International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 10, Issue 11 (November 2014), PP.17-22

Chemometrical Optimization for Fourier Transform Near Infrared Analysis of Sugar Brix in Cane Sugar Intermediate

Hua-Zhou Chen*, Ling-Hui Li, Ya-Juan Huo, Jiang-Bei Wen, Jie-Chao Chen
College of Science, Guilin University of Technology, Guilin, 541004, China

Abstract:- Real-time evaluation of sugar quality requires determining the content of sugar brix in the steps of the cane sugar process. Sugar brix is a key indicator for evaluating sugar quality. Fourier transform near infrared (FTNIR) spectroscopy is a simple, rapid and non-destructive technology on the analysis of material contents. In this study, the chemometric algorithm of parameter-combined tuning of Savitzky-Golay (SG) smoother and Partial Least Squares (PLS) regression was utilized for FTNIR analysis of sugar brix content in sugarcane clarified juice, an important intermediate product in cane sugar industry. The algorithms of combined optimization of SG smoother and PLS regression was achieved and the calibration models were established optimized by screening the expanded 540 SG smoothing modes and the 1-30 latent valuables (LV). The optimized models have high predictive accuracy. These results confirm that the combined optimization of SG smoothing modes and PLS LVs is effective in the quantitative determination of sugar brix contents in sugarcane clarified juice, and that the FTNIR spectroscopic technology with its chemometric algorithms have the potential in the analysis of cane sugar intermediates.

Keywords:- Sugarcane clarified juice; Sugar brix; Fourier transform near infrared spectroscopy; Partial least squares regression; Savitzky-Golay smoother; Chemometrical optimization

I. INTRODUCTION

This document is a template. An electronic copy can be downloaded from the conference website. For questions on paper guidelines, please contact the publications committee as indicated on the website. Information about final paper submission is available from the website.

Fourier transform near infrared (FTNIR) spectroscopy has the advantages of simple, rapid, non-destructive and reagent-free measurement, multi-component simultaneous determination, etc. It is a well-performed technology, widely applied to many fields, such as agriculture, food, environment, biomedicine [1-3]. In recent years, there are preliminary studies on the application of FTNIR to cane sugar industry [4-5]. Clarified juice is one of the important intermediate products, and sugar brix is an important indicator for evaluating sugar quality. Real-time evaluation of sugar quality requires determining the content of sugar brix in the steps of cane sugar process. Conventional chemical methods in laboratory cannot achieve the fast (or online) determination. Online fast detection of cane sugar intermediates is expected to be achieved by using FTNIR spectroscopic technology.

FTNIR spectroscopy with its chemometric methods owns the ability to output the perspective detection results in just a few minutes [6-7]. Partial Least Squares (PLS) is a common chemometric analytical algorithm integrating principal component analysis and multivariate linear regression. It can effectively eliminate spectral collinearity by creating comprehensive latent valuables. The tuning of latent valuables (LV) is the core procedure of PLS regression for noise reduction and information extraction [8-11]. Model predictive results will be reduced if LV is adopted too small or too large. Thus a reasonable LV should be selected by taking it as a tunable parameter.

As FTNIR spectroscopy includes kinds of noises generated in the detecting process. It is necessary to study the chemometric methods of data pretreatment to reduce spectral noises [12-14]. Savitzky-Golay (SG) smoother is a famous and widely used method for spectral data pretreatment [15-16]. Its major steps contain smoothing and differential, in which the smoothing mode is quite important for model improvement. There are many smoothing modes, determined by the three parameters of Order of Differential (OD), Degree of Polynomial (DP) and Number of Points (NP), and a specific smoothing mode outputs a corresponding calculation equation with its specific coefficients. Thus it is necessary to select a suitable SG smoothing mode and the best way to find it out is to screen it by tuning the three parameters combined with the optimization of PLS latent valuables.

The aim of this study is to determine the sugar brix content in sugarcane clarified juice by using the FTNIR spectrometry. Savitzky-Golay smoother is employed for data pretreatment and PLS is utilized for establishing calibration models. Model improvement is achieved by the combined tuning the LV of PLS and the parameters of SG smoother. For a wide-range optimization, we expand the tuning range of the three SG

smoothing parameters, and establish FTNIR calibration models with PLS regressions. The reasonable smoothing modes and the optimal PLS LV are simultaneously determined according to the model predictive results, in the combined computational algorithm platform. The selected pretreatment and modeling methods are examined by the prediction sample set, to have the potential of FTNIR modeling enhancement.

II. EXPERIMENT AND METHODS

A. Materials and Instruments

Eighty-three samples of sugarcane clarified juice were collected. The sugar brix content of each sample was measured using the traditional chemical methods and the measured values were used as the modeling reference values for FTNIR quantitative analysis. The sugar brix content ranges from 14.0 to 18.4 (%Bx) for the 83 samples.

We detected the FTNIR spectra of clarified juice samples using Nexus 870 spectrometer (Thermo Nicolet Corporation) with a 2-mm pathlength quartz cuvette. The whole scanning range was 4000-10000 cm^{-1}. An average of 32 scans per spectrum was made with a resolution of 8 cm^{-1}. Because the rotation of the cuvette cell can effectively reduce the unevenness, and multiple scans can effectively reduce the influence of background noise, we designed to measure the spectra while the cuvette cell is rotating. The temperature was controlled at 25±1°C and the relative humidity was at 46±1% RH throughout the spectral scanning process.

B. KS algorithm for Calibration-Prediction Partitioning

FTNIR spectroscopic analysis requires partitioning the samples into calibration set and prediction set. Calibration samples are used for model establishment and prediction samples for model evaluation. A suitable partition will lead to perspective modeling results. The Kernard-Stone (KS) algorithm is common used for sample partitioning in the spectroscopic field [17-18]. The classic KS algorithm is aimed at selecting a representative subset from the sample pool (*N* samples). In order to ensure a uniform distribution of such a subset along the *x* matrix (spectral response), KS follows a stepwise procedure in which new selections are taken in regions of the space far from the samples already selected. For this purpose, the algorithm employs the Euclidean distances $dx(j, k)$ between the *x*-vectors of each pair (j, k) of samples calculated as

$$dx(j,k) = \sqrt{\sum_{p=1}^{P} (x_j(p) - x_k(p))^2}\ ,\ \ j,k \in [1, N].$$

For spectral data, $x_j(p)$ and $x_k(p)$ are the instrumental responses at the *p*-th wavelength for samples *j* and *k*, respectively. *p* denotes the number of wavelengths in the spectra. KS algorithm selects the sample that exhibits the largest minimum distance with respect to any sample already selected.

C. Extension of Savitzky-Golay Smoother

SG smoother is a famous and widely-used pretreatment method to eliminate spectral noise. SG smoothing parameters include Order of Differential (OD), Degree of Polynomial (DP) and Number of Points (NP). For convenience, we denoted that the original spectral smoothing is 0th order differential. And NP is usually an odd number, denoted as NP=2*m*+1. It means that 2*m*+1 consecutive spectral data as a window, the spectral data in the window were fitted by using polynomial function whose independent variable was the serial number *i* of the spectral data, (*i*=0, ±1, ±2… ±*m*), and the polynomial coefficients were determined. Then the smoothing value and each order derivative value at the center point (*i*=0) of the window were calculated by using the determined polynomial coefficients. By moving the window in the whole spectral collecting region, the SG smoothed spectra and SG derivative spectra were obtained.

According to the above method, the smoothing value and each order derivative value at the center point of the window can be expressed as a linear combination of the measured spectral data in the window. The coefficients of the linear combination (i.e. smoothing coefficients) were uniquely determined by number of smoothing points (i.e. the number of points in the window), degree of polynomial, and order of derivatives. In Savitzky and Golay's paper [15], it was set that OD=0, 1, 2, 3, 4, 5, DP＝2, 3, 4, 5, and NP=5, 7… 25 (odd numbers). Different combinations of parameters correspond to different smoothing modes, and further correspond to different smoothing coefficient sets. There were a total of 117 smoothing modes (i.e. 117 sets of smoothing coefficients). The appropriate smoothing mode can be selected according to different analytes.

However, for the spectroscopic analysis of cane sugar intermediates, if the interval between spectral points was very small and number of points was small, then the window was narrow and the information in the window for smoothing was not sufficient. In this case, it was difficult to get satisfying smoothing effect. Hence, it was very necessary to expand the range of NP. In this paper, NP was expanded to 5, 7… 81 (odd), DP was expanded to 2, 3, 4, 5, 6, and the corresponding sets of new smoothing coefficients were calculated, so that a total of 540 smoothing modes were obtained including the original 117 modes, which is a SG smoothing group with a wider application scope.

D. Model Indicators

The model evaluation indicators mainly include correlation coefficient of predication (Rp), root mean squared error of predication (RMSEP) and the relative RMSEP (RRMSEP), which are calculated by the followings,

$$\mathrm{Rp}=\frac{\sum_{i=1}^{M}(C_{ip}-C_{mp})(C'_{ip}-C'_{mp})}{\sqrt{\sum_{i=1}^{M}(C_{ip}-C_{mp})\sum_{i=1}^{M}(C'_{ip}-C'_{mp})}},\quad \mathrm{RMSEP}=\sqrt{\frac{\sum_{i=1}^{M}(C'_{ip}-C_{ip})^2}{M-1}},\quad \mathrm{RRMSEP}=\frac{\mathrm{RMSEP}}{C_{mp}}\times 100\%,$$

where C'_{ip} and C_{ip} were predictive value and chemical values of the sample i in the prediction set, C'_{mp} and C_{mp} were the mean predicted value and mean chemical value of all samples in the prediction set, and M was the sample number in the prediction set.

The value of Rp is in coherent with RMSEP, usually that a higher Rp corresponds to a lower RMSEP. And, RRMSEP is always proportional to RMSEP. Thus, we take Rp and RMSEP as the main indicators for model optimization.

III. RESULTS AND DISCUSSIONS

The FTNIR spectra of 83 sugarcane clarified juice were showed in Fig. 1. The spectral responses contain the absorption information of many hydrocarbon groups in the clarified juice samples, such as sucrose, organic acids, amino acids and etc. In Fig. 1, we have the absorption of water molecules around 4200 cm^{-1} and 5250 cm^{-1}, and the absorption other than water appeared in the region of 5400-6000cm^{-1} and 6600-7200cm^{-1}. In order to reduce the interference of the water molecules, it is necessary to use SG smoother to deal with data pretreatment in spectral modeling.

The partitioning of calibration samples and prediction samples has to be finished before model establishment. Using the KS algorithm, we have the 83 clarified juice samples divided into 56 (for calibration) and 27 (for prediction). The mean value and the standard derivation of the measured sugar brix content for all calibration/prediction samples were showed in Table 1.

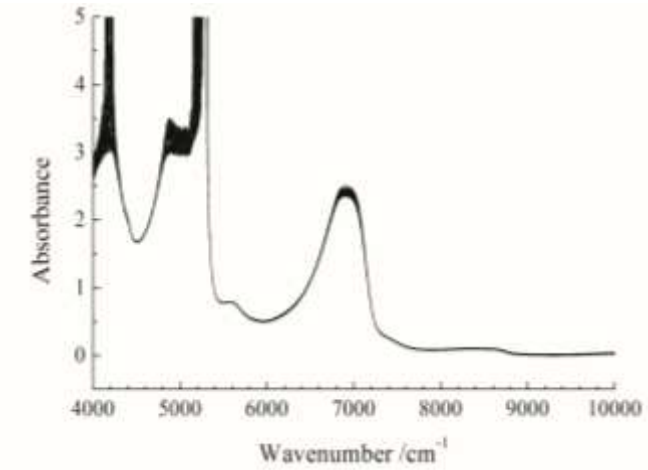

Fig. 1: FTNIR spectra of 83 clarified juice samples

Table 1: Mean value and standard derivation of the sugar brix content for calibration/prediction samples

	83 clarified juice samples	
	Mean value (%Bx)	**Standard deviation (%Bx)**
Calibration	**15.69**	**0.45**
Prediction	**15.85**	**0.58**

The optimized model parameters were selected for FTNIR analysis of clarified juice samples, by combining the 540 kinds of SG smoothing modes and the PLS LV tuning and optimizing, where the PLS LV was set changing from 1 to 30. The optimal models corresponding to each order of differential, with its parameters and predictive results, were showed in Table 2. In The non-smoothed full-range PLS modeling results was also listed in Table 2 for comparison. According to the maximum Rp (or minimum RMSEP), the optimal model FTNIR model of clarified juice output the predictive results of Rp=0.932, RMSEP=0.713%Bx and RRMSEP=4.5%, with the best smoothing mode of OD=2, DP=4 (or 5) and NP=61, and with the optimal LV=11.

It can be concluded from Table 2 that the model predictive results have high accuracy at each smoothing order of differential, when integrating the optimization of PLS models combined with SG smoother. And the modeling results are significantly superior in the PLS models with SG smoothing than without SG smoothing. The global optimal SG smoothing mode was 2nd order of differential and 4th (or 5th) degree of polynomial, and the optimal number of points was obviously larger than 25. These results indicated (1) the KS algorithm for sample partitioning lead to well-done calibration models; (2) the DP, NP of SG smoother and the LV of PLS altered in accordance with the varied OD; and (3) the NP of SG smoother is necessary to be expanded to the range of larger than 25.

Table 2: The optimal models corresponding to each order of differential, with its parameters and predictive results

	DP	NP	LV	Rp	RMSEP (%Bx)	RRMSEP
Non-smoothed	**—**	**—**	**10**	**0.885**	**1.131**	**7.1%**
0th order	**4, 5**	**53**	**10**	**0.906**	**0.821**	**5.2%**
1st order	**5, 6**	**79**	**9**	**0.924**	**0.741**	**4.7%**
2nd order	**4, 5**	**61**	**11**	**0.932**	**0.713**	**4.5%**
3rd order	**3, 4**	**69**	**9**	**0.917**	**0.797**	**5.0%**
4th order	**6**	**81**	**10**	**0.895**	**0.823**	**5.2%**
5th order	**3, 4**	**39**	**11**	**0.905**	**0.804**	**5.1%**

In tuning of SG smoothing parameters, we sketch the curves showing the RMSEP corresponding to each order of differential and each number of points, optimized from different DP of SG smoother and different LV of PLS (see Fig. 2). Fig. 2 also demonstrated that a lower-than-25 NP of SG cannot reach the minimum values of RMSEP and the expansion of NP would output the much optimal results.

LV of PLS is another key parameter for FTNIR modeling investigation. Fig. 3 showed the RMSEP values corresponding to the varied LV of PLS, optimized by SG smoothing mode, with 2nd order of differential and 4th or 5th degree of polynomial. The figure confirmed that the optimized LVs were larger than 10.

We drew out the comparative relationship between the FTNIR predicted values and the chemical measured values of the 27 prediction samples (see Fig. 4). We have the Rp larger than 0.9 and the RRMSEP smaller than 5%, which verified the potential of FTNIR spectroscopic technology applying to the quantitative analysis of cane sugar intermediates.

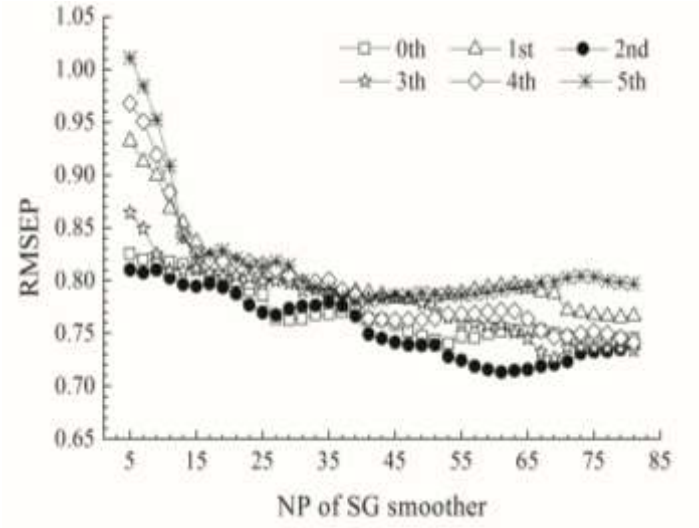

Fig. 2: RMSEP corresponding to each order of differential and each number of points (optimized from different DP of SG smoother and different LV of PLS)

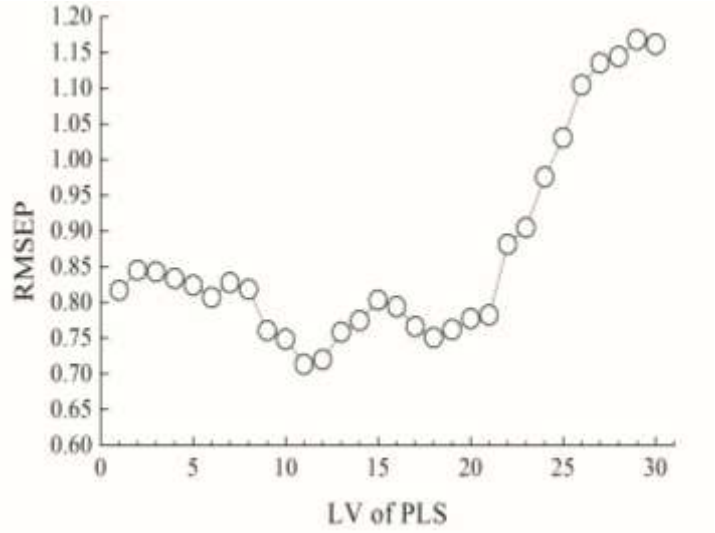

Fig. 3: RMSEP values corresponding to the varied LV of PLS (optimized by SG smoothing mode, with 2nd order of differential and 4th or 5th degree of polynomial)

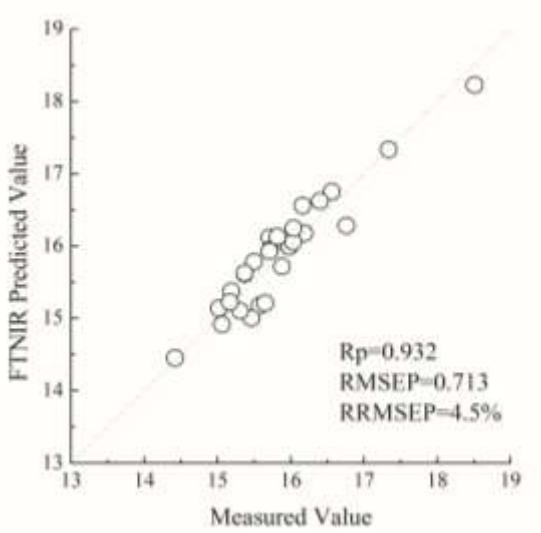

Fig. 4: The comparative relationship between the FTNIR predicted values and the Chemical measured values of the prediction samples

IV. CONCLUSIONS

The chemometric algorithm combined parameter-tuning of SG smoother and PLS regression was utilized for FTNIR spectroscopic analysis of sugar brix contents in sugarcane clarified juice, to establish and screen for the optimized calibration model. KS algorithm was used smoothly and seemed much effective in the partition of calibration and prediction samples. The algorithms of combined optimization of SG smoother and PLS regression was achieved and the calibration models were optimally established by screening the expanded 540 SG smoothing modes and the 1-30 LVs. The combined optimized calibration models have high predictive accuracy, and the optimized modeling results were quite appreciated. These results confirm that the expansion of SG parameters is quite necessary, and the combined optimization of SG smoothing modes and PLS LVs is an important method for the quantitative determination of sugar brix contents in sugarcane clarified juice. Our conclusions demonstrated that the FTNIR spectroscopic technology with its chemometric algorithms has the potential in the analysis of cane sugar intermediates. This rapid, non-destructive and reagent-free technology has practical meanings and is perspective in the online detection for cane sugar industry.

REFERENCES

[1]. D.A. Burns, E.W. Ciurczak, Handbook of near-infrared analysis, 3rd ed., Boca Raton: CSC Press LLC, 2006.

[2]. J.G. Wu, Modern Fourier Transform Near-Infrared Spectroscopy and Applications, Beijing: Science and Technology Literature Press, 1995.

[3]. A. Alishahi, H. Farahmand, N. Prieto, and D. Cozzolino, "Identification of transgenic foods using NIR spectroscopy: a review," Spectrochim Acta A., vol. 75, pp. 1-7, 2010.

[4]. N. Berding, and G.A. Brotherton, "Near Infrared Reflectance Spectroscopy for Analysis of Sugarcane from Clonal Evaluation Trials: II. Expressed Juice," Crop Science, vol. 31, l024-l028, 1991.

[5]. J.H. Meyer, "Near infrared spectroscopy (NIRS) research in the South African sugar industry," International Sugar Journal (Cane Sugar Ed.), vol. 100, pp. 279-286, 1998.

[6]. G. Cao, Z.W. Tan, J.N. Liang, and et al., "Mathematic Models for Analysis of Quality Components in Sugarcane Juice with Fourier Transform Near Infrared Spectroscopy," Scientia Agricultura Sinica, vol. 36, pp. 254-258, 2003.

[7]. H.Z. Chen, W. Ai; Q.X. Feng, Z. Jia, Q.Q. Song, "FT-NIR Spectroscopy and Whittaker Smoother Applied to Joint Analysis of Duel-Components for Corn," Spectrochim Acta A., vol. 118, pp. 752-759, 2014.

[8]. X.L. Chu, Y.P. Xu, and W.Z. Lu, "Research and Application Progress of Chemometrics Methods in Near Infrared Spectroscopic Analysis," Chinese Journal of Analytical Chemistry, vol. 36, pp. 702-709, 2008.

[9]. M. Daszykowski, M.S. Wrobel, H. Czarnik-Matusewicz, and B. Walczak, "Near-infrared reflectance spectroscopy and multivariate calibration techniques applied to modelling the crude protein, fibre and fat content in rapeseed meal," Analyst, vol. 113, pp. 1523-1531, 2008.

[10]. B. Igne, J.B. Reeves, G. McCarty, and et al., "Evaluation of spectral pretreatments, partial least squares, least squares support vector machines and locally weighted regression for quantitative spectroscopic analysis of soils," J. Near Infrared Spec., vol. 18, pp. 167-176, 2010.

[11]. S. Kasemsumran, Y.P. Du, K. Maruo, and et al., "Improvement of partial least squares models for in vitro and in vivo glucose quantifications by using near-infrared spectroscopy and searching combination moving window partial least squares," Chemometr. Intell. Lab. Syst., vol. 82, pp. 97-103, 2006.

[12]. H Chen, G Tang, Q Song, and W. Ai, "Combination of Modified Optical Path Length Estimation and Correction and Moving Window Partial Least Squares to Waveband Selection for the Fourier Transform Near-Infrared (FT-NIR) Determination of Pectin in Shaddock Peel," Anal. Lett., vol. 46, pp. 2060-2074, 2013.

[13]. X.L. Chu, H.F. Yuan, and W.Z. Lu, "Progress and Application of Spectral Data Pretreatment and Wavelength Selection Methods in NIR Analytical Technique," Progress in Chemistry, vol. 16, pp. 528-542, 2004.

[14]. K. Wang, G.Y. Chi, R. Lau, and T. Chen, "Multivariate Calibration of Near Infrared Spectroscopy in the Presence of Light Scattering Effect: A Comparative Study," Anal. Lett., vol. 4, pp. 824-836, 2011.

[15]. A. Savitzky, and M.J.E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," Analytical Chemistry, vol. 36, pp. 1627-1637, 1964.

[16]. A. Rinnan, F. Vandenberg, and S.B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," Trac-Trends in Analytical Chemistry, vol. 28, pp. 1201-1222, 2009.

[17]. H.Z. Chen, Q.Q. Song, G.Q. Tang, Q.X. Feng, L. Lin, "The Combined Optimization of Savitzky-Golay Smoothing and Multiplicative Scatter Correction for FT-NIR PLS Models," ISRN Spectroscopy, Volume 2013, Article ID 642190, 2013.

[18]. R.W. Kennard, and L.A. Stone, "Computer Aided Design of Experiments," Technometrics, vol. 11, pp. 137-148, 1969.

[19]. Q.B. Li, G.J. Zhang, K.X. Xu, and Y. Wang, "Application of KS Algorithm to the Calibration Transfer in Near-Infrared Spectroscopy," Spectrosc. Spect. Anal., vol. 27, pp. 873–876, 2007.