Engineering in Agriculture, Environment and Food
published by AABEA
Research Paper

EAEF 6(3) : 111-117, 2013

Prediction of L-Ascorbic Acid using FTIR-ATR Terahertz Spectroscopy Combined with Interval Partial Least Squares (iPLS) Regression*

Diding SUHANDY*1, Meinilwita YULIA*2, Yuichi OGAWA*2, Naoshi KONDO*2

Abstract

In this study iPLS regression was used to select the efficient spectral regions and variables to develop a calibration model for L-ascorbic acid (L-AA) determination using FTIR-ATR terahertz (THz) spectroscopy. The objectives of using iPLS were to improve the prediction performance of L-AA determination and to show mapping of contribution of high and low frequency in determining L-AA. The result obtained by iPLS model with 5 PLS factors was superior than that of full-spectrum PLS model with 10 PLS factors when 7 spectral regions and 70 variables were selected. Prediction performance of L-AA can be improved by using iPLS model with higher ratio prediction to deviation (RPD) value.

[Keywords] L-ascorbic acid, FTIR-ATR THz spectroscopy, iPLS regression, variable selection, calibration model, inter-molecular bonding

I Introduction

L-ascorbic acid (L-AA), also known as vitamin C, is a water-soluble vitamin. It is found widely in most plant materials, such as fruits and vegetables. In humans, however, L-AA cannot be synthesized because humans have lost the ability to produce L-gulono-γ-lactone oxidase, the enzyme necessary for its production (Gershoff, 1994). For this reason, many recently developed food products, such as juices and sport drinks are fortified with vitamin C.

L-AA is an essential nutrient and antioxidant and therefore, it has an important role in health. Hence, there is increasing demand to know the levels of L-AA consumed in both fresh and processed product. However, the concentration of L-AA in fresh food, such as fruits and vegetables, is subject to degradation. It is also subject to degradation during processing like heating, etc. (Lee and Kader, 2000). For this reason, it is highly desired to be able to quantify L-AA concentrations in foods, especially in juice or sports drink, throughout production and the distribution supply chain for quality control purposes.

Several non-spectroscopic methods for L-AA determination have been reported. These conventional methods of L-AA determination include colorimetry, titrimetry, chemiluminescence, fluorometric, chromatographic and electrochemical methods (Arya *et al*., 2000). Each of these methods though has limitations. For example, although the titrimetry method using dichlorophenolidophenol as the titrant is rapid, the titrant itself is unstable and must be standardized before use. Chromatographic methods are accurate, but expensive and time consuming. Recently, the use of a FTIR (Fourier transform infrared) spectroscopic based method for L-AA determination has become popular. It is simple, fast and free from chemical waste. With the rapid development of computers and FTIR instrumentation the price of equipment has come down and enhanced capabilities have been established. Today, an FTIR instrument is the standard for organic compound identification work in modern analytical laboratories.

In the previous work, Suhandy *et al*. (2012a) used Fourier transform infrared-attenuated total reflectance terahertz (FTIR-ATR THz) spectroscopy for L-AA concentration determination in aqueous solution combined with full-spectrum PLS (FS-PLS) regression. However, for practical application it is important to develop a robust calibration model with a high prediction performance. When using a FS-PLS regression, it is possible to include unrelated and unimportant spectrum characteristics in the calibration model that results in over-fitting. If over-fitting occurs, the model will generally have a poor predictive performance. To avoid this problem, it is a very important step to construct a calibration model with only a selected subset of wavenumbers instead of using all the wavenumbers. A robust calibration model with high prediction performance can be developed with properly selected variables that contain only the

---

* Partly presented at the IEEE / SICE International Symposium on System Integration SII2011 at Kyoto University, Kyoto, Japan in December 20-22, 2011.

*1 JSAM Member, Corresponding author, Graduate School of Agriculture, Kyoto University, Sakyo ku, Kyoto 606-8502, Japan; diding2004@yahoo.com
*2 JSAM Member, Graduate School of Agriculture, Kyoto University, Sakyo ku, Kyoto 606-8502, Japan

important and relevant information to the target variables.

In general, there are two methods for variable selection. The first one is a single variable selection based method. The main objective of this method is to select proper the wavenumber or wavelength to include in the calibration model. In this method, the importance of each variable is directly calculated on the basis of the statistical features of the variables and the calibration model. Uninformative variable elimination by PLS (UVE-PLS) proposed by Centner *et al.* (1996), iterative predictor weighting (IPW) PLS by Forina *et al.* (1999) and competitive adaptive reweighted sampling (CARS) introduced by Li *et al.* (2009) are examples of this kind of variable selection method. UVE-PLS extended along with wavelet transformation was proposed by Shao *et al.* (2004).

The second one is a selection method based on a spectral region instead of using a single variable. It is based on the fact that a region of consecutive wavenumbers or wavelength has to be selected simultaneously, because the spectra have continuous features. In this method, one or more important and relevant spectral region selected for the calibration model that result in a statistically stable and robust calibration model. Interval PLS (iPLS), moving window PLS (MWPLS), genetic algorithm PLS (GA-PLS) are examples of this kind of variable selection method (Nørgaard *et al.*, 2000; Leardi and Nørgaard, 2004; Jiang *et al.*, 2002; Chen *et al.*, 2010). Another example is searching combination MWPLS (SCMWPLS) as an extension of the MWPLS method (Du *et al.*, 2004; Kang *et al.*, 2006; Kasemsumran *et al.*, 2006). In recent work, genetic algorithm-based wavelength selection (GAWLS) was introduced by Arakawa *et al.* (2011).

In this work, the spectral region selection method, based on iPLS, was used to develop a calibration model for L-AA concentration determination in aqueous solution with high prediction performance. Nørgaard *et al.* (2000) first proposed the iPLS method. 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. The RMSECV is calculated as follows:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{n}} \tag{1}$$

where n is the number of samples in the calibration sample set, y_i is the actual value for sample i and $\hat{y}_i$ is the predicted value for sample i when the model is constructed with sample i removed (Romia and Bernardez, 2009).

In this study, the FTIR-ATR THz spectroscopy technique was applied to predict the concentration of L-AA in aqueous solution. When calibrating the PLS model, iPLS regression was used. First, the spectral data was split into equidistant spectral subintervals by iPLS; then subintervals which had lower RMSECV than the average RMSECV were selected. Finally, the iPLS model was developed using these selected subintervals. The prediction performance of the iPLS model was tested with the samples independent from the prediction sample set. The overall results were compared and discussed in comparison with results obtained in the previous report using a full-spectrum PLS (FS-PLS) calibration model (Suhandy *et al.*, 2012a).

II Materials and Methods

1. L-Ascorbic Acid (L-AA)

L-AA powder (L (+) - Ascorbic Acid, Wako Pure Chemical Industries, Ltd., Japan) was used to prepare L-AA solutions by dissolving the powder in distilled water. The solutions were stirred well using a mixer (Tube Mixer TRIO HM-, AS ONE, Japan). In this study, 55 samples of L-AA solution were used as samples. The concentration of L-AA solutions ranged from 0 - 21 % (mass / mass (w/w)).

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

2. FTIR-ATR THz device

THz spectra of L-AA solution were acquired using a Fourier-transform infrared (FTIR) based spectrometer (FARIS-1S, JASCO Corp., Japan) (Suhandy *et al.*, 2012a). In this spectrometer, a special light source (high pressure mercury lamp) was used as a THz generating device. This lamp has a low spectral intensity at low frequency (Hangyo *et al.*, 2000). To solve this problem, we used a silicon beam splitter which has a high transmittance in the terahertz region instead of mylar. A sample chamber with an ATR unit was installed on the spectrometer. In this ATR unit, a silicon prism is used as an internal reflection element (IRE). This has a high refractive index in terahertz region. The spectrometer was also equipped with a room-temperature pyroelectric sensor made from deuterated L-alanine triglycine sulfate (DLTGS) as a detector. Software of spectral manager for windows (JASCO Spectral Manager, JASCO Corp., Tokyo, Japan) was used to control the spectral acquisition process.