Check for updates

RESEARCH ARTICLE

# REVISED Bio-oil production from biogenic wastes, the hydrothermal conversion step

[version 2; peer review: 2 approved]

Geert Haarlemmer[1], Anne Roubaud[1]

[1]CEA/LITEN/DTCH, Université Grenoble Alpes, Grenoble, 38000, France

**v2** **First published:** 20 Sep 2022, **2**:111
https://doi.org/10.12688/openreseurope.14915.1

**Latest published:** 15 Dec 2022, **2**:111
https://doi.org/10.12688/openreseurope.14915.2

## Abstract

**Background:** Food wastes are an abundant resource that can be effectively valorised by hydrothermal liquefaction to produce bio-fuels. The objective of the European project WASTE2ROAD is to demonstrate the complete value chain from waste collection to engine tests. The principle of hydrothermal liquefaction is well known but there are still many factors that make the science very empirical. Most experiments in the literature are performed on batch reactors. Comparison of results from batch reactors with experiments with continuous reactors are rare in the literature.

**Methods:** Various food wastes were transformed by hydrothermal liquefaction. The resources used and the products from the experiments have been extensively analysed. Two different experimental reactors have been used, a batch reactor and a continuous reactor. This paper presents a dataset of fully documented experiments performed in this project, on food wastes with different compositions, conditions and solvents. The data set is extended with data from the literature. The data was analysed using machine learning analysis and regression techniques.

**Results:** This paper presents experimental results on various food wastes as well as modelling and analysis with machine learning algorithms. The experimental results were used to attempt to establish a link between batch and continuous experiments. The molecular weight of bio-oil from continuous experiments appear higher than that of batch experiments. This may be due to the configuration of our reactor.

**Conclusions:** This paper shows how the use of regression models help with understanding the results, and the importance of process variables and resource composition. A novel data analysis technique gives an insight on the accuracy that can be obtained from these models.

**Open Peer Review**

**Approval Status** ✓ ✓

| | 1 | 2 |
|---|---|---|
| **version 2** (revision) 15 Dec 2022 | ✓ view | ✓ view |
| **version 1** 20 Sep 2022 | ? view | ? view |

1. **Daniele Castello**, Aalborg University, Aalborg, Denmark

2. **Yuanhui Zhang**, University of Illinois at Urbana-Champaign, Urbana, USA
University of Illinois at Urbana-Champaign, Urbana, USA

Any reports and responses or comments on the article can be found at the end of the article.

**Keywords**
hydrothermal liquefaction, biocrude, biofuel, food wastes, bio-oil, biochar, machine learning



This article is included in the Chemical Engineering gateway.



This article is included in the Analytical Chemistry gateway.




This article is included in the Horizon 2020 gateway.



This article is included in the Green Chemistry collection.

**Corresponding author:** Geert Haarlemmer (geert.haarlemmer@cea.fr)

**Author roles: Haarlemmer G**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Roubaud A**: Conceptualization, Investigation, Methodology, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This research was financially supported by the European Union's Horizon 2020 research and innovation programme under the grant agreement No 818120 (Biofuels from WASTE TO ROAD transport [WASTE2ROAD]).
*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2022 Haarlemmer G and Roubaud A. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Haarlemmer G and Roubaud A. **Bio-oil production from biogenic wastes, the hydrothermal conversion step [version 2; peer review: 2 approved]** Open Research Europe 2022, **2**:111 https://doi.org/10.12688/openreseurope.14915.2

**First published:** 20 Sep 2022, **2**:111 https://doi.org/10.12688/openreseurope.14915.1

**REVISED** **Amendments from Version 1**

This new version of the paper has seen some significant changes. Upon suggestion of one of the reviewers, the aqueous phase recirculation experiments were removed from the manuscript. This section was not really essential to the paper due to the nature of the resources treated. Certain

The experimental procedures have been better explained and better structured and many sections have been improved. An important improvement was done on the calculation of the gas yields.

**Any further responses from the reviewers can be found at the end of the article**


## Plain language summary

Food wastes are an abundant resource that can be valorised by hydrothermal liquefaction. Hydrothermal liquefaction is a process taking place in hot compressed water, typically around 300 °C and above 100 bar. Under these conditions, water is an effective medium to transform organic resources into a black oily product we call biocrude, which can be further upgraded to bio-fuel. The European funded project WASTE2ROAD project aims to demonstrate the full value chain. The principle of hydrothermal liquefaction is well known but there are still many factors that make the science very empirical.

There are essentially two types of reactors in laboratories: generic batch reactors (the majority), and specially designed continuous reactors. Comparison of batch results with experiments with continuous reactors are rare but essential to be able to valorise the large body of data in the literature. Prediction of the results by simulation is often limited as there are many factors that are not measured or cannot be accounted for. This paper presents a dataset of fully documented experiments on food wastes, with different compositions, conditions and solvents. The data from this study is also extended with literature data.

There have been many publications that present correlations or kinetic models for a specific resource. Extrapolation to other resources is hazardous. Recently a different type of modelling, heavily used in Artificial Intelligence and Data Science, makes its way into chemical engineering. These modelling tools are referred to as Machine Learning algorithms. A novel data analysis technique based on machine learning tools give an insight on the accuracy that can be obtained from these models. The experimental results are further used to attempt to establish a link between batch and continuous experiments.


## Introduction

The H2020 European project "Biofuels from WASTE TO ROAD transport" WASTE2ROAD aims to develop a new generation of cost-effective biofuels from a selected range of low cost and abundant biogenic residues and waste fractions. The established consortium covers the full value chain, from waste management, the technological process of transforming waste to advanced biofuels to the assessment of the end-use compatibility of the obtained biofuels. This will be achieved through transformation of a diverse range of waste (and fractions thereof) into intermediate bio-oil, deploying both fast pyrolysis and hydrothermal liquefaction (HTL).

HTL converts biomass compounds in hot compressed water into a biocrude. This biocrude is an oily material containing bio-oil and char. This process has already been known for some time. The developments started in the 80's in Europe[1,2] and in the United States[3]. The conversion takes place at temperatures between 300 and 400 °C and at pressures above the saturation pressure to ensure that water remains in the liquid phase, typically above 100 bar. Under these conditions the ionisation of water increases while its polarity decreases, favouring depolymerisation and dehydration of biomass biopolymers to produce hydrophobic compounds[4]. This process is well adapted for wet resources avoiding an energy consuming step to dry the resource prior to for example combustion, pyrolysis of gasification[5]. Our previous work on HTL of agro-industrial residues has shown that the biochemical composition of the initial matter is the major parameter influencing conversion efficiency and quality of the product[6–8].

There is a large volume of literature on the basic transformation in batch reactors. Results on continuous reactors show some subtle differences from batch reactors making the results often more difficult to interpret[6]. The chemistry is complex accentuating the differences between batch end continuous reactors. A significant amount (up to 40 % of the organic dry matter) of the resource is transferred to the aqueous phase product during the transformation. Aqueous phase recycling has an important effect on the results of hydrothermal liquefaction. This has been noted earlier by Déniel *et al.*[9] and Biller *et al.*[10]. From a practical point of view, recycling of the water phase is important to limit the volume of discharged water and to optimise the use of the resource. For wet resources, aqueous phase recycling present limited advantages.

The complexity of the resources and large number of process variables make it difficult to fully understand the transformation in one study. The literature is very rich in data that can be exploited to understand the conversion of a particular resource under particular conditions. Multiple studies have been exploited in meta studies where data from a large number of publications is analysed with machine learning tools[11–13]. The modelling approach in these papers allows the regression of powerful and accurate models and also allow some basic data analysis. These models are not in equation form and not readily transposable, as they are very complex and relatively meaningless when taken out of context[14]. The analysis of the data can be pushed further with a game theory approach as was demonstrated by Onsree *et al.*[15].

Different types of resources have been considered for HTL conversion in the European funded project WASTE2ROAD. These include food waste, the fermentable fraction of municipal solid waste and its digested counterpart. This paper presents experimental results from these resources under a variety of

conditions, and two different reactors. A literature study also identified compatible data that has been used to construct a large homogeneous data set of similar experiments that is analysed with machine learning tools. The objective is to identify plausible explanations for observations.

## Methods

### Resources

The resources used in the project are raw food residues as well as residues from the digestion of food waste. Raw food waste (FW) was sourced from a company restaurant at the Commisariat à l'Energie Atomique et aux Energies Alternatives (CEA) campus in Grenoble. The restaurant which is named 'H1' was selected as it is the nearest restaurant (out of three) to our laboratory. The organic fraction of municipal solid waste (FFOM) was supplied by Suez in Montpellier France. This resource is usually used as a feed for methanisation by anaerobic digestion. Energi Gjenvinnings Etatens (EGE) in Norway (Oslo city waste management company) provided anaerobically digested food waste (DFOR) from their methanisation plant.

The food wastes from CEA are directly taken from the restaurant after the daily service. This waste is a mixture of peels from the food preparation, coffee marc, food that was put on display but not consumed and residues from the plates, Figure 1. The majority of the waste was non-comestible wastes from the food preparation. For economic and legal reasons, as little as possible comestible residues are produced. There is an important daily variability in these wastes. The collection was done during three months period from October to December 2019, in batches of 10 to 20 kg. The collected resources were dried, ground and mixed to ensure they were homogeneous and representative. The DFOR (Figure 1) and FFOM samples are much more homogeneous and they come from large processing units, for both resources around 50 kg was supplied.

All resources have been analysed by standard methods applicable to foodstuffs, subcontracted to an accredited commercial laboratory Capinov (Landerneau, France). Lipids are quantified by n-hexane extraction[16]. This method first hydrolyses the resource with hydrochloric acid. Oil is extracted from resulting product by hexane extraction in a Soxhlet device. Fibres are quantified using the methodology as described in the international standards (ISO)[17,18]. The sample is first cleaned from lipids by acetone extraction and proteins with digestion with proteinase. The method is based on a series of extractions using a Neutral Detergent, an Acid Detergent a, extraction by sulphuric acid. The resulting values are Neutral Detergent Fibres (NDF), Acid Detergent Fibres (ADF) and Acid Detergent Lignin (ADL). The hemicellulose content is calculated from the difference between NDF and ADF, the cellulose content is the difference between ADF and ADL while ADL is the lignin content.

Proteins are quantified by multiplying Kjeldahl nitrogen by 6.25. The Kjeldahl method only doses ammoniacal and amine nitrogen (from degraded proteins) but does not quantify nitrates. All organic nitrogen is converted in ammonia that is then quantified by absorption in a boric acid solution[19].

The data include proximate (moisture and ash) and ultimate (elemental) analysis. FFOM and DFOR contain a relatively large amount of ash, some in the form of glass and porcelain particles. Table 1 presents the results of the analyses. The structural compositional analysis was performed by the accredited commercial laboratory Capinov.

Chemicals used in this experiment are 2-propanol, acetone, tetrahydrofuran, dichloromethane, ethyl acetate and n-hexane, were acquired from Sigma-Aldrich. Reagent for Karl-Fischer titration, Hydranal Composit 1, was purchased from Honywell. Polycal polystyrene standards for the gel permeation chromatography were purchased from Malvern Panalytics.

### HTL experiments

For each resource, efficiency and product properties were determined by batch experiments. All resources can be treated in batch experiments. The two industrial wastes derived from food waste, FFOM and DFOR, were rich in hard abrasive particles such as glass and ceramics. They cannot be pumped to high pressures without damaging the equipment. For this reason these were only evaluated in batch experiments.

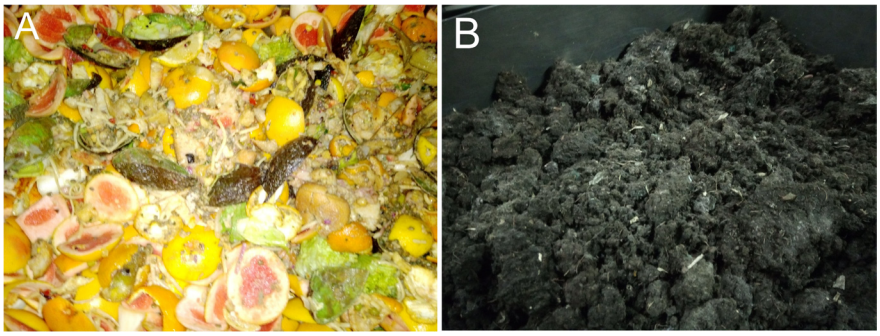



**Figure 1.** Example of food wastes (FW) from CEA restaurant (**A**) and Solid bio-residues from biogas reactors (DFOR) provided by EGE, Norway (**B**).

**Table 1. Results of analyses of organic wastes.** FFOM, the organic fraction of municipal solid waste; FW, food waste; DFOR, anaerobically digested food waste.

| Feedstock description | | FFOM | FW1 | FW2 | DFOR |
|---|---|---|---|---|---|
| **STRUCTURAL COMPOSITION** | | | | | |
| **Cellulose** | wt% | 30.1 | 8.6 | 7.7 | 8.8 |
| **Hemicellulose** | wt% | 5.5 | 25.3 | 12.5 | 9.8 |
| **Lignin** | wt% | 7.5 | 1.9 | 2.8 | 24.2 |
| **Proteins** | wt% | 9.6 | 18.2 | 16.5 | 29.5 |
| **Lipids** | wt% | 3.8 | 10 | 11.5 | 1.3 |
| **Sugars/dry basis (by difference)** | wt% | 13.0 | 27.9 | 40.1 | 0 |
| **PROXIMATE ANALYSIS** | | | | | |
| **Total moisture, as received** | wt% | 50 | 90 | 82 | 42 |
| **Ash 550 °C** | wt% | 22.5 | 5.1 | 4.9 | 28 |
| **ULTIMATE ANALYSIS** | | | | | |
| **Carbon (C)** | wt% | 37.7 | 47.3 | 43.8 | 37.6 |
| **Hydrogen (H)** | wt% | 5.2 | 6.3 | 8.1 | 5.9 |
| **Nitrogen (N)** | wt% | 1.7 | 3.2 | 3.2 | 4.8 |
| **Sulfur (S)** | wt% | 0.3 | 0.1 | 0.2 | 0.5 |
| **HEATING VALUE** | | | | | |
| **Higher heating value (HHV)** | MJ/kg | 13.8 | 20.3 | 20.5 | 15.9 |

Hydrothermal liquefaction experiments were performed in a 0.6 L stainless steel (SS316) stirred batch reactor (Parr Instruments). Figure 2 shows a photo and the schematics of the batch reactor. In a typical experiment, the reactor was filled with 300 g of biomass slurry prepared from 30 g resource and 270 g distilled water or aqueous phase recovered from a previous experiment. All experiments were performed at 300 °C and with different holding times. The pH was measured before and after each experiment using a pH meter. The autoclave was always leak tested, purged of air and pressurized to 10 bar with nitrogen gas. The latter is to ensure a sufficient pressure for gas analysis after the reaction even if not much gas is produced by the reaction. The pressure inside the reactor is a function of the initial amount of nitrogen and the reaction temperature. The amount of water plays a role in that the water occupies space in the reactor that cannot be occupied by the initial and produced gasses. The reactor was heated to the reaction temperature with a constant heating rate of 15 °C/min. Once the reactor had reached the reaction temperature, it was held during a specified time within ± 1 °C of the specified operating temperature. The reactor is stirred at s speed of 600 rpm. After the holding time, the reactor was rapidly cooled to room temperature in 20 min by an air quench.

The gas production is calculated with ideal gas law using its composition, pressure and temperature before and after each run. The ideal gas law is a simplication and the gas density is underestimated by about 1% compared to the Peng–Robinson or Soave–Redlich–Kwong equations of state, under these conditions.

Carbon dioxide is dissolved in significant quantities and taken into account with Henry's law. In a typical run, a partial pressure $CO_2$ of 2–4 bar is observed. The aqueous phase is slightly acidic (pH 4 to 4.5) making it possible to approach the $CO_2$ dissolution ignoring the acid base reactions. Under these conditions, molar concentration (C) of dissolved carbon dioxide can be calculated with Henry's law (Equation 1 and Equation 2).

$$C^{l}_{CO2} = P_{CO2} \cdot H_{CO2,H2O,T} \cdot \rho_{H_{2O}} \quad \text{Eq. 1}$$

$$n_{dissolved_{CO2}} = C^{l}_{CO2} \cdot V_{l} \quad \text{Eq. 2}$$

$C^{l}_{CO2}$ : molar concentration of dissolved $CO_2$ (mol·L$^{-1}$)

$P_{CO2}$: Partial pressure of $CO_2$ (bar)

$\rho_{H_2O}$: Density of water (kg·L$^{-1}$)

$H_{CO2,H2O,T}$: Henry's constant for a given temperature (mol·kg$^{-1}$·bar$^{-1}$)

$V_l$: Volume of liquid (L)

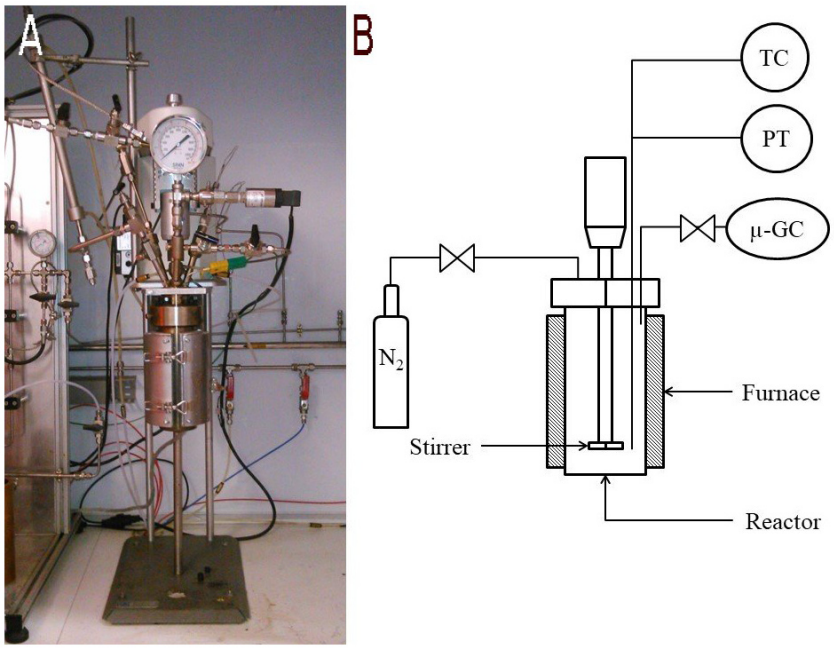



**Figure 2.** Batch hydrothermal reactor at CEA, photo (**A**) and schematic (**B**). TC (Thermocouple), PT (Pressure Transmitter), μ-GC (Connection for the gas analysis by micro chromatography).

Henry's coefficient is highly dependent on temperature. In the case where gas was analysed under a different temperature, Henry's coefficient can be calculated according to Equation 3.

$$H_{CO2,H2O} = H_{CO2,H2O}{}^{0} \cdot e^{\left(\frac{d\ln(H)}{d\frac{1}{T}} \cdot \left(\frac{1}{T} - \frac{1}{T^{\circ}}\right)\right)} \quad \text{Eq. 3}$$

$\frac{d\ln(H)}{d\frac{1}{T}}$: 2400 (K) [39]

$T^{0}$: 298.15 (K)

$H_{CO2,H2O,298.15\,K}{}^{0}$= 0.034 mol•kg$^{-1}$•bar$^{-1}$

The total amount of gas produced is the sum of gas measured from the pressure increase with ideal gas law and the amount of gas $CO_2$ dissolved in the water according Equation 4.

$$m_{gas_{produced}} = n_{gas_{measured}} \cdot \sum Mw_i x_i + n_{dissolved_{CO2}} \cdot MCO_2 \quad \text{Eq. 4}$$

$m_{gas_{produced}}$: Mass of gas produced (g)

$n_{gas_{measured}}$: Quantity of gas calculated by ideal gas law (mol)

$Mw_i$: Molecular weight of the species in the gas (g•mol$^{-1}$)

$x_i$: Molar fraction of the species *i*.

The final gas yield is defined as the ratio of the mass of gas produced over the mass of initial biomass introduced. In practice, taking into account the amount of dissolved carbon dioxide nearly doubles the gas yield in batch experiments.

Continuous experiments were performed on our 1.5 L/h test bench (TOP Industrie) in similar temperature conditions with selected resources based on the batch experiment results. The experiments were performed with 98 g/L FW2 enriched with 10 g/L of used cooking oil, named FW2CO. The suspension was stabilised by the addition of 2 g/L xanthane to avoid precipitation in the transfer lines. In the case of the continuous reactor, this means a volume flow rate of 1.5 L/h. Considering the effective volume of the reactor of 0.5 L, this leads to a residence time of 20 minutes. The temperature setpoint on the reactor is set to 300 °C. Figure 3 shows a photo and the schematics of the continuous reactor. The installations are described in more detail, including residence time distribution measurements, in Briand *et al.*[6] and Briand[20]. The reactor is heated while operating on a water feed. Once the installation is stable the switch to the biomass slurry tank is made. Operation is typically between 5 and 30 hours. The effective internal volume of the reactor is 0.5 L, leading to an averaged residence time of 20 min. Gas production is measured by the pressure increase on the product tank, followed by gas analysis by micro-chromatography.

An essential difference between the two reactors is that the batch reactor is heated up together with the reaction mixture while the continuous reactor is continuously heated, fluids are admitted in the already hot reactor.

## Product recovery

Before opening of the reactor or the product tank for the continuous reactor, the gas was first analysed. The reactor or tank was then opened and the products were recovered following the procedure given in Figure 4. The content of the reactor was first filtered on a Buchner filter (or a larger metallic filter

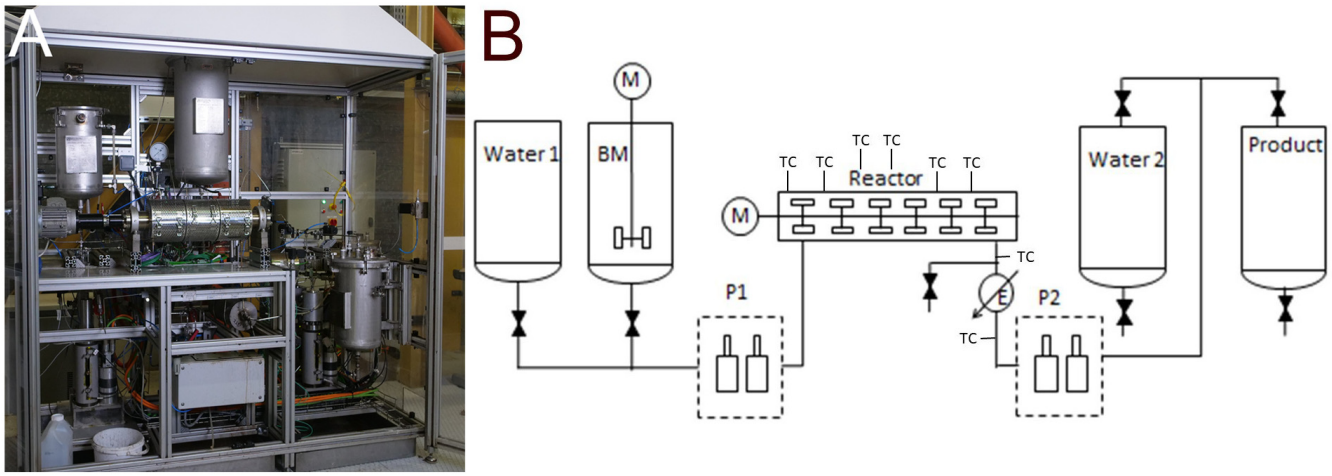



**Figure 3.** Continuous hydrothermal reactor at CEA, photo (**A**) and schematic (**B**). With the biomass tanks (BM), the location of the thermocouples (TC) and the pumps P1 and P2.

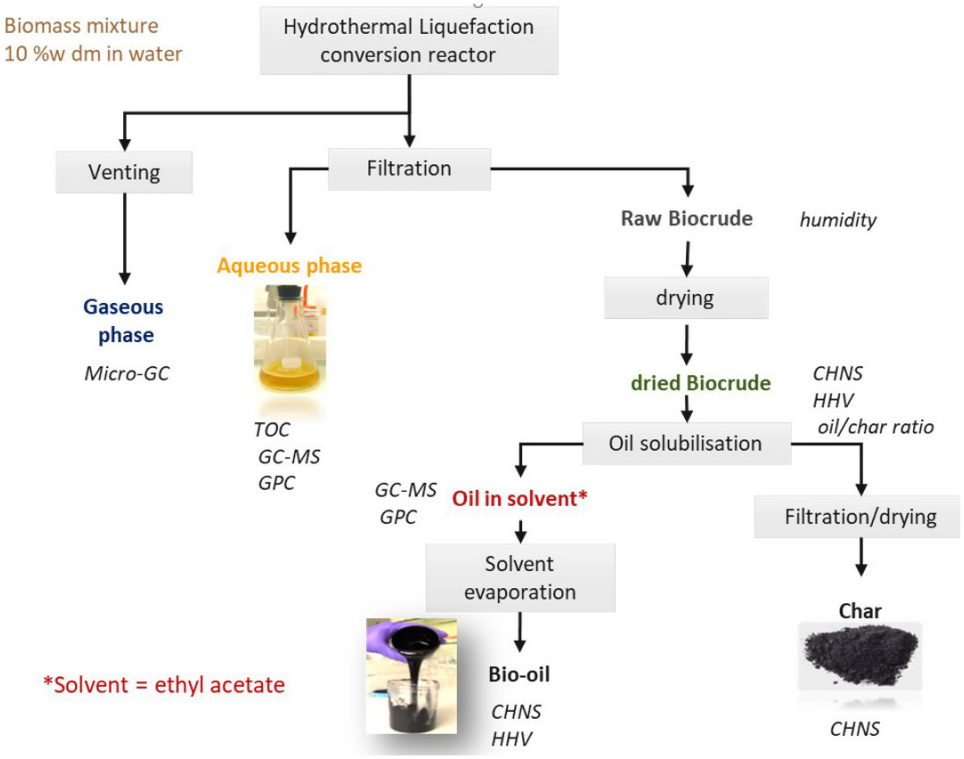



**Figure 4. Recovery procedure for products after hydrothermal liquefaction.** The specified analysis methods are Carbon-Hydrogen-Nitrogen-Sulphur (CHNS), Higher Heating Value (HHV), Gas Chromatography coupled with Mass Spectroscopy (GC-MS), Gel Permeation Chromatography (GPC) and Total Organic Carbon (TOC).

system for the continuous installation) to separate the aqueous phase from the raw organic residue. The raw organic residue is sometimes viscous and sticky but can also be a powder like product. The biocrude was removed from the batch reactor as best as possible. The empty reactor is dried in an oven at 70 °C, to remove any water. The biocrude remaining in the reactor is evaluated by comparing the weight of the empty reactor compared to the clean reactor before the experiment, typically less than one gram. The moisture content of the raw organic residue was estimated depending on the aspect of the product using one of two methods described as follows. Drying at room temperature under air circulation until a stable mass

was obtained if efficient for products that are not viscous, with the aspect of a dry powder. Many biocrudes are a viscous material containing bio-oil, char and some water. Air drying is not very effective as the top layer becomes quickly impermeable[21]. Karl Fischer titration was performed in these cases.

Depending on the proportion of bio-oil and char, the aspect of the raw organic residue can vary from an oily solid to a free-flowing viscous residue. When the char content is high, the bio-oil cannot be directly valorised and solvent extraction is necessary to separate the liquid from the solid fraction. To evaluate the char and oil yields individually, extractions are made on aliquots of the biocrude with different solvents as listed with the experimental data (see *Underlying data*). Two grams (weighed with a precision of 0.1 mg) of biocrude was washed with the respective solvent until the solvent runs off clear. Bio-oil can be recovered after evaporation of the solvent. The char was dried in an oven at 105 °C to remove any residual solvent until a stable weight was obtained (weighed with a precision of 0.1 mg). The proportion of solvent-soluble organics in the biocrude, and therefore the bio-oil yield is the biocrude yield minus its humidity (at the time of the extraction) and char content. This is by no means a substitute to an eventual industrial process, but purely an analytical technique to quantify yields.

Mass yields were calculated from the obtained experimental mass of the different phases after the experiments. Yield are defined as the mass ratios between the recovered phases and the dry biomass used in the experiment. In this paper we only report the bio-oil ($Y_{BO}$), char ($Y_C$) and gas ($Y_G$) yields. The quantity of the organic matter in the water phase is difficult to assess by simple drying as many compounds are volatile. In the literature, the aqueous phase yield is sometimes calculated by difference, closing the mass balance on the organic matter to 100%[22,23]. The aqueous phase yield includes the mass balance closure error. Hydration and dehydration reactions make that the overall organic mass balance does not necessarily close to 100%. For this reason we do not report the mass yield of organics in the aqueous phase together with the mass yield of the other phases, since it cannot be accurately determined.

## Product analysis

The gaseous phase was analysed by a micro-chromatograph (Varian Quad CP 4900) that samples the gas from the headspace of the reactor. Permanent gases ($O_2$, $H_2$, CO, $CO_2$ and $CH_4$) were analysed by a molecular sieve column using argon as carrier gas. Light hydrocarbons ($C_2H_2$, $C_2H_4$, $C_2H_6$, and $C_3H_8$), and sulphur species ($H_2S$ and COS) were analysed on a Poraplot-U column using helium as carrier gas.

The molecular composition of the bio-oil was analysed by a Gas Chromatograph coupled with a Mass Spectrometer, GC-MS (Clarus 500/ Clarus 600S, Perkin Elmer, USA) equipped with a DB-1701 capillary column 60 m × 0.25 mm, 0.25 µm film thickness. A 1 µL sample was injected into the instrument with a split ratio of 10:1. Helium was used as carrier gas. The GC oven temperature was programmed from 45 °C (10 min) to 230 °C at a rate of 6 °C/min, and held at 230 °C during 9.17 min. It was then raised to 250 °C at a rate of 10 °C/min, held at 250 °C during 20 min. The NIST database (NIST/EPA/NIH Mass Spectral Library version 2.0d) was used to identify the peaks.

Water content was determined by Karl-Fisher titration based on the reaction between water, sulphur dioxide and iodine on one side and sulphur trioxide and hydrogen iodate on the other (Equipment used Schott Titraline KF).

A portion for analysis is also subjected to azeotropic distillation with toluene to determine the amount of water that can be recovered from the bio-crude. The acidity (Total Acid Number, TAN) was determined by titration according to the method described by Anouti *et al.*[24]. A total organic carbon analyser (Shimadzu SSM-5000A) quantified the total carbon of the solid and oil samples. A total organic carbon analyser (Shimadzu TOC-L CSH/CSN) quantified total carbon of the aqueous phase.

Gel Permeation Chromatography (GPC), also often referred to as Size Exclusion Chromatography (SEC), was used to characterise the bio-oils in terms of molecular weight. The equipment used is a Viscotek TDA305-010 (Malvern Panalytical). The columns are the T1000, T2500 and T4000 with tetrahydrofuran as eluent. The data is presented in terms of averaged molecular weights in Dalton (Da, equivalent to g/mol). The calibration curve was made from 18 standards with molecular weights ranging from 162 Da to 400 kDa. The chromatogram obtained for the oil is compared to calibration curve to obtain the actual molecular weight distribution. The results are presented averaged by number (Mn) by weight (Mw) and the peak molecular weight (Mp).

## Data collection

The data collected in a single resource is often too small to be of real significance. To complete the dataset we have also included experiments from other authors working with similar food wastes. Even though the literature in HTL is very extensive, zooming in on food wastes, presenting complete data sets severely limits the available data. Data from Motavaf *et al.*[25], Bayat *et al.*[26], Aierzhati *et al.*[27], Evcil *et al.*[28] and Yang *et al.*[29–31] are also included. Experiments on soy protein are also included in the data to represent high protein resources[32–35]. Table 2 gives on overview of the features, or variables, considered in this study. There are 243 data points in the dataset, small in absolute terms for machine learning and artificial intelligence, but it represents a significant part (if not nearly all) of the available data in the literature. There are more published results on food wastes in the literature, but most of these papers do not present resource composition or fully documented experiments, rendering them unexplainable. It should be noted that Motavaf, Aierzhati and Evcil do not present data for char yield, only bio-oil.

The dry ash free composition of the full data set is presented in Figure 5. Lignin, typically absent or present in low

**Table 2. Range of the data in the included dataset.**

| Independent variables | Range |
|---|---|
| Temperature | 200-380 °C |
| Holding time | 0-80 min |
| Dry matter | 4–22 % |
| HHV | 13-32 MJ/kg |
| Ash content | 1–30 % |
| Protein content | 2–100 % |
| Lipids content | 1–72 % |
| Carbohydrates content | 1–64 % |
| Lignin content | 2–25 % |
| Solvents | Ethyl Acetate<br>Dichloromethane<br>Acetone<br>n-Hexane |
| Method | Order solvent extraction after or before water separation (values 1 and 2 respectively) |

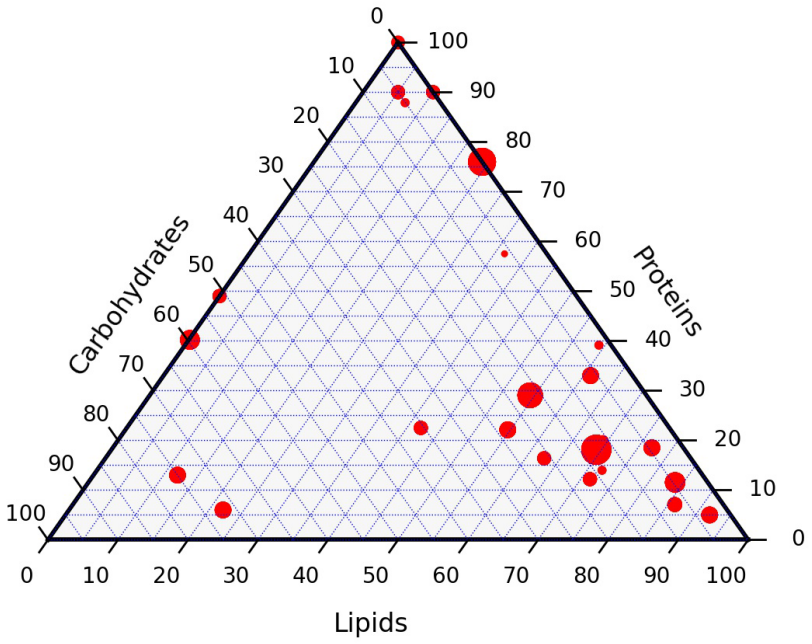



**Figure 5. Ternary diagram composition of the resources in this study.** Lignin is grouped with the carbohydrates.

proportions, is added to the carbohydrates for the sake of the graph. The composition of the resources are used as reported. The dataset covers a wide variety of resources that can be encountered in hydrothermal liquefaction. Individual data points have not been labelled. The size of each point is proportional to the number of data points with this particular composition. The points are not uniformly distributed on the ternary diagram but the resources are fairly representative for typical resources of this category. The full data is available from the *Underlying data*[36].

The solvent extraction in our experiments is after separation of the water phase. This is also the case for the data presented by Yang *et al.*[31] and Evcil *et al.*[28]. Other authors introduce the solvent directly into the reactor, this increases the oil yield as some of the organics in the aqueous phase is also extracted and included in the oil yield. A numerical value is attributed to the extraction method, this allows the extraction order to be included in the regression algorithm[12]. The value of 1 is used when extraction is after water separation, the value of 2 is used when the solvent is added to the reactor after the experiment. This approach allows the regression algorithm to include this parameter in the modelling equations. Different solvents are characterised by the relative polarity. The values can be found in the *Data availability* section.

### Data analysis with machine learning algorithms

Machine learning is a branch in the large family of artificial intelligence field. The tools in this field are very powerful modelling tools that complement, or even replace, classic polynomial correlations used to model the results from a design of experiment (DOE) approach. The advantage of the machine learning approach is that the data does not need to be highly structured. What counts is the variability of the data and the volume of the dataset. It is also possible to mix pure process parameters (such as the temperature) with indications on the method (as long as we use a numerical value).

Any model created from polynomial correlations, machine learning or any other regression framework devoid of first principles must be taken with caution. Larson's book on artificial intelligence[37] describes the story of the turkey (the original is a chicken from the English philosopher Bertrand Russel) that creates a model by inductive inference of its perfectly comfortable world by observing feeding times and accurately predicting feeding times day after day. One day, on Christmas Eve, this model suddenly no longer works with dramatic consequences. This shows us that any model inferred from observations, how valid they may be, is necessarily limited to the narrow scope of its data validity. The inconvenience of modelling of, what in the end is a chemical reaction without underlying chemical knowledge, is that extrapolation is hazardous. The risk is however somewhat limited when the data set is homogeneous and the model is used for interpolation.

The algorithms used in this study are well known algorithms from the freely available SciKit-Learn library (version 0.24.2)[38] and implemented in Python 3.9 (see *Extended data*). Two different regressors have been used in this study, the linear regressor (LinearRegression algorithm) and the random forest regressor (RandomForestRegressor algorithm). The linear regressor produces a very simple linear correlation that is not reputed for its accuracy for arbitrary (and nonlinear) problems but is easy to understand and not prone to overfitting. The random forest regressor is a robust ensemble method based on multiple decision trees. The built in boot-strapping algorithm samples data as it goes along and provides a good protection against overfitting.

Data from studies in any field are subject to uncertainties. In the machine learning and artificial intelligence field, process variables are referred to as features. Uncertainties come from experimental errors due to measurement accuracy and differences in analysis techniques that are not always specified in detail. We may have a very good repeatability of the HTL experiments; it is possible that the resource analyses are not repeatable. Variations also arise from (more or less) subtle differences between resources, as biomass is notoriously variable. Another significant problem arises from different analysis methods. Interpretation becomes much more difficult, when a same property is evaluated with different methods, especially when none of the methods yields an absolute value and are only based on estimates (case of protein content). Less appreciated are uncertainties due to unquantified or unrecognised variables, often referred to as latent features. The accuracy of any model is related to the characteristics of the regression model as well as the data used for the regression. The accuracy that can be obtained from modelling with the data in this study is evaluated using MAPIE (Model Agnostic Prediction Interval Estimator, version 0.3.1)[39]. This Python library allows the identification of confidence intervals on data modelling with an arbitrary regressor. The theoretical basis of this library is described by Kim *et al.*[40] and Barber *et al.*[41]. The training data contains features ($X$) and experimental results ($Y$) with an uncertainty ε, as expressed by Equation 5. The function $\mu$ is the model function.

$$Y = \mu(X) + \varepsilon \qquad \text{Eq. 5}$$

With α being the quartile, each new element has the probability $P$ to be in the confidence interval $1$-α, Equation 6.

$$P\left\{Y_{n+1} \in C_{n,\infty}(X_{n+1})\right\} \geq 1 - \infty \qquad \text{Eq. 6}$$

The naïve method, as coded in the Gradient Boosting Regressor in the SciKit-Lean Library, creates a model to fit the entire training set for a specified quantile, the fraction of the data being outside the distribution. This technique is prone to overfitting and generates large uncertainty intervals. MAPIE uses the Jacknife+ method, based on a leave-one-out approach. The model is fit successively on the full training set, with one left out. The residual of the left out data point is computed. The regression is then performed on the complete dataset with the confidence interval calculated from the leave-one-out residuals[41].

Analysis of the data is also performed using the SHapley Additive exPlanations (SHAP) library, version 0.40.0)[42] that supplies algorithms for interpretable AI. The library uses a game theory approach initially proposed by Lloyd Shapely[43] and developed by Lundberg *et al.*[44,45] as a Python library. The algorithms in this library allow the evaluation of individual variables and their interactions on the global results as well as individual experiments.

## Results and discussion

Experiments were performed in batch reactors and in a continuous reactor on the same resource at similar conditions. Product yields and oil analyses are presented here. We report in this section the effect of several process parameters such as the temperature and holding time. Experiments in a batch

reactor also served to screen more resources and a wider variety of conditions to obtain a larger picture of the transformation. All data underlying the results are available in *Data availability*.

### Continuous and batch experiments

Batch experiments are performed with the feedstocks presented in *Resource* section. Batch experiments were performed at different holding times, 0 and 30 minutes. To complete the data, experiments with a holding time of 20 minutes were performed in the batch reactor to allow comparison with the continuous reactor. The results of the batch and the average the continuous experiments are presented in Table 3. Typical run times in this project were 8 to 20 hours. About 12 kg of biocrude was produced in eight experiments. Precise mass balances have been somewhat difficult to establish in the continuous experiments. The carbon balance closes to 96 ± 12 % for the continuous experiments. The averages of the experiments are presented in Table 3. The resources FFOM and DFOR produce a lot of char and are not very interesting for HTL. The overall oil yield from FW2CO is somewhat lower for the continuous experiments, but the dispersion between the runs is also higher. The char production is lower in the continuous experiments. The ratio of oil to char is higher for continuous experiments.

The gas compositions from both batch and continuous experiments on FW2 are presented in Table 4. The gas produced is mainly carbon dioxide with traces of hydrogen and carbon monoxide. There are small differences in the gas quality. It is difficult to estimate a precision of the gas yield in the continuous experiments, as a minor leak is always possible. Reported here is the variability between the experiments without pretending to be an accuracy.

Briand[20] has shown that the residence time distribution is relatively flat, the reactor can be simulated by the equivalent of two or three continuously stirred ideal reactors. This means that part of the resources leave the reactor after a short while another faction stays for a longer time. The continuous reactor is at operating temperature and the injected resource is heated quickly. It is possible that high heating rates are more efficient in the depolymerisation of the biomass avoiding the production of primary char by slow pyrolysis of the resource, but no proof of this can be found.

### Bio-oil analysis

The biocrude was separated in bio-oil and char by solvent extraction with ethyl acetate. Gas chromatography couples with mass spectrometry (GC-MS) analysis was performed for the batch and continuous experiments. The results are presented in terms of areas and are not quantitative. Organic species in the aqueous phase are either oxygenates like alcohol, ketones, cyclic ethers, phenolic species or N species pyrazine and derivatives. The chromatograms and a full list of species are presented in the *Data availability* section. Figure 6 presents the families of molecules that can be identified in the bio-oil from the continuous experiments, with their relative peak areas. More GC-MS results and data can be found in the supplemental material[36]. The total acid number of this oil is high, 280 ± 25 mg KOH/g oil.

The Gel Permeation Chromatography results in Table 5, show that the bio-oil produced from batch experiments is lighter,

**Table 3. Results from the continuous experiments on FW2.** DFOR, digested food wastes; FFOM, The organic fraction of municipal solid waste; FW, food waste.

| Experiment | Resource | Conditions | Oil Yield (wt.%) | Char Yield (wt.%) | Oil to Char Ratio | Gas yield (wt.%) |
|---|---|---|---|---|---|---|
| Batch | DFOR | 300 °C, 30 min | 11.4 ± 0.8 | 47 ± 1 | 0.22 ± 0.2 | 14 ± 1 |
| Batch | FFOM | 300 °C, 30 min | 11 ± 2 | 31 ± 2 | 0.4 ± 0.1 | 13.2 ± 0.1 |
| Batch | FW2 | 300 °C, 30 min | 44 ± 2 | 23 ± 2 | 1.9 ± 0.1 | 15 ± 1 |
| Batch | FW2CO | 300 °C, 20 min | 40 ± 4 | 18 ± 0.6 | 2.1 ± 0.2 | 14.6 ± 0.1 |
| Continuous | FW2CO | 300 °C, 20 min | 35 ± 8 | 13.2 ± 3 | 2.8 ± 0.5 | 10.3 ± 2 |

**Table 4. Gas composition from typical experiments with food waste enriched with used cooking oil FW2CO.**

| Experiment | Resource | $CO_2$ | $H_2$ | CO | $CH_4$ | $C_2H_x$ | $C_3H_x$ | $H_2S$ |
|---|---|---|---|---|---|---|---|---|
| | | % vol | % vol | % vol | % vol | % vol | % vol | % vol |
| Batch | FW2CO | 89.9 | 0.8 | 7.9 | 0.8 | 0.22 | 0.28 | 0.02 |
| Continuous | FW2CO | 85.64 | 1.17 | 2.34 | 0.90 | 0.06 | 0.17 | 0.004 |

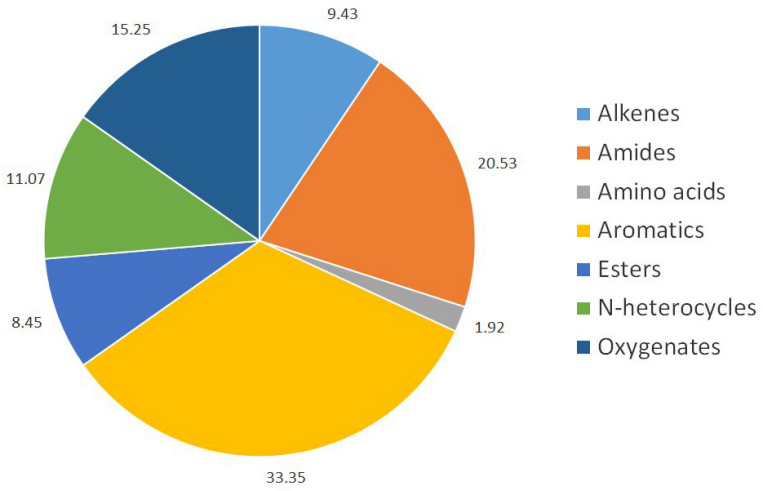



**Figure 6. Different chemical families in extracted air-dried bio-oil from the continuous experiments with *food waste enriched with used cooking oil* (FW2CO).**

**Table 5. Molecular weights batch and continuous experiments.** The results are presented averaged by number (Mn) by weight (Mw) and the peak molecular weight (Mp).

| Experiment | Resource | Conditions | Mn (Da) | Mw (Da) | Mp (Da) |
|---|---|---|---|---|---|
| Batch | FW2 | 300 °C, 30 min | 382 ± 1 | 698 ± 36 | 426 ± 10 |
| Batch | FW2CO | 300 °C, 20 min | 475 ± 3 | 745 ± 30 | 512 ± 4 |
| Continuous | FW2CO | 300 °C, 20 min | 418 ± 51 | 1287 ± 800 | 431 ± 37 |

of a smaller molecular weight, than those produced from continuous experiments. The observations in Table 3 and section 3.2 that suggested the higher heating rates limited primary char production. Following up on this it appears that the wide residence time distribution in the continuous reactor limits the cracking of the bio-oil compounds and they remain heavier compared to the batch experiments.

## Regression of batch experiments

The resources have been tested in a wide variety of conditions in a batch reactor to evaluate their potential for hydrothermal liquefaction and to better understand their conversion. As there are many experiments with different resources, conditions and solvents, the analysis of the product yields is presented with the aid of machine learning tools. The full data set of the experimental data is presented in the *Underlying data*. The data includes the analysis of the resources, the experimental conditions and the yields.

Figure 7 presents the results on the food waste labelled FW2 at various temperatures and two different holding times, 0 and 30 minutes. The solid lines represent the results of the linear regression model. In this case, the composition is constant and the only variation are temperature and holding time. As we can see, the linear model does moderately well with the data. The $R^2$ is low, 0.46 for the oil yield and 0.57 for the char yield. Even though the data presents a relatively high dispersion, the general trend is picked up. The Equation 7 and Equation 8 describe the linearized behaviour.

$$\text{Oil yield} = -15.3 + 0.105 * \text{HoldingTime} + 0.12 * \text{Temperature} \quad \text{Eq. 7}$$

$$\text{Char yield} = 29.32 - 0.035 * \text{HoldingTime} - 0.026 * \text{Temperature} \quad \text{Eq. 8}$$

To have a meaningful model with this type of regression a lot of data is needed, more than any typical experimental campaign can produce. Extending the dataset by including more data from other resources and authors makes the task slightly more complex as compositions and other experimental conditions also play a role. It is no longer possible to plot the results as a function of one particular feature. When we repeat the linear regression for the oil and char yields with the extended data set with all of the data, completed with the literature, we obtain models with relatively low $R^2$ values (84 % for the training data and 73 % for the test data) as shown in Figure 8. This to be expected as a linear model is too simple as has been shown in the past[7]. A linear model does not do justice to the complexity of the problem.

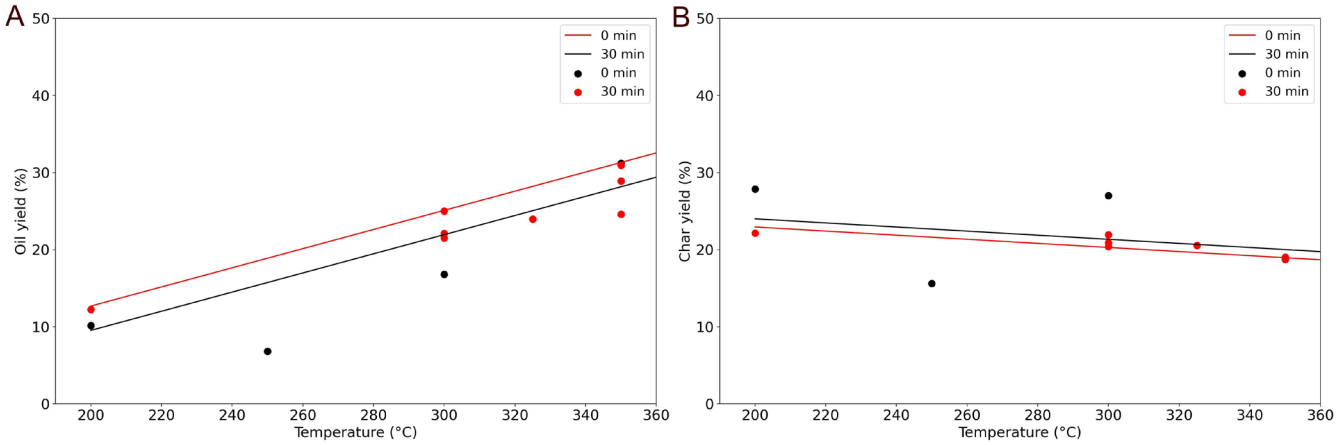



**Figure 7.** Experimental yields at different temperatures and holding times (oil yield **A**, char yield **B**).

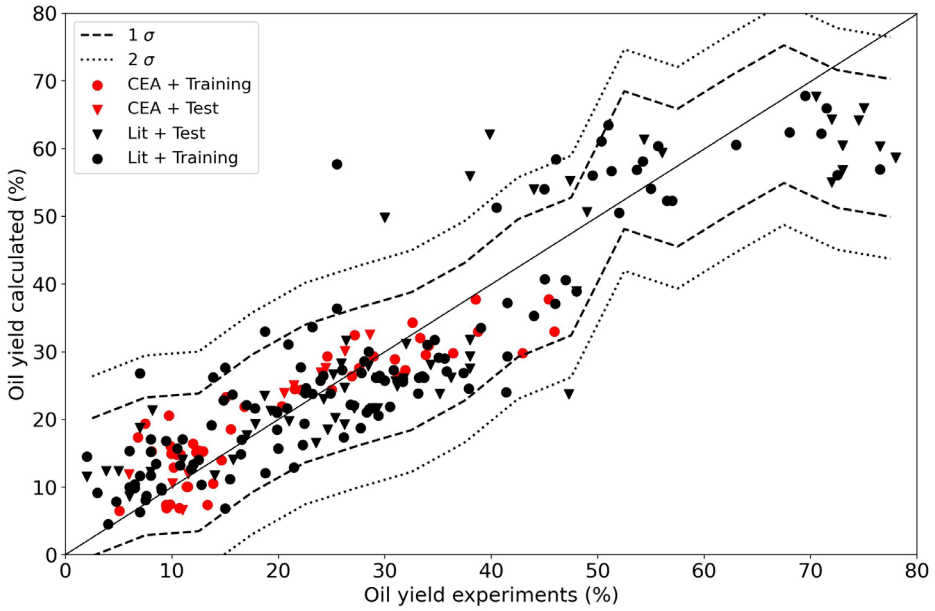



**Figure 8. Experimental yields compared to calclated yields using a linear regressor.** The data points from the current Waste2Road study are in red (CEA), the data from the literature are in black (Lit). Circles (●) denote training data (80% of the data); triangles (▼) denote test data (20% of the data).

The results for the random forest regressor are presented in Figure 9. The fit is obviously better with $R^2$ 98 % and 85 % for the training and test data respectively. The confidence interval shrank with the improved fit, 96 % of the data is in a ±5 % interval around the parity. The essence is that the data is not exact, and even though a machine learning algorithm can predict a result, it can only do so with a certain accuracy. The random forest algorithms does a better job than the linear regressor. The distribution of the data is presented in Figure 11. Table 3 shows that the dispersion in the data in this study is relatively high. No preference was given to the CEA data above literature data in the regression. Figure 9 suggests that the distribution in the CEA data is slightly lower than that of the literature data.

Figure 10 is the empirical cumulative distribution of the data set split into CEA and literature data. The x-axis is the Error, the deviation of the model from the experiment. The Proportion (also referred to as cumulative probability) represents the fraction of the data that lies below the value of the x-axis. The distribution plot in Figure 10 confirms the dispersion observed in Figure 9.

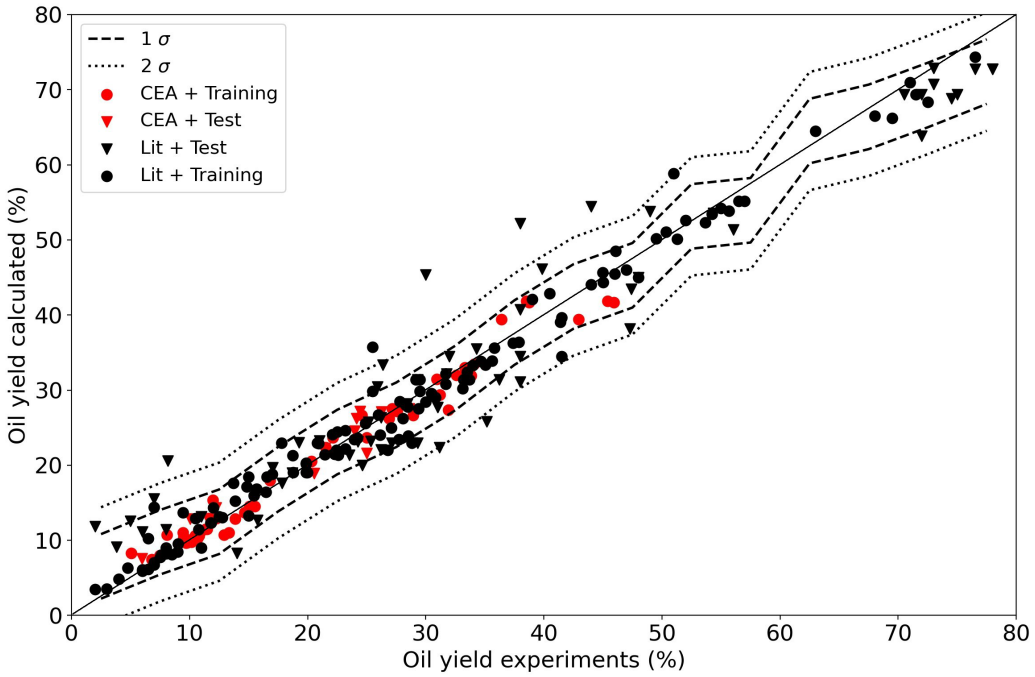



**Figure 9. Experimental yields compared to calculated oil yields using a random forest regressor.** The data points from the current Waste2Road study are in red (CEA), the data from the literature are in black (Lit). Circles (●) denote training data (80% of the data); triangles (▼) denote test data (20% of the data).

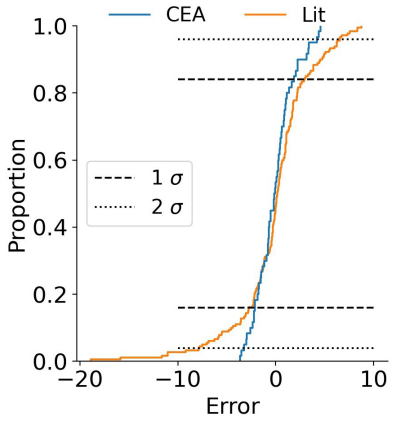



**Figure 10. Distribution plot with the random forest regressor of the data produced for this study (CEA) and the literature (Lit).**

From these modelling experiments we can conclude that combining datasets makes sense but the interpretation of this type of data should be subject to caution as there are many uncertainties in the characterisation of the experiments, non-quantified inputs as well as measurement error. Figure 11 presents the same data marked with the reaction temperature and the lipid content. The reaction temperature displays no obvious correlation with the oil yield presented in this form. The lipid content is however strongly correlated to the oil yield.

## Feature analysis

As we have seen, some features like lipid content are strongly correlated with the results; others display a much lower correlation as was shown for the temperature. The influence of each of the variables on the overall result of presented in Figure 12. As is expected, the resource composition, and in particular the lipid content plays a dominant role. Process parameters are less important. These results are valid for the food wastes included in this study, finally on a relatively small sample. Care should be taken to extrapolate these results to other studies and other resources. In a larger study, Li *et al.*[12] also found that the role of the lipid content was dominating.

The lipid content of the resource is overwhelmingly the most important factor in the HTL conversion in this dataset, before temperature and holding time. The SHAP library offers the possibility to go further in the analysis. The violin pot in Figure 13 shows the importance of features and *how* they influence the result. It can be seen that high lipid contents mostly yield higher than average oil yields. Very low lipid contents yield low oil yields. There is a zone with a moderately negative contribution to the oil yield with higher that averaged lipid yields, showing that there are interactions with other features. The holding time has clearly a positive effect for long times and a negative effect for short times. The analysis method appears to have a very low impact on the result. The ash content is also fairly neutral to the result, except for high values where it has a very negative effect.

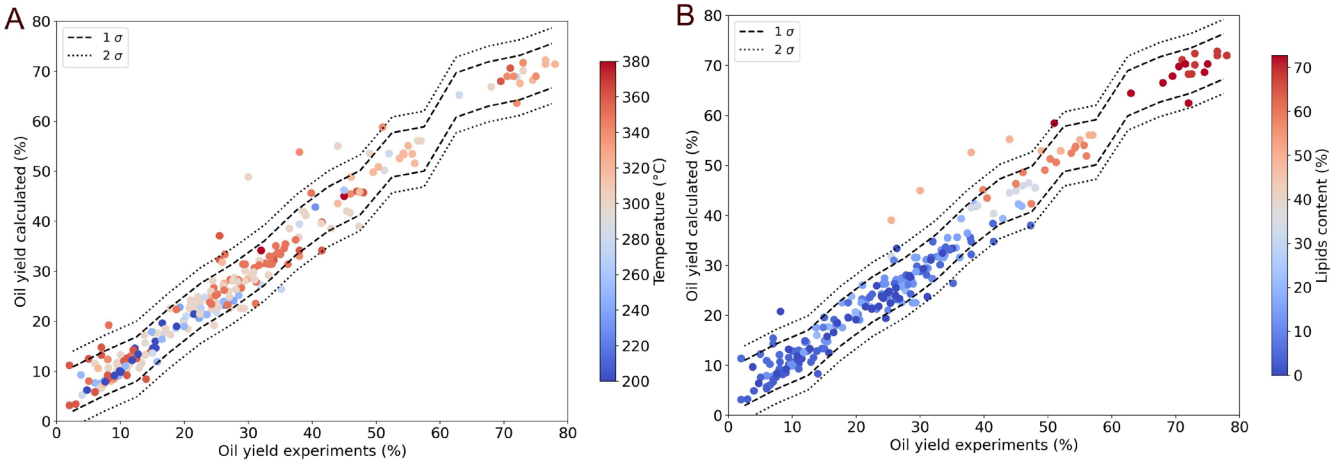



**Figure 11.** Modelling results (from Python machine learning model) for the oil yield using a random forest regressor with hues for the reactor temperature and lipid content of the resource (Temperature **A**, Lipid content **B**).

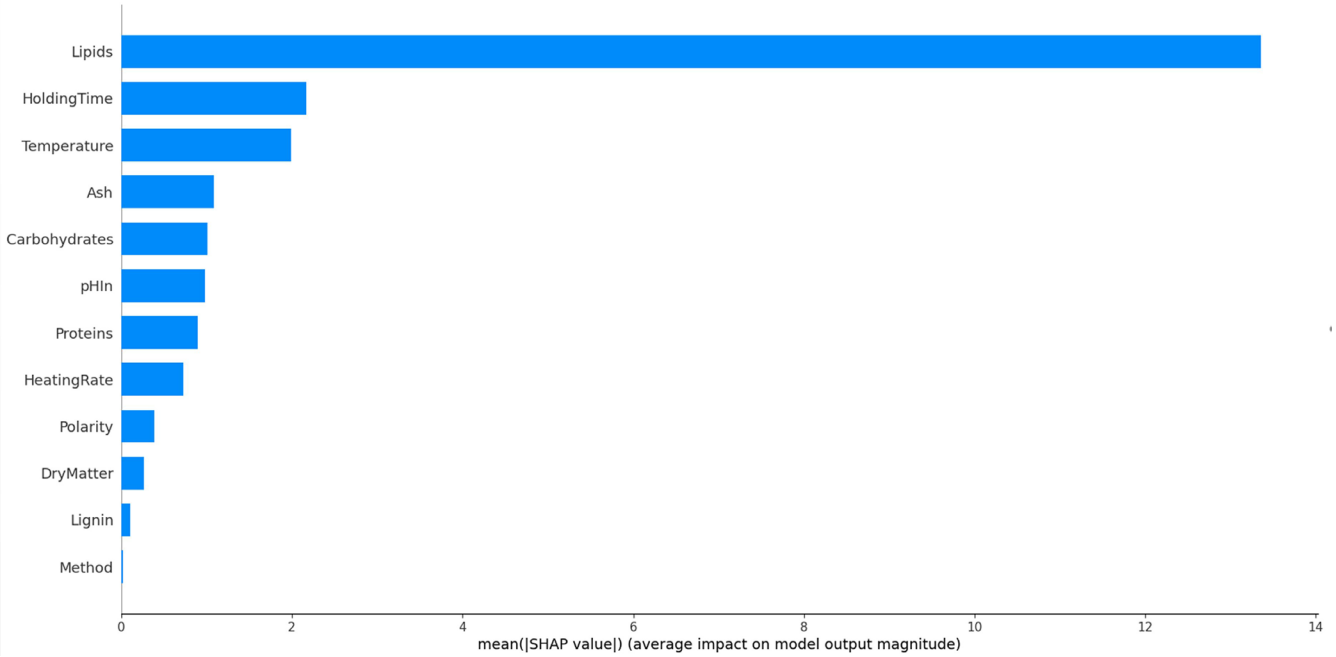



**Figure 12. Contributions of the experimental variables on the overall results for bio-oil.**

In Figure 13 we have seen how features influence the oil yield. For certain features the image is clear cut as was shown for the holding time. Other features show partial or complete multicolour surfaces, meaning that they do influence the result, but in collaboration with other features. This is especially true for the dry matter content. The lower the feature is ranked in importance the less pronounced the effects are, and subject to statistical noise.

SHAP (Shapely Additive exPlanations as described in the section *Data analysis with machine learning algorithms*) allows us to go deeper in the analysis. Figure 14 shows the dependency plot for proteins (A) and temperature (B). The protein strongly influences the oil yield. Low protein content resources (0–10%) show a positive SHAP value, meaning that this property contributes positively to the oil yield. The SHAP value then drops to negative values for intermediate values to terminate at around zero for high protein content. The method does not propose an explanation why this may be the case. The graph also shows that low protein content also corresponds to high carbohydrate content. While this may be obvious in itself, it can help correlate features. Figure 14B shows us that

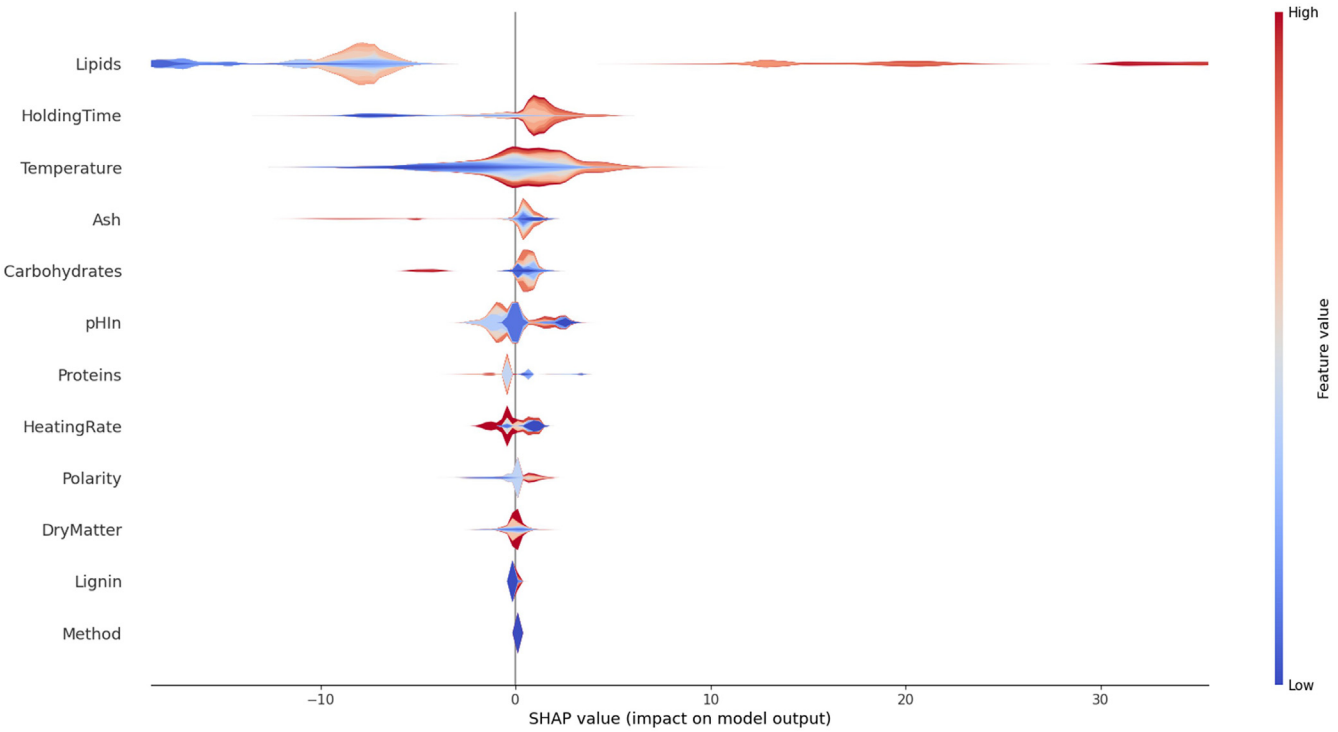



**Figure 13. Violin plot of contributions of the experimental variables on the overall results for bio-oil with more details.**

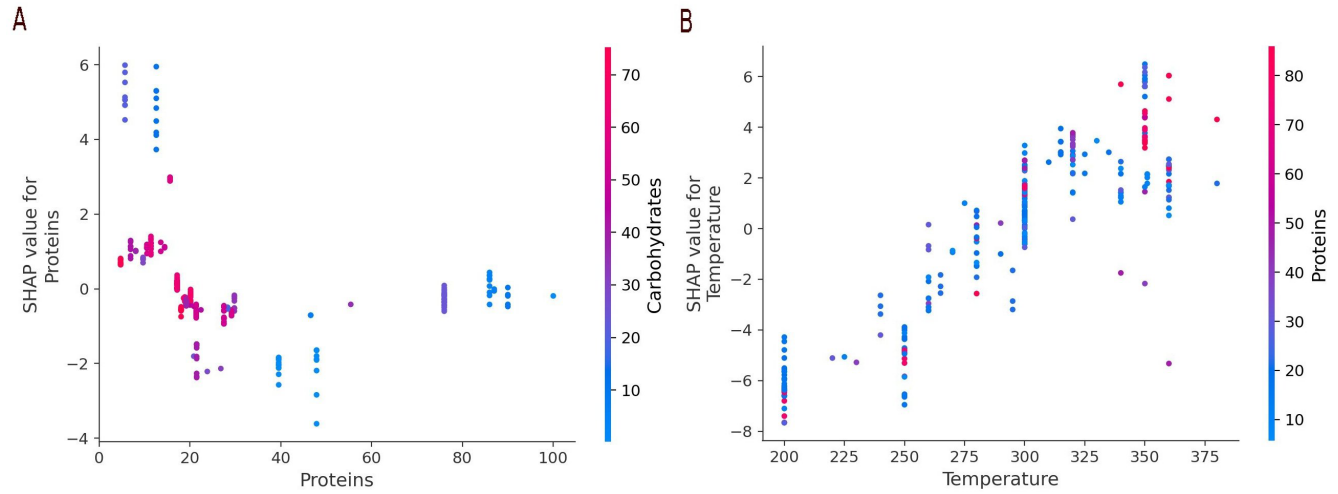



**Figure 14.** SHAP dependency plots for proteins mass fraction (**A**) and temperature in °C (**B**).

there is a group of carbohydrate rich, low protein resources that do produce more than average oil. Another group of high lipid, low protein resources do produce much oil (Figure not shown). Figure 14B shows the dependence plot for the temperature. The oil yield is globally proportional to the temperature. The colouring in the graph shows that a relatively high proportion of the data at high temperatures concerns high protein resources.

## Conclusions

Food wastes are an interesting resource rich in lipids and proteins. At temperatures above 300 °C they produce a sticky to fluid biocrude with an interesting yield of bio-oil, even at low holding times. Lower temperatures favour bio-char regardless of the holding time. The mass yield of the bio-oil obtained is mostly around 40 %. Ash rich resources as digested food wastes (DFOR) or organic fractions of municipal waste (FFOM) from

mechanical sorting produce low bio-oil yields and favour char formation.

Continuous experiments have shown that the yields are comparable to batch experiments. The oil to char ratio is an interesting quantity to compare batch and continuous reactor products. The continuous reactor yields an oil to char ratio of nearly three while batch experiments rarely produce a ratio much more than two. The bio-oil from continuous experiments present a higher mass averaged molecular weight (Mw). The higher heating rate may contribute to the higher oil production and oil to char ratio. Mixing in the reactor, leading to a relatively wide residence time distribution, may lead to shorter reaction times for part of the molecules, leading to a higher averaged molecular weight.

Batch experiments remain a useful tool in the comprehension of hydrothermal liquefaction. Numerous studies in the past have shown that the results from HTL experiments can be described by correlations obtained after carefully designed experimental plans. Data modelling with machine learning tools allow us to establish predictive models with confidence intervals from unstructured data. Experimental data can be enriched with external studies that contribute to the modelling results and increases the accuracy and the universality.

There are many variables that play a role in hydrothermal liquefaction; these include resource composition, process conditions but also product treatment and analysis. Biomass resources are very complex and an analysis in terms of carbohydrates, proteins, lipids and ash does not do justice to its complexity. It is however an analysis that can be easily performed on all resources. Any single study cannot hope to cover all these parameters in a meaningful way. An extremely high accuracy with any modelling tool or dataset cannot be achieved as biomass and the HTL chemistry is too complex. Combining datasets can be an interesting approach to draw more meaningful conclusions. For this to be possible, authors should take care to fully document their experiments, together with a full resource analysis.

## Ethics and consent

Ethical approval and consent were not required.

## Data availability

### Underlying data

Zenodo: Bio-oil production from biogenic wastes, the hydrothermal conversion step – Data. https://doi.org/10.5281/zenodo.6940211[36].

This project includes the following underlying data:

- HTLYieldData – Publi.xlsx (yield data from the experiments and the literature. Each experiment is labelled to find the corresponding analyser data).
- Bio-oil production from biogenic wastes, the hydrothermal conversion step - Supp.docx (data from the gas chromatograph with identification of the molecules by mass spectrometry).
- GCMS.7z (the raw chromatography data created with the program Tubomass 5.4.2 from Perkin Elmer).
- GPC.7z (raw data from the gel permeatography created with the program Omnisec 5.12.467 from Malvern).
- µGC Batch.7z (gas analysis of the experiments - Microsoft Excel files).
- µGC Cont.7z (gas analysis of the continuous experiments - Microsoft Excel files).

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

### Extended data

Archived analysis code at time of publication: https://doi.org/10.5281/zenodo.6940211

License: CC-BY 4.0

### Acknowledgements

The authors thank Bruno Lacaze for his help in the experimental work and Maria Marotta for her help with the product analyses.

## References

1. Goudrian F, van de Beld B, Boerefijn FR, *et al.*: **Thermal Efficiency of the HTU® Process for Biomass Liquefaction**. Blackwell Science Ltd, Oxford, 2001.
**Publisher Full Text**
2. Goudriaan F, Peferoen DGR: **Liquid fuels from biomass via a hydrothermal process.** *Chem Eng Sci.* 1990; **45**(8): 2729–2734.
**Publisher Full Text**
3. Elliott DC, Schiefelbein GF: **Liquid Hydrocarbon fuels from biomass**. In: *Division of Fuel Chemistry Annual Meeting Preprints.* American Chemical Society, 1989; 1160–1166.
**Reference Source**
4. Déniel M, Haarlemmer G, Roubaud A, *et al.*: **Hydrothermal liquefaction of blackcurrant pomace and model molecules: understanding of reaction mechanisms.** *Sustain Energy Fuels.* 2017; **1**(3): 555–582.
**Publisher Full Text**
5. Snowden-Swan L, Li S, Jiang Y, *et al.*: **Wet Waste Hydrothermal Liquefaction and Biocrude Upgrading to Hydrocarbon Fuels: 2021 State of Technology**. PNNL, Richland, Washington, 2022.
**Publisher Full Text**
6. Briand M, Haarlemmer G, Roubaud A, *et al.*: **Evaluation of the Heat Produced by the Hydrothermal Liquefaction of Wet Food Processing Residues and Model Compounds.** *ChemEngineering.* 2022; **6**(1): 2.
**Publisher Full Text**
7. Déniel M, Haarlemmer G, Roubaud A, *et al.*: **Modelling and Predictive Study of Hydrothermal Liquefaction: Application to Food Processing Residues.**

*Waste Biomass Valor.* 2017; **8**(6): 2087–2107.
**Publisher Full Text**

8. Haarlemmer G, Guizani C, Anouti S, *et al.*: **Analysis and comparison of bio-oils obtained by hydrothermal liquefaction and fast pyrolysis of beech wood.** *Fuel.* 2016; **174**: 180–188.
**Publisher Full Text**

9. Déniel M, Haarlemmer G, Roubaud A, *et al.*: **Bio-oil Production from Food Processing Residues: Improving the Bio-oil Yield and Quality by Aqueous Phase Recycle in Hydrothermal Liquefaction of Blackcurrant (*Ribes nigrum* L.) Pomace.** *Energy Fuels.* 2016; **30**(6)**:** 4895–4904.
**Publisher Full Text**

10. Biller P, Madsen RB, Klemmer M, *et al.*: **Effect of hydrothermal liquefaction aqueous phase recycling on bio-crude yields and composition.** *Bioresour Technol.* 2016; **220**: 190–199.
**PubMed Abstract** | **Publisher Full Text**

11. Katongtung T, Onsree T, Tippayawong N: **Machine learning prediction of biocrude yields and higher heating values from hydrothermal liquefaction of wet biomass and wastes.** *Bioresour Technol.* 2022; **344**(Pt B): 126278.
**PubMed Abstract** | **Publisher Full Text**

12. Li J, Zhang W, Liu T, *et al.*: **Machine learning aided bio-oil production with high energy recovery and low nitrogen content from hydrothermal liquefaction of biomass with experiment verification.** *J Chem Eng.* 2021; **425**: 130649.
**Publisher Full Text**

13. Zhang W, Li J, Liu T, *et al.*: **Machine learning prediction and optimization of bio-oil production from hydrothermal liquefaction of algae.** *Bioresour Technol.* 2021; **342**: 126011.
**PubMed Abstract** | **Publisher Full Text**

14. Wolpert DH, Macready WG: **No Free Lunch Theorems for Optimization.** *IEEE Trans Evol Comput.* 1997; **1**(1): 67–82.
**Publisher Full Text**

15. Onsree T, Tippayawong N, Phithakkitnukoon S, *et al.*: **Interpretable machine-learning model with a collaborative game approach to predict yields and higher heating value of torrefied biomass.** *Energy.* 2022; **249**: 123676.
**Publisher Full Text**

16. **Arrêté du 8 septembre 1977 relatif aux méthodes officielles d'analyse des produits diététiques et de régime.** *Journal Officiel de la république Française.* 2016.
**Reference Source**

17. ISO16472: **Animal feeding stuffs - Determination of amylase-treated neutral detergent fibre content (aNDF)**. 2006.
**Reference Source**

18. ISO13906: **Animal feeding stuffs — Determination of acid detergent fibre (ADF) and acid detergent lignin (ADL) contents.** ISO, 2008.
**Reference Source**

19. ISO1871: **Food and feed products — General guidelines for the determination of nitrogen by the Kjeldahl method**. 2009.
**Reference Source**

20. Briand M: **Conception et évaluation d'un procédé de liquéfaction hydrothermale en vue de la valorisation énergétique de résidus agroalimentaires.** L'Université Claude Bernard Lyon1, Grenoble, 2021.
**Reference Source**

21. Wang Y, Zhang Y, Liu Z: **Effect of Aging in Nitrogen and Air on the Properties of Biocrude Produced by Hydrothermal Liquefaction of *Spirulina*.** *Energy Fuels.* 2019; **33**(10): 9870–9878.
**Publisher Full Text**

22. Zhou D, Zhang L, Zhang S, *et al.*: **Hydrothermal Liquefaction of Macroalgae *Enteromorpha prolifera* to Bio-oil.** *Energy Fuels.* 2010; **24**(7): 4054–4061.
**Publisher Full Text**

23. Zhang B, von Keitz M, Valentas K: **Thermochemical liquefaction of high-diversity grassland perennials.** *J Anal Appl Pyrolysis.* 2009; **84**(1): 18–24.
**Publisher Full Text**

24. Anouti S, Haarlemmer G, Déniel M, *et al.*: **Analysis of Physicochemical Properties of Bio-Oil from Hydrothermal Liquefaction of Blackcurrant Pomace.** *Energy Fuels.* 2016; **30**(1): 398–406.
**Publisher Full Text**

25. Motavaf B, Savage PE: **Effect of Process Variables on Food Waste Valorization via Hydrothermal Liquefaction.** *ACS EST Engg.* 2021; **1**(3): 363–374.
**Publisher Full Text**

26. Bayat H, Cheng F, Dehghanizadeh M, *et al.*: **Hydrothermal Liquefaction of Food Waste: Bio-crude oil Characterization, Mass and Energy Balance.** *American Society of Agricultural and Biological Engineers.* St. Joseph, MI, 2019; 1.
**Publisher Full Text**

27. Aierzhati A, Stablein MJ, Wu NE, *et al.*: **Experimental and model enhancement of food waste hydrothermal liquefaction with combined effects of biochemical composition and reaction conditions.** *Bioresour Technol.* 2019; **284**: 139–147.
**PubMed Abstract** | **Publisher Full Text**

28. Evcil T, Tekin K, Ucar S, *et al.*: **Hydrothermal liquefaction of olive oil residues.** *Sustain Chem Pharm.* 2021; **22**: 100476.
**Publisher Full Text**

29. Yang C, Wang S, Ren M, *et al.*: **Hydrothermal Liquefaction of an Animal Carcass for Biocrude Oil.** *Energy Fuels.* 2019; **33**(11): 11302–11309.
**Publisher Full Text**

30. Yang L, He QS, Havard P, *et al.*: **Co-liquefaction of spent coffee grounds and lignocellulosic feedstocks.** *Bioresour Technol.* 2017; **237**: 108–121.
**PubMed Abstract** | **Publisher Full Text**

31. Yang L, Nazari L, Yuan Z, *et al.*: **Hydrothermal liquefaction of spent coffee grounds in water medium for bio-oil production.** *Biomass Bioenergy.* 2016; **86**: 191–198.
**Publisher Full Text**

32. Biller P, Ross AB: **Potential yields and properties of oil from the hydrothermal liquefaction of microalgae with different biochemical content.** *Bioresour Technol.* 2011; **102**(1): 215–225.
**PubMed Abstract** | **Publisher Full Text**

33. Sheng L, Wang X, Yang X: **Prediction model of biocrude yield and nitrogen heterocyclic compounds analysis by hydrothermal liquefaction of microalgae with model compounds.** *Bioresour Technol.* 2018; **247**: 14–20.
**PubMed Abstract** | **Publisher Full Text**

34. Teri G, Luo L, Savage PE: **Hydrothermal Treatment of Protein, Polysaccharide, and Lipids Alone and in Mixtures.** *Energy Fuels.* 2014; **28**(12): 7501–7509.
**Publisher Full Text**

35. Luo L, Sheehan JD, Dai L, *et al.*: **Products and Kinetics for Isothermal Hydrothermal Liquefaction of Soy Protein Concentrate.** *ACS Sustainable Chem Eng.* 2016; **4**(5): 2725–2733.
**Publisher Full Text**

36. Haarlemmer G, Roubaud A, Lacaze B: **Bio-oil production from biogenic wastes, the hydrothermal conversion step - Data.** *Zenodo*. 2022.
**http://www.doi.org/10.5281/zenodo.6940211**

37. Larson EJ: **The myth of artificial intelligence: why computers can't think the way we do.** The Belknap Press of Harvard University Press, Cambridge, Massachusetts, 2021.
**Reference Source**

38. **SciKit-Learn**. 2022.
**Reference Source**

39. **MAPIE - Model Agnostic Prediction Interval Estimator**. 2022.
**Reference Source**

40. Kim B, Xu C, Barber RF: **Predictive Inference Is Free with the Jackknife+-after-Bootstrap.** 2020.
**Publisher Full Text**

41. Barber RF, Candes EJ, Ramdas A, *et al.*: **Predictive inference with the jackknife+.** *Ann Stat.* 2021; **49**: 486–507.
**Publisher Full Text**

42. **SHAP (SHapley Additive exPlanations)**.
**Reference Source**

43. Shapley LS: **A Value for N-Person Games.** The RAND Coorporation, Santa Monica California, 1952.
**Publisher Full Text**

44. Lundberg SM, Erion G, Chen H, *et al.*: **Explainable AI for Trees: From Local Explanations to Global Understanding.** *ArXiv.* 2019.
**Publisher Full Text**

45. Lundberg SM, Lee SI: **A Unified Approach to Interpreting Model Predictions.** In: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.) *Advances in Neural Information Processing Systems 30.* Curran Associates, Inc., 2017; 4765–4774.
**Reference Source**

# Open Peer Review

## Current Peer Review Status: ✔ ✔

**Version 2**

Reviewer Report 04 January 2023

https://doi.org/10.21956/openreseurope.16647.r30492

© **2023 Zhang Y.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Yuanhui Zhang**
[1] Department of Agricultural and Biological Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[2] Department of Mechanical Science Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

The authors have made sufficient revision and rational discussion addressing previous review comments. The manuscript could provide useful information in the area of HTL of food waste. It is recommended to be accepted for indexing.

***Competing Interests:*** No competing interests were disclosed.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 04 January 2023

https://doi.org/10.21956/openreseurope.16647.r30493

© **2023 Castello D.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Daniele Castello**
AAU Energy, Aalborg University, Aalborg, Denmark

I have read the authors's replies to my previous reviews and found all answers adequate.
I would therefore advise the indexing of the manuscript in its revised form.

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Biomass, hydrothermal liquefaction, hydroprocessing, upgrading to fuels, drop-in fuels

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Version 1**

Reviewer Report 20 October 2022

https://doi.org/10.21956/openreseurope.16118.r30148

© **2022 Zhang Y.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

**Yuanhui Zhang**
[1] Department of Agricultural and Biological Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[2] Department of Mechanical Science Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[3] Department of Agricultural and Biological Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[4] Department of Mechanical Science Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

This manuscript reports an experimental and analytical/modeling study of bio-oil production from food waste via hydrothermal liquefaction (HTL). The experiment was well designed and new experimental data were obtained. An innovative part of the paper is using SHAP (a software) analysis to model the HTL bio-oil production affected by operating conditions (reaction temperature and holding time) and biochemical compositions of food waste. The SHAP model includes 12 variables which is very comprehensive and intriguing. The results show good agreement of the experiment data and the SHAP analysis. Given the complexity of the chemical reactions in HTL, most of HTL studies did cover such a comprehensive range of variable. To make the SHAP analysis more convincing, it is suggested by this reviewer to include more data, especially those of related to the variables affecting the HTL performance.

The SHAP database could be further expanded to other feedstocks similar to food waste such as manure and algae, with similar biochemical composition (category-wise) for HTL. There are a wealth of lab-scale HTL experimental studies, and a few pilot-scale HTL studies in the past five years, which were largely missed in the current version of the manuscript. Works of a few groups with sustained HTL research could help to strengthen the SHAP analysis – those groups including my own lab at University of Illinois at Urbana-Champaign, Savage's lab in Penn State University,

and so on. I listed some examples of publications related to the HTL work from my own lab (for my own simplicity), to show there are a broad range of data and work related to evaluate the HTL performance, which by no means is required or exclusive.

Listed below are some example references for HTL data collection based on my own work, but the authors should engage more broadly with the literature.

**References**

1. Chen W, Qian W, Zhang Y, Mazur Z, et al.: Effect of ash on hydrothermal liquefaction of high-ash content algal biomass. *Algal Research*. 2017; **25**: 297-306 Publisher Full Text
2. Chen WT, Ma J, Zhang Y, Gai C, et al.: Physical pretreatments of wastewater algae to reduce ash content and improve thermal decomposition characteristics.*Bioresour Technol*. 2014; **169**: 816-820 PubMed Abstract | Publisher Full Text
3. Qiu Y, Aierzhati A, Cheng J, Guo H, et al.: Biocrude Oil Production through the Maillard Reaction between Leucine and Glucose during Hydrothermal Liquefaction. *Energy & Fuels*. 2019; **33** (9): 8758-8765 Publisher Full Text
4. Lu J, Li H, Zhang Y, Liu Z: Nitrogen Migration and Transformation during Hydrothermal Liquefaction of Livestock Manures. *ACS Sustainable Chemistry & Engineering*. 2018; **6** (10): 13570-13578 Publisher Full Text
5. Chen W, Tang L, Qian W, Scheppe K, et al.: Extract Nitrogen-Containing Compounds in Biocrude Oil Converted from Wet Biowaste via Hydrothermal Liquefaction. *ACS Sustainable Chemistry & Engineering*. 2016; **4** (4): 2182-2190 Publisher Full Text
6. Zhang Y, Chen W: Hydrothermal liquefaction of protein-containing feedstocks. 2018. 127-168 Publisher Full Text
7. Watson J, Lu J, de Souza R, Si B, et al.: Effects of the extraction solvents in hydrothermal liquefaction processes: Biocrude oil quality and energy conversion efficiency. *Energy*. 2019; **167**: 189-197 Publisher Full Text
8. Li H, Lu J, Zhang Y, Liu Z: Hydrothermal liquefaction of typical livestock manures in China: Biocrude oil production and migration of heavy metals. *Journal of Analytical and Applied Pyrolysis*. 2018; **135**: 133-140 Publisher Full Text
9. Aierzhati A, Watson J, Si B, Stablein M, et al.: Development of a mobile, pilot scale hydrothermal liquefaction reactor: Food waste conversion product analysis and techno-economic assessment. *Energy Conversion and Management: X*. 2021; **10**. Publisher Full Text
10. Li H, Zhu Z, Lu J, Watson J, et al.: Establishment and performance of a plug-flow continuous hydrothermal reactor for biocrude oil production. *Fuel*. 2020; **280**. Publisher Full Text
11. Zhu Z, Liu Z, Zhang Y, Li B, et al.: Recovery of reducing sugars and volatile fatty acids from cornstalk at different hydrothermal treatment severity.*Bioresour Technol*. 2016; **199**: 220-227 PubMed Abstract | Publisher Full Text
12. Carpio R, Zhang Y, Kuo C, Chen W, et al.: Effects of reaction temperature and reaction time on the hydrothermal liquefaction of demineralized wastewater algal biomass. *Bioresource Technology Reports*. 2021; **14**. Publisher Full Text

**Is the work clearly and accurately presented and does it cite the current literature?**

Partly

**Is the study design appropriate and does the work have academic merit?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**
Partly

**Are all the source data underlying the results available to ensure full reproducibility?**
Partly

**Are the conclusions drawn adequately supported by the results?**
Partly

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** hydrothermal liquefaction of wet waste.  waste-to-energy.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 28 Nov 2022

**Geert Haarlemmer**

*Dear Reviewer,*
*Thank you for taking the time to read and review our paper. We have made significant improvements to the paper. Please find responses to your remarks below.*

This manuscript reports an experimental and analytical/modeling study of bio-oil production from food waste via hydrothermal liquefaction (HTL). The experiment was well designed and new experimental data were obtained. An innovative part of the paper is using SHAP (a software) analysis to model the HTL bio-oil production affected by operating conditions (reaction temperature and holding time) and biochemical compositions of food waste. The SHAP model includes 12 variables which is very comprehensive and intriguing. The results show good agreement of the experiment data and the SHAP analysis. Given the complexity of the chemical reactions in HTL, most of HTL studies did cover such a comprehensive range of variable. To make the SHAP analysis more convincing, it is suggested by this reviewer to include more data, especially those of related to the variables affecting the HTL performance.

*Response: Variables affecting the HTL performance are mostly included in this study, the main being composition, temperature and holding time. Other less important parameters such as heating rate and dilution are also included. We were surprised to see that finally the solvent and extraction order have a limited effect on the results.*

The SHAP database could be further expanded to other feedstocks similar to food waste

such as manure and algae, with similar biochemical composition (category-wise) for HTL. There are a wealth of lab-scale HTL experimental studies, and a few pilot-scale HTL studies in the past five years, which were largely missed in the current version of the manuscript. Works of a few groups with sustained HTL research could help to strengthen the SHAP analysis – those groups including my own lab at University of Illinois at Urbana-Champaign, Savage's lab in Penn State University, and so on. I listed some examples of publications related to the HTL work from my own lab (for my own simplicity), to show there are a broad range of data and work related to evaluate the HTL performance, which by no means is required or exclusive.

*Response: Yes, point taken, the more data the better. In this paper we want to limit ourselves to a particular type of feedstock that does not receive much attention. There are many HTL papers on algae. Microalgae are however quite different from food wastes in terms of composition and physical aspect. We prefer to stick to food wastes and they are a relevant waste that does not receive much attention in the literature.*

Listed below are some example references for HTL data collection based on my own work, but the authors should engage more broadly with the literature.

*Response: Thank you for the references. There are some that I did not know.*

***Competing Interests:*** No competing interests

Reviewer Report 17 October 2022

https://doi.org/10.21956/openreseurope.16118.r30147

**© 2022 Castello D.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

**Daniele Castello**
[1] AAU Energy, Aalborg University, Aalborg, Denmark
[2] AAU Energy, Aalborg University, Aalborg, Denmark

The present manuscript reports a comprehensive analysis of HTL biocrude production from a number of organic waste, including fresh food waste, organic fraction of municipal solid waste and fermented digestate.

The manuscript is in general well-written and presents a number of interesting results. However, there are some parts that need to be presented in a much more concise fashion, while some others are excessively condensed and lack the necessary space to be fully developed. The manuscript is indeed trying to summarize several topics, i.e. the differences between batch and continuous processes, the effect of aqueous phase recirculation and finally a machine learning

model for HTL built from a large dataset. Probably it is too much information for a single manuscript.

1. An important issue is the study on aqueous phase recirculation. In my opinion, this is a not well-placed research question. Indeed, the kind of organic waste taken into account are by their nature wet (potentially even up to 70-80%). In these conditions, aqueous phase recirculation cannot be established, because it would bring to accumulate water in the system. On the other hand, a crucial advantage of HTL over other thermochemical technologies is that the feedstock does not need pre-drying. The authors could explore aqueous phase recirculation because their feedstock was dried beforehand, but this does not correspond to what anyone would do in practice. The authors therefore need to better explain the motivation behind aqueous phase recirculation and how it could be realistically established in a real plant.

2. Regarding the experimental procedure, there are some weak points. How could the biocrude remaining in the reactor be evaluated just by weight difference? Is it assumed that there are no residual solids in the reactor? This operation is normally achieved by washing the internal part of the reactor with a solvent and then separating and quantifying the products.

3. How was the amount of produced gas quantified?

4. What was the original moisture content of the feed, before pre-drying?

5. “Biocrude with high char contents could be considered a product from hydrothermal carbonization” (p. 6). This definition is quite weak and I suggest removing or rephrasing, as there are no fixed boundaries between different processes and, especially, the utilization of the products depends on the solutions utilized on the production plant and not merely by their amounts. For example, scaling up the process to a continuous flow unit could feature some form of on-line filtration allowing to recover important amounts of biocrude.

6. The section “Product analyses” tends to be redundant with respect to the section “Product recovery batch experiments”. For example, Karl-Fischer titration is described again. Moreover, azeotropic distillation is not defined. There seem to be two almost similar TOC apparatuses: please have it checked.

7. “A numerical value is attributed to the extraction method (…)”: this is not very clear, especially it is not clear how this value affects the regression algorithm.

8. The section “Data analysis with machine learning algorithms” seems quite didactical, even with some funny anecdote, like the story of the turkey. I would advise more conciseness, reporting only information necessary to an average-skilled reader to understand and reproduce the methodology.

9. In Fig. 5 caption, it should be specified that carbohydrates include lignin.

10. The first part of the section “Continuous and batch experiments” belongs to Materials and Methods and not to Results.

11. Units in Table 3 are wrong. For example, oil to char ratio is not a percentage, while char and gas yields in the last row are. The unit for yields should be made explicit (wt. %). A carbon balance to verify closure is highly advised.

12. One of the differences between batch and continuous is the long-lasting thermal transient in batch operations, which is not present in continuous.

13. Fig. 6 is clearly wrong, as it is identical to Fig. 4. It should report the GC-MS analysis, instead.

14. The caption in Fig. 7 is insufficient. Please explain what the left (a) and right (b) plots represent.

15. Eq. 3 and 4 look misleading. There should be a precise indication of the validity limit, especially for residence time. Moreover, it is awkward that a dependence is stated for residence time, when only two values of this variable were tested.

16. At page 13, the paragraph: “From table 3 one could conclude (...) makes is even clear” is not clear and needs to be rephrased.

17. The legend in Fig. 8 is unclear. Its caption should report (a) and (b) instead of “left” and “right” (same also in Fig. 12).

18. Fig. 9 and 10 could be improved by drawing the parity line. In Fig.9 caption correct “calclated". Somewhere in the manuscript, a full account of the sources utilized to produce the dataset should be given.

19. The plot in Fig. 11 is not clear. What do the axes mean? What are their units of measure?

20. What is “SHAP”?

21. The discussion in section “Feature analysis” is too much condensed and would probably need more space to be properly expanded. I would suggest reducing the number of graphs and selecting the information to comment.

22. The reported explanation on why low values of lipids coincide with high values of carbohydrates is nicely obvious: “low values for lipids coincide with high values of carbohydrates due to the fact that high values of lipids correspond to low values of carbohydrates”. A more convincing explanation should be found, if any.

23. Fig. 15 and 16 have insufficient captions, as each subplot must be identified and described. Moreover, the units of the represented variables must be given.

24. In the conclusions, “recycling of the HTL aqueous process water seems a good way to spare water resource”. This is misleading, as in a real process you would receive wet biomass, so no external water should be needed. On the contrary, there is a problem with the disposal of process wastewater according to environmental standards.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and does the work have academic merit?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Partly

**Are the conclusions drawn adequately supported by the results?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Biomass, hydrothermal liquefaction, hydroprocessing, upgrading to fuels, drop-in fuels

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 28 Nov 2022

**Geert Haarlemmer**

*Dear Reviewer,*
*Thank you for taking the time to read and review our paper. We have made significant improvements to the paper. Please find responses to your remarks below.*

The present manuscript reports a comprehensive analysis of HTL biocrude production from a number of organic waste, including fresh food waste, organic fraction of municipal solid waste and fermented digestate. The manuscript is in general well-written and presents a number of interesting results. However, there are some parts that need to be presented in a much more concise fashion, while some others are excessively condensed and lack the necessary space to be fully developed. The manuscript is indeed trying to summarize several topics, i.e. the differences between batch and continuous processes, the effect of aqueous phase recirculation and finally a machine learning model for HTL built from a large dataset. Probably it is too much information for a single manuscript.
*Response: Thank you for your review. You highlighted quite a few points that needed addressing. We feel that the manuscript has been greatly improved. You find a point by point response below.*

1. An important issue is the study on aqueous phase recirculation. In my opinion, this is a

not well-placed research question. Indeed, the kind of organic waste taken into account are by their nature wet (potentially even up to 70-80%). In these conditions, aqueous phase recirculation cannot be established, because it would bring to accumulate water in the system. On the other hand, a crucial advantage of HTL over other thermochemical technologies is that the feedstock does not need pre-drying. The authors could explore aqueous phase recirculation because their feedstock was dried beforehand, but this does not correspond to what anyone would do in practice. The authors therefore need to better explain the motivation behind aqueous phase recirculation and how it could be realistically established in a real plant.

*Response: You are right of course. We are in an academic context where the resources have been dried to allow preconditioning of the resources. The initial idea was to shed some light on the chemistry but after review we have decided to remove this section and it is not really important. The data will remain available.*

2. Regarding the experimental procedure, there are some weak points. How could the biocrude remaining in the reactor be evaluated just by weight difference? Is it assumed that there are no residual solids in the reactor? This operation is normally achieved by washing the internal part of the reactor with a solvent and then separating and quantifying the products.

*Response: This only concerns the batch reactor. The text was amended to explain the reactor was dried in an oven at around 50 °C before weighing. A relatively small amount of biocrude remains, typically less than one gram. This amount is taken into account in the mass balances.*

3. How was the amount of produced gas quantified?

*Response: The gas production was calculated by the ideal gas law, from the pressure increase during the experiment as well as the dissolved carbon dioxide as calculated by Henry's law. These equations have been included in the text.*

4. What was the original moisture content of the feed, before pre-drying?

*Response: You can find this information in Table 2.*

5. "Biocrude with high char contents could be considered a product from hydrothermal carbonization" (p. 6). This definition is quite weak and I suggest removing or rephrasing, as there are no fixed boundaries between different processes and, especially, the utilization of the products depends on the solutions utilized on the production plant and not merely by their amounts. For example, scaling up the process to a continuous flow unit could feature some form of on-line filtration allowing to recover important amounts of biocrude.

*Response: OK you are right, the phrase was removed.*

6. The section “Product analyses” tends to be redundant with respect to the section “Product recovery batch experiments”. For example, Karl-Fischer titration is described again. Moreover, azeotropic distillation is not defined. There seem to be two almost similar TOC apparatuses: please have it checked.

*Response: Yes, the text was a bit overly complicated. The section « Product recovery batch experiments » was renamed to « Product recovery» to describe product recovery only. I have effectively simplified the text to limit descriptions of the actual equipment and procedures in the « Product analysis » section. The two TOC you refer to, are two different elements of the same system, one for liquids the other for solids.*

7. “A numerical value is attributed to the extraction method (…)”: this is not very clear, especially it is not clear how this value affects the regression algorithm.

*Response : This is just a modelling trick, as long as we have a numerical value we can use it in any regression correlation. Of course this works when there are two, a third would be complicated as there we need to be a physical sense to the variation in the value. An explanation was added to the text.*

8. The section “Data analysis with machine learning algorithms” seems quite didactical, even with some funny anecdote, like the story of the turkey. I would advise more conciseness, reporting only information necessary to an average-skilled reader to understand and reproduce the methodology.

*Response : After re-reading this section, we do believe that most of the text should stay. The concepts are not easy, and some theoretical background avoids most readers of having to dive into the references.*

9. In Fig. 5 caption, it should be specified that carbohydrates include lignin.

*Response: Done.*

10. The first part of the section “Continuous and batch experiments” belongs to Materials and Methods and not to Results.

*Response: Ok, as the conditions are globally the same for all experiments, some of it can be (and has been) moved to the Materials and Methods section.*

11. Units in Table 3 are wrong. For example, oil to char ratio is not a percentage, while char and gas yields in the last row are. The unit for yields should be made explicit (wt. %). A

carbon balance to verify closure is highly advised.

*Response: Changes made. The carbon balance is presented in the text for the continuous experiment. Measurements were not made for batch experiments.*

12. One of the differences between batch and continuous is the long-lasting thermal transient in batch operations, which is not present in continuous.

*Response: Yes, this is mentioned in the text, in the section HTL Experiments. The text is moved towards the end, to make it more obvious.*

13. Fig. 6 is clearly wrong, as it is identical to Fig. 4. It should report the GC-MS analysis, instead.

*Response: Yes, this error was introduced between the proof approval and publication, the editor was notified. It is correct on the web version. Very strange.*

14. The caption in Fig. 7 is insufficient. Please explain what the left (a) and right (b) plots represent.

*Response: This section was removed from the manuscript.*

15. Eq. 3 and 4 look misleading. There should be a precise indication of the validity limit, especially for residence time. Moreover, it is awkward that a dependence is stated for residence time, when only two values of this variable were tested.

*Response: This section was meant to introduce the modelling part. It is to show how we can easily perform a linear regression with its limitations. You are of course right in that no firm conclusions can be drawn from this.*

16. At page 13, the paragraph: “From table 3 one could conclude (...) makes is even clear” is not clear and needs to be rephrased.

*Response: OK, rephrased.*

17. The legend in Fig. 8 is unclear. Its caption should report (a) and (b) instead of “left” and “right” (same also in Fig. 12).

*Response: This change was implemented on all double figures.*

18. Fig. 9 and 10 could be improved by drawing the parity line. In Fig.9 caption correct “calclated". Somewhere in the manuscript, a full account of the sources utilized to produce the dataset should be given.

*Response: Parity lined have been added to the graphs 9 and 10, old numbering. In the graphs in Figure 11 they have not been added as these graphs are already cluttered in the centre. The references to the data used are given in the Data Collection section. The actual data is given in the Excel file in the Zenodo reference as explained in is presented in the Data availabity/ Undelying Data section.*

19. The plot in Fig. 11 is not clear. What do the axes mean? What are their units of measure?

*Response: The text was amended to better explain this plot.*

20. What is “SHAP”?

*Response: SHAP means “Shapely Additive exPlanations” as is explained in the last paragraph of the section “Data analysis with machine learning algorithms”. Some more explanations were added to the text.*

21. The discussion in section “Feature analysis” is too much condensed and would probably need more space to be properly expanded. I would suggest reducing the number of graphs and selecting the information to comment.

*Response: The section was reworked and simplified with some more detailed explanations. The statistical noise is quite high and ferm conclusions are difficult to obtain. This in itself is however also a result.*

22. The reported explanation on why low values of lipids coincide with high values of carbohydrates is nicely obvious: “low values for lipids coincide with high values of carbohydrates due to the fact that high values of lipids correspond to low values of carbohydrates”. A more convincing explanation should be found, if any.

*Response: Yes as we mentioned in the previous response, this part has been reworked.*

23. Fig. 15 and 16 have insufficient captions, as each subplot must be identified and described. Moreover, the units of the represented variables must be given.

*Response: Figures have been removed after reconsidering this section.*

24. In the conclusions, “recycling of the HTL aqueous process water seems a good way to spare water resource”. This is misleading, as in a real process you would receive wet biomass, so no external water should be needed. On the contrary, there is a problem with the disposal of process wastewater according to environmental standards.

*Response: We have removed this section as it is not really important for the message we want to convey.*

***Competing Interests:*** No competing interests