

APPLICATION OF CHEMOMETRIC METHODS COUPLED WITH INFRARED SPECTROSCOPY FOR DETERMINATION OF ETHANOL IN PRESENCE OF METHANOL IN AQUEOUS SOLUTIONS

Petar Petrov, PhD

Pro M Consulting,

Silistra, Bulgaria, www.pro-m-consulting.com

E-mail: consulting.pro.m@gmail.com

Abstract: A fast and cost-effective method, based on Fourier Transformed mid-Infrared Spectroscopy measurement coupled with chemometric method was developed for determination of ethanol in aqueous solutions at room temperature and presence of methanol. In order to facilitate the calibration process, and to minimize the effects of peaks' overlapping, an appropriate preprocessing of the IR signals was performed. The calibration was performed using 43 samples, with accuracy enough high, so that this method can be used in routine determinations of ethanol. The model was also tested on a separate set of other 9 validation samples, where it showed great performance with R^2 of prediction >0.99 . Consequently, the procedure can be used as a fast and reliable determination of ethanol in aqueous solutions containing methanol as a second alcohol. The method can therefore find application in different areas of the chemical and food industry, avoiding use of chemicals, consumables or expensive equipment as HPLC. In fact, this approach could be used in other applications where distinguishable signals in FTIR exist, as for example quality control in production of fuels, essential oils, biodiesel, drugs etc.

Keywords: Partial least squares, chemometric, FTIR, calibration, ethanol, methanol

INTRODUCTION

Liquid aqueous mixtures of ethanol and methanol are frequently met in the industrial practices: distillation processes, production of fuel ethanol from bioethanol and its dilution with the cheaper methanol, fermentation processes in food and beverage industries, production of technical and household cleaners, automotive liquids. The determination of the ethanol in such liquids is important from the quality control point of view. The ethanol quantification in such mixtures is traditionally performed by gas chromatography, near infrared spectroscopy coupled with partial least squares (PLS) regression (Meng, Q.F., 2005), electrochemical methods with enzyme electrodes (Belghith H., 1987).

Today's laboratory instruments produce a huge amount of data. It is not uncommon for processes to have hundreds or thousands of measured variables. Some analytical instruments measure thousands of variables. For instance, in a typical FTIR spectrometer, the absorbance is measured at over 2000 - 3000 frequencies. Chemical processes are becoming more heavily instrumented and the data are recorded more frequently. Furthermore, in the presence of noise signals, it would be desirable to take advantage of some sort of signal averaging or preprocessing (Wise B, 2006). Not surprisingly, in the last two decades the chemometric methods continued to spread rapidly, providing efficient tools for analyzing large amounts of data, produced by different chemical analytical methods. In addition, the computers became very fast and inexpensive, giving us the possibility to process the experimental data quickly and accurately. Nowadays the use of Fourier transformed (FT) near infra red (NIR) or mid infra red (IR) spectroscopy coupled with some regression model, as for example Artificial Neural Networks (Agatonovic-Kustrin, 2013), single or multiple linear regression, partial least squares regression (Sim S.F., 2018) becomes more and more accepted and applied for online and fast analysis in chemical and food industry. Based on previous experience, and having in mind the possibilities and limitations of the PLS method - described later, this work deals with application of mid-IR FT spectroscopy combined with PLS regression. In contrast with the most used analytical methods for determination of ethanol, this approach eliminates the use of solvents or pretreatment of the sample, and is fast, relatively inexpensive and simple analysis.

Nowadays, the pre-processing methods of experimental data, as well the theories of PLS regression, are well developed and tested in many different areas as FTIR, NIR, NMR, UV-VIS, and even non-chemical systems (Faber N.M., 2003; Wold S., 2001; Rinnan A., 2009). This fact, as well the general trend to implement more IT based knowledge is the reason to focus on this technique in this work. In general, PLS regression is recommended and used in various disciplines where predictive linear modeling, especially with a large number of predictors, is necessary. There are some examples in the literature regarding use of FTIR or NIR and chemometrical methods for: determination of ethanol in water, mainly in alcohol beverages (Debebe A., 2017; Anjos O., 2016; Ramasami P., 2005); ethanol and methanol in fuels (Fernandes H. L., 2008); ethanol in gel hand sanitizers (Fonseca F. S. Jr., 2020); ethanol in presence of methanol in aqueous solutions (Meng Q.F., 2005; Coldea T. E., 2013). In this article we aim to develop an efficient and cost-effective method for fast determination of ethanol in presence of methanol based on a simple FTIR spectrum measurement. This could find application in routine laboratory analysis of ethanol containing products such as gel sanitizers, antifreezes, cleaners, alcohol beverages, fermentation processes. The approach could be used in other fields of laboratory practice where fast determination of the target substance is needed, with the prerequisite that it has enough strong and clearly distinguishable signal in FTIR or NIR spectra.

EXPOSITION

Reagents, apparatus and raw FTIR spectra

All IR spectra of the ethanol samples were recorded with FTIR-ATR PerkinElmer FT-IR Spectrometer Frontier. The use of attenuated total reflectance (ATR) technique has some advantages – small sample volume, speed of the analysis of no more than one minute, data reproducibility and very simple instrument operation. With this technique, the IR beam first passes through an internal reflection element (crystal) and after this goes into the sample, with a penetration depth of several micrometers. This gives more reproducibility in comparison to the transmission case, and minimizes effects related to air bubbles and other non-homogeneities. The spectra were taken in the region $3900 - 580 \text{ cm}^{-1}$, with resolution of 4 cm^{-1} , and accumulation of 4 scans, although for the data regression analysis only a part of it was selected. The crystal of the ATR unit is diamond/ZnSe, and all spectra were with suppressed $\text{CO}_2 / \text{H}_2\text{O}$ absorption. Using a small vertical plastic tube mounted above the ATR diamond it was ensured that all its surface is covered with liquid sample. The background spectra were collected automatically each time the device asked, normally after 6-8 measurements. The collection of the spectral data was performed using the PerkinElmer Spectrum software, version 10.03, and exported in appropriate format for further analysis. The chemicals used were methanol and ethanol Sigma Aldrich, assay 99.8%; and water purified by reverse osmosis. The solutions for calibration and validation of the model were prepared on a scale with 0.0001g resolution, adding all three components in a small glass vessel. We prepared 43 solutions for calibration of the model, with concentrations of methanol from 2 to 53 weight % and ethanol from 5 to 81w.%, which corresponds to a typical composition of liquids targeted in this study. The validation was done using 9 separate samples, within same concentration range. The both calibration and validation sets are shown with different colors on Fig.1, where the rest of the mixture to 100w.% is water.

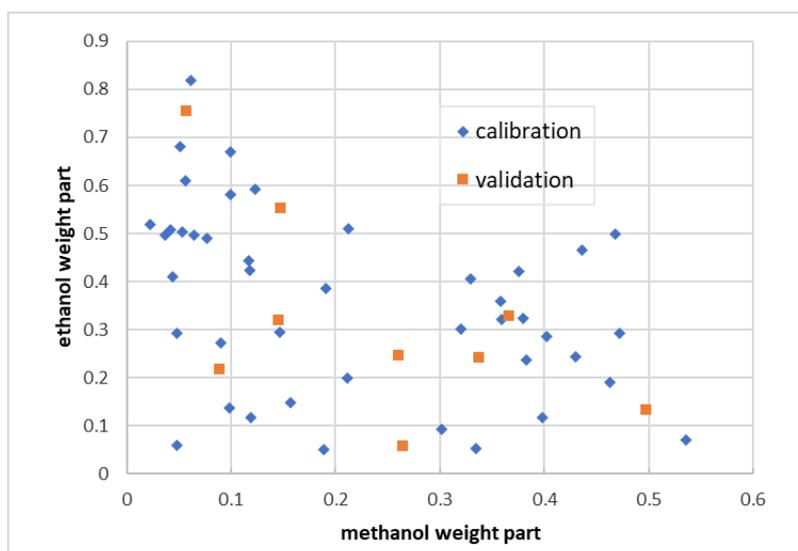


Fig.1 Concentrations of ethanol and methanol solutions used for calibration and validation of the model.

The raw FTIR spectra were exported from the spectromometer's software in appropriate form, so they can be utilized further with pre-processing algorithms and PLS regression model. When looking on the raw spectra (Fig.2) we observe the typical bands of the alcohols and water: an extensive band at 3100 – 3500 cm⁻¹ belonging to the OH group, but we focus mainly on the region 1136 – 852 where the more characteristic vibrations of methanol and ethanol are observed (Coldea T. E., 2013; Debebe A.,2017). The more sensitive bands are 1044 and 1087 cm⁻¹ for ethanol, 1015 – 1022 and 1114 cm⁻¹ for the methanol.

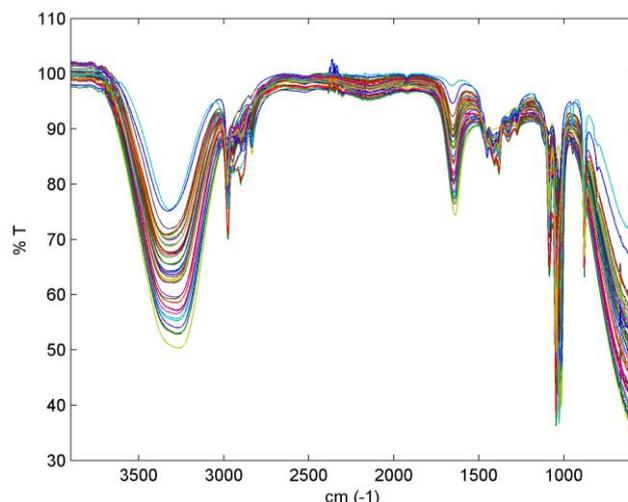


Fig.2 FTIR Spectra of ethanol / methanol / water solutions at 20deg C.

Pre-processing, calibration and validation of the regressin model

The chemometric analysis was done using the stand-alone version of software Solo, Eigenvector Research, Inc., v. 8.7, which allows direct import of the spectra data files produced by the PerkinElmer Spectrum software. After the loading of the experimental data, they were preprocessed and submitted to the model builder with a various of options. The raw spectral data of 43 mixtures, consisting of the X block (43 x 285, raw spectra) and the Y block (43 x 1, ethanol concentrations) were preprocessed in order to reduce the baseline variation, systematic noise, and to enhance the contribution of the desired spectral peaks. It is not purpose of this study to go deeper in the theory and application of pre-processing methods, as they are tested in many applications and well documented elsewhere (Rinnan Å.,2009). In this study we found that best results for the

calibration model, in term of prediction error, are obtained using 2-nd derivative of the spectra followed by mean centering which removes the mean offset of each variable. The concentrations block Y were mean centered and each variable was scaled to unit standard deviation, so that for both methanol and ethanol concentrations we have values within the range of appr. -1.5 to 2.5. In order to perform the internal validation of the calibration model, we performed a cross validation type “venetian blinds” with 10 data splits, which results in average 10% of the data left out, namely 4 or 5 samples. The root-mean-square error of calibration (RMSEC) tells us about the fit of the model to the calibration data, and together with the root-mean-square-error of cross validation (RMSECV), is used as criterion to select the number of the latent variables in PLS regression model. RMSEC, a measure of how well the model fits the data is defined as:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_i^* - y_i)^2}{n}}. \quad (1)$$

Later we present the RMSEP, the **prediction** error, defined in the same way as in (1), which is based on the application of the model on new samples, not used in the calibration process.

In fact, PLS regression constructs set of latent variables – vectors, that performs simultaneous decomposition of the both X (spectra) and Y (ethanol concentration) matrices, with the additional requirement that these variables explain as much as possible the covariance between the spectra and the ethanol’s concentrations. Although the liquid phase FTIR spectra of methanol and ethanol in water are very similar, and the presence of water and its broad peaks in the range 2000 – 800 cm^{-1} makes the ethanol’s quantitative determination very difficult, we succeeded to find a PLS regression model which gives accurate prediction within broad concentration range.

First, we present the RMSEC and RMSECV as a function of the latent variables used to calibrate the model (Fig.3), and following the rule to have both minimal number of latent variables and smallest error, we concluded that 6 latent variables are enough to describe the data.

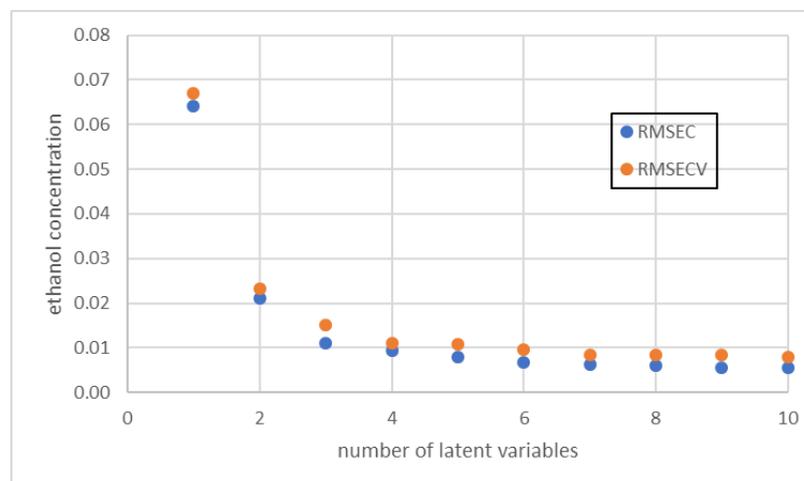


Fig. 3 Root mean square error of calibration and cross validation for ethanol concentration in absolute units – weight parts against number of latent variables.

For this model we have statistical data as follows: R^2 (cal.) = 0.998; RMSEC = 0.0067, average relative error for calibration calculated as $RMSEC * 100 / \sum(y_{measured}/n) = 1.87\%$. In order to check it on a separate set of 9 validation FTIR spectra, taken within the same concentration range of ethanol in presence of methanol. The performance is excellent, as we have R^2 (val.) = 0.999, RMSEP = 0.006 and average error of prediction, calculated as above for the calibration, and giving us the realistic estimate of the performance 1.86%. Fig.4 shows the plot calculated through calibration and validation ethanol concentrations (weight parts) against the experimental values. Obviously, they are very close to the 45deg line, meaning no significant deviations exist.

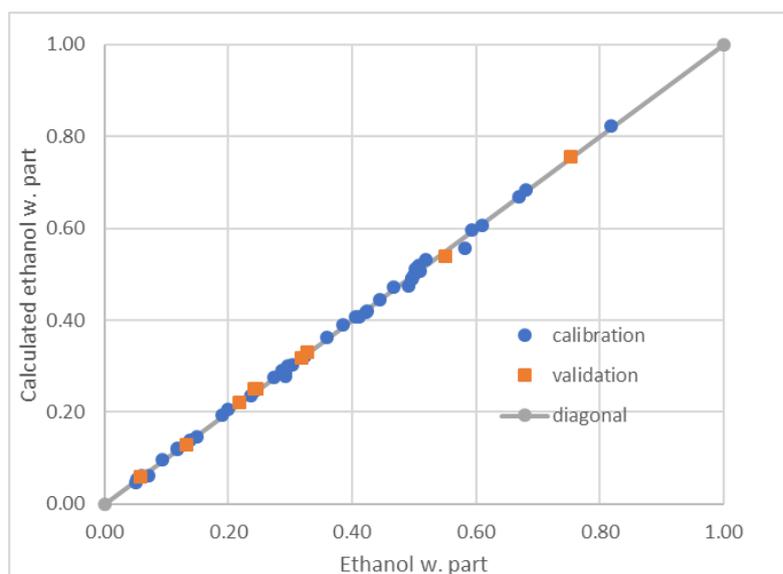


Fig. 4 Comparison of calculated and experimental values of ethanol concentration for both calibration procedure and validation of the model

CONCLUSION

The FTIR spectroscopy in ATR execution combined with partial least squares regression is demonstrated to predict with acceptable accuracy ethanol concentrations in aqueous mixtures containing methanol. The problem arising from the similar and partially overlapped alcohols' spectra, as well as the water influence, is easily resolved by performing appropriate preprocessing, namely second derivative per Savitzky-Golay algorithm, followed by mean centering of the spectra. The training data – X matrix consists of 43 samples. After applying the PLS regression algorithm with internal cross validation, we built a regression model with 6 latent variables for prediction of ethanol content based on FTIR spectra with values of $R^2 > 0.99$. The model was checked on a separate data set of 9 spectra of samples with different compositions, not used during calibration, and again showed $R^2 > 0.99$ and relative average error of only 1.86%. Consequently, this approach could find broad application in practice, as fast and precise enough to determine ethanol content in different liquids and sanitizing gels.

REFERENCES

- Agatonovic-Kustrin S., Morton D.W., Yusof A.P.M., The Use of Fourier Transform Infrared (FTIR) Spectroscopy and Artificial Neural Networks (ANNs) to Assess Wine Quality, *Mod Chem appl* 1, 2013, 110
- Anjos O., Santos A. J.A., Estevinho L. M., Caldeira I., FTIR–ATR spectroscopy applied to quality control of grape-derived spirits, *Food Chemistry*, 205, 2016, 28-35
- Belghith H., Romette J.L., Thomas D., An enzyme electrode for online determination of ethanol and methanol, *Biotechnol Bioeng*, 30, 1987, 1001-1005
- Coldea T. E., Socaciu C., Fetea F., Ranga F., Pop R. M., Florea M., Rapid Quantitative Analysis of Ethanol and Prediction of Methanol Content in Traditional Fruit Brandies from Romania, using FTIR Spectroscopy and Chemometrics, *Not Bot Horti Agrobo*, 41, 1, 2013, 143-149
- Debebe A., Redi-Abshiro M., Chandravanshi B. S., Non-destructive determination of ethanol levels in fermented alcoholic beverages using Fourier transform mid-infrared spectroscopy, *Chemistry Central Journal* 11:27, 2017, DOI 10.1186/s13065-017-0257-5
- Debebe A., Temesgen S., Redi-Abshiro M., Chandravanshi B. S., Partial least squares–near infrared spectrometric determination of ethanol in distilled alcoholic beverages, *Bull. Chem. Soc. Ethiop.* 31(2), 2017, 201-209

Faber N.M., Song X.-H., Hopke P.K., Sample-specific standard error of prediction for partial least squares regression, *Trends in Analytical Chemistry*, 22, 5, 2003, 330

Fernandes H. L., Raimundo I. M., Pasquini C., Rohwedder J. J.R., Simultaneous determination of methanol and ethanol in gasoline using NIR spectroscopy: Effect of gasoline composition, *Talanta* 75, 2008, 804–810

Fonseca F. S. Jr., Brito L. R., Pimentel M. F., Leal L. B., Determination of Ethanol in Gel Hand Sanitizers Using Mid and Near Infrared Spectroscopy, *J. Braz. Chem. Soc.*, 2020, 31, No. 9, 1759-1763

Klein O., Roth A., Dornuf F., Scholler O, and Mantele W., The Good Vibrations of Beer. The Use of Infrared and UV/Vis Spectroscopy and Chemometry for the Quantitative Analysis of Beverages, *Z. Naturforsch.* 2012, 67b, 1005 – 1015

Meng Q.F., Teng L.R., Lu J.H., Jiang C.J., Gao C.H., Du T.B., Wu C.G., Guo X.C., and Liang Y.C., Determination of Methanol and Ethanol Synchronously in Ternary Mixture by NIRS and PLS Regression in O. Gervasi et al. (Eds.), *Computational Science and Its Applications - ICCSA 2005*, 1040-1045

Ramasami P., Jhaumeer-Laulloo1 S., Cadet F., Rondeau P., Soophul Y., Quantification of alcohol in beverages by density and infrared spectroscopy methods, *International Journal of Food Sciences and Nutrition*, 2005, 56(3), 177-183

Rinnan Å., Nørgaard L., van den Berg F., J. Thygesen , Bro R., Engelsen S. B., *Data Pre-processing, in Infrared Spectroscopy for Food Quality Analysis and Control*, Elsevier Inc. 2009, ISBN: 978-0-12-374136-3

Rinnan A., van den Berg F., Engelsen S. B., Review of the most common pre-processing techniques for near-infrared spectra, *Trends in Analytical Chemistry*, 28, 10, 2009, 1201

Sim S.F., Chai M. X. L., Kimura A. L. J., Prediction of Lard in Palm Olein Oil Using Simple Linear Regression (SLR), Multiple Linear Regression (MLR), and Partial Least Squares Regression (PLSR) Based on Fourier-Transform Infrared (FTIR), Hindawi, *Journal of Chemistry*, 2018, Article ID 7182801

Wise B., Shaver J., Gallagher N., Windig W., Bro R., Scott R., *Chemometrics Tutorial for PLS_Toolbox and Solo*, Eigenvector Research Incorporated, 2006

Wold S., Sjostrom M., Eriksson L., PLS-regression: a basic tool of chemometric, *Chemometrics and Intelligent Laboratory Systems*, 58, 2001, 109–130