Prediction of Vitamin C Using FTIR-ATR Terahertz Spectroscopy Combined with Interval Partial Least Squares (iPLS) Regression

Diding Suhandy, Meinilwita Yulia, Yuichi Ogawa, and Naoshi Kondo

Abstract—In this study FTIR-ATR terahertz (THz) spectroscopy combined with interval PLS (iPLS) regression was used to measure concentration of vitamin C in aqueous solution. iPLS regression was used to select and develop the efficient spectral regions and variables in model calibration. The performance of the iPLS model than compared to that of full-spectrum PLS model. The result showed that iPLS model with 5 PLS factors was superior than that of full-spectrum PLS model with 11 PLS factors when 7 spectral regions and 70 variables were selected. Prediction performance of vitamin C can be improved by using iPLS model with RPD value of 4.570 could be obtained. This work demonstrated that concentration of vitamin C in aqueous solution can be predicted by FTIR-ATR THz spectroscopy method, and iPLS regression method revealed its superiority in model calibration.

I. INTRODUCTION

Vitamin C, also known as ascorbic acid, is one of the water-soluble vitamins. It can be found widely in most plant materials such as fruits and vegetables. In humans, however, vitamin C cannot be synthesized because we have lost the ability to produce *L-gulono- γ -lactone oxidase*, the enzyme necessary for its production [1]. For this reason, many of food products such as juices and sport drinks also contain vitamin C as one of its ingredients. Vitamin C is an essential nutrient and antioxidant and therefore has important role to keep health in good condition. That is why; there is increasing demand in vitamin C consumption both in fresh and processed product. Concentration of vitamin C in fresh food such as fruits and vegetables is subject to degradation. This concentration then will decrease due to some process like heating, cooking and etc.. For this reason, it is highly desired to quantify vitamin C present in foods especially in juice or sport drink for quality control purposes.

Many reported papers have established several non-spectroscopic methods for vitamin C determination. These conventional methods of vitamin C determination are including colorimetry, titrimetry, chemiluminescence,

Diding Suhandy is PhD student at Graduate School of Agriculture, Kyoto University. Kitashirakawa Oiwakecho Sakyo-ku, Kyoto 606-8185 Japan (corresponding author, e-mail: diding.suhandy@ fs2.ecs.kyoto-u.ac.jp).

Meinilwita Yulia is Master course student at Graduate School of Agriculture, Kyoto University. Kitashirakawa Oiwakecho Sakyo-ku, Kyoto 606-8185 Japan.

Yuichi Ogawa is Associate Professor at Graduate School of Agriculture, Kyoto University. Kitashirakawa Oiwakecho Sakyo-ku, Kyoto 606-8185 Japan.

Naoshi Kondo is Professor at Graduate School of Agriculture, Kyoto University. Kitashirakawa Oiwakecho Sakyo-ku, Kyoto 606-8185 Japan.

fluorometric, chromatographic and electrochemical method [2]. Each of these methods has limitation. For example titrimetry method using dichlorophenolidophenol as the titrant is rapid method for vitamin C concentration determination. However, the titrant itself is unstable and must be standardized before use. Chromatographic method is accurate, however it is expensive and time consuming.

In the previous work, Suhandy et al. [3] successfully used FTIR-ATR THz spectroscopy for vitamin C concentration determination in aqueous solution combined with full-spectrum PLS (FS-PLS) regression. However, for practical consideration it is important to develop a robust calibration model with high prediction performance. Using FS-PLS regression, it is possible to include irrelevant and unimportant wavenumber in calibration model and resulted in over-fitting. If over-fitting occurs, the prediction accuracy for samples that are not included in the calibration sample set becomes considerably low. To avoid this problem, it is very important step to construct a calibration model using only a selected subset of wavenumber instead of using full wavenumber. It is expected that robust calibration model with high prediction performance can be developed by using proper selected variables that contain only important and relevant information to the target variables.

In general, there are two kind of variable selection methods. The first one is single variable selection based method. The main objective of this method is to select proper wavenumber or wavelength to include in calibration model. In this method, the importance of each variable is directly calculated on the basis of the statistical features of the variables and calibration model. Uninformative variable elimination by PLS (UVE-PLS) proposed by Centner et al. [4] and competitive adaptive reweighted sampling (CARS) introduced by Li et al. [5] are example of this kind of variable selection method. The second one is selection method based on spectral region instead of using a single variable. It is based on the fact that a region of consecutive wavenumber or wavelength has to be selected simultaneously because spectra have continuous features. In this method, one or more important and relevant spectral region is selected for calibration model resulted in statistically stable and robust calibration model. Interval PLS (iPLS), moving window PLS (MWPLS), genetic algorithm PLS (GA-PLS) are example of this kind of variable selection method [6-8]. In recent work, genetic algorithm-based wavelength selection (GAWLS) was introduced by Arakawa et al.[9].

In this work, spectral region selection method based on iPLS was proposed to develop calibration model for vitamin C concentration determination in aqueous solution with high

prediction performance. Nørgaard et al., firstly proposed iPLS method [6]. The objective of iPLS is to split the spectra into some smaller equidistant subintervals, and develop PLS models on each subinterval. Then, the best subintervals are determined on the basis of the root mean squared error of cross-validation (RMSECV) values. In this study, FTIR-ATR THz spectroscopy technique was applied to predict concentration of vitamin C in aqueous solution. In calibration of PLS model, iPLS regression was attempted. Firstly, spectral data was split into some equidistant spectral subintervals by iPLS; then, subinterval which has lower RMSECV than the average of RMSECV was selected. Finally, iPLS model was developed using these selected subintervals. The prediction performance of iPLS model was tested by the independent samples in prediction sample set. The overall results were compared and discussed with previous report using full-spectrum PLS (FS-PLS) calibration model [3].

II. MATERIALS AND METHODS

A. Vitamin C

Vitamin C powder (L (+) - Ascorbic Acid, Wako Pure Chemical Industries, Ltd., Japan) were used to prepare vitamin C solution by dissolving powder in distilled water. The solutions were stirred well using mixer (Tube Mixer TRIO HM-, AS ONE, Japan). In this study, 55 samples of vitamin C solution will be used as samples. The concentration range of vitamin C solution was 0-21% (mass/mass (w/w)). For FTIR-ATR THz measurement, 300μL of vitamin C solution sample were dropped quickly onto the silicon prism surface using a micro pipette.

The samples were divided into two sample sets, calibration and cross-validation sample set (35 samples) and prediction sample set (20 samples). Using calibration and cross-validation sample set, calibration models were developed using PLS regression and full cross validation were conducted for each developed calibration models. Performance of calibration model finally was evaluated based on prediction result using prediction sample set.

B. FTIR-ATR THz spectral acquisition method

FTIR-ATR THz spectra of vitamin C solution were acquired in the range 20-450 cm^{-1} using 16 cm^{-1} of resolution. Each spectrum was the average of 100 scanning spectra. The reference of air was measured for every 5 samples. The intensity of sample and reference was obtained in single beam (SB) unit. During THz spectral measurement, the temperature and the related humidity were kept around 25°C and 70% in the laboratory, respectively. Using equation 1 we calculated the absorbance value of the sample.

$$A(\nu) = -\log_{10} \frac{S(\nu)}{R(\nu)} \dots (1)$$

Where:

A (ν) = Absorbance at wavenumber ν
S (ν) = Intensity of sample at wavenumber ν
R (ν) = Intensity of air reference at wavenumber ν

The calculated absorbance value was corrected using the ATR correction function provided in the software (JASCO Spectral Manager, JASCO Corp., Tokyo, Japan). The corrected value was used for further analysis.

C. Calibration model using interval PLS (iPLS) regression

Calibration and validation was developed using selected intervals. The average of RMSECV of subintervals was used as threshold value. Only subinterval which has lower RMSECV than the threshold was selected for iPLS regression model.

III. RESULTS AND DISCUSSIONS

A. Developing and evaluating calibration model using interval PLS (iPLS) regression

Using iPLS regression, calibration models were developed based on pre-processing of Savitzky-Golay first derivative spectra for 11 subintervals. Each subinterval consists of 10 variables. Table 1 shows the calibration results. In general, it is clear that lower frequency (less than 200 cm^{-1}) has lower RMSEC and RMSECV values than that of higher frequency, except for subinterval 1. It can be said that the performance of calibration model in lower frequency is better than that of

TABLE I
CALIBRATION RESULT USING iPLS REGRESSION

Subintervals	Wavenumber Range	R^2_{cal}	RMSEC	RMSECV
1	19–54	0.365	5.254	6.084
2	58 – 93	0.835	2.676	3.013
3	96 – 131	0.799	2.955	3.651
4	135 – 170	0.630	4.012	4.557
5	174– 208	0.736	3.383	3.901
6	212 – 247	0.610	4.118	4.493
7	251– 285	0.461	4.839	5.087
8	289 – 324	0.566	4.345	5.179
9	328 – 363	0.209	5.867	6.246
10	366 – 401	0.209	5.867	7.229
11	405 – 440	0.068	6.368	6.711

higher frequency.

The results presented in Table 1 suggested that inter-molecular vibration mode of THz in lower frequency has significant role in determining vitamin C concentration. It was consistent with a previous report on vitamin C and glucose concentration determination in aqueous solution using THz spectroscopy [3, 10]. In aqueous solution system of vitamin C, THz wave spectrum was dominated strongly by water absorbance. The presence of vitamin C has developed inter-molecular vibration mode between water and vitamin C and it results in weaker water absorbance. From this point, it

is clear that determination of vitamin C concentration in aqueous solution using THz wave has different in mechanism with other spectroscopy method. Using mid and near infrared spectroscopy, determination of vitamin C in aqueous solution samples was mainly driven by intra-molecular vibration mode arising from specific molecular bonding such as O-H and C-H.

B. Wavenumber selection based on iPLS regression

Based on Table 1, a variation of RMSECV and coefficient of determination (R^2) in 11 subintervals was plotted (Fig.1). In this study, the average value of RMSECV was used as a threshold for subinterval selection. Subinterval which has lower RMSECV than the threshold will be selected for iPLS regression model. As seen in Fig. 1, subintervals 2, 3, 4,5,6,7 and 8 has lower or almost same RMSECV values with the threshold. For this reason, these intervals are selected for iPLS calibration model.
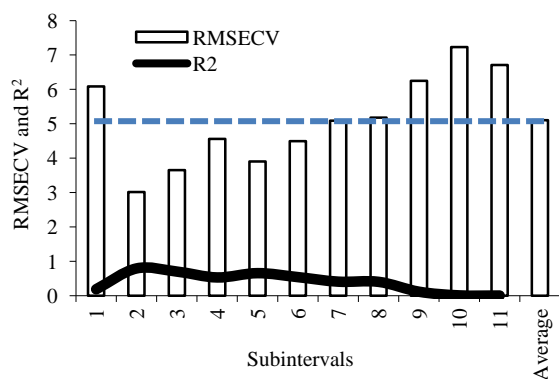

Fig. 1. Variation of RMSECV and R^2 calculated from PLS regression in each subinterval.

Calibration models which are developed based on iPLS regression method are prone to be over-fitting. To avoid this, determination of number of PLS factor to include in the calibration model is very crucial step. Several studies have been reported on the use of some statistical parameters for evaluating number of PLS factors such as root mean square error of cross-validation (RMSECV) and Durbin Watson (DW) values [11]. In this study, RMSECV was used for evaluation number of PLS factors. Fig. 2 shows a plot between RMSECV and number of PLS factors. It is clear that iPLS calibration model using 5 PLS factors has the lowest value of RMSECV. This calibration model was selected as the best iPLS calibration model and it will be used for prediction vitamin C concentration.
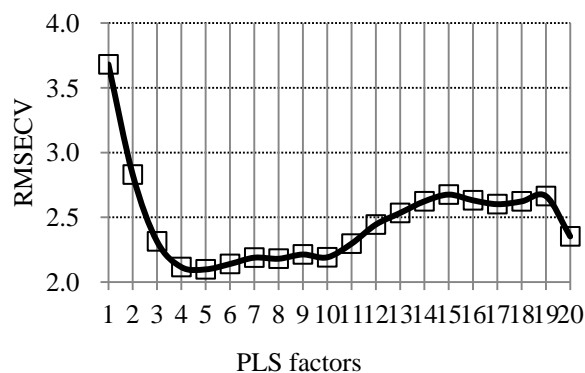

Fig. 2. Number of PLS factors versus RMSECV for the iPLS calibration model for vitamin C prediction.

C. Predicting vitamin C using iPLS calibration model

Fig. 3 shows the result for vitamin C concentration determination predicted based on iPLS calibration model. By a 95% confidence *paired t-test*, there were no significant differences between the actual vitamin C concentrations measured using the reference method and that predicted by FTIR-ATR THz spectroscopy. This result shows that a calibration model for spectroscopy-based determination of vitamin C concentration using FTIR-ATR THz spectroscopy could be developed.
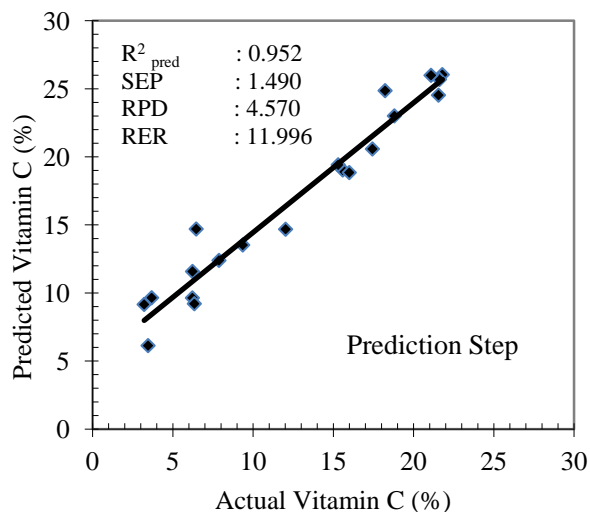

Fig. 3. Scatter plot of actual vs. predicted vitamin C calculated using iPLS calibration model.

The prediction result was better than that of vitamin C determination using full-spectrum PLS (FS-PLS) regression method. The SEP was 2.24% for vitamin C prediction using FS-PLS calibration model [3]. Using iPLS calibration model the SEP was improved to be 1.490%. The RPD and RER values were also better than that using FS-PLS calibration model.

IV. CONCLUSION

In this study, the feasibility of using FTIR-ATR THz spectroscopy combining with effective wavenumber selection algorithm used to measure vitamin C in aqueous solution was successfully demonstrated. iPLS results showed that lower frequency has better performance in vitamin C determination comparing to higher frequency. The influence of inter-molecular vibration mode in lower frequency is higher than that of intra-molecular vibration mode in higher frequency. Using iPLS regression with selected subintervals, prediction of vitamin C can be optimized with SEP was improved from 2.246 to 1.490. The RPD and RER values were also improved.

REFERENCES

[1] S.N. Gershoff, "Vitamin C (ascorbic acid): New roles, new requirements? ",Nutrition Review, 1994, vol. 52, pp.313-326.

[2] S, Arya, P.M. Mahajan, P. Jain, "Non-spectrophotometric methods for the determination of Vitamin C", Anal. Chim. Acta, 2000, vol. 417, pp.1–14.

[3] D. Suhandy, M. Yulia, Y. Ogawa, N. Kondo, "Vitamin C prediction in aqueous solution based on FTIR-ATR terahertz spectroscopy" (Submitted for publication)

[4] V. Centner, B.M., D.L. Massart, O.E. de Noord, S. de Jong, B.M., Vandeginste, C. Sterna, "Elimination of uninformative variables for multivariate calibration", Anal. Chem., 1996, vol. 68, pp. 3851-3858.

[5] H. Li, Y. Liang, Q. Xu, D. Cao, "Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration", Anal. Chim. Acta, 2009, vol. 648, pp. 77-84.

[6] L. Nørgaard, S. Saudland, J. Wagner, J.P. Nielsen, L. Munck, S.B. Engelsen, "Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy", Appl. Spectros., 2000, vol. 54, pp. 413-419.

[7] J.H. Jiang, R.J. Berry, H.W. Siesler, Y. Ozaki, "Wavelength interval selection in Multicomponent spectral analysis by moving window partial least-squares regression with application to mid-infrared and near-infrared spectroscopic data", Anal. Chem., 2002, vol. 74, pp. 3555-3565.

[8] Q. Chen, P. Jiang, J. Zhao, "Measurement of total flavones content in snow lotus (Saussurea involucrate) using near infrared spectroscopy combined with interval PLS and genetic algorithm", Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2010, vol. 76, pp. 50-55.

[9] M. Arakawa, Y. Yamashita, K. Funatsu, "Genetic algorithm-based wavelength selection method for spectral calibration", J. Chemom., 2011, vol. 25, pp. 10-19.

[10] D. Suhandy, T. Suzuki, Y. Ogawa, N. Kondo, H. Naito, T. Ishihara, Y. Takemoto, W. Liu," A Quantitative Study for Determination of Glucose Concentration Using Attenuated Total Reflectance Terahertz (ATR-THz) Spectroscopy", (Submitted for publication).

[11] A.A. Gowen, G. Downey, C. Esquerre, C.P. O`Donnell, "Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients", J. Chemom., 2011, vol. 25, pp. 375-381.