J Am Oil Chem Soc (2009) 86:581–585
DOI 10.1007/s11746-009-1374-7

ORIGINAL PAPER

^{1}H NMR and Multivariate Calibration for the Prediction of Biodiesel Concentration in Diesel Blends

Marcos R. Monteiro · Alessandra R. P. Ambrozin · Luciano M. Lião · Elisangela F. Boffo · Edenir R. Pereira-Filho · Antonio Gilberto Ferreira

Received: 13 August 2008 / Revised: 13 March 2009 / Accepted: 18 March 2009 / Published online: 21 April 2009

Abstract In this work, the use of ^{1}H-NMR spectroscopy and a statistical approach to the analysis of biodiesel concentrations in blends with conventional diesel is described. For this, we performed ^{1}H-NMR analyses using distinct mixtures of biodiesel from soybean and castor oil in mineral diesel, in concentrations ranging from 0.5 to 30%, and then we applied partial least squares regression (PLS) and principal components regression (PCR) to such data. So, six models were designed and they were evaluated through statistical parameters and through the analysis of four samples prepared in the laboratory. Briefly, a PLS model, obtained through the selection of aromatic, aliphatic and methoxy spectral regions, was quite suitable for the prediction of biodiesel concentrations greater than 2.0%. Deviations of real and predicted values were found to B2 commercial blends, indicating that this model can only be applied to blends exceeding a 2.0% level of biodiesel in petroleum diesel. In conclusion, the ^{1}H-NMR-PLS method is fairly useful for the quality control of biodiesel–diesel blends, whose commercialisation has increased in the last few years.

M. R. Monteiro (✉) · A. R. P. Ambrozin
Centro de Caracterização e Desenvolvimento de Materiais, Departamento de Engenharia de Materiais, Universidade Federal de São Carlos, São Carlos 13560-971, Brazil
e-mail: monteiro@ccdm.ufscar.br

L. M. Lião
Instituto de Química, Universidade Federal de Goiás, Campus Samambaia, Goiânia 74001-970, Brazil

E. F. Boffo · E. R. Pereira-Filho · A. G. Ferreira
Departamento de Química, Universidade Federal de São Carlos, São Carlos 13560-605, Brazil

Keywords Biodiesel · Diesel · Blends · NMR · Chemometrics · Partial least squares regression · Principal components regression

Introduction

Biodiesel, which is an alternative fuel basically composed of fatty acid mono-alkyl esters, is the major substitute for petroleum-derived diesel since their physical properties are very similar, allowing the use of pure or blended biodiesel without any modification to the diesel engine and in the existing fuel distribution and storage infrastructure [1, 2]. Recently, blends of biodiesel with mineral diesel became commercially available all over the world. In the United States, the use of a blend of 20% (v/v) of biodiesel in diesel, called B20, is common [3]. ASTM D 975, the mineral diesel standard in the United States, allows for up to B5 while ASTM D 7467, the biodiesel–diesel blend standard, allows from B6 to B20. Europe, which is the world's largest producer of biodiesel, employs B2 blends (2% of biodiesel in diesel) in their engines and intends to increase this amount to 5.75% in 2010, and to 20% in 2020 [4]. EN 590, the European diesel standard, allows for up to B5. In Brazil, the commercialisation of B2 became mandatory from January 2008 [5], however, the use of B3 was established by the National Council of Energy Policy (CNPE) through its resolution No. 2 on 14 March 2008.

Some analytical methods have been developed to determine the amount of biodiesel in biodiesel–diesel blends. In general, these analyses include the use of spectroscopic methods, mainly infrared spectroscopy [6]. Nuclear Magnetic Resonance (NMR) is a versatile spectroscopic method that has become one of the most powerful techniques for the elucidation of the structure of chemical

compounds. Although ^{1}H NMR is frequently used for monitoring biodiesel synthesis and quality [6], few reports describe the use of NMR for biodiesel blend level determination [3, 7, 8]. In the first works [7, 8], the integrals of the peaks at 3.5–3.7 ppm (methoxy group of biodiesel), 0.5–3.0 ppm (methyl and methylene hydrogens from biodiesel and diesel), and 5.3–5.4 ppm (olefin hydrogens of biodiesel) were used to quantify biodiesel in biodiesel–diesel blends. Recently, Monteiro et al. [8] showed that NMR can be used to quantify different types of biodiesel in any diesel. For this, some relationships of integrals related to peaks at 3.30–0.42, 3.65–3.55, and 8.80–6.50 ppm were employed. The results indicated that the quantification of biodiesel in diesel by ^{1}H NMR is not affected by either biodiesel or diesel types and thus this technique is especially valuable for such determinations. Therefore, these two-first works showed that the quantification of biodiesel from soybean in biodiesel–diesel blends was possible and the latter demonstrated that NMR is a valuable technique that allows quantification of any biodiesel type in biodiesel–diesel blends.

In the present work, we proposed, for the first time, the use of ^{1}H-NMR spectroscopy and multivariate calibration in the analysis of fatty acid methyl esters (FAME) in blends with conventional diesel. The use of NMR for this purpose first involves the development of calibration models to relate the spectra of biodiesel blends with the analytical data. So, chemometric tools such as partial least squares regression (PLS) and principal components regression (PCR) were used to build calibration models. In fact, as a huge amount of NMR data is produced, chemometric analysis is frequently needed to extract the desired information. Chemometrics, which is the application of mathematical, statistical and logical–mathematical methods to chemical issues, is capable of treating large quantities of information [9]. Chemometric methods have been applied to several spectroscopy data in order to predict biodiesel content in mineral diesel [6]. Recently, NMR and chemometrics was used to determine adulteration of biodiesel–diesel blends with vegetable oils [10].

Therefore, in the present work, we describe the use of ^{1}H-NMR spectroscopy in combination with a statistical approach for the determination of biodiesel concentration in conventional diesel. We performed ^{1}H-NMR analyses using distinct mixtures of biodiesel from soybean and castor oil in mineral diesel and then applied PCR and PLS modeling methods to the ^{1}H-NMR data. The models were evaluated according to statistical parameters as well as through the analysis of four samples prepared in the laboratory. The predictive ability of the best model obtained was also investigated in four B2 commercial samples.

Experimental Procedures

Biodiesel from castor and soybean oils were prepared on a laboratory scale according to procedures described by Oliveira et al. [11] and Conceição et al. [12], respectively. Two commercial diesel samples were acquired from different gas stations in the São Paulo state and Distrito Federal (Brazil). B2 commercial samples were randomly obtained from different gas stations in the São Paulo state.

Both biodiesel and diesel samples had been previously analysed by ^{1}H NMR to verify their purity, but they were not used in the construction of the models. Biodiesel samples were considered as 100% pure since no significant contaminants (glycerides, glycerol, methanol, among others) were observed. The ^{1}H-NMR analyses of diesel samples allowed us to certify that biodiesel was not present in them. In addition, the sulphur content, distillation profile, flash point, and specific mass of these samples were evaluated according to ASTM D 4294, ASTM D 86, NBR 14598, and NBR 14065 standards, respectively. Both diesel samples from the São Paulo and Distrito Federal met specifications established by the Brazilian National Agency of Petroleum, Natural Gas and Biofuels (ANP) in resolution No. 15 from 19 July 2006.

Biodiesel blends were prepared by mixing each biodiesel (from castor and soybean oils) with the diesel from Distrito Federal to define the following blend levels: 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0, 8.0, 9.0, 10.0, 15.0, 20.0, 25.0, and 30.0% (v/v). Therefore, 40 samples were prepared for the construction of the calibration models.

An additional series of B2, B5, B10, and B20 samples in diesel from São Paulo were prepared in the laboratory following the same procedure previously described. Such samples were used for external validation.

All ^{1}H-NMR experiments were carried out at room temperature on a Bruker DRX 400—9.4 Tesla spectrometer, using a 5-mm inverse-detection probehead with a z-gradient. The spectra were obtained at 400.21 MHz for ^{1}H, using $CDCl_3$ as solvent, and TMS as the internal standard. For each analysis, 300 μL of the pure or blended sample were dissolved in the same volume of solvent. Thirty-two pulses were employed to the acquisition of the spectra, with an acquisition time of 7 s, a spectral width of 4596 Hz, and a relaxation delay of 7 s. The T_1 relaxation time (1.5 s) was also obtained using the inversion recovery pulse sequence.

Spectra were processed with 32 k data points and using an exponential weighing factor corresponding to a line broadening of 0.3 Hz. The phase and baseline were automatically corrected.

All statistic calculations were carried out using the Pirouette® software (v. 3.11, InfoMetrix, Woodinville,

Washington, USA). NMR spectra were transformed into ASCII files and the resulting data matrices were imported into the Origin software (v. 5.0, Microcal, USA) and, thus to the Pirouette®. All dataset were employed for the calibration. Initially, an exploratory analysis of the calibration set was carried out, employing principal component analysis (PCA). This allowed the optimisation of pre-processing and transformation parameters, which would be applied to the development of PLS and PCR calibration models. The best parameters for both were obtained using autoscaling, normalisation and first order derivative (every 25 data points). Cross-validation was also used to evaluate the models.

Results and Discussion

Forty biodiesel–diesel blends and one pure diesel were analysed by ^{1}H NMR. Two types of biodiesel (from castor and soybean oil) and diesel from Distrito Federal were used for preparing such mixtures. The obtained spectral data were employed for the construction of six calibration models (Table 1) in order to predict FAME concentration in biodiesel–diesel samples. In this investigation, we intended to evaluate the influence of the spectral region and the calibration model type on such quantification. The final goal was to develop a reliable and versatile statistical NMR method, which could quantify any biodiesel in a mixture with conventional diesel. So, the proposed PCR and PLS models were classified according to spectral regions (Table 1). Noise and non-informative ranges of the spectra were excluded. The selected spectral regions (Fig. 1), 3.30–0.42 (R_1), 3.65–3.55 (R_2), and 8.80–6.50 ppm (R_3), are related to aliphatic, methoxy and aromatic hydrogens, respectively. The singlet near 3.6 ppm (from the methyl ester moiety) is characteristic of a biodiesel spectrum and could be used to distinguish biodiesel from diesel [3]. It is worth mentioning that the correct choice of spectral region in PLS and PCR design is a very important issue, since it will determine both the precision and accuracy of the analytical method [13].

Some statistical parameters were obtained for each model (Table 1), namely: principal components (PCs); variance percent (Var%); standard error of validation

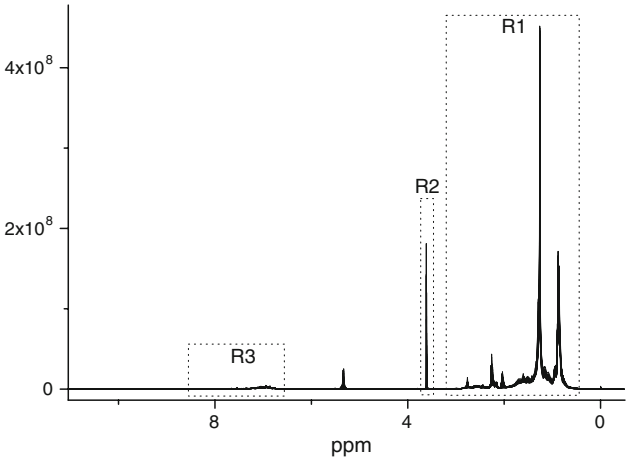

Fig. 1 Overloaded spectra of biodiesel–diesel blends with the selected regions used for the construction of the models

Table 1 Some statistics related to classification models

Model	Multivariate calibration method	Spectral region[a]	PCs[b]	Var%[c]	SEV[d]	PRESS Val[e]	r Val[f]	SEC[g]	PRESS Cal[h]	r Cal[i]	SEC/SEV[j]
1	PLS	R_3 and R_2	4	87.57	0.83	28.32	0.9946	0.53	10.27	0.9980	0.64
2	PCR	R_3 and R_2	5	90.97	0.90	33.52	0.9936	0.81	22.77	0.9956	0.90
3	PLS	R_1 and R_2	3	81.75	0.54	11.84	0.9977	0.41	6.19	0.9988	0.76
4	PCR	R_1 and R_2	3	83.44	0.60	14.75	0.9972	0.54	10.69	0.9980	0.90
5	PLS	R_1, R_2, and R_3	4	81.33	0.50	10.21	0.9981	0.30	3.32	0.9994	0.60
6	PCR	R_1, R_2, and R_3	4	83.95	0.71	20.93	0.9960	0.63	14.51	0.9972	0.89

[a] R_1 = 3.30–0.42 ppm, R_2 = 3.65–3.55 ppm, R_3 = 8.80–6.50 ppm

[b] *PCs* principal components

[c] *Var%* variance percent

[d] *SEV* standard error of validation

[e] *PRESS Val* predicted residual error sum of squares of validation

[f] *r Val* coefficient of correlation between the real concentration and the predicted concentration during the validation

[g] *SEC* standard error of calibration

[h] *PRESS Cal* predicted residual error sum of squares of calibration

[i] *r Cal* coefficient of correlation between the real concentration and the predicted concentration during the calibration

[j] Similarity criterion

(SEV); predicted residual error sum of squares of validation (PRESS Val); coefficient of correlation between the real concentration and the predicted concentration during the validation (r Val); standard error of calibration (SEC); predicted residual error sum of squares of calibration (PRESS Cal); coefficient of correlation between the real concentration and the predicted concentration during the calibration (r Cal); and, the similarity criterion obtained from the SEC/SEV relationship. The number of PCs for each model was chosen through the lowest obtained PRESS values.

The six calibration models were compared in terms of such parameters. The coefficient of correlation values (r Cal and r Val) near 0.99 showed that there is good agreement between the real and predicted concentrations. In general, these values must be greater than 0.9. Moreover, the SEC/SEV relationships indicated that all models were well adjusted, since the values lay within the 0.5–1 range. Table 1 also showed that model 5, which employs R_1, R_2 and R_3 regions, was the best one. Using four principal components, this model captured 81.33% of the overall variance and it showed the lowest values for SEC, SEV, PRESS Cal, and PRESS Val. The coefficient of correlation values showed high correlation between real and predicted concentrations. In fact, Table 2 shows that predicted (by model 5) and real concentrations of the forty biodiesel–diesel blends were quite similar.

Furthermore, the results from Table 1 pointed out that PLS models were better than all PCR ones. This was evidenced by the lowest values of the statistical parameters of all PLS models, when the comparison is performed between two equivalent models, such as 1 and 2, 3 and 4, and so on. In fact, PLS regression is currently used for the analysis of mixtures.

Moreover, these parameters indicated that the utilisation of R_1 and R_2 or all regions (R_1, R_2, and R_3) provided more suitable models than the ones obtained from R_2 and R_3 regions. Therefore, such regions provide important information concerning the biodiesel content and can be used for the calibration.

Besides cross-validation, the predictive abilities of the models were evaluated through external validation, using four biodiesel–diesel blends prepared in the laboratory. This allowed assessment of the performance of each model in the prediction of biodiesel concentration (Table 3). The samples had 2% (LS1, laboratory sample 1), 5% (LS2, laboratory sample 2), 10% (LS3, laboratory sample 3), and

Table 2 Predicted concentration by model 5 of biodiesel–diesel blends

Real concentration	Predicted concentration of biodiesel from soybean oil–diesel blends	Predicted concentration of biodiesel from castor oil–diesel blends
0.5	1.3	0.6
1	0.4	1.3
1.5	1.1	1.3
2	2.4	2.2
2.5	2.4	2.7
3	2.9	3.3
3.5	3.5	3.0
4	4.1	4.1
4.5	4.7	5.0
5	4.6	4.9
5.5	5.2	6.0
6	5.3	6.0
7	7.1	6.4
8	8.2	7.9
9	9.3	9.1
10	10.4	9.7
15	14.5	14.2
20	21.1	19.1
25	24.1	26.2
30	29.3	30.4

Results expressed as volume%

Table 3 External validation of models through the analysis of samples prepared in laboratory

Sample	Real concentration	Predicted concentration by					
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
LS1[a]	2	2.6	2.3	2.5	2.4	2.4	2.2
LS2[b]	5	4.1	3.7	5.4	4.8	5.0	4.0
LS3[c]	10	9.4	10.4	10.5	10.5	9.8	10.7
LS4[d]	20	20.2	20.5	20.6	20.5	20.0	20.2

Results expressed as volume %

[a] *LS1* laboratory sample 1

[b] *LS2* laboratory sample 2

[c] *LS3* laboratory sample 3

[d] *LS4* laboratory sample 4

20% (LS4, laboratory sample 4) of biodiesel (v/v) in diesel from São Paulo state. As we expected, model 5 predicted the amount of biodiesel with the lower relative error when compared to the others. This model allowed the prediction of biodiesel concentration with relative errors of 19.0, 0.4, 1.6, and 0.05% for samples LS1, LS2, LS3, and LS4, respectively. This indicated that such model is only suitable for the quantification of samples that have biodiesel concentration greater than 2.0%. Moreover, the higher the concentration, the lower the prediction error. The same results were observed in the analysis of four B2 commercial samples. The predicted concentrations by model 5 were 1.5, 1.8, 2.1, and 2.4%, corresponding to relative errors from 5 to 25%.

Acknowledgments We are grateful to the Centro de Caracterização e Desenvolvimento de Materiais (CCDM,UFSCar/UNESP), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Financiadora de Estudos e Projetos (FINEP) for the financial support.

References

1. Agarwal AK (2007) Biofuels (alcohol and biodiesel) applications as fuels for internal combustion engines. Prog Energ Combust Sci 33:233–271
2. Tat ME, Van Gerpen JH (1999) The kinematic viscosity of biodiesel and its blends with diesel fuel. J Am Oil Chem Soc 76:1511–1513
3. Knothe G (2001) Determining the blend level of mixtures of biodiesel with conventional diesel fuel by fiber-optic near-infrared spectroscopy and 1H nuclear magnetic resonance spectroscopy. J Am Oil Chem Soc 78:1025–1028
4. Directive 2003/30/EC (OJ L 123, 17.5.2003) of the European Parliament and of the Council on the promotion of the use of biofuels or other renewable fuels for transport
5. Brazilian federal law no 11.097/05 from 13/01/2005
6. Monteiro MR, Ambrozin ARP, Lião LM, Ferreira AG (2008) Critical review on analytical methods for biodiesel characterization. Talanta 77:593–605
7. Diehl B, Randel G (2007) Analysis of biodiesel, diesel and gasoline by NMR spectroscopy–a quick and robust alternative to NIR and GC. Lipid Technol 19:258–260
8. Monteiro MR, Ambrozin ARP, Lião LM, Ferreira AG (2009) Determination of biodiesel blend levels in different diesel samples by ^{1}H NMR. Fuel 88:691–696
9. Ferreira MMC, Antunes AM, Melgo MS, Volpe PLO (1999) Quimiometria I: calibração multivariada, um tutorial. Quim Nova 22:724–731
10. Monteiro MR, Ambrozin ARP, Santos MS, Boffo EF, Pereira-Filho ER, Lião LM, Ferreira AG (2008) Evaluation of biodiesel–diesel quality using ^{1}H NMR and chemometrics. Talanta 78:660–664
11. Oliveira JS, Montalvão R, Daher L, Suarez PAZ, Rubim JC (2006) Determination of methyl ester contents in biodiesel blends by FTIR-ATR and FTNIR spectroscopies. Talanta 69:1278–1284
12. Conceição MM, Candeia RA, Silva FC, Bezerra AF, Fernandes VJ Jr, Souza AG (2007) Thermoanalytical characterization of castor oil biodiesel. Renew Sust Energ Rev 11:964–975
13. Oliveira FCC, de Souza ATPC, Dias JA, Dias SCL, Rubim JC (2004) A escolha da faixa espectral no uso combinado de métodos espectroscópicos e quimiométricos. Quim Nova 27: 218–225