Talanta 81 (2010) 1413–1417

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Quantitative analysis of mixed hydrofluoric and nitric acids using Raman spectroscopy with partial least squares regression

Gumin Kang[a], Kwangchil Lee[a], Haesung Park[a], Jinho Lee[a], Youngjean Jung[a], Kyoungsik Kim[a,*], Boongho Son[b], Hyoungkuk Park[b]

[a] School of Mechanical Engineering, Yonsei University, 262 Seongsanno, Seodaemun-gu, Seoul 120-749, Republic of Korea
[b] POSCO Instrumentation Research Group,Technical Research Laboratory, Geodong-dong, Nam-gu, Pohang, Gyeongbuk 790-785, Republic of Korea

ARTICLE INFO

Article history:
Received 6 December 2009
Received in revised form 11 February 2010
Accepted 15 February 2010
Available online 24 February 2010

Keywords:
Raman spectroscopy
Hydrofluoric acid (HF)
Nitric acid (HNO_3)
Partial least squares regression (PLSR)
Net analyte signal (NAS)
Figures of merit (FOM)

ABSTRACT

Mixed hydrofluoric and nitric acids are widely used as a good etchant for the pickling process of stainless steels. The cost reduction and the procedure optimization in the manufacturing process can be facilitated by optically detecting the concentration of the mixed acids. In this work, we developed a novel method which allows us to obtain the concentrations of hydrofluoric acid (HF) and nitric acid (HNO_3) mixture samples with high accuracy. The experiments were carried out for the mixed acids which consist of the HF (0.5–3 wt%) and the HNO_3 (2–12 wt%) at room temperature. Fourier Transform Raman spectroscopy has been utilized to measure the concentration of the mixed acids HF and HNO_3, because the mixture sample has several strong Raman bands caused by the vibrational mode of each acid in this spectrum. The calibration of spectral data has been performed using the partial least squares regression method which is ideal for local range data treatment. Several figures of merit (FOM) were calculated using the concept of net analyte signal (NAS) to evaluate performance of our methodology.

1. Introduction

In the manufacturing process of stainless steel, oxide layer which is caused by heat treatment or welding shows inferior protective properties. The oxide layer is the source of the low corrosion resistance and surface quality. Thus, pickling process which is removal process based on chemical procedure is required. In general, mixed acid which is comprised of hydrofluoric acid (HF) and nitric acid (HNO_3) is widely used to pickle both iron contamination and high temperature oxide scales on the surface of the worked steel [1]. In order to reduce cost of the process and decrease the generation of corrosive waste acid, recycling or recovering process of pickling liquor is widely used. In the waste acid recycling system, acid supplementation process is crucial to maintain efficient pickling performance. Thus, fresh acids should be added to make up for HF and HNO_3 which were consumed during pickling process. For above reason, it is important to frequently measure and preserve the optimized individual acidic concentrations of the pickling liquid. In order to analyze the HF and HNO_3 concentrations quantitatively in the pickling liquid, Galvez et al. suggested a scheme involving several steps of titration [2]. Lindroos also proposed a method which combines direct potentiometry with an alkalimetric titration and a fluoride-selective electrode [3]. However, these methods present several disadvantages to be overcome: time-consuming sample preparation for the titration procedure, complicated measurement processes, and the problem of possible exposure to the toxic acid. On the other hand, great advantages can be gained from the development of an optical measurement system of each component in mixed acid in a non-invasive manner with high accuracy.

This optical measurement system can be realized by Raman spectroscopy [4] and near-infrared (NIR) spectroscopy [5]. Because water shows weak Raman scattering, it does not interfere with Raman signal in contrary to the IR absorption signal. Thus, Raman spectroscopy is a predominant technique for analysis of individual components in a aqueous solutions with minimum sample preparation. Furthermore, it is more powerful than NIR spectroscopy owing to the sensitivity of solvent bands to the presence of solute. However, as yet, no information exists on Raman spectroscopy for the detection of each component in mixed acids containing HF and HNO_3. Thus, we developed a novel method to obtain the concentrations of the HF and HNO_3 mixture samples with high degree of accuracy by using Raman spectroscopy. Because, above technique generates multivariate response for each analyzed sample, univariate calibration was not sufficient to extract chemical information from spectral data. Therefore, we coupled FT-Raman spectroscopy with the multivariate calibration techniques and partial least squares regression (PLSR) [6,7] to analyze the con-

* Corresponding author. Tel.: +82 2 2123 5815; fax: +82 2 312 2159.
E-mail address: kks@yonsei.ac.kr (K. Kim).

0039-9140/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2010.02.045

centration of each component in mixed acids. Net analyte signal concept was applied to our calibration model for estimation of figures of merit such as sensitivity, selectivity, signal to noise ratio, limit of detection and limit of quantification [8–10]. Repeatability is calculated for evaluation of precision of proposed method [11].

2. Experiments

2.1. Pickling solution

The experimental samples of Raman spectroscopy were designed by a dilution series of standard solutions. In this work, 36 samples were prepared by mixing appropriate amounts of HF, HNO_3 and H_2O. Pickling solution normally involves 12 wt% HNO_3 and 3 wt% HF mixtures which were optimized concentrations for pickling process. Employment of orthogonal design is crucial to minimize the degree of relationship between any concentrations of the mixtures and make exact calibration model. We employ l-level fractional factorial design to make mutually orthogonal design. The orthogonal design requires minimum l^2 experiments (36 experiments) for 6 concentration levels ($l = 6$) [12]. Concentration ranges of mixtures for HF and HNO_3 were 0.5–3 wt% and 2–12 wt%, respectively. Deionized (DI) water was used in this experiment to remove other ions (impurities) in H_2O because these impurities could affect the Raman spectra as noise signals.

2.2. Instrumentation

Raman spectra were obtained by a FT-Raman spectrometer (Bruker optics Inc., FRA106/s) using a Nd:YAG laser source (1064 nm) with 500 mW power. The spectrometer is equipped with a liquid nitrogen (N_2) cooled germanium (Ge) diode detector to obtain high sensitivity. A CaF_2 beam splitter is employed as an interferometer. During the experiments, the spectra were recorded over 100 scans in the 100–4500 cm^{-1} range and the spectral resolution was 4 cm^{-1}. Approximately 3 min was spent on obtaining each spectrum.

2.3. Data processing

For instrument control and data acquisition, the OPUS software (Bruker optics Inc.) was employed. Data treatment was performed using MATLAB 7.4.0 (The Mathworks, Inc.) with the PLS toolbox 5.2.1 (Eigenvector Research, Inc.). Several multivariate analysis methods such as PCA and PLSR algorithm were utilized to build a calibration model of the collection of Raman spectra.

2.4. Data analysis and interpretation

A chemometrics techniques such as the PCA and PLSR were employed to analyze the 36 spectra data through calibration modeling. We detect and remove the outlier using the Q-residual, leverage, Y Studentized residual values. After excluding 4 outliers, all 32 samples were split into two subsets—a calibration set and a prediction set (generally with a ratio of 2:1). Thus, 22 samples are in a group as a calibration set, which will be used to build a calibration model and the remaining 10 samples are grouped as a prediction set, which is used to evaluate the prediction performance of each model. In the calibration process, it is important to optimize the number of PLS factors. While the characteristics of the spectrum could not be accurately represented by too small a number of PLS factors, too large a number of PLS factors could draw measurement noises from the data into the calibration model. Because the number of measurements which consists of the data set were not big enough to group into a training and test set, a leave-one-out cross-validation (LOOCV) method was applied. This method is an iterative method in which each sample is used once for prediction and removed from the model formulation. The method can eliminate the dominant effect of a single outlier, perturb the data set and maintain statistical robustness.

In order to estimate performance of the calibration model, we employed the root mean squared error (RMSE) for the calibration (RMSEC) and for the cross-validation (RMSECV). To validate the prediction performance of the calibration model, we employed the root mean square error of prediction (RMSEP) for the prediction set.

2.5. Figures of merit (FOM)

Lorber defined net analyte signal (NAS) to be the part of a measured signal which is orthogonal to the interferents [8]. NAS calculation is a useful method for the evaluation of figures of merit (FOM). FOM in multivariate calibration represents the quality of a given analytical method. In this study, we calculated several FOM such as sensitivity (SEN), analytical sensitivity (γ), effective resolution (γ^{-1}), selectivity (SEL), signal to noise ratio (SNR), limit of detection (LOD) and limit of quantification (LOQ). Throughout this work, vectors and matrix are denoted by boldface letters. For our purpose, it suffices to regard all vectors as living in the three-dimensional Euclidean space $\mathbb{R}^3$. The notation $\|\mathbf{u}\|$ stands for $(\mathbf{u}\cdot\mathbf{u})^{1/2}$, the (Euclidean) length of $\mathbf{u}$ where summation convention applies to repeated indices.

NAS vector can be calculated as:

$$\mathbf{NAS}_i = (\mathbf{x}_i\mathbf{b})(\mathbf{b}^T\mathbf{b})^{-1}\mathbf{b}^T$$

where $\mathbf{x}_i$ is the row spectrum vector of the i th sample and $\mathbf{b}$ is the regression column vector from the calibration model [9,10].

Sensitivity (SEN) is defined to be:

$$\mathbf{SEN}_i = \frac{\mathbf{NAS}_i}{y_i}, \qquad SEN_i = \|\mathbf{SEN}_i\|$$

where y_i is the measured concentration for the i th sample. Sensitivity for specific components is the mean value of the univariate sensitivity values for all calibration samples [13]. However, it is impossible to compare sensitivities of multiple analytical method due to each sensitivity of its instrumental intensity. Calculation of analytical sensitivity (γ) and the inverse of its value (γ^{-1}) can be used to normalize sensitivities for comparison of analytical methods by estimating a minimum concentration difference regardless of the specific technique. The normalization is performed by the following equation:

$$\gamma = \frac{SEN_i}{\delta r}$$

where δr is the measure of instrumental noise which is the standard deviation in the predicted concentration for blank samples [10].

Selectivity is defined in order to inform how unique the spectrum of analyte is, compared with the other components. Thus, it represents the quantity of the signal which is not lost due to spectral overlap and is defined as:

$$SEL_i = \frac{\|\mathbf{NAS}_i\|}{\|\mathbf{x}_i\|}.$$

Finally, we calculated signal to noise ratio, limit of detection and limit of quantification which are defined as [13]:

$$SNR_i = \frac{NAS_i}{\delta r}, \qquad LOD = \frac{3\delta r}{SEN}, \qquad LOQ = 3.3LOD$$

where LOD is the minimum detectable concentration of analyte in the sample which can be distinguished from noise level. LOQ is the limit from which a reliable quantification is possible and is defined

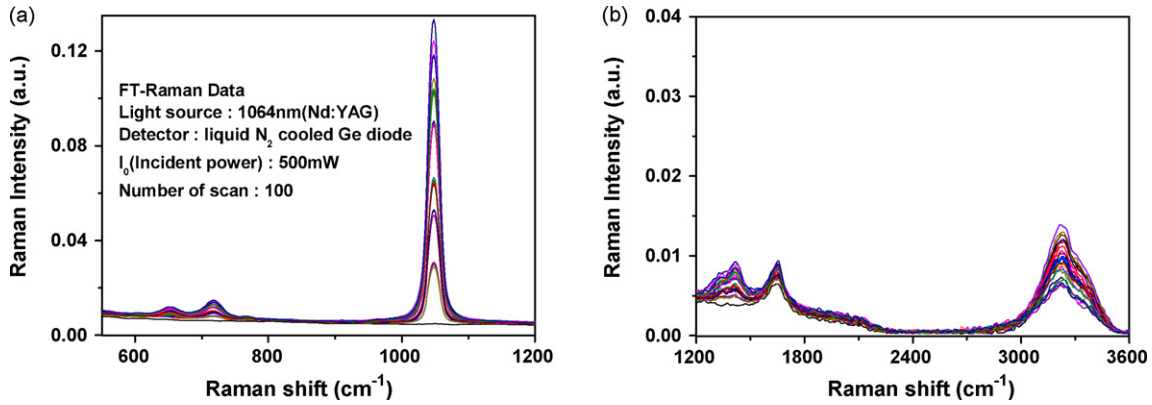

Fig. 1. FT-Raman spectra of DI water and mixed acids which have different mixture combinations of HF and HNO_3 in the following spectral regions: (a) 550–1200 cm^{-1} and (b) 1200–3600 cm^{-1}.

to be the minimum level as 3.3 times the LOD, or 10 times the standard deviation which is used to calculate LOD [14,15].

3. Results and discussion

3.1. FT-Raman spectra

Fig. 1 shows the FT-Raman spectra of 32 different acids which contain HF (between 0.5 wt% and 3 wt%) and HNO_3 (between 2 wt% and 12 wt%) compared with the spectrum of DI water in the range of 100–4500 cm^{-1}. Fig. 2 represents the Raman spectra of DI water (H_2O), 3 wt% HF and 3 wt% HNO_3. As shown in Fig. 2, H_2O has the Raman bands around 3215 cm^{-1} and 1651 cm^{-1}. These bands are related to the OH stretching mode and bending mode, respectively [16]. Raman spectra of 3 wt% HNO_3 show a distinctive spectral feature at 1048 cm^{-1} which corresponds to the symmetric stretch band of nitrate ion (NO_3^-). The Raman bands at 1413 cm^{-1} and 716 cm^{-1} correspond to the asymmetric stretch mode and the O–N–O bending mode of the nitrate ion, respectively [17].

Christy et al. measured the Raman spectra of solid phase $NaHF_2$ [18]. As the pressure increases, the Raman peak position of the hydrogen bifluoride ion (FHF^-) shifts from 633 cm^{-1} to 658 cm^{-1}. Thus, the Raman band of liquid phase HF at 655 cm^{-1} is associated with the symmetric stretch mode of hydrogen bifluoride ion. Because HNO_3 has several Raman active modes which do not overlap with the characteristic modes of HF, we can distinguish and analyze each component. The most intense and non-overlapping bands, 655 cm^{-1} for HF and 1048 cm^{-1} for HNO_3 are utilized for quantitative analysis. Fig. 3(a) represents the Raman spectra of the mixed acids where HF concentration varies from 0.5 wt% to 3.0 wt% with the fixed HNO_3 concentration. As the content of HF increases, the relative intensity of the symmetric stretch band corresponding to the hydrogen bifluoride ion (FHF^-) increases. Fig. 3(b) shows the Raman spectra of mixed acids where HNO_3 concentration varies from 2 wt% to 12 wt% while the HF concentration is fixed. As shown in Fig. 3(b), the increase in relative Raman intensity of the symmetric stretch band corresponding to the nitrate ion is observed with the increase of nitric acid. It is experimentally observed that the intensity of Raman spectra is linearly proportional to the concentrations of each component. This relation will be demonstrated through regression analysis in Section 3.2.

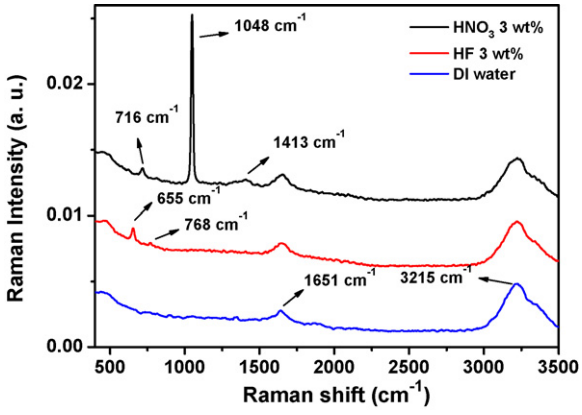

Fig. 2. FT-Raman spectra of DI water (H_2O), 3 wt% HF and 3 wt% HNO_3. The spectra have been offset on the *Y*-axis for clarity.

3.2. Analysis of results via PLS calibration model

Each acid of mixture samples shows comparatively distinctive spectral features as shown in Fig 2. Therefore, we can obtain quite robust calibration results with the use of univariate analysis. However, in case of HF, there exists some overlapping Raman bands caused by the O-N–O bending mode of the nitrate ion (716 cm^{-1}). Furthermore, when we need to monitor mixture samples with high concentration, the range of spectral overlap will broaden. Besides, there exists some impurities in the waste acid and they produce more interfering signals. Then, the univariate calibration is not sufficient to extract quantitative information from spectral data caused by interferences. Therefore, we performed multivariate analysis to build a less sensitive calibration model to the external variables and to apply this model into actual pickling process in the steel manufacturing line [19].

Prior to building a calibration model, we employed several preprocessing methods such as mean center, derivative and smoothing. The mean center method is used to remove constant background contributions, which are considered of little interest for data variance interpretation [20]. Hence, the data were preprocessed with the 15 point Savitzky–Golay smoothing filter to reduce the spectral noise [21] and the first derivative method was used to remove the effect of baseline fluctuation.

To evaluate the prediction performance of the calibration model, the root mean squares regression error of cross-validation (RMSECV) has been calculated for each acid. As the number of PLS factors increases, RMSEC always decreases, while RMSECV does not decrease in some cases. As a rule, the optimal number of PLS factors is determined by the first minimum value of RMSECV. If the increase of PLS factors does not reduce the RMSECV significantly, the former PLS factor is selected as the optimal calibration model. The optimal number of factors and the corresponding RMSE are summarized in Table 1. We obtained the RMSEP values of 0.22 wt% for HF and 0.29 wt% for HNO_3 and the correlation coefficients of

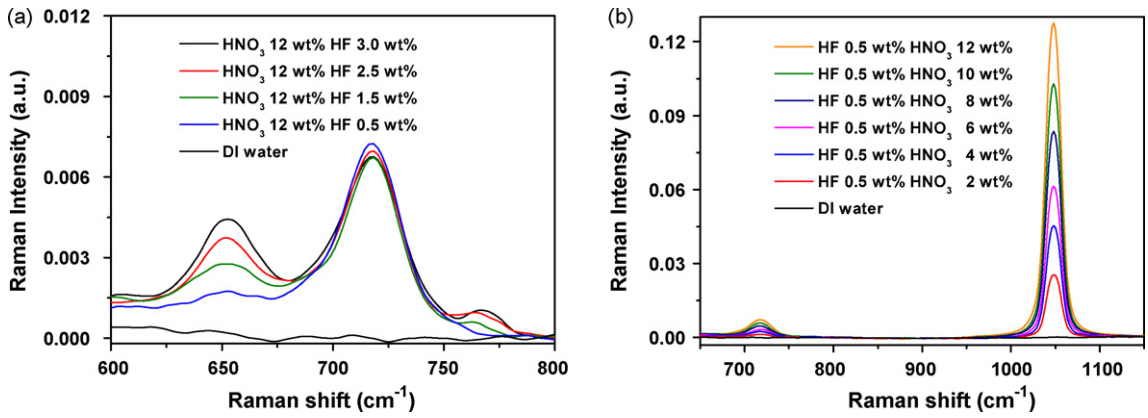

Fig. 3. FT-Raman spectra for (a) mixed acids of four different HF contents (0.5–3.0 wt%) with a fixed HNO_3 concentration (12 wt%) and (b) mixed acids of six different HNO_3 contents (2.0–12.0 wt%) with a fixed HF concentration (0.5 wt%).

Table 1
PLS calibration results and figures of merit (FOM) for HF and HNO_3 with Raman spectra.

	Calibration method		PLS	
	Preprocessing method		Mean center, first derivative smoothing (Savitzky–Golay)	
	Cross-validation (CV)		Leave-one-out CV (LOOCV)	
	Spectral region (cm^{-1})		4500–100	
	Components		HF	HNO_3
	# of PLS factors		3	3
Accuracy	RMSEC	wt%	0.15	0.41
	RMSECV	wt%	0.37	0.51
	RMSEP	wt%	0.22	0.29
	R^2 (calibration)	Unitless	0.97	0.98
	R^2(CV)	Unitless	0.89	0.97
	R^2 (prediction)	Unitless	0.96	0.99
FOM	Sensitivity (SEN)	Intensity/wt%	0.0002	0.0039
	Analytical sensitivity (γ)	1/wt%	18.79	387.09
	Effective resolution (γ^{-1})	wt%	0.053	0.003
	Selectivity (SEL)	Unitless	0.032	0.393
	Signal to noise ratio (SNR)	Unitless	17.87	314.98
	Limit of detection (LOD)	wt%	0.15	0.01
	Limit of quantification (LOQ)	wt%	0.49	0.03
Repeatability	HF 1.5 wt% HNO_3 6 wt%	coefficient of variation (% CV)	1.6	0.2
	HF 2.5 wt% HNO_3 10 wt%		1.2	0.4

0.96 for HF and 0.99 for HNO_3. Fig. 4 represents the predicted versus actual concentration plots in the optimized spectral range obtained by PLSR. The open and filled circles represent the set of calibration and prediction data, respectively. The predicted concentrations are in excellent agreement with actual concentrations. We also analyze three real samples which are used in pickling process by using Raman spectroscopy. And in order to validate our method, we employ the reference method by using the potentiometry with an alkalimetric titration and fluoride ion selective electrode. Fig. 5 shows the Raman spectra of real samples and Table 2 presents the results for three real samples analyzed by Raman spectroscopy and reference methods. As shown in Table 2, predicted results for real

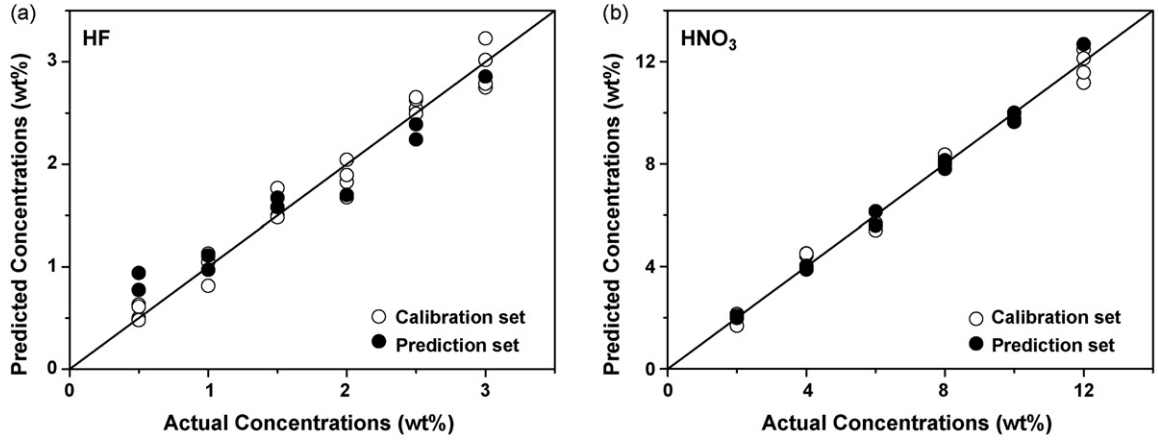

Fig. 4. Predicted versus true concentrations of (a) HF and (b) HNO_3. The open and filled circles represent the calibration and prediction data, respectively.

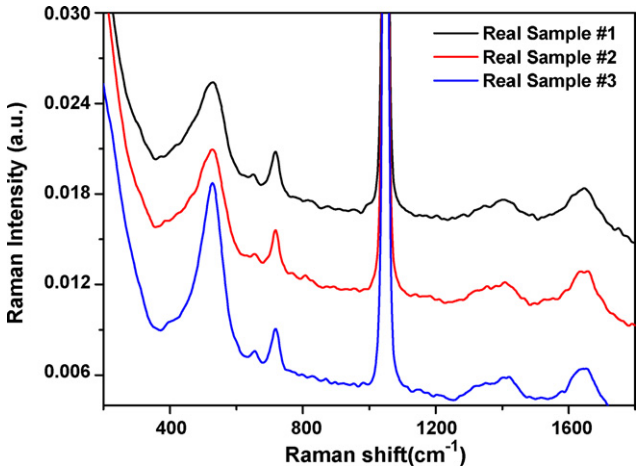

Fig. 5. FT-Raman spectra of three real samples. The spectra have been offset on the *Y*-axis for clarity.

Table 2
Results for three real samples analyzed by Raman spectroscopy and reference methods.

	HF		HNO_3	
	Raman [a]	Reference [a]	Raman [a]	Reference [a]
Real sample 1	0.53	0.45	7.08	6.90
Real sample 2	0.50	0.36	7.07	7.10
Real sample 3	1.01	1.02	7.44	7.70

[a] Results in weight percent (wt%).

samples are quite accurate. Therefore, these results confirm that the proposed method enables us to make a quantitative analysis of individual components in a mixture sample with high robustness by optical measurements.

3.3. Figures of merit (FOM)

We calculated several FOM such as sensitivity (SEN), analytical sensitivity (γ), effective resolution (γ^{-1}), selectivity (SEL), signal to noise ratio (SNR), limit of detection (LOD) and limit of quantification (LOQ) as described in Section 2.5. Results for figures of merit based on optimal PLS factors are shown in Table 1. Sensitivity for HF is much smaller compared to those of HNO_3. By calculating analytical sensitivity and effective resolution values, we can compare the sensitivity of our method with different analytical method. Selectivity value is a measure of quantity which is not overlapped with other interferences and the value varies between 0 and 1. As shown in Table 1, HNO_3 shows superior selectivity over HF. This means that the Raman signal of HNO_3 shows more distinctive spectral features compared to HF as mentioned in Section 3.2. The repeatability was evaluated by using the coefficient of variation (CV%), through the measurement of 6 replicates of the same sample [11]. We also calculated signal to noise ratio, limit of detection and limit of quantification. These repeatability and FOM results support that the calibration quality of the present work is sufficient to analyze pickling liquid.

4. Conclusion

This research has shown that FT-Raman spectroscopy can be applied to measure the concentrations of mixed acids consisting of HF and HNO_3. In contrast to chemical analysis methods such as titration or the ion selective electrode scheme, the current study shows the potential for predicting concentrations of mixed acid in a non-contact and non-destructive manner. Therefore, the method can be used to control and optimize the pickling process for obtaining high surface quality and superior corrosion resistance. PLSR was employed to design a calibration model for the determination of each acid concentrations in pickling liquid. Through employing the PLSR model with mean center, first derivative and smoothing, we obtained the good prediction results. We demonstrated that FT-Raman spectroscopy coupled with the PLSR algorithm can be used to analyze concentrations of pickling liquid in an on-line manner. Finally, we calculated several FOM to diagnose the limits of our method and those results showed the present method has sufficient performance in analyzing pickling liquid.

Acknowledgements

We thank POSCO for the loan of equipment. This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Nos. 2009-0077097 and 2009-0083512), the Low Observable Technology Research Center program of Defense Acquisition Program Administration and Agency for Defence Development and the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2008-313-C00353).

References

[1] L. Narvaez, E. Cano, J.M. Bastidas, Corrosion 61 (2005) 21–29.
[2] J.L. Galvez, J. Dufour, C. Negro, F. Lopez, ISIJ Int. 46 (2006) 281–286.
[3] K. Lindroos, Analyst 112 (1987) 71–73.
[4] M.G. Orkoula, C.G. Kontoyannis, C.K. Markopoulou, J.E. Koundourellis, Talanta 73 (2007) 258–261.
[5] C.J. Thompson, J.D. Danielson, J.B. Callis, Anal. Chem. 69 (1997) 25–35.
[6] J.J. Kim, J.Y. Hwang, H.I. Chung, Anal. Chim. Acta 629 (2008) 119–127.
[7] J.J. Kim, J.T. Han, J.G. Noh, H.I. Chung, Appl. Spectrosc. 61 (2007) 686–693.
[8] A. Lorber, Anal. Chem. 58 (1986) 1167–1172.
[9] R. Bro, C.M. Andersen, J. Chemometr. 17 (2003) 646–652.
[10] S.M. Short, R.P. Cogdill, C.A. Anderson, AAPS PharmSciTech 8 (2007) E1–E11 (Article 96).
[11] J.W.B. Braga, R.J. Poppi, J. Pharm. Sci. 93 (2004) 2124–2134.
[12] R.G. Brereton, Analyst 125 (2000) 2125–2154.
[13] A. Lorber, K. Faber, B.R. Kowalski, Anal. Chem. 69 (1997) 1620–1626.
[14] M. Zorn, R. Gibbons, W. Sonzogni, Environ. Sci. Technol. 33 (1999) 2291–2295.
[15] M.A. Cantarelli, I.G. Funesc, E.J. Marchevskyb, J.M. Camina, Talanta 80 (2009) 489–492.
[16] D.M. Carey, G.M. Korenowski, J. Chem. Phys. 108 (1998) 2669–2675.
[17] N. Minogue, E. Riordan, J.R. Sodeau, J. Phys. Chem. A 107 (2003) 4436–4444.
[18] A.G. Christy, J. Haines, S.M. Clark, J. Phys.: Condens. Matter 4 (1992) 8131–8140.
[19] C.N.C. Corgozinho, V.M.D. Pasa, P.J.S. Barbeira, Talanta 76 (2008) 479–484.
[20] E. Pere-Trepat, M. Petrovic, D. Barcelo, R. Tauler, Anal. Bioanal. Chem. 378 (2004) 642–654.
[21] A. Savitzky, M.J.E. Golay, Anal Chem. 36 (1964) 1627–1639.