



Bromine number prediction for Colombian naphthas using near-infrared spectroscopy and chemometric methods

Predicción del número de bromo de naftas colombianas usando espectroscopía infrarrojo cercano y métodos quimiométricos

Previsão do número de bromo de naftas colombianas usando espectroscopia NIR e métodos quimiométricos

Carlos Aparicio¹, Jader Guerrero², Rafael Cabanzo¹, Enrique Mejía-Ospino¹

¹Laboratorio de Espectroscopía Atómica y Molecular – LEAM, Centro de Materiales y Nanociencias

²Grupo de Óptica y Tratamiento de Señales. Universidad Industrial de Santander, A.A. 678; Parque Tecnológico Guatiguará, Km 2; Piedecuesta, Santander, Colombia.

*emejia@uis.edu.co

Fecha Recepción: 23 de junio de 2015

Fecha Aceptación: 11 de noviembre de 2016

Abstract

Thirty-eight naphtha samples were used to develop a chemometric method to predict bromine number. All samples were characterized by Fourier transform near infrared spectroscopy (FT-NIR), and their spectra were correlated by similarity using principal component analysis (PCA). The models for bromine number determination (BN) were established by Partial Least Squares regression (PLS) and Multiple Polynomial Regression (MPR). PCA allowed classifying the samples into the light and heavy, determining the most significant spectral variables. These variables are located in the regions between 4000-4800 and 5200-6350 cm^{-1} . The results determined by combining FT-NIR spectroscopy and chemometrics were very close to those obtained by standardized methods. This approach may be an alternative for analysis of BN, which requires sample turnarounds lower than five minutes and lower cost compared to traditional methods.

Keywords: FT-NIR, naphtha, bromine number, PCA, PLS, MPR.

Resumen

En este trabajo se emplearon treinta y ocho muestras de nafta para desarrollar un método quimiométrico para predecir el número de bromo. Las muestras se caracterizaron por espectroscopía infrarrojo cercano con transformada de Fourier (FT-NIR) y sus espectros fueron correlacionados por similitud utilizando análisis de componentes principales (PCA). Los modelos de calibración del número de bromo (BN) se llevaron a cabo por regresión de mínimos cuadrados parciales (PLS) y regresión polinomial múltiple (MPR). El PCA clasificó las muestras en livianas y pesadas y permitió determinar las variables espectrales más importantes. Estas variables están localizadas en las regiones entre 4000-4800 y 5200-6300 cm^{-1} . Los resultados obtenidos empleando esta metodología fueron muy cercanos a los obtenidos por los métodos standard. Por tanto, esta metodología puede ser una alternativa para el análisis del BN, por su bajo costo y menor tiempo de análisis.

Palabras clave: FT-NIR, nafta, número de bromo, PCA, PLS, MPR

Resumo

Neste trabalho, foram utilizados trinta e oito amostras de gasolina para desenvolver um método quimiométrica para prever o número de bromo. As amostras foram caracterizadas por Fourier no infravermelho próximo transformar espectroscopia (FT-NIR) espectros e foram correlacionados,

utilizando análise de componentes principais semelhante (PCA). Modelos de calibração, o número de bromo (BN) é realizada por mínimos quadrados parciais de regressão (PLS) e regressão polinomial múltipla (MPR). As amostras de PCA classificadas em leves e pesados e deixou-se determinar as variáveis espectrais mais importantes. Estas variáveis estão localizados nas regiões entre 4000-4800 e 5200-6300cm⁻¹. Os resultados obtidos usando esta metodologia são muito próximos daqueles obtidos por métodos padrão. Assim, este método pode ser uma alternativa para a análise de BN, o seu baixo custo e menor tempo de análise.

Palabras-chave: FT-NIR, a nafta, o número de bromo, PCA, PLS, MPR.

Introduction

In the Petrochemical Industry exists considerable interest in the development of analytical methods that provide high-quality information in short time. Petrochemical industries routinely monitor and control transformation processes using analytical information. Traditionally, standardized methods are used to determine physicochemical properties of petroleum crude oils and their fractions. These methodologies are time-consuming and require a significant amount of sample. The development of methods for accurate and fast analysis has become an urgent need to control the quality of petroleum crude oils and their fractions or cuts production processes. A large number of analytical techniques including high-performance liquid chromatography (HPLC), nuclear magnetic resonance (NMR), mass spectrometry (MS), fluorescence spectroscopy, Raman and infrared spectroscopy have been widely applied to the analysis of hydrocarbons and their derivatives [1-8]. These techniques provided excellent results in comparison with traditional methodologies of analysis; however, some of them are expensive and not always available in all laboratories. Fourier Transform Near Infrared Spectroscopy (FT-NIR) has been used as a promising alternative to replace traditional American Society for Testing and Materials methods (ASTM) to determine physicochemical properties of the petroleum crude oils and their fractions, because it requires a minimum sample treatment, the acquisition of spectra takes a few minutes and offers an acceptable repeatability for liquid and solid samples [6, 9].

Bromine Number (BN) is standardized methodology (ASTM D 1159 [10]), to determine olefin contents in light fractions of petroleum (distillates <350°C). The BN value is the amount of bromine which reacts with 100 grams of sample.

Olefins are not present in virgin crude oils; they

are produced during oil refining schemes which imply thermal and catalytic processing such as fluid catalytic cracking (FCC) to obtain gasoline and naphtha. Although olefins contained in gasoline can increase the octane number, they may have negative effects on engine deposits and exhaust emissions because of their active chemical properties, particularly about the formation of gums and photochemical smog [11]. Hence, the bromine number is an important quality parameter of light fractions of petroleum crude oils.

The combination of FT-NIR and chemometrics models has allowed developing robust methods for the prediction of properties of complex samples, such as petroleum crude oil and their fractions. The principal component analysis (PCA) allows discriminating samples through the analysis of score and loading plots [12]; while partial least squares regression (PLS-R) provides models for the prediction of different properties from spectroscopic data [13-14]. In this work, we developed a methodology to predict the bromine number analysis of Colombian naphtha samples, based on the combined use of FT-NIR, PCA, and PLS regression. The accuracy of the PLS regression for predicting bromine number was established by evaluating the root mean square error of prediction (RMSEP) and root mean square error of calibration (RMSEC) as well as the correlation coefficient (R^2). Although PLS regression is the chemometric method most widely used in combination with spectroscopic data, it is well known that the greatest problem in the methodology is the assumption of linear spectral data–property relationship. This assumption may be unacceptable for systems with strong intermolecular or intramolecular interactions, including π -stacking [15-18]. Therefore, also, we performed a Multiple Polynomial Regression to develop the MPR, we used principal components (PC) of the NIR spectral data of the naphtha samples, as inputs or X-variables

Experimental

Samples

Thirty-eight Colombian naphtha samples were used in this work. Conventional bromine number analysis was determined by electrometric titration (ASTM D1159) [10]. One Mettler Toledo model T90 titrator was used. The temperature for the analysis spanned 0-5°C, according to the standardized method. Table 1 shows results of bromine number and olefins obtained via the ASTM method.

Acquisition of NIR Spectra

FT-NIR spectra were recorded on a Bruker VERTX 70v spectrometer with a spectral resolution of 8cm⁻¹ over the range of 4000-13500cm⁻¹, at 20±0.1°C and 32 scans acquisition. Despite the fact that in the majority of works the acquisition of spectra is reported with 4cm⁻¹, the aim of selecting a resolution of 8 cm⁻¹ was to reduce the acquisition time to one-half, an important aspect for analyzing a high number of samples. The spectrometer is equipped with a deuterated L-alanine doped triglycine sulfate detector (DLATGS) and a transmittance sample cell of quartz with 0.5mm optical path. The acquisition time for 32 scans was approximately 30s.

The background spectrum of air was collected for every sample before the analysis in the single-beam mode. The spectral files were transformed to ASCII format using the IR Solution software and exported to The Unscrambler software, version X. Five naphtha samples were randomly chosen; ten spectra were recorded for each sample on ten different aliquots. Standard deviation determined was approximately 1x10⁻³ (in absorbance units) for any wavenumber.

Data analysis

Processing of spectroscopic data obtained including spectral pre-treatment, PCA, PLS, and validations was performed with the UnscramblerX and MPR was carried out with MATLAB® software.

Principal components analysis

PCA permits to find out samples pattern and which spectral intensities contribute most to this difference. PCA was done through the Unscrambler Software Version X by using Nonlinear Iterative Partial Least Squares algorithm (NIPALS). This algorithm handles missing values and is suitable for computing only the first few factors of a data set. The low information was extracted from the three principal components.

Table 1. Bromine number and w% of olefins in naphtha samples.

Samples	Bromine Number (Br ₂ /100 g)	Olefins* (wt%)	Samples	Bromine Number (Br ₂ /100 g)	Olefins (wt%)
N1	119	52.0	N20	12.5	ND
N2	131	63.1	N21	72.7	19.7
N3	113	43.4	N22	41.2	6.10
N4	89.2	19.2	N23	28.3	3.33
N5	59.2	9.12	N24	24.5	0.28
N6	46.1	2.02	N25	22.2	0.20
N7	44.0	22.2	N26	22.2	0.02
N8	112	51.1	N27	19.6	0.04
N9	128	64.8	N28	45.5	5.70
N10	121	50.9	N29	32.7	0.18
N11	101	29.8	N30	28.9	0.33
N12	55.4	9.63	N31	24.4	0.01
N13	27.6	1.42	N32	145	68.5
N14	25.7	14.1	N33	125	63.9
N15	90.2	41.9	N34	115	51.2
N16	82.1	43.4	N35	90.8	32.2
N17	62.7	26.8	N36	57.0	12.5
N18	43.6	11.1	N37	24.7	4.02
N19	21.3	ND [‡]	N38	15.1	0.21

* wt% of olefins was determined using PIANO analysis. ‡ ND: No determined

PLS regression

PLS-R was used to establish the relationship between NIR spectral intensities and the bromine number of naphtha samples. PLS-R is a useful procedure to find a common structure between one or more characteristic properties of the samples, called independent variables or simply Y variables, and the spectral intensities, called dependent variables or X variables. In our case, the bromine number in the naphtha samples is the Y variable, while the NIR spectral intensities are X variables. The linear PLS model finds a few variables, which are estimates of the latent variables (LVs) or their rotations. These variables are called X scores, and they are predictors of Y , and also model X ; i.e., both Y and X are assumed to be, at least partly, shaped by the same LVs. X scores are estimated as linear combinations of the original matrix of X variables with the weighted coefficients, W_{ka}^* ($a = 1, 2, \dots, A$). This could be more clearly seen in Equation 1.

$$t_{ia} = \sum W_{ka}^* X_{ik} \quad (1)$$

Where k denotes the number of X variables or spectral intensities, and i is the number of samples [19]. The regression model can be written according to the Equation 2

$$Y = b_0 + b_1 X_1 + \dots + b_k X_k \quad (2)$$

meaning that the observed response values are approximated by a linear combination of the

values of the predictors. The coefficients of that combination are called regression coefficients or b coefficients. These coefficients are calculated from X scores and Y score. A deeper analysis of the relationship between scores is beyond the scope of this paper, but it is correctly explained by Wold *et al.* [19]. Regression coefficients summarize the relationship between all predictors and a given response. For PLS, the regression coefficients can be computed for any number of components; i.e., for every latent variable, there are different b coefficients. When X variables have been centered, all predictors are brought back to the same scale; the coefficients show the relative importance of the X variables in the model. Predictors with a significant regression coefficient play a major role in the regression model; a positive coefficient shows a definite link with the response, and a negative coefficient indicates a negative relationship. Predictors with a small coefficient are negligible.

Multiple Polynomial Regression (MPR)

MPR is a particular form of linear regression in which the correlation between independent variable X and the dependent variable Y is considered as an n th order polynomial. In a view to performing the regression, principal components (PC) for each sample were used as independent variables and bromine number (BN) values as dependent variables. So, multiple polynomial regression can be represented by the following Equation;

$$BN_i = a_{i1} PC1_i + a_{i2} PC2_i + a_{i3} PC3_i + a_{i4} (PC1_i)^2 + a_{i5} (PC2_i)^2 + \dots + \varepsilon \quad (3)$$

Where a_{ir} ($i, j = 1, 2, 3, \dots$) are the coefficients that model the relationship between BN and PC for each sample [20]. The goal of MPR is to determine the values of a_{ir} and minimizing the error, ε .

NIR analysis

Figure 1 shows the NIR spectra of naphtha samples, all showing very similar spectra. In Figure 1, it is though to differentiate between the NIR spectra of heavy and light naphtha samples. However, it is possible to find small differences in the region spanning from 4400-4800 cm^{-1} , in particular between the heaviest and light naphtha samples. The bands spanning the region 4400 to

4800 cm^{-1} region correspond to a combination of the fundamental vibrational modes =CH- stretch and C=C stretch. Also, appreciable differences in the region spanning from 5200 to 6500 cm^{-1} can be observed, due to overtones of the vibrational modes corresponding to asymmetric and symmetric stretching of C-H bonds in methyl and methylene group in aliphatic and aromatic compounds. However, even though it is possible to differentiate the naphtha samples through six bands located at 4258, 4335, 4393, 5716, 5866 and 6113 cm^{-1} , only after modeling is it possible to know if these differences are associated with the Y variable (bromine number).

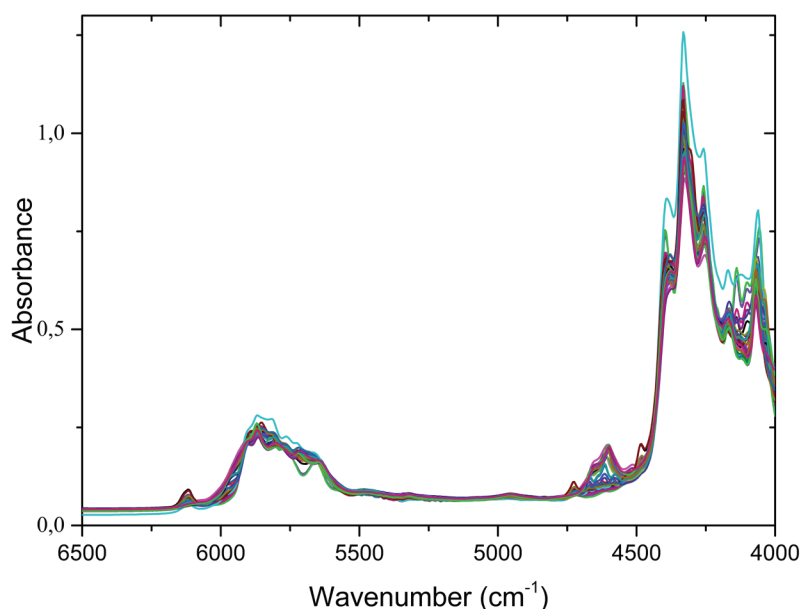


Figure 1. NIR spectra of naphtha samples.

Figure 2a and 2b shows average FT-NIR spectra of heavy and light naphtha samples in the 6500–4000 cm^{-1} range. The overall spectral features were similar to each other. However, slight differences are observed around 4600 cm^{-1} , corresponding to aromatic bands. Light naphtha samples show

a higher concentration of aromatic compounds. These bands corresponding to olefin compounds, particularly from a combination of =C–H stretching at 3100–3000 cm^{-1} and C=C ring stretching at 1600–1450 cm^{-1} .

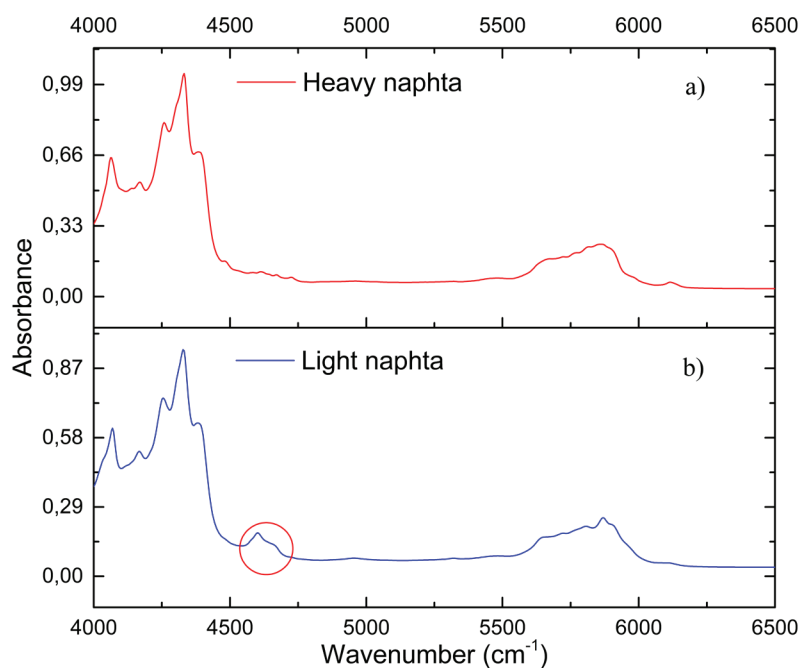


Figure 2. Average of FT-NIR spectra of (a) heavy and (b) light naphtha samples, respectively.

Principal component analysis

Principal component analysis was performed using three spectral regions: 4000-4800cm⁻¹, 5200-6350cm⁻¹, and 4000-6350cm⁻¹. In all cases, the first three PCs explained more than 85% of the total variance. These three PCs were used as X-variables in multiple polynomial regression. Figs. 3 show the PCA performed on the three spectral regions.

The plots show two perfectly distinct groups of naphtha samples; twenty-four heavy naphtha samples with a range of bromine number spanning from 12.5 to 72.7 and a group of fourteen light naphtha samples with bromine number between

82.1 and 145. However, N14 y N32 have BN values out of this range. According to Figure 3a, 3b, and 3c, the N32 sample could be considered an outlier, while N14 has a percent of olefin similar to samples in the first range (Table 1). The discrimination between heavy and light naphtha samples occur in PC1; for the region spanning from 4000 to 4800 cm⁻¹, heavy naphtha samples have negative values and light naphtha samples have positive values of PC1, while for the region spanning from 5200 to 6350cm⁻¹, heavy naphtha samples have positive values and light naphtha samples negative values of PC1. The behavior of PC2 is entirely contrary.

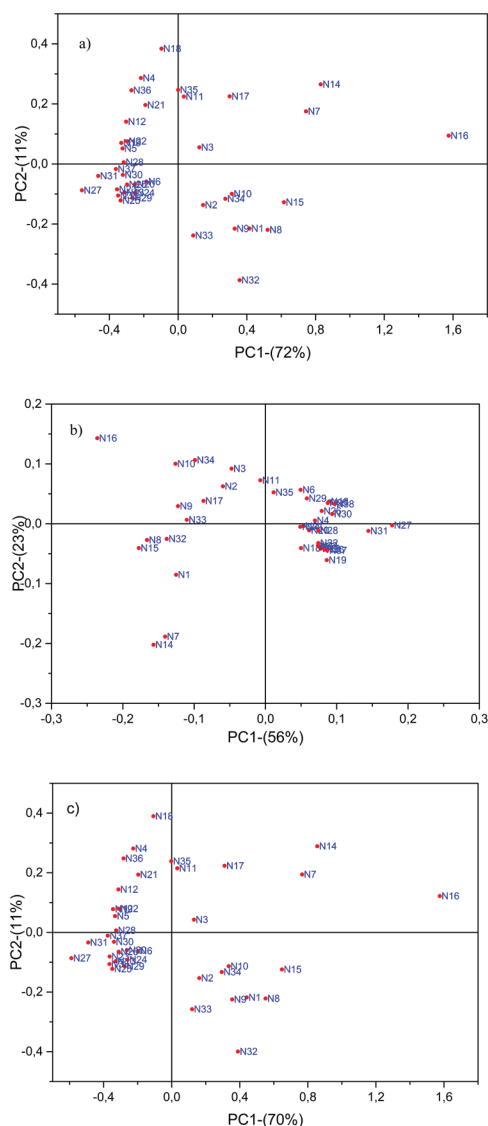


Figure 3. Scores plots of principal components PC1 and PC2 of thirty-eight naphtha samples in the region (a) 4000-4800cm⁻¹ (b) 5200-6350cm⁻¹ and (c) 4000-6350cm⁻¹

Figures 4a and 4b show loading for PC1 of NIR spectral regions from 4000 to 4800 cm^{-1} and from 5200 to 6350 cm^{-1} . As mentioned above, the bands in the spectral region between 4000 and 4800 cm^{-1} mainly correspond to a combination of the fundamental vibrational modes C=C stretch of olefins in naphtha samples. While the spectral region from 5200 to 6350 essentially corresponds to overtones of the vibrational modes of symmetric and asymmetric stretching of C-H and =CH- bond in saturated unsaturated aliphatic compounds. Therefore, the first principal component

(PC1) may be attributed to the olefin part in naphtha samples and the second one to aliphatic compounds.

PLS analysis

Intensities from FT-NIR spectra and bromine number for the studied naphtha samples were correlated using PLS regression. After dividing samples into the two groups, i.e., light and heavy naphtha samples according to PCA results, PLS regression was employed to build quantitative calibration models for each group.

Table 2. PLS calibration and validation results for bromine number analysis

Samples	Wavenumber range/ cm^{-1}	LV	R ²	RMSEC (Br ₂ /100g)	RMSEP (Br ₂ /100g)	Excluded samples
Light (24)	5200-6350	7	0.98	1.68	3.88	1
Heavy (14)	5200-6350	7	0.99	1.38	4.69	0

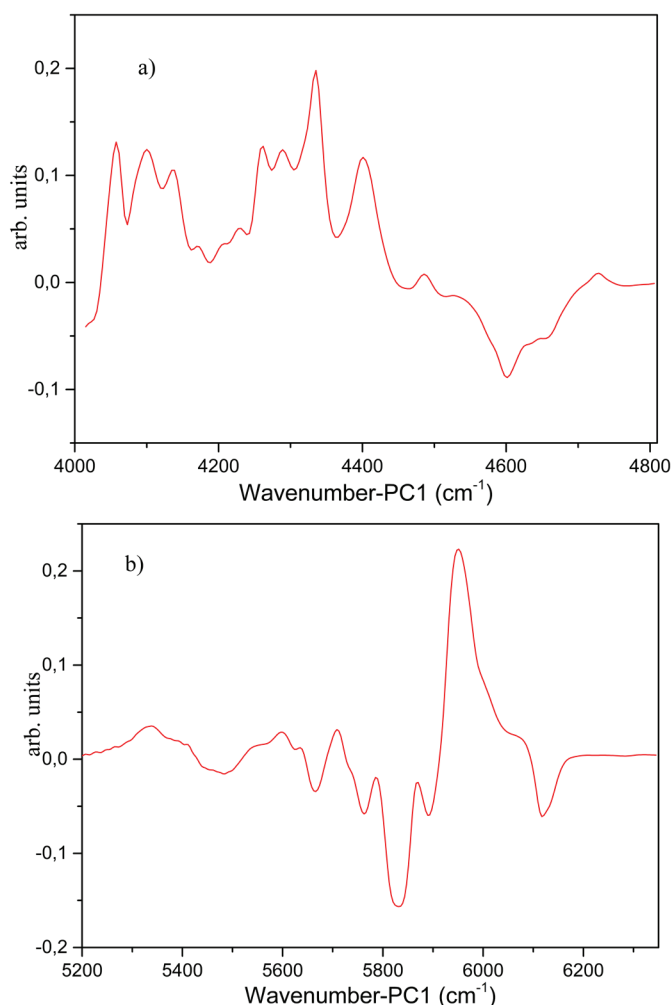


Figure 4. Loadings for PC1 of NIR spectral regions from (a) 4000 to 4800 cm^{-1} and (b) 5200 to 6350 cm^{-1} .

The spectral region that best related to BN was between 5200 and 6350 cm^{-1} . The development of predictive models was conducted in two stages; the first one consisted of building calibration models (Fig.4), and the second stage was the validation of the models. The evaluation of the calibration performance was assessed by the root mean squared error of calibration (RMSEC). The standard root mean squared error of prediction (RMSEP) for validation gives an evaluation of the predicted performance of the quantitative model obtained previously. The RMSEC and RMSEP were 1.68 and 3.88 for light naphtha samples, and 1.68 and 4.69 for heavy naphtha samples, respectively. The RMSEC and RMSEP are shown

in Table 2 together with the NIR spectral regions, explained variance and factor numbers employed. In all cases, seven factors were used and just one sample (N17), in the model for light naphtha samples was eliminated as an outlier. When the spectral region from 5200 to 6350 cm^{-1} was selected, the best RMSEP was obtained. If many variables (all wavenumber signals) are used, the model may be unstable; it means that small changes in the spectrum of the order of the spectral noise may produce statistically significant changes in the prediction. These results found in this research indicate that the NIR spectral data can be actually used for the bromine number estimation of naphtha samples.

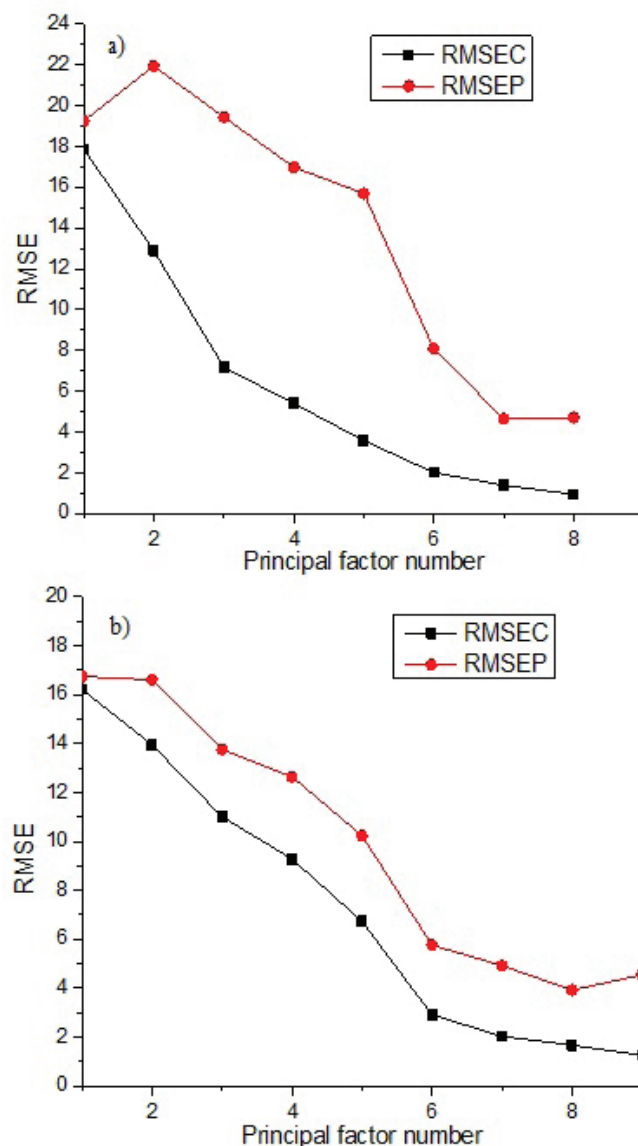


Figure 5. Number of Principal factor versus RMSE.

PMR

To compare with PLS-R, multiple polynomial regression calibrations (MPR) was obtained. We used the same wavenumber range selected for PLS-R. Three first principal components (PC), were used as X variables and they explained more than 85% of the data variance. Figure 5 shows the best results when a third order polynomial function and twenty terms (PC1 PC2 PC3 PC1PC2 PC1PC3 PC2PC3 PC1²PC²PC3² PC1PC2PC3 PC1²PC2 PC2²PC3 PC3²PC1

PC1²PC3 PC2²PC1 PC3²PC2 PC1³ PC2³ PC3³ 1) were used. The RMSEC value is 7.44Br₂/100 g. This value is very high compared to the one obtained via PLS-R. It is possible to use a fourth order polynomial function. However, thirty-five terms will be necessary. When a fourth order polynomial function is used, RMSEC can be decreased to as low as 1.00; however, this result can be due to over-fitting. The preceding results show that the spectral data–bromine number property relationship is linear.

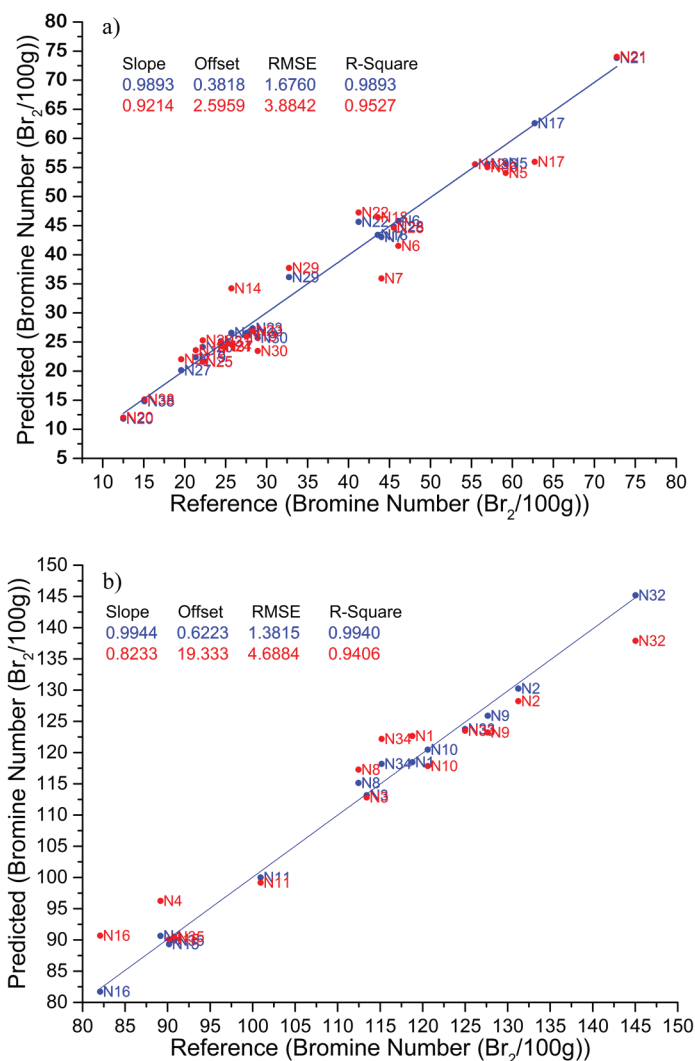


Figure 6. Scatter plots of NIR-PLS prediction values vs reference values for BN both in calibration (blue) and validation (red). (a) Light naphtha samples; (b) Heavy naphtha samples.

Conclusion

In this study, the determination of bromine number of naphtha samples by FT-NIR coupled to PCA,

PLS and MPR was studied. The PCA allowed identifying two groups of naphtha samples, i.e., light and heavy naphtha samples according to their bromine number. The PLS multivariate

calibration allowed obtaining prediction model for determination of bromine number. RMSEC and RMSEP for light naphtha samples were 1.6760 and 3.8842, respectively. While for heavy naphtha samples, RMSEC and RMSEP were 1.3815 and 4.6884, respectively. Finally, results obtained using multiple polynomial regression were not successful, and its error of prediction is higher than the error of prediction obtained with PLS regression, probably because assumed hypothesis about the non-linear relation between NB and FT-NIR signals is unacceptable. The validation results using PLS indicate that there are high consistencies between the FT-NIR prediction and those measured by the reference standard method (ASTM D1159). The proposed method is much faster than the reference method. This allows the determination of BN from a single spectrum, with the feasibility of further application for online monitoring.

Acknowledgements

This work was supported by the Research Agreement between ECOPETROL S.A. and Universidad Industrial de Santander. This work was also financially supported by COLCIENCIAS under the project no. FP44842-039-2015 and UIS through the Vicerrectoría de Investigación y Extensión

References

- [1] Shi TP, Xu ZM, Cheng M, Yun-Xiang H, Ren-An W. Characterization Index for vacuum Residua and their Subfractions. *Energy Fuels*. 1999;13(4):871–6.
- [2] Falla FS, Larini C, Le Roux GAC, Quina FH, Moro LFL, Nascimento CAO. Characterization of crude oil by NIR. *J Petrol Sci Eng*. 2006;51(1–2):127–37.
- [3] Borges B, Acevedo S. Caracterización estructural de distintas fracciones aisladas del crudo extrapesado de Carabobo. *Rev. Latinam. Metal. Mater*. 2007;27(2):83–94.
- [4] Mullins OC. The Modified Yen Model. *Energy Fuels*. 2010;24(6):2179–207.
- [5] Bezerra De Lira L, Vera F, Fernandes C, Silveira AP, Stragevitchb L, Pimentel MF. Prediction of properties of diesel/biodiesel blends by infrared spectroscopy and multivariate calibration. *Fuel*. 2010;89:405–9.
- [6] Orrego-Ruiz JA, Guzmán A, Molina D, Mejía-Ospino E. Mid-infrared Attenuated Total Reflectance (MIR-ATR) predictive models for asphaltene contents in vacuum residua: asphaltene structure–functionality correlations based on Partial Least-Squares Regression (PLS-R). *Energy Fuels*. 2011;25:3678–86.
- [7] Orrego-Ruiz JA, Mejía-Ospino E, Carbognani L, López-Linares F, Pereira-Almao P. Quality Prediction from Hydroprocessing through Infrared Spectroscopy (IR) *Energy Fuels*. 2012;26:586–93.
- [8] Meléndez LV, Lache A, Orrego-Ruiz JA, Pachón Z, and Mejía-Ospino E. Prediction of the SARA analysis of Colombian crude oils using ATR–FTIR spectroscopy and chemometric methods. *J. Pet. Sci. Eng*. 2012;90-91:56–60.
- [9] Chung H, Ku M, Lee JS. Comparison of near-infrared, infrared, and Raman spectroscopy for the analysis of heavy petroleum products *Appl. Spectrosc*. 2000;54:239–45.
- [10] ASTM D1159–07, Standard Test Method for Bromine Numbers of Petroleum Distillates and Commercial Aliphatic Olefins by Electrometric Titration, ASTM International, West Conshohocken, PA, 2012.
- [11] Carbognani L, Lopez-Linares F, Trujillo M, Wu Q, Carbognani J, Pereira-Almao P. On the determination of olefins in heavy oil, bitumen and their upgraded products. *Div. Energy Fuels. ACS Energy Fuels Preprints*. 2012;57:685-8.
- [12] Fusheng O, Xu PXuhong, Z, Xuan L, Huixin W. Effect of Operating Conditions on Olefin Distribution in FCC Gasoline as Part of an Olefin Reduction Process *China Pet. Process. Petrochem. Process*. 2010;12:34-42
- [13] Wold S. Chemometrics; what do we mean with it, and what do we want from it? *Chemom. Intell. Lab. Syst*. 1995;30:109–15.
- [14] N. Aske, H. Kallevik, J. Sjöblom. Determination of saturate, aromatic, resin and asphaltenic (SARA) components in crude oils using infrared and near-infrared spectroscopy. *Energy Fuels*. 2001;15:1304–12.
- [15] Satya S, Roehner RM, Deo MD, Hanson FV. Estimation of properties of crude oil residual fractions using chemometrics. *Energy. Fuels*. 2007;21:998–1005.
- [16] Syunyaev RZ, Balabin RM. Frequency dependence of oil conductivity at high pressure. *J. Disp. Sci. Technol*. 2007;28:419-27.
- [17] Syunyaev RZ, Balabin RM. Polarization of Fluorescence of Asphaltene Containing Systems. *J. Disp. Sci. Technol*. 2008;29:1505-14.

- [18] Kaplan IG. Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials. México: John Wiley & Sons; 2006.
- [19] Wold S, Sjöström M, Eriksson L. PLS-regression, a basic tool of Chemometrics. Chemom. Intell. Lab. Syst. 2001;58:109-30.
- [20] Cheung V. A comparative study of the characterization of colour cameras by means of neural networks and polynomial transforms. Color. Technol. 2004;120:19-25