

Research Article

Quantitative Analysis of Multiple Components in Wine Fermentation using Raman Spectroscopy

Qiaoyun Wang, Zhigang Li, Zhenhe Ma and Guangyuan Si

College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning Province 110819, China

Abstract: Glucose and ethanol are critical quality control components in the wine fermentation process. In this study, we present a novel method in which Fourier Transform (FT)-Raman spectroscopy and chemometric techniques are used to quantitatively analyze ethanolic beverages produced by fermentation. Chromium (VI) Oxide (CrO_3) was flame-sealed into a fused silica cuvette and used as external standard. Band ratios between the Raman bands of the target molecule and that of CrO_3 were calculated and found to be proportional to the concentration of ethanol and glucose. This method can eliminate factors such as laser power or instrumental effects. After pre-processing the spectra, Principal Component Analysis (PCA) and Partial Least Squares (PLS) were selected as the multivariate calibration models. The prediction models proved to be robust resulting in a desirable mapping between the spectra and output attributes. This method could predict the ethanol and glucose concentrations simultaneously and produced a more linear calibration curve. As a result, there is a great potential to use Raman spectroscopy in wine fermentation process and on line fermentation control.

Keywords: Partial least squares, principal component analysis, quantitative analysis, Raman spectroscopy, wine fermentation

INTRODUCTION

In winemaking, fermentation is the primary process by which grape juice is converted into an ethanolic beverage. The ethanol and glucose concentrations play critical roles in the fermentation process. Therefore, a reliable and rapid on-line method is needed to measure the ethanol and glucose concentration during wine fermentation.

Gas chromatography and liquid chromatography are generally used to detect the individual components in wine. Chromatography can provide qualitative and quantitative information on a given sample by isolating its many components. However, this method is rather time-consuming and therefore it cannot be used as an on line detection method. Chromatography is also difficult to apply to water-rich samples. To minimize sample pre-treatment, these time-consuming methods have been replaced by vibration spectroscopic techniques for qualitatively identifying target compounds (Boyaci *et al.*, 2012; Di Egidio *et al.*, 2010; Gallego *et al.*, 2011; Numata and Tanaka, 2011). Infrared (IR) spectroscopy is a simple and reliable technique that is used in both research and industry to detect organic compounds. However, the OH stretching vibration produces very intense absorption bands in the spectra. Therefore, IR spectroscopy cannot be used for water-rich samples.

Real samples must be separated and concentrated prior to analysis, making the aforementioned analyses difficult to apply. Our primary goal is to develop a cheap, rapid and accurate analytical procedure to quantify ethanol and glucose concentrations for quality control purposes (Numata and Tanaka, 2011).

Raman spectroscopy is unique among spectroscopy techniques because liquid, gas or solid samples can be analyzed directly and rapidly without any sample preparation (Omar *et al.*, 2012). Raman spectroscopy offers several advantages over chromatography and IR spectroscopy in the analysis of water-rich and multi-component samples. For example, sample preparation is not required in Raman spectroscopy, which is therefore suitable for on-line quantitative analysis. The band intensities caused by the OH stretching vibration are also weak in Raman spectra, such that a water-rich sample can be directly analyzed (Numata *et al.*, 2011). Although Raman spectra can be collected quite easily, the procedure for quantifying the spectra is rather cumbersome because the intensity of the Raman spectra depends on the sample concentration as well as other factors, such as the laser power and the instrumental effects. Therefore, the correlation between the absolute intensity of the Raman spectra and the concentration is not good. And the standard is needed to eliminate the effects of the laser power and the instrumental. To

Corresponding Author: Qiaoyun Wang, College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

remove these effects, a quantitative method has been developed in which a reference of known concentration is used as an internal or external standard. The band ratios between the Raman band of a target molecule and that of the standard can be calculated and used in quantitative analysis (Favors *et al.*, 2005; Nah *et al.*, 2007; Numata *et al.*, 2011). The internal standard is the solvent or an added component. So the chemical interaction may be occurred between the analyte and the standard. An external standard can also be used to eliminate factors such as laser power or instrumental effects. The Raman spectroscopy in which both the external standard and the target sample can be measured simultaneously have been developed already (Favors *et al.*, 2005; Nah *et al.*, 2007). However, since these instruments are very complicated and expensive. The method in which the Raman spectroscopy of the external standard is measured in advance was adopted in (Numata *et al.*, 2011). But this method cannot measure the external standard and the target sample spectrum simultaneously.

In this study, we used a different method to eliminate the effects of the laser power and the instrumental effects. CrO₃ was used as an external standard in developing a quantitative Raman method. The Raman spectra were related to the concentrations using Principal Component Analysis (PCA) and Partial Least Squares (PLS) analysis. Several pre-processing and correlation methods were applied to the data to obtain optimized models (Diakabana *et al.*, 2014). The Root-Mean-Square Error of Prediction (RMSEP), the Root-Mean-Square Error of Cross Validation (RMSECV), the correlation coefficient (r^2) and the Residual Predictive Deviation (RPD) were examined to optimize the model. In this study, RMSECV = 0.003614, RMSEP = 0.004282, $r^2 = 0.99827$ and RPD = 6.3427 were obtained for the ethanol concentration, RMSECV = 0.00327,

RMSEP = 0.003052, $r^2 = 0.99861$ and RPD = 19.0575 were obtained for the glucose concentration. These results represent the potential and reliability of Raman spectroscopy for the on line quantification of ethanol and glucose concentrations that is critical for quality control in wine fermentation.

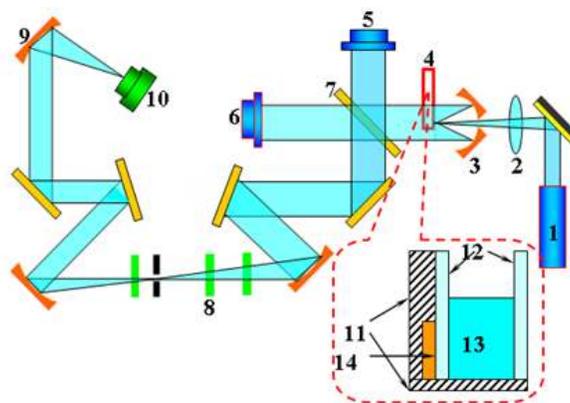
METHODOLOGY

Preparation of standard solution and calibration curves:

Absolute ethanol and glucose were locally purchased as commercial products and used without further purification. Standard solutions at several concentrations were obtained by diluting the pure products with water.

Forty samples (juice and ferments) with different ethanol and glucose concentrations (ranging from 0.05-0.38 V/V for ethanol and 3-50 g/L for glucose) were prepared to produce the calibration curves. All of the wine samples were stored below 5°C to prevent any changes in their characteristics during spectral acquisition and laboratory testing.

Raman spectroscopy: The Raman spectra of samples were obtained with a Bruker MultiRAM (Bremen, Germany) Fourier Transform (FT) Raman spectrometer equipped with a germanium detector using liquid nitrogen as the coolant. The excitation light was generated by a near infrared Nd:YAG laser at 1064 nm. The laser light with a power of 500 mw was introduced and focused on the sample and the scattered radiation was collected at 1800. All of the Raman spectra recorded in the 4000-400 cm⁻¹ range using a spectral resolution of 6 cm⁻¹ and a total of 512 scans were averaged for each spectrum. The OPUS 7.0 (Bruker Optics, Germany) software program was used for Raman spectral data acquisition. A quartz cuvette



1.Nd:YAG laser 2.optical lens 3.parabolic focusing mirror 4.quartz cuvette with sample 5.fixed mirror 6.moving mirror 7.beam splitter 8.optical filter 9.focusing mirror 10.Ge detector 11. quartz glass with mirror and single side polished 12. quartz glass with double side polished 13. liquid sample 14. Chromium (VI) oxide

Fig. 1: The structure of FT-Raman system and the cuvette

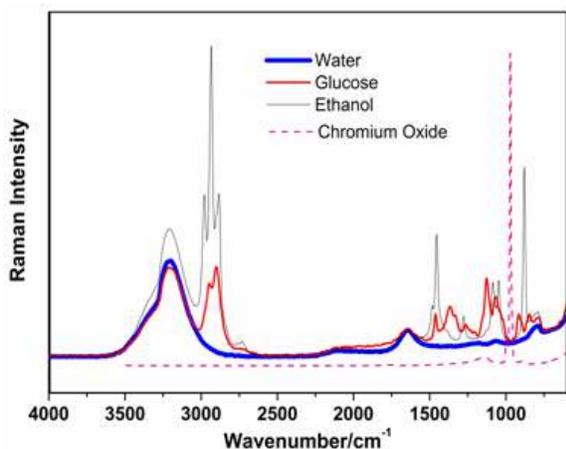


Fig. 2: Raman spectrum of 10% V/V ethanol, 25 g/L glucose, water and chromium oxide

with optically flat faces, a mirrored backside and near Infrared (IR) transparency was used as the cell for the liquid samples. CrO_3 was flame-sealed into the cuvette and used as an external standard to correct for the effects on the Raman intensity from the variations in the laser power. This procedure allowed for the mixed Raman spectrum of the CrO_3 standard and the sample to be collected synchronously without any extra optical configuration. The top of cuvette was sealed with a Teflon plastic cap to prevent the liquid from evaporating. The structure of system is shown in Fig. 1. The thicknesses of the quartz glass with double side polished and single side polished are 0.5 and 1 mm, respectively. The thickness of the CrO_3 is about 0.1 mm.

To eliminate the effects of the test environment, all of the experiments were performed at 25°C. Before acquiring the Raman spectra, the samples were warmed from 5 to 25°C and left at 25°C for 1 h. Then, a sample was placed in a cuvette sample holder and its Raman spectrum was obtained. To minimize the vaporization of ethanol in the samples, the Raman spectra were run immediately after sample preparation.

The Raman spectrum for the ethanol solution and glucose solution, CrO_3 and water are shown in Fig. 2. The ethanol and glucose bands can be clearly observed. The intense bands at approximately 2900 cm^{-1} correspond to water. However, there is generally little interference between the water, glucose and ethanol and therefore this region cannot be neglected in the quantitative analysis.

HPLC measurements for reference data: In our experiments, the ethanol and glucose concentrations were determined using High-Performance Liquid Chromatography (HPLC) (Merck Hitachi, Japan). Before sample detection, all of the samples were filtered through a 0.45 μm pore size membrane filter. Then, the analysis was performed via isocratic elution using 0.01

N sulfuric acid at a flow rate of 0.7 mL/min. The column temperature was 60°C and the injection volume was 20 μL . The analysis was carried out in triplicate. The HPLC results were used as reference data in the Raman analysis.

Data pre-processing: The spectra data were obtained using an OPUS 7 (Bruker Optics Inc.). Data pre-processing is a key step in multivariate analysis because appropriate data treatment is needed to develop the best-fit model. The most common pre-processing techniques were applied to enhance the spectra, such as Multiplicative Signal Correction (MSC), Standard Normal Variate correction (SNV), Savitzky-Golay (SG) smoothing, Direct Orthogonal Signal Correction (DOSC) and filtering and first and second order derivation. In addition to these pre-processing techniques, baseline correction, spectra normalization and combinations of these two methods have been used (Omar *et al.*, 2012). To obtain the best model, different mathematical pretreatments were used to remove or minimize unwanted spectral contributions. In this study, SG Filter, SNV, MSC and derivative transformations were combined with a baseline correction.

The pre-treated spectral data were processed using Principal Component Analysis (PCA). PCA identifies the orthogonal directions of the components with the maximum variance in the original dataset in decreasing order and projects the data onto a lower-dimensionality space formed by a subset of the highest-variance components. The orthogonal directions are linear combinations of the original variables and each component corresponds to a part of the total variance of the data. The first significant component corresponds to the largest percentage of the total variance, the second significant component to the second largest percentage of the total variance and so forth (Di Egidio *et al.*, 2010). And the PCA was used to eliminate the defective spectra outlier. Scores, loadings and explained variance were studied for the first 4 Principal Components (PC). And the regression coefficients of principal component score and composition content are shown in Table 1.

PLS was combined with data pre-processing to predict the ethanol and glucose concentrations simultaneously. The samples were separated into two groups by employing the Kennard-Stone algorithm (Kennard and Stone, 1969) and the optimum number of latent variables was chosen by the root mean square error of cross validation obtained from the calibration set by internal validation (leave one out). A total of 50 samples were randomly separated for calibration (= 80%) and validation (= 20%). The sets spanned the entire range of values; thus, there was no need to arbitrarily assign the samples to the sets. In our study, the following ranges of Raman spectra were applied in PLS model construction: 2500-3500 and 700-1500 cm^{-1} .

Table 1: Regression coefficient of principal component score and composition content

| Composition content | Principal component | | | | | | | | | |
|---------------------|---------------------|---------|---------|---------|--------|--------|---------|---------|--------|---------|
| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 | PC10 |
| Ethanol | 0.25820 | -0.0189 | -0.1376 | 0.39529 | 1.8080 | 0.2142 | 0.0817 | -0.3894 | 0.3887 | -3.9248 |
| Glucose | 0.10263 | 0.1550 | -0.1089 | 0.04370 | 0.0287 | 0.1770 | -0.0522 | -0.1170 | 0.0940 | 0.1045 |

In this study, the prediction capacity of the calibration models was evaluated using several parameters:

- The Root Mean Square Error of Prediction (RMSEP), which was used to determine the average prediction error or accuracy of the calibration model
- The Root Mean Square Error of Cross Validation (RMSECV), i.e., the prediction error of the calibration model, which is defined as the standard deviation in the differences between the spectral data and the reference values in the cross-validation sample set
- The correlation coefficient (r^2)
- The Residual Predictive Deviation (RPD) that is related to the precision of the PLS model (Niu *et al.*, 2012). These parameters are defined by Eq. (1) (Ozturk *et al.*, 2012):

$$\begin{aligned}
 RMSEP &= \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)_{pred}^2}{n_p}} \\
 RMSECV &= \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)_{calib}^2}{n_c}} \\
 RPD(\%) &= \frac{SD}{RMSEP} \\
 SD &= \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}
 \end{aligned} \tag{1}$$

where,

- y_i = The actual value obtained from HPLC methods
- \hat{y}_i = The predicted value by Raman spectra for the same sample
- \bar{y} = The mean of each set
- n_c = The number samples used in calibration set
- n_p = The number samples used in validation set
- n = The number samples used in each set
- S.D. = Standard Deviation in each set

Accuracy of the prediction models were evaluated by using R^2 and RPD determined for the calibration sets. In addition to these criteria; to construct a good model, the absolute values of RMSECV and RMSEP and differences between them should be small. The RPD value was used to check the robustness of the model and the higher RPD values were used for prediction purposes. A cut-off RPD value of 3 is recommended by researchers to ensure that a model is robust. The calibration models with the highest r^2 , the lowest

RMSECV and RMSEP were considered to be the optimal models in this study.

RESULTS AND DISCUSSION

Spectra and sample selection: The Raman spectra of the ethanol-glucose-water liquid mixtures and the wine samples were measured. Figure 3 consists of 50 spectra of wine samples with ethanol and glucose concentrations ranging from 0.04-0.40 v/v and 10-85 g/L, respectively. CrO_3 was used as the standard and its Raman spectrum was measured synchronously with that of the samples. The spectra for the ethanol-glucose mixture and the wine samples have very similar shapes and all show intense absorption bands at approximately 2974, 2929 and 2885 cm^{-1} , respectively which are mainly related to the C-H stretching modes of ethanol (Numata *et al.*, 2011). In the normal Raman spectrum of an aqueous saturated glucose solution, the peaks at 1462, 1365, 1268, 1126, 915 and 850 cm^{-1} , respectively correspond to peaks for crystalline glucose (Lyandres *et al.*, 2005). The peak at 3200 cm^{-1} corresponds to that of water.

First, the samples were divided into two sets, a calibration set and a prediction set, in a proportion of 4 to 1. The performances of the different models were compared using the same calibration set and prediction set that were used. HPLC was used to obtain reference data for ethanol and glucose. The calibration ranges were larger than those used for the prediction sets for both ethanol and glucose, thereby enabling stable and robust calibration models to be developed for the two components.

Multivariate calibration models: The most accurate and precise calibrated models were obtained by evaluating the prediction capacity of several PLS models. First, the calibration model was developed using the spectral range:

- 2500-3500
- 700-1500 cm^{-1} , which is according to the PCA

Thus, the time taken to collect the spectra and perform the analysis was considerably shorter. The performances of the differential regression models were compared for ethanol and glucose, as shown in Table 2. The prediction capacity of the PLS models was compared using the r^2 , RMSECV, RMSEP and RPD (%) values. The lower RMSEP and RMSECV values

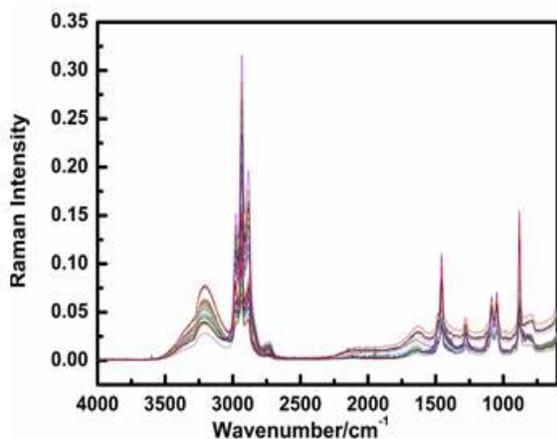
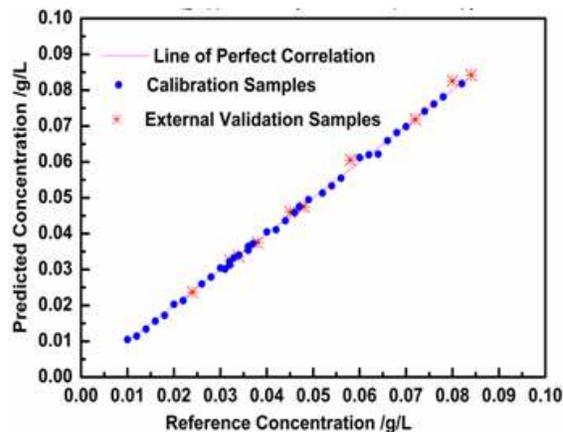


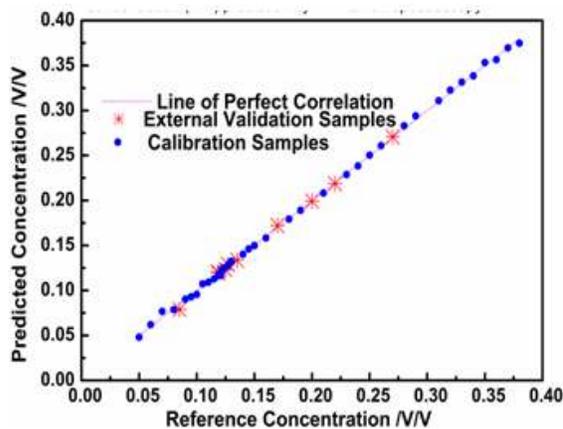
Fig. 3: Original Raman spectra of 40 ethanol-glucose liquid and 10 wine samples

show a better predictive ability and the higher r^2 and RPD values demonstrate the strong robustness and universality of the calibration. The optimal number of latent variables was decided by F-test. The selected spectra and the pre-processing method that are shown in bold were used for the quantitative measurements in Table 2. The prediction performance of the baseline correction and DOSC model was better than the other models for ethanol and glucose. Compared with the best PLS regression models, the highest-performing PLS model for ethanol had a RPD value of 6.3427, for glucose had a RPD value of 19.0575, this model produced accurate predictions and was considered to have a good predictive ability. Thus, the ethanol and glucose concentrations could be accurately predicted using PLS regression.

The calibration samples were previously analyzed by HPLC and the resulting concentrations were used as reference values. The concentrations obtained from Raman spectroscopy were used as the predicted values. The predicted and reference values are compared in Fig. 4. The calibration samples are marked by the * symbols, whereas the standard samples are marked by



(a)



(b)

Fig. 4: Calibration and validation of PLS model for ethanol concentration (V/V); (a): glucose concentration (g/L); (b): prediction by FT-Raman spectroscopy

round dots. The validation results exhibited excellent linearity, indicating that the model was linear over a wide range of ethanol or glucose concentrations. Thus, combining FT-Raman spectroscopy with a baseline correction and the PLS method is a powerful means of

Table 2: Prediction models obtained with different data treatment

| Component | Spectral pre-processing | RMSECV (%) | RMSEP (%) | r^2 | RPD |
|---------------|---------------------------|------------|--------------|---------|---------|
| Alcohol (V/V) | None | 0.003900 | 0.006000 (3) | 0.99780 | 4.4900 |
| | Baseline | 0.003700 | 0.004800 (3) | 0.99700 | 5.6500 |
| | Baseline+SNV | 0.003900 | 0.005400 (3) | 0.99600 | 5.0100 |
| | Baseline+MSC | 0.004500 | 0.004200 (4) | 0.99500 | 6.3800 |
| | Baseline+DOSC | 0.003614 | 0.004282 (3) | 0.99827 | 6.3427 |
| | Baseline+SG-21points | 0.003200 | 0.005240 (3) | 0.99760 | 5.1833 |
| | Baseline+first derivative | 0.003543 | 0.002961 (3) | 0.99786 | 9.1716 |
| Sugars (g/L) | None | 0.004100 | 0.002730 (5) | 0.99787 | 21.3000 |
| | Baseline | 0.004400 | 0.004200 (4) | 0.99680 | 14.0100 |
| | Baseline+SNV | 0.003600 | 0.005700 (5) | 0.99674 | 10.1640 |
| | Baseline+MSC | 0.003800 | 0.003300 (6) | 0.99780 | 17.1510 |
| | Baseline+DOSC | 0.003270 | 0.003052 (5) | 0.99861 | 19.0575 |
| | Baseline+SG-21points | 0.004800 | 0.002870 (4) | 0.99816 | 20.2759 |
| | Baseline+first derivative | 0.004770 | 0.002700 (4) | 0.99713 | 21.5359 |

The values in parentheses refer to the number of latent variables

determining the concentrations of ethanol or glucose in wine. The baseline correction can be combined with PLS to quantify the concentrations of ethanol and glucose in the fermentation process obtained using Raman spectroscopy.

CONCLUSION

In this study, FT-Raman spectroscopic methods were used to efficiently measure ethanol and glucose concentrations in fermentation. This method not only minimizes the amount of laboratory equipment required, eliminates sample preparation and reduces collection and processing time but also eliminates the need for chromatographic separation and the measurement errors resulting from a large number of experimental steps. Raman spectra obtained from different samples were normalized using the Raman scattering intensity of chromium oxide, which was used as an external standard. The models developed for the quantitative analysis have shown that the use of Raman spectroscopy, together with multivariate calibration regression based on PLS, allows the quantification of ethanol and glucose concentration in wine fermentation. The results clearly show that Raman spectroscopy can be used to identify the ethanol and glucose concentration in wine fermentation.

ACKNOWLEDGEMENT

This study was partly supported by the National Natural Science Foundation of China (11404054, 31170956), the Natural Science Foundation of Hebei Province, China (F2014501137), the Fundamental Research Funds for the Central Universities (N120323004), the Fundamental Research Funds for the Central Universities (N120223001), Science and technology research and development project funds of Shenzhen (JCYJ20120618142137681) and the Educational Commission of Hebei Province (Z2013040).

REFERENCES

- Boyacı, I.H., H.E. Genis, B. Guven, U. Tamer and N. Alper, 2012. A novel method for quantification of ethanol and methanol in distilled alcoholic beverages using Raman spectroscopy. *J. Raman Spectrosc.*, 43: 1171-1176.
- Diakabana, P., R.F. Niama, A.C. Mvila, I. Louzolo, S.C. Kobawila and D. Louembé, 2014. Behavior of fermentable sugars in the traditional production process of cassava bioethanol. *Adv. J. Food Sci. Technol.*, 6(10): 1104-1111.
- Di Egidio, V., N. Sinelli, G. Giovannelli, A. Moles and E. Casiraghi, 2010. NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation. *Eur. Food Res. Technol.*, 230: 947-955.
- Favors, R.N., Y. Jiang, Y.L. Loethen and D. Ben-Amotz, 2005. External Raman standard for absolute intensity and concentration measurements. *Rev. Sci. Instrum.*, 76: 033108-033108-033105.
- Gallego, Á.L., A.R. Guesalaga, E. Bordeu and A.S. González, 2011. Rapid measurement of Phenolics compounds in red wine using Raman spectroscopy. *IEEE T. Instrum. Meas.*, 60(2): 507-512.
- Kennard, R.W. and L.A. Stone, 1969. Computer aided design of experiments. *Technometrics*, 11: 137-148.
- Lyandres, O., N.C. Shah, C.R. Yonzon, J.T. Walsh, M.R. Glucksberg and R.P. Van Duynne, 2005. Real-time glucose sensing by surface-enhanced Raman spectroscopy in bovine plasma facilitated by a mixed decanethiol/mercaptohexanol partition layer. *Anal. Chem.*, 77: 6134-6139.
- Nah, S., D. Kim, H. Chung, S.H. Han and M.Y. Yoon, 2007. A new quantitative Raman measurement scheme using Teflon as a novel intensity correction standard as well as the sample container. *J. Raman Spectrosc.*, 38: 475-482.
- Niu, X., Z. Zhao, K. Jia and X. Li, 2012. A feasibility study on quantitative analysis of glucose and fructose in lotus root powder by FT-NIR spectroscopy and chemometrics. *Food Chem.*, 133: 592-597.
- Numata, Y. and H. Tanaka, 2011. Quantitative analysis of quercetin using Raman spectroscopy. *Food Chem.*, 126: 751-755.
- Numata, Y., Y. Iida and H. Tanaka, 2011. Quantitative analysis of alcohol-water binary solutions using Raman spectroscopy. *J. Quant. Spectrosc. Ra.*, 112: 1043-1049.
- Omar, J., A. Sarmiento, M. Olivares, I. Alonso and N. Etxebarria, 2012. Quantitative analysis of essential oils from rosemary in virgin olive oil using Raman spectroscopy and chemometrics. *J. Raman Spectrosc.*, 43: 1151-1156.
- Ozturk, B., D. Yucesoy and B. Ozen, 2012. Application of mid-infrared spectroscopy for the measurement of several quality parameters of alcoholic beverages, Wine and Raki. *Food Anal. Method.*, 5: 1435-1442.