Construction and exploration of a dilute acid pretreatment dataset on poplar wood to propose trade-offs of chemicals evolution

Julien du Pasquier, Patrick Perré, Gabriel Paës

Construction and exploration of a dilute acid pretreatment dataset on poplar wood to propose trade-offs of chemicals evolution

Julien du Pasquier[1, 2], Patrick Perré[1], Gabriel Paës[2]*

[1] Université Paris-Saclay, CentraleSupélec, Laboratoire de Génie des Procédés et Matériaux, Centre Européen de Biotechnologie et de Bioéconomie (CEBB), 51110 Pomacle, France

[2] Université de Reims Champagne-Ardenne, INRAE, FARE, UMR A 614, Reims, France

* patrick.perre@centralesupelec.fr

Abstract

This work provides one of the most comprehensive datasets to date on the chemical evolution of poplar wood polysaccharides and their degradation products during dilute acid pretreatment. Based on the literature, an unprecedented set of 38 conditions was generated using a design of experiment approach within the following parameter ranges: 2-60 min, 120-190°C, 0-4%wt H_2SO_4. Solid wood residues and pretreatment liquid, recovered separately, were analyzed for no less than 12 compounds: sugars, inhibitors, and lignin. Particular attention was paid to the sampling and method description so that the dataset, fully available, can be easily shared and reused. This unique dataset was explored using models to follow the evolution of each chemical species. A more original and in-depth analysis was carried out by combining the models using desirability functions, which enabled us to propose trade-offs between the evolution of the different compounds according to the targeted production objectives.

Keywords

Lignocellulosic biomass, Design of experiment, Dataset, Modeling, Optimization, Desirability function

1. Introduction

In the context of the ongoing transition to a sustainable society, it is necessary to produce energy from renewable sources which represented only 16% of global energy consumption in 2021 (Ritchie et al., 2022). Biofuels contribute marginally (0.6%) to this global energy mix, with 124 billion liters produced in 2021, mainly by the USA (68 billion) and Brazil (33 billion) (RFA, 2023; Ritchie et al., 2022). But most of these biofuels are produced from starch-based and sugar-based feedstocks, which compete with food industry (Chen et al., 2021). Besides, lignocellulosic biomass (LB) coming from global forestry wastes and agricultural residues represent each year an estimated equivalent quantity of 442 billion liters of bioethanol (Chen et al., 2021). This source of energy is therefore underutilized and represents an immense potential: it might rise from 124 to 442 billion liters over the coming decades, i.e. a 350% increase of biofuels production using only LB residues (Chen et al., 2021; RFA, 2023).

The potential of LB for producing biochemicals comes mostly from its main constituent, the cellulose, a polysaccharide composed of glucose units (linear β-1,4-glucan chains) (Chundawat et al., 2011) which can be used in fermentation processes to produce biofuels or other molecules of interest (Yoo et al., 2020). Other polysaccharides present in the LB in the form of a mixture of branched pentose and hexose units (mainly xylose and mannose), called hemicelluloses, are also of interest in fermentation (Acosta et al., 2021). The degradation of these polysaccharides into fermentable monomers requires an enzymatic hydrolysis step. Lignin, the third principal component of LB, is a complex polyphenol heteropolymer with a wide range of potential applications in the field of

chemicals, materials, fuels, and healthcare products (Becker & Wittmann, 2019). The main difficulty in processing these different polymers lies in the natural recalcitrance of LB, which refers to the difficulty of separating and deconstructing these components in building blocks (Chundawat et al., 2011; Pu et al., 2013). This mainly originates from the complex structure of LB: cellulose is polymerized into a crystalline structure to form microfibrils, packed together in fibrils themselves embedded in a matrix composed of hemicelluloses and lignin, and this structure inhibits the action of enzymes (Chundawat et al., 2011; Pu et al., 2013; Zoghlami & Paës, 2019).

To overcome recalcitrance, a common industrial strategy is to apply a pretreatment before enzymatic hydrolysis (Tan et al., 2021). This pretreatment eases the penetration of enzymes inside the LB material and thus enhances the enzymatic yield (Beig et al., 2021). Several pretreatment types exist, organized into four categories: physical, physico-chemical, chemical, and biological (Beig et al., 2021). This article focuses on dilute acid pretreatment (DAP), a thermo-chemical pretreatment that consists in submitting LB at a high temperature (120-210°C) in a liquid phase containing a low percentage of acid (<10% wt) (Beig et al., 2021; du Pasquier et al., 2023). The economic interest of DAP is to use lower temperature levels in comparison to thermal pretreatments, thanks to the acidic conditions. Although this process is already used industrially in Brazil and USA (Carpio et al., 2021), there is a need to further improve the enzymatic yield after DAP while limiting the production of fermentation inhibitors, as attested by studies published in recent years (Baral & Shah, 2017; Carpio et al., 2021; Chen et al., 2021; Vieira et al., 2020). To enhance the knowledge on DAP impact on LB, the use of Design of Experiment (DoE) and Response Surface Methodology (RSM) is recommended, because they allow an increased precision of the results, while making a large exploration of the DAP conditions and limiting the number of tests to be carried out (Pereira et al., 2021).

However, less than 10% of the published articles dealing with DAP optimization use these methods, in favor of a less complex approach, or simply by working on reduced fields

of study, changing only one or two factors at a time. This limitation was highlighted in a recent review (du Pasquier et al., 2023), which prevents the emergence of an overall interpretation of the effect of the pretreatment. Another difficulty to compare the results of different works lies in the variety of devices, often insufficiently described (du Pasquier et al., 2023; Yang et al., 2016; Zhi et al., 2013). All these issues are real obstacles to significant progress in quantifying the mechanisms involved in LB degradation by DAP.

To address these limitations, the present work proposes, to our knowledge, one of the most comprehensive, rigorous and innovative dataset on DAP. To this purpose, an unprecedented dataset of 38 DAP conditions defined by an experimental design with a wide range of parameters has been generated (2-60 min, 120-190°C, 0-4%wt H_2SO_4). These conditions were applied to carefully selected poplar samples using an in-house, fully instrumented DAP device, designed to accurately apply the operating conditions. Solid wood residues and treatment liquid were analyzed for a total of 12 compounds: sugars, inhibitors, and lignin. All these data have been compiled into a fully available dataset. The quality of this dataset is supported by advanced descriptions of the equipment and methods. As main outcome of the work, the data were analyzed to create chemical models for all compounds over the entire set of experiments and, again, available as all expressions are detailed in tables. Finally, desirability functions were derived from the models to describe the trade-offs between the evolution of different chemical species. This work therefore provides a decision-support tool targeted at the desired objectives.

2. Materials and methods

2.1. Sample preparation

2.1.1. Collection and cutting of samples

The raw poplar wood was supplied by INRAE Val de Loire (Orléans, France) from a field tree with a diameter of about 15 cm at the base. The tree was cut into logs, while

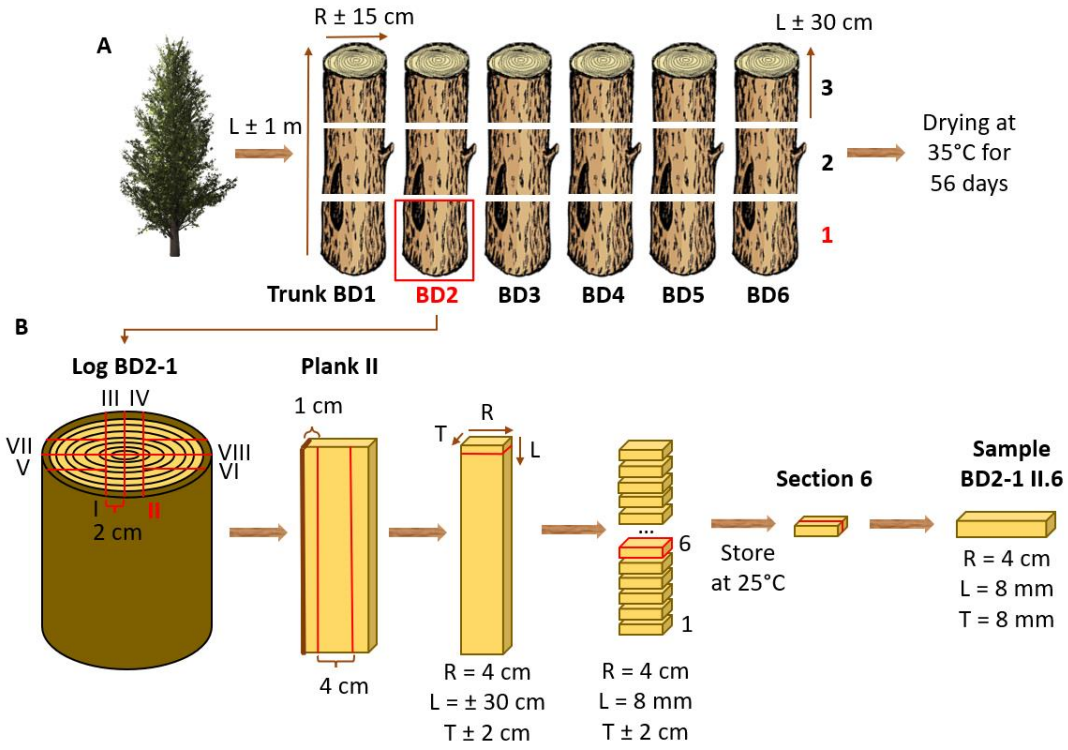

Figure 1: Cutting and sampling poplar wood samples. A: Felling of the tree in the field and cutting into six trunks of 1m, then into 3 logs of approximately 30 cm. B: Precise cutting of logs into samples. The sampling takes as an example log BD2-1, plank II and section 6.

keeping track of the position of each part in the tree (Figure 1A). All logs were dried in an oven at 35°C for 56 days down to a moisture content below 10% and stored in appropriate conditions.

In the sample preparation, consecutive logs were selected to guarantee normal wood based on two criteria: no bending to avoid tension wood and a limited number of small knots. Eight longitudinal planks (I-VIII) were prepared with a miter saw from each log. Planks were debarked, then the pith was removed for planks I to IV (Figure 1B). They were cut into thick sections, numbered from 1 (bottom of the tree) and stored at 25°C. Before their use for pretreatment, the sections were cut again with a micro-cutting machine (Secotom-50, Stuers, Denmark) to obtain the final dimensions of $0.8 \times 0.8 \times 4$ cm^3 suited to the reactor size (Figure 1B). This cutting produce thin samples in the longitudinal direction,

which allows the treatment to be homogeneous and facilitates the extraction of liquid. The samples can be considered as twins as they are intentionally aligned along the longitudinal dimension which has little impact on wood properties over such a small length.

2.1.2. Wood treatment

For each DAP, sets of 8 samples are used to obtain enough material for the analyses. As this work is focused on the effect of treatment conditions, all sets of 8 samples must be similar to each other, despite the biological variability of wood. For that purpose, each set contains one sample of the same eight planks (Set 1: I-1 to VIII-1, Set 2: I-2 to VIII-2, etc) (Figure 1). The reactor size can hold 4 samples, so the 8 samples were separated into two groups and pretreated in two successive experiments (Figure 2).

Three off-cuts from the cutting of the 8 samples were weighed and dried for 24 hours in an oven at 105°C to determine the moisture content (NewClassicMF, Mettler Toledo, Switzerland). The 8 samples were weighed and then immersed in a sulfuric acid solution (AnalaR Normapur, VWR, USA) at the desired concentration. Vacuum-pressure cycles of 30 min were carried out to saturate the samples until 1.5-1.8 g of liquid per g of wood.

A patented prototype WAVE^{T} 2 reactor was used for DAP pretreatment (Lancha et al., 2021; Mokdad et al., 2018). It allowed pretreatment to be performed up to 190 °C under 13 bar of saturated vapor pressure and 15 bar of total pressure with a rigorous control of temperature level and residence time. The temperature is measured by a K-thermocouple near the sample and controlled by a PID system (Eurotherm 3216, UK). The signal from the thermocouple is transmitted to an acquisition unit (Agilent HP 34970A, USA).

Four samples were placed in the reactor with 100 mL of sulfuric acid at the concentration studied (Figure 2). Once the reactor was closed, the nominal treatment

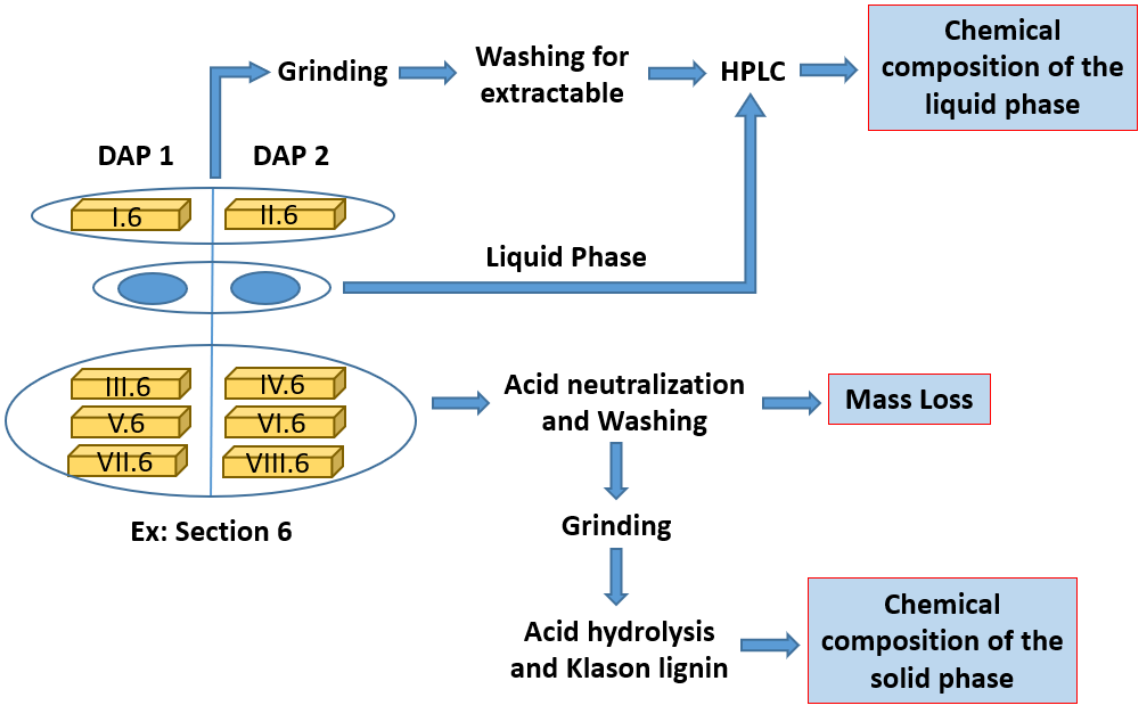

Figure 2 : Sample pretreatment and analysis workflow. The samples taken as examples are from cutting section 6 of log BD1-3.

temperature was set thanks to two electric heating collars (total of 370 W) placed at the outer surface of the reactor. The reaction is considered to start when the reactor temperature reaches the set point minus 2°C, which takes about 15 minutes depending on the nominal temperature. At the end of the plateau, two external fans cool down the reaction chamber temperature down to a temperature of 40° C, which takes about 15 minutes. The liquid fraction at the bottom of the reactor is frozen at -20°C before being analyzed. All the samples are immediately weighed and then stored at 5°C.

2.1.3. Recovery of extractives

One part of the solubilized compounds remained inside the sample at the end of the treatment. This fraction needed to be added to the liquid phase for a correct balance of extractives. Two samples, one per batch treatment, were finely grounded by a ball mill with liquid nitrogen (CryoMill, Retsch, Germany), 2 cycles of 2 min at frequency of 30 s^{-1},

intermediate cooling for 30 s at 5 s^{-1}. The resulting powder underwent 3 successive washing cycles: after suspension in approximately 40 mL of double-distilled water, the sample was placed in an ultrasonic bath for 5 min followed by 5 min of centrifugation at 500 g before recovering the supernatant. The 3 supernatants were pooled and mixed, then frozen at -20°C before being analyzed (Figure 2). The concentrations obtained were then reduced to the wet mass of the ground samples, then extrapolated to all the pretreated samples, assuming they have the same moisture content.

2.1.4. Acid neutralization and washing

The six remaining samples were gathered and treated with an excess of sodium bicarbonate (anhydrous, 99%) (Sigma Aldrich, Germany) to neutralize the acid inside the material and stop any further chemical evolution (Figure 2). Approximately 2 g of sodium bicarbonate was dissolved in 2.5 L of double distilled water, and the samples were immersed in the solution. Vacuum-pressure cycles were carried out every hour for 6 hours, then the samples were left to soak overnight at atmospheric pressure. The residual salts were then washed by three successive baths of 1 hour in 500 mL double distilled water (Figure 2). Neutralization and washes were done in an orbital shaker at 30 rpm. The samples were then dried for 5 days at 35°C, weighed for the mass loss, then 5 of them taken randomly were finely ground by the ball mill with liquid nitrogen. The sixth one was kept intact for structural analysis (data not shown). The powder obtained was analyzed for dry mass (100g at 105°C) and the leftover was kept with the entire sample at 25°C until analysis (Figure 2).

2.2. Chemical analysis

The two liquid fractions from the two pretreatments and the supernatant conserved for the extractives were analyzed for their chemical composition by Ultima 3000 HPLC (Thermo Fisher Scientific, USA) coupled with:

- a Refractive Index Detector RI-101 (Shodex) with an Aminex HPX-87H column (300 × 7.8 mm, Biorad) for xylose, glucose, acetic acid, formic acid, levulinic acid and succinic acid.

- a UV detector on an Acclaim Polar Advantage II C18 column (4.6 × 150 mm, 3 µm, 120 Å, Thermo Fisher Scientific) for HMF and furfural.

The weight of each compound in the liquid fractions and in the extractives was calculated and added to give the final composition of the liquid fraction in g per 100 g of initial dry matter (% DM_i), corrected for mass loss.

Solid samples were analyzed by acid hydrolysis: 10 mg of powder were treated with 12M sulfuric acid for 2 h at 20°C and then diluted to 1 M for 2 h at 100°C. All samples were then filtered (polytetrafluoroethylene, 0.45 µm) and injected on an HPAEC (Dionex©, US). Glucose, xylose and other monosaccharides (fucose, arabinose, rhamnose, galactose, mannose, fructose) were detected by pulsed amperometry (PAD 2, Dionex©, US) on a CarboPac PA-1 anion exchange column (4 × 250 mm, Dionex©, US)(Belmokhtar et al., 2013). Klason lignin content was determined after a two-step sulfuric acid hydrolysis but only on part of the samples at severities presenting contrasting results (Machinet et al., 2011). The composition of the solid phase in monosaccharides and lignin was then calculated in g of compound per 100g of initial (% DM_i) and final (% DM_f) dry matter corrected for mass loss.

2.3. Pretreatment conditions and data modeling

To investigate a wide range of DAP conditions, this work combined the techniques of Design of Experiment and the Response Surfaces Methodology to plan and analyze the data (Yildirim et al., 2021).

2.3.1. Design of experiment

The selected parameter ranges account for recommendations of a recent review paper (du Pasquier et al., 2023), considering three factors: time (2-60 min), temperature (120-190°C) and sulfuric acid concentration (0-4 % wt). These conditions cover pretreatment severities ranging from 1 to 4, according to the Combined Severity Factor (CSF) criterion commonly used for thermo-chemical pretreatment in industry (Auxenfans et al., 2017):

$$CSF = log_{10}\left(t \times e^{\frac{T-Tr}{14.75}}\right) - pH \quad (1)$$

where t is the reaction time (min), T the operating temperature (°C), T_r the reference temperature (100°C). *pH* is that of the liquid phase used to soak biomass.

The experimental design comes from a D-Optimal matrix made with Azurad software (Azurad, 2019; Manzon et al., 2020). Candidate points have been generated from an 8-level grid created using the limits of the domain. The selection of the experimental points was based on statistical criteria to homogeneously cover the domain of the study for the generated model to be as reliable as possible. The model contained 25 different conditions, including 3 in triplicate, and 7 conditions with different acid levels were also carried out: 5 to test the model and 2 to test higher acid conditions (Table 1), for a total of 38 pretreatment conditions. Five raw samples, taken at different heights to check the homogeneity of the wood, were also analyzed to determine its initial composition.

2.3.1. Data modeling

The processing and modeling of the data was performed with the software Azurad by RSM using second order polynomial models, except for lignin content due to the lack of some measurements. Equation 2 depicts the structure of the corresponding formula, where Y is the observed response, b_x the polynomial parameters and X_x the DAP conditions:

Table 1: Time, temperature, and acid conditions of the 38 conditions selected by DoE. Triplicates: samples 3-5, 10-12 and 26-28; Model test: samples 32-36; Higher acid conditions: samples 37 and 38. No CSF calculation for the 2 acid-free conditions.

Sample	Time (min)	Temperature (°C)	Acid (%wt)	CSF
1	20.0	120	0.1	0.20
2	10.0	140	0.5	1.19
3	20.0	160	0.1	1.38
4	20.0	160	0.1	1.38
5	20.0	160	0.1	1.38
6	45.0	120	1.5	1.73
7	5.0	160	1.0	1.78
8	20.0	120	4.0	1.81
9	60.0	140	0.4	1.87
10	30.0	140	1.0	1.97
11	30.0	140	1.0	1.97
12	30.0	140	1.0	1.97
13	2.0	180	1.5	2.14
14	2.0	170	3.0	2.15
15	45.0	120	4.0	2.16
16	10.0	175	0.5	2.22
17	20.0	150	2.0	2.39
18	30.0	140	3.0	2.45
19	15.0	170	1.0	2.55
20	60.0	140	2.0	2.57
21	60.0	170	0.4	2.75
22	25.2	180	1.5	3.25
23	25.2	170	3.0	3.25
24	45.0	160	4.0	3.34
25	60.0	170	2.0	3.45
26	35.0	180	3.0	3.69
27	35.0	180	3.0	3.69
28	35.0	180	3.0	3.69
29	15.0	190	4.0	3.75
30	25.2	190	0.0	/
31	2.0	190	0.0	/
32	10.0	140	0.86	1.42
33	45.0	120	2.72	1.99
34	10.0	175	0.86	2.45
35	60.0	140	3.56	2.82
36	25.2	180	2.72	3.51
37	20.0	120	6.84	2.04
38	15.0	190	6.84	3.98

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{11}{X_1}^2 + b_{22}{X_2}^2 + b_{33}{X_3}^2 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 \quad (2)$$

The DAP conditions are dimensionless values: each time, temperature and acid factors are reduced to a dimensionless scale between -1 and 1 corresponding, respectively, to the less severe conditions (2 min, 120°C, 0%) and to the most severe conditions (60 min, 190°C, 4%).

The desirability functions combine all the models obtained to provide zones able to fulfill a combination of desired constraints. These outcomes are trade-off zones depending on the selected values of maximum and minimum concentrations for each chemical species. An infinity of scenarios can be envisaged, depending on the targets for each chemical produced. In all interpretations, it is important to keep in mind that the RSM method gives the best possible representation using a second order polynomial. For example, such a function is not able to correctly represent a large region with constant value.

3. Results and Discussion

3.1. Data

For further use by the scientific community, all the data collected in this study, chemical analyses and mass losses, are available on an open data repository (du Pasquier, 2022).

3.2. Raw materials

Based on the analysis of the five raw samples, the average composition of dry poplar is 46.0% cellulose, 21.5% hemicelluloses (including 17.3% xylose) and 19.9% lignin, which explains almost 90% of the total mass. The samples taken at different heights of the logs have a similar composition, with standard deviations of 1.66% for cellulose, 0.41% for hemicelluloses, and 0.18% for lignin. This confirms the assumption of twin samples.

3.3. Models and validation of the data

A model, depicting the evolution of the chemical content as a function of the operating conditions (time, temperature and acid concentration), was obtained for each chemical species measured in both liquid and solid phases (see supplementary material). To obtain these models, some of the 31 initial experimental points have been partially disabled due aberrant values and were therefore discarded to calculate the model (Table 2). This concerns experiment 16 (Exp_{16}) for xylose in liquid phase (Xylose-L) with 17.42% experimental *vs.* 7.52% calculated, and experiment 22 (Exp_{22}) for HMF with 1.04% experimental *vs.* 0.51% calculated (p-value<0.0005). Exp_{11} carried out in triplicate gives also aberrant results compared to the other two identical tests for Xylose-L (Exp_{10}: 13.26%, **Exp_{11}: 15.77%**, Exp_{12}: 12.53%, Figure 3A) and Furfural (Exp_{10}: 1.06%, **Exp_{11}: 0.40%**, Exp_{12}: 0.93%) and was deleted for these compounds. Exp_1 has been disabled for Furfural because

Table 2 : Model coefficients of chemical species measured. OM: Other monosaccharides; -L: in liquid phase; -S: in solid phase; DM_f: Over final dry mass; DM_i: Over initial dry mass; Center accuracy: Accuracy of the models at the center of the study domain (31 min / 155°C / 2%); Liquid phase: Xylose-L to Succinic acid; Solid phase: Xylose-S DM_f to OM DM_i.

Chemical species	Disabled experiments	b_0	b_1	b_2	b_3	b_{11}	b_{22}	b_{33}	b_{12}	b_{13}	b_{23}	Center accuracy
Xylose-L	11, 16	11.85	-1.19	-5.75	-0.85	-0.08	-7.79	-2.62	-7.12	-6.78	-5.37	± 1.25
Glucose-L	/	4.93	2.31	4.49	2.68	0.41	1.00	-3.57	4.44	2.49	0.62	± 2.58
Acetic acid	/	5.41	0.39	0.52	1.12	-0.45	-0.51	-2.00	-0.80	-1.34	-1.08	± 0.87
HMF	22	0.29	0.16	0.35	0.17	-0.03	0.19	-0.17	0.29	0.10	0.11	± 0.08
Furfural	1, 11	3.84	0.45	2.06	1.06	-0.52	-2.14	-1.87	-1.27	-1.39	-1.69	± 0.57
Formic acid	/	0.82	0.66	1.31	0.63	-0.26	0.74	0.03	0.39	0.11	1.04	± 0.19
Levulinic acid	/	0.63	1.06	3.07	2.01	-0.38	2.81	0.12	1.21	0.73	3.13	± 1.44
Succinic acid	/	0.22	0.26	0.35	0.15	0.13	0.19	-0.17	0.39	0.16	0.11	± 0.10
Xylose-S DM_f	31	0.71	-1.04	-4.38	-2.64	0.54	3.46	2.63	1.83	0.88	2.21	± 0.92
Glucose-S DM_f	/	59.56	-4.87	-16.21	-8.63	6.10	-19.26	-3.00	-7.91	-6.12	-22.00	± 4.44
OM DM_f	31	0.71	-0.22	-1.23	-0.45	0.05	0.58	0.45	0.25	0.08	0.30	± 0.55
Xylose-S DM_i	31	-0.14	-1.02	-3.71	-2.28	0.97	3.95	2.19	2.45	1.25	2.61	± 0.75
Glucose-S DM_i	30	35.61	-6.16	-18.95	-8.24	6.62	-10.52	2.41	-5.80	-2.89	-10.28	± 2.93
OM DM_i	31	0.34	-0.24	-1.05	-0.43	0.17	0.73	0.41	0.40	0.17	0.41	± 0.40

its zero value influences the polynomial model too strongly. The two experiments 30 and 31 without acids have also been often deleted because of aberrant results for the solid phase models: Exp_{30} for glucose in solid phase (Glucose-S) DM_i with 46.79% experimental Vs. 29.15% calculated and Exp_{31} for xylose in solid phase (Xylose-S) DM_f with 10.88% experimental Vs. 3.53% calculated (p-value<0.0005).

3.4. Equipment limitations

Conditions with very short residence times (less than 5 minutes) are difficult to control. As the reactor takes about fifteen minutes to reach the set temperature, the thermal effect during the last instants when the temperature approaches is no longer negligible. This effect is stronger as the residence time is shorter. Conditions with short times are therefore less precise and difficult to reproduce with different experimental devices. The CSF is thus most of the time underestimated for these experiments because the temperature rise is not included in the calculation.

3.5. General trends

Most of the models depict either a monotonic increase with severity for the liquid phase (Glucose-L, HMF, Formic, Levulinic and Succinic acid), or a monotonic decrease for the solid phase (except Glucose-S DM_f) (Figure 3). Note however that some intermediate components are first produced and then degraded. This is the case for example for Xylose and Furfural in the liquid phase (Figure 3)

3.5.1. Liquid phase analysis

The monotonic increase in Glucose-L and HMF exhibited by the model (Figure 3B, 3C) is contradicted by the dataset where degradation is observed at a very high severity (du Pasquier, 2022). It reveals that the second order polynomial used is not precise enough to consider these degradations at the limits of the model. These models are therefore no longer reliable beyond a severity of 3-3.5.

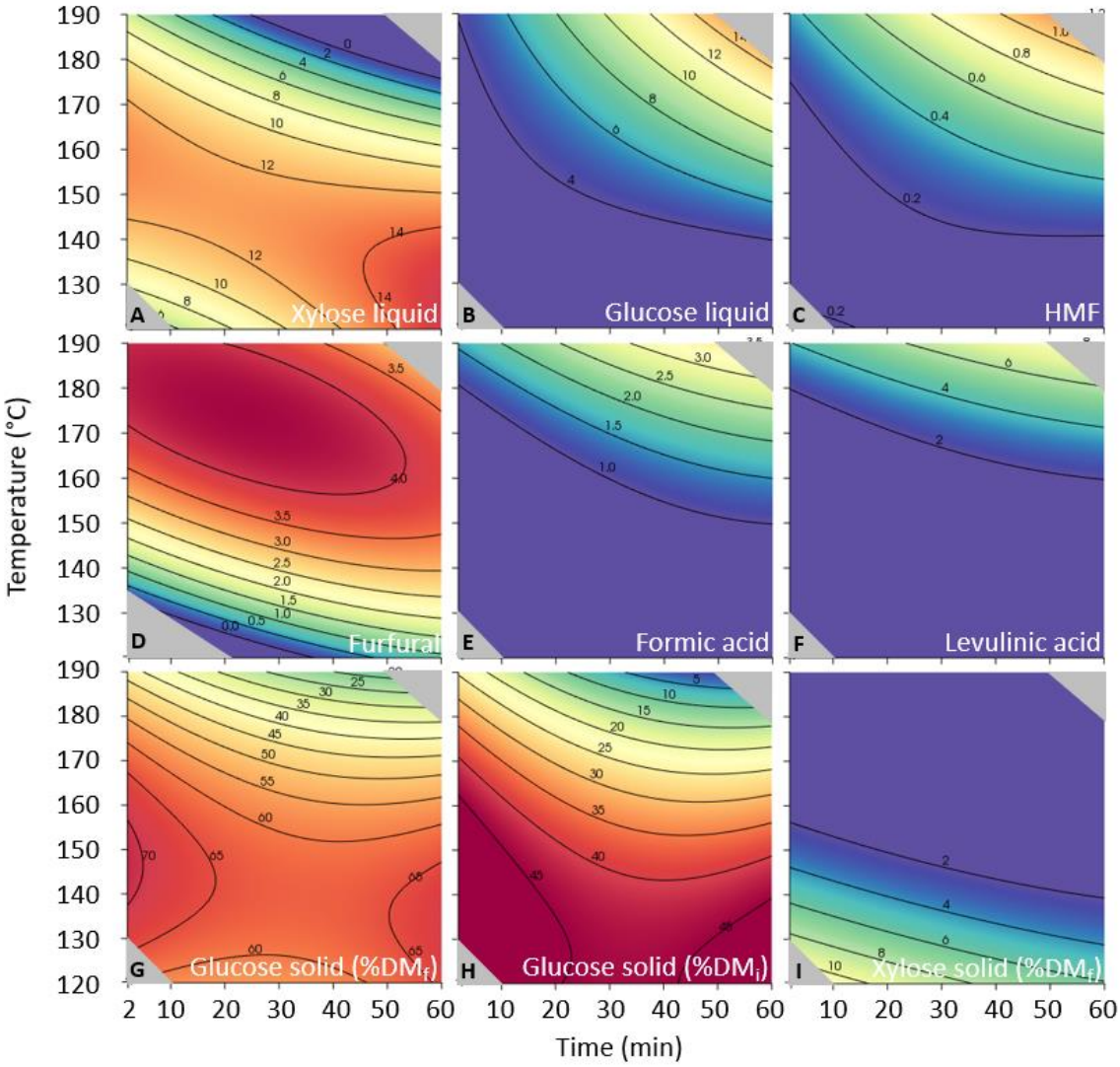

Figure 3: 2D plots obtained for some chemical species concentrations at acid levels of 2% depending on time and temperature. The gray areas indicate the accuracy limits of the models in which the predicted values are no longer reliable. A: Xylose-L; B: Glucose-L

This problem does not arise for Xylose-L and Furfural: their degradation kinetics take place at more moderate severities, far from the limits of the models (Figure 3A, 3D). The maximization of Xylose-L is possible over a very wide area for highly variable conditions up to a maximum of 160-170°C, beyond which degradation is unavoidable. For Furfural, a maximum is formally identified: 4.41% DM_i at 15min / 175°C / 2.5% H_2SO_4, which corresponds to a CSF of 3.1. This furfural optimum has been described by Swiatek *et al.*

(2020) with beech wood with a maximum level obtained for 10-15min / 200°C / 0.5% H_2SO_4. The temperature and acid levels are different, but these conditions also represent a CSF between 3 and 3.1 (Swiatek et al., 2020).

For the other inhibitors (Formic, Levulinic and Succinic acid, Figure 3E, 3F), their increase is consistent: these compounds come from the degradation of HMF and Furfural which occur at very high severities, so only the beginning of their formation can be observed. This is confirmed by previous observations showing an increase in their levels up to severities of 3.5-3.8 (Swiatek et al., 2020). Yet, the levels of HMF and succinic acid remain very low. HMF is a very unstable molecule in water under acidic conditions and is rapidly degraded (Kim et al., 2021). However, literature indicates that it is possible to obtain larger quantities at severity 3.5-3.8 but for temperatures of 200°C and above, which might represent threshold temperatures of deep changes in chemical reactions (Swiatek et al., 2020). The formation of succinic acid is a complex reaction usually requiring a catalyst (Dulie et al., 2021; Kim et al., 2021). At high severity and without catalyst, it is possible that other degradation pathways with lower activation energy are favored.

3.5.2. Solid phase analysis

Considering the solid residue, hemicelluloses degrade very quickly, including at low severities, whether in terms of proportion in the pretreated sample (DM_f, Figure 3I) or in terms of overall yield (DM_i). The cellulosic part is much more resistant, and it is possible to keep a high overall yield even at high severities until 2.5 where it starts to deteriorate (Figure 3H). These two phenomena are commonly observed in the literature, with degradation of cellulose starting between CSF 2 and 3 for poplar, beech and spruce wood, even at higher temperature (Swiatek et al., 2020; Yan et al., 2014). This dual effect of rapid degradation of hemicelluloses and resilient cellulose causes an increase in the relative content of cellulose in the pretreated sample that can reach 70% DM_f (Figure 3G), the

remainder fraction consists mainly of lignin. The pretreated sample is ultimately composed of 100% lignin and pseudo-lignin at a severity of 4 (Yan et al., 2014).

Experiments 30 and 31, corresponding to acid-free conditions, were repeatedly disabled in models for the solid fraction due to aberrant results (Table 2). It means that the degradation of solid samples without acid is fundamentally different from that with acid, even at very low acid level. Models for the solid fraction are therefore probably not correct for a zero-acid level.

Considering all models, it is possible to notice that, beyond 2% acid, further addition of acid seems to have less impact on the evolution of the compounds. The two experiments carried out at 6.84% (Table 1) are not statistically sufficient to confirm this observation. Nevertheless, literature shows that high-temperature levels increase the effect of the acid, which could explain that high concentrations are not needed (du Pasquier et al., 2023).

3.6. Desirability function

As seen before, the RSM expressions should be used carefully and must be considered as a convenient way to summarize the main trends. However, they can be combined and used jointly to propose a global approach of the effects of the pretreatment. This is the purpose of the desirability functions: all models are combined to fulfill different constrains simultaneously, and a maximization - or a minimization - of a combined criterion is determined. The resulting surface shows all areas where the desired combined criteria are satisfied.

3.6.1. Initial scenarios

As an example, two scenarios have been selected here (Figure 4). Both are based on maximizing the separation and recovery of fermentable sugars: conservation of cellulose in the solid phase and degradation of hemicellulose into xylose in the liquid phase (Glucose-S above 35% DM_i and 50% DM_f and Xylose-L above 8% DM_i). Other criteria have been added

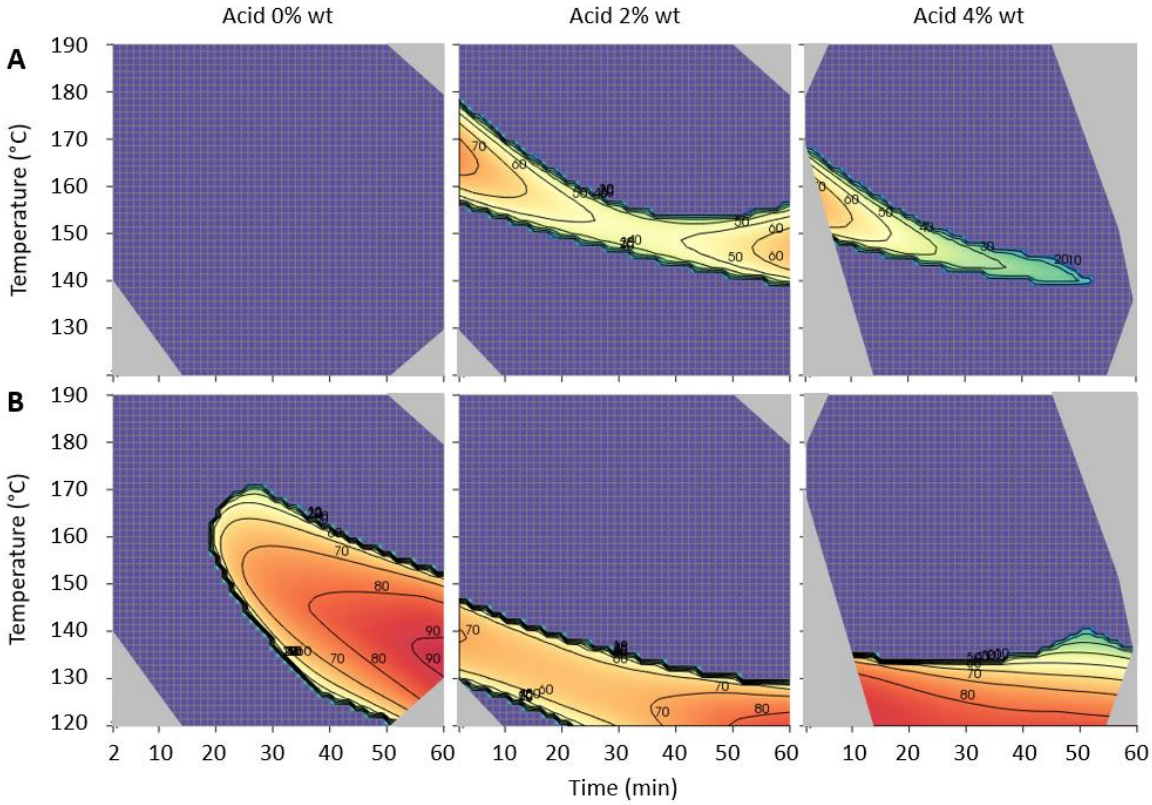

Figure 4: Desirability scenarios A and B, both maximizing xylose in the liquid phase and glucose in the solid phase. In addition, scenario A promotes the separation of cellulose and hemicelluloses in the solid phase and scenario B minimizes inhibitors. Gray areas mask the model accuracy limits.

to differentiate them: scenario A promotes the separation of cellulose and hemicelluloses in the solid phase (Xylose-S below 2% DM_i*)* (Figure 4A) and scenario B limits the production of inhibitors in the liquid phase (HMF and Succinic acid below 0.4%, Formic acid below 1%, Furfural and Levulinic acid below 2% DM_i*)* (Figure 4B).

These two scenarios have no overlapping zones and are therefore not compatible (Figure 4): it is then impossible to separate cellulose and hemicelluloses without producing any inhibitor. The complete degradation of hemicelluloses in scenario A needs severe conditions (140-180°C, minimum 0.8% acid) with CSF between 2 and 3 leading to the degradation of compounds into inhibitors, mainly Furfural between 3 and 4% DM_i. Conversely, preventing inhibitor production in scenario B requires low CSF from 1.5 to 2

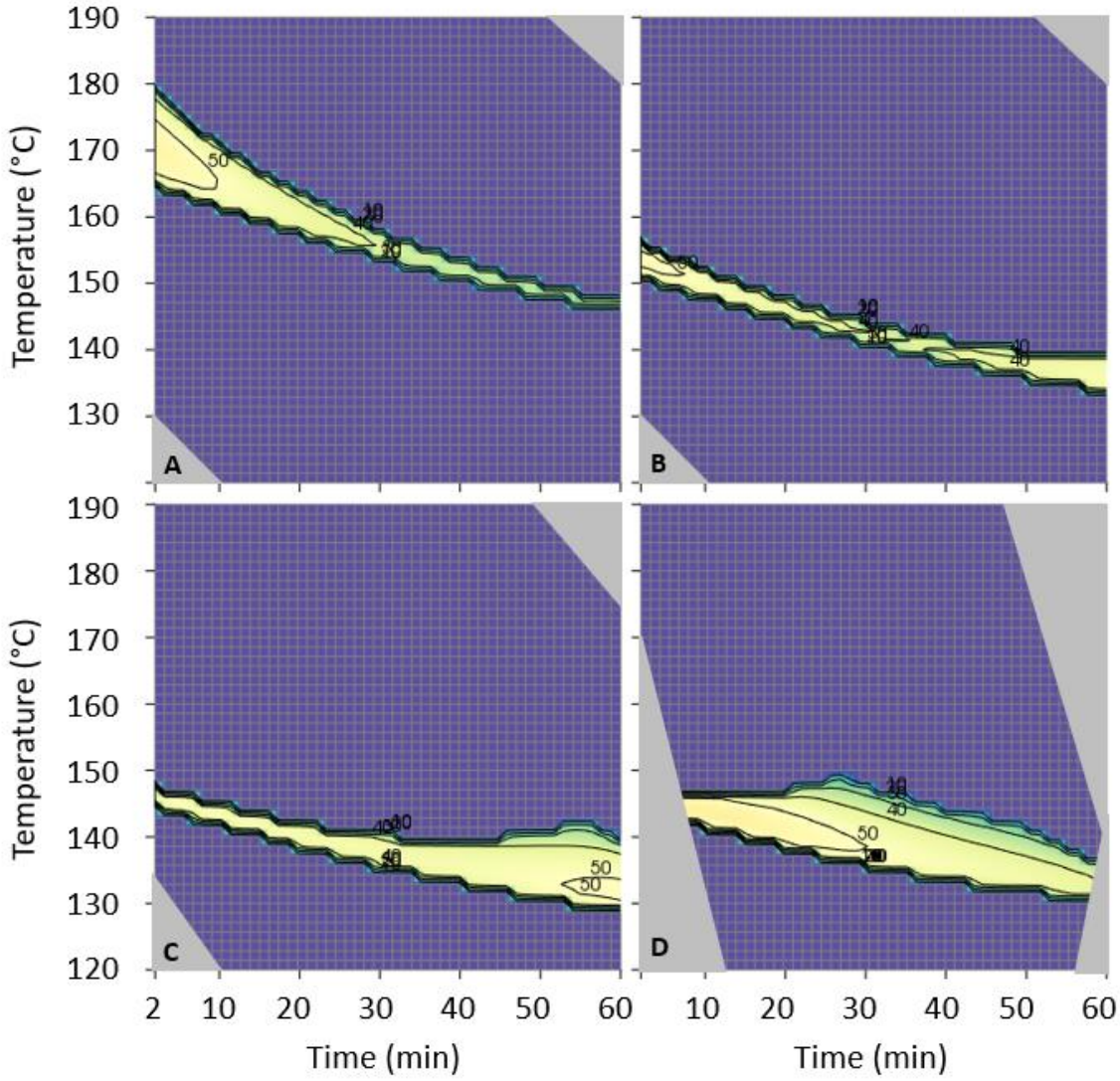

Figure 5: Desirability scenario AB' combining scenarios 1 and 2 and minimizing Xylose-S and Furfural below 3% DMi instead of 2% for different acid levels: A: 1%; B: 2%; C: 3%; D: 4%. Gray areas mask the model accuracy limits.

(120-150°C) which leaves more than 30% of the initial xylose in the form of hemicelluloses (5% DM_i).

3.6.1. Trade-off between all models

A trade-off could be found by increasing the acceptable levels of Xylose-S and Furfural (below 3% instead of below 2% DM_i) respectively in scenarios A and B. The

combination of these two scenarios defines a third scenario, AB', which respects all the criteria of the different chemical species (Xylose-L above 8% DM_i and Glucose-S beyond 35% DM_i and 50% DM_f, Xylose-S below 3%, HMF and Succinic acid below 0.4%, Formic acid below 1%, Levulinic acid below 2% and Furfural below 3% DM_i, Figure 5).

Scenario AB' reveals three trade-off zones, with CSF all between 2 and 2.5: a first one at 2-10 min / 150-170°C / 0.8-2 % H_2SO_4, a second one at 50-60 min / 130-135°C / 2-3 % H_2SO_4 and the last one at 10-30 min / 140-145°C / 3-4 % H_2SO_4 (Figure 5). All these zones offer similar compound levels, with furfural and xylose in the solid phase greater than 2% DM_i. Overall, two scenarios can be anticipated: a “flash” pretreatment at high temperature (170°C) during short time, which is very aggressive but requires less than 2% acid, or, on the contrary, soft pretreatments, with longer residence times and lower temperatures, but with higher acid level (2 to 4%). It is interesting to note that the required temperature no longer varies much beyond 2% acid, only the preprocessing time varies. In any case, this scenario shows that a temperature higher than 170°C and an acid concentration lower than 0.8% are ineffective (data not shown).

4. Conclusion

This work delivers to the scientific and industrial community a dataset of unprecedented size (38 DAP conditions ranging on 2-60 min / 120-190°C / 0-4%wt H_2SO_4) for the evolution of 12 main chemical species involved in the DAP of poplar wood. Thanks to the attention paid to sampling and to the description of the methods used, these data can be reused more easily than most of those from existing publications and reports to optimize production processes, whatever their objectives. An initial study of these data, carried out in this article by using multivariable polynomial models and desirability functions, has allowed us to propose examples of their exploitation: the search for production trade-offs between hemicelluloses degradation, cellulose preservation and the formation of inhibitors. The results show that it is impossible to completely separate and

maintain the integrity of the cellulosic and hemicellulosic fractions simultaneously without producing inhibitors. Some trade-offs exist, using high temperatures, low acid contents and fast treatments (2-10 min / 150-170°C / 0.8-2% H_2SO_4) or longer treatments at lower temperatures with more acid (50-60 min / 130-135°C / 2-3% H_2SO_4 or 10-30 min / 140-145°C / 3-4% H_2SO_4). These conditions correspond to severities between 1.5 and 3. Below this limit, hemicelluloses are not sufficiently degraded, while above, large amounts of inhibitors are produced.

Data availability

Datasets related to this article can be found at https://www.data.gouv.fr/fr/datasets/chemical-composition-of-solid-and-liquid-phases-resulting-from-dilute-acid-pretreatment-of-poplar-wood-under-different-conditions/, an open-source online data repository hosted at Data.gouv (du Pasquier, 2022).

Supplementary material

E-supplementary data for this work can be found in e-version of this paper online

CRediT authorship contribution statement

Julien du Pasquier: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization.

Gabriel Paës: Conceptualization, Writing - review & editing, Visualization, Project administration, Funding acquisition.

Patrick Perré: Conceptualization, Writing - review & editing, Visualization, Project administration, Funding acquisition.

Acknowledgment

The authors are grateful to Pr. Michelle Sergent and Dr. Magalie Claeys-Bruno for their help during the development of the DoE, the analysis of the models resulting from the RSM and for making Azurad software available, to Mr. François Gaudard for his help with the Klason lignin analyzes and to Mrs. Annabelle Dejardin and her colleagues at INRAE Val de Loire for supplying the raw material.

Formatting of funding sources

J. du Pasquier's PhD was funded by Grand Reims and Chaire de Biotechnologie de CentraleSupéelec. The authors would like to thank the Département de la Marne, Grand Reims, Région Grand Est and the European Union along with the European Regional Development Fund (ERDF Champagne Ardenne 2014-2020) for their financial support of the Chair of Biotechnology of CentraleSupélec.

References

1. Acosta, A.M., Cosovanu, D., Lopez, P.C., Thomsen, S.T., Gernaey, K.V., Canela-Garayoa, R., 2021. Co-cultivation of a novel Fusarium striatum strain and a xylose consuming Saccharomyces cerevisiae yields an efficient process for simultaneous detoxification and fermentation of lignocellulosic hydrolysates. *Chem. Eng. J*. **426**, 131575. 10.1016/j.cej.2021.131575.
2. Auxenfans, T., Cronier, D., Chabbert, B., Paës, G., 2017. Understanding the structural and chemical changes of plant biomass following steam explosion pretreatment. *Biotechnol Biofuels*. **10**, 36. 10.1186/s13068-017-0718-z.
3. Azurad, 2019. Intuitive and powerful software for the construction and processing of DOE, Marseille, France. http://www.azurad.fr/logiciel-plans-experiences.php (accessed 06/09/2023).

4. Baral, N.R., Shah, A., 2017. Comparative techno-economic analysis of steam explosion, dilute sulfuric acid, ammonia fiber explosion and biological pretreatments of corn stover. *Bioresour. Technol*. **232**, 331-343. 10.1016/j.biortech.2017.02.068.

5. Becker, J., Wittmann, C., 2019. A field of dreams: Lignin valorization into chemicals, materials, fuels, and health-care products. *Biotechnol adv*. **37**, 107360. 10.1016/j.biotechadv.2019.02.016.

6. Beig, B., Riaz, M., Naqvi, S.R., Hassan, M., Zheng, Z.F., Karimi, K., Pugazhendhi, A., Atabani, A.E., Chi, N.T.L., 2021. Current challenges and innovative developments in pretreatment of lignocellulosic residues for biofuel production: A review. *Fuel*. **287**, 119670. 10.1016/j.fuel.2020.119670.

7. Belmokhtar, N., Habrant, A., Lopes Ferreira, N., Chabbert, B., 2013. Changes in Phenolics Distribution After Chemical Pretreatment and Enzymatic Conversion of Miscanthus x giganteus Internode. *BioEnergy Res.* **6**, 506-518. 10.1007/s12155-012-9275-2.

8. Carpio, R.R., Miyoshi, S.D., Elias, A.M., Furlan, F.F., Giordano, R.D., Secchi, A.R., 2021. Multi-objective optimization of a 1G-2G biorefinery: A tool towards economic and environmental viability. *J. Clean Prod.* **284**, 125431. 10.1016/j.jclepro.2020.125431.

9. Chen, J.X., Zhang, B.Y., Luo, L.L., Zhang, F., Yi, Y.L., Shan, Y.Y., Liu, B.F., Zhou, Y., Wang, X., Lu, X., 2021. A review on recycling techniques for bioethanol production from lignocellulosic biomass. *Renew. Sust. Energ. Rev.* **149**, 111370. 10.1016/j.rser.2021.111370.

10. Chundawat, S.P.S., Beckham, G.T., Himmel, M.E., Dale, B.E., 2011. Deconstruction of Lignocellulosic Biomass to Fuels and Chemicals. in: *Annual Review of Chemical and Biomolecular Engineering*, (Ed.) J.M. Prausnitz, Vol. 2, Annual Reviews. Palo Alto, pp. 121-145.

11. [dataset] du Pasquier, J., 2022. Chemical composition of solid and liquid phases resulting from dilute acid pretreatment of poplar wood under different conditions. Data.gouv, v1. https://www.data.gouv.fr/fr/datasets/chemical-composition-of-

solid-and-liquid-phases-resulting-from-dilute-acid-pretreatment-of-poplar-wood-under-different-conditions/.

12. du Pasquier, J., Paës, G., Perré, P., 2023. Principal factors affecting the yield of dilute acid pretreatment of lignocellulosic biomass: A critical review. *Bioresour. Technol*. **369**, 128439. 10.1016/j.biortech.2022.128439.
13. Dulie, N.W., Woldeyes, B., Demsash, H.D., Jabasingh, A.S., 2021. An Insight into the Valorization of Hemicellulose Fraction of Biomass into Furfural: Catalytic Conversion and Product Separation. *Waste Biomass Valorization*. **12**, 531-552. 10.1007/s12649-020-00946-1.
14. Kim, J.-H., Choi, J.-H., Kim, J.-C., Jang, S.-K., Kwak, H.W., Koo, B., Choi, I.-G., 2021. Production of succinic acid from liquid hot water hydrolysate derived from Quercus mongolica. *Biomass Bioenergy*. **150**, 106103. 10.1016/j.biombioe.2021.106103.
15. Lancha, J.P., Colin, J., Almeida, G., Guerin, C., Casalinho, J., Perré, P., 2021. A validated Distributed Activation Energy Model (DAEM) to predict the chemical degradation of biomass as a function of hydrothermal treatment conditions. *Bioresour. Technol*. **341**, 125831. 10.1016/j.biortech.2021.125831.
16. Machinet, G.E., Bertrand, I., Barrière, Y., Chabbert, B., Recous, S., 2011. Impact of plant cell wall network on biodegradation in soil: Role of lignin composition and phenolic acids in roots from 16 maize genotypes. *Soil Biol. Biochem*. **43**, 1544-1552. 10.1016/j.soilbio.2011.04.002.
17. Manzon, D., Claeys-Bruno, M., Declomesnil, S., Carite, C., Sergent, M., 2020. Quality by Design: Comparison of Design Space construction methods in the case of Design of Experiments. *Chemometr Intell Lab Syst*. **200**, 104002. 10.1016/j.chemolab.2020.104002.
18. Mokdad, S.A., Casalinho, J., Almeida, G., Perre, P., 2018. Assessment of biomass alterations during hydrothermal pretreatment by in-situ dynamic mechanical analysis. *Biomass Bioenerg*. **108**, 330-337. 10.1016/j.biombioe.2017.11.014.

19. Pereira, L.M.S., Milan, T.M., Tapia-Blacido, D.R., 2021. Using Response Surface Methodology (RSM) to optimize 2G bioethanol production: A review. *Biomass Bioenerg*. **151**, 106166. 10.1016/j.biombioe.2021.106166.

20. Pu, Y.Q., Hu, F., Huang, F., Davison, B.H., Ragauskas, A.J., 2013. Assessing the molecular structure basis for biomass recalcitrance during dilute acid and hydrothermal pretreatments. *Biotechnol Biofuels*. **6**, 15. 10.1186/1754-6834-6-15.

21. RFA, Renewable Fuels Association, 2023. Annual Ethanol Production: U.S. and World Ethanol Production. https://ethanolrfa.org/markets-and-statistics/annual-ethanol-production (accessed 23/03/2023).

22. Ritchie, H., Roser, M., Rosado, P., 2022. Energy, Our World in Data. https://ourworldindata.org/energy (accessed 24/08/2023).

23. Swiatek, K., Gaag, S., Klier, A., Kruse, A., Sauer, J., Steinbach, D., 2020. Acid Hydrolysis of Lignocellulosic Biomass: Sugars and Furfurals Formation. *Catalysts*. **10**, 437. 10.3390/catal10040437.

24. Tan, J.Y., Li, Y., Tan, X., Wu, H.G., Li, H., Yang, S., 2021. Advances in Pretreatment of Straw Biomass for Sugar Production. *Front Chem*. **9**, 696030. 10.3389/fchem.2021.696030.

25. Vieira, S., Barros, M.V., Sydney, A.C.N., Piekarski, C.M., de Francisco, A.C., Vandenberghe, L.P.d.S., Sydney, E.B., 2020. Sustainability of sugarcane lignocellulosic biomass pretreatment for the production of bioethanol. *Bioresour. Technol*. **299**, 122635. 10.1016/j.biortech.2019.122635.

26. Yan, L., Zhang, L., Yang, B., 2014. Enhancement of total sugar and lignin yields through dissolution of poplar wood by hot water and dilute acid flowthrough pretreatment. *Biotechnol Biofuels*. **7**, 76. 10.1186/1754-6834-7-76.

27. Yang, H.S., Duncan, S.M., Hafez, I., Schilling, J.S., Tze, W.Y., 2016. Hydroxyl Availability in Aspen Wood After Dilute Acid Pretreatment and Enzymatic Saccharification. *BioResources*. **11**, 7490-7499. 10.15376/biores.11.3.7490-7499.

28. Yildirim, O., Ozkaya, B., Altinbas, M., Demir, A., 2021. Statistical optimization of dilute acid pretreatment of lignocellulosic biomass by response surface methodology to obtain fermentable sugars for bioethanol production. *Int. J. Energy Res.* **45**, 8882-8899. 10.1002/er.6423.

29. Yoo, C.G., Meng, X., Pu, Y., Ragauskas, A.J., 2020. The critical role of lignin in lignocellulosic biomass conversion and recent pretreatment strategies: A comprehensive review. *Bioresour. Technol*. **301**, 122784. 10.1016/j.biortech.2020.122784.

30. Zhi, S.L., Yang, J., Yao, Y., Zhang, S.T., Lu, X.B., 2013. Optimization of dilute acid hydrolysis of corn stover for separate production of xylose and glucose by response surface methodology. in: *Progress in Renewable and Sustainable Energy, Points 1 and 2*, (Eds.) Y.G. Li, Y. Li, W.G. Pan, Vol. 608-609, Trans Tech Publications Ltd. Durnten-Zurich, pp. 298-+.

31. Zoghlami, A., Paës, G., 2019. Lignocellulosic Biomass: Understanding Recalcitrance and Predicting Hydrolysis. *Front Chem*. **7**, 874. 10.3389/Fchem.2019.00874.