

Methodological approach to response surfaces and Bayesian optimization using the Gaussian process as predictive tools for lignin extraction from *Pterocarpus tinctorius*

Njouond Kamdem Donald

donaldkamdem6@gmail.com

University of Douala

Kenmogne Sidonie Beatrice

University of Douala

Wansi Jean Duplex

University of Douala

Research Article

Keywords: Response surface methodology, Bayesian optimization, *Pterocarpus tinctorius*, lignin, alkaline hydrolysis

Posted Date: February 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8781383/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations:

No competing interests reported.

Table 1 to 6 are available in the Supplementary Files section.

Methodological approach to response surfaces and Bayesian optimization using the Gaussian process as predictive tools for lignin extraction from *Pterocarpus tinctorius*.

*Njouond Kamdem Donald^{a,b}, Kenmogne Sidonie Beatrice^a, Wansi Jean Duplex^a,

^aDepartment of Chemistry, Faculty of Sciences, University of Douala, Douala - Cameroon
P.O. Box 24157

^bDepartment of Process Engineering, Saint Jerome Catholic University Institute, Douala - Cameroon. Email: donaldkamdem6@gmail.com

***Corresponding author:** Njouond Kamdem Donald

Email: donaldkamdem6@gmail.com

ABSTRACT

An annual yield of wood dust is produced, with a notable contribution arising from African padauk (*Pterocarpus tinctorius*). This wood species, a by-product of the forestry industry, has heretofore been underutilized. The objective of this study was to optimize the conditions for alkaline hydrolysis for the extraction of lignin from *Pterocarpus tinctorius* using two different approaches: response surface methodology (RSM) and Bayesian methodology using the Gaussian process (GP). A comprehensive general design was employed to evaluate the influence of five key factors in the process of lignin extraction by basic hydrolysis. These factors were as follows: the concentration of the NaOH solution (X1: 3 to 7% w/v), the stirring speed of the medium (X2: 150 to 250 rpm), the solid/liquid ratio (X3: 0.05 to 0.1% w/v), the temperature (X4: 60 to 100 °C) and the duration (X5: 2 to 6 hours). The objective was to maximize lignin yield. The models were evaluated using a range of coefficient of determination (R^2), mean square error (MSE), mean absolute error (MAE) and Pearson's X^2 metrics. The results indicated that Bayesian Optimization (BO) demonstrated superior performance in comparison to the response surface model (RSM) ($R^2=99.962$; $MSE=0.049$; $MAE=0.051$; and $X^2=0.078$). The BO model exhibited a significantly higher R^2 value (99.962%) compared to the RSM ($R^2=97.014\%$), as well as a lower MSE (0.049) and MAE (0.051). Additionally, the BO model exhibited a lower X^2 value (0.078) in comparison to the RSM (5.756). The disparities observed between the various validation coefficients are sufficiently pronounced to substantiate the conclusion that BO is by far the most suitable method for this process. In accordance with the predictions made by the BO, the optimal conditions were determined as follows: The experimental conditions comprised a soda percentage of 5%, a rotation speed of 178 rpm, a solid/liquid ratio of 0.074, a temperature of 72.8°C and a duration of 4 hours and 36 minutes. These conditions were optimized to maximize lignin yield, which corresponded to a polyphenolic content of 182.065 mg gallic acid equivalent/g extract with a desirability of 98.29%. The analysis of the functional groups present in the sample was carried out by Fourier transform infrared spectroscopy (FTIR). This analysis revealed the presence of lignin and clearly distinguished the FTIR spectrum of sawdust from that of lignin. X-ray fluorescence spectroscopy was utilized to ascertain the presence of oxides, including Na₂O, K₂O, Al₂O₃, SiO₂, CaO, MgO, Fe₂O₃, SO₃, P₂O₅, TiO₂, MnO, and ZrO₂, with a content of 1.20%, which was found to be proximate to the ash content of 1.76%. X-ray diffraction analysis was performed in order to ascertain the crystalline nature of the sawdust (CrI = 57.56%) and the amorphous nature of the lignin obtained, with an average nanoparticle size of 0.53 nm. Morphological analysis by EDX revealed a highly porous structure of *Pterocarpus tinctorius* sawdust (which explains the efficiency of the extraction) and an irregular and compact structure of the lignin obtained. This work contributes to the optimization of environmentally friendly extraction techniques and provides a basis for the potential valorization of lignin in the pharmaceutical, cosmetics and coating materials industries. Subsequent research will concentrate on its utilization in sustainable (bio-based) epoxy resin synthesis systems.

Keywords: Response surface methodology, Bayesian optimization, *Pterocarpus tinctorius*, lignin, alkaline hydrolysis

INTRODUCTION

Lignocellulosic biomass, which is both abundant and renewable, represents a promising source for the production of high value-added chemical compounds and sustainable bio-based materials[1]. In Cameroon, the process of logging results in the generation of substantial quantities of wood residues, particularly sawdust derived from local species such as *Pterocarpus tinctorius* (African padauk). This species is distinguished by its abundant extracts and distinctive lignocellulosic composition [2]. This underestimated biomass is a potential source of lignin, a complex phenolic polymer with intrinsic properties that are of interest for various industrial applications[3]. Lignin, the second most abundant organic polymer after cellulose, is an essential structural component of plant cell walls, providing them with rigidity and resistance to degradation [4]. Despite the fact that the extraction of lignin from lignocellulosic biomass has been the focus of extensive research [5], further optimization of the process is required to maximize the yield and quality of lignin obtained. Basic hydrolysis, a technique that utilizes alkalis such as sodium hydroxide, is a frequently employed method for the solubilization of lignin from the cellulose and hemicellulose matrix [6]. However, the efficiency of this process is strongly influenced by parameters such as alkali concentration, temperature, reaction time and the nature of the biomass [7]. In industry, experimental optimization tasks are often approached with design of experiment (DOE) methods [8], [9], which provide general strategies for planning data collection and modeling the experimental output. A shortcoming of these approaches is that they largely employ a sampling of design space that is pre-determined. Since experiments must typically be performed sequentially or in batches, however, information from new measurements cannot be utilized to optimize the sampling strategy. Another issue concerns complex processes where many design variables lead to a high-dimensional design space that cannot be efficiently explored using traditional DOE methods. The last decade has seen machine learning methods enter a variety of natural science domains to solve challenging optimization and modeling problems[10][11][12]. For applications in DOE, Bayesian optimization (BO) is a promising machine learning method that couples model building, data collection and global optimization of black-box functions[13][14]. A key feature of BO is an integrated data collection policy, known as the acquisition function, which ensures that the new samples are selected to be as informative as possible given the current state of the model. Consequently, BO is sample efficient and provides better scaling with the dimension of the design space than traditional DOE methods. The BO process can also incorporate experimental errors, prior knowledge as well as batch experiments[15], further increasing its potential for DOE tasks. BO can be extended to deal with a large category of design problems including both constraints and multiple objectives [16]. Recently, BO has become a popular tool in computational materials science, where it has been used both as a means of efficient optimization [17][18][19] and to guide materials discovery .[20] BO is also increasingly being used for DOE, [21][22][23][24][25] although the overall number of studies remains modest. This holds true in particular for applications in the biomaterial's community, where the adoption of machine learning methods is still in its infancy. Existing applications of BO to experiments tend furthermore to focus on optimization problems rather than mapping experimental outputs to design variables. As a result, limited attention is paid to BOs potential for predictive modeling and the insight into experiments and materials this can yield.

In Cameroon, research on the valorization of lignin from local resources remains limited, despite the abundance of these biomasses. To date, the majority of research has concentrated on the general characterization of local biomasses [26], with more limited interest

in the extraction of cellulose for paper manufacturing or bioethanol applications [27]. Nevertheless, lignin extracted from these Cameroonian sources is attracting increasing interest, particularly in the context of the circular economy and the development of bio-based materials [28]. However, previous studies have frequently been limited to descriptive analyses of extraction under empirical conditions, with a paucity of in-depth modelling to optimise the process in order to obtain specific lignin yields and qualities.

The present study aims to address this lacuna by comparing traditional modelling (MSR) and machine learning methods applied to the process of extracting lignin from *Pterocarpus tinctorius* sawdust by basic hydrolysis.

Moreover, a paucity of literature exists on the subject of the preparation of *Pterocarpus tinctorius* lignin for advanced chemical applications, particularly the synthesis of bio-based epoxy resins. The primary objective of this study is to develop a mathematical model capable of predicting lignin extraction yield based on key operational parameters (NaOH concentration, medium agitation speed, solid/liquid ratio, temperature, reaction time, etc.) following a comparison between two specific methods (MSR and OB). It is imperative to understand and improve this extraction process, as the quality and reactivity of the extracted lignin have a direct impact on its potential for further applications [29][30]. The significance of this research lies in its contribution to the valorisation of Cameroon's local wood resources, thereby reducing waste and paving the way for the production of higher value-added bio-based materials. Utilising a rigorous modelling approach, the present study aims to overcome the limitations of previous research by determining an ideal predictive tool for optimizing lignin extraction from *Pterocarpus tinctorius*. Furthermore, it provides a comprehensive physicochemical characterization of sawdust and lignin within a specific Cameroonian context.

Subsequent to this study, the lignin extracted from *Pterocarpus tinctorius* will undergo further chemical modification to increase the reactivity of its functional groups, with particular emphasis on phenolic groups [31]. The ultimate objective is to utilize this modified lignin as a biological component for the synthesis of epoxy resins, which are polymers that are widely utilized in various industries, including coatings, adhesives and composites [32]. The development of bio-based epoxy resins from local Cameroonian resources signifies a substantial advancement in the utilization of more sustainable materials.

Subsequent to this study, the lignin extracted from *Pterocarpus tinctorius* will undergo further chemical modification to increase the reactivity of its functional groups, with particular emphasis on phenolic groups [31]. The ultimate objective is to utilize this modified lignin as a biological component for the synthesis of epoxy resins, which are polymers that are widely utilized in various industries, including coatings, adhesives and composites [33]. The development of bio-based epoxy resins from local Cameroonian resources signifies a substantial advancement in the utilization of more sustainable materials, thereby reducing reliance on fossil resources. This endeavour addresses a pivotal gap in current research in Cameroon.

Materials and methods

Materials

This study was conducted on samples of sawdust from the African Padouk species (*Pterocarpus tinctorius*) obtained from a forest in eastern Cameroon. The identification of the samples was confirmed by a botanical expert at the Cameroon National Herbarium. The wood was subjected to an air-drying process, followed by grinding to a particle size of less than 1 mm via a hammer mill (PX-MFC 90D, POLYMIX, from the UK). This resulted in a powder with a diameter of less than 1.5 mm. The powder was then dried at 100°C for 24 hours, with the objective of achieving a water content of 4.37% (w/w). It is imperative to consider the water content, as it significantly impacts the total mass of the sawdust and influences the extraction processes for phenolic compounds in general. A high-water content may necessitate prior drying to optimise the extraction of phenolic compounds.

Lignin extraction was conducted via a combination of reagents, namely, sodium hydroxide (NaOH, purity $\geq 98\%$, from Sigma–Aldrich), hydrochloric acid (HCl, 37%, from Merck), ethanol (purity $\geq 99.9\%$, from Merck), and distilled water.

Physicochemical characterisation

Determination of the moisture content of the sawdust

The percentage moisture content (PMC) was determined by weighing 1.5 g of the sawdust sample, placing it in a crucible of known mass and placing it in an oven at $105^{\circ}\text{C} \pm 5^{\circ}\text{C}$ for 1 hour. The crucible and its contents were removed from the oven, allowed to cool to room temperature and reweighed. This process was repeated until the weight after cooling became constant, and this value was recorded as the final weight. The moisture content of each sample was determined via equation (1).

$$\%TMC = \frac{W_1 - W_2}{W_2} \times 100 \quad (1)$$

where W_1 is the initial weight of the sawdust sample and W_2 is the final weight of the briquette sample.

Determination of the volatile matter of the sawdust

The percentage of volatile matter (PVM) was determined by placing 1.5 g of the briquette sample in a crucible and keeping it in a furnace for 8 minutes at a temperature of $550^{\circ}\text{C} \pm 5^{\circ}\text{C}$, after which it was weighed after cooling. The percentage of volatile matter in the sample was determined via equation (2):

$$\%PVM = \frac{W_2 - W_3}{W_3} \times 100 \quad (2)$$

where W_2 is the weight of the oven-dried sample (g) and W_3 is the weight of the sample after 8 min in the furnace at 550°C (g).

Determination of the ash content of the sawdust

A total of 1.5 g of each sawdust sample was placed in a closed furnace and burned completely. The weight of the residue was measured with an electronic balance. The percentage weight of residue gives the ash contained in the sample and is determined via equation (3).

$$PAC = \frac{W_4}{W_2} \times 100 \quad (3)$$

X-ray fluorescence (XRF)

X-ray fluorescence analysis was performed with an Axios-type wavelength dispersion spectrometer. The sample was taken in the form of pellets, which were prepared from the fine powder of sawdust.

X-ray Diffraction (XRD)

X-ray diffraction analysis of the sawdust powder was carried out with an XPERT-PRO diffractometer using a Cu-K α line with a wavelength of $\lambda = 1.54060 \text{ \AA}$. The X-ray source was a ceramic tube provided with a copper anode and powered by a voltage current of 30 kV and an intensity of 10 mA.

Estimating the crystallinity index and size of crystalline nanoparticles via X-ray diffraction (XRD) is the subject of this study.

To determine the discrepancy between the crystallinity of the starting material (sawdust) and the extracted lignin, the crystallinity index was determined via the empirical method of determining the height of the DRX peak. This method was developed by Nam et al. (2015)[34] and Safou-Tchiama S (2005)[35]. The process involved examining the alterations in the DRX spectra of the two samples. The CrI crystallinity index was calculated from the ratio of the height of the 002 peak (I_{002}) and the height of the minimum value (I_{AM}) between the 002 and 101 peaks via equation (4) [36]. This method is useful for comparing the relative differences between samples of *Pterocarpus tinctorius* sawdust and lignin obtained after extraction.

$$CrI = \left(\frac{I_{002} - I_{AM}}{I_{002}} \right) \times 100 \quad (4)$$

The sampling technique used was diffuse reflectance. Powder samples were scanned for wavenumbers ranging from 0–4000 cm⁻¹ and 2 θ values ranging from 0–50°. The X-ray diffractometer used was a BRUKER D8 Advance, which uses monochromatic copper K α 1 radiation (CuK α radiation) ($\lambda = 0.154 \text{ nm}$) in 45 kV and 40 mA modes and continuous scanning.

The diameter (D_{002}) of the crystalline nanoparticles of *Pterocarpus tinctorius* sawdust and lignin were calculated from the width at mid-height of the peak (002) via the formula 5 proposed by Scherrer [37].

$$D_{002} = \frac{0.9 \lambda}{B \cos \theta} \quad (5)$$

where λ is the wavelength of the copper K α 1 line ($\lambda = 0.15406$); B is the width halfway up the peak (002); and θ is the diffraction angle of the peak (002).

Fourier Transform Infrared Spectroscopy (FTIR)

The infrared analyses were carried out using a Perkin Elmer brand spectrometer and an FTIR 2000 model. The samples analysed were dispersed in the form of pellets obtained by vacuum compression of a mixture of 1% biomass, finely ground and 99% KBr previously dried in an oven. The IR absorption range used was between 40000 and 400 cm^{-1} .

Scanning electron microscopy (SEM)

The morphological characterisation of lignin and sawdust was performed with a Philips XL30i scanning electron microscope at two magnifications (250 and 1000) using samples with diameters less than 1 mm.

Alkaline extraction of lignin from *Pterocarpus tinctorius* sawdust

Alkaline hydrolysis of *Pterocarpus tinctorius* sawdust was performed in an automated reactor of the ASET XMTG 3000 brand of 2 L. A total of 50 g of *Pterocarpus tinctorius* sawdust was amalgamated with sodium hydroxide solutions of varying concentrations (5% to 9%) and varying solid/liquid ratios (1:20 to 2:20). These solutions were subjected to temperatures ranging from 60 to 100°C and durations ranging from 2 to 6 h. The stirring speeds used ranged from 150 to 250 rpm, as detailed in Table 1. The extracts were filtered through a nylon membrane filter with a pore size of 0.45 μm to separate the black liquor (fig. 1) containing the lignin from the solid particles. The lignin subsequently precipitated through acidification of the black liquor to a pH range of 2--3, utilising chloric acid as the acidifying agent. The precipitated lignin was then recovered by filtration, followed by washing with distilled water to achieve neutrality. The lignin was then oven-dried at 60°C to a constant weight. The extraction yield was calculated, and the optimum extraction conditions were determined according to the polyphenol content (assessed via UV–Vis spectroscopy).

Total phenolic content (TPC)

The total polyphenol content (TPC) of the lignin extracts was determined via the following-Ciocalteu method with slight modifications. To carry out the experiment, 0.2 ml of lignin extract (3 mg/L) was mixed with 0.8 ml of 0.2 N Folin-Ciocalteu reagent. The mixture was then incubated at room temperature in the dark for 20 minutes. Thereafter, 1 ml of 8.5% (w/v) sodium carbonate was added to the mixture, which was left to stand for 90 minutes. The extract was then filtered through 45 μm pore size filter paper. The above samples were measured via a JENWAY 7310 UV/Vis spectrophotometer at a wavelength of 765 nm. Gallic acid was utilised as a standard, and the results are expressed as milligrams of gallic acid equivalent per gram of extract (GAE/g).

Lignin contents

The acid-insoluble lignin contents in the untreated sawdust and alkaline-treated sawdust prepared under the optimal extraction conditions were determined according to the TAPPI standard T 222 om-02 [36] with slight modifications. The difference in the lignin contents between both samples represented the amount of lignin that was extracted. Before analysis, the materials were prepared according to the TAPPI standard T 264 om-88 [38] to obtain extractive-free samples. Briefly, the samples were extracted with 150 ml of an ethanol–toluene mixture (1:2 v/v) for 6 h followed by 50 ml of 95% ethanol for 4 h in a Soxhlet apparatus. Then, 15 ml of 72% v/v sulfuric acid at 10–15°C was added to 1.0 ± 0.1 g of the prepared samples, and the mixture was kept at $20 \pm 1^\circ\text{C}$ for 2 h. Then, it was diluted with distilled water to obtain 3% w/w sulfuric acid solution, boiled, and refluxed for 4 h. Lignin was filtered through nylon membrane filters with 0.45 μm pore diameters and washed with hot distilled water until the pH of the rinse was close to 7.0. The mixture was subsequently dried at 60°C to a constant weight. The lignin content was calculated via Eq 6.

$$\text{Lignin content}(\%) = W_L \times \frac{100}{W_S} \quad (6)$$

where W_L and W_S are the dry weights of lignin and *Pterocarpus tinctorius* sawdust, respectively.

Experimental design

Response surface methodology (RSM)

Response surface methodology (RSM) was utilised to optimise the conditions for the extraction of lignin from *Pterocarpus tinctorius* sawdust through alkaline hydrolysis. A five-variable (factor) Box–Behnken design was employed to investigate the combined effects of the percentage of sodium hydroxide (X_1), medium stirring speed (X_2), solid/liquid ratio (X_3), temperature (X_4), and extraction time (X_5) at three levels (-1, 0, +1). The experimental levels are expressed in Table 1. The design comprised a total of 30 experimental trials, including 3 focal points. The response variable was the total polyphenol content (TPC, Y), and the data were integrated into a second-order polynomial regression equation described by equation (7):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{14} X_1 X_4 + \beta_{15} X_1 X_5 + \beta_{23} X_2 X_3 + \beta_{24} X_2 X_4 + \beta_{25} X_2 X_5 + \beta_{34} X_3 X_4 + \beta_{35} X_3 X_5 + \beta_{45} X_4 X_5 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 + \beta_{44} X_4^2 + \beta_{55} X_5^2 \quad (7)$$

where Y_i is the response of interest, β_0 is the constant coefficient, $\beta_1, \beta_2, \beta_3, \beta_4$ et β_5 are the linear coefficients, $\beta_{12}, \beta_{13}, \beta_{14}, \beta_{15}, \beta_{23}, \beta_{24}, \beta_{25}, \beta_{34}, \beta_{35}$ et β_{45} are the interaction coefficients and $\beta_{11}, \beta_{22}, \beta_{33}, \beta_{44}$ et β_{55} are the quadratic coefficients.

Bayesian Optimisation with a Utilitarian Approach via Gaussian Processes

BO involves two main components, namely a surrogate model that approximates the objective function and an acquisition function that provides a data collection policy. During a BO iteration, the surrogate model is fit to the current data set using Gaussian process regression. The posterior mean of the Gaussian process represents the most probable approximation of the objective, and the posterior variance provides a measure of the model uncertainty. By

minimizing the acquisition function, a new sampling location is subsequently determined and used to augment the existing data set. Acquisition functions come in many flavors and provide trade-offs between exploitation and exploration. Here, the former refers to the sampling regions of design space where the objective is likely to achieve a minimum or maximum, while the latter refers to the sampling of regions where model uncertainty is high and where data has not been sampled before. Common choices of acquisitions function include the lower confidence bound (LCB) function [37] and expected improvement (EI) function.

We carried out the BO using JMP software; In our experimental context, the objective was to maximize the amount of lignin extracted through the total polyphenol content (Y) in the extracts.

In order to maximize lignin extraction and optimize the final polyphenol content (Y), a Bayesian Optimization (BO) strategy was adopted. This sequential method is particularly well-suited to objective functions that are costly to evaluate and of the "black box" type. The experimental space was defined by five controlled factors: sodium hydroxide concentration (X₁), stirring speed (X₂), solid/liquid ratio (X₃), temperature (X₄), and extraction time (X₅). The experimental levels are delineated in Table 1.

The underlying relationship between these input parameters x and the yield $f(x)$ was modelled by a Gaussian Process (GP), assuming noisy observations of the form $y = f(x) + \varepsilon$, with Gaussian noise $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$ [39][14]. At each iteration t , the posterior predictive distribution for a new point x is found to follow a normal distribution, conditional on the set of observed data D_t . The predicted mean (return estimate) and the variance (model uncertainty) are calculated respectively by the following equations (8 and 9):

$$\mu_t(x) = k_*^T (K + \sigma_n^2 I)^{-1} Y_t \quad (8)$$

$$\sigma_t^2(x) = k(x, x) - k_*^T (K + \sigma_n^2 I)^{-1} k_* \quad (9)$$

In this study, the observed points are represented by the K matrix of the observed points and the k_* vector of the observed points and the new point x . In order to arbitrate the exploration-exploitation trade-off within the 5-dimensional space, the Expected Improvement (EI) acquisition function was maximized to select each new experimental condition[14]. The EI is analytically defined by:

$$\alpha_{EI}(x) = (\mu_t(x) - f(x^+) - \xi) \Phi(z) + \sigma_t(x) \phi(z) \quad (10)$$

$$Avec \quad Z = \frac{\mu_t(x) - f(x^+) - \varepsilon}{\sigma_t(x)} \quad (11),$$

where $f(x^+)$ is the best yield observed to date, and Φ and ϕ are respectively the cumulative distribution function and the probability density function of the standard normal distribution. This iterative approach was conducted on a total of 247 experimental trials (table 2), allowing robust convergence towards optimal operating conditions. The elevated number of iterations (N=247) facilitated the identification of the global optimum with high precision and concomitantly led to a substantial reduction in the predictive variance $\sigma^2(x)$ of the Gaussian Process model across the entire 5-dimensional experimental space.

To facilitate comparisons between the two models, the root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), mean percentage error (MPE) and Pearson's chi-squared measure (χ^2) were determined using equations 12-17:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \frac{1}{n} \sum_{i=1}^n y_i)^2} \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (13)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (14)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (15)$$

$$MPE = \frac{100\%}{n} \sum \left(\frac{y - \hat{y}}{y} \right) \quad (16)$$

$$\chi^2 = \sum_{i=1}^n \frac{(y - \hat{y})^2}{\hat{y}} \quad (17)$$

where n , y_i and $\hat{y}_i$ are the sample size and the actual and predicted values, respectively. R^2 evaluates the fitness between the actual and predicted outputs. Greater values for this indicator indicate perfect fitness. The RMSE, MAPE, MAE, MPE and (χ^2) are expected to be very small (table3).

Results and discussion

Optimal extraction conditions for lignin from *Pterocarpus tinctorius* sawdust

The present study adopted two methodologies with the aim of comparing models in order to optimize the conditions for alkaline hydrolysis for the extraction of lignin from *Pterocarpus tinctorius* sawdust with the greatest precision. The two methodologies that were adopted were response surface methodology and Bayesian optimization (BO). Preliminary single-factor experiments were conducted to identify the ranges of the independent variables in the process, namely the concentration of the NaOH solution, the stirring speed, the extraction temperature, the solid/liquid ratio, and the extraction time. Initially, the effects of different NaOH solution concentrations (5 to 9%) were studied at a constant extraction temperature (100°C), a constant solid/liquid ratio (0.1), a constant stirring speed (200 rpm) and a constant extraction time (4 hours). The findings indicated that the TPC of lignin extracts exhibited an upward trend with increasing NaOH concentration up to 5%, and that a subsequent increase in concentration did not result in a substantial enhancement of the value obtained. The second step was to determine the effects of different extraction temperatures (65-105 °C) at a fixed concentration of 5% NaOH solution, obtained in the first step, at a constant solid/liquid ratio (0.1), a constant stirring speed (200 rpm) and a constant extraction time (4 hours). The highest TPC was obtained at 90 °C. *Pterocarpus tinctorius* sawdust was then extracted at varying times (2 to 6 hours) with a fixed concentration of 5% NaOH solution at a constant extraction temperature (90 °C), a constant solid/liquid ratio (0.1) and a constant stirring speed (200 rpm). The TPC increased with time until it reached a plateau after 4 hours. In the subsequent experiment, *Pterocarpus tinctorius* sawdust was extracted at varying stirring speeds (150-250 rpm) with a constant concentration of 5% NaOH solution, a constant temperature of 90 °C and a constant solid/liquid ratio (0.1) for a duration of 4 hours. The highest TPC was obtained at

200 rpm. Finally, extraction was conducted at varying solid/liquid ratios (0.05-0.1), while maintaining constant parameters such as temperature (90 °C), stirring speed (200 rpm), duration (4 hours), and NaOH concentration (5%). The highest TPC was obtained with a solid/liquid ratio of 0.085. Preliminary results have indicated the necessity of establishing three levels (low, -1; medium, 0; and high, +1) for each process variable. The categorization is illustrated in Table 1.

Response surface methodology model and its adequacy

The TPC values of the lignin extracts obtained from 247 runs are shown in Table 2. These values ranged from 125.25–179.81 mg GAE/g extracts. The second-order polynomial model for the TPC was constructed from the experimental data. It is expressed as Eq.12:

$$TPC(Y) = 231.1 - 486X_1 - 0.35X_2 - 828X_3 + 0.517X_4 + 3.68X_5 + 2430X_1^2 + 3530X_3^2 - 0.17X_5^2 + 1.24X_1X_2 + 5078X_1X_3 - 8.74X_1X_4 + 15.9X_1X_5 + 0.32X_2X_3 - 2.69X_3X_4 - 7.8X_3X_5 - 0.02X_4X_5 \quad (18)$$

Analysis of variance (ANOVA) and the coefficient of determination (R^2) of the TPC were used to check the adequacy of the RSM models. The ANOVA revealed significant results of the model at the 95% confidence interval. The significant contributions were determined on the basis of the probability values (P values) of the F test, as shown in Table 4. The P value for response was lower than 0.0001, revealing that the model was highly significant. The P value of lack of fit for response was not significant ($P < 0.05$), confirming the suitability of the model. The R^2 value of the TPC was 96.26. As it was greater than 0.80, the observed and predicted values of all the variables were considered highly correlated [40]. Therefore, the proposed model was adequate for representing the relationships between the independent and response variables.

Bayesian Optimization

The distribution of the 247 experimental points demonstrates a significant positive correlation between the predictions of the Gaussian Process model and the actual observations ($R^2 \approx 0.6-0.7$, visual estimate), validating the model's capacity to capture the topology of the response surface despite the complexity of the vegetation matrix (see fig.3). A notable densification of points is observed in the high yield region ($Y > 160$). This distribution is not random but is characteristic of the Bayesian Optimization strategy, whereby the algorithm favours the exploration and exploitation of areas of the experimental space identified as optimal (EI acquisition function). The residual dispersion around the diagonal is attributed to the natural variability of the plant substrate and experimental uncertainties, which are integrated into the model via the Gaussian noise parameter (σ_n^2).

The Bayesian Optimization approach facilitated effective convergence towards the optimum (as indicated by the density of the parity graph), whilst the underlying Gaussian Process model achieved near-ideal predictive performance, as demonstrated by an R^2 of 0.9996 and an average error of only 0.2234 (RMSE) in Jackknife cross-validation. The findings of this study provide maximum confidence in the optimal extraction conditions recommended by the algorithm. An analysis of the jackknife cross-validation metrics reveals the exceptional predictive capabilities of the Gaussian Process model. The coefficient of determination (R^2) of 0.9996 indicates that almost all of the variance in polyphenol yield is explained by the five extraction factors. The low value of the nugget ($\sigma_n^2 = 0.0188$), which represents the variance of

the measurement noise, confirms the high fidelity of the experimental data. These findings are substantiated by a negligible mean absolute error (MAE = 0.051) and a minimal mean relative error (MAPE = 3.1). These results provide robust validation of the model's reliability, thereby substantiating the confidence attributed to the utilization of Bayesian Optimization in identifying the optimal conditions.

Comparative study between the two approaches.

While both methodologies yielded robust models, Bayesian Optimisation exhibited a marked superiority in absolute accuracy when compared with MSR. The Gaussian Process model demonstrated a near-perfect fit ($R^2 = 0.9996$) and an average prediction error (RMSE = 0.2234) approximately nine times lower than that obtained by MSR (RMSE = 1.93). These results provide validation for the superiority of the non-parametric probabilistic model in accurately mapping the complex response surface of extraction, thus ensuring maximum confidence in the identified global optimum[10].

Given that the MSR's MAPE is 0.92 (compared to 3.1% for BO) (Table 5), this could indicate that the polynomial model is slightly more accurate than the GP only at the centre of its design, or that the BO dataset of 247 trials, including a broad exploration phase, has peripheral points that increase its overall relative error, even though its overall R^2 and RMSE are better. However, in the context of optimization, the low root mean square error (RMSE) of Bayesian optimization (BO) is the determining criterion of superiority [41]

The combination of high absolute accuracy ($RMSE_{BO} \ll RMSE_{MSR}$) and low noise ($\sigma_n^2 \approx 0.0188$) provides the model with a high degree of confidence in each sampling decision, thereby facilitating faster convergence towards maximum yield[42].

Effects of alkaline hydrolysis conditions on the TPC

Three-dimensional response surface plots obtained from the predictive model were used to illustrate the main and interactive effects of the independent variables on the response variables. Fig. 2 A–F (see fig.2) shows the response surface plots of the TPC of the lignin extracts as a function of two process variables, while the other variables were kept constant at zero. The TPC was critically affected by the NaOH concentration and solid/liquid ratio according to the ANOVA results, as shown in Table 4. Although the linear and quadratic effects of both parameters were significant, their interactions were not significant ($P > 0.05$). Fig. 2B shows the effects of the NaOH concentration and solid/liquid ratio on the TPC of the lignin extracts. The plot trend indicated that the TPC increased when the NaOH concentration and solid/liquid ratio decreased. These results indicated that a high solid/liquid ratio had a positive influence on the reaction rate of alkaline hydrolysis, but exposure to the highest NaOH concentration and high solid/liquid ratio resulted in a decrease in the TPC. This might be attributed to the undesirable change in the lignin structure of the phenolic groups [43].

As demonstrated in Table 3, the interaction between the NaOH concentration and the solid/liquid ratio had a significant effect on the TPC. Fig. 2B shows the impact of the NaOH concentration (X_1) and solid/liquid ratio (X_3) on the TPC of the lignin extracts. The trend of the

interaction between the NaOH concentration and extraction time (Fig. 2 D) indicates that the maximum TPC was reached for an extraction time of less than 5 h and that as the extraction time increased, the TPC decreased. This phenomenon persists even in instances of heightened NaOH concentrations. The decline in TPC observed in the presence of elevated NaOH concentrations, concomitant with an extended extraction time, can be explained by the chemical modification of phenolic groups under the influence of temperature, duration, and sodium hydroxide.

Optimal conditions for alkaline hydrolysis and model accuracy

The numerical optimization of the conditions for extracting lignin from *Pterocarpus tinctorius* sawdust was conducted utilizing JMP PRO17 software, with the process informed by experimental results following Bayesian optimization. The software in question is utilized for the purpose of identifying all potential combinations of factor levels whose requirements are placed on the response and the factor concurrently. In order to ascertain the optimal conditions, five independent variables (NaOH solution concentration, stirring speed, solid/liquid ratio, temperature and duration) were defined within the ranges indicated in Table 1, and the TPC was set to the maximum value. The optimal conditions for lignin extraction were as follows: The NaOH solution was prepared at a concentration of 5% w/v, with a stirring speed of 177.7 rpm. The solid-to-liquid ratio was maintained at 0.074 w/v, and the temperature was set to 82°C. The duration of the extraction process was 4.59 hours, yielding 182.06 mg GAE/g extract, with a desirability of 0.98. The adequacy of the model was verified by comparing the experimental and predicted values of the responses obtained under the conditions predicted from the analysis (see Table 6). The relative errors between the predicted and experimental values were found to be less than 5%. The findings indicate that Bayesian optimization is a viable and more efficacious approach than RSM for the optimization of the alkaline extraction of lignin from *Pterocarpus tinctorius* sawdust.

Characteristics of lignin extracts prepared under optimal conditions

Lignin content, porosity, moisture content, ash content and density

Under the optimal extraction conditions, a high extraction efficiency (74.78%) of lignin from *Pterocarpus tinctorius* sawdust was achieved, which is in good agreement with those (70–90%) reported in previous studies[44][45][46]. The lignin content was reduced from $27.14 \pm 0.35\%$ in the *Pterocarpus tinctorius* sawdust to $7.02 \pm 0.54\%$ in the residue. This corresponded to 0.2714 g of lignin in 1 g of *Pterocarpus tinctorius* sawdust, 0.0192 g of lignin in the residue, and 0.2522 g of lignin in the extract.

It is evident from the observations of porosity, water content, ash content and density that *Pterocarpus tinctorius* has a porosity of 0.498, a water content of 14.12%, an ash content of 1.76% and a density of 196.5 g/L. This density value appears to be in perfect alignment with the porosity value obtained. A relatively low density is indicative of a porous structure, which may facilitate solvent access to phenolic compounds during extraction[47]. Adequate drying is imperative to optimise the efficiency of polyphenol extraction[48], and the low ash and mineral oxide contents indicate that sawdust is predominantly composed of organic matter, a favourable composition for polyphenol extraction[46][48]. The mineral composition of the material has been shown to have a significant effect on the optimisation of extraction and purification

processes for polyphenols. Furthermore, the compatibility of sawdust with other industrial applications can be assessed[49].

FT-IR spectra

The characterisation of the raw material (*Pterocarpus tinctorius* sawdust) and the extracted lignin was accomplished through the utilisation of Fourier transform infrared spectroscopy (FTIR). FTIR spectroscopy can provide reliable insights into the structural and chemical features of lignocellulosic materials[50]. Therefore, to clarify the efficiency and specificity of the extraction method used, the presence/absence of lignin in the lignin-rich solid extract was evaluated via FTIR, and the spectra of *Pterocarpus tinctorius* sawdust (red line) and lignin extract (black line) are depicted in Figure 5. As alkaline hydrolysis extraction destroyed the structure of *Pterocarpus tinctorius* sawdust, it had a direct effect on the functional groups present in the lignin extract. The FT-IR technique was used to identify the functional groups of the extract prepared under the optimal conditions (5% NaOH, an agitation speed of 150 rpm, a solid/liquid ratio of 0.05, and a temperature of 100°C for 3.95 hours (237 min)) and to study the effects of alkaline treatment on the structural changes in *Pterocarpus tinctorius* sawdust.

The spectral bands of the lignin extracts are shown in Fig. 5. They were analysed on the basis of the wavenumber assignments reported by Boeriu et al[51], Abdelaziz and Hultberg [52] and Nitchanan et al[46]. The bands at 1592, 1511, and 1420 cm^{-1} corresponded to the aromatic skeletal vibrations of lignin. The band at 1461 cm^{-1} was due to C–H deformations combined with aromatic ring vibrations. The vibration characteristics of phenolic hydroxyl groups (O–H in-plane deformation) and conjugated carbonyl groups (C=O stretching) appeared at 1383 cm^{-1} and 1640 cm^{-1} , respectively. The phenolic hydroxyl groups might result from the cleavage of weak β -O-4 and α -O-4 aryl-ether interunit linkages, a dominant reaction during alkaline hydrolysis [33]. The band representing C–H stretching vibrations in aromatic methoxyl groups at 2850 cm^{-1} was also detected. The characteristic fingerprints of carbohydrate and lignin functional groups appear in the 1800–1200 cm^{-1} region, whereas the OH and CH vibration modes are detected in the 3800–2700 cm^{-1} region [46]. The vibrational band ca. 1530–1480 cm^{-1} is attributed mostly to the stretching of C=C bonds, which are part of the aromatic skeleton of lignin molecules. Furthermore, the C=O stretching attributed to the aromatic ring bending mode is assigned to the band at approximately 1720–1660 cm^{-1} . Finally, the bands at 1490 and 1510 cm^{-1} can be assigned to the vibration modes of the aromatic rings in lignin[53]. Furthermore, an increase in the intensity of the C=O (carboxylic acid and carbonyl) infrared bands under alkaline conditions was also observed. This phenomenon is attributable to the cleavage of β -O-4 aryl ether bonds (Fig. 4) under alkaline conditions, which results not only in the formation of phenols but also in the formation of carboxylic acids[54]. The weak bands appearing at 1364 cm^{-1} for the lignin sample can originate from nonetherified phenolic OH groups, which are liberated from the cleavage of β -O-4 and α -O-4 linkages during the cooking process; this finding supports extensive phenol ether cleavage during extraction[52]. The previously highlighted bands can also be identified in the starting *Pterocarpus tinctorius* sawdust, as well as in the extracted fraction (lignin-rich portion). However, in the *Pterocarpus tinctorius* sawdust spectrum, these bands have very low intensities, thus confirming the very high differences between the two samples. The extract has a greater concentration of lignin and aromatic skeletons of lignin molecules than the initial raw material (sawdust), confirming that our extract could be lignin at a fairly high purity.

X-ray fluorescence spectrometry (XRF) analysis

X-ray fluorescence (XRF) revealed a total mineral oxide percentage of 1.20%, which was composed of various mineral oxides, such as Na₂O, K₂O, Al₂O₃, SiO₂, CaO, MgO, Fe₂O₃, SO₃, P₂O₅, TiO₂, MnO and ZrO₂. This analysis provides a detailed breakdown of the elements present in the ash[49]. The ash content was estimated at 1.76%.

The difference between the total ash content and the percentage of total mineral oxides is due to the presence of residual carbon and other organic compounds that are not completely oxidised during combustion. The mineral oxides represent the pure inorganic portion of the ash[48].

X-ray diffraction (XRD) analysis

The diffractogram of *Pterocarpus tinctorius* sawdust shows the presence of semicrystalline material, and three diffraction peaks at approximately $2\theta = 16, 22$ and 34° are characteristic of cellulose. According to Ciolacu et al. and El Hajam et al., those peaks correspond to the diffractions of the (110) planes of amorphous cellulose and the (002) and (004) planes of crystalline cellulose[55], respectively[56] (Fig. 6). X-ray powder diffraction (XRD) analysis was performed to study the lignin morphology in more detail and to determine the differences from *Pterocarpus tinctorius* sawdust. The XRD diffractogram of lignin revealed broad diffraction of the amorphous halo, with a maximum at approximately $2\theta = 22^\circ$ (Fig. 6), which confirmed its natural amorphous character. However, previous research has demonstrated that differences in the original source of lignin, the chemical structure and the extraction method used may explain the variation observed in the maximum peak of lignin. For example, Mohammed et al.[57] and Goudarzi et al[58]. and Kubo et al[59]. reported average peaks for hardwood located at 20° , $21.2 \pm 0.15^\circ$ and 22.7° [59], respectively.

The X-ray diffraction (XRD) profiles of *Pterocarpus tinctorius* sawdust and extracted lignin (Fig. 6) revealed four distinct peaks with diffraction angles (2θ) of approximately 14.86° , 22.37° , and 45.86° , which are characteristic of the (101), (200), (004), and (040) planes of cellulose I[60]. The position of the peak at 45.86° varies depending on the extraction technique employed.

The peak at approximately 30.03° , by contrast, is less directly attributable to the main crystalline forms of cellulose. It has been hypothesised that this may correspond to reflections of less predominant planes in the cellulose or, alternatively, to the presence of other components in the *Pterocarpus tinctorius* sawdust, such as minerals, as revealed via X-ray fluorescence (XRF). However, following the extraction process, a significant decline in the intensity of the crystalline peaks at 14.86° and 22.37° was observed in the lignin curve (black curve). In some cases, these peaks even disappeared at approximately 30.3° and 45.86° .

The mean nanoparticle sizes for sawdust and lignin were determined to be 7.19 and 0.53 nanometres, respectively, and the crystallinity indices for sawdust and lignin were 57.56% and 32.33%, respectively. The results obtained demonstrate a significant discrepancy in the crystallinity index between the raw material (sawdust) and the extract (lignin). This is attributable to the fact that sawdust is predominantly composed of cellulose, which is recognised as a crystalline structure, whereas lignin is significantly more amorphous[61].

Morphological analysis

Electron diffraction (EDX) was used to characterise the surface microstructure of the sawdust sample, whereas scanning electron microscopy (SEM) was used to characterise the extracted lignin. Morphological examination yielded an EDX microphotograph of the sawdust sample since FTIR and XRD analyses confirmed the functional and structural differences between the two samples (see Figure 7). The image, magnified at 250x and displayed in Fig. 7(a), reveals a surface with an irregular morphology and heterogeneous structure, as previously observed in the sawdust sample. This figure shows the presence of intercellular voids in the form of partially exposed cavities in the sawdust sample. A horizontal section of the sawdust, as illustrated in image 6(a), reveals the presence of tracheids, which are defined as follows. The porous nature of the surface is evident and is a key factor in understanding the efficacy of the solvents employed in the extraction process. This porosity is advantageous in facilitating the movement of the solvents to access and extract the target molecules (lignin). Furthermore, hidden micropores were exposed in the walls of longitudinal cavities, as illustrated in Fig. 7 (a). As illustrated in Figure 7(b), the particles of lignin appear smooth and relatively compact, exhibiting significant variations in size and shape.

Conclusion

This study provides a comprehensive analysis of the modelling and optimisation, as well as the physicochemical characterisation, of lignin extracted from African padauk (*Pterocarpus tinctorius*). A comparative study between response surface methodology and Bayesian optimisation was conducted, which resulted in the conclusion that the most suitable predictive method for studying the extraction of plant molecules (specifically lignin from *Pterocarpus tinctorius*) was Bayesian optimisation, with a near-perfect fit ($R^2 = 0.9996$) and an average prediction error (RMSE = 0.2234) approximately nine times lower than that obtained by RASM (RMSE = 1.93). Subsequent optimisation of this extraction process enabled the extraction of a maximum amount of lignin, accompanied by the identification of the presence of phenolic groups in maximum quantity (182.06). The experiment yielded an equivalent gallic acid concentration of 71 mg/g of extract, with a desirability greater than 98.3% for the optimal factors. The experimental factors included a NaOH concentration of 5% w/v, a stirring speed of 177.7 rpm, a solid/liquid ratio of 0.074 w/v, a temperature of 82°C and an extraction time of 4.59 hours. The lignin content was estimated to be $27.14 \pm 0.35\%$ in sawdust and $7.02 \pm 0.54\%$ in residues. The analysis of the functional groups present in the sample was performed by Fourier transform infrared spectroscopy (FTIR). This analysis revealed the presence of lignin and clearly distinguished the FTIR spectrum of sawdust from that of lignin. X-ray fluorescence spectroscopy was utilised to ascertain the presence of oxides, including Na₂O, K₂O, Al₂O₃, SiO₂, CaO, MgO, Fe₂O₃, SO₃, P₂O₅, TiO₂, MnO and ZrO₂, with a content of 1.20%, which is proximate to the ash content of 1.76%. X-ray diffraction analysis was performed in order to verify the crystalline nature of the sawdust (CrI = 57.56%) and to determine the amorphous nature of the lignin obtained, with an average nanoparticle size of 0.53 nm. Morphological analysis by EDX revealed a highly porous structure in *Pterocarpus tinctorius* sawdust (which explains the efficiency of the extraction process) and an irregular and compact structure in the lignin obtained. This work contributes to the optimization of environmentally friendly extraction techniques and lays the foundations for the potential use of lignin in the pharmaceutical, cosmetics and coating materials industries. Subsequent research will concentrate on the utilization of these catalysts in sustainable (bio-based) epoxy resin synthesis systems.

Authors' contribution statement

The lead author, NJOUOND KAMDEM Donald, was responsible for all experimental work, data analysis and the composition of the entire manuscript.

Co-author Kemmogne Sidonie Beatrice contributed to experimental planning, methodology definition and results validation.

The supervision of all work was overseen by co-author Wansi Jean Duplex, who also provided regular scientific guidance and validated the final version of the manuscript.

All authors have perused and endorsed the final version of the document, and consent to its submission to the journal Biomass Conversion and Biorefinery.

- [1] A. J. Ragauskas *et al.*, “Lignin valorization: Improving lignin processing in the biorefinery,” *Science* (80-.), vol. 344, no. 6185, 2014, doi: 10.1126/science.1246843.
- [2] J.-B. S. Tchinda, “Caractérisation et valorisation des substances extractibles de cinq essences camerounaises majeures de l’industrie du bois : Ayous, Moabi, Movingui, Padouk et Tali,” *Thèse en cotutelle présentée pour obtenir le grade Docteur/Ph.D, Univ. Lorraine, Ec. Dr. Sci. Ingénierie Ressources Procédés Prod. Environ.*, no. October 2015, p. 161, 2015, [Online]. Available: <https://hal.univ-lorraine.fr/tel-01751831>
- [3] J. Zakzeski, A. L. Jongerius, P. C. A. Bruijninx, and B. M. Weckhuysen, “Catalytic lignin valorization process for the production of aromatic chemicals and hydrogen,” *ChemSusChem*, vol. 5, no. 8, pp. 1602–1609, 2012, doi: 10.1002/cssc.201100699.
- [4] W. Boerjan, J. Ralph, and M. Baucher, “Lignin Biosynthesis,” *Annu. Rev. Plant Biol.*, vol. 54, pp. 519–546, 2003, doi: 10.1146/annurev.arplant.54.031902.134938.
- [5] K. Zhang, X. Yang, N. Zhu, Z. C. Wang, and H. Yan, “Environmentally benign paints for superhydrophobic coatings,” *Colloid Polym. Sci.*, vol. 295, no. 4, pp. 709–714, Apr. 2017, doi: 10.1007/s00396-017-4053-5.
- [6] D. Watkins, M. Nuruddin, M. Hosur, A. Tcherbi-Narteh, and S. Jeelani, “Extraction and characterization of lignin from different biomass resources,” *J. Mater. Res. Technol.*, vol. 4, no. 1, pp. 26–32, 2015, doi: 10.1016/j.jmrt.2014.10.009.
- [7] D. Stewart, “Lignin as a base material for materials applications: Chemistry, application and economics,” *Ind. Crops Prod.*, vol. 27, no. 2, pp. 202–207, 2008, doi: 10.1016/j.indcrop.2007.07.008.
- [8] “DOE-RAF.pdf.”
- [9] D. C. Montgomery, *Design and Analysis of Experiments Eighth Edition*.
- [10] C. Methods and A. Karl, “NEURAL NETWORKS AS TOOLS TO SOLVE PROBLEMS,” pp. 1–14, 1994.
- [11] A. W. Senior *et al.*, “Improved protein structure prediction using potentials from deep learning,” vol. 577, no. January, 2020, doi: 10.1038/s41586-019-1923-7.
- [12] B. Tang, Z. Pan, K. Yin, and A. Khateeb, “Recent Advances of Deep Learning in Bioinformatics and Computational Biology,” vol. 10, no. March, pp. 1–10, 2019, doi: 10.3389/fgene.2019.00214.
- [13] J. Mockus, “ON BAYESIAN METHODS FOR SEEKING THE EXTREMUM,” p. 5, 2018.
- [14] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, “Taking the Human Out of the Loop : A Review of Bayesian Optimization,” no. 1, pp. 1–24.
- [15] D. Ginsbourger, R. Le Riche, and L. Carraro, “Kriging is well-suited to parallelize optimization,” pp. 1–27.
- [16] A. Shah, Z. Ghahramani, and M. L. Group, “Pareto Frontier Learning with Expensive Correlated Objectives,” vol. 48, 2016.
- [17] G. Río, J. Jørgen, K. Wedel, E. Garijo, J. J. Mortensen, and K. W. Jacobsen, “Local Bayesian optimizer for atomic structures Local Bayesian optimizer for atomic

structures,” 2019, doi: 10.1103/PhysRevB.100.104103.

- [18] Y. Chen, L. Wang, and H. Huang, “An effective surrogate model assisted algorithm for multi-objective optimization : application to wind farm layout design,” no. September, pp. 1–16, 2023, doi: 10.3389/fenrg.2023.1239332.
- [19] M. Todorovi, “Bayesian inference of atomistic structure in functional materials,” no. February, pp. 1–7, 2019.
- [20] T. Lookman, “Accelerated Search for Materials via Adaptive Learning Introduction : Approach to Learning Examples : NiTi alloys Pb- free piezoelectrics”.
- [21] R. D. C. Leal, C. Li, D. Rubín, and D. Cel, “Rapid Bayesian optimisation for synthesis of short polymer fiber materials,” vol. 7, pp. 1–10, 2017, doi: 10.1038/s41598-017-05723-0.
- [22] P. B. Wigley, P. J. Everitt, A. Van Den Hengel, J. W. Bastian, and M. A. Sooriyabandara, “Fast machine-learning online optimization of ultra-cold-atom experiments,” pp. 1–6, 2015, doi: 10.1038/srep25890.
- [23] S. Greenhill, S. Rana, S. Gupta, and P. Vellanki, “Bayesian Optimization for Adaptive Experimental Design : A Review,” vol. 8, pp. 13937–13948, 2020.
- [24] T. H. P. Hartono , Anuj Goyal, “A data fusion approach to optimize compositional stability of halide perovskites”.
- [25] Q. Liang *et al.*, “Benchmarking the Performance of Bayesian Optimization across Multiple Experimental Materials Science Domains,” no. NeurIPS, pp. 1–7, 2020.
- [26] C. Kapseu, “Énergies renouvelables en Afrique subsaharienne,” no. January, p. 363, 2012, [Online]. Available: <http://www.editions-harmattan.fr/index.asp?navig=catalogue&obj=livre&no=37936>
- [27] H. A. Rasheed, A. A. Adeleke, P. Nzerem, A. I. Olosho, T. S. Ogedengbe, and S. Jesuloluwa, “Isolation, characterization and response surface method optimization of cellulose from hybridized agricultural wastes,” *Sci. Rep.*, vol. 14, no. 1, pp. 1–15, 2024, doi: 10.1038/s41598-024-65229-4.
- [28] A. V. Bridgwater, “Renewable fuels and chemicals by thermal processing of biomass,” *Chem. Eng. J.*, vol. 91, no. 2–3, pp. 87–102, 2003, doi: 10.1016/S1385-8947(02)00142-0.
- [29] G. W. Huber and A. Corma, “Synergies between Bio- and Oil Refineries for the Production of Fuels from Biomass Angewandte,” pp. 7184–7201, 2007, doi: 10.1002/anie.200604504.
- [30] G. Arimura and D. P. W. Huber, “Forest tent caterpillars (*Malacosoma disstria*) induce local and systemic diurnal emissions of terpenoid volatiles in hybrid poplar (*Populus trichocarpa* × *deltoides*): cDNA cloning , functional characterization , and patterns of gene expression of (*A*),” pp. 603–616, 2004, doi: 10.1111/j.1365-313X.2003.01987.x.
- [31] H. Hatakeyama, R. Kosugi, and T. Hatakeyama, “Thermal properties of lignin-and molasses-based polyurethane foams,” *J. Therm. Anal. Calorim.*, vol. 92, no. 2, pp. 419–424, 2008, doi: 10.1007/s10973-007-8963-1.
- [32] M. Fache, A. Viola, R. Auvergne, B. Boutevin, and S. Caillol, “Biobased epoxy

- thermosets from vanillin-derived oligomers,” *Eur. Polym. J.*, vol. 68, pp. 526–535, Oct. 2015, doi: 10.1016/j.eurpolymj.2015.03.048.
- [33] M. Fache, B. Boutevin, and S. Caillol, “Vanillin Production from Lignin and Its Use as a Renewable Chemical,” *ACS Sustain. Chem. Eng.*, vol. 4, no. 1, pp. 35–46, 2016, doi: 10.1021/acssuschemeng.5b01344.
- [34] S. Nam, A. D. French, B. D. Condon, and M. Concha, “Segal crystallinity index revisited by the simulation of X-ray diffraction patterns of cotton cellulose I β and cellulose II,” *Carbohydr. Polym.*, vol. 135, pp. 1–9, 2016, doi: 10.1016/j.carbpol.2015.08.035.
- [35] R. Safou and T. Tchiama, “Valorisation des produits connexes de la transformation industrielle des bois tropicaux d ’ Afrique : étude fondamentale et appliquée sur les composés extractibles et les polymères lignocellulosiques Rodrigue Safou Tchiama Tchiama To cite this version : H,” 2023.
- [36] H. Number and K. Number, “Acid-insoluble lignin in wood and pulp,” *Test*, pp. 3–7, 2011.
- [37] P. H. Elchinger, P. A. Faugeras, C. Zerrouki, D. Montplaisir, F. Brouillette, and R. Zerrouki, “Tosylcellulose synthesis in aqueous medium,” *Green Chem.*, vol. 14, no. 11, pp. 3126–3131, 2012, doi: 10.1039/c2gc35592b.
- [38] M. Schwanninger and B. Hinterstoisser, “Comparison of the classical wood extraction method using a Soxhlet apparatus with an advanced extraction method,” *Holz als Roh - und Werkst.*, vol. 60, no. 5, pp. 343–346, 2002, doi: 10.1007/s00107-002-0312-2.
- [39] C. E. Rasmussen, C. K. I. Williams, G. Processes, M. I. T. Press, and M. I. Jordan, *Gaussian Processes for Machine Learning*. 2006.
- [40] W. Tinsson, *Plans d’expérience: constructions et analyses statistiques (Mathématiques et Applications, 67)*. 2011.
- [41] A. Hebbal, “Deep Gaussian Processes for the Analysis and Optimization of Complex Systems Application to Aerospace System Design A dissertation submitted by,” 2021.
- [42] T. C. Jian Qing, *Gaussian Process Regression Analysis for Functional Data*. new york, 2011.
- [43] S. M. W. S. Y. Y. W. L. X. Q. X. Z. F. Fangong, “Preparation of chemical modified lignin and its application in advanced functional materials.pdf,” *Ressources Chem. Mater.*, p. 23, 2025, doi: <https://doi.org/10.1016/j.recm.2025.100104>.
- [44] L. D. Khuong, R. Kondo, R. De Leon, T. Kim Anh, K. Shimizu, and I. Kamei, “Bioethanol production from alkaline-pretreated sugarcane bagasse by consolidated bioprocessing using *Phlebia* sp. MG-60,” *Int. Biodeterior. Biodegrad.*, vol. 88, pp. 62–68, 2014, doi: 10.1016/j.ibiod.2013.12.008.
- [45] D. M. de Carvalho, J. H. de Queiroz, and J. L. Colodette, “Assessment of alkaline pretreatment for the production of bioethanol from eucalyptus, sugarcane bagasse and sugarcane straw,” *Ind. Crops Prod.*, vol. 94, pp. 932–941, 2016, doi: 10.1016/j.indcrop.2016.09.069.
- [46] N. Ratanasumarn and P. Chitprasert, “Cosmetic potential of lignin extracts from alkaline-treated sugarcane bagasse: Optimization of extraction conditions using

- response surface methodology,” *Int. J. Biol. Macromol.*, vol. 153, pp. 138–145, 2020, doi: 10.1016/j.ijbiomac.2020.02.328.
- [47] P. Ikenyiri, F. Abowei, C. Ukpaka, and S. Amadi, “Characterization and Physicochemical Properties of Wood Sawdust in Niger Area, Nigeria,” *SSRN Electron. J.*, vol. 5, no. 3, pp. 190–197, 2020, doi: 10.2139/ssrn.3407342.
- [48] D. Smółka-Danielowska and M. Jabłońska, “Chemical and mineral composition of ashes from wood biomass combustion in domestic wood-fired furnaces,” *Int. J. Environ. Sci. Technol.*, vol. 19, no. 6, pp. 5359–5372, 2022, doi: 10.1007/s13762-021-03506-9.
- [49] M. O. Asare and M. Hejman, “Effect of tree species on the elemental composition of wood ashes and their fertilizer values on agricultural soils,” *GCB Bioenergy*, vol. 14, no. 12, pp. 1321–1335, 2022, doi: 10.1111/gcbb.13002.
- [50] C. Chen, D. Jin, X. Ouyang, L. Zhao, X. Qiu, and F. Wang, “Effect of structural characteristics on the depolymerization of lignin into phenolic monomers,” *Fuel*, vol. 223, no. November 2017, pp. 366–372, 2018, doi: 10.1016/j.fuel.2018.03.041.
- [51] C. G. Boeriu, D. Bravo, R. J. A. Gosselink, and J. E. G. Van Dam, “Characterisation of structure-dependent functional properties of lignin with infrared spectroscopy,” *Ind. Crops Prod.*, vol. 20, no. 2, pp. 205–218, 2004, doi: 10.1016/j.indcrop.2004.04.022.
- [52] O. Y. Abdelaziz and C. P. Hulteberg, “Physicochemical Characterisation of Technical Lignins for Their Potential Valorisation,” *Waste and Biomass Valorization*, vol. 8, no. 3, pp. 859–869, 2017, doi: 10.1007/s12649-016-9643-9.
- [53] O. Derkacheva and D. Sukhov, “Investigation of lignins by FTIR spectroscopy,” *Macromol. Symp.*, vol. 265, no. 1, pp. 61–68, 2008, doi: 10.1002/masy.200850507.
- [54] G. H. Delmas, B. Benjelloun-Mlayah, Y. Le Bigot, and M. Delmas, “Biolignin TM based epoxy resins,” *J. Appl. Polym. Sci.*, vol. 127, no. 3, pp. 1863–1872, 2013, doi: 10.1002/app.37921.
- [55] D. Ciolacu, F. Ciolacu, and V. I. Popa, “Amorphous cellulose - Structure and characterization,” *Cellul. Chem. Technol.*, vol. 45, no. 1–2, pp. 13–21, 2011.
- [56] M. El Hajam, N. Idrissi Kandri, A. Harrach, A. El khomsi, and A. Zerouale, “Physicochemical characterization of softwood waste ‘Cedar’ and hardwood waste ‘Mahogany’: Comparative study,” *Mater. Today Proc.*, vol. 13, pp. 803–811, 2019, doi: 10.1016/j.matpr.2019.04.043.
- [57] M. Saber, L. El Hamdaoui, M. El Moussaouiti, and M. Tabyaoui, “Extraction and Characterization of Lignin From Moroccan Thuya. Its Application As Adsorbent of Methylene Blue From Aqueous Solution,” *Cellul. Chem. Technol.*, vol. 56, no. 1–2, pp. 69–81, 2022, doi: 10.35812/CelluloseChemTechnol.2022.56.06.
- [58] A. Goudarzi, L. T. Lin, and F. K. Ko, “X-ray diffraction analysis of kraft lignins and lignin-derived carbon nanofibers,” *J. Nanotechnol. Eng. Med.*, vol. 5, no. 2, pp. 1–6, 2014, doi: 10.1115/1.4028300.
- [59] S. Kubo and J. F. Kadla, “Lignin-based carbon fibers: Effect of synthetic polymer blending on fiber properties,” *J. Polym. Environ.*, vol. 13, no. 2, pp. 97–105, 2005, doi: 10.1007/s10924-005-2941-0.

- [60] G. Cheng *et al.*, “Transition of cellulose crystalline structure and surface morphology of biomass as a function of ionic liquid pretreatment and its relation to enzymatic hydrolysis,” *Biomacromolecules*, vol. 12, no. 4, pp. 933–941, 2011, doi: 10.1021/bm101240z.
- [61] S. Laurichesse and L. Avérous, “Chemical modification of lignins: Towards biobased polymers,” *Prog. Polym. Sci.*, vol. 39, no. 7, pp. 1266–1290, 2014, doi: 10.1016/j.progpolymsci.2013.11.004.

Figures



Fig 1: Black liquor after extraction (a), after acidification (b), and after separation of the insoluble fraction of lignin in an acidic medium

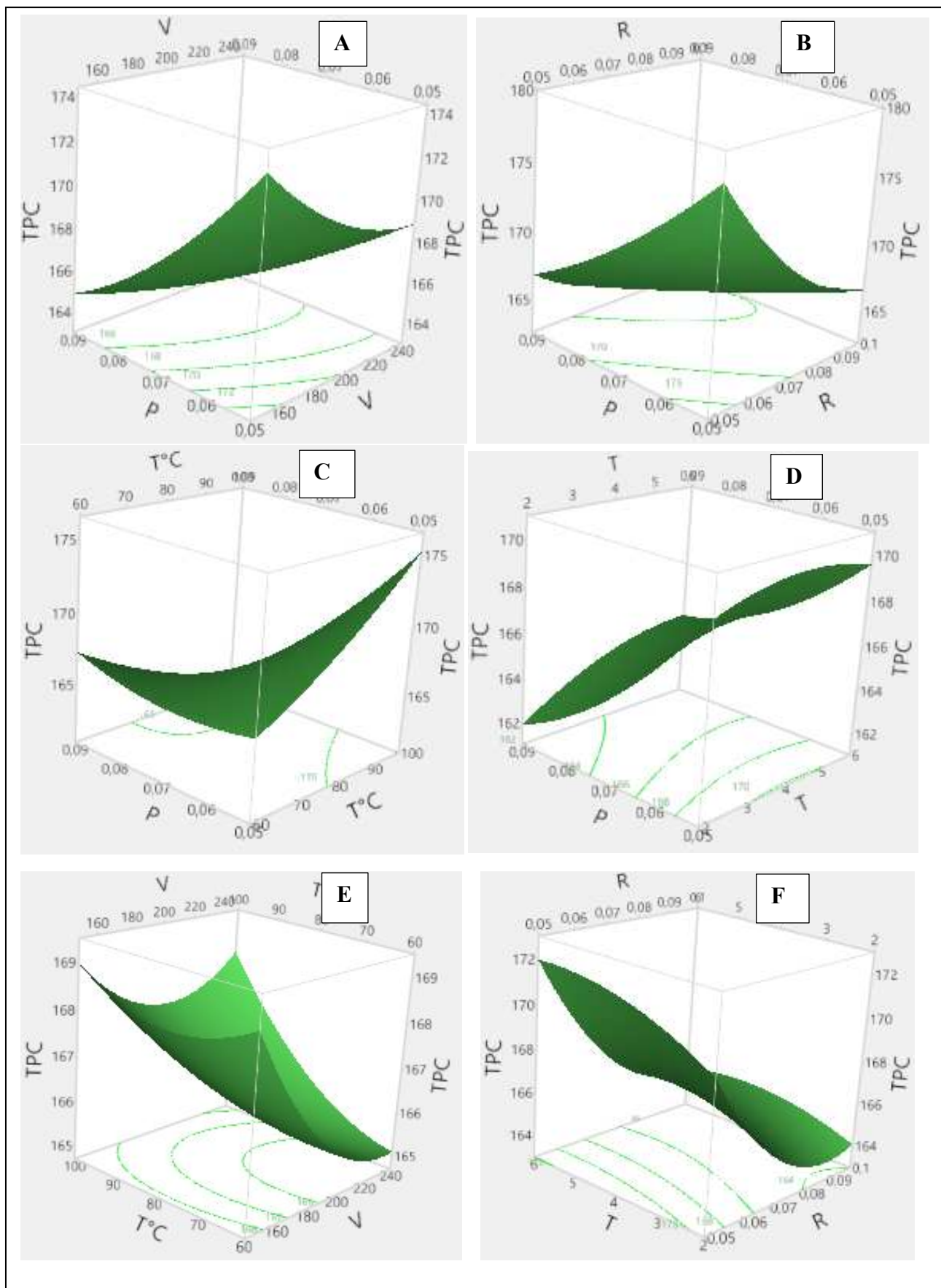


Fig 2:Response surface plots showing effects of two factors on total phenolic contents: (A)effects of NaOH concentration and agitation speed; (B) effects of NaOH concentration and solid/liquid ratio; (C) effects of NaOH concentration and extraction temperature; (D)effects of NaOH concentration and extraction time; (E) effects of agitation speed and temperature; (F) effects of solid/liquid ratio and extraction time

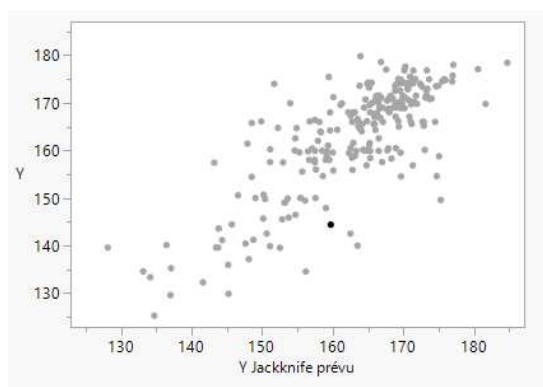


Figure 3: Parity diagram comparing experimental polyphenol yields (Y) with values predicted by Jackknife cross-validation.

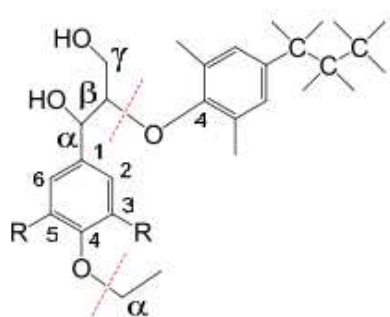


Fig 4: Cavage of an α -O-4 and β -O-4 aryl ether bond in lignin

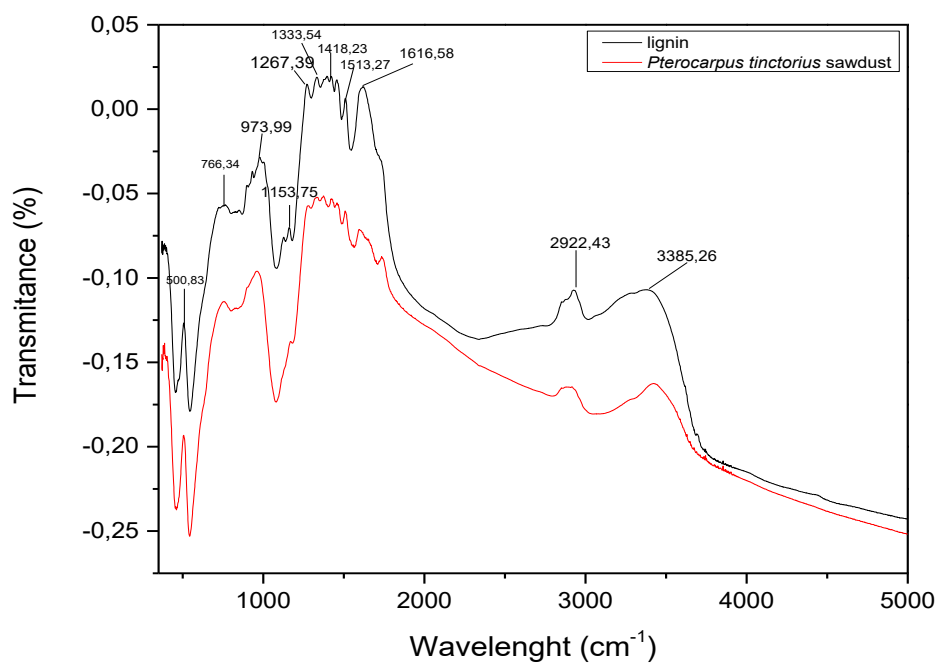


Fig 5: FTIR comparison of sawdust and lignin

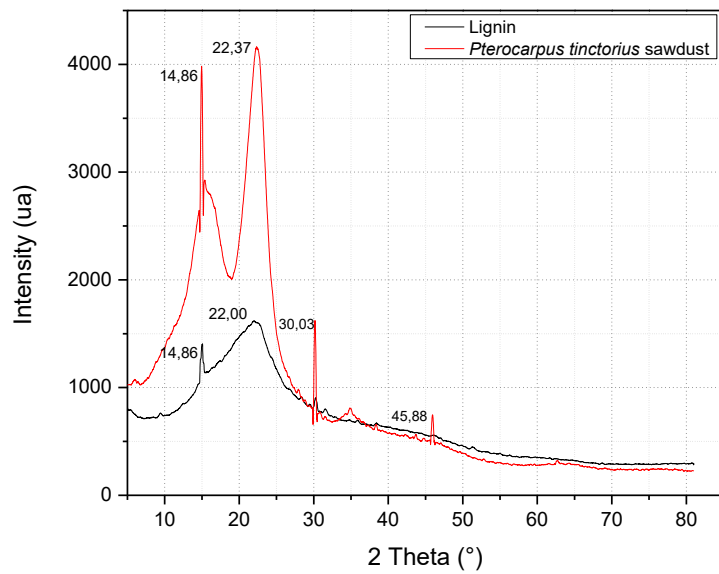


Fig 6: X-ray diffraction (XRD) patterns of *Pterocarpus tinctorius* sawdust (red curve) and lignin extract (black curve)

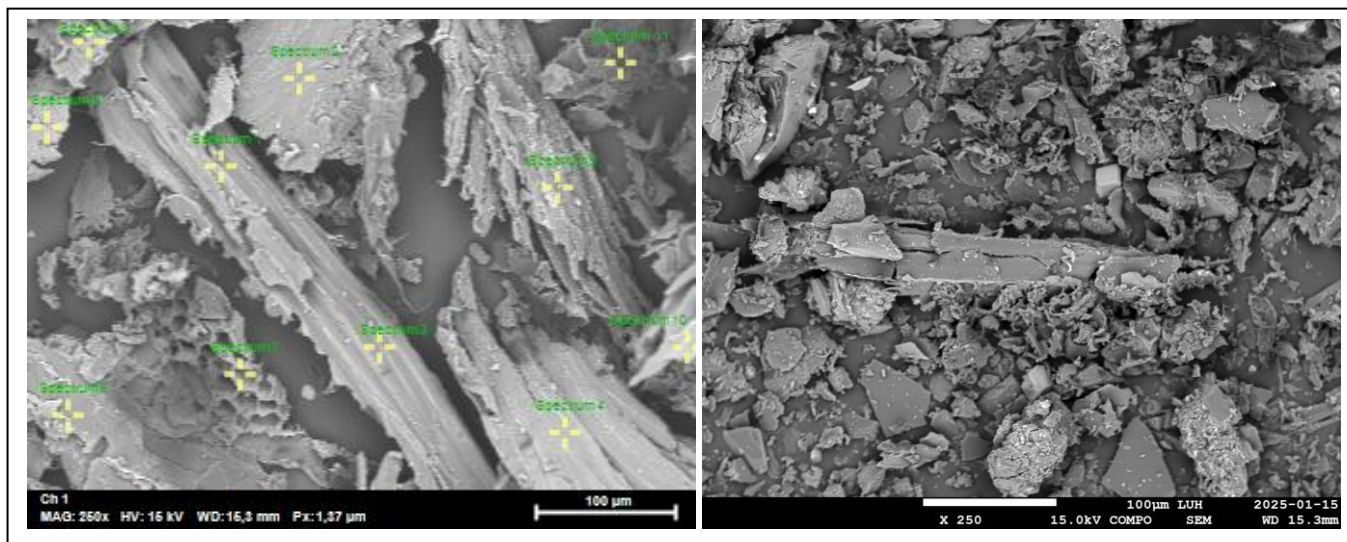


Fig 7: Morphology of Sawdust (by EDX) and lignin extract (by SEM)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.docx](#)