Determining Bio-oil Composition *via* Chemometric Tools based on Infrared Spectroscopy

Tomás García[†], *Alberto Veses*[†], *José Manuel López*[†], *Begoña Puértolas*[‡], *Javier Pérez-Ramírez*[‡] *and María Soledad Callén*[†]*

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

[‡] Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zurich, Vladimir-Prelog-Weg 1, 8093 Zurich, Switzerland.

*Corresponding author. E-mail: marisol@icb.csic.es

ABSTRACT

The development of rapid and accurate techniques to predict the composition of crude bio-oils obtained *via* the pyrolysis of lignocellulosic biomass is a prerequisite for their industrial implementation. Here, we demonstrate the potential of the Fourier Transform Infrared Spectroscopy to replace the gas chromatography-mass spectrometry (GC-MS) in determining the compositional groups of bio-oils. Using mid-infrared spectroscopic technique as predictor, chemometric tools based on partial least squares regression models were contrasted with GC-MS results to foresee the various families of organic compounds. A broad data set, consisting of more than one hundred samples obtained from the thermal and catalytic pyrolysis of woody biomass and from the upgrading of bio-oil vapors by catalytic cracking over zeolites and metal oxides was used. The

applicability of the developed model was assessed by external validation using the Kennard-Stone algorithm, showing that more than 90 wt% of the bio-oil composition was accurately determined. These results pave the path for the on-line monitoring of the forthcoming manufacture system of second-generation biofuels through rapid and cost-effective characterization of the pyrolysis bio-oils, thus enabling industrial producers to make timely decisions.

KEYWORDS

Chemometric tools, Bio-oil composition, Fourier Transform Infrared Spectroscopy, Gas Chromatography-Mass Spectrometry, Partial least squares regression

INTRODUCTION

The global dependency and depletion of fossil fuels together with the associated impact related to their use raise a serious threat for the environment, driving the scientific community towards the search of sustainable solutions for the production of fuels and commodity chemicals. Lignocellulosic biomass is deemed as the only carbon-based renewable source on Earth with real potential to partly replace fossil fuels. Among different conversion technologies for the valorization of lignocellulosic biomass, pyrolysis is generally considered as the most simple and economic platform to obtain alternative liquid bio-fuels. Pyrolysis oils accomplish a complex mixture of organic compounds,[1] which are characterized by their high water content (up to 30% by weight), strong acidity associated to the presence of phenolic species and carboxylic acids, and low calorific value (*ca.* 16-19 MJ kg^{-1}).[2] The high oxygen content of crude bio-oil confers poor stability, corrosiveness and prevents the blending with commercial fuels,[3] which restricts its direct application in current infrastructures.[4] To overcome this

limitation, the catalytic upgrading into value-added chemicals with relative similar characteristics than those obtained from fossil fuels,[3] but also to the so-called second-generation bio-fuels is prerequisite. Therefore, this strategy involves different steps, thus demanding a fast, cost-efficient and non-destructive technique for the real-time characterization of pyrolysis bio-oils.

Traditionally, GC-MS has been applied to determine the composition of pyrolysis bio-oils[5-7] in both qualitative and quantitative manners.[8] The identified components are subsequentially grouped into different organic families, which generally include phenols, acids, aldehydes, ketones, furans, cyclic hydrocarbons (HCs), aromatics and esters. However, the lack of a standard GC-MS analytical method jointly with the fact that this is a time and cost-demanding technique, drive the development of alternative strategies.

In this regard, infrared spectroscopy (IR) associated to rotational-vibrational structure within a molecule emerges as a fast, cost-efficient and non-destructive technique to determine bio-oil composition. IR has been extensively used for the qualitative characterization of bio-oil samples obtained from the pyrolysis of lignocellulosic biomass[9-11] and allows the identification of the different organic groups in the bio-oil without providing quantitative information on the composition. A step forward in the capacities of this analytical technique has been recently achieved in combination with chemometric tools, demonstrating to be an effective, rapid and accurate methodology for predicting the composition as well as different physical properties of solid biomass feedstocks.[12] In this sense, this methodology has been used for the prediction of lignin, cellulose and hemicellulose fractions[13] over a wide range of woody biomass types. Likewise, lignin, xylose, mannose, galactose, glucose and arabinose contents[14] have been accurately determined in a set of agricultural biomass feedstocks. The capacities of

chemometric tools in combination with Fourier Transform Infrared Spectroscopy (FTIR) are also extended to the analysis of first generation biofuels. Thus, ester content in biodiesel samples can be accurately determined by near-IR spectroscopy in combination with partial least square (PLS) analysis.[15] Additionally, a standard test method for the assessment of the biodiesel content in conventional diesel fuel oil using mid-IR spectroscopy with PLS analysis is already implemented in biodiesel production processes.[16] Against this background, although the above-described models show good predictive ability and could be used to predict the above properties at laboratory and industrial scales, no previous attempt to develop a comprehensive and general model coupling IR spectroscopy and multivariate analysis as a rapid quantitative characterization technique of pyrolysis oils has been reported so far.

Herein, we show the development and validation of a multivariate method based on PLS regression in combination with IR spectroscopy for the quantification of the different compositional groups in pyrolysis oils. A data set consisting of one hundred and eleven samples obtained from the thermal and catalytic pyrolysis of woody biomass and the upgrading of bio-oil vapors by catalytic cracking over zeolites and metal oxides enables the accurate determination of more than 90% of the bio-oil composition.

EXPERIMENTAL

Materials

Bio-oils

A total of 111 bio-oil samples were produced *via* thermal and catalytic pyrolysis of pine woodchips and catalytic upgrading of the bio-oil vapors. Thermal and catalytic pyrolysis of woody biomass was performed in an auger reactor operated at 450ºC with N_2 as the carrier gas resulting in 10 bio-oil samples. A detailed description of the

experimental system used can be found elsewhere.[17,18] Thermal pyrolysis was conducted using merely silica sand as heat carrier with sand-to-silica mass ratio of 3:1. For the catalytic pyrolysis, a list of the catalysts and catalyst-to-biomass ratios used in the experiments can be found in **Table S1**.

Catalysts

Catalytic upgrading of the organic fraction of the crude bio-oil was performed in a fixed-bed reactor at 450ºC over a set of catalysts consisting of zeolites and metal oxides resulting in 101 bio-oil samples. Some of the zeolites were previously reported in bio-oil upgrading.[19-21] The rest of the catalysts were included to improve the performance of the predictive model. Briefly, commercial zeolites with different framework topologies (MFI and FAU) and Si/Al ratios (**Table S2**) were converted into the protonic form by calcination prior to use. Hierarchical ZSM-5 and faujasite zeolites were prepared by desilication of the bulk zeolites in stirred aqueous NaOH solution using an Easymax^{TM} 102 reactor system (Mettler Toledo).[22,23] In some cases, a sequential mild acid treatment with either HCl or Na_2H_2EDTA was applied to restore a similar bulk Si/Al atomic ratio to that of the bulk zeolite. Metal loading (Mg, Ga, Fe, Ca, Ag, Zn, Cu, or Co) of selected bulk and hierarchical zeolites was attained by ion exchange and wet impregnation.[19,20] Carbon-templating ZSM-5 zeolite was prepared by hydrothermal treatment from a gel of the following molar composition: SiO_2 : 0.46 TPAOH : 0.025 $Al(CH(CH_3)_2)_3$: 51.25 H_2O.[24] CeO_2 and Fe_2O_3 metal oxides of different morphologies were prepared following the experimental protocol reported in refs. 25-27. Detailed synthesis conditions used to obtain the different catalysts are provided in Supporting Information.

Bio-oil characterization

The chemical composition of the organic phase of bio-oil was analyzed by GC-MS using a Varian CP-3800 gas chromatograph connected to a Saturn 2200 ion trap mass spectrometer. A capillary column (CP-Sil 8 CB, Agilent, low bleed: 5% phenyl, 95% dimethylpolysiloxane, 60 m × 0.25 mm i.d. × 0.25 mm film thickness) was used. The oven temperature was initially kept at 40ºC for 4 min. Then, a heating rate of 4ºC min^{-1} was implemented to reach a final column temperature of 300ºC, which was maintained for 16 min. He (BIP quality) was used as carrier gas (1 mL min^{-1}). The temperatures of the injector, detector, and transfer line were 300, 220, and 300ºC, respectively. Sample volumes of 1 μL (1:25 wt%, in a mixture of 1:1 CH_2Cl_2/C_2H_6O) were injected (5:1 split mode, 7.5 min solvent delay). The MS was operated in electron ionization mode within 35-550 m/z range. Each peak was assigned to selected compounds according to the corresponding m/z, which were previously defined in the automatic library search NIST 2011. Each sample was analyzed twice, and the results were computed as an average. A total of 104 compounds were individually identified and they were categorized into eight major classes namely phenols, acids, aldehydes, ketones, furans, cyclic HCs, aromatics and esters (**Tables S3** and **S4**). The percentage of each compound in the bio-oil was determined by area normalization, *i.e.*, the quotient between the area of each peak and the total area.[28] The repeatability of GC-MS results was tested by injecting twice the samples, obtaining relative standard deviations lower than 10% (**Table 1**).

A Bruker Vertex 70 spectrometer was used to record the FTIR spectra of the bio-oil samples. 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. Due to limitations in the equipment configuration, two spectra were acquired for each sample corresponding to Medium-IR (MIR) and Near-IR (NIR). Whilst a Globar source and a deuterated L-alanine doped triglycine sulphate detector

were used for MIR configuration (4000 to 400 cm^{-1}), a halogen source and an indium gallium arsenide detector were set up for NIR conditions (9000-4000 cm^{-1}). Each spectrum was acquired in transmission mode using 32 co-averaged scans and a spectral resolution of 2 cm^{-1}.

Chemometric methods and data analysis

Before building optimal PLS calibration models for each chemical family, an optimization of the spectral treatment and the wavelength was carried out. 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. The initial data matrix used as predictor was composed of 111 spectra collected from the different bio-oils with 8918 variables per spectrum (NIR-MIR) (9000-400 cm^{-1}) whilst the response matrix consisted of 111 samples and eight variables corresponding to the percentages of compositional groups obtained by GC-MS, i.e., phenols, acids, aldehydes, ketones, furans, cyclic HCs, aromatics and esters (Scheme 1). Full cross-validation (FCV), also known as leave-one-out cross validation, was run as a validation procedure to determine the optimum number of latent variables (LV).

FTIR data pre-treatment

Spectral data (**Figure 1**) were imported from OPUS software into the Unscrambler X 10.3 (Camo Inc., Oslo, Norway) software. All other spectral processing and chemometric tools were performed using this program. Three different spectral pre-processing techniques were separately evaluated to minimize those physical effects that were not representing chemical phenomena and to remove any irrelevant information

(Scheme 1): (1) Standard Normal Variate (SNV), a row-oriented transformation which centers and scales individual spectra, avoiding scatter effects from spectral data, (2) first derivative using Savitzky-Golay algorithms (SGolay) (polynomial order 2, window with 11 points) to remove additive effects from the spectra, such as baseline offsets or features, and (3) normalization to the highest peak, which attempts to correct for scaling differences (pathlength effects, scattering effects, source or detector variations, or other general instrumental sensitivity effects) by identifying some aspect of each sample which should be essentially constant from one sample to the next. The model parameters corresponding to the three spectral treatments for the whole set of samples (111) are provided in the Supporting Information (**Tables S5-S7**). Herein, we will only show SNV results since this data pretreatment method drove to the optimal model performance.

IR wavelength selection

Due to the high number of wavelengths, initially, a variable reduction process was manually performed[39] based on spectroscopic experience. Firstly, the removal of some spectral ranges associated with the NIR region (9000-4000 cm^{-1}), which showed very low contribution to the total data variance was evaluated. A significant improvement in the model performance was obtained based on parameters determined in subsection 2.3. Secondly, the exclusion of the broadest vibrations associated with the presence of water (3700-3000 cm^{-1}) was also appraised. Again, better model performance in terms of model prediction capacity was obtained. Finally, the removal of different wavelength ranges were systematically assessed (4000-3700, 3000-2700, 2700-2400, 2400-2000, 2000-1700, 1700-1400, 1400-1000 and 700-400 cm^{-1}). Those wavelength ranges decreasing the model performance in terms of the accuracy and precision of the model were removed. Finally, a range comprised between 2000 and 700 cm^{-1} was selected as

the optimal range since other wavelengths associated to NIR and MIR regions did not improve the model performance **(Table S5).** With this wavelength range eight PLS regression models were built, one for each GC-MS group, where the 1349 variables of the IR spectra comprised between 2000 and 700 cm^{-1} for the 111 bio-oils were used as predictor data (x) and the percentage of each chemical family as the response variable (y).

Partial least square (PLS) regression models

The ASTM E1655 standard[29] establishes a minimum number of samples for constructing PLS models using infrared spectroscopy. The minimum number of samples must be equal to 6(k + 1) in calibration set and to 4k in prediction set, being the k value the number of latent variables selected in the model. The PLS models were built with up to seven latent variables. A total of 74 samples, two thirds of the samples were used for constructing PLS calibration models and the remaining one third, 37 samples were used for external validation to predict, with a separate set of samples not used in the calibration, those model deviations that can be expected when real samples are assessed.[30] Samples were designated according to the Kennard-Stone algorithm,[31] which selects samples sequentially by maximizing the Euclidean distances between each other. PLS regression analysis with FCV determined the optimum number of LV with the calibration set (74 samples). Once obtained the optimal PLS calibration model, the prediction accuracy of this calibration model was evaluated on the prediction set (37 samples) (Scheme 1).

The performance of the calibration and the predictive ability of the models was evaluated by the following criteria: 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; the root mean square error of calibration ($RMSE_{Cal}$); the root

mean square error of prediction ($RMSE_{Pred}$) [32]; the standard error of calibration (SEC); the standard error of prediction (SEP), which is defined as the standard deviation of the predicted residuals; the ratio of performance to deviation (RPD), being calculated from the ratio between the standard deviation of the reference values and the standard error of prediction (SD/SEP)[33] and the linear correlation coefficients for calibration (R_{cal}^2) or prediction (R_{Pred}^2)[34] (see **Scheme 1**), which provide information regarding the fit of the model. The lower the bias, the RMSE, the SEP and the higher the R^2 (close to 1), the better the prediction ability of the PLS model. RMSE, SEP and bias provide information about the accuracy, the precision and the trueness of the model. These parameters were calculated according to the following equations:

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{I}(\bar{y}_i - y_i) \quad (1)$$

$$RMSE_{Pred} = \sqrt{\frac{\sum(\bar{y}_i - y_i)^2}{N}} \quad (2)$$

$$SEP = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\bar{y}_i - y_i - Bias)^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum(y_i - \bar{y}_i)2}{\sum(y_i - y_{mean})2} \quad (4)$$

where N is the number of samples, $\bar{y}_i$ is the value predicted by the model, y_i is the experimental value obtained by GC-MS and y_{mean} is the average value of the experimental values.

The ASTM E1655-05 [29] standard proposed the investigation of systematic errors using a t-test for validation samples at 95% confidence and degrees of freedom equal to the number of prediction samples. The t value was calculated using Eq. (5).

$$t_{bias} = \frac{|Bias|\sqrt{N}}{SEP} \quad (5)$$

When the t_{bias} is higher than the t-value from a standard t-table (t_{crit}), it can be concluded that there is a 95% probability that the prediction from the multivariate

model will not lead to the same average results as the reference method, pointing out that the differences between experimental and modelled results are significant.

An estimation of the model uncertainty was performed by a confidence interval (C.I.) which is determined from the Eq. (6) when bias is small and standard deviation is unknown:

$$\text{C.I.} = \bar{y}_i \pm z\, SEP \approx \bar{y}_i \pm z\, RMSE_{Pred} \quad (6)$$

where 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.

Absolute precision and relative precision will be calculated according to the Eq (7) and (8):

$$\text{Absolute Precision} = z\,.\,RMSE_{Pred} \quad (7)$$

$$\text{Relative Precision (\%)} = \frac{z\,.\,RMSE_{Pred}}{\bar{y}} * 100 \quad (8)$$

Residuals, which are of diagnostic value for the quality of a model, enabled identifying 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 "±3SD" was indicative of non-outliers. In contrast, a point that clearly deviates from this line and is outside "±4SD"[35] can be considered as outlier. In this study, all samples were included, except in the case of aldehydes and esters, where some outliers were detected.

In order to evaluate the precision of the PLS model based on repeatability for IR spectra, a total of five spectra for a same sample were determined and the compositional predictions for each replicate were obtained. Uncertainties expressed as relative precision were also calculated (**Table 2**).

RESULTS AND DISCUSSION

GC-MS results

Compositional intervals for the eight chemical families obtained by GC-MS as a function of the different validation procedures are shown in **Table 1**. The two groups of compounds showing a broader calibration range were phenols (24.66-65.06%) and aromatics (4.90-63.85%) whereas furans (1.10-5.94%) and acids (non-detected-5.77%) showed the narrowest ranges for the whole set of samples. This was also reflected in the Kennard-Stone calibration set whereas for the prediction set, acids and aldehydes were the two groups showing the lowest variations. In fact, precisions expressed as relative uncertainty showed high values for these two groups of compounds (acids and aldehydes) when samples were injected by GC-MS by duplicate due to these low compositional intervals. Although a high number of replicates should be analyzed in order to improve the precision of the method, this would be very time-consuming considering that the whole number of samples was 111 and only duplicate samples were determined by GC-MS in this study.

FTIR results

Original data set enclosed 8918 variables, covering the NIR and MIR regions of the IR spectra (**Figure 1**). The highest dissimilarities were obtained in the MIR region ranging from 4000 to 400 cm^{-1}, which involves 4254 variables. Briefly, several peaks with different intensities were found, pointing out different functionalities in the bio-oils. The highest and broadest vibrations were attained between 3675 and 3000 cm^{-1}, which are associated with O-H stretching due to the presence of water, phenols, alcohols,[36] 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}.[37] 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.[38] 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[1] between 840 and 700 cm^{-1}.

Implementation of the PLS regression models

Building PLS calibration models

Once selected the wavelength range of the infrared spectra and the spectral pre-processing (SNV), eight PLS regression models were built, one for each chemical family, where the 1349 variables of the IR spectra comprised between 2000 and 700 cm^{-1} for the 74 calibration bio-oils were used as predictor data (x) and the percentage of each chemical family as the response variable (y)..

FCV was used as validation procedure in order to determine the number of LV included in the PLS regressions models (**Scheme 1**). The PLS calibration models showed LV, which ranged between 2 for cyclic HCs and 7 for aldehydes. The different parameters showing the performance of the PLS regression models and the explained variance are shown in **Table 2**.

The bias values were zero for all families indicating no systematic errors. This was reflected on the $RMSE_{Cal}$ and SEC, which were similar for each chemical family. Therefore, RMSECal was used to determine precision. The highest $RMSE_{Cal}$ corresponded to aromatics (5.18%) (one of the groups with the widest range of

percentage values) and the lowest (0.42%) to acids. The lowest SEC was obtained for the aldehydes (0.42%) whereas the aromatics presented the highest value (5.22%).

Other parameter providing additional information on the expected accuracy of the PLS predictions is the RPD (**Table 2**). The RPD should be at least 3, between 2 and 3, 1.5 to 2 and lower than 1.5 for analytical, good, medium and poor performance quality, respectively.[39]

All the chemical families presented RPD higher than 2, with the exception of furans, indicating good performance quality with values higher than 3 for acids, aldehydes, cyclic HCs and aromatics (**Table 2**).

Regarding the fitting of the models, all of the different compositional families showed remarkable regression coefficients, except furans (**Table 2**, **Figure 2**) with R^2_{Cal}=0.55 indicating that 45% of the total variation could not be explained by the PLS model. This group included only few compounds with variable nature providing a marginal percentage of the total sample (see **Table S3**), which can be the main reason for the inferior results. The other seven groups, *i.e.*, phenols, acids, aldehydes, ketones, cyclic HCs, aromatics and esters, showed PLS regression models with R_{Cal}^2 higher or equal than 0.80. Remarkably, these models were built with a calibration set having an ample heterogeneity associated to bio-oils of different composition, which includes not only upgraded bio-oils obtained after catalytic cracking of pyrolysis vapors, but also those raw bio-oils obtained from the pyrolysis and catalytic pyrolysis of biomass without further upgrading. The use of a more homogenous sample set to generate calibrations models, *i.e.,* only upgraded bio-oils would probably result in reduced errors[40] although would decrease the applicability of the model.

Relative uncertainties showed high values for two chemical families, acids and aldehydes which will be commented in more detail in next section.

Prediction of chemical families by external validation

Once the optimal calibration models were built, these were applied to predict the compositional groups for a set of 37 samples used as external validation, which were not used in the calibration models. It is worth mentioning that although the prediction of furans was carried out, these data were not discussed due to the poor results achieved during the development of the calibration model. As previously commented, furans included only a few compounds with variable nature, covering a marginal percentage of the total sample (see **Tables S3-S4**). The prediction ability for the calibration models can be assessed according to different parameters, such as bias, $RMSE_{Cal}$, $RMSE_{Pred}$ and SEP, which are reported in **Table 2**. Firstly, bias can be used to determine the presence of systematic errors. The bias values for acids ($-4.6*10^{-2}$) and aldehydes ($-6.3*10^{-2}$) were almost zero, indicating that systematic errors were negligible for these two families. Similarly, although bias values indicated that the models over-predicted the concentration of phenols, ketones and cyclic hydrocarbons and under-estimated the composition of aromatics and esters, these values were again quite close to zero, pointing out that the presence of systematic errors can be ruled out.

Regarding the accuracy of the model, **Table 2** shows that $RMSE_{Pred}$ values ranged between 0.24% for acids and 3.79% for aromatics, being the highest $RMSE_{Pred}$ values obtained for those compositional groups with the broadest range of composition (aromatics and phenols). Additionally, the highest SEP value was obtained for aromatics (3.83%), whilst acids showed the lowest (0.24%). $RMSE_{Pred}$ and SEP exhibited similar values, which corroborated the absence of bias and thus $RMSE_{Pred}$ could be used to determine the relative uncertainty. All the chemical families showed RPD values higher than 2.5 (**Table 2**), which further corroborated that bio-oil composition could be predicted with an acceptable degree of accuracy. With regard to

the fit of the models, R_{Pred}^2 values between 0.8 and 0.9 point out a notable predictive ability, and values higher than 0.9 denote an excellent predictive ability.[41] In addition, a reasonable model predictive ability was suggested by Chen et al. [42] when differences between R_{cal}^2 and R_{Pred}^2 are lower than 0.2. Larger differences between both values indicate over-fitting of the models,[35] *i.e.,* a well-fitting model with limited or no predictive ability.

The prediction of the chemical families based on the independent set of samples led to R^2_{Pred} values ranging between 0.68 for acids and 0.90 for aromatics and esters. Remarkable predictions were attained for phenols (0.85), aldehydes (0.84), ketones (0.87), cyclic HCs (0.83), aromatics (0.90) and esters (0.90) (**Figure 2**), which were in line with the results obtained by the PLS calibration model, thus suggesting the applicability of the developed models in predicting the bio-oil composition.[43] However, a very poor fitting between the prediction and the reference compositional values was attained for acids (R^2_{Pred} = 0.68) although the linear correlation coefficient for the calibration was notable (R^2_{cal} = 0.93). This poor prediction could be associated to the fact that R_{cal}^2 value for this chemical family was over-fitted, likely related to own nature of the bio-oil samples, where 10 highly acidic bio-oils together with 64 samples with a very low percentage of acid compounds, even null in some cases, was modeled. Indeed, R_{cal}^2 would decrease from 0.93 to 0.58 if the former samples were considered as outliers. This result prevented the potential application of the PLS model for the prediction of this compositional group in the current set of samples. Additionally, it is worth mentioning that the presence of two groups greatly different in their acid composition highlighted the importance of the selection algorithm between calibration and prediction samples, since the selection procedure could lead to inconsistences in the concentration ranges between both sets of samples, as clearly observed for acids in

Figure 2. Similar results have been found by other authors,[44,45] pointing out that a bigger set of samples covering a broader range of compositions could be needed for this compositional group. In line with this, although the same issues might appear when the aldehydes and the cyclic HCs were predicted, R_{cal}^2 for the aldehydes and cyclic HCs hardly decreased if the highly concentrated samples would be considered as outliers. Then, consistent results were obtained between R_{cal}^2 and R_{Pred}^2 although the application of the Kennard-Stone selection algorithm did not include the highly concentrated samples in the prediction set.

Statistical tests based on paired t-test at 95% confidence level were also performed for each chemical family in order to assess whether significant differences were found between the values predicted by the PLS models and the experimental values obtained by GC-MS (**Table S8-S10**). No significant statistical differences were found between both methods for each chemical family. Only in the case of furans, a weak correlation was found between experimental and modeled results confirming the poor model prediction for this chemical family. Moreover, the bias included in the model were not statistically significant at 95% level since t_{bias} calculated by the PLS models were always lower than t_{crit} (2.03) for the degree of freedom equal to the number of prediction samples (**Table 2**), again pointing out no systematic errors as suggested by the ASTM E1655 [29]. Once proved that bias were negligible, confidence intervals (C.I.) for predictions were also calculated for the different chemical families, providing information regarding the precision of the PLS model. **Table 2** shows the uncertainties expressed as relative precision (%) for the different predicted chemical families. For comparative purposes, the relative uncertainties, obtained for the different compositional groups by GC-MS are reported in **Table 1**The PLS model uncertainties were very high for furans, acids and aldehydes due to the own nature of the samples

and low range of compositions The other chemical families (phenols, ketones, cyclic HCs, aromatics and esters) showed relative precisions ranging from 14% for phenols to 32% for esters, similar to the ones obtained by GC-MS (**Table 2**) concluding that the PLS modeling based on infrared spectra could be considered as a first approximation to semi-quantitatively determine the composition of phenols, ketones, cyclic hydrocarbons, aromatics and esters with reasonable precision. These results were compared with studies of repeatability for one sample analysed by FTIR-PLS five times where the highest relative uncertainties were also obtained for acids and aldehydes. However, higher precision for each chemical family was obtained by studying the repeatability of samples by FTIR when compared to GC-MS. This could be explained by the different procedure to determine the chemical families by both methods. Whereas FTIR allowed obtaining spectra in a fast way providing directly information regarding the organic composition of the bio-oils, GC-MS required a more tedious method in which each chemical family was composed by several individual organic compounds (30 for the aromatics). On the other hand, prediction uncertainties based on a constant $RMSE_{Pred}$ could not lead to correct interval estimates, in particular when a reference value based on GC-MS is taken as good and object-specific prediction uncertainties would be required in future. Special caution should be taken into account when the mean value of the data set is close to zero as it happens for acids and aldehydes since the value of the relative precision increases.

More research should be focused on developing ruggedness models able to predict the composition of these chemical families by including other spectral ranges and/or spectral treatments. In particular, an optimization of the spectral range by using more sophisticated programs based on algorithms could improve results shown in this work. Also, a higher number of samples could reduce the uncertainties obtained for some

specific families. However, it is worth mentioning that the percentage of acids, aldehydes and furans was always lower than 5%, which makes their modelling highly demanding. Additionally, a reduction in the uncertainties could be obtained in future work by performing more studies of repeatability and reproducibility for IR spectra. Still, the main advantage of implementing IR in combination with PLS regression models over the chromatographic procedure relies on the efficiency of the analyses. In particular, in this study, each chromatogram runs for 90 minutes and the further treatment of data as individual organic compounds, with a previous developed method, would extend for at least one hour. Additionally, to ensure the repeatability of the results, this protocol has to be repeated at least by duplicate, which significantly increases the analysis time. In contrast, the IR spectrum is acquired in less than 1 minute and no additional sample preparation is required, as the liquid samples are used as-received. The following treatments of the IR data and the subsequent application of chemometric models are also performed in a few minutes. Thus, total analysis time can be greatly decreased from hours to minutes whereas similar accuracy to that obtained by the GC-MS analysis is attained in determining more than 90% of the bio-oil composition in terms of compositional groups. In addition, the loss of information related to individual organic compounds can be considered as negligible since it is not required for the monitoring and quality control of production processes. Finally, it is important to note that the trends identified in this work have been derived from pine-pyrolysis oils and therefore these observations can differ under different process conditions and/or with other feedstocks.

CONCLUSIONS

In this work, we demonstrate that mid-infrared spectroscopy based on partial least squares regression models can be used as a fast technique to predict the composition of

five chemical families: phenols, ketones, cyclic hydrocarbons, aromatics and esters with reasonable accuracy in 111 pine-wood-derived pyrolysis oils including those obtained via catalytic upgrading and thermal and catalytic pyrolysis. Standard normal variate was the spectral pre-processing technique leading to the best multivariate regression. Performance of the calibration models were validated by external validation, obtaining $RMSE_{Pred}$ between 0.40-3.79% for cyclic hydrocarbons and aromatics with relative uncertainties decreasing when repeatability was studied by FTIR versus GC-MS. Good fitting between the experimental and the predicted values with coefficients of determination for the prediction higher than 0.8 were obtained for these five chemical families. The attained results have successfully proved that infrared spectroscopy in combination with the chemometric technique can drive the development of enhanced tools for the online monitoring and quality control during the production of second-generation bio-fuels in a fast and secure manner.

ASSOCIATED CONTENT

Supporting Information

Details on the catalyst preparation. List of compounds identified by GC-MS analysis. Relative area (%) of the different chemical families for each bio-oil sample obtained by GC-MS. Results of the PLS regression models with different spectral treatments.

AUTHOR INFORMATION

Corresponding Author

*Address correspondence to marisol@icb.csic.es

ORCID

Mª Soledad Callén: 0000-0001-6063-7386

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Authors would like to thank the Spanish MINECO and European Union FEDER funds for providing support for this work (projects CTQ2012-37984-C02-01 and ENE2015-68320-R). B.P. thanks the Spanish Ministry of Education through the Temporary Transfer Program (TRA13/00003).

REFERENCES

(1) Kanaujia, P.K.; Sharma, Y.K.; Garg, M.O.; Tripathi, D.; Singh, R. Review of analytical strategies in the production and upgrading of bio-oils derived from lignocellulosic biomass. *J. Anal. Appl. Pyrol*. **2014**, *105*, 55–74.

(2) Bridgwater, A.V.; Peacocke, G.V.C. Fast pyrolysis processes for biomass. *Renew. Sust. Energy Rev.* **2000**, *4*, 1–73.

(3) Yildiz, G.; Pronk, M.; Djokic, M.; Van Geem, K.M.; Ronsse, F.; Van Duren, R. Validation of a new set-up for continuous catalytic fast pyrolysis of biomass coupled with vapor phase upgrading. *J. Anal. Appl. Pyrol*. **2013**, *103*, 343–351.

(4) Bridgwater, A. V. Review of fast pyrolysis of biomass and product upgrading. *Biomass Bioenergy* **2012**, *38*, 68–94.

(5) Adjaye, J.D.; Sharma, R.K.; Bakhshi, N.N. Characterization stability analysis of wood-derived bio-oil. *Fuel Process. Technol.* **1992**, *31*, 241–256.

(6) Karagoz, S.; Bhaskar, T.; Muto, A.; Sakata, Y. Comparative studies of oil compositions produced from sawdust, rice husk, lignin and cellulose by hydrothermal treatment. *Fuel* **2005**, *84*, 875–884.

(7) Luo, Z.; Wang, S.; Liao, Y.; Cen, K. Mechanism study of cellulose rapid pyrolysis. *Ind. Eng. Chem. Res.* **2004**, *43*, 5605–5610.

(8) Zhang, J.; Toghiani, H.; Mohan, D.; Pittman Jr. C.U.; Toghiani, R.K. Product analysis and thermodynamic simulations from the pyrolysis of several biomass feedstocks. *Energy Fuels* **2007**, *21*, 2373–2385.

(9) Li, J.; Wu, L.; Yang, Z. Analysis and upgrading of bio-petroleum from biomass by direct deoxy-liquefaction. *J. Anal. Appl. Pyrol.* **2008**, *81*, 199–204.

(10) Eide, I.; Neverdal, G. Fingerprinting Bio-Oils from Lignocellulose and Comparison with Fossil Fuels. *Enery Fuels* **2014**, *28*, 2617–2623.

(11) Zhang, L.; Chenjie, S.; Liu, R. GC-MS and FT-IR analysis of the bio-oil with addition of ethyl acetate during storage. *Front. Energy Res.* **2014**, *2*, 1–6.

(12) Chadwick, D.T.; McDonnell, K.P.; Brennan, L.P.; Fagan, C.C.; Everard, C.D. Evaluation of infrared techniques for the assessment of biomass and biofuel quality parameters and conversion technology processes: A review. *Renew. Sust. Energ. Rev.* **2014**, *30*, 672–681.

(13) Jin, S. Y.; Chen, H. Z. Near-infrared analysis of the chemical composition of rice straw. *Ind. Crops Prod.* **2007**, *26* (2), 207–211.

(14) Kelley, S. S. Rowell, R. M.; Davis, M.; Jurich, C.K.; Ibach, R. Rapid analysis of the chemical composition of agricultural fibers using near infrared spectroscopy and pyrolysis molecular beam mass spectrometry. *Biomass Bioenerg.* **2004**, *27* (1), 77–88.

(15) Baptista, P.; Felizardo, P.; Menezes, J.C.; Correia, M.J.N. Multivariate near infrared spectroscopy models for predicting the methyl esters content in biodiesel. *Anal. Chim. Acta.* **2008**, *607* (2), 153–159.

(16) ASTM Standard D7371, 2007. Standard test method for the determination of biodiesel (fatty acid methyl esters) content in diesel fuel oil using mid-infrared spectroscopy (FTIR-ATR-PLS method). ASTM International: West Conshohocken, PA, 2007; http://dx.doi.org/10.1520/D7371-12.

(17) Veses, A.; Aznar, M.; Martínez, I.; Martínez, J.D.; López, J.M.; Navarro, M.V.; Callén, M.S.; Murillo, R.; García, T. Catalytic pyrolysis of wood biomass in an auger reactor using calcium-based catalysts. *Biores. Technol*. **2014**, *162*, 250–258.

(18) Veses, A.; Aznar, M.; López, J.M.; Callén, M.S.; Murillo, R.; García, T. Production of upgraded bio-oils by biomass catalytic pyrolysis in an auger reactor using low cost materials. *Fuel* **2015**, *141*, 17–22.

(19) Veses, A.; Puértolas, B.; López, J. M.; Callén, M. S.; Solsona, B.; García, T. Promoting deoxygenation of bio-oil by metal-loaded hierarchical ZSM-5 zeolites. *ACS Sustainable Chem. Eng.* **2016**, *4* (3), 1653–1660.

(20) Veses, A.; Puértolas, B.; Callén, M.S.; García, T. Catalytic upgrading of biomass derived pyrolysis vapors over metal-loaded ZSM-5 zeolites: Effect of different metal cations on the bio-oil final properties. *Microp. Mesop. Mat*. **2015**, *209*, 189–196.

(21) Puértolas, B.; Veses, A.; Callén, M. S.; Mitchell, S.; García, T.; Pérez-Ramírez, J. Porosity-acidity interplay in hierarchical ZSM-5 zeolites for pyrolysis oil valorization to aromatics. *ChemSusChem* **2015**, *12*, 2383–3293.

(22) Milina, M.; Mitchell, S.; Michels, N.-L.; Kenvin, J.; Pérez Ramírez, J. Interdependence between porosity, acidity, and catalytic performance in hierarchical ZSM-5 zeolites prepared by post-synthetic modification. *J. Catal.* **2013**, *308*, 398–407.

(23) Verboekend, D.; Keller, T.C.; Mitchell, S.; Pérez-Ramírez, J. Hierarchical FAU and LTA-type zeolites by post-synthetic design: a new generation of highly efficient base catalysts. *Adv. Funct. Mater.* **2013**, *23*, 1923–1934.

(24) Koo, J.-B.; Jiang N.; Saravanamurugan, S.; Bejblova, M.; Musilova, Z.; Cějka, J.; Park, S.-E. Direct synthesis of carbon-templating mesoporous ZSM-5 using microwave heating. J. Catal. **2010**, *276*, 327–334.

(25) López, J. M.; Arenal, R.; Puértolas, B.; Mayoral, A.; Taylor, S. H.; Solsona, B.; García, T. Au deposited on CeO_2 prepared by a nanocasting route: A high activity catalyst for CO oxidation. *J. Catal.* **2015**, *317*, 167–175.

(26) Torrente-Murciano, L.; Gilbank, A.; Puértolas, B.; Garcia, T.; Solsona, B.; Chadwick D. Shape-dependency activity of nanostructured CeO_2 in the total oxidation of polycyclic aromatic hydrocarbons. *Appl. Cat., B* **2013**, *132-133*, 116–122.

(27) Solsona, B.; García, T.; Sanchis, R.; Soriano, M.D.; Moreno, M.; Rodríguez-Castellón, E.; Agouram, S.; Dejoz, A.; López Nieto, J. M. Total oxidation of VOCs on mesoporous iron oxide catalysts: Soft chemistry route versus hard template method. *Chem. Eng. J.* **2016**, *290*, 273–281.

(28) Dupuy, N.; Molinet, J.; Mehl, F.; Nanlohy, F.; Le Dréau, Y.; Kister, J. Chemometric analysis of mid infrared and gas chromatography data of Indonesian nutmeg essential oils. *Ind. Crops Prod.* **2013**, *43*, 596–601.

(29) ASTM Standard E1655. Standard practices for infrared multivariate quantitative analysis. ASTM International: West Conshohocken, PA, 2005; http://www.astm.org.

(30) Lupoi, J.S.; Singh, S.; Davis, M.; Lee, D.J.; Shepherd, M.; Simmons, B.A.; Henry, R.J. High-throughput prediction of eucalypt lignin syringyl/guaiacyl content using

multivariate analysis: a comparison between mid-infrared, near-infrared, and Raman spectroscopies for model development. *Biotechnol. Biofuels* **2014**, *7*, 93.

(31) Kennard, R. W.; Stone, L. A. Computer aided design of experiments. *Technometrics* **1969**, *11*(1), 137–148.

(32) Naes, T., Isaksson, T., Fearn, T., Davies, T. *A user friendly guide to multivariate calibration and classification*; NIR Publications: Chichester, UK, 2002.

(33) Williams, P.C. Implementation of near-infrared technology. In *Near-Infrared Technology in the Agricultural and Food Industries*; Williams, P.C., Norris, K.H., Eds.; American Association of Cereal Chemists: St Paul, MN., 2001, 145–169.

(34) Sills, D.L.; Gossett, J.M. Using FTIR spectroscopy to model alkaline pretreatment and enzymatic saccharification of six lignocellulosic biomasses. *Biotechnol. Bioeng.* **2012**, *109*(4), 894–903.(35) Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemometrics Intelligent. Lab. Syst.* **2001**, *58*(2), 109–130.

(36) Schnitzer, M. I.; Monreal, C. M.; Facey, G. A.; Fransham, P. B. The conversion of chicken manure to biooil by fast pyrolysis I. Analysis of biooils by FTIR and NMR spectroscopy. *J. Environ. Sci. Health, Part B* **2007**, *42*, 71–77.

(37) Coury, C.; Dillner, A.M. A method to quantify organic functional groups and inorganic compounds in ambient aerosols using attenuated total reflectance FTIR spectroscopy and multivariate chemometric techniques. *Atmos. Environ.* **2008**, *42*, 5923–5932.

(38) Asadieraghi, M.; Wan Dau, W.M.A. In-depth investigation on thermochemical characteristics of palm oil biomasses as potential biofuel sources. *J. Anal. Appl. Pyrol.* **2015,** *115*, 379–391.

(39) Janik, L.J.; Merry, R.H.; Forrester, S.T. Rapid prediction of soil water retention using mid infrared spectroscopy. *Soil Sci. Soc. Am. J.* **2007**, *71*, 507–514.

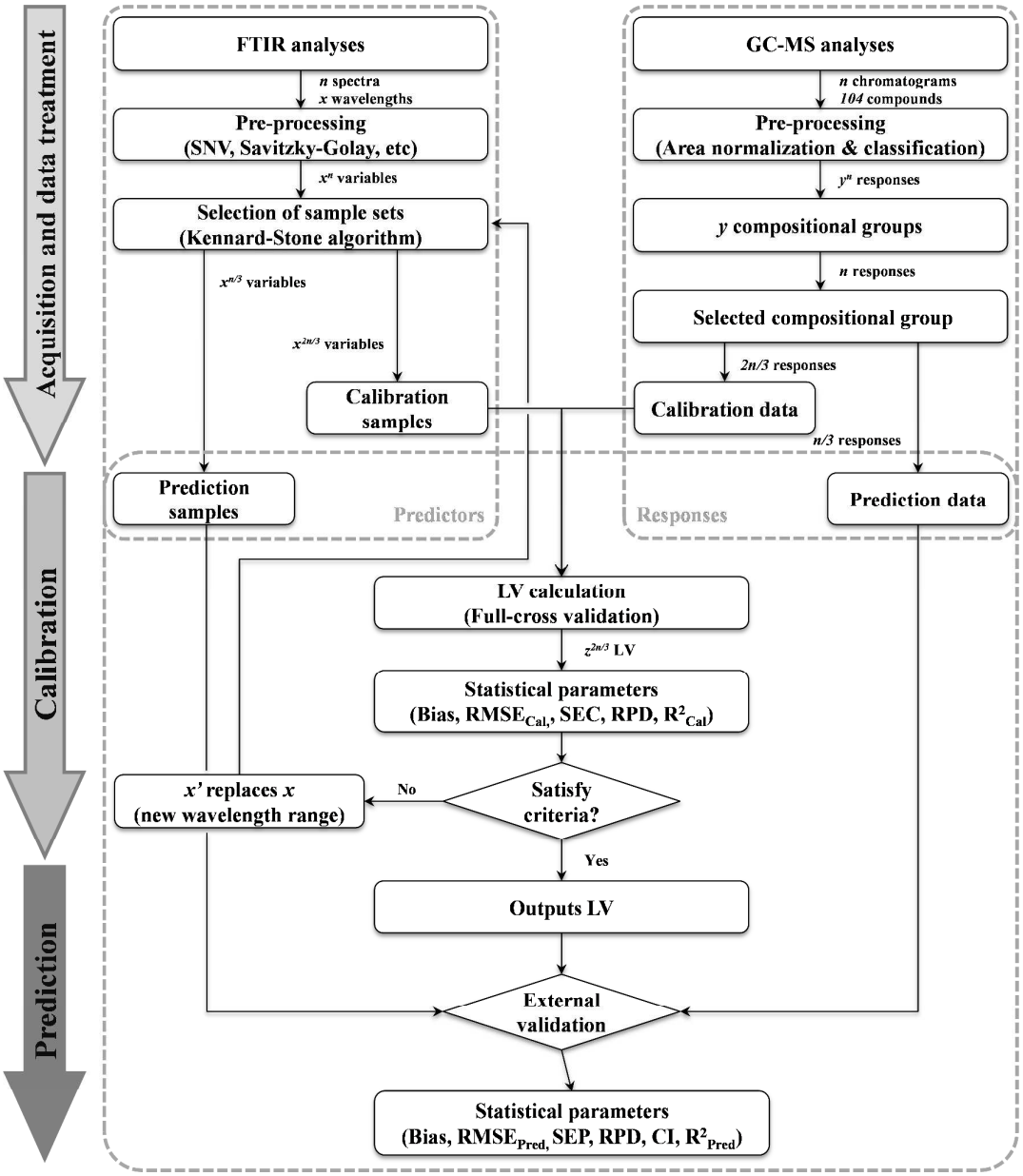
(40) Romera-Fernández, M.; Berrueta, L.A.; Garmon-Lobato, S.; Gallo, B.; Vicente, F.; Moreda, J.M. Feasibility study of FT-MIR spectroscopy and PLS-R for the fast determination of anthocyanins in wine. *Talanta* **2012**, *88*, 303–310. (41) Tamaki, Y.; Mazza, G. Rapid determination of lignin content of straw using attenuated total reflectance Fourier transform mid-infrared spectroscopy. *J. Agric. Food Chem.* **2011**, *59*, 504–512.

(42) Chen, H.; Ferrari, C.; Angiuli, M.; Yao, J.; Raspi, C.; Bramanti, E. Qualitative and quantitative analysis of wood samples by Fourier transform infrared spectroscopy and multivariate analysis. *Carbohydr. Polym.* **2010**, *82*, 772–778.

(43) Meissl, K.; Smidt, E.; Schwanninger, M. Prediction of humic acid content and respiration activity of biogenic waste by means of Fourier transform infrared (FTIR) spectra and partial least squares regression (PLS-R) models. *Talanta* **2007**, *72*, 791–799.

(44) Bellon-Maurel, V.; McBratney, A. Near-Infrared (NIR) and Mid-Infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils. Critical review and research perspectives. *Soil Biol. Biochem.* **2011**, *43*, 1398–1410.

(45) D'Acqui, L.P.; Pucci, A.; Janik, L.J. Soil properties prediction of western Mediterranean islands with similar climatic environments by means of mid-infrared diffuse reflectance spectroscopy. *Eur. J. Soil Sci.* **2010**, *61*, 865–876.



Scheme 1. Algorithm of a PLS regression model for the calculation of the statistical parameters from FTIR and GC-MS results.

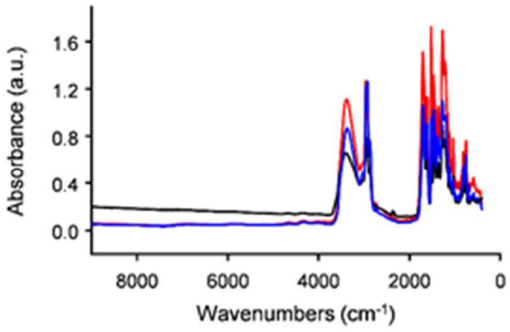


Figure 1. FTIR of three selected bio-oils. Black line corresponds to the bio-oil obtained from biomass pyrolysis using silica sand as heat carrier. Red and blue lines correspond to bio-oils obtained after catalytic upgrading with a commercial ZSM-5 zeolite (Si/Al = 40) at 500ºC and 550ºC, respectively.

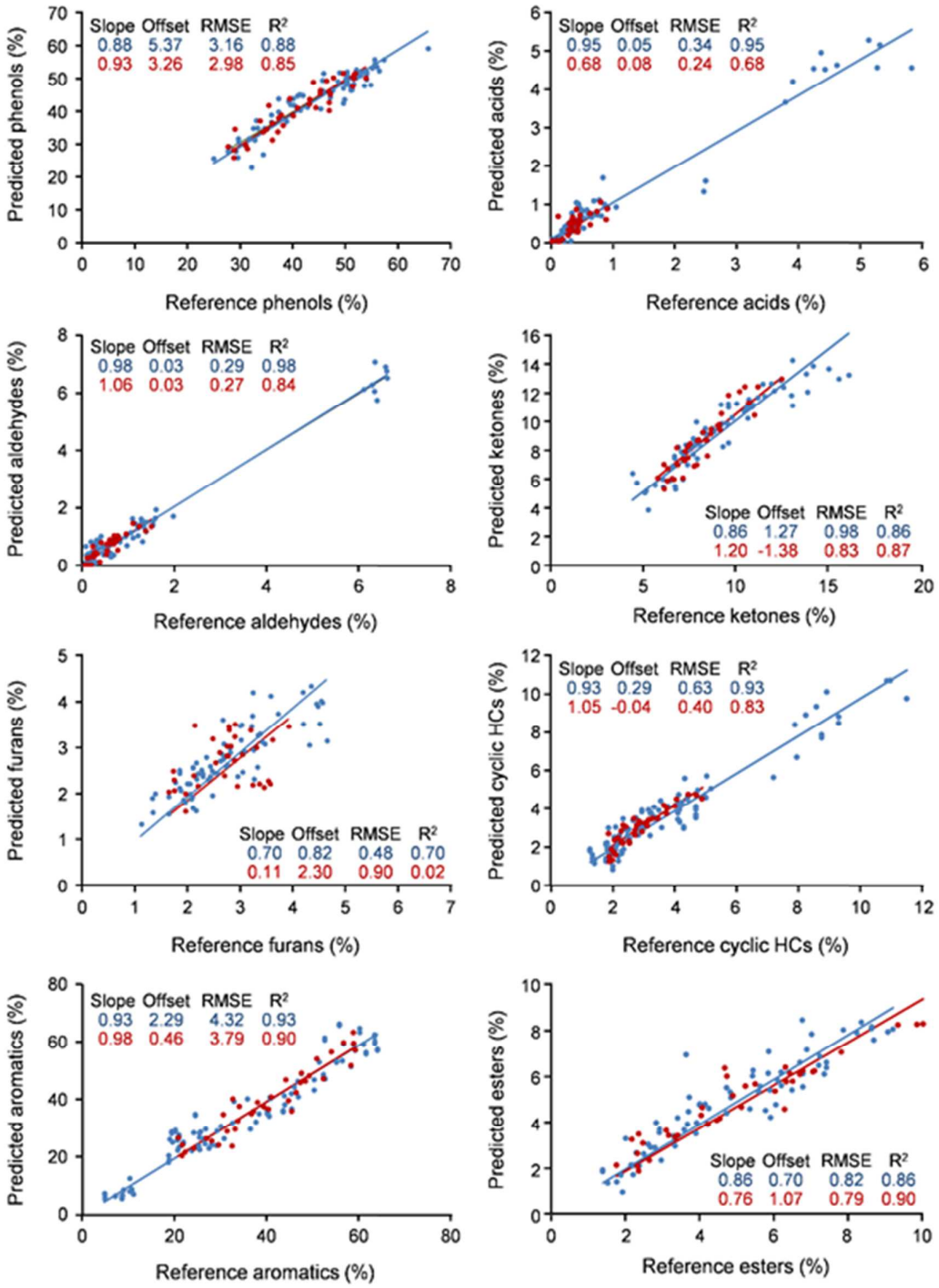


Figure 2. Predicted *versus* GC-MS results for the different chemical families obtained by the PLS model with external validation (target lines; blue dots=calibration, N=74; red dots=prediction N=37).

Table 1. Descriptive statistics of all samples, calibration and prediction sets according to the Kennard-Stone algorithm for the different chemical families (%) obtained by GC-MS. N: number of samples; Cal.: Calibration; Pred.: Prediction; KS: Kennard-Stone algorithm; Rel. Unc.: relative uncertainty (%).

	Samples	N	Phenols	Acids	Aldehydes	Ketones	Furans	Cyclic HCs	Aromatics	Esters
	All	111								
Mean			43.09	0.74	1.13	8.81	2.75	3.66	34.93	4.89
SD[a]			1.11	0.12	0.15	0.24	0.11	0.23	0.89	0.46
Max.[b]			65.06	5.77	6.62	15.95	5.94	11.43	63.85	10.01
Minim.[c]			24.66	n.d.[f]	0.70	4.40	1.10	1.26	4.90	1.01
Rel. Unc.[d]			11	73	60	12	17	28	11	42
C.I.[e]			43.09±4.95	0.74±0.54	1.13±0.67	8.81±1.07	2.75±0.47	3.66±1.01	34.93±3.96	4.89±2.07
	Cal. KS	74								
Mean			44.38	0.96	1.44	9.15	2.73	4.06	32.47	4.82
SD[a]			1.12	0.07	0.06	0.23	0.11	0.21	0.76	0.48
Max.[b]			65.06	5.77	6.63	15.95	4.60	11.43	63.85	9.21
Minim.[c]			24.66	n.d.[f]	0.07	4.40	1.10	1.26	4.91	1.01
Rel. Unc.[d]			11	32	20	11	18	23	10	45
C.I.[e]			44.38±5.00	0.96±0.31	1.44±0.28	9.15±1.01	2.73±0.50	4.06±0.92	32.47±3.41	4.82±2.16
	Pred. KS	37								
Mean			40.52	0.30	0.51	8.13	2.80	2.86	39.86	5.02
SD[a]			1.09	0.22	0.33	0.27	0.09	0.26	1.13	0.43
Max.[b]			53.33	0.88	1.53	12.36	5.94	4.86	59.02	10.01
Minim.[c]			27.38	n.d.[f]	0.08	5.76	1.62	1.85	20.82	1.78
Rel. Unc.[d]			12	333	285	15	15	41	13	38
C.I.[e]			40.52±4.85	0.30±1.00	0.51±1.45	8.13±1.19	2.80±0.41	2.86±1.19	39.86±5.06	5.02±1.90

[a] Average sample standard deviation based on duplicate injections by GC-MS for N samples
[b] Maximum value obtained for each chemical family by GC-MS
[c] Minimum value obtained for each chemical family by GC-MS
[d] Relative uncertainty for a confidence level of 90%
[e] Confidence interval at a confidence level of 90%
[f] non-detected

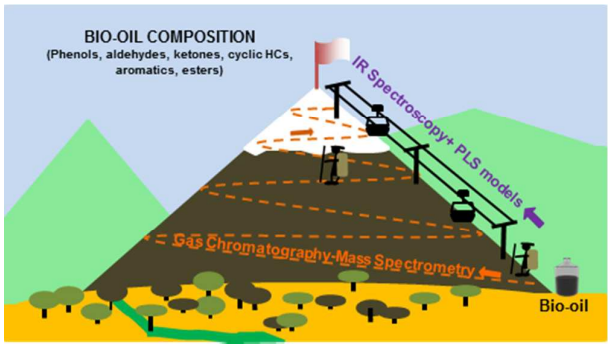
Table 2. Results of the PLS regression models for each chemical family.

Dependent variable	N	Model type	LV	Variance		Bias	RMSE	SEC, SEP	RPD	R^2	Rel Uncertainty (%)	C.I.	t_{bias}
				x	*y*		(%)	(%)					
Phenols	74	Cal.	5	89.97	83.24	2.6×10^{-3}	3.77	3.80	2.41	0.83	17	44.37±7.54	
	37	Pred.	5	91.69		5.5×10^{-1}	2.98	3.17	2.88	0.85	14 (0.39)[c]	41.08±5.96	1.06
Acids	74	Cal.	4	84.01	92.65	3.3×10^{-3}	0.42	0.42	3.64	0.93	83	0.99±0.84	
	37	Pred.	4	80.43		-4.6×10^{-2}	0.24	0.24	6.41	0.68	142 (37)[c]	0.33±0.48	1.18
Aldehydes	73[a]	Cal.	7	95.37	95.28	2.4×10^{-3}	0.42	0.42	4.67	0.96	60	1.30±0.80	
	37	Pred.	7	94.34		-6.3×10^{-2}	0.27	0.27	7.30	0.84	104 (9)[c]	0.51±0.54	1.43
Ketones	74	Cal.	6	93.50	79.84	1.6×10^{-2}	1.20	1.21	2.20	0.80	26	9.15±2.40	
	37	Pred.	6	93.56		2.2×10^{-1}	0.83	0.93	2.85	0.87	19 (0.82)[c]	8.36±1.66	1.47
Furans	74	Cal.	6	93.10	54.97	1.2×10^{-2}	0.59	0.59	1.47	0.55	42	2.73±1.18	
	37	Pred.	6	93.78		-1.9×10^{-1}	0.90	0.89	0.98	0.07	67 (0.40)[c]	2.61±1.80	1.32
Cyclic HCs	74	FCV	2	73.96	91.95	6.3×10^{-3}	0.68	0.68	3.48	0.92	33	4.06±1.36	
	37	Pred.	2	64.30		1.2×10^{-1}	0.40	0.45	5.24	0.83	26 (1.97)[c]	2.98±0.80	1.58
Aromatics	74	Cal.	5	89.28	90.15	4.5×10^{-2}	5.18	5.22	3.14	0.90	31	32.47±10.36	
	37	Pred.	5	91.93		2.9×10^{-1}	3.79	3.83	4.28	0.90	19 (0.52)[c]	39.57±7.58	0.45
Esters	73[b]	Cal.	5	92.83	79.63	3.8×10^{-3}	0.98	0.99	2.40	0.80	40	4.87±1.98	
	37	Pred.	5	91.20		-1.6×10^{-1}	0.79	0.78	2.77	0.90	32 (1.99)[c]	4.82±1.58	1.21

[a,b] After removing some outliers

[c] Results of relative uncertainty (%) for the repeatability of one sample determined five times by FTIR with a confidence level of 95%

For Table of Contents Use Only



Infrared spectroscopy in combination with partial least squares regression models enable the fast and accurate prediction of 90 wt% of bio-oil composition.