Highlights

- Elemental analysis, water content and heating value were modeled.
- Predictions were checked to prove the robustness of the models
- Good predictions were obtained for all properties with $R^2_{\text{pre}} \geq 0.82$
- Differences between predictions and experimental results were not statistically significant

Prediction of elemental composition, water content and heating value of upgraded biofuel from the catalytic cracking of pyrolysis bio-oil vapors by infrared spectroscopy and partial least square regression models.

A. Veses, J.M. López, T. García, M.S. Callén*

Instituto de Carboquímica (ICB-CSIC), C/ Miguel Luesma Castán, 50018 Zaragoza, Spain.

Abstract

The elemental composition, heating value and water content, are important properties to be characterized for pyrolysis bio-oils, providing information on their quality. These properties are mainly determined according to ASTM standards by using three different analytical techniques requiring time and cost. This research was focused on a simple method to determine the weight content of carbon, hydrogen, oxygen and water as well as the heating value, by Fourier transform infrared spectroscopy (FT-IR) using models based on partial least squares regressions (PLS). Samples were classified into two sets according to Kennard-Stone algorithm. The first set of samples was used to develop the calibration models for each physical parameter, where the number of latent variables was determined by full cross validation procedure. The second set of samples was employed as an external prediction set, assessing the quality of the models. External predictions confirmed that robust models were developed since elemental analysis, heating value and water content of the upgraded biofuels obtained by the catalytic cracking of pyrolysis bio-oil vapors could be determined with good predictive ability with a root mean square error of prediction of carbon content =0.963 wt.% ($R^2=0.836$, range=70.9-78.8 wt.%), hydrogen content = 0.101 wt.% ($R^2=0.815$, range=8.01-8.75 wt.%), oxygen content=0.910 wt.% ($R^2=0.873$, range=12.3-20.9 wt.%), water content=0.416 wt.% ($R^2=0.829$, range= 2.79-7.24 wt.%) and heating value=0.539 MJ

kg⁻¹(R²=0.874, range= 31.9-36.9 MJ kg⁻¹) by chemometric tools joined to medium infrared spectrum.

Keywords: upgraded biofuels; infrared spectroscopy; elemental composition; heating value; water content; partial least square regressions.

Corresponding author: M^a Soledad Callén. E-mail: marisol@icb.csic.es. Phone number: +34 976 733977, Fax number: +34 976733318

1. Introduction

The applications of biomass for power and heat generation are becoming of increasing interest due to its renewable character and the CO₂ neutral overall balance in the atmosphere [1]. There are different classes of biomass for biofuel production. Whilst sugars and vegetable edible oils are already implemented to produce the so-called “first generation” biofuels, lignocellulosic pyrolysis bio-oils are postulated for the production of the “second generation” biofuels. It is generally accepted that first generation biofuels have some important disadvantages not only related to the food versus fuel debate [2], but also to their negative environmental impact and carbon balance [3]. These issues convert the second generation biofuels as potential alternatives.

In general, pyrolysis bio-oils can be considered as a complex mixture of chemicals containing varied functional groups [4, 5] with numerous undesirable properties for fuel applications like high water content, low stability, and remarkable acidity [6, 7]. Since these bio-oils cannot be directly employed in the present infrastructures for power and heat generation, bio-oils need to be upgraded by using catalytic processes, which enhance the bio-oil quality by reducing the oxygen content as well as improving other biofuel properties. As for other conventional fuels, their physical and chemical properties provide information regarding the biofuel quality, stability and potential uses

119
120
121 51 as alternative fuels in conventional infrastructures [8]. These properties include water
122
123 52 content, pH, total acid number (TAN), heating value, elemental composition, density
124
125 53 and viscosity, among others.
126
127 54 Fuel properties are commonly measured by standards procedures ensuring quality and
128
129 55 comparability of results. These measurements require time and costly equipment to
130
131 56 perform the different analyses. One of the most important characteristics of upgraded
132
133 57 biofuels is the heating or calorific value, which is mainly measured by a bomb
134
135 58 calorimeter requiring a tedious, time-consuming process and advanced technical skills
136
137 59 in the handling of equipment [9-11]. The heating value can also be calculated by an
138
139 60 empirical model requiring the elemental analysis, in particular, the carbon, hydrogen
140
141 61 and oxygen content [12-14]. Precisely, the elemental analysis is one of the main
142
143 62 properties characterizing biofuels. In general, the procedure uses an analyzer which is
144
145 63 based on the flash combustion of dried and ground samples in combination with
146
147 64 chromatographic methods and a thermal conductivity detector (TCD) [15]. Other
148
149 65 important biofuel property is related to the water or moisture content, which requires its
150
151 66 determination with Karl Fischer titration [16-17]. This is a chemical analysis procedure
152
153 67 based on the oxidation of sulfur dioxide by iodine in a methanolic hydroxide solution.
154
155 68 This titration can be performed volumetrically or coulometrically, requiring an adaption
156
157 69 of the working method to the specific sample. Therefore, there is a request for general
158
159 70 analytical tools that allow predicting sample properties in a faster and easier way than
160
161 71 conventional techniques. One of these predicting tools is based on infrared spectroscopy
162
163 72 (IR) [18-21] combined with chemometric techniques [22-24]. Basically, IR needs trace
164
165 73 sample and it is a non-destructive analysis with minimal sample preparation, providing
166
167 74 fast and simultaneous analysis of multiple components from the same sample in a single
168
169 75 instrumental analysis. These advantages allow IR joint to chemometric tools to be
170
171
172
173
174
175
176
177

successfully employed for control and monitoring operations in several sectors such as agricultural, food and pharmaceutical industries [25, 26]. IR spectroscopy and multivariate analysis have also shown to be an effective, rapid and accurate method for predicting the composition of different biomass and biodiesel samples, as recently shown in a comprehensive review by Chadwick et al. [27]. Additionally, heating value, moisture content and ash content are other important parameters that have been successfully determined for biomass quality [28]. The determination of hydrogen content and heating value in diesel fuels has effectively been accomplished by Al-Ghouti et al. [29] using Fourier transform infrared spectroscopy (FTIR) and PLS models. IR coupled to PLS have also been applied to determine the total acid number (TAN) in mineral gas engine oils [30]. The results from these studies have shown that IR combined to chemometric tools may indeed be used to predict all of the above properties at laboratory scale and for industrial samples, with the potential for online monitoring and quality control purposes for the final product. However, there is not any attempt to develop chemometric models to predict the main properties of pyrolysis lignocellulosic biofuels in the literature.

Herein, the use of FTIR as a fast tool to predict different properties of biofuels obtained from the catalytic upgrading of pyrolysis lignocellulosic bio-oils has been investigated for the first time. Water content, elemental analysis (%C, %H and %O) and heating value were predicted using chemometric tools based on linear regression models such as partial least squares regressions, providing fast and cost-effective methods that reduce time and expenses when compared to traditional methodologies.

2. Materials and methods

2.1. Upgraded bio-oils

The biomass used in this work was forest pine woodchips (*Pinus halepensis*), including bark, obtained from the North-East area of Spain. Raw bio-oil used in the catalytic upgrading tests was obtained from the catalytic pyrolysis of biomass in an auger reactor operated at 450°C with N₂ as the carrier gas. The residence time of solids in the reactor was 7 min. Tests were carried out with a catalyst (90% CaO, Calcinor) to biomass mass ratio of 1:6 and catalysts were diluted with sand, keeping a (sand + catalyst) to biomass mass ratio of 3:1. The liquid obtained was separated in two different layers by centrifugation at 1500 rpm for 1 hour: the upper or aqueous layer and the bottom or organic layer, which was selected as raw bio-oil. This organic layer is the most interesting to be upgraded for its potential use as biofuel due to its properties, both less water and oxygen content, and higher heating value in comparison with the aqueous layer. A detailed description of the experimental system used can be found elsewhere [31-34].

Upgraded biofuels were produced *via* catalytic upgrading of raw bio-oil vapours (see Supplementary Material) in a fixed-bed reactor at 450 °C over several commercial and lab-made zeolites, resulting in 70 biofuel samples. Details about the different zeolite samples can be found in the Supplementary material (Tables S1-S2, Supplementary material).

2.2. Biofuel characterization

The complete characterization of the organic phase for the upgraded biofuels was carried out to determine elemental composition (C, H, O, S, N weight percent, as received) (Carlo-Erba EA1108 according to UNE-EN ISO 16948:2015) [35], higher heating value or HHV (IKA C-2000, according to UNE 164001 EX) [36] and water content by Karl Fischer titration (Crison Titromatic KF, according to ASTM E203-16;

weight percent) [17]. In addition, FTIR spectra of biofuels were obtained with a Bruker Vertex 70 spectrometer corresponding to medium-infrared (MIR) covering the range 4000-400 cm^{-1} . 50 μl liquid samples were applied on disposable real crystal IR sample cards (KBr sample support substrate) with 15 mm of aperture provided by International Crystal Laboratories. A Globar source and a deuterated L-alanine doped triglycine sulphate detector were used. Each spectrum was acquired in transmission mode using 32 scans and a spectral resolution of 2 cm^{-1} .

2.3. Chemometric methods and data analysis

2.3.1. Data pre-treatment for FTIR

Spectral data were imported from OPUS into the Unscrambler X 10.3 (Camo Inc., Oslo, Norway) and they were converted to absorbance. All other spectral processing and chemometric tools were performed using this program. The original data set contained 3735 variables (each one of the wavelengths obtained in the IR spectra), covering the MIR region. Before building optimal PLS calibration models for each property, data were centered by calculating the average value for each variable, and then, subtracting this from each of the original variables. All variables were weighted at constant value equal to 1 in order to avoid the influence of the different scales used for the variables. Although different ranges of wavelengths were tested in order to assess the accuracy of the PLS regression models obtained for each parameter under study, the best results for the five bio-oil properties led to select the full MIR range: 4000-400 cm^{-1} . Additionally, combinations of three different spectral treatments were evaluated. The pre-processing treatments eliminate or reduce the impact of the non-relevant spectral information and often lead to simpler and more robust calibration models. These pre-processing treatments were: a) Normalization to the highest peak, which attempts to correct for

scaling differences, b) Standard Normal Variate (SNV), a row-oriented transformation which centers and scales individual spectra, avoiding scatter effects from spectral data and c) first derivative using Savitzky-Golay algorithms (SGolay) (polynomial order 2, window with 11 points), which removes additive effects from the spectra, such as baseline offsets or features. In this research, initial spectral treatments were limited to the application of a SNV spectral treatment [37] (Fig. 1) since optimal results were obtained, the same than in a previous work [38].

2.3.2. Partial least square regression models

Five PLS regression models were built, one for each biofuel property, where the variables of the IR spectra comprised between 4000 and 400 cm^{-1} for the biofuel oils were used as predictor data (X) and the elemental analysis, water content and heating value as the response variable (Y).

Samples were classified into two sets: half of samples (35 samples) were used to develop the calibration models for each biofuel property, where full cross validation (FCV), also known as leave-one-out cross validation, was selected to determine the optimum number of latent variables (LV). The other half of the samples (35 samples) was used as external prediction set to evaluate the predictive capacity of the model between the predicted values from the calibration model and those from the reference methods, using samples not used in the calibration model. Selection of the samples was performed according to Kennard-Stone algorithm [39]. In general, external validation is preferentially required [40] since full cross validation usually provides a too optimistic assessment of the model predictive capacity.

The calibration models were performed by considering the correlation between the IR spectral data and the five biofuel properties determined by the corresponding

reference methods. The predictive capacity of the models obtained for each biofuel property were performed by assessing the following criteria: the determination coefficients (R^2), which provide information regarding the fit of the model, the root mean square errors (RMSE), for both calibration and prediction regression models [41] and the bias, which can be interpreted as the average difference between the reference value and the predicted value in the prediction set and refers to systematic errors. Additionally, the ratio of performance to deviation (RPD) was calculated for each model, which was determined by the coefficient between the standard deviation of the input data and the standard error of the prediction (SD/SEP) [42, 43]. RMSE also enabled to identify the presence of outliers that did not fit the model. A normal probability plot of the Y-residuals of the model showing a fairly straight line with all values within “ $\pm 3SD$ ” was also used to identify potential outliers. RMSE, SEP and bias provide information about the accuracy, the precision and the trueness of the model, respectively. The higher the RPD and the R^2 values and the lower the RMSE, SEP and bias values, the better the predictive capacity of the model.

A t-test for validation samples at 95% confidence level and degrees of freedom equal to the number of prediction samples was carried out in order to establish whether there were significant differences between the experimental and modelled results for each biofuel property (Table S3, Supplementary material).

As estimation of the model uncertainty, a confidence interval (C.I.) was calculated by the Eq. (1) when bias is small and standard deviation is unknown:

$$C.I. = \bar{y}_i \pm z \, SEP \approx \bar{y}_i \pm z \, RMSE_{Pred} \quad (1)$$

where $\bar{y}$ is the value predicted by the model and z-value must be replaced by a value from a Student’s t-distribution table which is estimated as 2 (1.960 is used for the 95%

confidence interval) when N is high. This is an important measure, indicating the uncertainty of the model.

The precision of each model, both absolute precision and relative precision, were calculated according to the Eq (2) and (3) considering that bias values were close to zero:

$$\text{Absolute Precision} = z \cdot RMSE_{Pred} \quad (2)$$

$$\text{Relative Precision (\%)} = \frac{z \cdot RMSE_{Pred}}{\bar{y}} * 100 \quad (3)$$

According to Saeys et al. [44], RPD and R² greater than 3.0 or 0.90, respectively, are indicative of an excellent prediction. Values from 2.5 to 3.0 (RPD) and 0.82 to 0.90 (R²) denote good prediction. RPD values between 2.0 and 2.5 and R² values in the range from 0.66 to 0.81 indicate approximate quantitative predictions. Finally, unsuccessful predictions have RPD and R² values lower than 1.5 or 0.50, respectively, whilst poor prediction have RPD and R² values of 1.5-2.0 and 0.50-0.65, respectively.

3. Results and discussion

3.1. Upgraded biofuel properties

A total of 70 upgraded biofuels (Table S4, Supplementary material) were produced. For these samples, C, H and O elemental composition (wt.%), water content (wt.%) and heating value (MJ kg⁻¹) were determined according to reference methods mentioned in Section 2.2. Biofuel characterization. Statistical descriptive of these properties for the calibration and prediction sets are shown in Table 1.

Comparable values were observed for both groups of samples, confirming the reliability of the selection algorithm. It is worth commenting that although C, H and O were the main biofuel components, small amounts of nitrogen with a narrow compositional range from 0.1 to 0.5 were also detected. According to the literature [45], standard deviation

should be considered as a critical parameter in the development of any prediction model since a small value together with a high standard error of the reference method hinder the development of a good prediction model. Unfortunately, the presence of some volatile components in the biofuels led to very relevant uncertainties during the determination of the nitrogen content for these samples. This fact together with a low variation for this parameter determined that nitrogen content would not be further considered for modeling.

3.2. FTIR results

Basically, infrared energy can be absorbed at the atomic level originating alterations in rotational-vibrational movements and giving rise to IR bands only if they cause a change in the dipole moment of the molecule. These vibrational movements can be classified as stretching (changes in the bond length) and bending (changes in bond angles) and they can only reflect information of molecular functional groups but not identify a specific chemical compound [46].

The FT-MIR spectra of the 70 biofuel samples are shown in Fig. 1 indicating a varied organic composition. There were several broad peaks at different intensities related with different functionalities in the biofuel and arising from activating molecular vibrations. Two main zones were particularly remarkable in the spectra: 3600-2750 cm^{-1} and 1800-700 cm^{-1} . The first zone included the highest and broadest vibrations between 3600 and 3000 cm^{-1} , indicating O-H stretching due to the presence of water and different functional groups also related to phenols, alcohols [47], carboxylic acids, carbohydrates and amino acids. Unsaturated and aromatic C-H bonds absorb from 3100 to 3000 cm^{-1} and aliphatic C-H bond absorbance bands were detected from 3000 to 2850 cm^{-1} [48]. A broad and intense peak around 1700 cm^{-1} indicates C=O stretching vibration of free carbonyl groups of aldehydes, ketones, carboxylic acids and esters and subsequent

bands around 1650-1550 cm^{-1} represent C=C stretching vibrations caused by aliphatic or aromatic structures. The spectral region of 1488-1400 cm^{-1} contains bands in the O-H bending region, which were most probably associated with alcohols and carboxylic acids [49]. The peak around 1260 cm^{-1} indicated C-O stretching peak (in acids, esters, ethers and alcohols) and the peaks around 1030 cm^{-1} were associated with aromatic C-H in plane bending. Aromatic rings could also be determined by the presence of C-H out of plane bands [50] between 840 and 700 cm^{-1} .

All the spectral information regarding different bonds seemed to be quite promising in order to provide information regarding the biofuel properties: elemental composition and water content. Because the heating value of biofuel oils is associated with the content of C, H, O, N and S and their covalent carbon bonds, all these bands could also be useful in determining this property.

3.3. Prediction of carbon content

A PLS calibration model was built from the set of samples selected by the Kennard-Stone algorithm. It was observed that no more than two LV were required to explain most of the variance (71.3% of the X variance and 86% of the Y variance)(Table 2). The most important variables for the carbon content were shown in Fig. 2a) where the weighted regression coefficients are shown for each LV as a function of the wavenumbers. The highest coefficients involved in the modeling of this property were included in the regions: 3080-2823; 2739-2723; 1809-1690; 1526-1506; 1263-1034 and 888-685 cm^{-1} that, as commented in section 3.2 FTIR results, were related with C-H bonds, C=O and C-O stretching vibrations and C-H out of plane bands mainly caused by aliphatic and aromatic structure.

273 An assessment of the PLS regression prediction was carried out by means of the R^2_{Pred}
 274 and the $\text{RMSE}_{\text{Pred}}$ parameters (Table 2), which were 0.800 and 1.040 wt.%, respectively,
 275 pointing out that the model could predict the carbon content of the upgraded biofuel
 276 samples with approximate accuracy. $\text{RMSE}_{\text{Pred}}$ and SEP were quite similar and the bias
 277 value was close to zero, indicating that systematic errors were negligible. Accordingly,
 278 RPD values higher than 2.0 and differences between R^2_{Cal} , R^2_{FCV} and R^2_{Pred} lower than
 279 0.2 [41] were obtained for the calibration and the prediction, showing that a robust
 280 model was obtained with no overfitting.
 281 However, it should be highlighted that the removal of one outlier in the prediction set
 282 drove to a better model performance. Thus, a slight improvement in the R^2_{Pred} values
 283 (Fig. 3) was obtained. This value increased from 0.800 to 0.836 whilst a decrease on the
 284 $\text{RMSE}_{\text{Pred}}=0.963$ wt.% was observed.
 285 Relative precision was also assessed for the prediction set based on one specific value
 286 for the prediction set ($\text{RMSE}_{\text{Pred}}$), obtaining values lower than 3% for the carbon
 287 content (Table 3). Despite the elemental analysis is an analytical technique providing
 288 quite good results regarding precision, these results seem to be quite promising taking
 289 into account that infrared with PLS is a rapid, non-destructive, reliable and low-cost
 290 technique. However, it is remarkable to say that the use of only one value for $\text{RMSE}_{\text{Pred}}$
 291 has some limitations and it can overestimate the model precision. In fact, when
 292 repeatability of one specific sample was carried out by plotting five IR spectra, it was
 293 observed that the relative precision based on the standard deviation was 0.11% (Table
 294 3). Finally, a statistical analysis comparing the prediction results obtained by the PLS
 295 regression model to those obtained by the standard analytical method was carried out
 296 using a paired t-test at 95% level of confidence (Table S3, Supplementary material). T-
 297 test results showed that differences between real and prediction values were not

statistically significant and null hypothesis could not be rejected. Therefore, it could be concluded that a PLS regression model could be accurately used to predict the carbon content in upgraded biofuels obtained from the catalytic cracking of lignocellulosic pyrolysis biofuels for a specific confidence interval between 73.74 and 77.51 wt.% (Table 2).

3.4. Prediction of the hydrogen content

Regarding the hydrogen content, four LV were necessary to explain 88% of the X variance and 83% of the Y variance in the calibration model. The weight regression coefficients in the calibration model are shown in Fig. 2b) indicating that the most interesting wavebands were the regions: 2985-2820 cm^{-1} , 1524-1499 cm^{-1} and 1286-1195 cm^{-1} associated with C-H, C=C and O-H stretching and bending vibrations. The PLS regression model indicated that approximate predictions could be obtained for this property, since $R^2_{\text{Pred}}=0.757$, $\text{RMSE}_{\text{Pred}}=0.117$ wt.% and RPD values higher than 2 were obtained (Table 2). The removal of two outliers (Fig. 3) allowed increasing the R^2_{Pred} value up to 0.815, again leading to better predictions and higher precision ($\text{RMSE}_{\text{Pred}}=0.101$ wt.%). The relative uncertainties were also similar to those found for carbon content prediction and lower than 3% (Table 3). Again, no statistically significant differences were found between those values obtained by the standard analytical method and the PLS regression model (Table S3, Supplementary material), allowing the prediction of the hydrogen content with a reasonable degree of accuracy by non-destructive techniques. Bias value was also close to zero, indicating similarity between the $\text{RMSE}_{\text{Pred}}$ and SEP and discarding systematic errors. In fact, very good precision was obtained for the hydrogen content with values lower than 3% and 0.19%, based on the $\text{RMSE}_{\text{Pred}}$ and the standard deviation, respectively, for a specific sample.

Herein, despite hydrogen content showed a narrow compositional range, this property could be predicted with approximate accuracy within a confidence interval from 8.23 to 8.63 wt.% by our standard reference method, as shown in Table 2, allowing the development of a consistent PLS regression model.

3.5. Prediction of the oxygen content

For this property, the calibration model was able to explain with two LV 71% and 87% of the X and Y variances, respectively. The most important wavelengths associated with this property were reflected in Fig. 2c) indicating that the regression coefficients were affected by different spectral regions: 3600-2750 and 1800-500 cm^{-1} . The region between 3400-3680 cm^{-1} , corresponding to O-H stretching vibrations, was mainly caused by acid and/or alcohol structures. The bands between 3080-2821 cm^{-1} affected negatively over the model and C-C stretching vibrations caused by aliphatic/aromatic structure showed peaks at 1500-1527 cm^{-1} and C-O peaks at 1644-1800 cm^{-1} possibly due to carboxylic acid/ester, aldehydes/ketone groups. At 1000-1300 cm^{-1} , C-O groups from acids, esters, ethers and alcohols and aromatic C-H bonds would be contributing to the oxygen content modeling.

The oxygen content was the biofuel compositional element showing one of the best PLS regression model performances regarding the fitting of the model. $R^2_{\text{Pred}} = 0.812$, $\text{RMSE}_{\text{Pred}} = 1.087$ wt.% and $\text{RPD} = 2.48$ values were attained. According to these values, approximate predictions could be obtained for this property. The removal of two outliers (Fig. 3) allowed increasing the R^2_{Pred} up to 0.873, leading to good predictions. Accordingly, a decrease of $\text{RMSE}_{\text{Pred}}$ (0.910 wt.%) was obtained, probing that the accuracy and the precision of the method was increased. Relative uncertainties were lower than 15% (Table 3) for the prediction set, whilst an excellent precision was

obtained based on the repeatability of IR spectra (0.48%). The confidence interval for the prediction presented a wide compositional range from 14.00 to 17.57 wt.%. Results of comparing the predicted data to those values obtained by the standard analytical method did not provide any significant difference, pointing out that PLS regression models could be used to foresee the oxygen content (Table S3, Supplementary material) with notable accuracy.

3.6. Prediction of the water content

As shown in Table 2, two LV were required in the calibration model to explain 72% of the X variance and 64% of the Y variance. The weight regression coefficients are reflected in Fig. 2d) where the most important variables were associated basically with different C-H, O-H, C=O, C-O functional groups. The highest coefficients were obtained at 3519, 2870, 1516, 1271, 1034, 752 and 692 cm^{-1} , respectively. The water content was the property showing the lowest accuracy for the prediction model with R^2_{Pred} and $\text{RMSE}_{\text{Pred}}$ values of 0.705 and 0.577 wt.%, respectively (Table 2). These parameters together with the narrow range of data modeled, pointed out that water content could be only predicted with approximate quantitative accuracy. For this property, relative uncertainties were the highest, about 26% (Table 3) within a confidence interval for the prediction ranging from 3.36 to 5.62 wt.%. However, the removal of two outliers in the prediction set significantly increased the R^2_{Pred} value up to 0.829 and good predictions were then obtained (Fig. 3). A RPD value higher than 3.0 was obtained and, again, no statistically significant differences were found between the predictions with the PLS regression model and the experimental values, obtained by Karl Fischer valorization (Table S3, Supplementary material).. Herein, for the first time in the literature, it is shown that chemometric tools joint to medium-IR allow predicting

the water content in upgraded lignocellulosic biofuels, as a fast screening tool instead of the standard Karl Fischer valorization method.

3.7. Prediction of the heating value

The calibration model established two LV to explain 71% of the X variance and 90% of the Y variance. The most important variables describing the heating value were shown in Fig. 2e) where the weight regression coefficients were plot over wavenumbers. It is remarkable the similarity of the important wavelengths (3505, 2922, 2855, 1712, 1516, 1271, 1035, 752 and 692 cm^{-1}) with the ones obtained for the carbon content (Fig. 2a)) indicating that the heating value is quite depending on the carbon content for these biofuel samples, as expected.

Heating value was the fuel property showing the best fitting parameters, since R^2_{Pred} and $\text{RMSE}_{\text{Pred}}$ values were 0.836 and 0.599 MJ kg^{-1} , respectively. A good prediction was obtained, which was notably improved after the removal of one outlier (Fig. 3), increasing the $R^2_{\text{Pred}} = 0.874$ and decreasing the $\text{RMSE}_{\text{Pred}}=0.539 \text{ MJ kg}^{-1}$. Relative uncertainties were also lower than 4% with values of 0.10% for the repeatability of IR spectra (Table 3). RPD value was 2.56 and as for the other modeled properties, no statistically significant differences were found between the prediction by the external validation and the real values obtained by the standard analytical technique (calorimetric bomb). Summarizing, this PLS regression model could be successfully used to predict the heating value with notable accuracy, especially taking into account that this is the first time that the heating value of upgraded biofuels has been predicted based on chemometric tools, using PLS regression and MIR spectra. These results are quite promising to advance in the use of non-destructive and sensing techniques that involve more rapid analysis and less inversion costs than other traditional procedures in

the characterization of second generation biofuels obtained from the catalytic upgrading of pyrolysis lignocellulosic bio-oils.

4. Conclusions

The search of techniques that allow determining specific properties of second generation biofuels produced from lignocellulosic biomass in a fast way are of great concern. In this research, the coupling of mid-infrared with chemometric tools based on partial least square regressions have proven to be a promising tool in order to predict quantitatively some relevant properties (C, O, H, water content and heating value) associated to upgraded biofuels obtained from the catalytic cracking of pyrolysis lignocellulosic bio-oils. Robust models were built for each property, always achieving good predictions, $R^2_{\text{Pred}} \geq 0.82$. In this way, an easy assessment of these properties could be performed by determining the MIR spectra of upgraded biofuels, avoiding more tedious traditional procedures that require more time and expenses. These results could be considered in a first approach as a promising and useful tool in order to perform quality controls of second generation biofuels production technologies. More research should be conducted to increase the number of upgraded bio-oils and to study how the application of other catalysts could affect the prediction of the PLS models for these five properties.

Appendix A. Supplementary material

Supplementary material related to this article can be found at:

Acknowledgements

Authors would like to thank the Spanish MINECO and European Union FEDER funds for supporting this work (projects CTQ2012-37984-C02-01 and ENE2015-68320-R).

References

- [1] A. Friedl, E. Padouvas, H. Rotter, K. Varmuza, Prediction of heating values of biomass fuel from elemental composition, *Analytica Chimica Acta* 544 (2005) 191-198.
- [2] W. Laursen, Students take a green initiative, *Chem. Eng.* (2006) 32-34.
- [3] S.N. Naik, V.V. Goud, P.K. Rout, A.K. Dalai, Production of first and second generation biofuels: A comprehensive review. *Renew. Sustainable Energy Rev.* 14(2) (2010) 578-597.
- [4] A.V. Bridgwater, D. Meier, D. Radlein, An overview of fast pyrolysis of biomass, *Org. Geochem.* 30 (1999) 1479.
- [5] D. Meier, O. Faix, State of the art of applied fast pyrolysis of lignocellulosic materials—a review, *Biosource Technol.* 68(1) (1999) 71–77.
- [6] D. Mohan, Jr. C.U. Pittman, P.H. Steele, Pyrolysis of wood/biomass for bio-oil: A critical review, *Energy Fuels* 20 (2006) 848-889.
- [7] L. Qiang, L. Wen-Zhi, Z. Xi-Feng, Overview of fuel properties of biomass fast pyrolysis oils, *Energy Convs. Manage.* 50 (2009) 1376-1383.
- [8] I. Graça, J.M. Lopes, H.S. Cerqueira, M.F. Ribeiro, Bio-oils upgrading for second generation biofuels, *Ind. Eng. Chem. Res.* 52 (2013) 275-287.
- [9] S. Kathiravale, M.N. Mudhd-Yunus, K. Sopian, A.H. Samsuddin, R.A. Rahman, Modeling the heating value of municipal solid waste, *Fuel* 82(9) (2003) 1119-1125.
- [10] CEN/TS 14918: 2005. Solid biofuels. Method for the determination of calorific value.
- [11] R.S. Jessup, Precise Measurement of Heat Combustion with a Bomb Calorimeter, U.S. Bureau of Standards Monograph No. 7, US Government Printing Office, Washington, D.C., 1970.

- 449 [12] C. Sheng, J.L.T. Azevedo, Estimating the higher heating value of biomass fuels
450 from basic analysis data, *Biomass Bioenergy* 28 (2005) 499–507.
- 451 [13] R. Saidur, E.A. Abdelaziz, A. Demirbas, M.S. Hossain, S. Mekhilef, A review on
452 biomass as a fuel for boilers, *Renew. Sustain. Energy Rev.* 15 (2011) 2262–89.
- 453 [14] J.F. Saldarriaga, R. Aguado, A. Pablos, M. Amutio, M. Olazar, J. Bilbao, Fast
454 characterization of biomass fuels by thermogravimetric analysis (TGA), *Fuel* 140
455 (2015) 744–51.
- 456 [15] CEN/TS 15104: 2005. Solid biofuels. Determination of total content of carbon,
457 hydrogen and nitrogen. Instrumental methods.
- 458 [16] K. Fischer, Karl (1935). Neues verfahren zur maßanalytischen bestimmung des
459 wassergehaltes von flüssigkeiten und festen körpern. *Angew. Chem.* 48 (26): 394–396,
460 doi:10.1002/ange.19350482605.
- 461 [17] ASTM E203 – 16. Standard test method for water using volumetric Karl Fischer
462 titration.
- 463 [18] D.B. Dahlberg, S.M. Lee, S.J. Wenger, J.A. Vargo, Classification of vegetable oils
464 by FT-IR. *Appl. Spectrosc.* 51 (1997) 1118–1124.
- 465 [19] Z. Droussi, V. D’Orazio, M.R. Provenzano, M. Hafidi et al. Ouattmane Study of the
466 biodegradation and transformation of olive-mill residues during composting using FTIR
467 spectroscopy and differential scanning calorimetry, *J.Hazard. Mater.* 164 (2009) 1281–
468 1285.
- 469 [20] W. Terouzi, M. De Luca, A. Bolli, A. Oussama, M. Patumi, G. Ioele, G. Ragno, A
470 discriminant method for classification of Moroccan olive varieties by using direct FT-IR
471 analysis of the mesocarp section, *Vib. Spectro.* 56 (2011) 123–128.

- [21] W. Terouzi, Z. Ait Yacine, F. Kzaiber, A. Boulli, A. Oussama, Kernel olive: Marker of tracing the variety by using the FTIR spectra and chemometrics. *Int. J. Chemtech. Appl.* 3(4) (2014) 1-13.
- [22] C.E. Miller, Chemometrics for on-line spectroscopy applications theory and practice, *J. Chemom.* 14(5–6) (2000) 513–528.
- [23] K.H. Esbensen, Multivariate data analysis-in practice. An introduction to multivariate data analysis and experimental desing. Oslo: CAMO; 2000.
- [24] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S.D.E. Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier, Amsterdam, 1998, p 88.
- [25] P.R. Wiley, G.J. Tanner, P.M. Chandler, R.S. Anderssen, Molecular classification of barbey (*Hordeum vulgare* L.) mutants using derivative NIR spectroscopy, *J. Agric. Food Chem.* 57 (2009) 4042-4050.
- [26] S. Vonhoff, J. Condliffe, H. Schiffter, Implementation of an FTIR calibration curve for fast and objective determination of changes in protein secondary structure during formulation development, *J. Pharm. Biomed. Anal.* 51 (2010) 39-45.
- [27] D.T. Chadwick, K.P. McDonnell, L.P. Brennan, C.C. Fagan, C.D. Everard, Evaluation of infrared techniques for the assessment of biomass and biofuel quality parameters and conversion technology processes: A review, *Renew. Sustainable Energy Rev.* 30 (2014) 672-681.
- [28] G.D. Gillespie, C.D. Everard, K.P. McDonnell, Prediction of biomass pellet quality indices using near infrared spectroscopy, *Energy* 80 (2015) 582-588.
- [29] Al-Ghouti M, Al-Degs Y, Mustafa F. Determination of hydrogen content, gross heat of combustion and net heat of combustion of diesel fuel using FTIR spectroscopy and multivariate calibration. *Fuel* 2010; 89: 193-01.

- 497 [30] Y. Felkel, N. Dörr, F. Glatz, K. Varmuza, Determination of the total acid number
498 (TAN) of used gas engine oils by IR and chemometrics applying a combined strategy
499 for variable selection, *Chemometr. Intell. Lab.* 101 (2010) 14–22.
- 500 [31] A. Veses, M. Aznar, J.M. López, M.S. Callén, R. Murillo, T. García, Production of
501 upgraded bio-oils by biomass catalytic pyrolysis in an auger reactor using low cost
502 materials, *Fuel* 141 (2015) 17-22.
- 503 [32] A. Veses, B. Puértolas, M.S. Callén, T. García, Catalytic upgrading of biomass
504 derived pyrolysis vapors over metal-loaded ZSM-5 zeolites: Effect of different metal
505 cations on the bio-oil final properties, *Microp. Mesop. Mat.* 209 (2015) 189-196.
- 506 [33] A. Veses, B. Puértolas, J.M. López, M.S. Callén, B. Solsona, T. García, Promoting
507 Deoxygenation of Bio-Oil by Metal-Loaded Hierarchical ZSM-5 Zeolites, *ACS*
508 *Sustainable Chem. Eng.* 4(3) (2016) 1653–1660.
- 509 [34] A. Veses, M. Aznar, I. Martínez, J.D. Martínez, J.M. López, M.V. Navarro, M.S.
510 Callén, R. Murillo, T. García, Catalytic pyrolysis of Wood biomass in an auger reactor
511 using calcium-based catalysts, *Bioresour. Technol.* 162 (2014) 250-258.
- 512 [35] UNE-EN ISO 16948:2015. Solid biofuels. Determination of total content of
513 carbon, hydrogen and nitrogen.
- 514 [36] UNE 164001 EX. Solid Biofuels. Method for the determination of calorific value.
- 515 [37] V.R. Nadadoor, H. De la Hoz Siegler, S.L. Shah, W.C. McCaffrey, A. Ben-Zvi,
516 Online sensor for monitoring a microalgal bioreactor system using supportvector
517 regression, *Chemometr. Intell. Lab.* 110 (2012) 38-48.
- 518 [38] T. García, A. Veses, J.M. López, B. Puértolas, J. Pérez-Ramírez, M.S. Callén,
519 Determining bio-oil composition via chemometric tools based on infrared spectroscopy,
520 *ACS Sustain. Chem. Eng.* 5(10) (2017) 8710-8719.

- [39] R.W. Kennard, L.A. Stone, Computer aided design of experiments, *Technometrics* 11(1) (1969) 137-148.
- [40] G.C.S. Smith, S.R. Seaman, A.M. Wood, P. Royston, I.R. White, Correcting for optimistic prediction in small data sets, *Am. J. Epidemiol.* 180(3) (2014) 318-324.
- [41] D.L. Sills, J.M. Gossett, Using FTIR spectroscopy to model alkaline pretreatment and enzymatic saccharification of six lignocellulosic biomasses, *Biotechnol. Bioeng.* 109(4) (2012) 894-903.
- [42] P.C. Williams, D.C. Sobering, How do we do it: a brief summary of the methods we use in developing near infrared calibrations, in: A.M.C. Davies, P. Williams (Eds.), *Near Infrared Spectroscopy: the future waves*, NIR Publications, Chichester, UK, 1996, pp. 185-188.
- [43] T. Fearn, Assessing calibrations: SEP, RPD, RER, and R², *NIR News* 13 (2002) 12-14.
- [44] W. Saeys, A.M. Mouazen, H. Ramon, Potential for onsite and online analysis of pig manure using visible and near infrared spectroscopy, *Biosyst. Eng.* 91(2005) 393-402.
- [45] J. Mata Sánchez, J.A. Pérez Jiménez, M.J. Díaz Villanueva, A. Serrano, N. Núñez, J. López Giménez, Assessment of near infrared spectroscopy for energetic characterization of olive byproducts, *Renew. Energy* 74 (2015) 599-605.
- [46] C.C. Sousa, J.A. Lopes, Infrared spectroscopy detection coupled to chemometrics to characterize foodborne pathogens at a subspecies level, in: D. Granato (Ed.), *Mathematical and Statistical Methods in Food Science and Technology*, John Wiley & Sons, Ltd, Chichester, UK, 2014, DOI: 10.1002/9781118434635.ch20.

- 1299
1300
1301 544 [47] M.I. Schnitzer, C.M. Monreal, G.A. Facey, P.B. Fransham, The conversion of
1302
1303 545 chicken manure to biooil by fast pyrolysis I, Analysis of biooils by FTIR and NMR
1304
1305 546 spectroscopy, J. Environ. Sci. Health, Part B 42 (2007) 71–77.
1306
1307
1308 547 [48] C. Coury, A.M. Dillner, A method to quantify organic functional groups and
1309
1310 548 inorganic compounds in ambient aerosols using attenuated total reflectance FTIR
1311
1312 549 spectroscopy and multivariate chemometric techniques, Atmos. Environ. 42 (2008)
1313
1314 550 5923–5932.
1315
1316 551 [49] M. Asadieraghi, W.M.A. Wan Dau, In-depth investigation on thermochemical
1317
1318 552 characteristics of palm oil biomasses as potential biofuel sources, J. Anal. Appl. Pyrol.
1319
1320 553 115 (2015) 379–391.
1321
1322 554 [50] P.K. Kanaujia, Y.K. Sharma, M.O. Garg, D. Tripathi, R. Singh, Review of
1323
1324 555 analytical strategies in the production and upgrading of bio-oils derived from
1325
1326 556 lignocellulosic biomass, J. Anal. Appl. Pyrol. 105 (2014) 55–74.
1327
1328
1329 557
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357

Table 1. Statistical descriptive of the different properties for the set of samples (N= number of samples; Cal.= Calibration; Pred.= Prediction, KS= Kennard-Stone algorithm).

	Samples	N	C (wt.%)	O (wt.%)	H (wt.%)	N (wt.%)	Water Content (wt.%)	Heating value (MJ kg ⁻¹)
	Cal. KS	35						
Mean			75.1	16.4	8.38	0.2	4.72	34.5
SD _{Cal.}			2.5	2.7	0.25	0.1	1.26	1.6
Maxim			78.2	21.0	8.72	0.5	7.04	36.4
Minim			70.8	13.2	7.89	0.1	3.11	31.6
	Pred. KS	35						
Mean			75.4	16.0	8.40	0.2	4.45	34.8
SD _{Pred.}			2.3	2.4	0.22	0.1	1.04	1.4
Maxim			78.8	20.9	8.75	0.4	7.24	36.9
Minim			70.9	12.3	8.01	0.1	2.79	31.9

N= number of samples; SD_{Cal.}= Standard deviation of the calibration set; SD_{Pred.}= Standard deviation of the prediction set, wt.%=weight percent; Maxim=maximum value and Minim=minimum value.

Table 2. Summary of the partial least squares regressions models developed to predict ultimate analysis (wt.%), water content (wt.%) and heating value (MJ kg⁻¹) (N= number of samples, LV=latent variables, RMSE=root mean square error, SEC/SEP=standard error of calibration or prediction, RPD=ratio of performance to deviation, R²= determination coefficient, C.I.=confidence interval).

Dependent	Wavelength	N	Model	LV	Variance	Bias	RMSE	SEC/SEP	RPD	R ²	C.I.	
			type									
C	4000-400	35	Cal.	2	71.3	86.0	-1.1 x10 ⁻²	0.934	0.947	2.60	0.860	73.19-76.85
		35	Pred.	2	-	-	2.2 x10 ⁻¹	1.040	1.123	2.19	0.800	73.60-77.68
		34	Pred.	2			3.0 x10 ⁻¹	0.963	1.110	2.22	0.836	73.74-77.51
H	4000-400	35	Cal.	4	87.7	82.5	1.8 x10 ⁻³	0.104	0.106	2.32	0.825	8.18-8.59
		35	Pred.	4	-	-	1.3 x10 ⁻²	0.117	0.120	2.05	0.757	8.19-8.65
		33	Pred.	4			3.3 x10 ⁻²	0.101	0.119	2.06	0.815	8.23-8.63
O	4000-400	35	Cal.	2	71.2	87.0	1.1 x10 ⁻²	0.971	0.986	2.70	0.870	14.47-18.28
		35	Pred.	2	-	-	-2.6 x10 ⁻¹	1.087	1.071	2.48	0.812	13.58-17.83
		33	Pred.	2			2.8 x10 ⁻¹	0.910	0.877	3.40	0.873	14.00-17.57
Water	4000-400	35	Cal.	2	72.0	63.7	-9.5 x10 ⁻³	0.770	0.782	1.61	0.637	3.21-6.23
		35	Pred.	2	-	-	4.4 x10 ⁻²	0.577	0.590	2.14	0.705	3.36-5.62
		33	Pred.	2			5.6 x10 ⁻²	0.416	0.418	3.02	0.829	3.60-5.23
Heating value	4000-400	35	Cal.	2	71.2	90.2	-3.3 x10 ⁻³	0.502	0.509	3.11	0.902	33.49-35.46
		35	Pred.	2	-	-	1.1 x10 ⁻¹	0.599	0.639	2.47	0.836	33.69-36.04
		34	Pred.	2			1.6 x10 ⁻¹	0.539	0.618	2.56	0.874	33.80-35.91

Table 3 Relative precision (%) at 95 % confidence level in the calibration and prediction sets obtained with the PLS regression models.

Properties	Relative precision
Calibration N=35	
C	2.46
H	2.41
O	11.67
Water content	27.82
Heating value	2.93
Prediction N=35	
C	2.69 (0.11)
H	2.78 (0.19)
O	13.85 (0.48)
Water content	25.70 (1.90)
Heating value	3.44 (0.10)

In brackets the relative precision at 95% confidence level based on the repeatability of five IR spectra of the same sample calculated according to standard deviation.

Fig.1 FTIR spectrum of the calibration and prediction biofuel samples corresponding to the MIR region: 4000-400 cm^{-1} without any spectral treatment (left side) and after standard normal variate spectral treatment (SNV)(right side).

a)

b)

c)

d)

e)

Fig. 2. Weighted regression coefficients of the FTIR-PLS calibration model of: a) C, b) H, c) O, d) water content and e) heating value in upgraded biofuel samples.

Fig. 3 Plot of the prediction set (left graph for all samples, N=35 samples, right graph after removing outliers) for the different properties obtained by PLS regression model.