# Analyst

## RSCPublishing

# Simultaneous determination of hydrocarbon renewable diesel, biodiesel and petroleum diesel contents in diesel fuel blends using near infrared (NIR) spectroscopy and chemometrics

Julio Cesar Laurentino Alves* and Ronei Jesus Poppi

Highly polluting fuels based on non-renewable resources such as fossil fuels need to be replaced with potentially less polluting renewable fuels derived from vegetable or animal biomass, these so-called biofuels, are a reality nowadays and many countries have started the challenge of increasing the use of different types of biofuels, such as ethanol and biodiesel (fatty acid alkyl esters), often mixed with petroleum derivatives, such as gasoline and diesel, respectively. The quantitative determination of these fuel blends using simple, fast and low cost methods based on near infrared (NIR) spectroscopy combined with chemometric methods has been reported. However, advanced biofuels based on a mixture of hydrocarbons or a single hydrocarbon molecule, such as farnesane (2,6,10-trimethyldodecane), a hydrocarbon renewable diesel, can also be used in mixtures with biodiesel and petroleum diesel fuel and the use of NIR spectroscopy for the quantitative determination of a ternary fuel blend of these two hydrocarbon-based fuels and biodiesel can be a useful tool for quality control. This work presents a development of an analytical method for the quantitative determination of hydrocarbon renewable diesel (farnesane), biodiesel and petroleum diesel fuel blends using NIR spectroscopy combined with chemometric methods, such as partial least squares (PLS) and support vector machines (SVM). This development leads to a more accurate, simpler, faster and cheaper method when compared to the standard reference method ASTM D6866 and with the main advantage of providing the individual quantification of two different biofuels in a mixture with petroleum diesel fuel. Using the developed PLS model the three fuel blend components were determined simultaneously with values of root mean square error of prediction (RMSEP) of 0.25%, 0.19% and 0.38% for hydrocarbon renewable diesel, biodiesel and petroleum diesel, respectively, the values obtained were in agreement with those suggested by reference methods for the determination of renewable fuels.

## 1    Introduction

The environmental impact caused by the use of fossil fuels and the need to decrease the amount of pollutants such as greenhouse gas (GHG) and particulate matter emitted into the atmosphere due to combustion emissions,[1,2] and the desire to reduce the importation of petroleum and its derivatives and to promote rural economic activity have led the governments of many countries[1,3] to encourage or make mandatory the use of the so-called biofuels or renewable fuels, produced from vegetable or animal biomass, in order to replace the use of more pollutant fuels from non-renewable resources, such as petroleum derivatives, mainly in the transport economic sector.

The liquid biofuels may be classified as first generation biofuels, second generation biofuels, and third generation biofuels. First generation biofuels are primarily produced from feedstocks that are food crops, in another way the goal of second generation biofuel processes is to extend biofuel production capacity by incorporating residual biomass. The latest generation of biofuels research is now directing the attention to microorganisms-based advanced technologies, which are considered to be a viable alternative energy resource that is devoid of the major drawbacks associated with first and second-generation biofuels.[3–7]

The European Union, United States and Brazil in 2007 contributed 15%, 43% and 32%, respectively, of the total liquid biofuels produced in the world. In the European Union the main contribution was from biodiesel and in the United States and Brazil it was ethanol.[8] In the European Union the policy with indicative objectives regarding the use of renewable energy in general, and of biofuels in the transport sector in particular, has its goals specified in the Directive 2009/28/EC (ref. 9) and in the United States the mandate minimum usage requirements of

*Institute of Chemistry, University of Campinas – UNICAMP, P.O. Box 6154, 13083-970, Campinas, SP, Brazil. E-mail: julio@iqm.unicamp.br*

biofuels in the transport sector is described in the expanded Renewable Fuel Standard (RFS2).[10] In Brazil there is the mandated ethanol blend of 18–25% (v/v) in gasoline and 5% (v/v) of biodiesel in diesel oil (B5)[1,11] and currently a government study is being carried out with the objective of making it mandatory to have a 10% (v/v) biodiesel content in diesel fuel blend (B10) by 2020. Moreover, private companies initiatives and/or municipal ordinances have led to an increase in the amount of biofuels used in the transport sector, such as in urban transport bus fleets, and the experiences in the cities of São Paulo, Rio de Janeiro and Curitiba stand out. In São Paulo a municipal ordinance, started in 2009, determines that the total petroleum diesel fuel used by the urban transport bus fleet will be gradually replaced by biofuels or other renewable energy sources by 2018.

Different mixtures of petroleum diesel and biofuels are commonly used to make experimental diesel fuel blend with "drop-in" characteristics, which means that the experimental diesel fuel can be used without any change in diesel engines. Some of these experimental diesel fuel blends have been used in Brazil, such as diesel fuel blend (DFB) #1, which is a mixture of biodiesel and petroleum diesel; and diesel fuel blend (DFB) #2 which is a mixture of biodiesel, hydrocarbon renewable diesel and petroleum diesel.

The diesel fuel produced from renewable raw material can be classified as biodiesel (fatty acid alkil esters) or as renewable diesel (hydrocarbons).[12] Different types of renewable diesel based on a mixture of hydrocarbons or a single hydrocarbon molecule can be produced through processes such as biomass-to-liquid (BTL) Fischer–Tropsch synthesis or fermentation of sugars using engineered microorganisms followed by hydrogenation of the intermediate product, respectively.[7,12]

Currently (the first half of 2013) in the city of São Paulo approximately 10% (1500 buses) of the urban transport bus fleet uses DFB #1 or DFB #2 daily. Hydrocarbon renewable diesel and biodiesel have different costs and are more expensive, due to the higher production costs, than for petroleum diesel. Therefore, the use of these biofuels commonly becomes viable only with financial subsidy from the government.[3] Quality control to determine the relative amounts of each type of biofuel (hydrocarbon renewable diesel and/or biodiesel) and petroleum diesel in DFB #1 and DFB #2 is needed in order to assure the good operational performance of diesel engines; eventually to enable the calculation and negotiation of carbon credits[13] due to the reduction of GHG emission by combustion engines; and also to demonstrate to the government that the money spent on biofuels is a good use of financial subsidies passed to public transport private companies. It can be seen that there are financial justifications for the accurate identification and determination of the concentration of biofuels.

### 1.1 Determination of the content of biofuels in diesel fuel blends

The determination of the hydrocarbon renewable diesel concentration in mixtures with petroleum diesel can be performed using the American Society for Testing and Materials (ASTM) standard method D6866 (ref. 14) which measures the radiocarbon ($^{14}$C) content of a diesel fuel blend, that is directly related to the biofuel content. Such a standard method determines the bio-based content of a sample with a maximum total absolute error of 3%. Some methods based on a modification of this ASTM standard method have recently been reported[15,16] and use the same analytical techniques, which are relatively expensive and time-consuming (about 180 to 360 minutes per sample), leading to analytical results with an absolute error of 0.4% for mixtures up to 20% (v/v) of hydrocarbon renewable diesel.[15] The analytical techniques used in these methods do not permit the discrimination and quantification of two different biofuels in a mixture with petroleum diesel.

Commonly used methods for the determination of biodiesel and petroleum diesel blends use calibration models with partial least squares (PLS) regression[17] applied to mid infrared (MIR) spectroscopy data, such as the ASTM standard method D7371 (ref. 18) and Associação Brasileira de Normas Técnicas (ABNT) NBR 15568.[19] The ABNT NBR standard method limits the root mean square error of prediction (RMSEP) to 0.1% and 1% for the determination of mixtures in the analytical ranges of 0–8% (v/v) and of 8–30% (v/v) of biodiesel content in petroleum diesel blends, respectively. However, the use of near infrared (NIR) spectroscopy combined with chemometric methods such as PLS and support vector machines (SVM)[20,21] also have been used with success for biodiesel content determination in mixtures with petroleum diesel,[22] as well as for quality parameters determination of petroleum diesel[23–25] and its mixtures with biodiesel.[26]

The use of NIR spectroscopy for the quantitative determination of a ternary fuel blend with two different biofuels as well as two hydrocarbon-based fuels can be of great interest from the quality control point of view. This study present the development of an accurate, simple, fast and low cost method based on NIR spectroscopy data combined with chemometric methods to obtain multivariate calibration models suitable for the discrimination and quantification of two different biofuels (hydrocarbon renewable diesel and biodiesel) and petroleum diesel in ternary diesel fuel blends such as DFB #2.

## 2 Experimental

### 2.1 Materials

The hydrocarbon renewable diesel used in DFB #2 is a single iso-alkane compound, farnesane (2,6,10-trimethyldodecane), produced by the fermentation of sugarcane juice using a genetically modified microorganism (GMM) such as yeast, *e.g. S. cerevisiae*, leading to an olefin intermediate product which is then transformed *via* a hydrogenation process into diesel fuel molecules.[12,27] The biodiesel used in DFB #1 and DFB #2 is a mixture of fatty acid methyl esters (FAME) produced by the transesterification reaction of triglycerides from vegetable oil or animal fat with methanol, in the presence of a catalyst.[28] The soybean biodiesel is mainly a mixture of methyl esters derived from triglycerides with C16 to C18 fatty acid side chains with 0 to 3 unsaturated carbon bonds. The petroleum diesel is a mixture of mainly linear saturated hydrocarbons with C10 to C18 carbon chains, naphthenic hydrocarbons and aromatic hydrocarbons.

**Table 1** Reference quality parameters of farnesane, biodiesel and petroleum diesel

| | Renewable diesel (farnesane) | Biodiesel | Petroleum diesel (ULSD) | ASTM Standard method |
|---|---|---|---|---|
| Density, 20 °C/g ml$^{-1}$ | 0.773 | 0.850–0.900 | 0.820–0.850 | D4052 |
| Viscosity, 40 °C/mm$^2$ s$^{-1}$ | 2.3 | 3.0–6.0 | 2.0–4.5 | D445 |
| Cetane number, mín. | 58 | 48 | 48 | D6890 |
| CFPP,[a] max./°C | −48 | 19 | 0 | D6371 |
| Flash point, mín./°C | 101 | 100 | 38 | D93 |

[a] cold filter plugging point

The relative amounts of these different compounds in the petroleum diesel can vary due to the crude petroleum characteristics and the refining processes used.[29] Some reference quality parameters of hydrocarbon renewable diesel (farnesane), biodiesel and petroleum diesel are shown in Table 1.

Hydrocarbon renewable diesel, soybean biodiesel and petroleum diesel (ultra low sulfur diesel – ULSD) were supplied by the local distributor Petrobras Distribuidora S.A., from its Barueri, SP, Brazil facilities. The hydrocarbon renewable diesel was produced by Amyris Brasil S.A., from Campinas, SP, Brazil and has 93% (w/w) of farnesane. The soybean biodiesel is a mixture of many production batches from two manufacturers, Camera S.A., from Ijuí, RS, Brazil and BS Bios S.A., from Passo Fundo, RS, Brazil and has 97% (w/w) of FAME. The petroleum diesel was produced by Petrobras in its refineries in São Paulo state and the local distributor receives the production continuously in its facilities.

The transflectance spectra in the NIR region were measured using a Perkin Elmer Spectrum 100 MIR/NIR spectrometer with a halogen source and a deuterated triglycine sulphate (DTGS) detector. A Petri dish combined with an aluminum reflector with 0.5 mm pathlength was used as the transflectance cell.

## 2.2 Sample sets and experimental procedure

The analytical range was 1–100% (v/v) of hydrocarbon renewable diesel, 0–21% (v/v) of biodiesel, and 0–96% (v/v) of petroleum diesel. The experimental design included one calibration sample set and three validation sample sets.

The calibration set comprised 47 samples prepared according to the following experimental design: the hydrocarbon renewable diesel content in each sample increased by 1% in the range of 1–41% and increased by 10% in the range of 50–100%; the biodiesel content in each sample decreased by 0.5% in the range of 21–1% and in samples with increasing contents of hydrocarbon renewable diesel from 50% (and decreasing contents of petroleum diesel from 40%) the biodiesel content was kept constant at 10%; the petroleum diesel content in each sample decreased by 0.5% in the range of 78–58% and decreased by 10% in the range of 40–0%.

Validation set #1 comprised 24 samples and was prepared with hydrocarbon renewable diesel, biodiesel and petroleum diesel contents that differed from that of the calibration set, but were within the calibration analytical range. Validation set #2 comprised 9 samples, prepared with hydrocarbon renewable diesel content at three concentration levels, such as 10%, 40% and 70%, and for each of these hydrocarbon renewable diesel concentration levels there were three samples with different levels of biodiesel content, such as 5%, 10% and 15%. Petroleum diesel was used to complete the volume of each sample. Validation set #3 comprised 11 samples and mixtures with only hydrocarbon renewable diesel and petroleum diesel. The hydrocarbon renewable diesel content varied in the range of 4–95% and the petroleum diesel content varied in the range of 96–5%.

Fig. 1 illustrates the experimental design using a ternary diagram with the calibration and validation samples compositions throughout the analytical ranges. The aim of this experimental design was to include the current use of the diesel fuel blends in Brazil, such as diesel fuel blend (DFB) #1: 2% (v/v) to 20% (v/v) of biodiesel + 98% (v/v) to 80% (v/v) of petroleum diesel; and diesel fuel blend (DFB) #2: 5% (v/v) of biodiesel + 10% (v/v) to 30% (v/v) of hydrocarbon renewable diesel + 85%
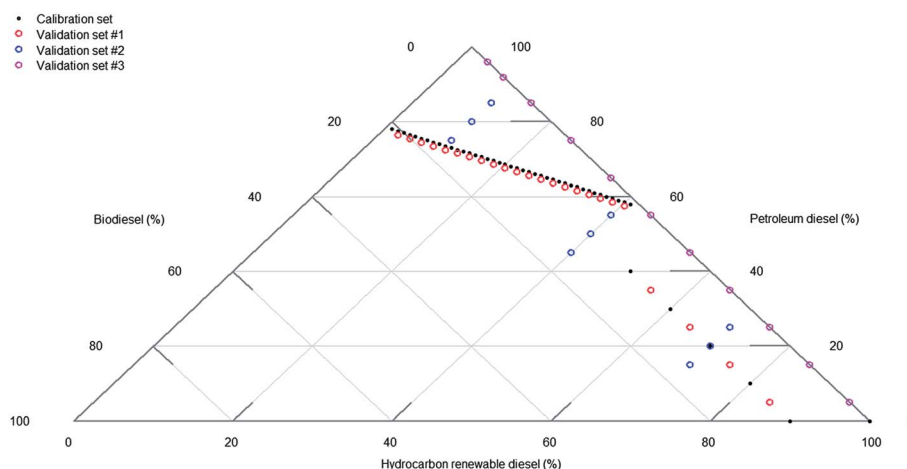


**Fig. 1** Experimental design of calibration and validation samples.

(v/v) to 65% (v/v) of petroleum diesel. Higher concentrations of the hydrocarbon renewable diesel could be used in mixtures with petroleum diesel and/or biodiesel, by extending the analytical range for this diesel fuel blend component. Moreover, the validation set #2 has samples with different relative contents of hydrocarbon renewable diesel and biodiesel in relation to the calibration samples, and the validation set #3 has samples without biodiesel in the mixtures, in order to confirm the prediction ability of the models.

The samples were prepared by mixing a total volume of 20 ml placed in dark glass bottles of 100 ml and all the sample analyses were run in random order and at constant temperature of 23 °C.

NIR spectra were obtained in the range of 3850–9000 cm$^{-1}$ for the 91 samples and each spectrum was obtained as an average of 32 scans with 4 cm$^{-1}$ resolution.

### 2.3  Data treatment and data analyses methods

Different data preprocessings were carried out to verify which provides the best model. The preprocessings tested were: baseline correction and mean centering; standard normal variate (SNV); and second derivative. A blocked cross-validation of the calibration set was used for model development.

For the development of PLS and SVM models, PLS toolbox version 4.0 (ref. 30) and LIBSVM package version 2.88 (ref. 31) were used, respectively. All the programs are ready for Matlab version 7.7 from Mathworks. The SVM calibration models were developed using the algorithm $\nu$-support vector regression ($\nu$-SVR).[32,33] For the $\nu$-SVR models development different kernel functions[25,34] such as radial basis function (RBF) and linear function were tested and the data set was previously scaled between 0 and 1. The LIBSVM default value of $\gamma$ parameter for the RBF kernel ($\gamma = 1/k$, where $k$ means the number of variables in calibration data set) was used. The $\nu$-SVR parameters $C$ and $\nu$ were selected in the ranges of $0$–$10^4$ and $10^{-4}$–$1$, respectively.

In order to gain further insight into the accuracy of the developed calibration models, linear regression analyses of prepared concentrations values *versus* PLS and $\nu$-SVR predicted concentrations values for the three validation sets were applied. The estimated intercept ($b$) and slope ($a$) were compared with their ideal values of 0 and 1, respectively, using the elliptical joint confidence region (EJCR) test, in this case by using an ordinary least squares fitting of the prepared concentration values *versus* predicted concentration values for each model. The boundary of the ellipse is determined by the magnitude of the experimental errors and by the degrees of confidence chosen, and is described by the following equation:

$$n(b - \beta)^2 + 2(\sum y_i)(b - \beta)(a - \alpha) + \sum y_i^2 (a - \alpha)^2 = 2s^2 F_{2,d} \quad (1)$$

where $n$ is the number of data points, $y_i$ are the prepared concentrations values, $s^2$ the regression variance and $F_{2,d}$ is the critical $F$ value with 2 and $d = n - 2$ degrees of freedom at a given confidence level. In this work the 95% confidence level was used. The centre of the ellipse is $(b,a)$ and any point $(\beta,\alpha)$ that lies inside the EJCR is compatible with the data at the chosen confidence level. In order to check the constant

(translational) or proportional (rotational) bias, the values $\beta = 0$ and $\alpha = 1$, respectively, are compared with the estimates $b$ and $a$ using EJCR. If the point $(0,1)$ lies inside the EJCR, then biases are not present.[35,36]

## 3  Results and discussion

The main qualitative and quantitative differences between the molecules of biofuels in the diesel fuel blend under study in relation to the hydrocarbon mixture of the petroleum diesel are as follows: farnesane has a relative higher content of methyl groups of branched carbon chain alkane; the biodiesel has a carbonyl group and a terminal methyl group near the carbonyl group, moreover the fatty acid side chains of methyl esters have unsaturated carbon bonds and a relative higher content of methylene groups.

The NIR spectral region used of 3850–9000 cm$^{-1}$ has the occurrence of combination bands, first and second overtones of vibrational modes of C–H bond in methyl and methylene groups and C=C bond of unsaturated compounds.[37] We want to emphasize some bands related to vibrational modes of these groups such as the C–H bond stretching of the methyl group in branched alkanes near 5905 cm$^{-1}$, 5872 cm$^{-1}$, 4400 cm$^{-1}$ and 4100 cm$^{-1}$ and the C–H bond stretching of the methylene group in alkanes near 5800 cm$^{-1}$, 5680 cm$^{-1}$ and 4336 cm$^{-1}$.[37] There is also a combination band of C–H bond stretching and C=O bond stretching near 4650 cm$^{-1}$.[37] Moreover the difference in NIR spectra in the region of 4425 cm$^{-1}$ and 6005 cm$^{-1}$, probably related to stretching of the terminal methyl group near the carbonyl group, where methyl esters have peaks while triglycerides exhibit only shoulders, provides the biodiesel quantification in a selective manner, as demonstrated in previous studies,[22,38] without interference from the possible presence of vegetable oil in the mixture. Fig. 2(a) illustrates the difference between the spectrum of hydrocarbon renewable diesel, biodiesel and petroleum diesel using spectra with SNV preprocessing in the spectral range of 3900–6150 cm$^{-1}$ where there is the occurrence of combination bands and first overtone bands.

The spectral region used allows the calibration of hydrocarbon renewable diesel, biodiesel and petroleum diesel contents in diesel fuel blends due to its different compositions in terms of compounds with linear, branched, cyclic and aromatic carbon chains and due to the presence of a carbonyl group and a terminal methyl group near the carbonyl group in fatty acid methyl esters. In this manner, with relative variation in the amounts of each diesel fuel blend component, a relative variation of methyl and methylene groups of linear, branched, cyclic and aromatic carbon chains, and of carbonyl group of esters and terminal methyl group near the carbonyl group also occurs, and the respective intensities of the absorbance signals in the NIR spectrum enable the quantitative determination of each diesel fuel blend component.

### 3.1  PLS and $\nu$-SVR calibration models

For the development of the PLS models instead of the use of the full spectral range of 3850–9000 cm$^{-1}$, a variable selection
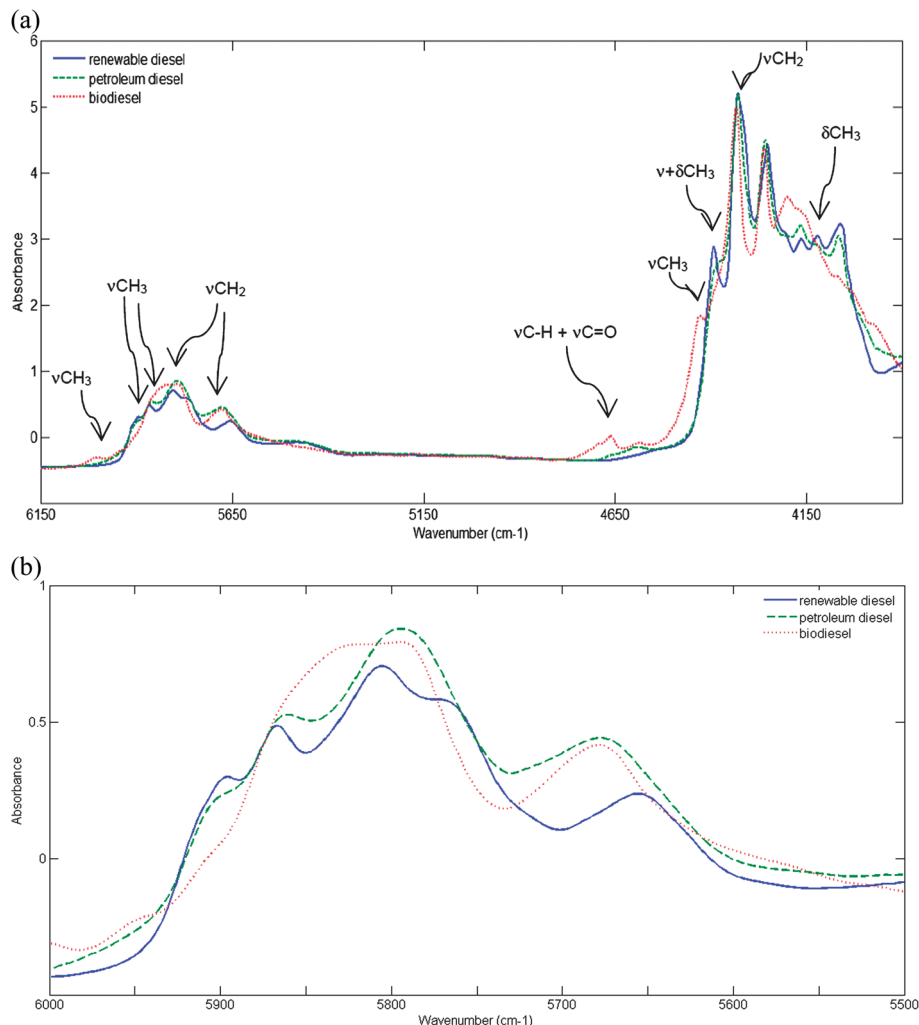
**Fig. 2** Characteristic NIR spectra of hydrocarbon renewable diesel (farnesane), biodiesel and petroleum diesel after SNV preprocessing. (a) spectral region of 3900–6150 cm$^{-1}$ and (b) spectral region of 5500–6000 cm$^{-1}$.

was performed by interval partial least squares (iPLS)[39] selecting the spectral region that provides the PLS model with the lowest value of root mean square error of cross validation (RMSECV). The iPLS variable selection was tested using 51, 17 and 10 intervals (which correspond to blocks with 100, 300 and 500 variables, respectively) and different data preprocessings. The use of iPLS with 10 intervals and SNV preprocessed data allows the spectral region of 5500–6000 cm$^{-1}$ to be selected, which provides a PLS model with a suitable root mean square error of calibration (RMSEC). Fig. 2(b) illustrate in detail the selected spectral region of the individual spectra of hydrocarbon renewable diesel, biodiesel and petroleum diesel and Fig. 3 shows the calibration set spectra with SNV preprocessed data. In the SNV transformation each spectrum is centered and then scaled by dividing by its standard deviation, which corrects for both baseline shift and global intensity variations. The SNV-corrected spectra contain positive and negative values.

Very similar results were obtained using the same spectral region and data preprocessing by using the PLS calibration model for simultaneous determination or individual PLS calibration models for the determination of each diesel fuel blend component, for this reason we only mention the results of the PLS simultaneous determination here. The best PLS model use four latent variables, which explains 99.9% of the data variance in the Y-block. Table 2 shows the PLS model results and Fig. 4 illustrates the PLS model absolute residual distribution for calibration and validation sample sets for hydrocarbon renewable diesel, biodiesel and petroleum diesel predictions. It appears that the PLS model provides a good fit throughout the analytical ranges with a constant variance or homoscedasticity of the absolute residual values and low residues for calibration and validation sample sets. In PLS regression modeling assessment of the importance of the wavelength region for the multivariate calibration model can be performed based on PLS loadings and regression coefficients. The most important variables can be identified by large PLS loadings and regression coefficients.[17] Fig. 5 shows the loadings of the first and second latent variables (LV), which explain 98.66% and 1.30% of variance in the X-block, respectively. We can compare
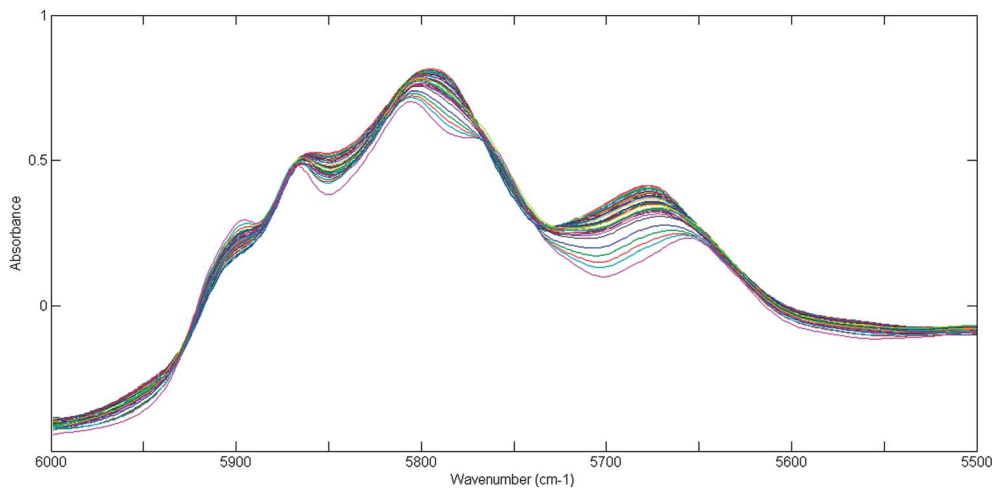
**Fig. 3** NIR spectra of the 47 calibration samples within the spectral region of 5500–6000 cm$^{-1}$ after SNV preprocessing.

**Table 2** PLS and ν-SVR models results

| Model | Diesel fuel blend component | RMSEC (%) | RMSEP$^a$ (%) | $R^2$ |
|-------|------------------------------|-----------|--------------|-------|
| PLS | Hydrocarbon renewable diesel | 0.19 | 0.25 | 0.9999 |
| | Biodiesel | 0.20 | 0.19 | 0.9988 |
| | Petroleum diesel | 0.31 | 0.38 | 0.9997 |
| ν-SVR | Hydrocarbon renewable diesel | 0.16 | 0.26 | 0.9965 |
| | Biodiesel | 0.21 | 0.19 | 0.9999 |
| | Petroleum diesel | 0.22 | 0.32 | 0.9972 |

$^a$ Considering samples of validation sets #1, #2 and #3.

these results with pure spectra (Fig. 2(b)) and recognize that the whole spectral range is important. Because the three diesel fuel blend components presents the NIR spectra with over-lapping band profiles and because of its simultaneous determination, the PLS loadings profiles does not enable discrimination of individually the most important spectral regions for the calibration of each diesel fuel blend component. Fig. 6 shows the PLS regression coefficients for renewable diesel, biodiesel and petroleum diesel. The most important wavelength regions for the calibration model of the different components of diesel fuel blend are as follows: regression coefficients for renewable diesel show distinct maximums near 5680 (−), 5750, 5800 (−) and 5905 cm$^{-1}$; regression coefficients for biodiesel show distinct maximums near 5750 (−), 5840, 5900 (−) and 6000 cm$^{-1}$; regression coefficients for petroleum diesel show distinct maximums near 5600 (−), 5680, 5800, 5840 (−) and 5872 (−) cm$^{-1}$. As discussed earlier, these wavelengths are assigned to spectral regions of C–H bond stretching of methyl and methylene groups, and for this type of mixture some specific spectral regions are important for regression purposes of more than just one diesel fuel component, but the use of the selected spectral range gives an adequate calibration model due to some relative variations in NIR spectra profiles.

For the development of ν-SVR models the spectral region of 5500–6000 cm$^{-1}$ was tested, and the spectral region of 4250–4600 cm$^{-1}$, which includes the spectral region used in a previous study[22] of biodiesel content determination in petroleum diesel fuel blends using ν-SVR and linear kernel function, was also tested. The best results were obtained using the spectral region of 5500–6000 cm$^{-1}$, the SNV preprocessed data and the linear kernel function. The ν-SVR parameters, $C$ and $ν$, and the number of support vectors for each calibration model were: 10, 0.0035 and 22; 2.5, 0.0050 and 10; and 15, 0.0030 and 21, for hydrocarbon renewable diesel, biodiesel and petroleum diesel calibration models, respectively. The better results provided by the use of linear kernel function instead of RBF kernel function were in agreement with the previous results obtained[22] for the determination of biodiesel contents in petroleum diesel fuel blends, moreover, the use of ν-SVR and linear kernel function provided better results than PLS for diesel fuel parameters determination,[24] although in such a case the use of the RBF kernel function provided the best results due to some relationship particularities of the studied problem that suggest some degree of nonlinearity. Table 2 shows the results for the ν-SVR models and Fig. 7 illustrates the ν-SVR models absolute residual distribution for calibration and validation sample sets for hydrocarbon renewable diesel, biodiesel and petroleum diesel predictions. It is possible to see the good fit of ν-SVR models throughout the analytical ranges, especially for hydrocarbon renewable diesel and biodiesel, with homoscedasticity of the absolute residual values and low residues for calibration and validation sample sets.

The statistical significance of the regression models were assessed through the analysis of variance (ANOVA).[40] The ratio of the mean square due to regression (MQ$_R$) and the mean square due to residuals (MQ$_r$) was calculated for each regression model in order to know if the linear regression models are statistically significant at the 95% confidence level. Considering the critical value of $F_{1,45} = 4.08$ and the calculated value of MQ$_R$/MQ$_r$ > 2000 for each regression model, there is statistical evidence for the existence of a linear relationship between the
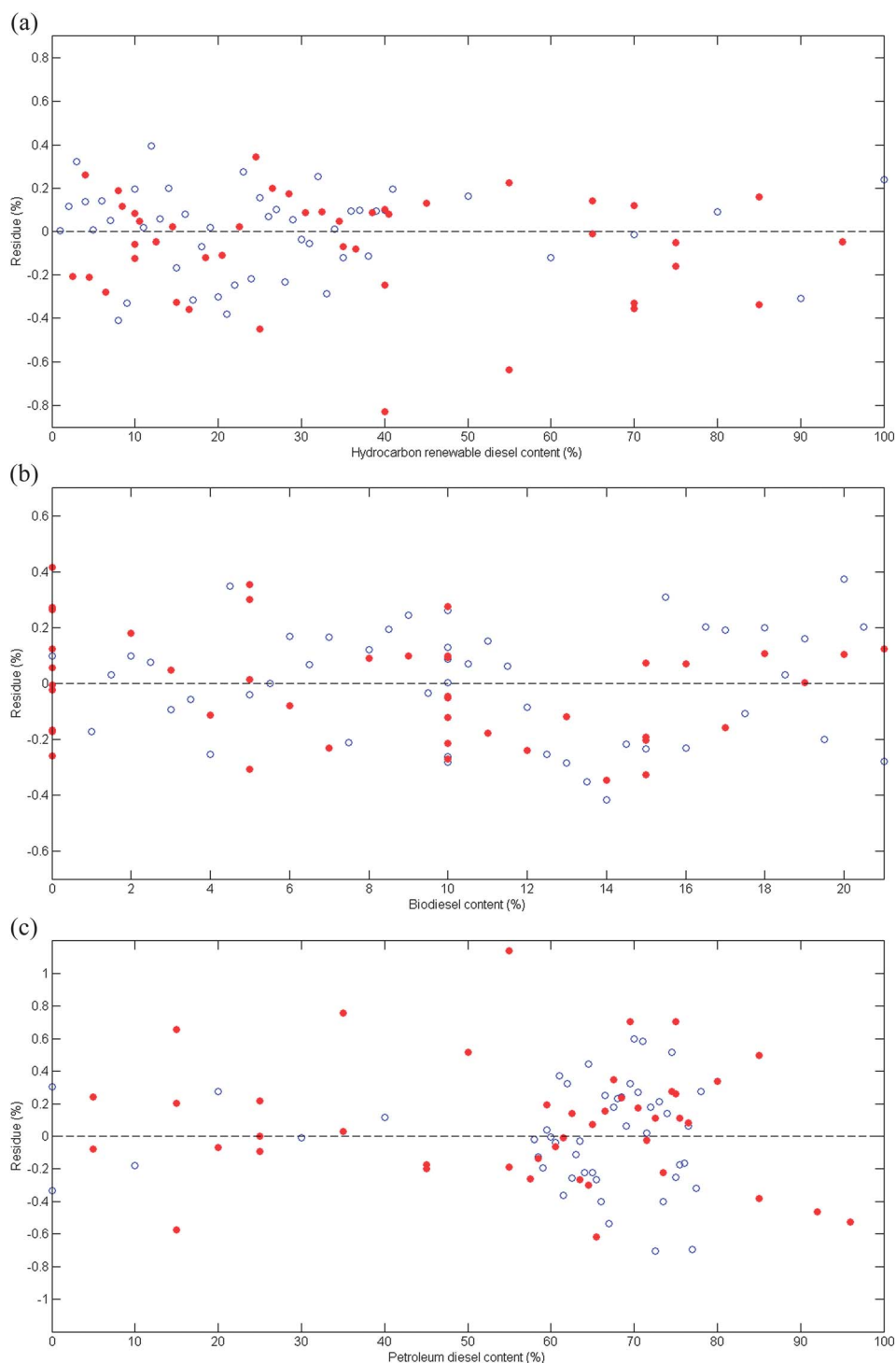
**Fig. 4** Absolute residual distribution of PLS model predictions for the calibration (O) and validation (●) samples for (a) hydrocarbon renewable diesel, (b) biodiesel and (c) petroleum diesel.

analytical signal and concentration. The values of the coefficient of determination $(R^2)$ near to 1 also indicate the good fit of linear regression model for each diesel fuel blend component, as is shown Table 2.

The obtained values of root mean square error of prediction (RMSEP) for PLS and ν-SVR calibration models for hydrocarbon renewable diesel and biodiesel are lower than the maximum

error suggested by the ASTM D6866 standard method and are lower or similar to the absolute errors reported in recent studies[15,16] using an analytical procedure based on a modification of the ASTM D6866 standard method. Moreover, the obtained values of RMSEP for PLS and ν-SVR calibration models for biodiesel are in agreement with the requirement of the ABNT NBR 15568 standard method. The *F*-test was used to
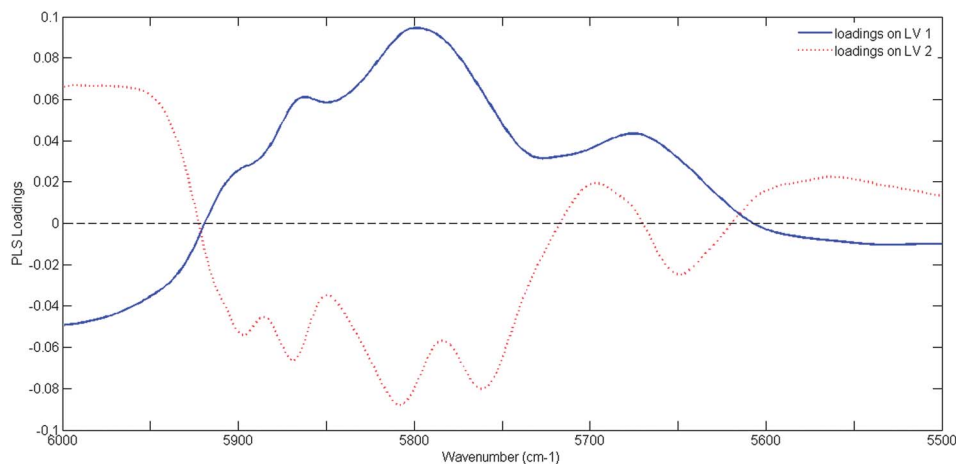
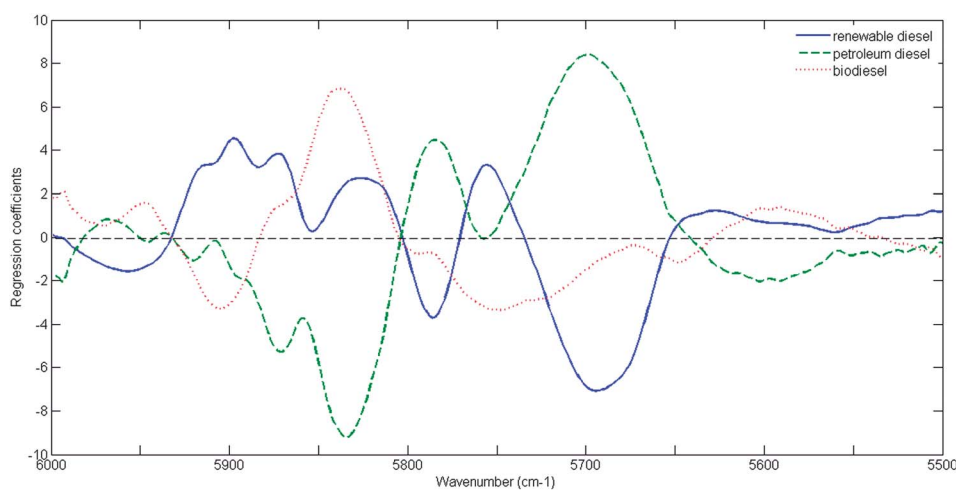**Fig. 5** PLS loadings of the first and second latent variables (LV).



**Fig. 6** PLS regression coefficients for renewable diesel, biodiesel and petroleum diesel.

statistically compare the RMSEP values obtained with PLS and ν-SVR models using the ratio of the squared RMSEP being compared and the critical value of $F_{43,43} = 1.69$ at the 95% confidence level. It was found that both PLS and ν-SVR models provide similar average errors of prediction for each diesel fuel blend component, but this comparison approach does not provide information on the possible occurrence of bias for the predicted results of the models.

The EJCR test applied to the PLS and ν-SVR calibration models results for hydrocarbon renewable diesel, biodiesel and petroleum diesel are shown in Fig. 8. There are no significant differences between the prepared concentration values and the predicted concentration values for PLS and ν-SVR models for the three validation sets and there is no evidence of bias with the 95% confidence level, except for the ν-SVR petroleum diesel calibration model which has a theoretical point (0,1) outside the boundary of joint confidence region, probably due to a positive tendency of the residual values of the upper extreme values of the analytical range. It is important to consider that due to the experimental design of

validation sets #2 and #3 some samples extrapolate the calibration analytical range of petroleum diesel in the case of contents higher than 78% of this diesel fuel blend component. Previous studies[41,42] demonstrate the good generalization ability of PLS and SVR (using nonlinear kernel-based models) algorithms applied to spectroscopy data of crude petroleum, petroleum derivative fuels and biodiesel, for both interpolation and extrapolation of the calibration analytical range, but in the present work the PLS model provides the best prediction results.

The choice of a chemometric method for a real-world application must take into consideration more than just the accuracy of the predictions. Other relevant issues are the ease of model development, implementation in routine analyses and interpretability. In this manner, we need to emphasize that ν-SVR modeling requires an adequate choice of kernel function and optimization of parameters, which can be more time-consuming than selecting the adequate number of latent variables in PLS modeling. Moreover, because ν-SVR employs a kernel function, it has the drawback that the information
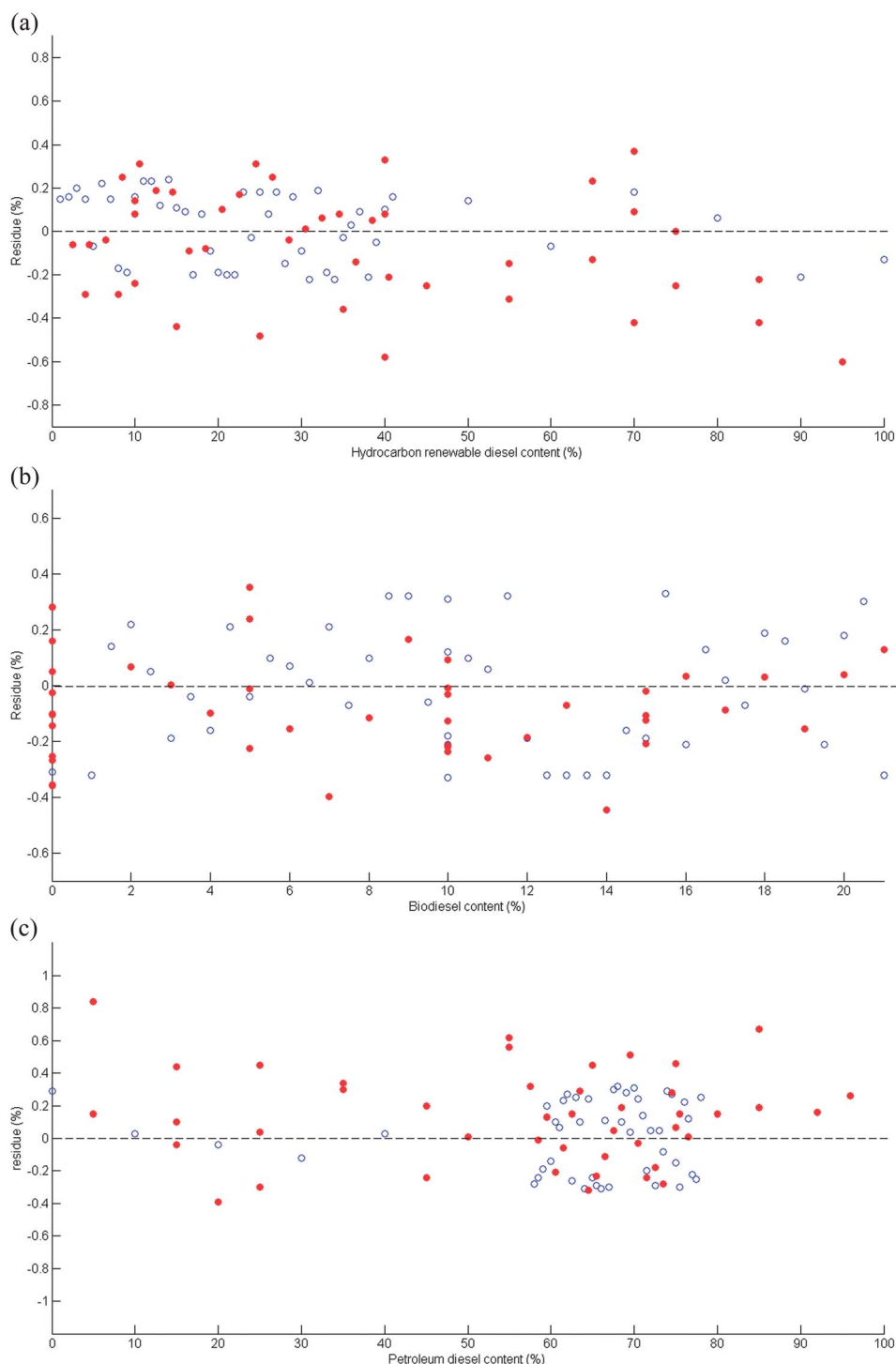
**Fig. 7** Absolute residual distribution of ν-SVR models predictions for the calibration (O) and validation (●) samples for (a) hydrocarbon renewable diesel, (b) biodiesel and (c) petroleum diesel.

about the original input variables is lost and direct interpretation of the final ν-SVR model in relation to the input variables involved is not possible.[43]

The main attraction of ν-SVR is the possibility of adequately treating nonlinear relationships (using a nonlinear kernel function such as RBF) with high generalization performance, this has proved to be very interesting for example in petroleum

refinery applications.[23,24] But in this study statistical evidence demonstrated the good fit of linear models and that the generalization performance of PLS was adequate. Although both ν-SVR based on linear kernel function and PLS models provide good predictions, the simpler model development for simultaneous analyses and easier interpretability of PLS model is remarkable in this work.
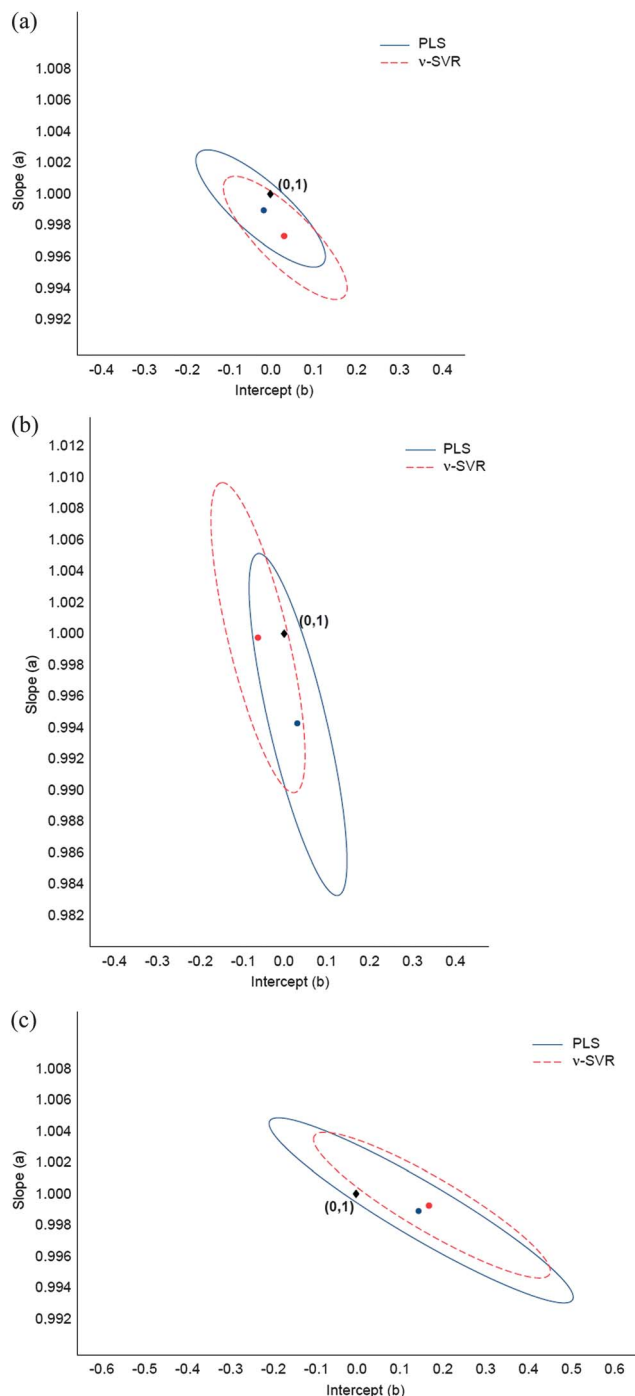
**Fig. 8** Elliptical joint confidence regions for the intercept and slope corresponding to regressions of prepared concentration values *versus* PLS and ν-SVR model predicted concentration values for (a) hydrocarbon renewable diesel, (b) biodiesel and (c) petroleum diesel. The estimated (*b*,*a*) for PLS (●) and ν-SVR (●) models.

## 4 Conclusion

This study demonstrated the development of a multivariate calibration method based on NIR spectroscopy combined with chemometric methods to discriminate and quantify two different biofuels as well as two different hydrocarbon-based fuels in a ternary diesel fuel blend. The spectral region of 5500–6000 cm$^{-1}$

provides suitable calibration models for quantitative determination of hydrocarbon renewable diesel (farnesane), biodiesel and petroleum diesel in diesel fuel blends. Similar prediction results were obtained using PLS and ν-SVR calibration models for both biofuels. Nevertheless, the simpler PLS calibration model enables simultaneous determination with high accuracy and without the occurrence of bias for validation sample set predictions for the three diesel fuel blend components, for this reason it is considered to be the best choice for this analytical problem.

The RMSEP values of PLS and ν-SVR calibration models for hydrocarbon renewable diesel, biodiesel and petroleum diesel are less than 0.4% and are in agreement with the values required by the ASTM D6866 and ABNT NBR 15568 standard methods, for hydrocarbon renewable diesel and biodiesel determination, respectively. The developed calibration models provide accurate quantitative determination of both biofuels in the diesel fuel blend in a simpler, faster and cheaper manner, in relation to the ASTM D6866 standard method. Moreover, the advantage of simultaneous determination using the PLS model, and the determination of biodiesel in a selective manner, without possible interference from vegetable oil and in a broad analytical range must be highlighted.

Although this method has been developed for the determination of hydrocarbon renewable diesel based on farnesane, further studies can lead to similar methodology for determination of other types of hydrocarbon renewable diesel in petroleum diesel fuel blends.

## Acknowledgements

## References

1 J. L. Sawin, *REN 21 – Renewable energy policy network for the 21st century – Renewables 2012 – Global status report*, Paris, France, 2012, http://www.ren21.net.

2 *A comprehensive analysis of biodiesel impacts on exhaust emissions – Draft technical report*, US Environmental Protection Agency – EPA, 2002.

3 G. Fiorese, M. Catenacci, E. Verdolini and V. Bosetti, *Energy Policy*, 2013, **56**, 293–311.

4 P. S. Nigam and A. Singh, *Prog. Energy Combust. Sci.*, 2011, **37**, 52–68.

5 S. N. Naik, V. V. Goud, P. K. Rout and A. K. Dalai, *Renewable Sustainable Energy Rev.*, 2010, **14**, 578–597.

6 J. J. Cheng and G. R. Timilsina, *Renewable Energy*, 2011, **36**, 3541–3549.

7 T. G. Smagala, E. Christensen, K. M. Christison, R. E. Mohler, E. Gjersing and R. L. McCormick, *Energy Fuels*, 2013, **27**, 237–246.

8 R. W. Howarth and S. Bringezu, *UNESCO-SCOPE-UNEP Policy Briefs*, 2009, vol. 9, pp. 1–7.

9 Directive 2009/28/EC of the European Parliament and of the Council, of 23 April 2009, on the promotion of the use of

energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC, Brussels, 2009.

10 R. Schnepf and B. D. Yacobucci, *Renewable Fuel Standard (RFS): Overview and issues, Congressional Research Service*, 2012.

11 J. O. B. Carioca, J. J. H. Filho, M. R. L. V. Leal and F. S. Macambira, *Biotechnol. Adv.*, 2009, **27**, 1043–1050.

12 P. J. Westfall and T. S. Gardner, *Curr. Opin. Biotechnol.*, 2011, **22**, 344–350.

13 J. A. Mathews, *Energy Policy*, 2008, **36**, 3633–3639.

14 ASTM Standard D6866, 2012, *Standard test methods for determining the biobased content of solid, liquid, and gaseous samples using radiocarbon analysis*, ASTM International, West Conshohocken, PA, 2012, DOI: 10.1520/D6866-12, http://www.astm.org.

15 G. A. Norton and M. X. Woodruff, *J. Am. Oil Chem. Soc.*, 2011, **89**, 797–803.

16 M. Oinonen, H. Hakanpaa-Laitinen, K. Hamalainen, A. Kaskela and H. Jungner, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2010, **268**, 1117–1119.

17 S. Wold, M. Sjostrom and L. Eriksson, *Chemom. Intell. Lab. Syst.*, 2001, **58**, 109–130.

18 ASTM Standard D7371, *Standard test method for determination of biodiesel (fatty acid methyl esters) content in diesel fuel oil using mid infrared spectroscopy (FTIR-ATR-PLS method)*, ASTM International, West Conshohocken, PA, 2007, DOI: 10.1520/D7371-07, http://www.astm.org.

19 Norma brasileira ABNT NBR 15568, *Biodiesel – determinação do teor de biodiesel em óleo diesel por espectroscopia na região do infravermelho médio*, Associação Brasileira de Normas Técnicas, Rio de Janeiro, RJ, 2008, http://www.abnt.org.br.

20 A. J. Smola and B. Scholkopf, *Stat. Comput.*, 2004, **14**, 199–222.

21 R. G. Brereton and G. R. Lloyd, *Analyst*, 2010, **135**, 230–267.

22 J. C. L. Alves and R. J. Poppi, *Talanta*, 2013, **104**, 155–161.

23 J. C. L. Alves, C. B. Henriques and R. J. Poppi, *Fuel*, 2012, **97**, 710–717.

24 J. C. L. Alves and R. J. Poppi, *J. Near Infrared Spectrosc.*, 2012, **20**, 419–425.

25 H. Li, Y. Liang and Q. Xu, *Chemom. Intell. Lab. Syst.*, 2009, **95**, 188–198.

26 L. F. B. Lira, F. V. C. Vasconcelos, C. F. Pereira, A. P. S. Paim, L. Stragevitch and M. F. Pimentel, *Fuel*, 2010, **89**, 405–409.

27 M. A. Rude and A. Schirmer, *Curr. Opin. Microbiol.*, 2009, **12**, 274–281.

28 J. M. Marchetti, *Process Saf. Environ. Prot.*, 2012, **90**, 157–163.

29 C. Song, Introduction to chemistry of diesel fuels, in *Chemistry of diesel fuels*, ed. C. Song, C. S. Hsu and I. Mochida, Taylor and Francis, New York, 2000.

30 B. M. Wise, N. B. Gallagher, R. Bro, J. M. Shaver, W. Windig and R. S. Koch, *PLS Toolbox version 4.0 for use with Matlab*, 2006, Eigenvector research, Inc., Wenatchee, WA, USA, http://www.eigenvector.com.

31 C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, http://www.csie.ntu.edu.tw/~cjlin/libsvm.

32 B. Scholkopf, A. J. Smola, R. C. Williamson and P. L. Bartlett, *Neural Comput.*, 2000, **12**, 1207–1245.

33 A. Chalimourda, B. Scholkopf and A. J. Smola, *Neural Networks*, 2004, **17**, 127–141.

34 A. J. Smola, P. Bartlett, B. Scholkopf and D. Schuurmans, Introduction to large margin classifiers, in *Advances in Large Margin Classifiers*, ed. A. J. Smola, P. Bartlett, B. Scholkopf and D. Schuurmans, MIT Press, Cambridge, MA, 2000.

35 J. Mandel and F. J. Linning, *Anal. Chem.*, 1957, **29**, 743–749.

36 A. G. Gonzalez, M. A. Herrador and A. G. Asuero, *Talanta*, 1999, **48**, 729–736.

37 J. Workman Jr and L. Weyer, *Practical guide to interpretative near infrared spectroscopy*, CRC Press, Boca Raton, 2008.

38 G. Knothe, *J. Am. Oil Chem. Soc.*, 2001, **78**, 1025–1028.

39 C. M. Andersen and R. Bro, *J. Chemom.*, 2010, **24**, 728–737.

40 M. C. Ortiz, M. S. Sánchez and L. A. Sarabia, Quality of analytical measurements: univariate regression, in *Comprehensive chemometrics. Chemical and biochemical data analysis*, ed. S. D. Brown, R. Tauler and B. Walczak, Elsevier, 2009.

41 R. M. Balabin and S. V. Smirnov, *Analyst*, 2012, **137**, 1604–1610.

42 C. B. Henriques, J. C. L. Alves, R. J. Poppi, R. Maciel Filho and M. I. M. S. Bueno, *Energy Fuels*, 2013, **27**, 3014–3021.

43 B. Ustun, W. J. Melssen and L. M. C. Buydens, *Anal. Chim. Acta*, 2007, **595**, 299–309.