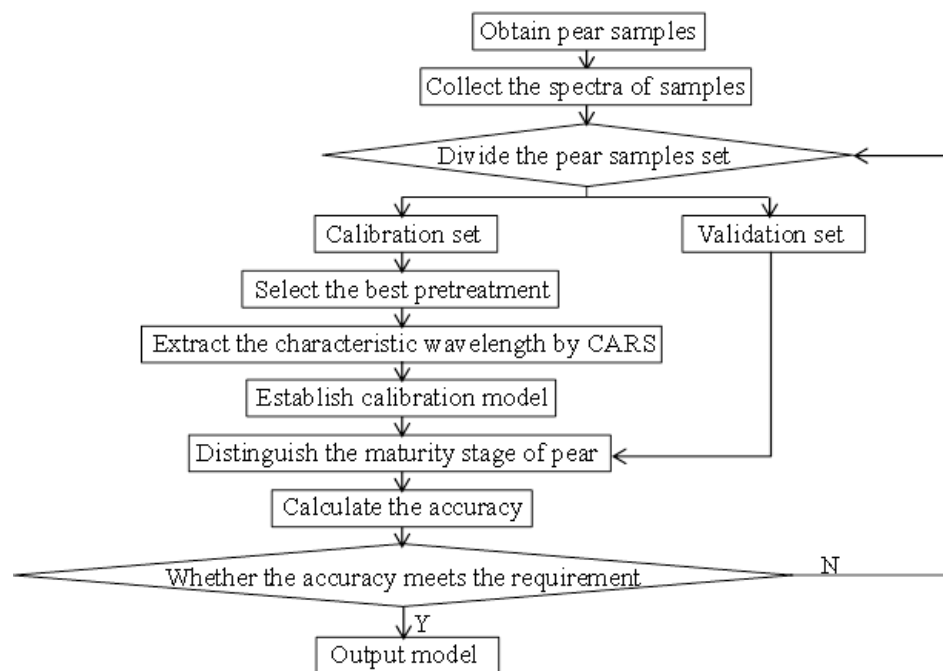


Figure 1. The flowchart of establishing a maturity stage discrimination model of pears

3.2. Sample Collection

The pear samples for experiment is from Hangzhou binjiang fruit industry co., LTD. There are totally 144 pear samples which should be numbered with 108 randomly selected as the calibration set and the rest of the 36 as the prediction set. The pear samples are divided into three maturity stages. As shown in Figure 2, it is the diffuse reflectance spectra Comparison of pear with three different maturity stages. The spectral range of the pear diffuse reflection spectra of the three maturity levels is 450nm - 1650nm.

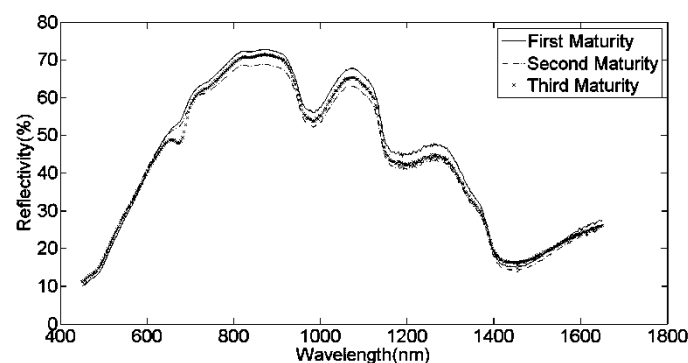


Figure 2. The diffuse reflectance spectra comparison of pear with three different maturity stages

3.3. Spectra Collection

The spectra collection instrument is a German Zeiss spectrometer with MCS600, and the spectrum collection area is 190nm - 2150 nm. The experiment is conducted at room temperature. In the experiment, each sample was scanned 12 times at the equator, and Ba₂SO₄ was used as the reference sample to take the mean spectrum. To avoid stray light interference, a 1.5mm gasket was used to protect the spectrum. The spectrum acquisition software is Aspect Plus.

3.4. Spectra Pretreatment

To reduce the noise in the spectrum and the useless information, this subject adopts four different spectral preprocessing methods, respectively, for the Savitzky-Golay smoothing, first derivative

correction, second derivative correction, and multiplicative scatter correction. The calibration model of the maturity discrimination of pear was established by using the Partial least squares discrimination analysis method, and 36 pear samples of the forecast set are used to test the effect of the model.

During the experiment, it is found that combining the original pretreatment method with autoscale will influence the model effect. Therefore, the results of the model prediction of single pretreatment mode and further autoscale processing will be discussed together. In addition, two kinds of multiple scattering correction are used to pretreat the spectrum: MSC (mean) and MSC (median). The former uses the average spectrum, while the latter uses the intermediate value spectrum. Partial modeling results are shown in table 1.

Table 1. The model effect of different pretreatment modes

Pretreatment mode	Filter width	Principal divisor number	Accuracy (%)
First derivative correction	11	6	80.56
First derivative correction &Autoscale	9	4	86.11
Second derivative correction	17	10	86.11
Second derivative correction &Autoscale	19	3	80.56
SG Smoothing	19	3	80.56
MSC(mean)	25	9	75.00
MSC(mean)&Autoscale	\	4	69.44
MSC(median)	\	7	75.00
MSC(median)&Autoscale	\	9	75.00

Table 2. The prediction result of the pear maturity identification model when pretreatment is the first derivative with filter width with 9 combined with autoscale

	Actual Class		
	Class 1	Class 2	Class 3
Predicted as Class 1	15	1	0
Predicted as Class 2	1	10	1
Predicted as Class 3	0	2	6

According to Table 1, the model effect of SG smoothing is not ideal in general. The recognition rate is low, and the model effect of first derivative correction pretreatment is significantly better than that of SG smoothing. The pretreatment method that combines first derivative correction with autoscale is superior to the single first derivative correction treatment, and the model effect of second derivative correction is inferior to the first derivative correction, but better than the SG smoothing. The effect of autoscale on second derivative correction pretreatment is negligible. However, for multiple scattering correction, further autoscale processing can be optimized. To sum up, the best pretreatment method is first derivative correction with filter width of 9 combined with autoscale, the predicted results are shown in Table 2, and Class1 is the first maturity stage, Class2 is the second kind of maturity, and Class3 is the third kind of maturity.

3.5. Characteristic Wavelength Extraction

The spectra of the first derivative correction (filter width of 9) with autoscale pretreatment are extracted and the characteristic wavelength will be picked out by the CARS algorithm. According to figure 3, with the increase of the number of running times, the number of wavelengths retained has decreased, indicating that the useless information in the spectrum has been removed, and it is a selection process. When the number of runs was 24, the RMSECV is the smallest, 0.4062%, and the retention wavelength number is 33. When the number of runs is greater than 24, the RMSECV becomes larger. It means that the valid information in the spectrum is also eliminated.

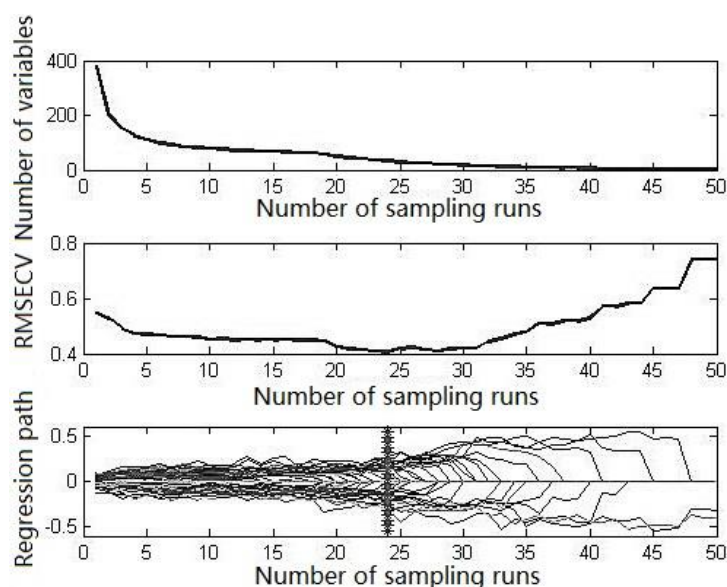


Figure 3. The variables selection of CARS with the first derivative correction (filter width of 9)

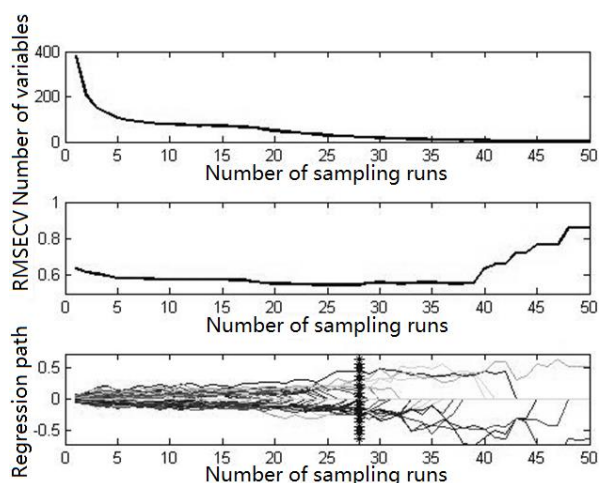


Figure 4. The variables selection of CARS with the first derivative correction (filter width of 17)

In addition, the model prediction result with the pretreatment method which is first derivative correction (filter width of 17) combined with autoscale is also good. Therefore, it is also carried out by the CARS algorithm to optimize the model, and the results are shown in figure 4. When the number of runs is 28, the RMSECV is the minimum, which is 0.5470%, and the number of retention wavelength is 21. The retention wavelength is less than that of the first derivative (filter width of 9), and the value of RMSECV is also greater.

3.6. Establish and Verify the Model

By using the CARS algorithm, the characteristic wavelength of the pretreated spectrum was screened and the model was reestablished. The pretreatment was no longer considered because the number of wavelengths was reduced. It can verify the model effect through forecast set, and when the pretreatment is first derivative correction (filter width of 9) combined with autoscale, the pear maturity identification model recognition correct rate is 97.22%. There is only a second stage pear sample is mistaken for the third, which is shown in table 3.

Table 3. The model prediction effect of the first derivative correction (filter width of 9) by CARS

	Actual Class		
	Class 1	Class 2	Class 3
Predicted as Class 1	16	0	0
Predicted as Class 2	0	12	0
Predicted as Class 3	0	1	7

However, the correct recognition rate of the pear maturity test model of the first derivative (filter width of 17) combined with autoscale is only 88.89%, which is shown in table 4.

Table 4. The model prediction effect of the first derivative correction (filter width of 17) by CARS

	Actual Class		
	Class 1	Class 2	Class 3
Predicted as Class 1	14	0	0
Predicted as Class 2	2	11	0
Predicted as Class 3	0	2	7

4. Conclusion

Difference of harvest period can cause noteworthy effect to the economic benefit. In order to improve the economic benefit of pears, this paper proposes a pear distinction method of near infrared detection which is based on CARS algorithm. It can help to arrange the best harvest of pear, as much as possible to guarantee the quality of the pear.

This paper compares the influence of four kinds of commonly used pretreatment methods on model, and concludes that the first derivative correction (filter width of 9) combined with autoscale pretreatment's effect is the best. Then it establishes a pear maturity stage identification model with good effect by using the pretreated spectrum, CARS algorithm and partial least squares discriminant analysis.

Contrast to the PLSDA model with full spectrum which does not extract characteristic wavelengths, those simplified by CARS algorithm has a greater accuracy of pear maturity identification, which raises from the original 86.11% to 97.22%. This shows that the CARS algorithm optimizes the model of the project.

5. Acknowledgment

This work is supported by National major scientific instruments and equipment development projects (2014YQ470377). The authors awfully thanks for anonymous reviewers who made constructive comments.

6. Reference

- [1] Zhou R and Li Y 2015 *J. Transactions of the CSAE*. **23** 255-259
- [2] He Y, Wei Y and Zheng Y 2010 *J. Northern Horticulture*. **33** 208-212

- [3] Ying Y 2006 *J. Spectroscopy and Spectral Analysis*. **26** 63-66
- [4] Li M 2013 *J. Journal of Green Science and Technology*. **50** 215-218
- [5] Xia J, Lu Y and Su Y 2015 *J. Acta Tabacaria Sinica*. **23** 19-22
- [6] Wu J and Xu Y 2011 *J. T Chin Soc Agric Mach*. **42** 162-166