

ARTICLE

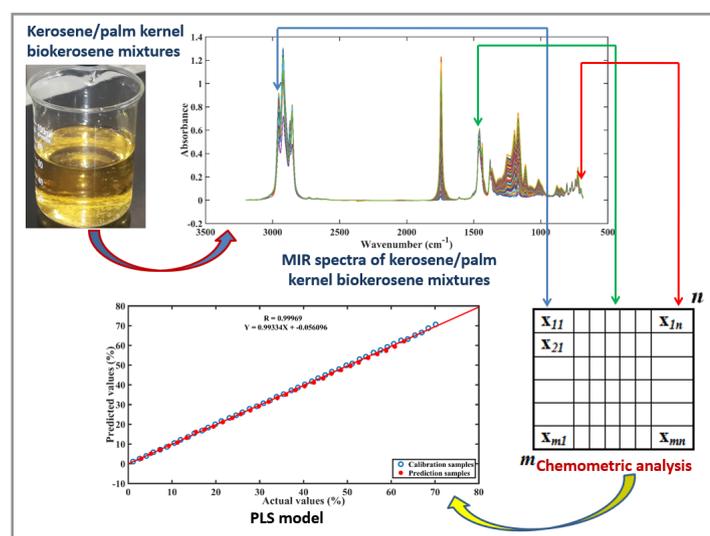
Rapid Quantification of the Palm Kernel Biokerosene Content in Mixtures with Aviation Kerosene using MIR Spectroscopy and Multivariate Regression by PLS

Ademar Domingos Viagem Máquina^{1,2*}  , Maria Teresa Carvalho Ferreira¹ , Edvando Souza Teles¹ , Douglas Queiroz Santos³ , Waldomiro Borges Neto¹ 

¹Instituto de Química, Universidade Federal de Uberlândia, Campus Santa Mônica, 38408-100, Uberlândia, MG, Brazil

²Universidade de Púnguè, Faculdade de Ciências Exactas e Tecnológicas, Rua Nacional 6, CEP 323, Chimoio, Manica, Moçambique

³Escola Técnica de Saúde, Universidade Federal de Uberlândia, Campus Umuarama, 38400-902, Uberlândia, MG, Brazil



An alternative methodology was developed to monitor the biokerosene content of palm kernel in blend with kerosene using medium infrared spectroscopy associated with partial least squares (PLS). The efficiency of this methodology was analyzed based on the parameters of accuracy and figures of merit. The values of root-mean-square error of cross-validation (RMSECV), root-mean-square error of calibration (RMSEC) and root-mean-square error of prediction (RMSEP) were in agreement because the RMSEP was higher than RMSECV and RMSEC. In addition, the RMSEP value is considered acceptable according to the Brazilian standard ABNT NBR 15568 because it is less than 1%.

The figures of merit were performed in agreement with the requirements established in the standard ASTM E1655-05. The linearity of the model was assessed based on the analysis of the model fit through the correlation of the actual and predicted values of the calibration and prediction sets, where a high correlation between the values was evidenced, with a correlation coefficient (R) exceeding 0.99. The good results of the application of MIR spectroscopy combined with multivariate regression by PLS suggest that this analytical methodology is feasible, efficient and suitable for use by inspection agencies to control the biokerosene content of palm kernel in mixture with diesel.

Keywords: Chemometric method, Monitoring, Biofuels, Fuel.

Cite: Máquina, A. D. V.; Ferreira, M. T. C.; Teles, E. S.; Santos, D. Q.; Borges Neto, W. Rapid Quantification of the Palm Kernel Biokerosene Content in Mixtures with Aviation Kerosene using MIR Spectroscopy and Multivariate Regression by PLS. *Braz. J. Anal. Chem.*, 2021, 8 (32), pp 155–164. doi: <http://dx.doi.org/10.30744/brjac.2179-3425.AR-26-2021>

Submitted 28 February 2021, Resubmitted 05 May 2021, Accepted 25 May 2021, Available online 30 June 2021.

INTRODUCTION

Aviation kerosene, commercially known as QAV, is a fuel derived from non-renewable energy sources used mainly in the airline sector. According to data provided by the International Air Transport Association - IATA, the use of this fuel by this sector generates approximately 2% of carbon dioxide emissions into the atmosphere, reaching a level of 3% by 2030 [1,2].

Several countries have sought to impose limits on current and future emissions to mitigate these polluting gases in the environment. These emissions represent a global concern in the 21st century and may increase with the expansion of the aviation industry. Thus, it is necessary to use environmentally sustainable fuel as an alternative to fossil fuel [3].

Aviation biokerosene (BioQAV) is a fuel derived from renewable sources with the potential to replace fossil kerosene. Some oilseeds such as jatropha (*Jatropha curcas L.*), camelina (*Camelina sativa*), babassu (*Attalea speciosa*), macaúba (*Acrocomia aculeata*) and palm kernel (*Elaeisguineensis*) stand out in the production of BioQAV because they have high levels of fatty oils in their composition whose carbon chain is similar to fossil kerosene [4].

The use of biokerosene in the aviation sector also produces carbon dioxide, but its production cycle reduces this environmental damage with the absorption of CO₂ by plants in the process of photosynthesis. This process allows the biofuel to be carbon neutral during its life cycle. Another advantage of biofuel is that it is sulfur-free, which is one of the elements responsible for acid rain [5,6].

In Brazil, there is the National Biokerosene Program created in 2009 by law No. 3213/2009, which establishes the development of research on renewable fuels from biomass without changes in the technologies established in the turbine engines. That is, the composition of these fuels must guarantee the safety of the aviation system [7,8]. In this context, it is necessary to develop analytical methods that provide fast and reliable responses to assess the biokerosene content in the kerosene/biokerosene mixture.

For quantification purposes, several studies report the use of spectroscopic techniques associated with multivariate regression by partial least squares (PLS). Some of these works present the quantification of adulterants in biodiesel / diesel mixtures, as is the case of the authors De Souza et al. [9], proposes a new method for the quantitative analysis of soybean oil and sunflower oil as adulterants in oil of extra virgin flaxseed, using MIR spectroscopy associated with PLS. The models obtained were built according to the standard ASTM E1655-05, having obtained acceptable error values and good correlation between the measured and predicted values of the calibration and prediction sets. Máquina et al. [10], developed two methodologies to quantify and classify the cotton biodiesel content in mixtures with diesel, using MIR spectroscopy associated with PLS and Discriminant Analysis by Partial Minimum Squares (PLS-DA) methods. The PLS model developed to determine the biodiesel content showed a good fit, with a correlation coefficient of the measured and predicted values exceeding 0.99.

However, there are no reports of published studies on the quantification of the biokerosene content of palm kernel in the kerosene/biokerosene mixture, using rapid and non-destructive analytical techniques associated with PLS regression methods. Thus, the present work aims to develop an analytical methodology that can be used by inspection bodies to quantify the biokerosene content of palm kernel in mixtures with kerosene, using MIR spectroscopy associated with the PLS chemometric method.

MATERIALS AND METHODS

Sample preparation

In this study, palm kernel biokerosene provided by LABIO (Biofuels Laboratory of the Chemistry Institute) of the Federal University of Uberlândia was used. Transpetro S/A (Brazil) supplied the pure kerosene used to prepare the kerosene/palm kernel biokerosene mixture, adding biokerosene to the kerosene in a concentration ranging from 1.00 to 70.00% (v/v). For the construction of the PLS model, 45 samples were used in the calibration set and 30 samples in the forecast set.

Acquisition of spectral data

The MIR spectra were acquired using a PerkinElmer Spectrum Two spectrometer (PerkinElmer, Waltham, MA, USA) equipped with a HATR accessory and ZnSe crystal (Pike Technologies). The spectra were recorded in the range of 4000–600 cm^{-1} with a 4 cm^{-1} resolution and were acquired using 16 scans for each of the quintuplicates.

Chemometrics analysis

MATLAB software, version R2018b (Mathworks, Inc.) and PLS_Toolbox, version 8.9.1 (Eigenvector Research) were used to pre-process the data and develop the PLS model. To perform the multivariate procedures, the MIR spectra data were organized in an ordered array of rows and columns, constituting a matrix \mathbf{X} , where each row corresponds one sample “ m ” and each column corresponds to one variable “ n ”, in the which $m = 1, 2, 3, \dots, 75$ and $n = 1, 2, 3, \dots, 1306$.

The spectral baseline was corrected in the bands of 4000–3100 and 2500–1850 cm^{-1} to minimize undesirable systematic variations in the data and then they were centered on the mean. Finally, a \mathbf{Y} matrix was created containing concentration values (from 1.00 to 70.00% (v / v)) of the samples.

In the construction of the PLS model, matrix \mathbf{X} is correlated with matrix \mathbf{Y} through mathematical operations to obtain Latent Variables (LVs) and Regression Coefficients, used to achieve the maximum covariance between the spectra and concentrations of the species of interest and determine the concentration value of each spectral profile, respectively [9]. The purpose of this process is to find a small number of relevant factors that are predictive of \mathbf{Y} and that use \mathbf{X} efficiently. To do this, matrix \mathbf{X} is decomposed into a set of orthogonal factors that are used to adjust \mathbf{Y} , according to Equation 1. Matrix \mathbf{Y} is decomposed according to Equation 2: [10]

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}_X = \sum t_h p_h^T + \mathbf{E}_X \quad (1)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{E}_Y = \sum u_h q_h^T + \mathbf{E}_Y \quad (2)$$

where, \mathbf{X} and \mathbf{Y} are matrices that contain instrumental measurement and response (concentration) data, respectively; \mathbf{T} and \mathbf{U} are the ($m \times A$) scores for the two data matrix; \mathbf{P} and \mathbf{Q} are the ($n \times A$) loadings respective, h is the latent variable number (LV), \mathbf{E}_X and \mathbf{E}_Y are the respective residues. However, if block \mathbf{Y} is univariate, weight \mathbf{Q} is set to 1 [11].

The linear relationship between the two matrices is established by the correlation of the \mathbf{X} and \mathbf{Y} scores for each LV (h), according to Equation 3: [12]

$$\mathbf{U}_h = \mathbf{b}_h \mathbf{T}_h + \mathbf{E} \quad (3)$$

where, \mathbf{U}_h is a matrix that contains the properties of all samples (concentration), \mathbf{b}_h is a vector that contains the model parameters, \mathbf{T}_h is a matrix that contains instrumental measurement data (spectrum) for the calibration samples and \mathbf{E} is a matrix that represents noise.

The PLS model was constructed by minimizing the waste matrices \mathbf{E}_X and \mathbf{E}_Y , at the same time, thereby obtaining a linear relationship between \mathbf{t} and \mathbf{u} , through the method of cross-validation by the venetian blind criterion, with 14 blocks of data division and with two samples per block. The number of LVs that provide the lowest RMSECV was selected, following the requirements of ASTM E1655-55 (2012) [13].

Once completed, the model was validated based on the determination of the following figures of merit: selectivity, sensitivity, analytical sensitivity, limit of detection, limit of quantification and test for systematic error (bias and t_{bias}), according to the equations presented in Table I [14–17]. The fit of the PLS model

was analyzed based on the correlation between the current and predicted concentration values of the calibration and forecast sets.

Table I. Equations used to calculate the merit figures of the PLS model*

Figure of merit	Equation
Selectivity	$S\hat{E}L_i = \frac{n\hat{a}s_i}{\ x_i\ }$
Sensitivity	$S\hat{E}N = \frac{1}{\ b_k\ }$
Analytical sensitivity	$\gamma = \frac{S\hat{E}N}{\ \delta_x\ }$
Limit of detection	$LD = 3.3\delta_x \frac{1}{S\hat{E}N}$
Limit of quantification	$LQ = 10\delta_x \frac{1}{S\hat{E}N}$
<i>bias</i>	$bias = \frac{\sum_{i=1}^{n_{val}}(y_i - \hat{y}_i)}{n_{val}}$
SDV	$SDV = \sqrt{\frac{\sum[(y_i - \hat{y}_i) - bias]^2}{n_{val} - 1}}$
t_{bias}	$t_{bias} = \frac{ bias \sqrt{n_{val}}}{SDV}$

*where, \hat{y}_i is the predicted value using the PLS model; y_i is the reference value for the sample i ; n_{val} is the number of validation samples; $n\hat{a}s_i$ is the norm of the NAS vector and $\|x_i\|$ is the norm for each spectrum; b_k is the vector of the final regression coefficients, δ_x is the standard deviation value of 9 kerosene samples.

RESULTS AND DISCUSSION

Spectrum MIR

Figure 1 shows the MIR spectra of kerosene/palm kernel biokerosene mixtures in the concentration range from 1.00 to 70.00% (v/v) before being pre-processed, where noise is observed in the region from 600 cm^{-1} to 700 cm^{-1} and low signal variation in the region from 3100 cm^{-1} to 4000 cm^{-1} .

Figure 2 shows the MIR spectra of (a) kerosene (b) palm kernel biokerosene and Figure 3 shows the MIR spectra of kerosene/palm kernel biokerosene mixtures in the concentration range from 1.00 to 70.00% (v/v) pre-processed. In these spectra, similarities are noted in the following significant absorption bands: at 2950 cm^{-1} – attributed to the asymmetric stretching vibrations of the C–H bond of methyl groups ($-CH_3$); at 2923 cm^{-1} – attributed to the asymmetric stretching vibration of the C–H bond of methylene group ($-CH_2$); at 2851 cm^{-1} – attributed to the symmetrical stretching vibration of the C–H bond of methylene group ($-CH_2$); at 1451 cm^{-1} – attributed to the asymmetric angular deformations of the C–H bond of methyl groups ($-CH_3$) and at 1379 cm^{-1} – attributed to the symmetric angular deformations of the C–H bond of methyl groups ($-CH_3$) [18,19].

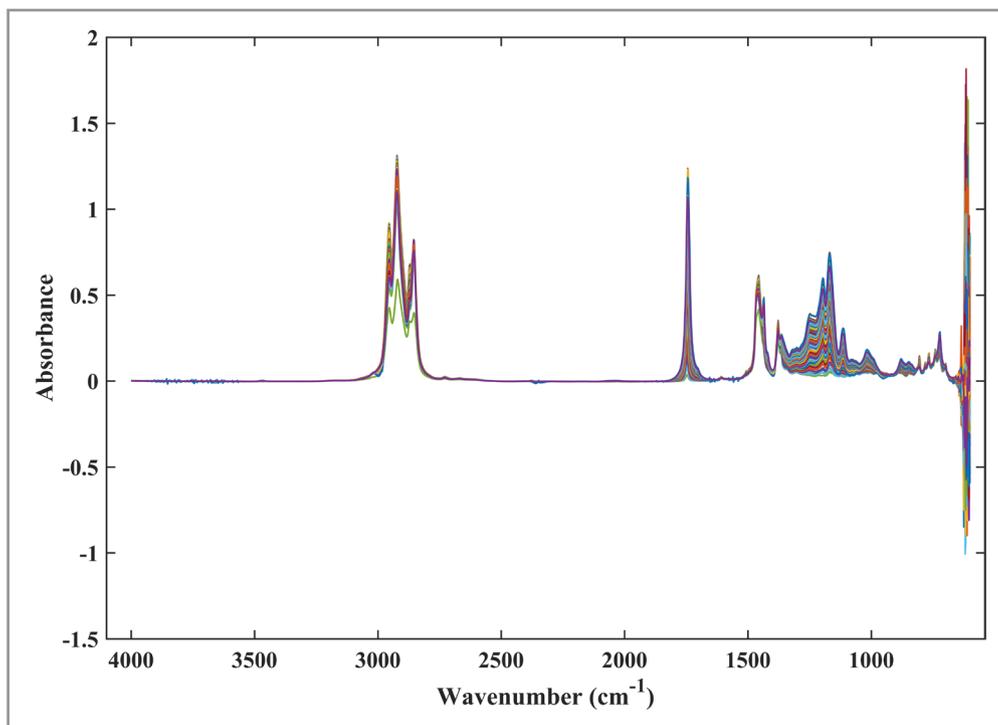


Figure 1. Unprocessed MIR spectra of kerosene/palm kernel biokerosene mixtures in the concentration range from 1.00 to 70.00% (v/v).

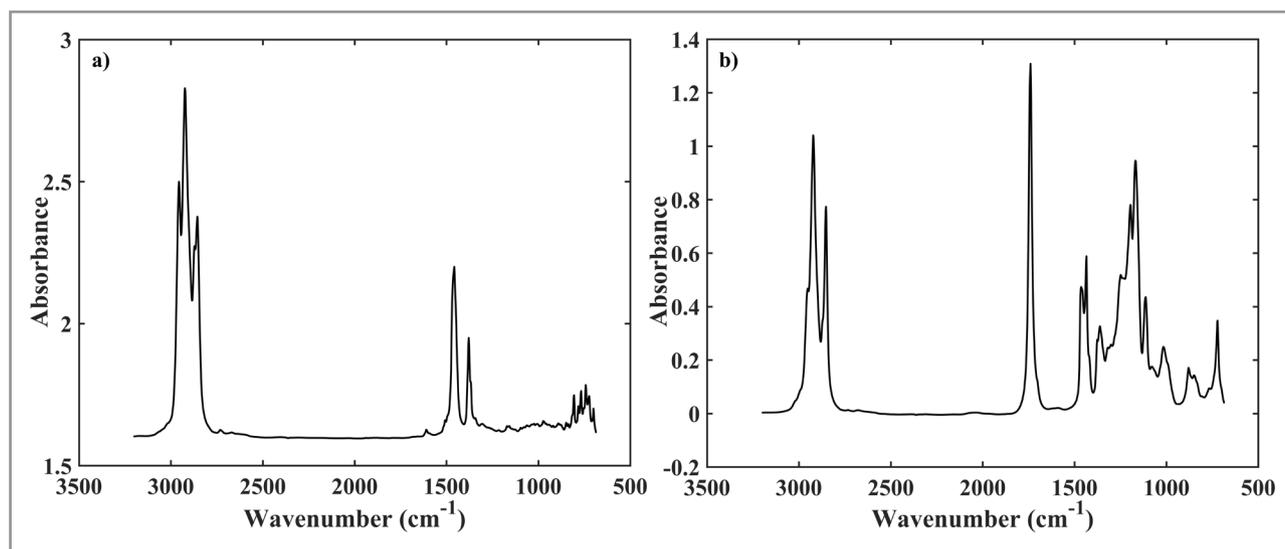


Figure 2. MIR spectra of (a) kerosene (b) palm kernel biokerosene.

However, the spectrum of kerosene differs from the spectrum of bio-kerosene, mainly in the characteristic bands of bio-kerosene, found in the 1744 cm^{-1} spectral regions, attributed to $\nu\text{C=O}$ of acylglycerols; from 1200 to 1119 cm^{-1} , attributed to the stretching vibrations of the C–O group bond in esters of type O=C-(OR) [18,19]. That is why, when mixed (Figure 3), these characteristic bands are evident. However, it is difficult to attribute each spectral profile to a specific concentration based on visual analysis. To this end, the use of chemometric tools is essential.

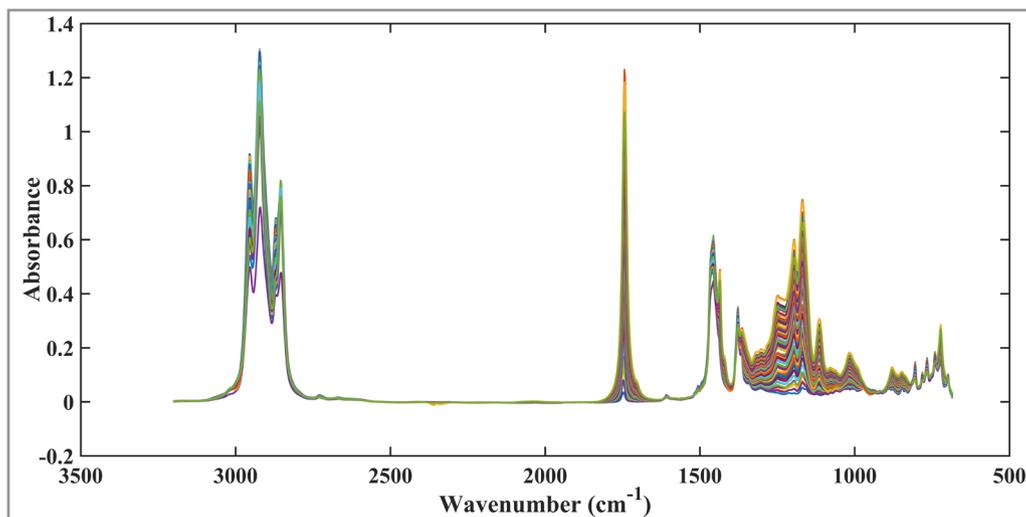


Figure 3. Pre-processed MIR spectra of kerosene/palm kernel biokerosene mixtures in the concentration range from 1.00 to 70.00% (v/v).

PLS Model

Figure 4 shows: a) the root-mean-square error of cross-validation (RMSECV) and root-mean-square error of Calibration (RMSEC); b) the accumulated variance captured in blocks **X** and **Y** obtained as a function of the amount of the Latent Variable (LV) chosen for construction of the PLS model. In this figure, it can be seen that the 5 LVs chosen to build the PLS model provide less errors and greater captured variance, that is, they represent 99.99% and 99.98% of the variance explained in blocks **X** and **Y**, respectively.

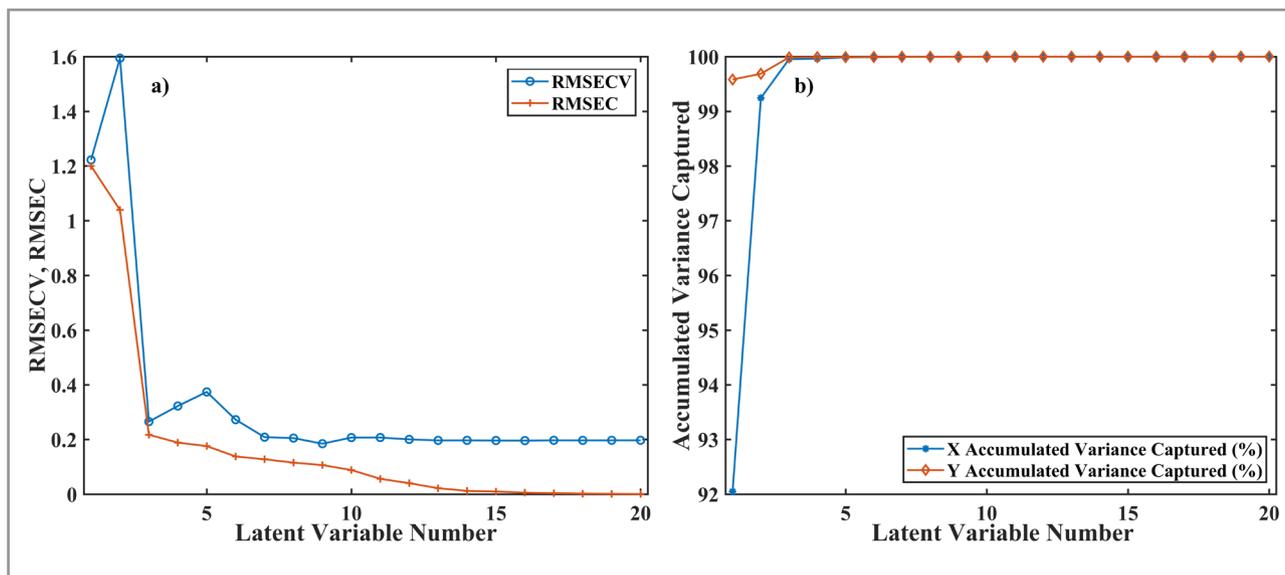


Figure 4. (a) Root-mean-square error of cross-validation (RMSECV) and root-mean-square error of Calibration (RMSEC), (b) Accumulated variance captured in blocks **X** and **Y**.

Table II presents results of the parameters and figures of merit, calculated for the PLS model, where it is observed that the values of root-mean-square error of cross-validation (RMSECV), root-mean-square error of calibration (RMSEC) and root-mean-square error of prediction (RMSEP) were in agreement because the RMSEP was higher than RMSECV and RMSEC. In addition, the RMSEP value is considered acceptable

according to the Brazilian standard ABNT NBR 15568 because it is less than 1% [20] and are close to the values obtained in some quantification models published in the literature [21,22]. It is also observed that a good agreement between the parameters is evident because RMSEC was lower than RMSEP. The calculated t_{bias} value was lower than $t_{critical}$, which means that there are no systematic errors in the model.

The value of the LQ obtained in the model is higher than the LD, demonstrating a good agreement between the parameters because the lowest concentration of the substance of interest that is measured with a maximum uncertainty of 10% is within what can be detected. The sensitivity expresses an increase in the signal fraction when the concentration of the analyte of interest has a high value for one unit, however, its value was estimated at 0.871. The analytical sensitivity of the model is relatively high, showing that the influence of residues in the prediction of unknown samples is low. The inverse of the analytical sensitivity value shows that the PLS model is able to distinguish differences between samples with a variation of 0.052. The selectivity was good, which means that the PLS model did not present significant overlap of the interference signal with the analyte.

Table II. Results of the accuracy parameters and figures of merit

	Figure of Merit	Value
Accuracy	RMSEC (% v/v)	0.175
	RMSECV (% v/v)	0.373
	RMSEP (% v/v)	0.708
	Analytical sensitivity/ % (v/v)	19.20
	Inverse of analytical sensitivity $/(v/v)^{-1}$	0.052
	Sensitivity/ % (v/v) ⁻¹	0.871
	Selectivity	0.375
	Limit of detection / % (v/v)	0.171
	Limit of quantification / % (v/v)	0.521
	bias	0.118
	SDV	0.364
	t_{bias}	1.782
	$t_{critico}$	2.051

The evaluation of the fit of the PLS model by correlating the measured and predicted values of the calibration and prediction sets is shown in Figure 5. In this figure, it can be seen that the PLS model showed a good fit because a high correlation between the two values was evident, with the correlation coefficient (R) exceeding 0.99. Figure 6, illustrates the plot of the real concentration versus absolute errors, showing uniform distribution in a horizontal range. Ideally, the residuals should show random behavior and constant variance in the concentration range, as seen in the calibration set samples. Although the residuals from the prediction set show a trend, the predicted concentrations are not severely affected.

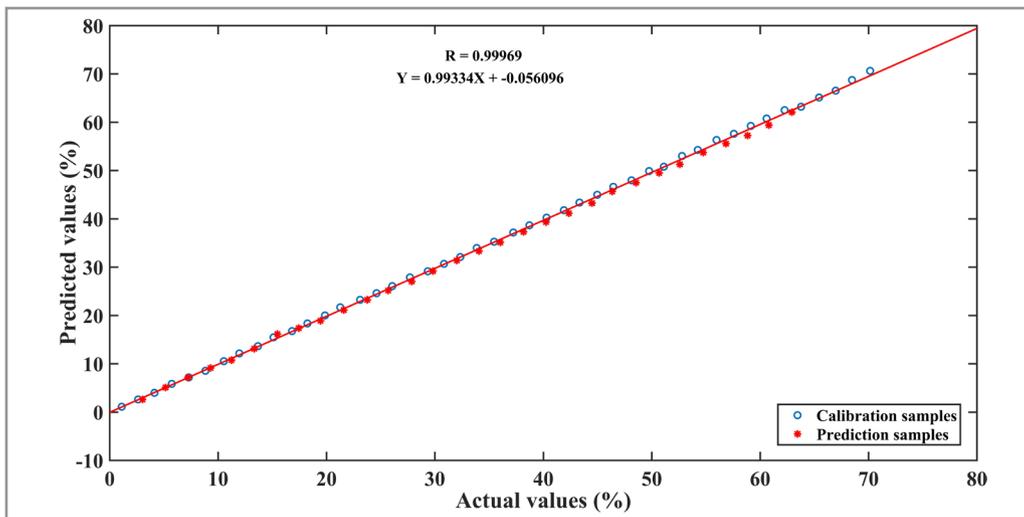


Figure 5. Fit of the PLS model through the actual values (experimental concentrations) versus predicted values of the calibration and forecast sets.

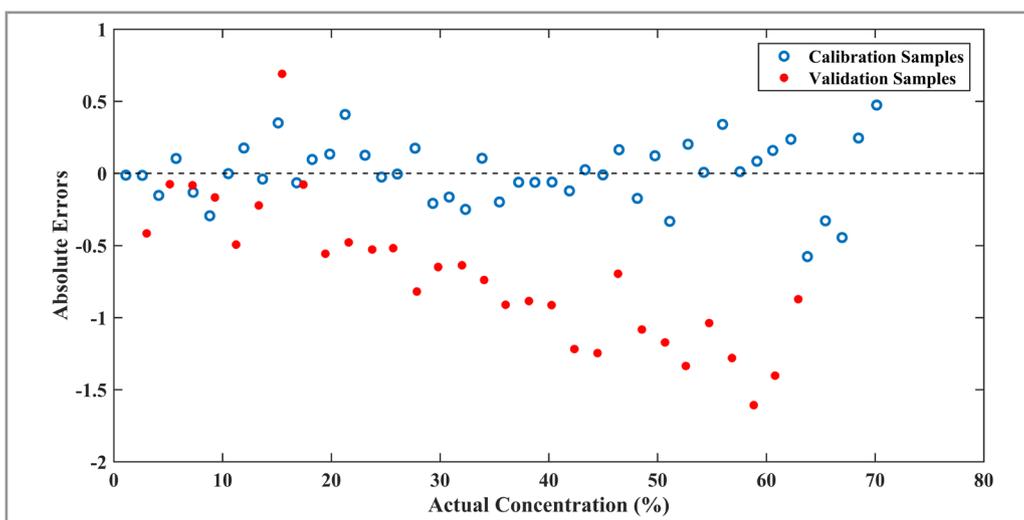


Figure 6. Residues of the PLS model for the calibration and prediction set

Figure 7 shows the weights of the latent variables of the PLS model, where it can be seen that the LV1 describes the variables of the characteristic bands of the biokerosene, with 92.05% of the explained variance. The remaining latent variables describe the variables of the common bands, with 7.94% of the explained variance.

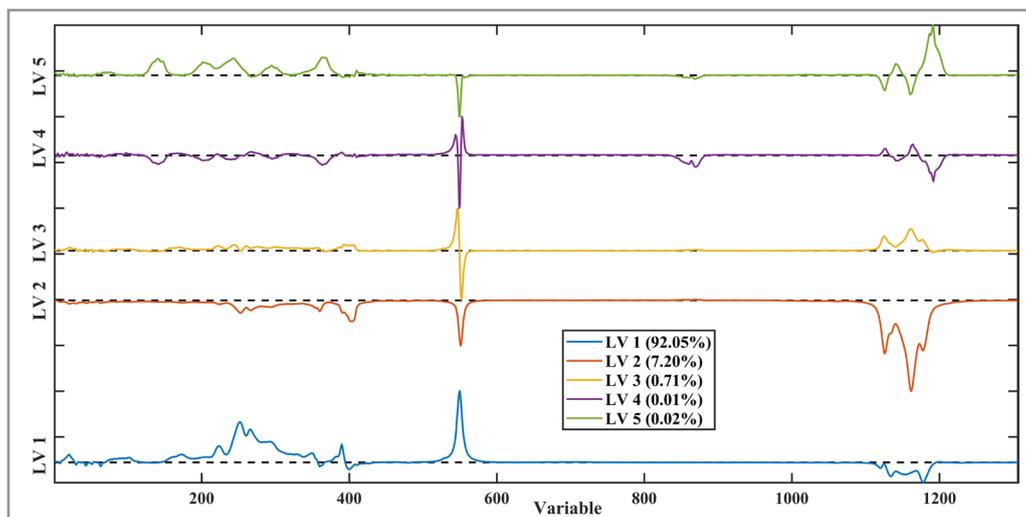


Figure 7. Plot of the loadings of LV1, LV2, LV3, LV4 and LV5 versus variables for the application of the PLS model.

CONCLUSIONS

The application of multivariate regression by PLS in the MIR spectra of kerosene/palm kernel biokerosene mixtures allowed the development of a methodology to quantify the content of this biokerosene. The efficiency of this methodology was analyzed based on the figures of merit and the fit of the model. The results of the figures of merit were in agreement with the requirements established in the standard ASTM E1655-05. The model fit showed a high correlation between actual and predicted concentration values of the calibration and prediction sets, with a correlation coefficient exceeding 0.99 and with relatively low errors.

These results demonstrate that the developed methodology has the potential to be explored by regulatory agencies to monitor the content of biokerosene in a mixture with kerosene, because it is relatively low cost and allows for quick, direct and in situ analyzes.

Conflicts of interest

Regarding conflicts of interest and on behalf of all authors, I declare that for this manuscript there are no conflicts of interest.

Acknowledgements

The authors acknowledge the partnership program PEC-PG (Process 190279/2014-1) for financial support; Transpetro, S/A. for the supply of pure kerosene and Biofuels Laboratory of the Federal University of Uberlândia for the supply of palm kernel biokerosene.

REFERENCES

1. Grampella, M.; Martini, G.; Scotti, D.; Zambon, G. *Transp. Res. Part A Policy Pract.*, **2016**, 92, pp 310–325 (<https://doi.org/10.1016/j.tra.2016.06.013>).
2. Hofer, C.; Dresner, M. E.; Windle, R. J. *Transp. Res. Part D Transp. Environ.*, **2010**, 15 (1), pp 37–45 (<https://doi.org/10.1016/j.trd.2009.07.001>).
3. Geller, H. S. *Revolução Energética: Políticas Para Um Futuro Sustentável*; Relume Dumará, Rio de Janeiro, **2003**.

4. Biocombustíveis Aeronáuticos: Progressos e Desafios. Série Documentos Técnicos do Centro de Gestão e Estudos Estratégicos (CGEE). **2010**, Nº 08. Available at: https://www.cgge.org.br/documents/10195/734063/biocombustiveis_aeronauticos_24012011_9559.pdf/378f8f90-fa5c-4e0c-aad7-7adcbf607063?version=1.5 [Accessed January 11, 2021].
5. Llamas, A.; Al-Lal, A. M.; Hernandez, M.; Lapuerta, M.; Canoira, L. *Energy and Fuels*, **2012**, 26 (9), pp 5968–5976 (<https://doi.org/10.1021/ef300927q>).
6. Reynol, F. *Inovação Uniemp*, **2007**, 3 (3). Available at: http://inovacao.scielo.br/scielo.php?script=sci_arttext&pid=S1808-23942007000300011&lng=en&nrm=iso&tling=pt [Accessed January 11, 2021].
7. Bonassa, G.; Schneider, L. T.; Damaris, K.; Frigo, D. A.; Severo, F.; Lins, M. A.; Frigo, E. P. *Rev. Bras. Energias Renov.*, **2014**, 3, pp 97–106.
8. Baroutian, S.; Aroua, M. K.; Raman, A. A. A.; Shafie, A.; Ismail, R. A.; Hamdan, H. *J. Taiwan Inst. Chem. Eng.*, **2013**, 44 (6), pp 911–916 (<https://doi.org/10.1016/j.jtice.2013.02.007>).
9. Máquina, A. D. V.; Siteo, B. V.; Buiatte, J. E.; Santos, D. Q.; Neto, W. B. *Fuel*, **2019**, 237, pp 373–379 (<https://doi.org/10.1016/j.fuel.2018.10.011>).
10. Máquina, A. D. V.; Souza, L. M.; Buiatte, J. E.; Santos, D. Q.; Neto, W. B. *Energy and Fuels*, **2017**, 31 (1), pp 571–577 (<https://doi.org/10.1021/acs.energyfuels.6b02079>).
11. Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; Jong, S. De; Lewi, P. J.; Smeyers, J. *Handbook of Chemometrics and Qualimetrics, Volume 20B*. Elsevier Science, Amsterdam, **1998**.
12. Ferreira, M. M. C. *Quimiometria: Conceitos, Métodos e Aplicações*; Editora da Unicamp, Campinas, SP, **2015** (<https://doi.org/10.7476/9788526814714>).
13. American Society for Testing and Materials (ASTM). Standard E1655-05. *Standard Practices for Infrared Multivariate Quantitative Analysis*. United States, **2012**, pp 1–29 (<https://doi.org/10.1520/E1655-05R12>).
14. Ferreira, M. H.; Braga, J. W. B.; Sena, M. M. *Microchem. J.*, **2013**, 109, pp 158–164 (<https://doi.org/10.1016/j.microc.2012.03.008>).
15. Valderrama, P.; Braga, J. W. B.; Poppi, R. J. *Quim. Nova*, **2009**, 32 (5), pp 1278–1287 (<https://doi.org/10.1590/S0100-40422009000500034>).
16. Silva, M. A. M.; Ferreira, M. H.; Braga, J. W. B.; Sena, M. M. *Talanta*, **2012**, 89, pp 342–351 (<https://doi.org/10.1016/j.talanta.2011.12.039>).
17. De Carvalho Rocha, W. F.; Nogueira, R.; Vaz, B. G. V. *J. Chemom.*, **2012**, 26 (8–9), pp 456–461 (<https://doi.org/10.1002/cem.2420>).
18. Barbosa, L. C. A. *Espectroscopia no Infravermelho*, 1th Ed. Editora da Universidade Federal de Viçosa (UFV), Viçosa, MG, Brazil, **2007**.
19. Silverstein, R. M.; Webster, F. X.; Kiemle, D. J.; Bryce, D. L. *Spectrometric Identification of Organic Compounds*, 8th Ed. John Wiley & Sons, Nova Jersey, **2014**.
20. Associação Brasileira de Normas Técnicas. ABNT NBR-15568. *Biodiesel – Determinação do Teor de Biodiesel em Óleo Diesel por Espectroscopia na região do Infravermelho Médio*. Brazil, **2008**.
21. Shi, T.; Zhu, M. T.; Zhou, X. Y.; Huo, X.; Long, Y.; Zeng, X. Z.; Chen, Y. *Food Chem.*, **2019**, 287, pp 46–54 (<https://doi.org/10.1016/j.foodchem.2019.02.072>).
22. Singh, K.; Kumar, S. P.; Blümich, B. *Fuel*, **2019**, 243, pp 192–201 (<https://doi.org/10.1016/j.fuel.2019.01.084>).