Rapid determination of water, total acid number and phenolic content in bio-crude from hydrothermal liquefaction of biomass using FT-IR

René B. Madsen[1], Konstantinos Anastasakis[2], Patrick Biller[2], Marianne Glasius[1*]

[1] Interdisciplinary Nanoscience Center and Department of Chemistry, Aarhus University, Langelandsgade 140, Aarhus, Denmark

[2] Biological and Chemical Engineering, Aarhus University, Hangøvej 2, Aarhus, Denmark

* Corresponding author: glasius@chem.au.dk

Abstract

This paper investigates the use of Fourier-transform infrared spectroscopy (FT-IR) for quantitative analysis of bio-crudes from hydrothermal liquefaction (HTL) of biomass. HTL is a versatile process rendering virtually all biomasses suitable for conversion into bio-crude and side-streams. However, continuous processes require rapid analytical methods applicable to highly diverse bio-crudes. Bio-crudes were obtained from two different continuous HTL reactors (lab scale and pilot scale) and in some cases recirculation of water. The bio-crudes originated from a diverse range of feedstocks including lignocellulosics (pine, *Miscanthus*), microalgae (*Spirulina*, *Chlorella vulgaris*), and residues (sludge, dried distillers grains with solubles). Quantitative analysis of water content, total acid number, and total content of phenolics was performed using FT-IR. Principal component analysis indicated a potential correlation between quantitative measurements and FT-IR. Partial least squares regression was used to develop predictive models that performed well considering the high diversity of the bio-crudes. The content of phenolics was 83.1 – 254.6 mg g^{-1}(gallic acid equivalent) and model calibration was good (RMSE = 19.7, slope = 0.81, y-exp = 81.2%). A diverse set of test samples were subjected to the models. The relative difference for measured and predicted phenolic content was generally < 15%. Total acid numbers (TAN) were 7 – 98 mg_{KOH} g^{-1} and model calibration was found to be satisfactory considering the titration method used (RMSE = 18.5, slope = 0.53, y-exp = 52.6%). The relative difference for measured and predicted TAN was generally < 20%. The

water content (Karl Fischer titration) was 1 – 24 % and model calibration was very good (RMSE = 2.0, slope = 0.93, y-exp = 92.6%). The water content was generally predicted within 1.5% and the relative difference for measured and predicted water content was large (2.7 – 16.6%) due to the small values. All models included samples that deviated and could be considered outliers, however, their deviations were explained from their composition and retained in the models. Overall, the results show the potential of FT-IR as a universal technique to obtain rapid quantitative results from a variety of bio-crudes processed using different reactors.

Keywords: Hydrothermal liquefaction; PLS regression; Quantitative analysis; Predictive modelling; FT-IR

1. Introduction

Hydrothermal liquefaction (HTL) is a promising technique for conversion of wet biomass into a bio-crude with potential for further upgrading into a combustible fuel equivalent to diesel or aviation fuel.[1] The HTL process is carried out in an aqueous slurry at subcritical conditions with typical solid loadings of 5-35 wt.%.[2, 3] The resulting bio-crude has decreased contents of nitrogen and oxygen and correspondingly increased carbon content and higher heating value compared to the biomass. Multiple process variables have been found to influence the outcome from HTL including; biomass type, biomass pretreatment, solid loading, catalyst loading, reactor type, reaction temperature, heating rate, reaction time, and work-up procedure.[4] Interpretation of the effects of these variables are typically based on yields and elemental composition. The elemental composition is determined from combustion where oxygen is most often determined by difference. The oxygen content may arise from different sources including water. The water content is determined by time-consuming Karl Fischer titration requiring the use of hazardous reagents and the need for sample volumes most often not obtained during batch experiments.

Lipid-containing feedstocks, such as sludge and some microalgae, result in bio-crudes abundant with fatty acids while lignocellulosics produce high amounts of small organic acids (although they are mostly water soluble) contributing to the oxygen content.[5, 6] The carboxylic acids are a potential source of corrosion and often measured as a total acid number (TAN) from the amount of potassium hydroxide needed to neutralize

one gram of bio-crude. Hence, time-consuming methods such as titration in organic solvent are needed to determine the TAN with consumption of sample volumes corresponding to the yields from micro batch experiments (~ 200mg). The ASTM method for TAN includes the use of an indicator for which the endpoint can be very difficult to determine, as bio-crudes are intensely black, which introduces potential error to the method. The method can be improved with the use of potentiometric titration but it remains time-consuming and involves the use of organic solvents.[7]

Lignin is the most important source of phenolics in bio-crudes but may also arise from degradation of carbohydrates and proteins.[8, 9] They are present in the volatile and semi-volatile fraction as derivatives of phenols, catechols, hydroquinones, and hydroxypyridines.[9, 10] While phenolics exhibit medium to high reactivity towards deoxygenation, hydrogenation of the benzene ring requires increased amounts of hydrogen increasing the expense of the upgrading process.[11] Furthermore, the phenolic compounds are reactive and studies have shown that instability of bio-crudes is linked to formation of cyclic and aromatic compounds from phenolics.[12] Determining the total amount of phenolics is rarely done for bio-crudes and requires extraction into organic solvents, addition of reagents, two hours reaction time, and subsequent colorimetric analysis.[13]

The laborious and time-consuming methods required for complete general analysis of a single bio-crude sample limits the number of samples being analyzed and introduces additional costs to the process while substantial volumes of organic solvents are required. The least laborious and time-consuming chemical analysis of bio-crude is the use of Fourier-transform infrared spectroscopy (FT-IR), which also does not involve the use of organic solvents. IR spectroscopy exploits the fact that molecules absorb light at resonant frequencies based on vibrational modes as long as it introduces a change in dipole moment. The resonant frequency is dependent on the bond strength and the mass of the atoms leading to a number of regions characteristic of functional groups. The number of vibrational modes in a given molecule is again dependent on the number of atoms. The characteristic resonant frequencies and the development of chemometrics in recent decades has led to a widespread use of FT-IR to predict results from time-consuming or expensive chemical analyses.[14-16] The use of partial least squares regression (PLS-R) is valuable in combination with methods for variable selection as the data set is greatly reduced to a few relevant latent variables used for

prediction. The use of aqueous slurry in HTL means that any biomass, which can be suspended in water is amenable to HTL resulting in a vast number of feedstock applications. The resulting bio-crudes hence vary in physical properties depending on feedstock type but also on reactor type and reaction parameter, which may also require different product separation techniques. This means that a truly universal predictive model from FT-IR has to handle the presence and absence of a variety of potential overlapping absorptions. Several studies have identified the relation between FT-IR and the phenolic content of bio-oils from pyrolysis where typically a specific wavenumber is chosen.[17] However, the versatility of HTL and the complexity of the resulting bio-crudes likely means that multiple sections of wavenumbers are required for HTL bio-crudes. Hence, hundreds of wavenumbers may be necessary to account for the variations.

In this work, we demonstrate the application of FT-IR and PLS-R to predict water content, TAN, and total phenolic content in HTL bio-crude. The bio-crudes were produced from varying feedstocks including dried distillers grains with solubles (DDGS), pine, *Miscanthus*, *Spirulina*, *Chlorella,* and primary sludge. DDGS was also processed with aqueous phase recirculation. The bio-crudes were chosen to gain the largest compositional variation in order to obtain models applicable to bio-crudes of a variety of feedstocks. Furthermore, the feedstocks were processed on two different continuous flow systems (lab scale and pilot scale) by two separate research groups to account for potential variation in processing and separation conditions. The results show the potential to apply FT-IR for high-throughput analysis of HTL bio-crudes to determine relevant industrial parameters within minutes instead of days and without the use of reagents and organic solvents.

2. Materials and methods

2.1 Material and reagents

Absolute ethanol and potassium hydroxide were obtained from Merck Chemicals. Phenolphthalein, potassium phthalate, sodium carbonate, gallic acid, and Folin-Ciocalteu phenol reagent were obtained from Sigma Aldrich. Hydranal® solvent CM and Hydranal® titrant 5 was from Honeywell.

2.2 Hydrothermal liquefaction

Biomasses were processed with two separate continuous flow reactors: 1) a lab scale reactor, 2) a pilot scale reactor. The lab scale reactor has been presented in detail elsewhere and only the biomass and their processing parameters and work-up procedures are presented.[18] The pilot scale reactor was operated at 350 ˚C with a flow rate of ~1 L min^{-1} (+/- 0.2 L min^{-1}). Residence time at 350 ˚C was approximately 10 minutes with a heat up time from ambient to 350 ˚C of 5 minutes and a cooling to 70 ˚C of 5.5 minutes. At 70 ˚C, the pressure was released from approximately 220 bar to atmospheric pressure. The reactor design is of plug flow type with an internal pipe diameter of 14 mm and a total reactor length of 65 meters (excluding 49.2 m heat exchanger, 12.6 m trim heater, 5.5 m cooling).

All biomasses were processed in water without the addition of co-solvents.

2.2.1. Lab scale HTL

DDGS was processed with 20 wt.% solid loading and 2 wt.% potassium carbonate at 350 ˚C and 20 min reaction time. In one case, DDGS was processed with 2 wt.% citric acid instead of potassium carbonate. No size reduction was needed for DDGS. *Miscanthus* (only stems) was chopped and milled prior to pretreatment with 1 M sodium hydroxide as previously reported.[19] The pretreated biomass was processed with 10 wt.% solid loading and 2 wt.% potassium carbonate at 350 ˚C and 20 min reaction time. The bio-crudes were separated from the aqueous phase by cooling in an ice bath and subsequent centrifugation. The aqueous phase from DDGS was subsequently recirculated for slurry preparation in two consecutive recirculation experiments described in detail elsewhere.[20] Bio-crudes were stored at 5 ˚C until further analysis.

2.2.2 Pilot scale HTL

Pine and *Miscanthus* (stems and leaves) were chopped and extruded prior to processing. The extruded biomass was processed with 15-19 wt.% solid loading and 1.4 wt.% potassium hydroxide and 0.5 wt.% carboxymethylcellulose (CMC). *Spirulina* and primary sludge was processed with 16 wt.% and 6 wt.% solid loading, respectively, both without the addition of alkali or CMC. All biomasses were processed at 350 ˚C and 10 min reaction time. Bio-crudes were gravimetrically separated from the aqueous phase. Bio-crudes were stored at 5 ˚C until further analysis.

2.3 Analytical methods

Varying amounts of bio-crude were obtained from the different experiments meaning that sampling was experiment dependent. Laboratory scale experiments sampling was performed on the entire sample while for pilot scale experiments a subsample of 10-40 g was obtained on which further sampling was conducted. All measurements were carried out in triplicates.

Water content was determined from Karl Fischer titration (two component). Bio-crude was dissolved in 5.0 ml of hydranal solvent to obtain sample solutions with 1-10 mg of water. A methanol free hydranal solvent was used to avoid formation of acetals, ketals, and resulting water especially from lignocellulosics. However, a methanol based titrant was used, which could be a potential source of error.

TAN was determined by dissolving approximately 20 mg of bio-crude in 50 ml absolute ethanol leading to an orange colored solution. Phenolphthalein (15 drops, 2 wt.% in ethanol) was added to the solution and it was titrated with potassium hydroxide (0.02 M in ethanol) until persistent color change. When titration is performed in organic solvents the end-point is often transient, which was also observed during titration of our samples making determination of the end-point highly subjective. TAN was reported as mg potassium hydroxide required to neutralize 1 g of bio-crude. The titrant was standardized with potassium phthalate.

For determination of total phenolic content bio-crude was first dissolved in absolute ethanol (10 mg ml^{-1}) and filtered with a syringe filter (millipore, 0.45 µm). An aliquot of 20 µl was transferred to a vial along with 1.58 ml deionized water and 100 µl Folin-Ciocalteu reagent.[13] The content was mixed and incubated at ambient temperature for 5 min. Sodium carbonate solution (300 µl) was added and the vial was stored at ambient temperature for 2 h. Sample absorbance was measured at 765 nm. Samples were appropriately diluted to obtain absorbances of 0.08 – 0.80. Lambert-Beers law was used to calculate the concentration based on gallic acid standards and the results are reported as gallic acid equivalent.

FT-IR was performed with a Techno Nicolet 380 with attenuated total reflectance (ATR) in the spectral region of 4000-400 cm^{-1}. Each sample spectrum was collected at a resolution of 2 cm^{-1} at 32 scans per sample. Prior to loading of the sample, the diamond was thoroughly cleaned with ethanol and a background spectrum was obtained before each sample. The diversity in sample viscosity meant that samples were either

placed as a droplet (e.g. *Spirulina*) or smeared across the diamond with a spatula (e.g. pine). A total of 28 samples were analyzed.

2.4 Partial least squares regression

The main goal of this study was to develop a multivariate model capable of predicting the quantitative measurements based on time-consuming methods. This is done with PLS-R, which determines a set of regression coefficients that correlate a set of dependent variables (X) to a set of independent variables (Y). The X data (FT-IR) and Y data (e.g. TAN) are decomposed to a set of scores (T and U) and loadings (P' and Q') along with residuals (E and F) where the covariance is maximized. Subsequently the correlation between the scores is maximized leading to a set of latent variables (LV) where X and Y are related through a set of regression coefficients (B).

$$X = TP' + E$$

$$Y = UQ' + F$$

$$Y = BX$$

Generally, it is necessary to preprocess spectroscopic data prior to PLS-R. The Savitzky-Golay algorithm was applied as preliminary preprocessing to enhance signal properties and for de-noising.[21, 22] The following settings were applied for the Savitzky-Golay algorithm; width = 15, first order polynomial, first derivative. The X-block and Y-block data were randomly split into a calibration set of 21 samples and a test set (7 samples). The test set was used to estimate the model performance on samples that were not part of the calibration data. Initially the wavelengths 2000-2500 cm^{-1} were removed as they contained a large degree of noise. Further variable selection was performed with selectivity ratios of the spectral variables.[23] The threshold for selectivity ratios varied depending on the y-block with an approximate threshold of 0.25. This means that any variable (wavenumber) with a selectivity ratio less than the threshold is omitted from the model development and calibration and prediction is performed on the remaining variables.

The following figures of merit have been calculated to evaluate and validate the different models: root mean squared error of calibration (RMSEC), root mean squared error of cross validation (RMSECV), and bias.

178 The models were further assessed based on predicted versus determined values for calibration, cross
179 validation, and test set. Root mean squared error of prediction (RMSEP) was found to be misleading since in
180 some cases a sample would deviate from the model, which significantly influences the RMSEP value.
181 The use of PLS-R for predicting quantitative measures is presented here as a proof of concept. The combined
182 application of data preprocessing and PLS-R means that the regression coefficients consist of several
183 hundred wavenumbers, which are obtained after the exact preprocessing outlined for the Savitzky-Golay
184 algorithm.

185 ## 3. Results and discussion

186 ### 3.1 Principal component analysis

187 Initially a principal component analysis (PCA) was carried out with the preprocessed data (without variable
188 selection) for exploration and to identify potential outliers. The initial PCA specifically modeled two
189 segments (460-400 cm^{-1}, 2500-2000 cm^{-1}) of variables that were mostly noise and were removed prior to
190 further analysis. A total of 8 components were required to explain 84% of the variance, however, the first
191 four components explained 69% of the variance.
192 PC1 (37.1%) and PC2 (15.9%) provided an almost complete grouping of bio-crudes based on the feedstock
193 used (Fig. 1). PC1 predominantly separated bio-crudes of lignocellulosics from bio-crudes of protein- and
194 lipid-rich feedstocks. *Miscanthus* was processed with both reactors and the bio-crude and aqueous phase
195 were separated with either cold centrifugation or gravimetrically. These samples were only just separated by
196 PC1 indicating that reactor and separation technique did not influence the specific bio-crude. PC1 effectively
197 separated pine from *Miscanthus,* and furthermore decreasing scores were obtained for pine with increasing
198 reactor operation time showing that steady-state conditions had not been reached. PC1 also effectively
199 separated bio-crudes of DDGS, *C. vulgaris*, and *Spirulina* with respectively increasing scores. PC2 clearly
200 separated bio-crude from DDGS with recirculation of water from the remaining samples with positive scores
201 increasing with the number of recirculations while negative scores were obtained for sludge. The ability of
202 FT-IR combined with PCA to distinguish between samples from different feedstocks, process parameters,

and the use of water recirculation indicates the potential use for process control, which will be highly relevant at commercial scale.

Many loadings were found to be important for PC1 and PC2 but positive loadings for PC1 and PC2 were especially found for regions of amide and nitrogen containing aromatics, while positive loadings for PC1 and negative loadings for PC2 were found for two regions of fatty acids corresponding mainly to saturated fatty acids. Negative loadings for PC1 were especially found for regions of increased unsaturation and markedly for regions of aromaticity. Additionally, negative loadings for PC1 were observed for regions of CH_2 stretching. Since phenolic compounds are mainly found as phenols, catechols, hydroquinones, and hydroxypyridines, the separations by PC1 and PC2 indicate a correlation with the total phenolic content when moving towards lower PC1 scores and higher PC2 scores, which is also clearly observed for PC1 and less obvious for PC2 in Fig. 1. PC2 instead seems to capture the difference in fatty acid composition. The TAN also appears to correlate particularly with PC1 (Fig. 1). Since PC1 was mainly influenced by wavenumbers of phenolic compounds, the correlation between TAN and PC1 is an inverse correlation to the phenolic content as microalgae, DDGS, and sludge contain limited lignin and have higher lipid contents than lignocellulosics.

The water content of the bio-crudes was generally found to increase from microalgae to lignocellulosics with a further increase by recirculation of water phase. The increased water content from microalgae to lignocellulosics is likely a combination of increased viscosity observed for bio-crudes of lignocellulosic and increased contents of polar compounds, such as small organic acids and phenolics,[6] leading to binding and trapping of water. The increasing water content with recirculation may occur due to increasing contents of polar compounds caused by a continuous increase of the same components in the aqueous phase until saturation is reached.[24]

PC3 (9.7%) and PC4 (6.7%) separated bio-crude based on their differences in water content. One sample of microalgae was an extreme sample but was not removed prior to variable selection. Additional PCs did not provide trends that could be traced back to specific regions of interest.

3.2 Total phenolic content

A number of studies have identified the presence of a range of phenolic compounds in the water phase, solid residue, and bio-crude from HTL.[8, 25, 26] Phenolics are distributed among the volatile, semi-volatile, and non-volatile fractions in the bio-crude. Phenolics are produced from lignin, carbohydrates, and proteins in descending order of abundance with lignin being the major contributor. The total phenolic content is typically not reported in literature but single phenolic compounds have been shown to comprise up to 0.5 wt.% in bio-crudes from both microalgae and lignocellulosics.[6] The greatest diversity of phenolics is obtained from lignocellulosics where alkylated phenolics, catechols, and hydroquinones are found of which phenolics and catechols are known for repolymerization reactions. In contrast, the phenolic compounds obtained from carbohydrates are predominantly alkylated phenols, which may also be present as hydroxypyridines depending on the biochemical content.[10]

PLS-R was performed to confirm the correlation observed between FT-IR and total phenolic content from PCA. Previous studies of pyrolysis bio-oil from pine have identified the correlation between the phenolic content and aromatic bands in FT-IR (1640-1500 cm^{-1}).[12] In this study, the most important variables were identified from selectivity ratios (threshold = 0.25), which included 6 specific regions, listed with approximate wavelengths; 1) 1140 cm^{-1}, 2) 1280 cm^{-1}, 3) 1515 cm^{-1}, 4) 1580 cm^{-1}, 5) 2850 cm^{-1}, 6) 2960 cm^{-1} (Fig. 2). The multiple regions of importance are likely a result of the higher complexity of HTL bio-crudes compared to pyrolysis bio-oils, specifically with the presence of polyaromatic hydrocarbons that are not detected by the method.[27] The first two regions may occur from C-O bending, which are both strong and wide for many phenol and hydroquinone derivatives while it is less intense for catechol derivatives. The third and fourth regions are from C=C stretching, which typically occur in a series of 3-4 bands but may also arise from non-heteroatom containing aromatics. The fifth and sixth region is associated with sp^3 hybridized stretching of CH_3 and CH_2 groups. More specifically the fifth region is likely from methoxy groups of partially degraded lignin while the sixth region may be long carbon chains associated with phenolics. It is noteworthy that the region at approximately 3600 cm^{-1} is not found to be important even though it strongly absorbs in one band from phenols and in two bands from catechols. This may be due to the presence of fatty amides in bio-crudes from protein-rich feedstocks or the presence of water.

Wavenumbers with selectivity ratios less than 0.25 were removed and calibration was performed with the six remaining regions. A total of 3 latent variables were selected based on RMSECV values and interpretation of predicted and measured values for cross validation and prediction. The calibration was good (RMSE = 19.7, slope = 0.81, y-exp = 81.2%) considering the number of samples while the statistics deteriorated with cross validation (RMSECV = 24.4, slope = 0.75, y-exp = 71.2%) indicating that the model may lack robustness. The cross validation statistics can be explained by the small data set. The test samples were predicted well apart from a single sample (RMSEP = 19.7, slope = 0.78, y-exp = 88.8%).

The measured phenolic contents were as expected since bio-crude of *Spirulina* and DDGS contain phenolics predominantly from degradation of protein containing phenylalanine and tyrosine and from dehydration of carbohydrates. The higher phenolic content of bio-crudes from *Miscanthus* and pine arises from the lignin content. Generally, the relative differences between predicted values and measured values were ≤ 15% except for a single sample, which showed a relative difference of 29.5% (Pine).

The relative differences may arise from a number of sources that include the analytical method and the sample composition. The analytical method has been standardized for phenolic compounds of pyrolysis bio-oil where the sample is dissolved in ethanol and passed through syringe filters. The diversity of HTL bio-crudes may lead to phenolic species, such as polymeric material, in some samples that are poorly dissolved in ethanol leading to lower measured values. We have previously identified modified residual lignin and repolymerized lignin in the solid residue from batch experiments of lignocellulosics,[26] which may explain the relatively large deviation for pine. The diversity of HTL bio-crudes also means that a variety of phenolics will be abundant in different amounts, which will vary in their specific absorptivities.[28] Another possible explanation is that the high lignin content of pine leads to a higher abundance of catechol-derived compounds leading to an under-prediction due to the high importance of region 1 and 2 (strongest in phenols and hydroquinones) and the small importance of wavelength 3600 cm^{-1} (strongest in phenols and catechols) observed from the selectivity ratios. Hence, development of high quality models requires a larger data set with a more normally distributed phenolic content and it may require separate models for bio-crudes from feedstocks with similar biochemical contents and at the same time a mixture of solvents may be required for dissolving the bio-crude.

The regressions coefficients in Figure 3 showed multiple sections with positive values, which included C-O stretching (1175-1130 cm^{-1}), C=C stretching (1590-1550 cm^{-1}), OCH_3 stretching from partially degraded lignin (2860-2850 cm^{-1}), and unfunctionalized C-H stretching (2950-2925 cm^{-1}). The negative regression coefficient were also numerous and could arise from indoles and O-H bending of alcohols (1290-1270 cm^{-1}), C=C stretching of polyaromatic hydrocarbons (1530-1510 cm^{-1}), and unsaturated cyclic ketones such as cyclopent-2-enones (1630-1610 cm^{-1}).

3.3 Total acid number

The TAN of bio-crudes is a consequence of the abundance of carboxylic acids mainly from fatty acids of C16, C18, and C20, which apart from palmitic acid can be found as mono-, di-, tri-, and tetraunsaturated in the 3, 6, 9, and 12 position (notation from the carboxylic acid group).[9] Some publications also report the presence of acetic acid and lactic acid.[6] TAN values were determined between 7 mg g^{-1} (pine) and 98 mg g^{-1}(DDGS), which are equivalent to TAN values from bio-crudes of lignocellulosics and high lipid-containing feedstocks published in the literature.[27, 29]

PLS-R confirmed the correlation between FT-IR and TAN. The selectivity ratios showed a strong resemblance to the selectivity ratios from the previous PLS-R with phenolics (Fig. 4). Although the regions were more diffuse, the regions of 1100 cm^{-1}, 1550 cm^{-1}, and 1610 cm^{-1} were comparative to phenolics. The strong resemblance shows that TAN may largely be inverse proportional to the total phenolic content, which can be extended to the feedstocks in the form of microalgae (lipid-rich) and lignocellulosic (lignin-rich). This confirms the observation made in PCA where the same PCs were used to explain the variance. The more characteristic absorbances of carboxylic acids were also observed at 3000-2500 cm^{-1} (O-H, extending into 3200 cm^{-1}) and 1725-1700 cm^{-1} (C=O) along with a more dense region at 1410-1260 cm^{-1} (O-H). The abundance of long chain fatty acids may also contribute with C-H stretching at 3000-2900 cm^{-1}. It is interesting to note that the model found the region of C=O stretching to be moderately important, which may result from the large number of unsaturated ketones typically encountered in bio-crudes.

Wavenumbers with selectivity ratios less than 0.3 were removed and calibration was performed with the remaining regions. A total of 2 latent variables was selected based on RMSECV values and interpretation of

predicted and measured values for cross validation and prediction. The calibration was found to be satisfactory (RMSE = 18.5, slope = 0.53, y-exp = 52.6%) while the statistics substantially deteriorated with cross validation (RMSECV = 20.2, slope = 0.47, y-exp = 43.7%), which was again explained by the small data set (Fig. 4). The test samples were predicted well except for one sample, which influenced the statistics (RMSEP = 19.6, slope = 0.52, y-exp = 45.7%).

The predictions of the test set were clearly divided into two groups. Samples with TANs < 50 $mg_{KOH}g^{-1}$ were predicted within 4.3 $mg_{KOH}g^{-1}$ difference while the difference was at least three times as high for samples with TANs > 50 $mg_{KOH}g^{-1}$ (Table 2). The poor predicting ability is also observed from the calibration set, which clearly show that TAN is poorly determined at higher values. Spirulina was very poorly predicted, which can be explained by the very low phenolic content along with a moderate content of fatty acids. This combination provides difficulty for the model since it is to some extent predicting based on the absence of phenolics. The fact that the model partly uses the absence of phenolics for prediction is unfortunate as phenolics have low acidity and will thus influence the endpoint. Improving the model will require a more reliable titration method than the use of an indicator, such as potentiometric titration. Additionally it should be noted that the TAN is routinely determined for fossil oils making the endpoint more easily observed and the organic solvents applied are known to fully dissolve the oil. The diversity of bio-crudes means that a single solvent mixture is less likely to fully dissolve all bio-crudes while the color of the solution is bio-crude dependent making determination of the endpoint less standardized.

The regression coefficients confirmed some degree of inverse proportionality of TAN to phenolics. Specifically, the regression coefficients at 1100 cm^{-1} and 1555 cm^{-1} were negative along with a large negative regression coefficient at 1880 cm^{-1}, which is the band used to assign substitution patterns for phenolics (Fig. 5). Negative coefficients were also found for 2960 cm^{-1} (CH_3) while positive coefficients were found at 2940 cm^{-1} (CH_2). The band ratios at 3000-2800 cm^{-1} have previously been used to evaluate the carbon chain length and degree of branching of petroleum fractions.[30] An increasing number of fatty acids will thus increase the number of CH_2 groups while a larger number of cyclic compounds, such as phenolics, will likely increase the number of CH_3 groups. In general, positive coefficients were observed at 2940-2750

334 cm^{-1}, which is part of the characteristic region for carboxylic acids. However, similar or larger coefficients
335 were obtained in the equally characteristic regions at 1270 cm^{-1} and 1700 cm^{-1}.

336 **3.4 Moisture content**

337 The water content of bio-crudes from continuous HTL is often higher than from batch studies due to the
338 separation technique used. The continuous HTL process employed for the samples in this experiment enables
339 gravimetric separation leaving behind 1-24% water in the bio-crude spanning the range of water content
340 reported from most other studies.[20, 31] Similar to TAN and total phenolic content, the water content is to some
341 extent feedstock-specific with protein-rich and lipid-rich feedstocks having the lowest water contents,
342 lignocellulosics having higher contents, and recirculation increasing the water content further. Although this
343 observation is clearly dependent on the separation technique.
344 Mid-IR is typically not well suited for determination of water due to the presence of other absorbing
345 compound classes at similar wavelengths, especially the region of 3500-3000 cm^{-1}. However, the very low
346 abundance of free amines and alcohols in bio-crudes means that a minor degree of co-absorption occurs.
347 Furthermore, the majority of carboxylic acids are present as long chain fatty acids, such as palmitic acid,
348 which typically have absorptions limited to 3000 cm^{-1}. A previous study of pyrolysis bio-oil from pine
349 showed correlation between water content and the band at 3400 cm^{-1} corresponding to the maximum
350 absorption of water from symmetrical and asymmetrical stretch.[17] In the current study, PLS-R also identified
351 a correlation between FT-IR and water content but selectivity ratios showed three regions of importance
352 (Fig. 6). The region around 700 cm^{-1} corresponds to librational movements (rocking motion) while the band
353 at 3670-2960 cm^{-1} corresponds to symmetrical and asymmetrical stretching. It is also observed that the
354 otherwise strong absorption from bending motion of water at 1640 cm^{-1} is not useful due to the presence of
355 numerous unsaturated compounds.
356 Wavenumbers with selectivity ratios less than 0.25 were removed and calibration was performed with the
357 remaining regions. A model with four latent variables was made excluding variables with selectivity ratios
358 less than the threshold. The model performance for the calibration data was very good (RMSE = 2.0, slope =
359 0.93, y-exp = 92.6%) apart from one sample that was processed with citric acid, whose absorption extends

until 3600 cm^{-1} and therefore may influence the model. Cross validation did not change the model performance markedly (RMSECV = 3.0, slope = 0.90, y-exp = 84.4%) showing the robustness of the model. The test samples were predicted well (RMSEP = 1.0, slope = 1.0, y-exp = 97.5).

The sample set for water included only 25 samples since there was not sufficient material to determine the water content of three samples. Therefore, the test set included only six samples. The difference of predicted and measured value was generally satisfactory (< 1.5%) considering the viscosity of bio-crudes, which is likely to affect the sampling especially if water is present in small pockets (Table 3). Water determination with Karl Fischer is widely applied to many different matrices. In bio-crudes water may be both bound and free and lack of homogenization may make sampling a potential source of error. The pine bio-crude sample was the most viscous and small droplets of water appeared when smearing out the sample, which may account for the higher predicted water content. Further investigation is required to identify best practice for analyzing particularly viscous HTL bio-crudes with ATR FT-IR. Water content of bio-crudes of protein-rich feedstock were generally underestimated, which may be caused by the presence of fatty amides, which absorb strongly in the same regions as water leading to a decrease in the predicted value (see regression coefficients).

The regression coefficients were positive for several bands related to water including 820-680 cm^{-1}, 3300-3000 cm^{-1}, and 3670-3560 cm^{-1} (Fig. 7). The positive coefficients for the bands of 860 cm^{-1} and 2900 cm^{-1} could not be explained but are likely due to specific compound classes such as carboxylic acids related to higher water contents. It is interesting to note that negative coefficients were obtained for the band of 3560-3300 cm^{-1}, which not only corresponds to absorption of water but also the absorption of fatty amides. A previous study of pyrolysis bio-oil from pine showed correlation between water content and the absorption at 3400 cm^{-1}.[17] In the current study, a broader approach to feedstocks was chosen and the presence of fatty amides shows that multiple bands are required for HTL bio-crudes when a universal model is required.

4. Conclusion

The versatility of HTL makes almost any biomass a potential feedstock, while continuous HTL leads to a large volume of bio-crude and samples that require high-throughput and quantitative analysis. In this work,

bio-crude from HTL of microalgae (*C. vulgaris*, *Spirulina*), DDGS, sludge, and lignocellulosics (pine, *Miscanthus*) obtained from two different continuous reactors were analyzed for total phenolics (gallic acid equivalent), TAN, and water content. PLS-R was used to correlate the results with FT-IR spectra providing good quality models for total phenolics and water content while TAN only performed well for the lower half of TAN concentrations. The models were used to satisfactorily predict quantitative values in test samples. The relative differences may be accounted for by a number of factors, which include low acidity compounds (TAN), specific absorptivities (total phenolics), and sampling (water content). Regression coefficients showed that multiple bands are required to develop the models and the regression coefficients do not necessarily follow the absorption of the specific compound class, which were related to the known composition of bio-crudes. The work shows that a universal model based on FT-IR can be constructed to quantitatively predict key characteristics of bio-crude, valuable for both the HTL process and downstream processing.

Acknowledgement

We would like to acknowledge Professor Bo B. Iversen and his research group for providing bio-crude samples from lab scale experiments. This study has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 764734 (project HyFlexFuel). The work was carried out under the auspices of Aarhus University Center for Circular Bioeconomy (CBIO).

References

1. Biller, P.; Sharma, B. J.; Kunwar, B.; Ross, A. B., Hydroprocessing of bio-crude from continuous hydrothermal liquefaction of microalgae. *Fuel* **2015,** 159, 197-205.
2. Peterson, A. A.; Vogel, F.; Lachance, R. P.; Fröling, M.; Antal, M. J.; Tester, J. W., Thermochemical biofuel production in hydrothermal media: A review of sub- and supercritical water technologies. *Energy Environ. Sci.* **2008,** 1, 32-65.
3. Elliott, D. C.; Biller, P.; Ross, A. B.; Schmidt, A. J.; Jones, S. B., Hydrothermal liquefaction of biomass: Developments from batch to continuous process. *Biores. Technol.* **2015,** 178, 147-156.
4. Dimitriadis, A.; Bezergianni, S., Hydrothermal liquefaction of various biomass and waste feedstocks for biocrude production: A state of the art review. *Renew. Sus. Energy Rev.* **2017,** 68, 113-125.
5. Wang, W.; Yu, Q.; Meng, H.; Han, W.; Li, J.; Zhang, J., Catalytic liquefaction of municipal sewage sludge over transition metal catalysts in ethanol-water co-solvent. *Biores. Technol.* **2018,** 249, 361-367.

6. Madsen, R. B.; Bernberg, R. Z. K.; Biller, P.; Becker, J.; Iversen, B. B.; Glasius, M., Hydrothermal co-liquefaction of biomasses – quantitative analysis of bio-crude and aqueous phase composition. *Sust. Energy Fuels* **2017,** 1, 789-805.
7. Ferrell, J. R.; Olarte, M. V.; Christensen, E. D.; Padmaperuma, A. B.; Connatser, R. M.; Stankovikj, F.; Meier, D.; Paasikallio, V., Standardization of chemical analytical techniques for pyrolysis bio-oil: history, challenges, and current status of methods. *Biofuels Bioprod. Bioref* **2016,** 10, 496-507.
8. Pedersen, T.; Rosendahl, L., Production of fuel range oxygenates by supercritical hydrothermal liquefaction of lignocellulosic model systems. *Biomass Bioenergy* **2015,** 83, 206-215.
9. Madsen, R. B.; Zhang, H.; Biller, P.; Goldstein, A. H.; Glasius, M., Characterizing Semivolatile Organic Compounds of Biocrude from Hydrothermal Liquefaction of Biomass. *Energy Fuels* **2017,** 31, 4122-4134.
10. Croce, A.; Battistel, E.; Chiaberge, S.; Spera, S.; De Angelis, F.; Reale, S., A Model Study to Unravel the Complexity of Bio-Oil from Organic Wastes. *Chem. Sus. Chem.* **2016,** 9, 1-12.
11. Jensen, C. U.; Hoffmann, J.; Rosendahl, L., Co-processing potential of HTL bio-crude at petroleum refineries. Part 2: A parametric hydrotreating study. *Fuel* **2016,** 165, 536-543.
12. Kosinkova, J.; Ramirez, J. A.; Ristovski, Z. D.; Brown, R.; Rainey, T. J., Physical and Chemical Stability of Bagasse Biocrude from Liquefaction Stored in Real Conditions. *Energy Fuels* **2016,** 30, 10499-10504.
13. Rover, M. R.; Brown, R. C., Quantification of total phenols in bio-oil using the Folin–Ciocalteu method. *J. App. Anal. Pyrolysis* **2013,** 104, 366-371.
14. Song, S. Y.; Lee, Y. K.; Kim, I., Sugar and acid content of Citrus prediction modeling using FT-IR fingerprinting in combination with multivariate statistical analysis. *Food Chem.* **2016,** 190, 1027-1032.
15. Paulo, J. M.; Barros, J. E. M.; Barbeira, P. L. S., A PLS regression model using flame spectroscopy emission for determination of octane numbers in gasoline. *Fuel* **2016,** 176, 216-221.
16. Guryanova, A.; Ermakov, V.; Galyanin, V.; Artyushenko, V.; Sakharova, T.; Usenov, I.; Bykov, D.; Bogomolov, A., Quantitative analysis of total hydrocarbons and water in oil-contaminated soils with attenuated total reflection infrared spectroscopy. *J. Chemometrics* **2017**, doi.org/10.1002/cem.2826.
17. Stankovikj, F.; McDonald, A. G.; Helms, G. L.; Garcia-Perez, M., Quantification of Bio-Oil Functional Groups and Evidences of the Presence of Pyrolytic Humins. *Energy Fuels* **2016,** 30, 6505-6524.
18. Mørup, A. J.; Becker, J.; Christensen, P. S.; Houlberg, K.; Lappa, E.; Klemmer, M.; Madsen, R. B.; Glasius, M.; Iversen, B. B., Construction and Commissioning of a Continuous Reactor for Hydrothermal Liquefaction. *Ind. Eng. Chem. Res.* **2015,** 54, 5935-5947.
19. Lappa, E.; Christensen, P. S.; Klemmer, M.; Becker, J.; Iversen, B. B., Hydrothermal liquefaction of Miscanthus Giganteus: Preparation of the ideal feedstock. *Biomass Bioenergy* **2016,** 87.
20. Klemmer, M.; Madsen, R. B.; Houlberg, K.; Mørup, A. J.; Christensen, P. S.; Becker, J.; Glasius, M.; Iversen, B. B., Effect of Aqueous Phase Recycling in Continuous Hydrothermal Liquefaction. *Ind. Eng. Chem. Res.* **2016,** 55, 12317-12325.
21. Savitzky, A.; Golay, M. J. E., Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964,** 36, 1627-1639.
22. Svecnjak, L.; Bubalo, D.; Baranovic, G.; Novosel, H., Optimization of FTIR-ATR spectroscopy for botanical authentication of unifloral honey types and melissopalynological data prediction. *Eur. Food Res. Technol.* **2015,** 240, 1101-1115.
23. Farrés, M.; Platikanov, S.; Tsakovski, S.; Tauler, R., Comparison of the variable importance inprojection (VIP) and of the selectivity ratio (SR)methods for variable selection andinterpretation. *J. Chemometrics* **2015,** 29, 528-536.
24. Biller, P.; Madsen, R. B.; Klemmer, M.; Becker, J.; Iversen, B. B.; Glasius, M., Effect of hydrothermal liquefaction aqueous phase recycling on bio-crude yields and composition. *Bioresour. Technol.* **2016,** 220, 190-199.

25. Panisko, E.; Wietsma, T.; Lemmon, T.; Albrecht, K.; Howe, D., Characterization of the aqueous fractions from hydrotreatment and hydrothermal liquefaction of lignocellulosic feedstocks. *Biomass Bioenergy* **2015,** 74, 162-171.
26. Madsen, R. B.; Jensen, M. M.; Glasius, M., Qualitative characterization of solid residue from hydrothermal liquefaction of biomass using thermochemolysis and stepwise pyrolysis-gas chromatography-mass spectrometry. *Sus. Energy Fuels* **2017,** 1, 2110-2119.
27. Jensen, C. U.; Rosendahl, L.; Olofsson, G., Impact of nitrogenous alkaline agent on continuous HTL of lignocellulosic biomass and biocrude upgrading. *Fuel Processing Technol.* **2017,** 159, 376-385.
28. Blainski, A.; Lopes, G. C.; Palazzo de Mello, J. C., Application and Analysis of the Folin Ciocalteu Method for the Determination of the Total Phenolic Content from Limonium Brasiliense L. *Molecules* **2013,** 18, 6852-6865.
29. Shakya, R.; Adhikari, S.; Mahadevan, R.; Shanmugam, S. R.; Nam, H.; Hassan, E. B.; Dempster, T. A., Influence of biochemical composition during hydrothermal liquefaction of algae on product yields and fuel properties. *Biores. Technol.* **2017,** 243, 1112-1120.
30. Pasadakis, N.; Livanos, G.; Zervakis, M., Deconvolving the absorbance of methyl and methylene groups in the FT-IR 3000-2800 cm-1 band of petroleum fractions. *Trends App. Spec.* **2013,** 10, 25-35.
31. Pedersen, T.; Grigoras, I. F.; Hoffmann, J.; Toor, S. S.; Daraban, I. M.; Jensen, C. U.; Iversen, S. B.; Madsen, R. B.; Glasius, M.; Arturi, K. R.; Nielsen, R. P.; Søgaard, E. G.; Rosendahl, L. A., Continuous hydrothermal co-liquefaction of aspen wood and glycerol with water phase recirculation. *App. Energy* **2016,** 162, 1034-1041.

Table 1. Measured values and predicted values for total phenolic content (mg g^{-1}, gallic acid equivalent) of test samples and their relative difference.

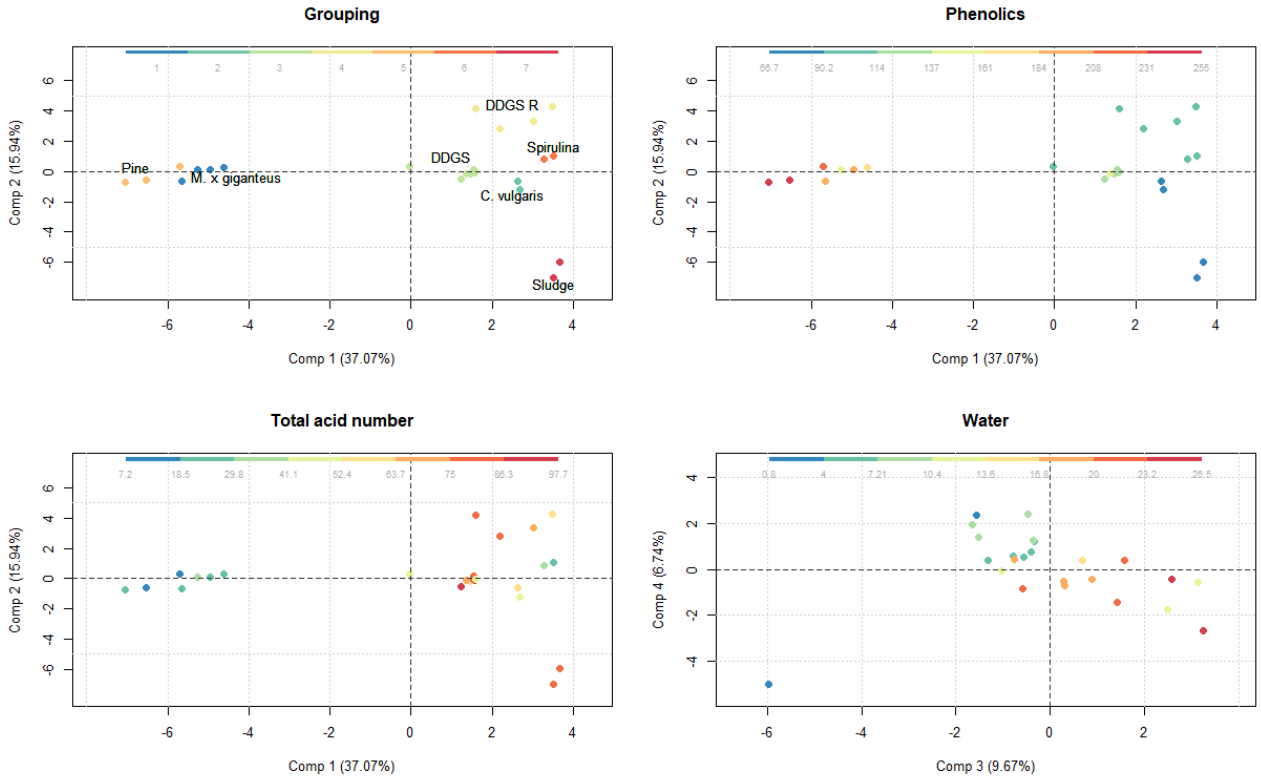
	Pine pilot	*Spirulina* pilot	Miscanthus lab	Miscanthus pilot	DDGS lab 350C	DDGS lab R	DDGS lab 330C
Measured	254.6	110.1	160.8	187.7	94.9	105.5	115.8
Predicted	221.1	112.1	186.3	194.6	124.4	102.7	112.7
Difference	33.5	2.0	25.5	26.9	29.5	2.8	3.1
Relative difference	14.1%	1.8%	14.7%	14.8%	26.9%	2.7%	2.7%

Table 2. Measured values and predicted values for TAN (mg_{KOH} g^{-1}) of test samples and their relative difference.

	Pine pilot	*Spirulina* pilot	Miscanthus lab	Miscanthus pilot	DDGS lab 350C	DDGS lab R	DDGS lab 330C
Measured	20.1	24.8	27.0	27.7	48.7	81.4	83.3
Predicted	17.0	69.5	29.3	28.0	44.4	66.7	61.5
Difference	3.1	44.7	2.3	0.3	4.3	14.7	21.8
Relative difference	16.7%	94.8%	8.2%	1.1%	9.2%	19.9%	30.1%

Table 3. Measured values and predicted values for water content (%) of test samples and their relative difference.

	Pine pilot	*Spirulina* pilot	Miscanthus lab	DDGS lab 350C	DDGS lab R	DDGS lab 330C
Measured	11.5	7.3	17.1	4.4	22.5	8.5
Predicted	12.9	6.2	16.6	3.9	21.9	7.2
Difference	1.4	1.1	0.5	0.5	0.6	1.3
Relative difference	11.5%	16.3%	3.0%	12.0%	2.7%	16.6%



514

515 *Fig. 1. PCA score plots for preprocessed data. The color gradient corresponds to different feedstocks or*

516 *values for TAN, total phenolics, and water content (blue-low, red-high). Top-left: PC1 (37.1%), PC2*

517 *(15.9%), and different feedstocks. Top-right: PC1, PC2, and total phenolics. Bottom-left: PC1, PC2, and*

518 *TAN. Bottom-right: PC3 (9.7%), PC4 (6.7%), and water content.*

519

520

521

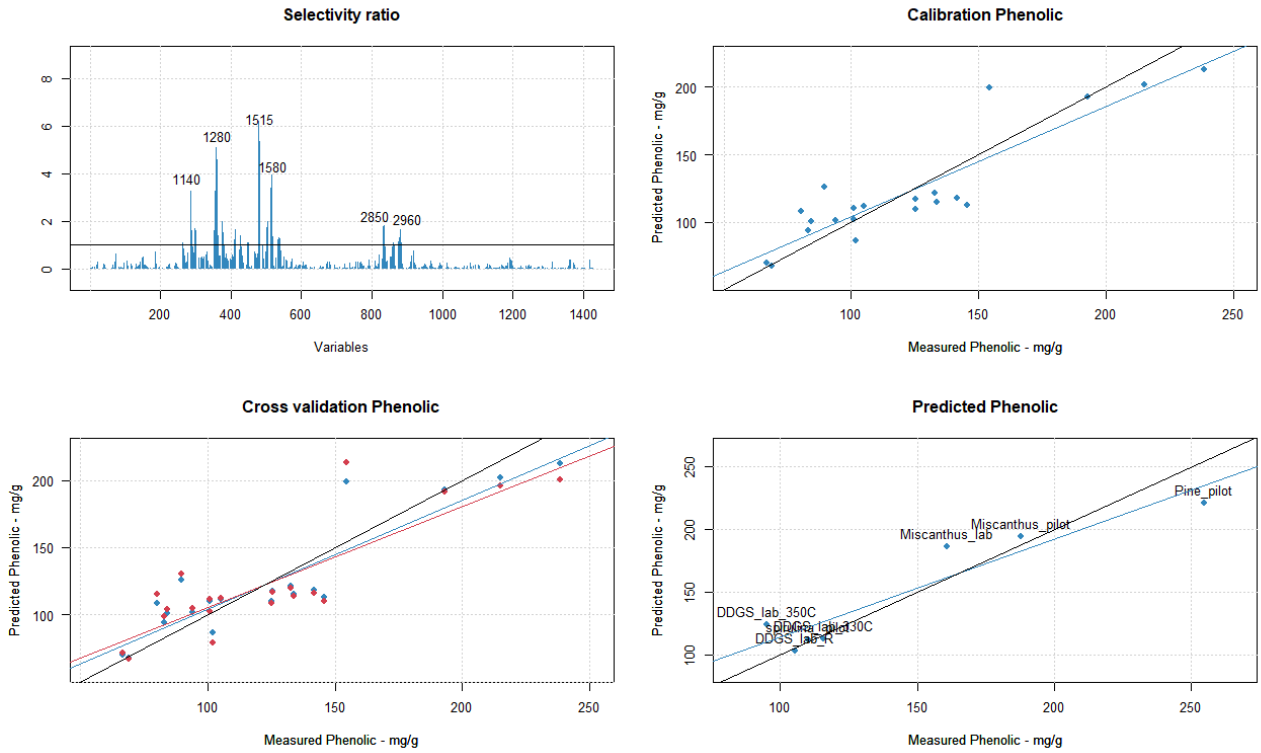
522

523

524

525

526



527

528 *Fig. 2. PLS-R plots for measured and predicted values for total phenolics and selectivity ratios. Top-left:*

529 *Selectivity ratios (black line – threshold of 0.25). Top-right: Calibration model (blue line – calibration fit,*

530 *black line – slope = 1). Bottom-left: Cross validation (red line – cross validation fit). Bottom-right: Test*

531 *data.*

532

533

534

535

536

537

538

539

540

541

542

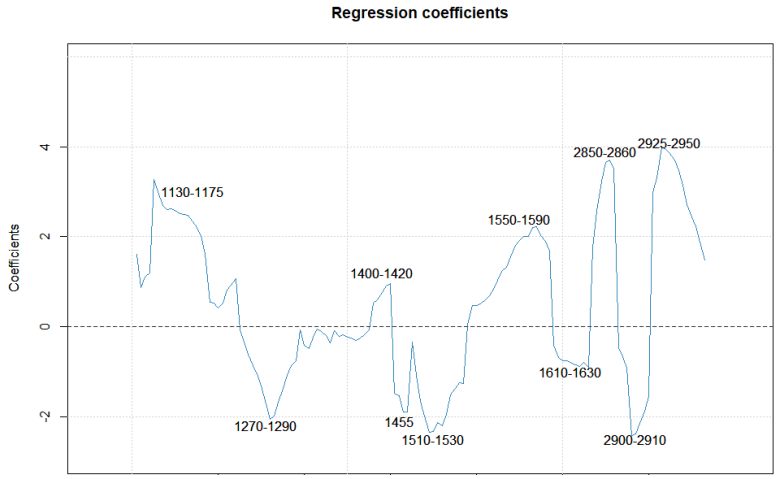
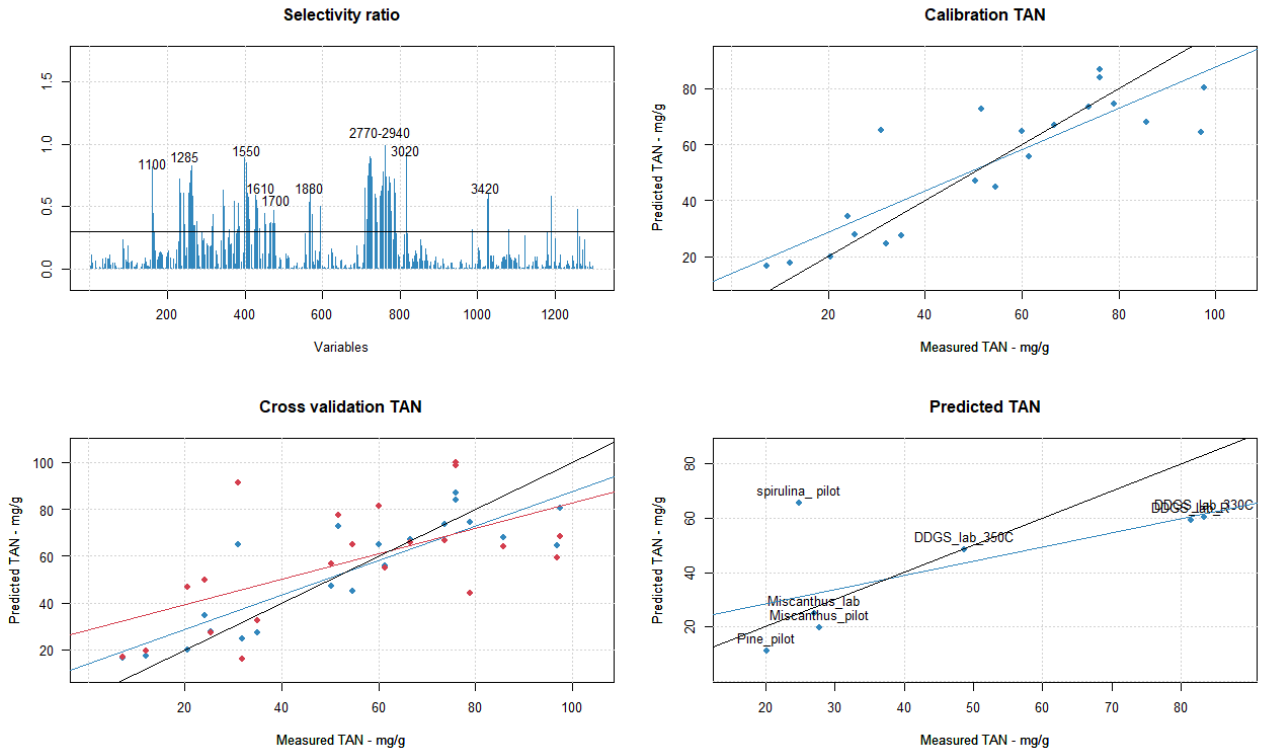


Fig. 3. Regression coefficients of PLS model for total phenolics.



560

561 *Fig. 4. PLS-R plots for measured and predicted values for total TAN and selectivity ratios. Top-left:*

562 *Selectivity ratios (black line – threshold of 0.3). Top-right: Calibration model (blue line – calibration fit,*

563 *black line – slope = 1). Bottom-left: Cross validation (red line – cross validation fit). Bottom-right: Test*

564 *data.*

565

566

567

568

569

570

571

572

573

574

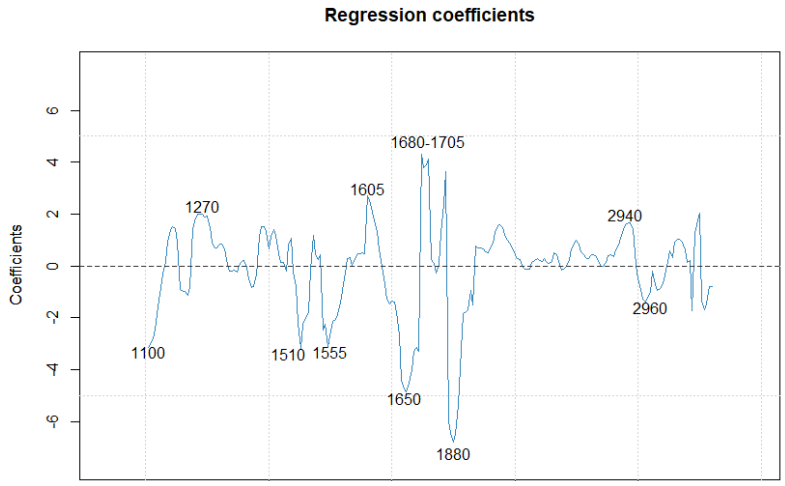


Fig. 5. Regression coefficients of PLS model for TAN.

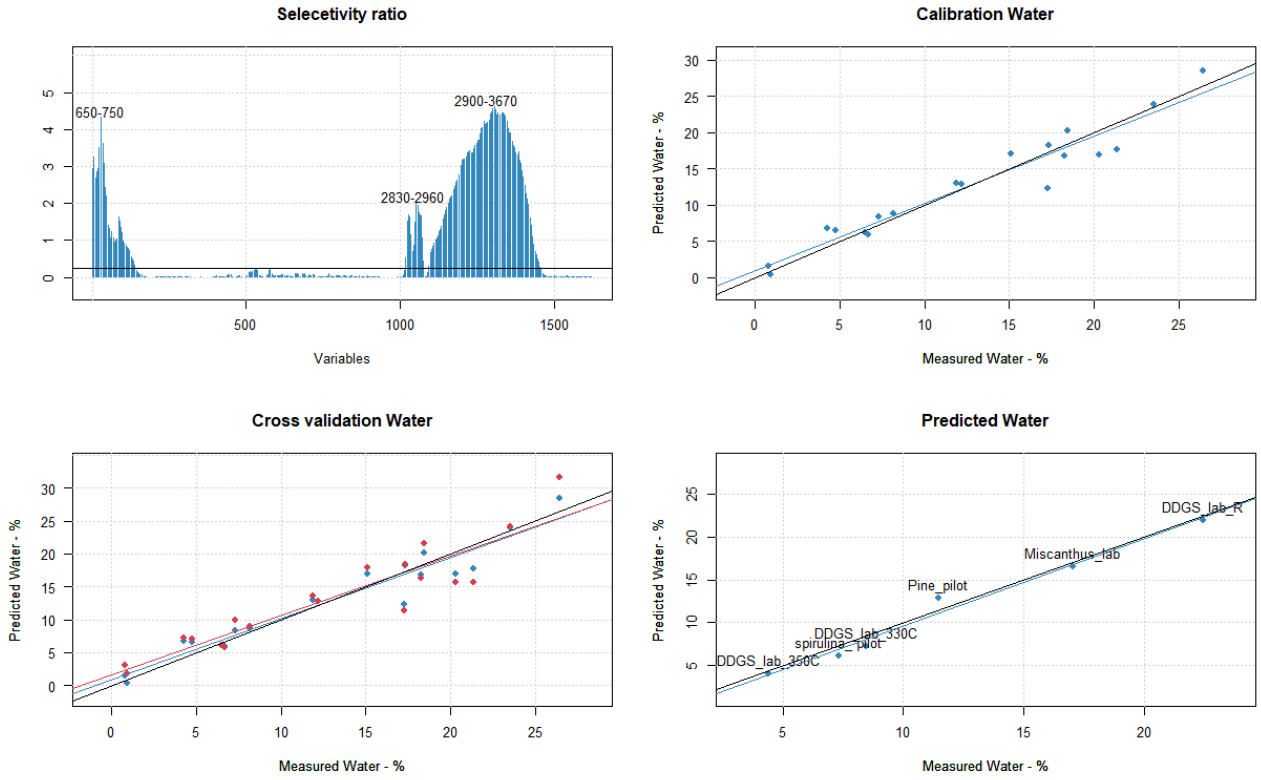
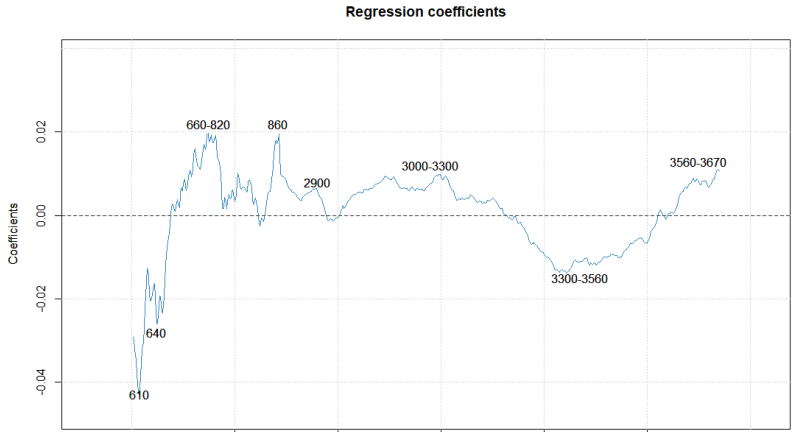


Fig. 6. PLS-R plots for measured and predicted values for water content and selectivity ratios. Top-left: Selectivity ratios (black line – threshold of 0.25). Top-right: Calibration model (blue line – calibration fit, black line – slope = 1). Bottom-left: Cross validation (red line – cross validation fit). Bottom-right: Test data.



608

609 *Fig. 7. Regression coefficients of PLS model for water content.*

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629