

# Exploring The Impact of Biomass Composition on High-Value Bio-Oil Components: Insights from Fast Pyrolysis Kinetic Simulation and Multivariate Analysis

Ingrid L. Motta<sup>a\*</sup>, Andressa N. Marchesan<sup>b</sup>, Henrique R. Guimarães<sup>c</sup>, Mateus F. Chagas<sup>c</sup>, Antonio Bonomi<sup>c</sup>, Maria Regina W. Maciel<sup>b</sup>, Rubens Maciel Filho<sup>b</sup>

<sup>a</sup>School of Mechanical Engineering, University of Campinas, Rua Mendeleev 200, Campinas, Brazil. Postcode: 13083-860

<sup>b</sup>School of Chemical Engineering, University of Campinas, Av. Albert Einstein 500, Campinas, Brazil. Postcode: 13083-852

<sup>c</sup>Brazilian Biorenewables National Laboratory, Brazilian Center for Research in Energy and Materials, R. Giuseppe Máximo Scolfaro, 10000, Campinas, Brazil. Postcode: 13083-970

[imotta@unicamp.br](mailto:imotta@unicamp.br)

Biomass consists of biodegradable and non-fossilized organic matter, including a wide range of materials such as forestry, agricultural, municipal, and industrial solid residues. Due to its renewable nature, high carbon content, and availability in several countries, biomass is a promising alternative to fossil fuels. In this context, thermochemical conversion processes such as fast pyrolysis can be an interesting option to produce heat, power, fuels, and chemicals of lower greenhouse gas emissions from biomass sources. Fast pyrolysis uses high temperatures (~500 °C), inert atmospheres (e.g., nitrogen), and short residence times (1 – 5 s) to convert biomasses mostly into bio-oil (liquid stream), also producing char (solid) and gas fractions. Biomass composition highly affects the bio-oil properties and, although much work has been done to understand such an effect focusing on heat and power production, little has been done aiming to design pyrolysis for high-added value chemicals. This work performed the kinetic simulation and multivariate analysis of a fast pyrolysis (FP) plant fed by multiple feedstocks to assess the effect of biomass composition on the bio-oil functional groups and provide guidelines to produce high-added value bio-oil components. The simulation was built in Aspen Plus™ v.10, validated against experimental data, and used to obtain a dataset correlating biomass compositions from 60 sources and FP outputs. The dataset was analyzed via hierarchical cluster analysis (HCA) and principal component analysis (PCA), showing the correlations between biomass and bio-oil properties and highlighting key biomass features to obtain specific high-value functional groups. Among the results, biomasses of higher cellulose and hemicellulose contents such as agricultural feedstocks may produce higher amounts of anhydro-sugars, ketones, and aldehydes. In contrast, feedstocks of higher lignin contents such as wood may generate increased concentrations of phenols and aromatics. This work shows how simulation and multivariate analysis tools can be used in the design of fast pyrolysis aiming at multiple bio-oil applications.

## 1. Introduction

Biomass feedstocks refer to organic materials obtained from living organisms (e.g., animals and plants). Biomass covers a wide range of sources, including wood, agricultural residues, energy crops, algae, and municipal solid wastes. The variety, abundance in many regions, and usual low costs of these sources make them interesting options for producing fuels, energy, and chemicals of lower greenhouse gas emissions (Motta et al., 2018). Among the possible biomass conversion routes, thermochemical processes (Guimarães et al., 2022) such as fast pyrolysis (FP) have gained attention over the past few decades due to their flexibility to biomass types and composition, low cost, and high energy efficiency (Figueirêdo et al., 2022).

Biomass FP is the thermal decomposition of high-carbon-containing materials into liquid, solid, and gas streams under inert conditions, temperatures in the 400–600 °C range, and residence times of 0.5–2 s. The main product is the bio-oil fraction (~50 wt% yields), a complex liquid mixture composed of oxygenated organic compounds such as acids, alcohols, aldehydes, esters, phenolics, and lignin oligomers (Demirbas, 2007), whose contents

are highly affected by the biomass features (Motta et al., 2023). Within the bio-oil composition, some functional groups such as anhydrosugars, aldehydes, aromatics, and phenols receive special attention, as these may form hydrocarbons, alcohols, platform chemicals, and building blocks of industrial interest (Figueirêdo et al., 2022). Thus, efforts towards understanding the effect of biomass composition on the bio-oil functional groups obtained via FP are required.

This work assessed, using simulation and multivariate analysis tools, the effect of biomass lignocellulosic compounds on the bio-oil functional groups obtained via FP, aiming to provide guidelines to obtain high-value bio-oil species. A rigorous kinetic simulation was built in Aspen Plus™ v.10 and used to obtain a dataset correlating biomass compositions from 60 sources and bio-oil functional groups. The dataset was analyzed via hierarchical cluster analysis (HCA) and principal component analysis (PCA), highlighting key biomass features to obtain specific high-value functional groups.

## 2. Methods

### 2.1 FP simulation and validation

The FP simulation was based on the kinetics of Ranzi et al. (2017), built from a database of >600 samples and including the steps of primary pyrolysis, char decomposition, and pyrolysis of extractives and heavy molecular-mass lignin. The reaction network consists of 28 kinetic expressions in which cellulose, hemicellulose, lignin, and extractives (here represented as triglycerides and tannins) are converted into intermediate lignocellulosic compounds, which later form species such as levoglucosan, char, C5–C6 tar, and decomposition products.

The simulation was built in Aspen Plus™ v.10 and adopted the following assumptions: steady-state and isothermal conditions; biomass was fed to the pyrolyzer with 5 wt% moisture; ash was inert; biomass Cl, S, and N were converted into Cl<sub>2</sub>, H<sub>2</sub>S, and HCN; heat loss in the pyrolyzer was neglected; effects of biomass particle size and density were not considered; Peng-Robinson with Boston-Mathias modifications (PR-BM) and NRTL were the thermodynamic models; steam tables (STEAM-TA) were used in water and steam streams; and physicochemical properties of the biomass constituents (molar weight, solid heat capacity coefficients, standard heat of formation, and solid molar volumes) were inserted in the simulation using data reported by Motta et al. (2023). The main simulated phenomena were divided into three different hierarchies as shown in Figure 1: fast pyrolysis (FP), product recovery (RECOVERY), and char combustion (COMBUST) (Motta et al., 2023).

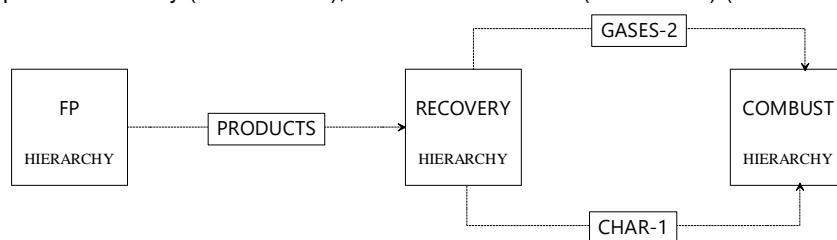


Figure 1: FP simulation hierarchies. Reproduced with permission from Elsevier.

The FP hierarchy models the biomass fast pyrolysis. Biomass is fed at 25 °C and 1 atm and inserted in the simulation as a nonconventional stream. Afterward, the nonconventional stream (characterized by its proximate and ultimate analysis) is converted into a conventional stream (featured by its lignocellulosic composition), which undergoes FP at 480 °C, 1.5 bar, 1 s residence time under the presence of compressed nitrogen following the kinetics of Ranzi et al. (2017). The FP products then go to the RECOVERY hierarchy, in which product separation occurs. The solids (char) are first recovered in a cyclone with 99% separation efficiency, followed by the separation of the condensables (bio-oil) and non-condensables (gases) in a two-stage condensation process. The char and gases are finally burned in the COMBUST hierarchy to provide heat for the FP process. Further details on how the simulation was built can be found elsewhere (Motta et al., 2023).

The simulation was validated against the experimental results of eucalyptus FP obtained in a bubbling fluidized bed pyrolyzer at 480 °C and 1 s residence time (Pienihäkkinen et al., 2022). In the validation, pyrolysis yields and bio-oil density, water content, higher heating value, elemental composition, and functional group concentrations were compared, obtaining satisfactory results (Motta et al., 2023). The validated simulation was finally run with the biomass compositions of the study, as explained in section 2.2.

## 2.2 Biomass composition

Nine agricultural and forestry biomasses – corn straw (CS), energy-cane (EC), eucalyptus (EU), pine sawdust (PS), rice straw (RS), soybean straw (SS), sugarcane bagasse (SCB), sugarcane straw (SCS), and wheat straw (WS) – were selected as feedstocks for the study. To account for composition variability in each species, several studies were collected (a total of 60 entries), and mean proximate, ultimate, and lignocellulosic analyses were calculated (Table 1) and used as inputs of the FP simulation.

Table 1: Mean biomass proximate, ultimate, and lignocellulosic analyses. db: dry basis. Moisture: 5 wt%.

Parameter	CS	EC	EU	PS	RS	SS	SCB	SCS	WS
Volatiles (wt%, db)	73.13	82.55	83.08	85.07	71.69	75.32	82.77	78.69	74.38
Fixed carbon (wt%, db)	19.65	12.55	15.30	14.39	13.80	19.06	11.41	13.42	16.52
Ash (wt%, db)	7.23	4.90	1.62	0.54	14.52	5.62	5.82	7.88	9.10
C (wt%, db)	45.51	49.19	48.61	51.20	42.34	47.90	45.45	45.64	45.31
H (wt%, db)	5.69	5.61	5.72	6.37	5.24	6.01	5.77	5.97	5.69
N (wt%, db)	0.87	0.33	0.16	0.82	0.90	1.51	0.34	0.56	0.78
Cl (wt%, db)	0.04	0.14	0.01	0	0	0	0	0	0.07
S (wt%, db)	0.21	0.17	0.06	0.12	0.24	0.38	0.13	0.09	0.17
O (wt%, db)	40.46	39.65	43.83	40.95	36.75	38.58	42.48	39.85	38.88
Cellulose (wt%)	41.90	36.65	47.51	41.60	44.80	41.81	45.27	45.50	38.91
Hemicellulose (wt%)	27.58	16.00	17.32	20.14	22.91	23.03	25.42	28.02	30.43
Lignin (wt%)	16.86	24.80	27.85	28.10	22.12	23.85	24.06	22.48	20.83
Extractives (wt%)	13.66	22.55	7.32	10.16	10.17	11.31	5.25	4.00	9.83

## 2.3 Dataset and multivariate analysis

The main FP inputs and outputs of the study were organized in a 9 × 11 matrix (9 biomasses × 4 inputs + 7 outputs, see Table 2), standardized, and analyzed via HCA and PCA in the software Statistica™ 7.0.

Table 2: 9 × 11 matrix used in HCA and PCA analyses to correlate biomass and bio-oil composition. \*CEL: cellulose; HCEL: hemicellulose; LIG: lignin; EXT: extractives.

Feedstock	Parameter										
	Inputs (biomass composition)				Outputs (bio-oil composition)						
	CEL*	HCEL*	LIG*	EXT*	Alcohols	Aldehydes& Ketones	Acids	Phenolics	Sugars	Ethers	Water
CS	41.90	27.58	16.86	13.66	3.17	22.56	19.98	3.39	30.56	0.95	41.90
EC	36.65	16.00	24.80	22.55	3.47	17.32	33.80	4.54	20.96	1.40	36.65
EU	47.51	17.32	27.85	7.32	4.60	23.37	20.29	2.11	26.72	1.71	47.51
PS	41.60	20.14	28.10	10.16	4.37	23.96	9.71	8.12	29.16	1.58	41.60
RS	44.80	22.91	22.12	10.17	3.85	23.39	14.03	4.76	30.05	1.31	44.80
SS	41.81	23.03	23.85	11.31	3.77	23.93	8.96	8.32	30.76	1.22	41.81
SCB	45.27	25.42	24.06	5.25	4.10	24.51	16.09	1.68	31.43	1.41	45.27
SCS	45.50	28.02	22.48	4.00	3.69	27.20	7.92	4.07	34.08	1.07	45.50
WS	38.91	30.43	20.83	9.83	3.49	24.07	11.13	5.43	33.59	1.13	38.91

HCA is a technique that arranges the data into tree diagrams, grouping the data into clusters in descending order of similarity based on their linkage distances (Xu and Wunsch, 2009). In this work, the variables (columns) were clustered using a single linkage amalgamation rule, and tree diagrams were generated with the Euclidean distances among the biomass samples.

The PCA adds information to the HCA. PCA reduces the dimensionality of a dataset, converting it into new sets of variables (*principal components*, PCs) that describe most of the original variation. The contributions of the variables to the PCs are measured by their *p* loadings (0–1 in module, with 1 being the highest). The *p* loadings and loading scatterplots in each PC reveal information on the correlations and anticorrelations between the parameters, which refer to the variables having *p* loadings of the same or opposing signals, respectively (Jolliffe, 1986). The 11 parameters were inserted as continuous variables (>80% of valid cases per variable and variables per case) and converted into a few uncorrelated PCs while still retaining most of the original information (>90%). The cross-validation method was used to determine the number of PCs that described the matrix, and the cumulative variance (cumulative R<sup>2</sup>X) was calculated to estimate to what extent the analysis described the data.

### 3. Results and Discussion

Figure 2 branches the biomass lignocellulosic components and bio-oil functional groups by similarity in an HCA tree diagram. In HCA diagrams, the lower the linkage distance, the higher the similarity between the parameters (Xu and Wunsch, 2009). Therefore, three main clusters have been observed, as follows.

Cluster C1 reveals the relationship between biomass extractives and bio-oil acids. In the simulation, extractives were represented as triglycerides and tannins (also known as tannic acid). Among the acids present in bio-oil, (di)carboxylic acids are of interest for applications in the polymer and food industry (Figueirêdo et al., 2022).

Cluster C2 shows the effect of biomass lignin on bio-oil phenolics, alcohols, and ethers. Lignin has –OH groups which, during degradation, form alcohols, phenols (Trubetskaya et al., 2016; Kumar et al., 2019), and aromatic and phenolic fragments (referred to as pyrolytic lignin) (Figueirêdo et al., 2022). Several depolymerization approaches involving reductive and/or oxidative conversion steps may result in considerable amounts of phenolic and aromatic building blocks, which can act as high-value platform chemicals (Figueirêdo et al., 2022). Therefore, such building blocks can be obtained from biomasses of high lignin contents.

Cluster C3 indicates the relationship between biomass cellulose and hemicellulose with aldehydes and ketones, sugars, and water in the bio-oil product. The degradation of cellulose and hemicellulose chains is known to produce oxygenates such as aldehydes and ketones (Li et al., 2022), as well as anhydrosugars (Figueirêdo et al., 2022), in a way that feedstocks of high cellulose and hemicellulose content such as agricultural and herbaceous biomasses may be suggested. Aldehydes and ketones are of interest to the solvent and paint industries. Anhydrosugars, especially levoglucosan, can be converted (via hydrolysis, fermentation, hydro-processing, aqueous phase reforming) into hydrocarbons, alcohols, sugar monomers such as glucose, gluconic acid, ketones, and platform chemicals such as hydroxymethylfurfural (Figueirêdo et al., 2022). Water in the cluster may be due to the presence of aldehydes and ketones in the aqueous phase up to a certain chain length.

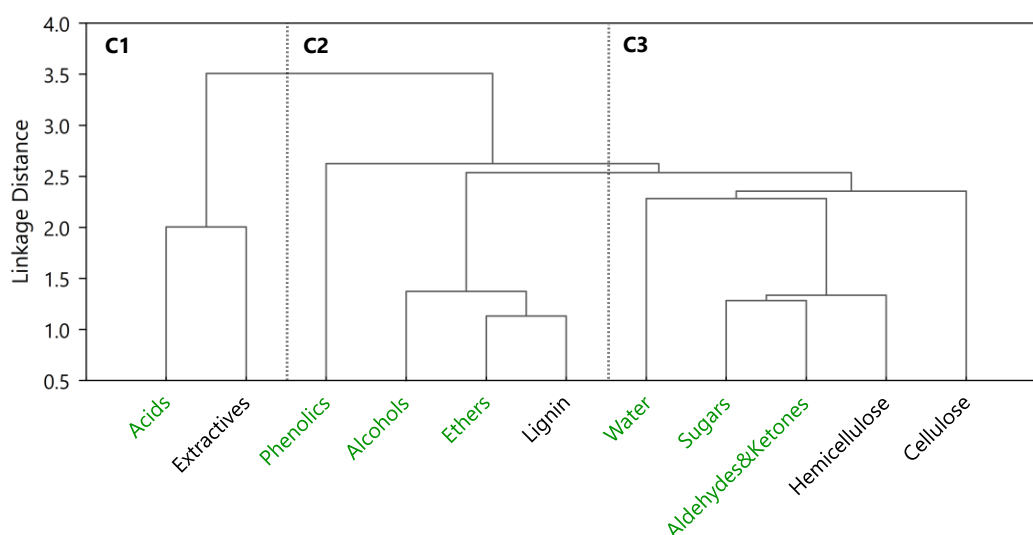


Figure 2: HCA tree diagram of biomass lignocellulosic components (black) and bio-oil functional groups (green).

Table 3: *p* loadings of the PCA analysis. The most influential parameters ( $|p| > 0.7$ ) are highlighted in bold.

Parameter		<i>p</i> loading		
		PC1	PC2	PC3
Inputs	Cellulose	0.60	0.47	-0.55
	Hemicellulose	0.68	<b>-0.70</b>	-0.08
	Lignin	-0.14	<b>0.94</b>	0.18
	Extractives	<b>-0.91</b>	-0.27	0.31
Outputs	Alcohols	0.21	<b>0.96</b>	-0.09
	Aldehydes&Ketones	<b>0.98</b>	0.07	-0.08
	Acids	<b>-0.94</b>	-0.04	-0.35
	Phenolics	0.09	0.03	<b>0.99</b>
	Sugars	<b>0.94</b>	-0.32	0.00
	Ethers	-0.26	<b>0.94</b>	-0.05
	Water	<b>0.72</b>	0.43	0.50

Despite the correlations provided by the HCA, it did not show the most influential variables or important anticorrelations between the variables, which are provided by the PCA. The PCA reduced the dimensionality of the dataset from 11 to 3 with a cumulative  $R^2X$  of 96.27%. This means that the dataset of 11 FP parameters can be represented by three PCs while retaining an explained variance  $>90\%$ . PC1 (see Table 3) retains most of the explained variance (45.6%), and its most contributing parameters ( $|p|>0.7$ ) are the biomass extractives and the bio-oil aldehydes/ketones, acids, sugars, and water. As biomass extractives and bio-oil acids both have negative  $p$  loadings (same signals) within PC1, these parameters are correlated, agreeing with cluster C1 of Figure 2. Similarly, aldehydes/ketones, sugars, and water are also correlated (all have positive  $p$  loadings), which agree with the branches of cluster C3. PC2 has the second highest explained variance (34.3%), having biomass hemicellulose and lignin, and bio-oil alcohols and ethers as its most influential parameters. Biomass lignin and bio-oil ethers and alcohols are correlated (all have positive  $p$  loadings, in line with cluster C2), and anticorrelated to hemicellulose. The anticorrelation between hemicellulose and lignin, represented by  $p$  loadings with opposing signals, agrees with the reports of reduced anhydrosugars production from cellulose and hemicellulose due to the presence of lignin (Kumar et al., 2019), an anticorrelation not shown in the HCA. As for PC3, it has the lowest explained variance (16.4%), with only one influential parameter (phenolics). The  $p$  loadings of PC1, PC2, and PC3 can be plotted in scatterplots (Figure 3), in which the three groups of the HCA (Figure 2) can be seen as three distinct groups, with the most influential parameters being those farthest from the origin.

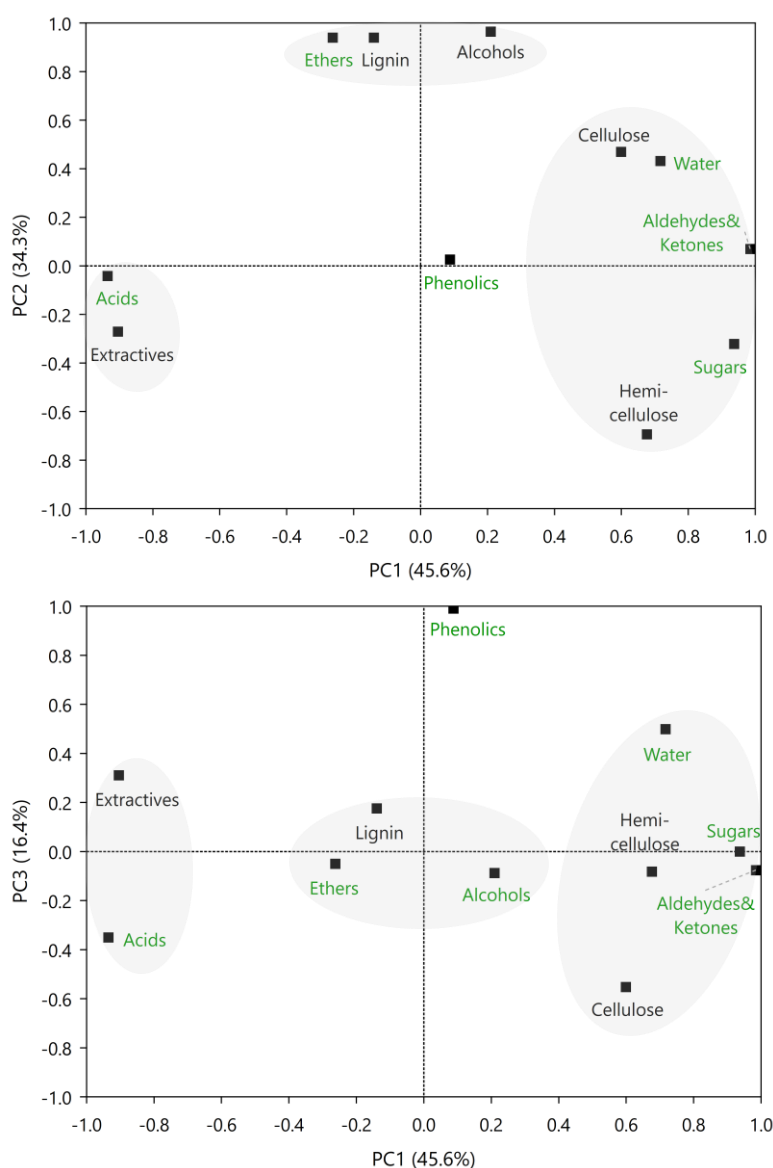


Figure 3: Component  $p$  loading scatterplots in PCA: (a) PC1 and PC2; (b) PC1 and PC3. Parameters in black and green refer to biomass components and bio-oil functional groups, respectively.

#### 4. Conclusions

This paper studied, using a rigorous kinetic-approach simulation and multivariate analysis, the relationship between biomass and FP bio-oil composition to recommend feedstock selection aiming to produce high-value functional groups. The observed relationships were compared to the literature, which enabled analyzing to what extent a statistical analysis could represent the phenomena involved. HCA evidenced several correlations, such as higher lignin contents being correlated with the presence of alcohols, phenols, and aromatic and phenolic fragments (pyrolytic lignin). Biomass cellulose and hemicellulose were correlated with bio-oil oxygenated compounds (aldehydes and ketones) and anhydrosugars such as levoglucosan. PCA complemented the HCA results by confirming the observed clusters, as well as evidencing anticorrelations and the most influential parameters. Hemicellulose and lignin are among the most influential biomass lignocellulosic compounds, and both variables are anticorrelated. All bio-oil functional groups have a great impact on a specific PC, and their specific correlations agree with the clusters obtained via HCA. The possible end uses for the high-value functional groups have also been indicated, mostly referring to applications as platform chemicals. This work hopes to serve as a guideline for future works on the development of pyrolysis biorefineries focused not only on fuels but also on chemicals of industrial interest.

#### Acknowledgments

The authors would like to thank the funding received from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, São Paulo Research Foundation (FAPESP) grants #2015/20630-4 and #2016/50403-2, and four Brazilian companies (EMBRAER S.A., KLABIN S.A., PETROBRAS S.A., and SUZANO S.A). This research was started by ILM during her period at the "Pesquisador Colaborador" program of the School of Chemical Engineering (University of Campinas) and finished at the School of Mechanical Engineering (University of Campinas).

#### References

- Demirbas A., 2007, The influence of temperature on the yields of compounds existing in bio-oils obtained from biomass samples via pyrolysis, *Fuel Processing Technology*, 88, 591-597.
- Figueirêdo M.B., Hita I., Deuss P.J., Venderbosch R.H., Heeres H.J., 2022, Pyrolytic lignin: a promising biorefinery feedstock for the production of fuels and valuable chemicals. *Green Chemistry*; 24, 4680–4702.
- Guimarães H.R., Bressanin J.M., Motta, I.L., Chagas M.F., Bonomi A. Maciel Filho R., Watanabe M.D.B., 2022, Techno-Economic and Environmental Assessments of Thermochemical Routes Integrated into the Brazilian Sugarcane Industry for the Production of Renewable Jet Fuel, *Chemical Engineering Transactions*, 92, 115–120.
- Jolliffe I.T., 1986, Introduction, Chapter In: I.T Jolliffe (Ed.), *Princ. Compon. Anal.*, Springer Science + Business Media, Kent, UK, 1–7.
- Kumar M., Sabbarwal S., Mishra P.K., Upadhyay S.N., 2019, Thermal degradation kinetics of sugarcane leaves (*Saccharum officinarum* L) using thermo-gravimetric and differential scanning calorimetric studies. *Bioresource Technology*; 279, 262–270.
- Li J., Shang Y., Wei W., Liu Z., Qiao Y., Qin S., Tian Y., 2022, Comparative study on pyrolysis kinetics behavior and high-temperature fast pyrolysis product analysis of coastal zone and land biomasses, *ACS Omega*, 7, 10144–10155.
- Motta I.L., Marchesan A.N., Real Guimarães H., Chagas M.F., Bonomi A., Wolf Maciel M.R., Maciel Filho R., 2023, Fast pyrolysis simulation via kinetic approach and multivariate analysis to assess the effect of biomass properties on product yields, properties, and pyrolyzer performance, *Energy Conversion and Management*, 296, 117676.
- Motta I.L., Miranda N.T., Maciel Filho R., Wolf Maciel M.R., 2018, Sugarcane Bagasse Gasification: Thermodynamic Modelling and Analysis of Operating Effects in a Steam-Oxygen-Blown Fluidized Bed using Aspen PlusTM, *Chemical Engineering Transactions*, 65, 169-174.
- Pienihäkkinen E., Leijenhörst E.J., Wolters W., Lindfors C., Lahtinen J., Ohra-aho T., Oasmaa A., 2022, Valorization of eucalyptus, giant reed arundo, fiber sorghum, and sugarcane, *Energy & Fuels*, 36, 12021-12030.
- Ranzi E., Debiagi P.E.A., Frassoldati A., 2017, Mathematical modeling of fast biomass pyrolysis and bio-oil formation. Note I: Kinetic mechanism of biomass pyrolysis, *ACS Sustain. Chem. Eng.*, 5, 2867-2881.
- Trubetskaya A., Jensen P.A., Jensen A.D., Garcia Llamas A.D., Umeki K., Glarborg P., 2016, Effect of fast pyrolysis conditions on biomass solid residues at high temperatures. *Fuel Processing Technology*, 143, 118-29.
- Xu R., Wunsch II D.C., 2009, Hierarchical Clustering. Chapter In: D.B. Fogel (Ed.), *Clustering*, IEEE Press, Hoboken, USA, 31–69.