

26 **Abstract**

27 Adulteration is a recurrent issue found in fuel screening. Commercial diesel contamination by kerosene is highly
28 difficult to be detected via physicochemical methods applied in market. Although the contamination may affect
29 diesel quality and storage stability, there is a lack of efficient methodologies for this evaluation. This paper assessed
30 the use of IR spectroscopies (MIR and NIR) coupled with partial least squares (PLS) regression, support vectors
31 machine regression (SVR) and multivariate curve resolution with alternating least squares (MCR-ALS) calibration
32 models for quantifying and identifying the presence of kerosene adulterant in commercial diesel. Moreover,
33 principal component analysis (PCA), successive projections algorithm (SPA) and genetic algorithm (GA) tools
34 coupled to linear discriminant analysis were used to observe the degradation behavior of 60 samples of pure and
35 kerosene-added diesel fuel in different concentrations over 60 days of storage. Physicochemical properties of
36 commercial diesel with 15% kerosene remained within conformity with Brazilian screening specifications; in
37 addition, specified tests were not able to identify changes in the blends' performance over time. By using
38 multivariate classification, the samples of pure and contaminated fuel were accurately classified by aging level into
39 two well-defined groups, and some spectral features related to fuel degradation products were detected. PLS and
40 SVR were accurate to quantify kerosene in the 2.5–40% (v/v) range, reaching RMSEC<2.59% and RMSEP<5.56%,
41 with high correlation between real and predicted concentrations. MCR-ALS with correlation constraint was able to
42 identify and recover the spectral profile of commercial diesel and kerosene adulterant from the IR spectra of
43 contaminated blends.

44 **Keywords:** Diesel fuel, Adulteration, Kerosene, Multivariate analysis, Storage stability.

45
46
47
48
49
50
51
52
53

54 **Introduction**

55 An increase in energy and fuel consumption worldwide has encouraged researchers to study new energy sources and
56 look for the best ways to use them. Nowadays, the most important problem faced in the fuel screening field is
57 adulteration, which occurs during the route between the refinery and gas stations to provide illegal profit to
58 scammers [1,2]. In Brazil, gasoline adulteration is currently well controlled in the Fuel Quality Monitoring Program
59 (PMQC) founded by the National Agency of Petroleum, Natural Gas and Biofuels (ANP), using specific analysis of
60 markers added to solvents that can be used as contaminants [3]; however, there is no specific methodology for
61 detecting adulterants in diesel, which is the most consumed fuel in the country [4,5].

62 Biodiesel, kerosene and vegetable oils are the main adulterants of diesel reported in literature due to its
63 affordability and lower cost in comparison to the original fuel [6]. Biodiesel is a renewable fuel composed of esters
64 that is blended into Brazilian diesel (10% v/v) due to ANP requirements for reducing the emission of harmful gases,
65 but scammers have illegally added a greater amount of biodiesel due to its cheaper production [7]. Kerosene solvent
66 is a cheap petroleum distillate that has similar hydrocarbon composition to diesel and is widely used for adulteration,
67 making it practically impossible to detect this contaminant via physicochemical property tests and other univariate
68 methods [8].

69 Fuel contamination can cause many problems to burning and storage quality, with the latter being directly
70 associated with the oxidative stability of fuel and signifies how much they resist degradation processes. Diesel and
71 mainly biodiesel components are susceptible to oxidation and hydrolysis reactions over time; thus, the composition
72 of the fuel blend changes over time and the presence of an uncontrolled substance used for adulteration can exert
73 some influence on the process, affecting its quality due to aging [9]. Despite this, ANP specifications do not regulate
74 tests to observe the degradation level of commercial diesel over time due to the lack of a methodology that performs
75 efficient quality screening of this parameter during fuel storage [5]. Therefore, mathematical tools provided by
76 chemometrics enable analyzing multivariate results generated by simple techniques such as infrared (IR)
77 spectroscopy in a way that statistical methods for univariate systems may be inadequate, with accurate, fast and
78 detailed responses [10,11].

79 Principal component analysis (PCA), successive projections algorithm (SPA) and genetic algorithm (GA) are
80 techniques that promote intelligent experimental data reduction. PCA reduces data to principal components, while
81 SPA and GA reduce it to selected variables. This procedure can improve the potential of the supervised linear

82 discriminant analysis (LDA) for discriminating the samples in their respective classes due to the lower data
83 complexity. The combinations PCA-LDA, SPA-LDA and GA-LDA are often used in combination for a wide range
84 of applications [12-15], but their potential is not widely explored for screening diesel quality [16].

85 In case of quantifying kerosene into commercial diesel fuel, the use of a calibration model capable to deal
86 efficiently with non-linear relationships and high dimensional input vectors as support vectors machine regression
87 (SVR) is crucial, since the widely used partial least squares (PLS) has limited performance with complex systems
88 [17]. A promising tool for analyzing fuel adulteration is the multivariate curve resolution with alternating least
89 squares (MCR-ALS). This technique stands out due to its capability to quantify and identify the analyte in the
90 presence of interferences in samples, since these interferences are presented in the calibration samples [18]. In
91 addition, this technique presents some advantages in relation to PLS, such as the smaller number of samples needed
92 and the capacity to quantify and identify interferences (adulterants) in samples without previous knowledge of them,
93 which can be called “second order advantage” [19].

94 This paper evaluates the efficiency of multivariate tools to solve important issues in diesel screening using NIR
95 and MIR spectroscopies. PLS, SVR and MCR-ALS were applied to quantify and identify the presence of kerosene
96 adulterant in commercial diesel. PCA-LDA, SPA-LDA and GA-LDA models were used to classify and observe the
97 degradation behavior of pure and kerosene-added diesel fuel samples in different concentrations over 60 days of
98 storage. Oxidative stability analysis were performed by PetroOxy accelerated oxidation method to compare with the
99 classification results, in addition to atmospheric distillation, crystallization, specific mass and viscosity tests
100 specified by ANP.

101

102 **Materials and Methods**

103 **Sample Preparation**

104 Diesel S10B10 samples with 10% (v/v) of Brazilian commercial biodiesel (soybean oil biodiesel) and 10 mg kg⁻¹
105 sulfur were kindly donated by Clara Camarão Potiguar Refinery (RPCC) and mixed up to different concentrations of
106 commercial kerosene solvent at the Energetic Technology Laboratory (LTEN) for simulating the adulteration
107 process. The samples were divided in two datasets. In the first one, which involves classification models, were used
108 concentrations in the range 5 to 25% (v/v), with increments of 2.5% (v/v) and precision of 0.05%. Each blend was
109 divided into six parts of the same volume and placed in six different amber flasks, making a total of sixty samples.

110 All flasks were encoded, sealed and stored in a closed box at room temperature for sixty days, along with samples of
 111 pure S10B10. Six analysis periods were defined for monitoring physicochemical properties and MIR/NIR features
 112 of the mixtures during storage, namely days 0 (beginning of storage), 7, 15, 30, 45 and 60.

113 For the second dataset, used for adulterant quantification with regression methods such as PLS, SVR and MCR-
 114 ALS were produced 16 samples ranging from 2.5 to 40% (v/v). MIR and NIR spectra were recorded for modeling.

115

116 Physicochemical Analysis

117 All samples were submitted to MIR and NIR analysis on monitoring days, but only the pure S10B10 and 15%
 118 kerosene (Q0 and Q15) samples had some of their physicochemical properties evaluated in order to observe possible
 119 changes in composition and properties as a result of the added kerosene and/or storage time. Table 1 shows the
 120 reference methods of specified ANP physicochemical analysis, in addition to oxidative stability PetroOXY test,
 121 employed for monitoring samples Q0 and Q15 following regulations determined by ASTM International. In addition,
 122 some properties inferred for diesel S10B10 and kerosene solvent are depicted.

123

124 **Table 1.** Physicochemical analysis and ASTM standards used in this work.

Analysis	Diesel S10B10	Commercial kerosene solvent	Standard [20-23]
PetroOXY induction period (IP) (min)	60.0	-	ASTM D7545
Atmospheric distillation (°C):			
10% distillate	192.9	-	ASTM D86
50% distillate	281.1		
Kinematic viscosity at 40 °C (mm ² /s)	3.21	2.86	ASTM D7042
Specific mass at 20 °C (kg/m ³)	828.0	786.0	ASTM D7042
Cloud point (°C)	0.40	-	ASTM D2500

125
 126

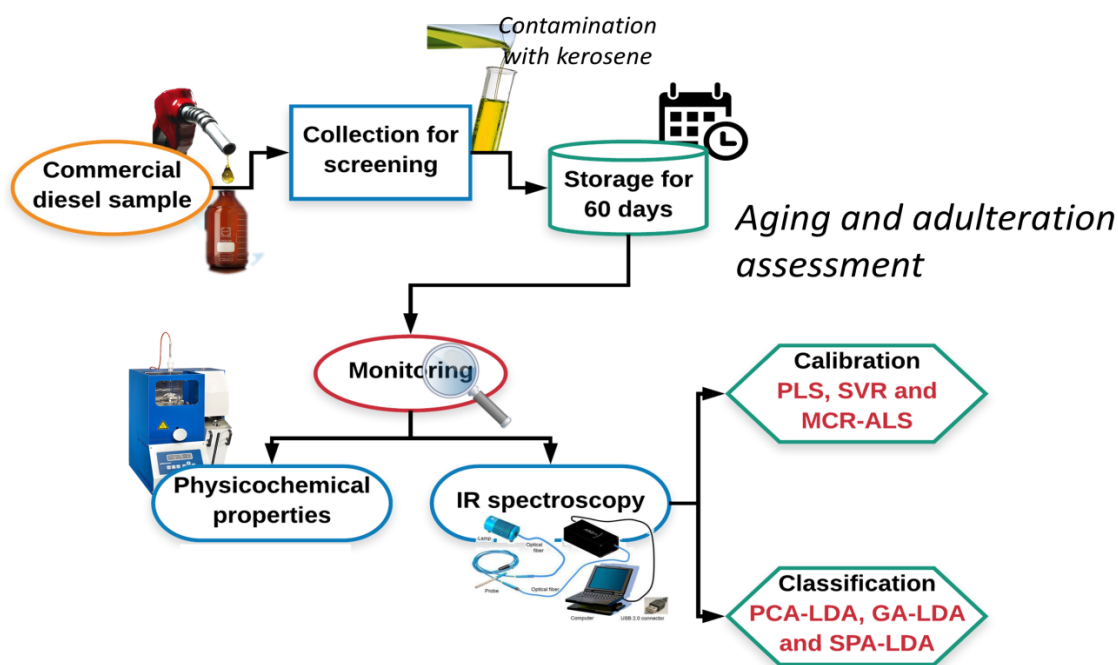
127 Spectral Data Analysis

128 MIR measurements were carried out on a Shimadzu IRAffinity-1 spectrometer equipped with an attenuated total
 129 reflectance (ATR) sample holder and ZnSe crystal. The results were obtained in a wavenumber range from 700-

130 4000 cm^{-1} with resolution of 4 cm^{-1} and 32 scans. NIR data were obtained using a spectrometer from ARCOptix with
131 a quartz cuvette of 1.00 mm in transmission mode. The NIR readings were performed using ARCSpectro software
132 version 1.6 in a wavelength range from 1000-2500 nm and resolution of 8 nm.

133 Data pre-treatment and construction of the classification and calibration models were performed using MATLAB
134 R2012b software (MathWorks Inc., Natick, USA) with PLS Toolbox version 7.9.3 (Eigenvector Research, Inc.,
135 Manson, USA). For the development of multivariate models, baseline correction, Savitzky-Golay smoothing
136 (window of 15 points) and mean center were performed on the MIR spectra dataset for denoising; whereas the NIR
137 data were pre-processed by using multiplicative scatter correction (MSC), Savitzky-Golay smoothing (window of 15
138 points) and mean center.

139 Before modeling, spectral data were divided into training (70%), validation (15%) and prediction (15%) sets for
140 applying LDA to PCA, SPA and GA scores; and calibration (70%) and prediction (30%) sets for PLS regression, by
141 applying the classic Kennard-Stone (KS) uniform sampling selection algorithm [24]. Cross-validation “leave-one-
142 out” was used for PLS. The overall methodology developed in this research is depicted in Scheme 1.



143
144 **Scheme 1.** Process flow diagram of the methodology used in this work.

145
146 **Chemometric Methods**

147 **Exploratory Analysis (PCA) and Variable Selection Methods (SPA and GA)**

148 Principal component analysis (PCA) is an unsupervised classification method that decomposes a data set into
 149 orthogonal variables called principal components (PCs). This reduces the size of the data while retaining the
 150 variance in the data set [25]. PCA is calculated based on the maximum variance contained in the dataset in a
 151 descending order of importance, where the first PC contains the larger explained variance, followed by the second,
 152 and so on. In addition, PCA solves collinearity problems to reduce the data size and improves efficacy to highlight
 153 and visualize the variations and heterogeneities among the samples. The PCA decomposition takes the form of [26]:

$$154 \quad X = TP^T + E \quad (1)$$

155 where X is the spectral data set with n rows (samples) and m columns (wavelengths); T are the scores for all
 156 principal components a ($a = 1, 2, 3, \dots, A$), with size of n rows and A columns; P are the loadings for all principal
 157 components a , with size of m rows and A columns; and E is the residual matrix.

158 Successive projection algorithm (SPA) and genetic algorithm (GA) are techniques used as variables selection.
 159 SPA is a technique that considers each spectral variable as a vector. In this analysis, an initial vector (initial variable)
 160 is used. Then, new vectors with their respective projections are added in a subset orthogonal to that initial vector. In
 161 this way, the SPA selects those variables with more differentiated projections. With this, collinearity problems are
 162 eliminated [27]. GA, on the other hand, has a process that mimics the principle of Darwin's theory of evolution. In
 163 this technique, a population of variables is chosen randomly. This population is composed of subsets of variables.
 164 For each subset a fitness value is assigned through the fitness function present in the GA routine. Based on this
 165 fitness the subsets of variables can be eliminated or "survived" in a step called selection. Then, the genetic operators
 166 mutation and crossover are triggered, where initially selected variables may become unselected (mutation) and
 167 characteristics of one subset can pass to another (crossover). This procedure is called generation. There are as many
 168 generations as requested, and finally, the best fit subset will be the one with the selected variables [28]. The
 169 reduction of the multicollinearity problems obtained by SPA is done through the minimum of the cost function G .
 170 The fitness of GA is also calculated with this function, but in this case the fitness is calculated as the inverse of the
 171 cost function G , which is defined as:

$$172 \quad G = \frac{1}{N_v} \sum_{N=1}^{N_v} g_n \quad (2)$$

173 with g_n being described as:

$$174 \quad g_n = \frac{r^2(x_n, m_{l(n)})}{\min_{l(m) \neq l(n)} r^2(x_n, m_{l(m)})} \quad (3)$$

175 where the numerator is the square of the Mahalanobis distance between the object x_n of the class $l_{(n)}$ and the mean of
 176 its true class $m_{l(n)}$; and the denominator is the square of the Mahalanobis distance between the object x_n and the
 177 center of the nearest wrong class.

178

179 **Linear Discriminant Analysis (LDA)**

180 Linear discriminant analysis (LDA) is a supervised classification technique that improves the segregation level and
 181 reveals clusters that are maximized based on the separation between multiple classes rather than variations within
 182 each group [29]. Since PCA, SPA and GA are exploratory analysis methods, they are only able to show a
 183 distribution pattern between samples. On the other hand, LDA is a supervised classification method capable of
 184 making an exact differentiation between the different data groups. Thus, the scores are utilized as discriminant
 185 variables for LDA technique in order to create a linear decision boundary between them [30]. The LDA
 186 classification score takes the form of:

$$187 \quad cf(x_i) = (x_i - \mu_k)^T \Sigma_{\text{pooled}}^{-1} (x_i - \mu_k) - 2 \ln \pi_k \quad (4)$$

188 where x_i is the measurement vector of sample i ; μ_k is the mean of class k ; Σ_{pooled} is the pooled covariance matrix; and
 189 π_k is the prior probability of class k . These parameters are calculated as [30]:

$$190 \quad \mu_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i \quad (5)$$

$$191 \quad \Sigma_{\text{pooled}} = \frac{1}{n} \sum_{k=1}^K n_k \Sigma_k \quad (6)$$

$$192 \quad \Sigma_k = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (x_i - \mu_k)(x_i - \mu_k)^T \quad (7)$$

$$193 \quad \pi_k = \frac{n_k}{n} \quad (8)$$

194 where Σ_k is the variance covariance matrix of class k ; n_k is the number of samples of class k ; n is the total number of
 195 samples in the training set; and K is the number of classes.

196

197 **Calibration Models**

198 Partial least squares (PLS) regression is a multivariate calibration technique that finds factors (latent variables, LVs)
 199 in the spectra set that explain its maximum variance by using the simultaneous decomposition of the spectral and
 200 concentration matrices. The spectra set X and the concentration set y are decomposed as follows [31]:

$$201 \quad X = TP^T + E \quad (9)$$

202 $y = Uq^T + f$ (10)

203 where T is the scores matrix of X ; P is the loadings matrix of X ; E is the residual matrix of X ; U is the scores matrix
204 of y ; q is the loading vector of y ; and f is the residual vector of y .

205 Support vector machines (SVM) is a supervised learning algorithm employed for training a computational system
206 to recognize patterns and to perform further predictions. SVM for regression, called support vector regression (SVR)
207 [32], is commonly employed in calibration problems for quantification purposes. SVR is based on estimating a
208 response function for each sample spectrum as [33]:

209 $f(x) = W\phi(x) + b$ (11)

210 where the sample spectrum x is non-linear mapped into a high-dimensional feature space Z by $W\phi(x)$, in which
211 $\phi: x_i \rightarrow z_i$; and b represents the bias parameters.

212 Multivariate curve resolution with alternating least squares (MCR-ALS) is a bilinear model that is the multi-
213 wavelength extension from Lambert-Beer's law, and can be described by Equation 12 [34]:

214 $D = CS^T + E$ (12)

215 where D ($n \times m$) is a data matrix containing the NIR or FTIR spectra of n samples for the m recorded wavelengths;
216 C ($n \times A$) and S^T ($A \times m$) are the matrices with the concentration and spectra profiles of A pure components in the
217 samples, respectively. E has the same size as D and contains the unexplained variance from the bilinear model,
218 related as the experimental error [35].

219 Some constraints must be applied to each iteration to reduce the number of possible solutions for C and S^T , and to
220 give chemical meaning to the results. Non-negativity constraint was applied in this work. This constraint forces the
221 concentration and/or spectral profile to be equal or larger than zero [36]. The correlation constraint allows the
222 construction of a model with a univariate internal calibration from the scores calculated by MCR against the
223 reference values concentration, being able to predict the concentration of calibration and unknown samples, and the
224 concentration of these samples has to be in the analytical range of the calibration set [18].

225

226 **Figures of Merit (FOMs)**

227 In order to evaluate the predictive capacity and accuracy of multivariate calibration and classification models, a set
228 of figures of merit was calculated (Table 2) [37].

229

230 **Table 2.** Equations for calculating FOMs.

Calibration	
Root mean square error of prediction (RMSEP)	$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$
Standard error of prediction (SEP)	$SEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$
Bias	$bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n}$
Relative percentage error in concentration prediction (RE%)	$RE(\%) = 100 \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n y_i^2}}$
Root mean square error of cross-validation (RMSECV),	$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_p - y_e)^2}{n}}$
Classification	
Sensitivity (SENS)	$SENS(\%) = \frac{TP}{TP + FN} \times 100$
Specificity (SPEC)	$SPEC(\%) = \frac{TN}{TN + FP} \times 100$
Correct classification (CC)	$CC(\%) = \frac{\{\sum y_1 + \sum y_2\}}{n} \times 100$

n : total number of samples in the set; y_i and \hat{y}_i : actual and predicted concentrations in sample i ; y_p and y_e : predicted and expected concentration values; FN: false negative; FP: false positive; TP: true positive; TN: true negative; y_1 and y_2 : number of samples incorrectly classified for each class.

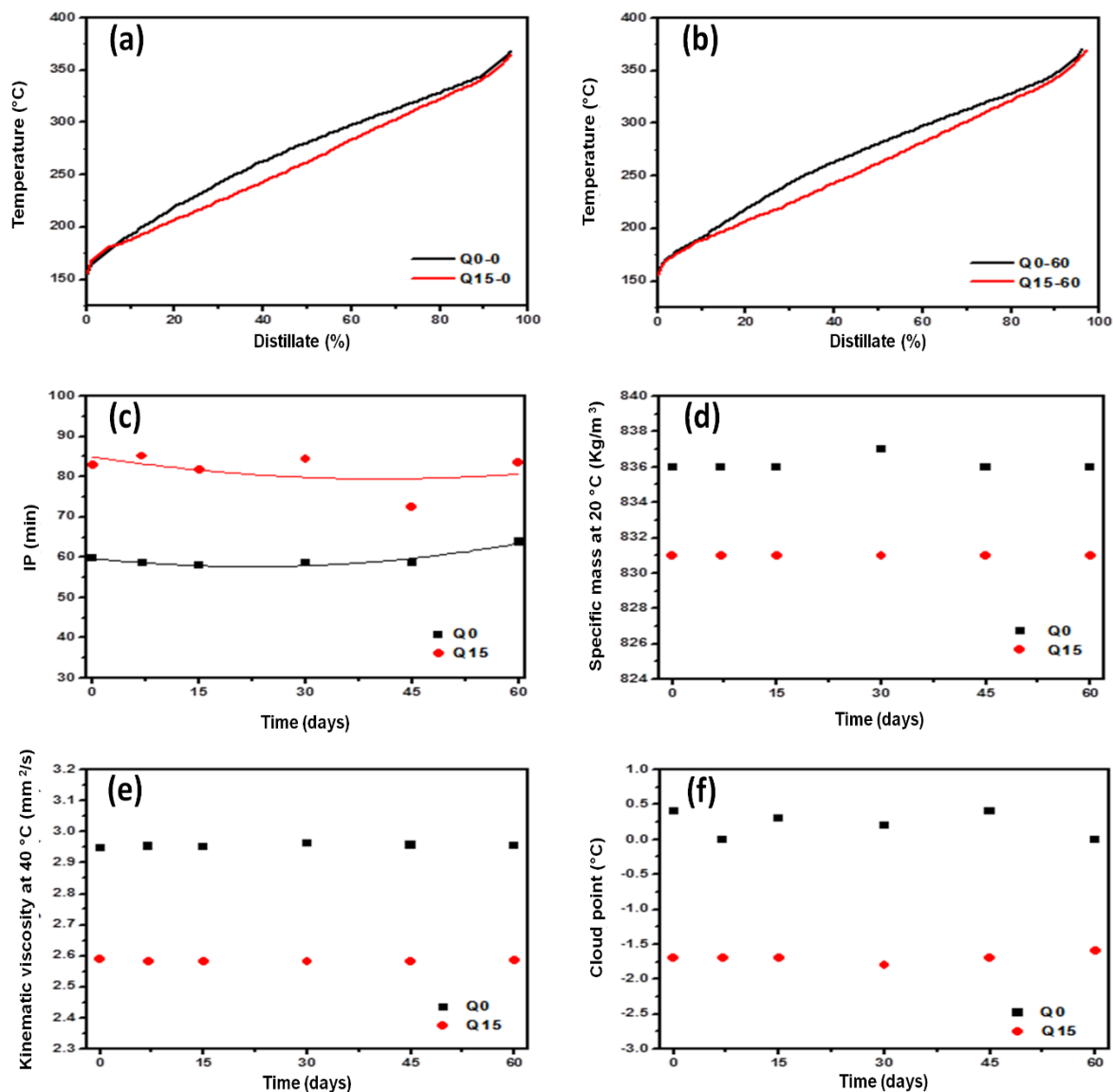
231

232

233 **Results and Discussion**

234 **Physicochemical Analysis**

235 Fig. 1 depicts the results for the physicochemical evaluations. The distillation curves of samples with 0 and 15%
 236 kerosene performed on storage days 0 and 60 are indicated by Q0-0 and Q15-0 (Fig. 1a), Q0-60 and Q15-60 (Fig.
 237 1b). Atmospheric distillation is one of the most important physicochemical properties to detect fuel adulteration in
 238 Brazil, and is based on the boiling temperature profile of the sample components [6]. Although the insertion of
 239 kerosene caused a slight decrease in boiling points of the intermediary hydrocarbon fraction of the blends, the
 240 temperatures obtained were still in accordance with ANP specifications for commercial diesel [5] and did not
 241 change over storage time.



242
 243 **Fig. 1.** Results for (a and b) atmospheric distillation, (c) PetroOxy, (d) specific mass at 20 °C, (e) kinematic
 244 viscosity at 40 °C and (f) cloud point tests for Q0 and Q15 samples during storage.

245
 246 The induction period (IP) of Q0 and Q15 blends (Fig. 1c) remained within a range of little variability between
 247 days 0 and 60 [38]. The addition of 15% kerosene to S10B10 diesel increased the IP of the mixtures, probably due to
 248 the initial dilution of biodiesel in the S10B10 mixture, thus reducing the number of unsaturated molecules and ester
 249 groups available to react and form oxidized by-products [39]. The decrease in specific mass at 20 °C and kinematic
 250 viscosity at 40 °C values (Fig. 1d and 1e) with kerosene contamination is an effect of its relatively lighter
 251 composition in comparison to diesel, and promoting the dilution of denser and more viscous diesel-biodiesel

252 components [40]. The cloud point of the mixtures also decreased after adding kerosene (Fig. 1f) by simply diluting
253 the paraffin waxes and biodiesel ester chains [41,42].

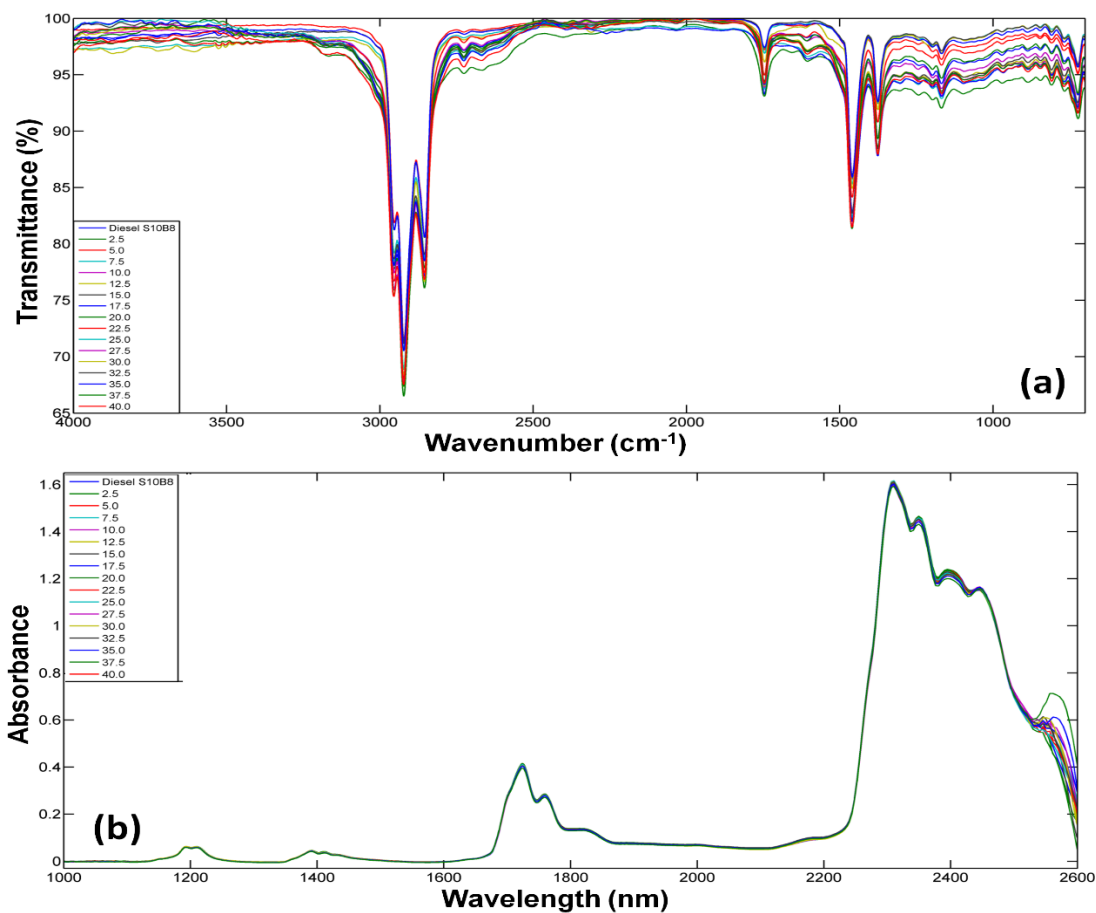
254 According to the results shown above in Fig. 1, none of the physicochemical properties evaluated in this work
255 were able to detect changes in the characteristics of the diesel-biodiesel blends, either pure or kerosene, during
256 storage. In addition, the contaminated diesel S10B10 samples remained within the quality parameters of ANP
257 Resolution N° 30 and the adulteration would easily go unnoticed by a common physicochemical evaluation. Thus,
258 chemometric tools were applied to IR data for elucidating these issues.

259

260 Infrared Spectroscopy

261 The spectra obtained by MIR and NIR analysis for samples with 0 to 40% (v/v) kerosene on the initial day of
262 storage are shown in Fig. 2a and 2b, respectively.

263



264

265 **Fig. 2.** (a) MIR and (b) NIR spectra for samples with 0 to 40% (v/v) kerosene on the first day of storage.

266 In the IR spectra (Fig. 2) there is the presence of some characteristic absorption features of biodiesel and the
267 petroleum distillates. For MIR (Fig. 2a), the bands at 2952-2853 cm^{-1} are related to anti-symmetric and symmetrical
268 stretching modes of CH_2 , CH_3 and CH biodiesel chains. The absorption feature at 1740 cm^{-1} refers to $\text{C}=\text{O}$ stretching
269 mode of saturated aliphatic esters, and those occurring at 1196 and 1168 cm^{-1} correspond to the $\text{C}-\text{O}$ stretching mode
270 of esters from the biodiesel. The bands at 1461 cm^{-1} , 1377 cm^{-1} and 722 cm^{-1} are referent to angular $\text{C}-\text{H}$ bond
271 deformations [43].

272 There are bands at 2130 nm and 2375 nm in the NIR spectra (Fig. 2b) referent to combination $\text{C}-\text{H}/\text{C}=\text{O}$
273 stretching and $\text{C}-\text{H}$ deformation bands, and a $\text{C}-\text{H}$ stretch/ $\text{C}-\text{C}$ stretching combination band, respectively. In addition,
274 there are band suppressions at 1690-1800 nm, 2150 nm, 2400 nm and 2450 nm, referent to the 1st overtone of CH_2
275 symmetric stretching, combination $\text{C}-\text{H}$ stretching/ $\text{C}=\text{O}$ stretching, combination $\text{C}-\text{H}$ stretching/ $\text{C}-\text{C}$ stretching
276 forming CH , and combination $\text{C}-\text{H}$ stretching/ $\text{C}-\text{C}$ stretching forming CH_2 , respectively [44].

277 Although the increase of kerosene content in the samples promotes the biodiesel dilution (as can be seen in Fig.
278 A1 of Online Resource 1), it does not linearly alter the intensity of the ester fingerprint region bands (2130 nm for
279 NIR, 1740 cm^{-1} for MIR), thus precluding univariate quantification.

280

281 **Multivariate Calibration for Kerosene Quantification**

282 As spectroscopic techniques do not resolve the components in a sample, chemical information about single
283 components is embedded in multiple bands in the spectra and spectroscopic instruments alone provide very limited
284 information toward unambiguous identification of unknown mixtures [6,45]. In this case, chemometric tools are
285 commonly employed.

286 PLS regression was applied to the data using 4 LVs (99.94% explained variance) for MIR and 4 LVs (99.99%
287 explained variance) for NIR data. SVR calibration models were obtained using 9 support vectors (SVs) for NIR and
288 11 SVs for MIR. Thus, the SVR model was obtained using C (100), m (0.01) and Gamma (10) parameters, for both
289 techniques, in order to find the best RMSEC value. The use of adequate parameters allows the adjustment of the e -
290 insensitive loss function and the e -tube, which prevents the model from overfitting [33]. The measured *versus*
291 predicted concentration of kerosene (%) plots of PLS and SVR calibration models are found in the Online Resource
292 1. Results for FOMs are depicted in Table 3.

293

294 **Table 3.** Figures of merit (FOMs) for PLS and SVR calibration models.

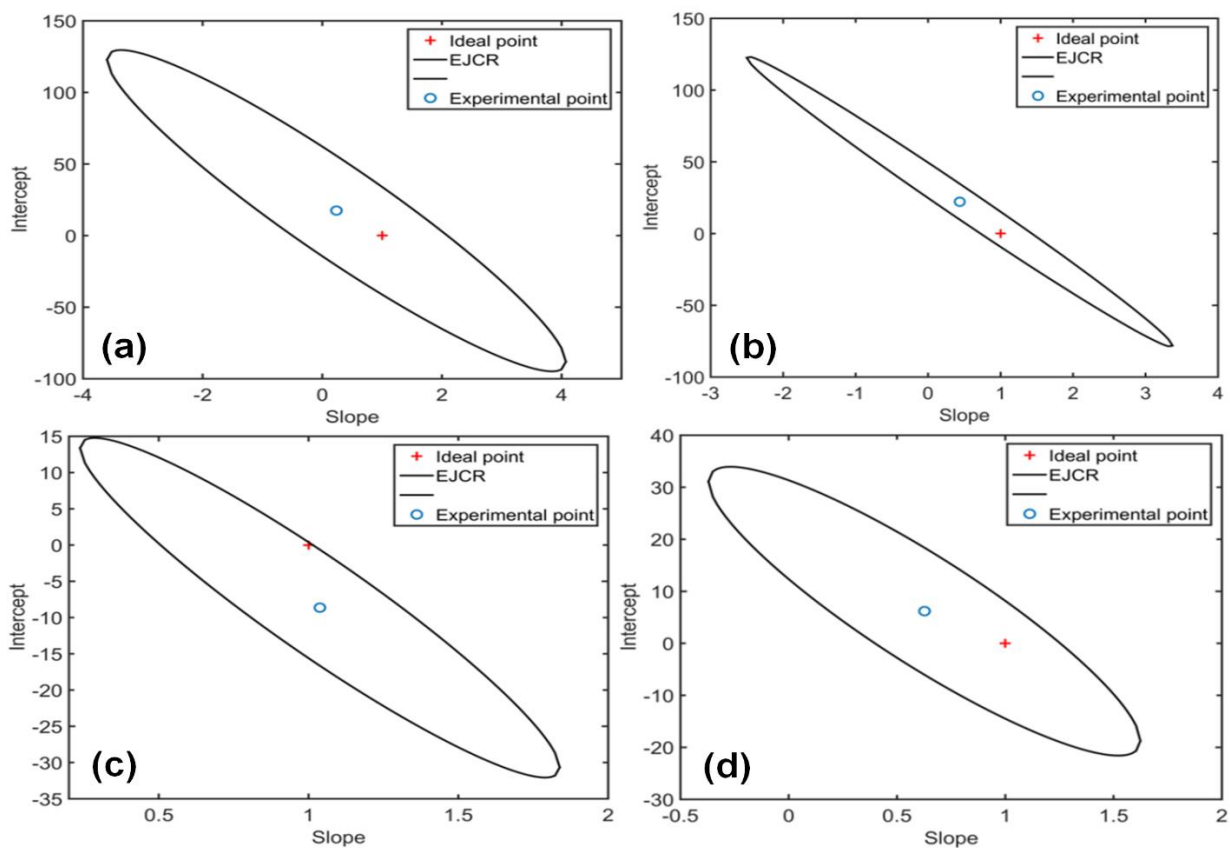
Model	FOM	MIR			NIR		
		Calibration	Cross-validation	Prediction	Calibration	Cross-validation	Prediction
PLS	RMSE (%)	2.35	2.35	3.21	2.59	2.59	3.74
	Bias (%)	-0.19	-0.01	0.12	0.00	5.61x10 ⁻⁵	-3.61
	RE (%)	10.21	-	12.14	11.09	-	14.48
	R ²	0.959	0.964	0.926	0.951	0.951	0.991
	t-test*	-	-	2.774	-	-	2.306
	F-test*	-	-	1.120	-	-	1.051
SVR	RMSE (%)	0.43	5.99	3.85	0.63	5.08	5.56
	Bias (%)	0.104	0.03	-2.99	-0.22	-1.26	3.16
	RE (%)	2.00	-	11.18	3.26	-	19.99
	R ²	0.999	0.714	0.988	0.998	0.784	0.999
	t-test*	-	-	2.446	-	-	2.776
	F-test *	-	-	0.999	-	-	1.028

*Tabulated value of t = 2.776; Critical F = 3.717 at 95% confidence level.

295
 296 The PLS model for MIR data shows satisfactory performance for quantifying the kerosene content in the
 297 adulterated samples, with a root mean square error of cross-validation (RMSECV) and prediction (RMSEP) equal to
 298 2.35% and 3.21%, respectively. Additionally, the model shows to be linear in the concentration range ($R^2 = 0.947$).
 299 Similar results are observed for NIR data, where the RMSECV and RMSEP were respectively equal to 2.59% and
 300 3.74%. The growth in the residues is associated with the higher complexity of NIR data in comparison with MIR.
 301 The NIR spectra are composed of superposed overtone and combinations bands; whereas the MIR spectra are
 302 associated to fundamental vibration modes of the samples' constituents, therefore being more sensible [6]. The
 303 SVR-NIR model presents a constant variance and low residues for both calibration and prediction samples, with a
 304 root mean square error of cross-validation (RMSECV) and prediction (RMSEP) equal to 5.07% and 5.56%,
 305 respectively, and a determination coefficient of 0.999. The same behavior occurs in SVR-MIR, where the RMSECV
 306 and RMSEP were equal to 3.85% and 5.99%, respectively, in addition to a determination coefficient of 0.988, which
 307 indicates good fit throughout the analytical ranges for both methods.

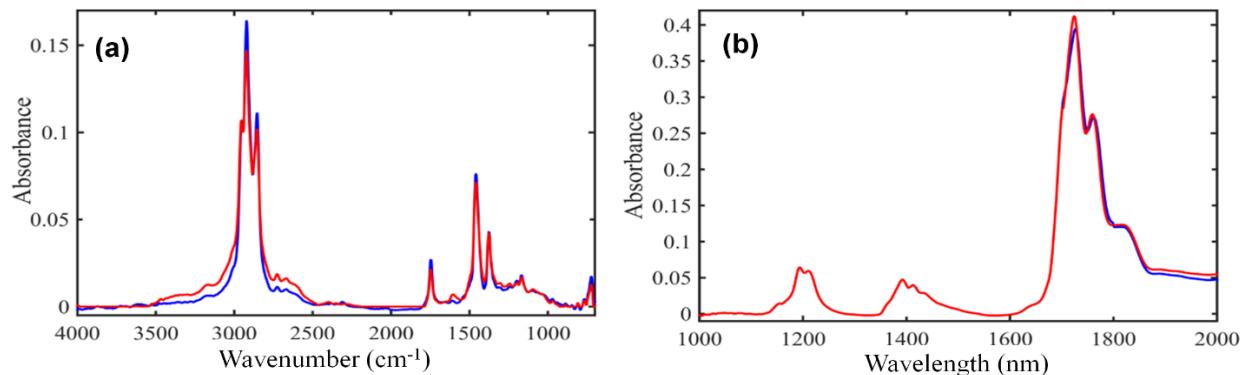
308 The values of calibration, cross-validation and prediction errors obtained for this methodology are close to those
 309 observed in the literature for quantifying kerosene in diesel by another spectroscopic technique [46]. Results of
 310 paired t-tests and F-tests (Table 3) confirm that predicted concentrations were statistically equal to the reference
 311 concentrations and all the calibration models used in this work are valid for a confidence level of 95%. The low
 312 RMSE values and the high correlation coefficients demonstrate that these PLS models may be applied to quantify

313 kerosene volumetric concentration in diesel for controlling adulteration issues. Fig. 3 depicts the elliptical joint
 314 confidence region (EJCR) at 95% confidence level for the slope and intercept of the regression line in predicted
 315 versus reference values. The ellipse contains the ideal point (1,0) for slope and intercept, respectively, showing that
 316 the reference and predicted values are not significantly different at 95% confidence level for PLS-MIR, PLS-NIR,
 317 SVR-NIR and SVR-MIR; thus, no systematic errors were detected in calibration.



318
 319 **Fig. 3.** EJCR for the calibrations models to quantifying kerosene with (a) PLS-MIR, (b) SVR-MIR, (c) PLS-NIR
 320 and (d) SVR-NIR.

321
 322 MCR-ALS was also applied to these data; however, it was not possible to quantify the concentration of kerosene
 323 with IR spectroscopy due to the non-correlation between samples. A low correlation coefficient and high errors were
 324 obtained for MIR, such as RMSEC and R^2 of 20.79% and 0.470, respectively. Meanwhile, the recovered spectral
 325 profile (Sopt) of the adulterant could be calculated and is shown in Fig. 4a.



326
 327 **Fig. 4.** Comparison between the original IR spectra (blue) and the Sopt obtained by MCR-ALS (red) for kerosene
 328 solvent using (a) MIR and (b) NIR data.

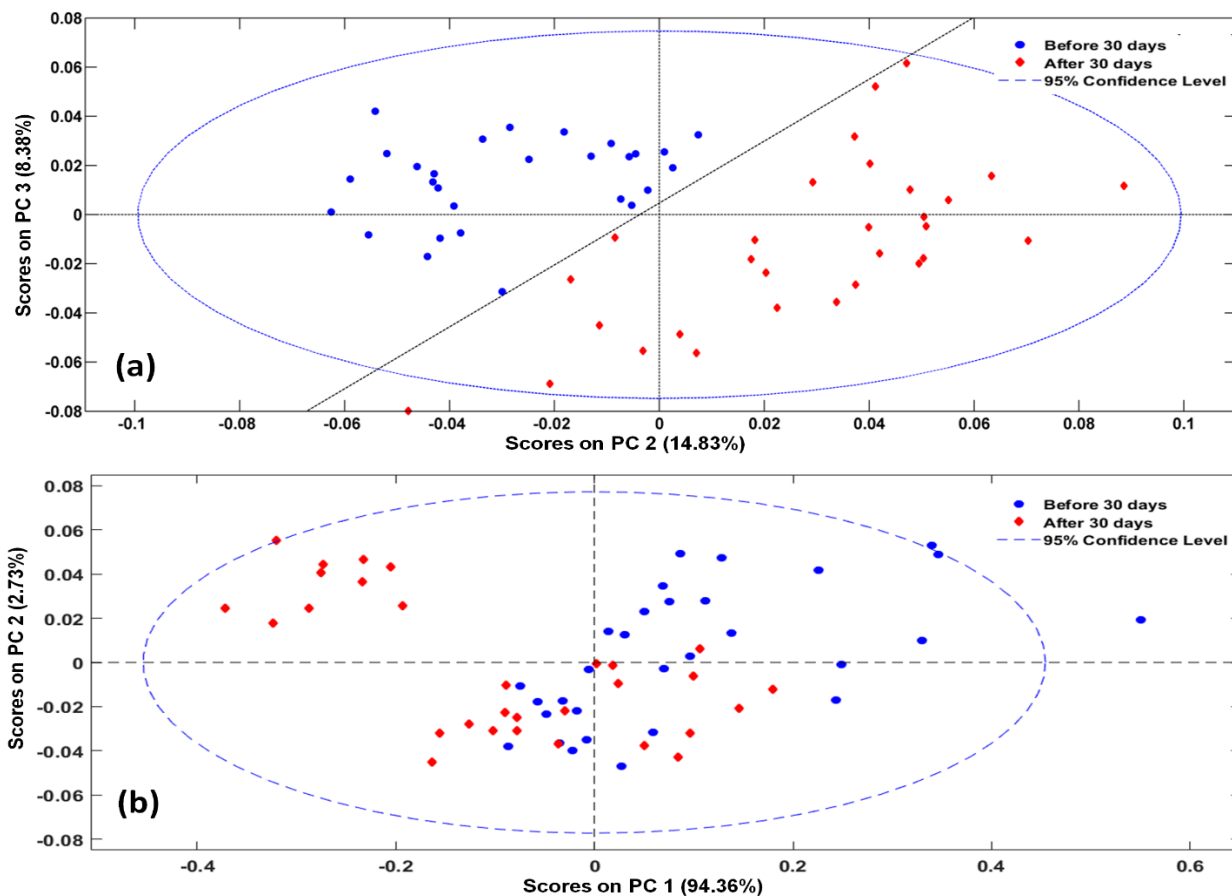
329
 330 The model was able to recover 3 Sopts, where Sopt1, Sopt2 and Sopt3 are the recovered profiles of diesel S10,
 331 kerosene and biodiesel, respectively, identifying the adulterant spectra despite the high chemical similarity between
 332 diesel and kerosene. This can be concluded because of the similarity between the spectral profiles and Sopts,
 333 presenting only a small difference in some band intensities due to the resemblance among the fuel and the adulterant.
 334 MCR-NIR presented lower RMSEC and higher R^2 values than MCR-MIR (15.94% and 0.620, respectively). These
 335 results indicate better precision in the quantification process; meanwhile, this model is not reliable for quantifying
 336 kerosene content and may only be used for the Sopt recuperation (Fig. 4b). It was possible to recover two spectral
 337 profiles (Sopt1 and Sopt2), corresponding to the diesel/biodiesel blend and kerosene, respectively.

338 Correlation analysis (R^2) between kerosene spectra and the recovered profile was also performed and showed the
 339 resemblance among them. The R^2 was of 0.977 and 0.990 for MCR-MIR and MCR-NIR. This method can be
 340 interesting to solve the big issue of identifying kerosene adulteration in commercial diesel, along with PLS and SVR
 341 calibration to efficiently quantify its content.

342
 343 **Multivariate Classification for Fuel Aging**

344 By applying PCA to the preprocessed data, three PCs (87.53% of explained variance) were selected for MIR and
 345 three PCs (96.67% of explained variance) for NIR. Fig. 5 depicts the scores graph for the PCA models without LDA.
 346 The scores graph that best separated the sample classes under and over 30 days of storage in PCA was built using
 347 PC2 x PC3 for MIR and PC1 x PC2 for NIR (Fig. 5a and 5b). PCA was able to separate the samples with aging time

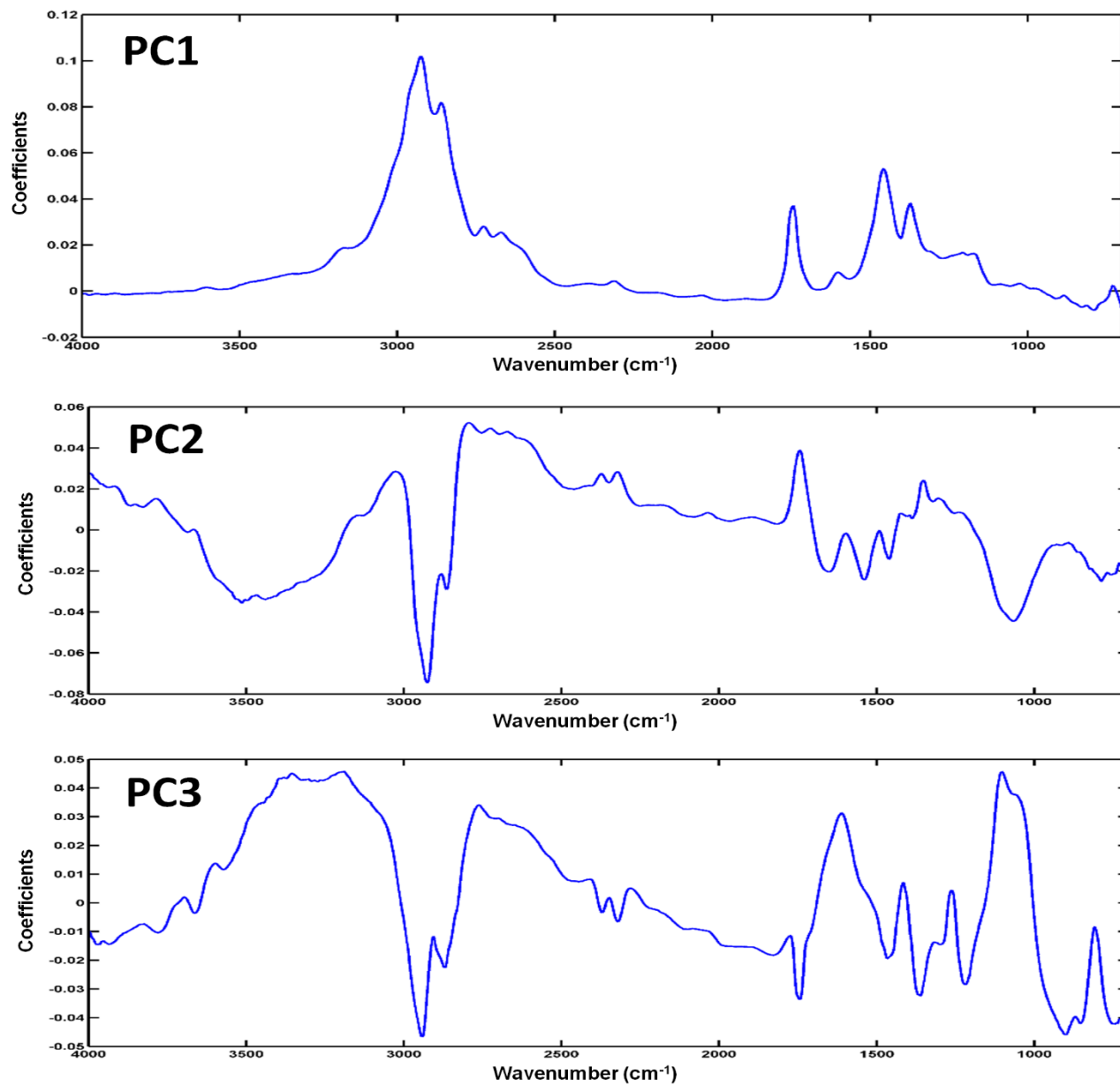
348 over and under 30 days of storage using MIR spectral data (Fig. 5a), detecting some compositional differences
349 caused by the aging process to distinguish both classes. For NIR (Fig. 5b), group distribution was not satisfactory
350 just with the exploratory analysis. Since PCA is used to get a view of the data in space and the important loadings
351 for each PC, LDA is needed for enabling an accurate classification of the samples.



352
353 **Fig. 5.** Scores plot for (a) MIR and (b) NIR spectra analyzed by PCA.

354
355 The loadings profile of PCA-MIR model (Fig. 6) depicts the absorption features of the IR spectra that most
356 influenced the segregation between the sample classes. The coefficients that most influenced PC1 were
357 wavenumbers at 1740 cm^{-1} , 1196 cm^{-1} and 1168 cm^{-1} , referent to biodiesel absorption bands (see Fig. 2a). This
358 occurs since biodiesel is composed of esters, which have different spectroscopic characteristics than the
359 hydrocarbons present in diesel and kerosene. Thus, the pattern recognition model was able to detect the influence of
360 kerosene by the changes in biodiesel content into the blends. It can be proposed that PC1 was able to identify the
361 influence of the increasing content of kerosene adulterant in the storage stability and in the degradation process of

362 the samples, as predicted by PetroOxy results. PC2 and PC3 were more influenced by wavenumbers that correspond
363 to compounds that may have been produced in the degradation of the blends, such as aldehydes, phenols and amides
364 (see Table 4), which justifies the good separation of classes with different storage times in the PCA scores (Fig. 5a).



365
366 **Fig. 6.** PCA loadings profile for MIR.

367
368
369

370

371 **Table 4.** Absorption features that most influenced PC2 and PC3 in MIR loadings.

Positive coefficients of PC2		Negative coefficients of PC2	
Wavenumber (cm ⁻¹)	Vibration	Wavenumber (cm ⁻¹)	Vibration
3070-3010	Aromatic $\nu(\text{C-H})$	3500	O-H
2800-2700	Aldehyde $\nu(\text{C-H})$	2921	$\nu(\text{C-H})$ of CH ₂
1800-1700	$\nu(\text{C=O})$	2853	$\nu(\text{C-H})$ of CH ₂
		1600-1450	Aromatic $\nu(\text{C=C})$
		1060-1020	$\nu(\text{C=S})$ or $\nu(\text{S=O})$
Positive coefficients of PC3		Negative coefficients of PC3	
Wavenumber (cm ⁻¹)	Vibration	Wavenumber (cm ⁻¹)	Vibration
3300	Carboxylic acid $\nu(\text{O-H})$ or amide $\nu(\text{N-H})$	2921	$\nu(\text{C-H})$ of CH ₂
2800-2700	Aldehyde $\nu(\text{C-H})$	2853	$\nu(\text{C-H})$ of CH ₂
1560-1530	Amide $\delta(\text{N-H})$ and $\nu(\text{C-N})$	1800-1700	$\nu(\text{C=O})$
1060-1020	$\nu(\text{S=O})$	1250	$\nu(\text{C-O})$
		722	$\rho_{\text{as}}(\text{C-H})$ of CH ₂

ν : axial deformation; δ : angular deformation; ρ_{as} : asymmetric deformation at the plane.

372

373

374 In PCA-LDA, 10 PCs were selected for MIR and NIR (96.53% and 97.87% of explained variance, respectively)

375 to classify the data according to its storage time (over and under 30 days of storage). SPA and GA algorithms

376 provide a set of selected variables that promotes the higher inter-class segregation to be used in LDA. The selected

377 variables for both SPA and GA are shown in Table 5. All the discriminant function (DF) plots for the models are

378 presented on Online Resource 2.

379

380

381

382

383

384

385

386

387

388 **Table 5.** SPA and GA selected variables for MIR and NIR data.

Spectroscopic technique	Model	Selected variables									
MIR (cm ⁻¹)	SPA (34) ^a	698	745	779	810	854	907	999	1204		
		1321	1360	1398	1422	1445	1485	1533			
		1578	1663	1721	1748	1794	2280	2336			
		2791	2886	2938	2959	3005	3518	3545			
		3651	3686	3794	3906	3998					
	GA (26) ^a	1022	1047	1059	1092	1142	1167	1254			
		1346	1416	1528	1744	1746	2045	2095			
		2253	2452	2675	2806	2860	2997	3009			
		3096	3406	3688	3711	3993					
NIR (nm)	SPA (31) ^a	922	925	928	934	941	946	953	957		
		964	966	973	983	987	997	1008	1017		
		1026	1035	1048	1095	1139	1238	1387			
		1411	1673	1689	1705	1765	1852	1926			
		2133									
	GA (27) ^a	954	977	1033	1065	1096	1152	1154			
		1160	1168	1244	1265	1309	1340	1356			
		1378	1397	1437	1478	1528	1660	1696			
		1748	1799	1814	1833	1926	2108				

389 ^anumber of selected variables.

390 As also seen in PCA loadings (Table 4) for the model with good aging classification before applying LDA (PCA-
391 MIR), some of the SPA and GA selected variables for MIR and NIR are related to biodiesel content (~1748 cm⁻¹ for
392 MIR and ~2133 nm for NIR, for example) and probable products of sample degradation, such as aldehydes (1700-
393 1800 cm⁻¹ for MIR), amides (~1530 cm⁻¹ for MIR and ~1430 nm for NIR) and carboxylic acids (~1700 cm⁻¹ for
394 MIR and ~1920 nm for NIR) [43,47]. After this selection, LDA was applied in order to classify the samples into
395 their correct classes. Sensitivity (Sens), specificity (Spec) and correct classification (CC) were calculated in order to
396 infer the prediction performance for these models. Table 6 summarizes the results for figures of merit for the
397 classification models. PCA-LDA and SPA-LDA reached 100% accuracy with both IR methods; although GA-LDA
398 presented some lower results of sensitivity and correct classification for the sample class before 30 days of storage,
399 its efficiency was still satisfactory (>85.7%).

400 **Table 6.** Figures of merit (FOMs) inferred for the classification work with PCA-LDA, SPA-LDA and GA-LDA
 401 models.

Spectroscopic technique	FOM	PCA-LDA	SPA-LDA	GA-LDA
MIR	Before 30 days			
	Sens (%)	100.0	100.0	85.7
	Spec (%)	100.0	100.0	100.0
	CC (%)	100.0	100.0	91.6
	After 30 days			
	Sens (%)	100.0	100.0	100.0
	Spec (%)	100.0	100.0	100.0
	CC (%)	100.0	100.0	100.0
	NIR	Before 30 days		
Sens (%)		100.0	100.0	85.7
Spec (%)		100.0	100.0	100.0
CC (%)		100.0	100.0	91.6
After 30 days				
Sens (%)		100.0	100.0	100.0
Spec (%)		100.0	100.0	100.0
CC (%)		100.0	100.0	100.0

402
 403 These results show that PCA-LDA, SPA-LDA and GA-LDA models for both IR spectroscopies are capable of
 404 differentiating the monthly storage time of these fuels with good accuracy, unlike what can be observed in the
 405 physicochemical analysis results during storage, which were not able to detect changes in diesel-biodiesel samples
 406 with or without kerosene during the monitoring. Furthermore, NIR data can be recorded by portable instruments,
 407 enabling faster “in loco” inspection procedures with an effective and simple methodology with the combined
 408 classification models. PCA loadings detected the presence of adulteration by observing variations on biodiesel
 409 concentration, in addition to detect chemical species from decomposition reactions of the diesel-biodiesel-kerosene
 410 mixtures as the main features responsible for aging class separation, as well as SPA and GA selected variables.

411
 412 **Conclusions**

413 The NIR and MIR spectra coupled to PLS and SVR models for quantifying kerosene content presented low RMSE
 414 values and high correlation between real and predicted concentrations, in spite of the similar chemical composition

415 of diesel and kerosene. MCR-ALS with correlation constraint was able to identify and recover the spectral profile of
416 commercial diesel and kerosene adulterant from the IR spectra of contaminated blends.

417 PCA-LDA, SPA-LDA and GA-LDA enabled correctly classifying diesel-biodiesel with kerosene in different
418 degradation levels, separating these samples into two well-defined groups under and over thirty days of storage. The
419 method was highly accurate and reliable for evaluating fuel storage stability. PCA loadings, as well as GA and SPA
420 selected variables, detected that spectroscopic features related to degradation products such as amides, carboxylic
421 acids and aldehydes were responsible for the classification by aging stage.

422 The multivariate classification methodology developed in this paper is an efficient tool for classifying commercial
423 diesel with kerosene adulterant by aging time and chemically observe the degradation phenomenon. Combining
424 MCR-ALS with PLS or SVR models is powerful to solve the great issue in quantifying and identifying this
425 adulteration, being interesting to improve the investigative process of adulteration in diesel fuel screening. The
426 results of the study that we performed demonstrated good results in the quantification using these techniques.
427 However, more in-depth studies with more sampling need to be performed in order to have a better validation of the
428 technique and to be more certain. However, our results are encouraging. The evaluation method is simple, fast, does
429 not require pretreatment of the samples, may be carried out “in loco” with portable NIR instruments and is low cost.

430

431 **Acknowledgements**

432 The authors acknowledge the support provided by the Post-Graduate Chemistry Program PPGQ/UFRN, the
433 Energetic Technologies Laboratory (LTEN), the Biological Chemistry and Chemometrics Group, the CAPES and
434 CNPQ – Brazil for the financial support.

435

436 **Conflict of Interest**

437 The authors declare that they have no conflict of interest.

438

439 **References**

440 [1] Obeidat SM, Al-Ktash MM, Al-Momani IF. Study of fuel assessment and adulteration using EEMF and
441 multiway PCA. *Energy Fuels*. 2014; <https://doi.org/10.1021/ef500718e> [CrossRef] [Google Scholar]

- 442 [2] Krakowska B, Stanimirova I, Orzel J, Daszykowski M, Grabowski I, Zaleszczyk G, Sznajder M. Detection
443 of discoloration in diesel fuel based on gas chromatographic fingerprints. *Anal Bioanal Chem.* 2015;
444 <https://doi.org/10.1007/s00216-014-8332-4> [CrossRef] [Google Scholar]
- 445 [3] Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – ANP. Resolução No. 3 de 08.02.2007. In:
446 DOU 09.02.2007. [http://legislacao.anp.gov.br/?path=legislacao-anp/resol-anp/2007/fevereiro&item=ramp-3--](http://legislacao.anp.gov.br/?path=legislacao-anp/resol-anp/2007/fevereiro&item=ramp-3--2007&export=pdf)
447 [2007&export=pdf](http://legislacao.anp.gov.br/?path=legislacao-anp/resol-anp/2007/fevereiro&item=ramp-3--2007&export=pdf). Accessed in 20 Oct 2018.
- 448 [4] Menezes EW, Silva R, Cataluña R, Ortega RJC. Effect of ethers and ether/ethanol additives on the
449 physicochemical properties of diesel fuel and on engine tests. *Fuel.* 2006; <https://doi.org/10.1016/j.fuel.2005.08.027>
450 [CrossRef] [Google Scholar]
- 451 [5] Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – ANP. Resolução No. 30 de 23.06.2016. In:
452 DOU 24.06.2016.
453 http://www.lex.com.br/legis_27160107_RESOLUCAO_N_30_DE_23_DE_JUNHO_DE_2016.aspx. Accessed in
454 20 Oct 2018.
- 455 [6] Câmara ABF, de Carvalho LS, Morais CLM, Lima LAS, Araújo HOM, Oliveira FM, Lima KMG. MCR-
456 ALS and PLS coupled to NIR/MIR spectroscopies for quantification and identification of adulterant in biodiesel-
457 diesel blends. *Fuel.* 2017; <https://doi.org/10.1016/j.fuel.2017.08.072> [CrossRef] [Google Scholar]
- 458 [7] Cunha IBS, Fernandes AMAP, Tega DU, Simas RC, Nascimento HL, Sá GF, Daroda RJ, Eberlin MN,
459 Alberici RM. Quantitation and quality control of biodiesel/petrodiesel (Bn) blends by easy ambient sonic-spray
460 ionization mass spectrometry. *Energy Fuels.* 2012; <https://doi.org/10.1021/ef3010866> [CrossRef] [Google Scholar]
- 461 [8] Gotor R, Tiebe C, Schilischka J, Bell J, Rurack K. Detection of adulterated diesel using fluorescent test
462 strips and smartphone readout. *Energy Fuels.* 2017; <https://doi.org/10.1021/acs.energyfuels.7b01538> [CrossRef]
463 [Google Scholar]
- 464 [9] Jose TK, Anand K. Effects of biodiesel composition on its long term storage stability. *Fuel.* 2016;
465 <https://doi.org/10.1016/j.fuel.2016.03.007> [CrossRef] [Google Scholar]
- 466 [10] Brereton RG, Jansen J, Lopes J, Marini F, Pomerantsev A, Rodionova O, Roger JM, Walczak B, Tauler R.
467 Chemometrics in analytical chemistry – part II: modeling, validation, and applications. *Anal Bioanal Chem.* 2018;
468 <https://doi.org/10.1007/s00216-018-1283-4> [CrossRef] [Google Scholar]

- 469 [11] Zhang J, Wei X, Huang J, Lin H, Deng K, Li Z, Shao Y, Zou D, Chen Y, Huang P, Wang Z. Attenuated
470 total reflectance Fourier transform infrared (ATR-FTIR) spectral prediction of postmortem interval from vitreous
471 humor samples. *Anal Bioanal Chem.* 2018; <https://doi.org/10.1007/s00216-018-1367-1> [CrossRef] [Google Scholar]
- 472 [12] Aboualizadeh E, Ranji M, Sorenson CM, Sepehr R, Sheibani N, Hirschmugl CJ. Retinal oxidative stress at
473 the onset of diabetes determined by synchrotron FTIR widefield imaging: towards diabetes pathogenesis. *Analyst.*
474 2017; <https://doi.org/10.1039/c6an02603f> [CrossRef] [Google Scholar]
- 475 [13] Theophilou G, Morais CLM, Halliwell DE, Lima KMG, Drury J, Martin-Hirsch PL, Stringfellow HF,
476 Hapangama DK, Martin FL. Synchrotron- and focal plane array-based Fourier-transform infrared spectroscopy
477 differentiates the basalis and functionalis epithelial endometrial regions and identifies putative stem cell regions of
478 human endometrial glands. *Anal Bioanal Chem.* 2018; <https://doi.org/10.1007/s00216-018-1111-x> [CrossRef]
479 [Google Scholar]
- 480 [14] Marques AS, Moraes EP, Júnior MAA, Moura AD, Neto VFA, Neto RM, Lima KMG. Rapid
481 discrimination of *klebsiella pneumoniae* carbapenemase 2 – producing and non-producing *klebsiella*
482 *pneumoniae* strains using near-infrared spectroscopy (NIRS) and multivariate analysis. *Talanta.* 2015;
483 <https://doi.org/10.1016/j.talanta.2014.11.006> [CrossRef] [Google Scholar]
- 484 [15] Hu J, Ma X, Liu L, Wu Y, Ouyang J. Rapid evaluation of the quality of chestnuts using near-infrared
485 reflectance spectroscopy. *Food Chem.* 2017; <https://doi.org/10.1016/j.foodchem.2017.03.127> [CrossRef] [Google
486 Scholar]
- 487 [16] Corgozinho CNC, Pasa VMD, Barbeira PJS. Determination of residual oil in diesel oil by
488 spectrofluorimetric and chemometric analysis. *Talanta.* 2008; <https://doi.org/10.1016/j.talanta.2008.03.003>
489 [CrossRef] [Google Scholar]
- 490 [17] Thissen U, Pepers M, Ustun B, Melssen WJ, Buydens LMC. Comparing support vector machines to PLS
491 for spectral regression applications. *Chem Intell Lab Syst.* 2004; <https://doi.org/10.1016/j.chemolab.2004.01.002>
492 [CrossRef] [Google Scholar]
- 493 [18] Dantas WFC, Alves JCL, Poppi RJ. MCR-ALS with correlation constraint and Raman spectroscopy for
494 identification and quantification of biofuels and adulterants in petroleum diesel. *Chemom Intell Lab Syst.* 2017;
495 <https://doi.org/10.1016/j.chemolab.2017.04.002> [CrossRef] [Google Scholar]

- 496 [19] de Juan A, Tauler R. Multivariate curve resolution (MCR) from 2000: progress in concepts and
497 applications. *Crit Rev Anal Chem.* 2006; <https://doi.org/10.1080/10408340600970005> [CrossRef] [Google Scholar]
- 498 [20] ASTM D 7545-14. Standard test method for oxidation stability of middle distillate fuels – rapid small scale
499 oxidation test (RSSOT). In: West Conshohocken (PA): ASTM International. 2014;
500 <https://www.astm.org/Standards/D7545.htm>. Accessed 23 Oct 2018.
- 501 [21] ASTM D 86-12. Standard test method for distillation of petroleum products at atmospheric pressure. In:
502 West Conshohocken (PA): ASTM International. 2013;
503 <https://www.astm.org/DATABASE.CART/HISTORICAL/D86-12.htm>. Accessed 23 Oct 2018.
- 504 [22] ASTM D 7042-14. Standard test method for dynamic viscosity and density of liquids by Stabinger
505 viscometer (and the calculation of kinematic viscosity). In: West Conshohocken (PA): ASTM International. 2014;
506 <https://www.astm.org/DATABASE.CART/HISTORICAL/D7042-14.htm>. Accessed 23 Oct 2018.
- 507 [23] ASTM D 2500-11. Standard test method for cloud point of petroleum products. In: West Conshohocken
508 (PA): ASTM International. 2011; <https://www.astm.org/DATABASE.CART/HISTORICAL/D2500-11.htm>.
509 Accessed 23 Oct 2018.
- 510 [24] Kennard RW, Stone LA. Computer aided design of experiments. *Technometrics.* 1969;
511 <https://doi.org/10.2307/1266770> [CrossRef] [Google Scholar]
- 512 [25] Bro R, Smilde AK. Principal component analysis. *Anal. Methods.* 2014;
513 <https://doi.org/10.1039/C3AY41907J> [CrossRef] [Google Scholar]
- 514 [26] Eftekhari A, Forouzanfar M, Moghaddam HA, Alirezaie J. Block-wise 2D kernel PCA/LDA for face
515 recognition. *Inform Process Lett.* 2010; <https://doi.org/10.1016/j.ipl.2010.06.006> [CrossRef] [Google Scholar]
- 516 [27] Pontes MJC, Galvão RKH, Araújo MCU, Moreira PNT, Neto ODP, José GE, Saldanha TCB. The
517 successive projections algorithm for spectral variable selection in classification problems. *Chemom Intell Lab Syst.*
518 2005; <https://doi.org/10.1016/j.chemolab.2004.12.001> [CrossRef] [Google Scholar]
- 519 [28] Broadhursta D, Goodacrea R, Jones A, Rowland JJ, Kell DB. Genetic algorithms as a method for variable
520 selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass
521 spectrometry. *Anal Chim Acta.* 1997; [https://doi.org/10.1016/S0003-2670\(97\)00065-2](https://doi.org/10.1016/S0003-2670(97)00065-2) [CrossRef] [Google Scholar]
- 522 [29] Dixon SJ, Brereton RG. Comparison of performance of five common classifiers represented as boundary
523 methods: Euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning

524 vector quantization and support vector machines, as dependent on data structure. *Chemometr Intell Lab Syst.* 2009;
525 <https://doi.org/10.1016/j.chemolab.2008.07.010> [CrossRef] [Google Scholar]

526 [30] Wu W, Mallet Y, Walczak B, Penninckx W, Massart DL, Heuerding S, Erni F. Comparison of regularized
527 discriminant analysis, linear discriminant analysis and quadratic discriminant analysis. Applied to NIR data. *Anal*
528 *Chim Acta* 1996; [https://doi.org/10.1016/0003-2670\(96\)00142-0](https://doi.org/10.1016/0003-2670(96)00142-0) [CrossRef] [Google Scholar]

529 [31] Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. *Anal. Chim. Acta.* 1986;
530 [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9) [CrossRef] [Google Scholar]

531 [32] Smola AJ, Schölkopf B. A tutorial on support vector regression. *Stat. Comput.* 2004;
532 <https://doi.org/10.1023/B:STCO.0000035301.49549.88> [CrossRef] [Google Scholar]

533 [33] Alves JCL, Henriques CB, Poppi RJ. Determination of diesel quality parameters using support vector
534 regression and near infrared spectroscopy for an in-line blending optimizer system. *Fuel.* 2012;
535 <https://doi.org/10.1016/j.fuel.2012.03.016> [CrossRef] [Google Scholar]

536 [34] Tauler R, Kowalski B, Fleming S. Multivariate curve resolution applied to spectral data from multiple runs
537 of an industrial process. *Anal Chem.* 1993; <https://doi.org/10.1021/ac00063a019> [CrossRef] [Google Scholar]

538 [35] Jaumot J, Igne B, Anderso CA, Drennen JK, de Juan A. Blending process modeling and control by
539 multivariate curve resolution. *Talanta.* 2013; <https://doi.org/10.1016/j.talanta.2013.09.037> [CrossRef] [Google
540 Scholar]

541 [36] Bro R, de Jong S. A fast non-negativity-constrained least squares algorithm. *J Chemom.* 1997;
542 [https://doi.org/10.1002/\(SICI\)1099-128X\(199709/10\)11:5<393::AID-CEM483>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1099-128X(199709/10)11:5<393::AID-CEM483>3.0.CO;2-L) [CrossRef] [Google
543 Scholar]

544 [37] Olivieri AC, Faber NM, Ferré J, Boqué R, Kalivas JH, Mark, H. Uncertainty estimation and figures of
545 merit for multivariate calibration. *Pure Appl Chem.* 2006; <https://doi.org/10.1351/pac200678030633> [CrossRef]
546 [Google Scholar]

547 [38] Botella L, Bimbela F, Martin L, Arauzo J, Sanchez JL. oxidation stability of biodiesel fuels and blends
548 using the Rancimat and PetroOxy methods. Effect of 4-allyl-2,6-dimethoxyphenol and catechol as biodiesel additives
549 on oxidation stability. *Front Chem.* 2014; <https://doi.org/10.3389/fchem.2014.00043> [CrossRef] [Google Scholar]

550 [39] Karavalakis G, Stournas S, Karonis D. Evaluation of the oxidation stability of diesel/biodiesel blends. *Fuel.*
551 2010; <https://doi.org/10.1016/j.fuel.2010.03.041> [CrossRef] [Google Scholar]

552 [40] Roy MM, Wang W, Alawi M. Performance and emissions of a diesel engine fueled by biodiesel-diesel,
553 biodiesel-diesel-additive and kerosene-biodiesel blends. *Energ Convers Manage.* 2014;
554 <https://doi.org/10.1016/j.enconman.2014.04.033> [CrossRef] [Google Scholar]

555 [41] Yadav SR, Murthy KV, Mishra D, Baral B. Estimation of petrol and diesel adulteration with kerosene and
556 assessment of usefulness of selected automobile fuel quality test parameters. *IJEST.* 2005;
557 <https://doi.org/10.1007/BF03325839> [CrossRef] [Google Scholar]

558 [42] Ziegler K, Manka J. The effect of mixing diesel fuels additized with kerosene and cloud point depressants.
559 SAE Technical Paper 2000-01-2884. 2000; <https://doi.org/10.4271/2000-01-2884> [CrossRef] [Google Scholar]

560 [43] Silverstein RM, Webster FX, Kiemle DJ. Spectrometric identification of organic compounds. 7th Ed. New
561 Jersey: John Wiley & Sons; 2005. [Google Scholar]

562 [44] Workman Jr, Weyer JL. Practical guide to interpretive near-infrared spectroscopy. 1st Ed. Boca Raton:
563 CRC Press; 2008. [Google Scholar]

564 [45] Yang C, Yang Z, Zhang G, Hollebone B, Landriault M, Wang Z, Lambert P, Brown CE. Characterization
565 and differentiation of chemical fingerprints of virgin and used lubricating oils for identification of contamination or
566 adulteration sources. *Fuel.* 2016; <https://doi.org/10.1016/j.fuel.2015.09.070> [CrossRef] [Google Scholar]

567 [46] Divya O, Mishra AK. Multivariate methods on the excitation emission matrix fluorescence spectroscopic
568 data of diesel-kerosene mixtures: a comparative study. *Anal Chim Acta.* 2007;
569 <https://doi.org/10.1016/j.aca.2007.03.079> [CrossRef] [Google Scholar]

570 [47] Monograph NIR spectroscopy. A guide to near-infrared spectroscopic analysis of industrial manufacturing
571 processes. In: Metrohm NIR Systems. 2017.
572 http://www.mep.net.au/wpcontent/uploads/2013/05/MEP_Monograph_NIRS_81085026EN.pdf. Accessed
573 29 Oct 2018. [CrossRef] [Google Scholar]

574

575