International Journal of Engineering and Technology Innovation, vol. 16, no. 2, 2026, pp. 211-238

Machine Learning–Driven Analysis of KOH and NaOH Performance in Biodiesel Production from Waste Frying Oil

Jehad Ahmad Abdalla Yamin[1,*], Zayed Al-Hamamre[2]

[1]Department of Mechanical Engineering, School of Engineering, The University of Jordan, Amman, Jordan

[2]Department of Chemical Engineering, School of Engineering, The University of Jordan, Amman, Jordan

Received 04 December 2025; received in revised form 19 February 2026; accepted 26 February 2026

DOI: https://doi.org/10.46604/ijeti.2026.15944

Abstract

This study aims to conduct a comparative multivariate analysis of potassium hydroxide (KOH) and sodium hydroxide (NaOH) catalysts for biodiesel production from waste frying oil using a multivariate analytical framework. A central composite rotatable design (CCRD) and response surface methodology (RSM) are integrated with machine learning (ML) techniques—random forest regression and principal component analysis (PCA)—to model, predict, and decode the complex interactions between process parameters. This proposed hybrid approach uncovers previously underexplored catalyst-specific mechanistic differences. The analysis indicates KOH's superior performance, achieving a higher mean biodiesel yield (90.7%) than NaOH (84.1%). Crucially, the ML models reveal a fundamental mechanistic divergence: the methanol-to-oil (MeOH) ratio is a dominant positive factor for KOH but a detrimental factor for NaOH. This catalyst-specific effect underpins distinct optimization strategies and confirms KOH as the essential catalyst for efficient biodiesel synthesis from waste frying oil.

Keywords: biodiesel, waste frying oil, Alkali catalyst, response surface methodology, random forest method

1. Introduction

The escalating global demand for energy, coupled with the depletion of fossil fuel reserves and mounting environmental concerns, intensifies the search for sustainable alternatives. Recent data indicate that global oil demand grew by approximately 0.7–0.8% year-on-year in 2025, reaching around 103–104 mb/d, with projections for modest increases through 2030, influenced by economic and geopolitical factors [1]. The transport sector contributes roughly 29–32% of global energy-related CO_2 emissions, with over 90% of its energy coming from fossil fuels. It also accounts for approximately 16–25% of total greenhouse gas emissions, based on recent assessments (2024–2025) [2].

In response, biofuels have emerged as promising substitutes. Biodiesel, in particular, has gained prominence for its renewable, biodegradable, and nontoxic properties, as well as its higher flash point and lower greenhouse gas emissions compared to petroleum diesel [3]. The International Energy Agency highlights that biofuels could play a pivotal role in achieving net-zero emissions by 2050. They may account for up to 27% of the transport fuel supply in hard-to-decarbonize sectors such as road transportation, shipping, and aviation [4].

Biodiesel, primarily composed of fatty acid methyl esters (FAME), is derived from various feedstocks, including edible and non-edible oils, as well as waste materials such as fats, oils, and grease sourced from restaurants and the food industry [5-6]. The production process typically involves transesterification, a straightforward and efficient reaction in which triglycerides in the feedstock react with alcohol (typically methanol or ethanol) in the presence of a catalyst to yield biodiesel and glycerol

* Corresponding author. E-mail address: yamin@ju.edu.jo

as by-products [7-8]. The use of waste feedstocks in biodiesel production offers both waste management and cost reduction, while also delivering environmental advantages, notably a lower carbon footprint [9-10].

Challenges persist, including feedstock variability, the need for extensive pretreatment to remove impurities such as water and particulates, and soap formation during processing [5]. Additionally, biodiesel's oxygen-rich composition enhances combustion efficiency in engines, reducing emissions of hydrocarbons, particulates, and carbon monoxide. By-products such as glycerol can be valorized into additives to bolster the industry's viability further [11].

Catalysts play a pivotal role in transesterification, with homogeneous alkali catalysts, such as KOH and NaOH, commonly employed due to their effectiveness in promoting the reaction [5, 12]. Key process parameters, including the alcohol-to-oil ratio, catalyst concentration, reaction time, and temperature, directly influence biodiesel yield and efficiency [12]. To address the complexities and optimize these variables while minimizing time, energy, and costs, traditional approaches such as response surface methodology (RSM) have been widely used. RSM, a mathematical tool for experimental design and modeling, enables the evaluation of multiple variables and their interactions with fewer experiments, thereby enhancing yields and sustainability [13].

To optimize the complex, multivariable transesterification process, statistical and computational tools are essential. RSM has been widely used to model and optimize biodiesel production, enabling the evaluation of variable interactions with fewer experiments [10]. For instance, studies have successfully applied RSM to optimize yields from Neem oil and waste frying oil [14]. Recent advances have integrated machine learning (ML) with traditional RSM to enhance predictive capability and insight. Bajwa et al. [15] combined RSM with an artificial neural network (ANN) to optimize biodiesel from mixed feedstocks, while Yamin et al. [16] used RSM to model the effects of key parameters on yield for different catalysts. Similarly, [17] used RSM with a Box-Behnken design to assess the impact of various parameters, including molar ratio (4:1–8:1), catalyst concentration (0.5–1.5% *w*/w NaOH), reaction temperature (40–60℃), and reaction duration (30–120 min) on the yield of biodiesel.

However, despite their widespread use, a critical research gap persists. Most studies treat KOH and NaOH as broadly interchangeable alkali catalysts, focusing on empirical optimization for a single catalyst. A comprehensive, comparative analysis that leverages advanced data-driven techniques to systematically investigate the fundamental mechanistic differences between KOH and NaOH is lacking. Specifically, a need exists for an integrated framework combining RSM with robust ML methods, such as random forest regression (RFR) and principal component analysis (PCA), to identify optimal conditions. This approach elucidates the underlying catalyst-specific parameter interactions and sensitivity landscapes that govern yield efficiency and stability.

For instance, RSM-based designs, such as Box-Behnken, were applied to optimize biodiesel production from Neem oil, achieving high yields while being cost-effective compared to other methods [14]. The study reported obtaining 96.36% Neem Biodiesel yield with an alcohol-to-oil molar ratio of 7.1:1 and 1.02 wt.% KOH catalyst, at a reaction temperature of 41℃ and a reaction duration of 98 minutes. The RSM model further predicts a maximum yield of 97.14% under similar conditions.

Reference [15] optimized biodiesel extraction from a combination of waste frying oil (WFO) and sesame seed oil (SSO) through RSM and ANN. They used an ANN model for prediction, while RSM was used for process optimization. The results indicate that the highest biodiesel yield of 94% was achieved at a stirring speed of 350 rpm, a reaction time of 3 min, a catalyst concentration of 1.05 w/w, and a 10:1 methanol-to-oil ratio.

Reference [16] also employed RSM to model the effects of reaction time, catalyst concentration, and oil-to-methanol ratio on biodiesel production from waste frying oil using three different catalysts. The study reported that the best yield is obtained at lower methanol-oil-to ratios (nearly 0.02) and higher catalyst concentrations (10%); however, reaction time varies across catalysts.

Reference [18] applied RSM to optimize the parameters for solar-assisted biodiesel production from linseed oil using Taguchi's L27 orthogonal array. Their results showed that Taguchi's approach yields a maximum biodiesel yield of 89.14%, while the RSM model predicts a slightly higher yield of 91.9%.

Despite the widespread use of KOH and NaOH catalysts, comprehensive comparative analyses across diverse operating conditions remain limited, and their relative performance in biodiesel production remains underexplored. This study aims to fill this gap by conducting a systematic, comparative investigation of KOH and NaOH catalysts in biodiesel production from waste frying oil. The specific objectives are:

(1) To optimize the transesterification process for both KOH and NaOH using a central composite rotatable design (CCRD) and RSM.

(2) To employ RFR and PCA to quantitatively assess and compare the influence and complex interactions of process parameters (catalyst loading, temperature, time, methanol-to-oil ratio) on yield for each catalyst.

(3) To identify and elucidate the optimal operational windows and mechanistic landscapes that define the performance and stability of each catalyst, thereby providing a data-driven basis for catalyst selection and process intensification.

2. Novelty and Research Gap

While RSM and ML techniques have been increasingly applied to optimize biodiesel production from waste frying oils, most prior investigations have focused on single-catalyst systems or treated KOH and NaOH as broadly interchangeable homogeneous alkali catalysts. These studies emphasized empirical yield maximization without probing catalyst-specific mechanistic behaviors under identical experimental frameworks.

For instance, studies employing RSM or ANN have successfully modeled transesterification parameters for waste oils using either KOH or NaOH individually, achieving high yields but often lacking direct comparative analyses that can reveal fundamental differences in parameter sensitivity and interaction landscapes. Similarly, integrations of RSM with machine learning have enhanced predictive accuracy and process understanding in mixed-feedstock or heterogeneous catalysis contexts; however, few studies have extended such hybrid approaches to systematically investigate catalyst-dependent effects in homogeneous base-catalyzed systems.

The present study addresses this critical gap by conducting a systematic, side-by-side comparative investigation of KOH and NaOH performance in the transesterification of the same low-FFA waste frying oil feedstock, utilizing a shared CCRD to ensure methodological consistency. The novelty of this work lies in integrating conventional RSM-based quadratic modeling and Analysis of Variance (ANOVA) with advanced machine learning tools. Specifically, RFR for quantitative feature importance ranking and PCA for decoupling multivariate latent structures.

This integrated approach enables the identification of catalyst-specific dominance patterns (e.g., methanol-to-oil ratio as the primary positive driver for KOH versus temperature as the overriding sensitivity factor for NaOH) and the elucidation of distinct catalytic response surfaces. This multifaceted data-driven framework not only confirms KOH's superior yield robustness and stability but also uncovers underlying mechanistic divergences, such as differential tolerances to excess methanol and thermal side reactions, which have been underexplored in comparative homogeneous catalysis studies.

By providing catalyst-tailored and actionable optimization strategies and deeper insights into parameter interactions, this work advances beyond conventional empirical optimization toward a more mechanistic and predictive understanding of alkali-catalyzed biodiesel synthesis from waste oils. These findings have implications for industrial process intensification, catalyst selection, and sustainable biofuel production.

3. Materials and Methods

WFO, originally derived from sunflower oil, is collected daily from the main restaurant at the University of Jordan in Amman, where it is used exclusively for frying beefsteak, French fries, and falafel. To minimize degradation and maintain consistent feedstock quality, the oil is renewed after each use. Before use, the WFO is pre-treated to remove moisture and impurities. Methanol (analytical grade, 99.5%), KOH, NaOH, anhydrous sodium sulfate (Na_2SO_4), hexane (GC grade), and calcium chloride ($CaCl_2$) are procured from Merck® (Darmstadt, Germany). All chemicals are used as received without further purification.

The transesterification reaction is conducted in a 250 mL three-necked round-bottom flask equipped with a reflux condenser, a digital thermometer (range: 0–150℃), and a magnetic stirrer. The reactor is immersed in a water bath maintained at a controlled temperature using a 1.5 kW heating plate with an adjustable stirring speed (up to 800 rpm) to ensure homogeneous mixing throughout the reaction. The instruments are used to characterize WFO and biodiesel properties in Table 1.

Table 1 Equipment used for physical and chemical property analysis

Equipment	Model / Type	Manufacturer	Purpose
Digital Density Meter	DMA 35	PAAR	Density measurement
Viscometer	TSML 210045	Fungilab	Kinematic viscosity
Adiabatic Bomb Calorimeter	Autobomb 3A3021	Galenkamk	Gross calorific value determination
Magnetic Stirrer	MR Hei-standard	Héldolph Instruments GmbH & Co. KG	Homogeneous mixing during reaction
Rotary Evaporator	IKA RV 3 pro V	IKA	Removal of residual methanol and water
Separating Funnel	500 mL	Borosilicate glass	Phase separation post-reaction

Collected WFO is filtered to remove food residues and then heated at 105–110℃ until a constant weight is achieved, thereby eliminating residual water. The pre-treated oil is then subjected to base-catalyzed transesterification in a 250 mL three-necked round-bottom flask using methanolic NaOH or KOH (1.0–2.0 wt.% relative to the mass of WFO). The methanol-to-oil ratio ranges from 6:1 to 12:1 (v/v), and the methanol-catalyst solution is preheated to the desired reaction temperature. Reactions are conducted under atmospheric pressure with constant stirring at 800 rpm for 0.5–5 h at 40, 50, or 60℃.

Following the reaction, the mixture is transferred to a separating funnel and allowed to settle for 8 h, forming a lower glycerol-rich layer and an upper biodiesel layer. The glycerol phase is removed, and the biodiesel is washed three times with warm distilled water (50℃) until neutral. It is then treated with 2–3 drops of glacial acetic acid to neutralize residual alkalinity and decompose soaps. The product is dried over anhydrous Na_2SO_4 and purified by rotary evaporation at 80℃ under reduced pressure.

The yield of biodiesel is determined gravimetrically as the ratio of the mass of the produced biodiesel to the mass of the initial oil feedstock. Final fuel properties, including density, kinematic viscosity, acid value, flash point, cloud/pour points, and calorific value, are measured in triplicate according to ASTM standards.

The main independent variables that influence biodiesel yield are the molar ratio (x_1) and reaction time (x_2). The variables significantly affect biodiesel yield, so it is essential to determine the optimal methanol-to-oil molar ratio and reaction time to minimize time and energy losses.

To increase and optimize the process variables and achieve the highest biodiesel yield, the RSM approach is used. The CCRD of a three-factor, five-level design is employed to conduct experimental trials, with the input variables coded as 2, 1, 0, -1, and -2, as shown in Table 2.

Table 2 Independent variables and coded levels

Biodiesel Input	Coded Stages (or variable levels)				
	-2	-1	0	1	2
Reaction Time (hr)	1	2	3	4	5
Reaction Temperature (℃)	-	40	50	60	-
Catalyst Loading (%)	-	1.0	1.5	2.0	-
Oil-to-Methanol Ratio	1:15	1:25	1:35	1:45	-

Ranges are selected to encompass practical industrial conditions while capturing non-linear effects (quadratic terms) and catalyst-specific sensitivities, based on preliminary trials and established transesterification kinetics. These ranges are essential for capturing non-linear effects and catalyst-specific sensitivities. Additionally, the chosen levels are aligned with established kinetics of base-catalyzed transesterification and with ranges frequently reported in the literature for similar feedstocks and heterogeneous/homogeneous catalysts. This alignment ensures a balance between industrial relevance and the statistical requirements for a reliable second-order response surface model (typically requiring axial points in central composite design (CCD) or appropriate coverage in BBD).

The general form of the second-order (quadratic) polynomial model employed in RSM to fit the experimental biodiesel yield data obtained from the CCRD is presented below. This second-order model is selected because it efficiently captures both linear and curvilinear (non-linear) effects, as well as two-factor interactions among the process variables. These interactions are critical in transesterification kinetics due to phenomena such as mass transfer limitations, equilibrium shifts, and catalyst deactivation at extreme levels.

The CCRD is specifically chosen for its rotatability, uniform precision in estimating response variance across the design space, and ability to fit a full quadratic model with a reasonable number of experimental runs. Multiple regression analysis via ANOVA is used to estimate the coefficients and assess model adequacy. High coefficients of determination (R^2 and adjusted R^2) together with a non-significant lack-of-fit, confirm the suitability of the quadratic approximation for describing and optimizing the transesterification process. The governing equation is expressed as follows:

$$y = \beta_o + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{1 \le i \le j}^{k} \beta_{ij} x_i x_j + \varepsilon \quad (1)$$

where k is the number of variables, β_o is the constant term, β_i represents the coefficients of the linear parameters, xi represents the variables, ε is the residual associated with the experiments, and β_{ii} represents the coefficients of the quadratic terms. To estimate the parameters in Eq. (1), the experimental design ensures that all studied variables are evaluated at least three factor levels.

Response surface analysis is performed by fixing one independent variable at a constant level (typically the center point, low, or high value) while simultaneously varying the other two variables across their experimental ranges. This approach generates a series of three-dimensional (3D) response surface plots and corresponding two-dimensional (2D) contour plots that visually illustrate the main effects, quadratic behavior (curvature), and two-factor interactions on biodiesel yield. By examining these plots, the nature of the interactions—such as synergistic, antagonistic, or near-additive effects—and the location of potential optima or saddle points within the design space can be readily interpreted. Such conditional slicing of the multidimensional response surface is a standard practice in RSM when dealing with three or more factors, as it facilitates understanding of how process variables jointly influence the response under practical constraints.

ANOVA is subsequently applied to the fitted quadratic model to statistically evaluate the significance of the linear, quadratic, and interaction terms, as well as the overall model adequacy. The ANOVA results identify which variables and their

interactions exert statistically significant effects on the response ($p < 0.05$). Furthermore, the analysis confirms the absence of a significant lack of fit and supports the model's reliability for prediction and optimization.

The CCRD method employed in this study systematically varies four key independent process parameters known to govern the kinetics, equilibrium, and side reactions in alkali-catalyzed transesterification of WFO: catalyst loading (A, wt.%), reaction temperature (B,℃), reaction time (C, min), and methanol-to-oil volumetric ratio (D, v/v; equivalent to molar ratios of approximately 4.5:1 to 9.5:1).

Catalyst loading (1.0–2.0 wt.%) is selected based on its role in generating methoxide ions (CH_3O^-) from methanol via dissociation of KOH or NaOH. These ions initiate nucleophilic attack on triglyceride carbonyl groups to form FAME. Insufficient loading limits the reaction rate, while excess promotes saponification (soap formation from free fatty acids or ester hydrolysis). This increases emulsion stability, and viscosity, and phase separation difficulties are amplified in waste oils with residual moisture or FFAs. The tested range ensures coverage of optimal to supra-optimal levels, allowing quadratic modeling of non-linear yield declines due to side reactions.

Reaction temperature (40–60℃) critically influences Arrhenius-dependent kinetics: higher temperatures accelerate transesterification by increasing molecular collisions and reducing oil viscosity, thereby improving mass transfer. However, exceeding approximately 60℃ (near methanol's boiling point of 64.7℃) risks methanol evaporation, accelerated saponification/hydrolysis, reverse transesterification, and glycerol solubility issues that drive equilibrium backward. This narrow window was chosen to probe the thermal sensitivity threshold observed in low-FFA WFO systems.

Reaction time (60–300 min, coded as 1–5 h in the initial design) captures the progression from kinetically limited early stages to equilibrium or side-reaction dominance. Shorter times favor efficiency in industrial settings, while prolonged exposure exacerbates soap/gel formation, particularly with NaOH.

Methanol-to-oil ratio (0.2–0.4 v/v) drives the reversible transesterification toward completion per Le Chatelier's principle by shifting the equilibrium and diluting glycerol (which inhibits the forward reaction). The stoichiometric ratio is approximately 3:1 molar, but an excess (typically 6:1–12:1) is used to compensate for mass-transfer limitations in heterogeneous phases. However, very high ratios enhance glycerol solubility in the methanol-rich phase, promoting the reverse reaction and emulsion formation—effects that are catalyst-dependent, as evidenced by RFR results showing positive dominance for KOH but greater sensitivity for NaOH.

Ranges are determined from preliminary single-factor experiments and from literature on similar sunflower-derived WFOs. This ensures rotatability ($\alpha = \pm 1.414$) for uniform precision across the design space and enables reliable estimation of quadratic/interaction terms via ANOVA.

4. Results and Discussion

The results of this study are presented in the following sections. The discussion is organized into three main parts: first, the properties of the biodiesel obtained are discussed; second, the effects of operating conditions are analyzed; and finally, the results of the statistical analysis, e.g., RSM and ANOVA modelling, are presented.

4.1 Oil and biodiesels characteristics

The waste vegetable oil (WVO) feedstock is characterized by a high content of unsaturated fatty acids, primarily linoleic (57.3 wt.%) and oleic (28.1 wt.%) acids. Other saturated fatty acids are palmitic (9.2 w/w%) and stearic (5.6 w/w%) acids. As detailed in Table 3, the transesterification process successfully converts this WVO into biodiesel that meets the critical international standards ASTM D6751 [6] and EN 14214 [19-20]. The process yields a substantial ~70% reduction in kinematic viscosity, from 19.24 mm^2/s to 5.86 mm^2/s, achieving the requisite range for proper engine operation.

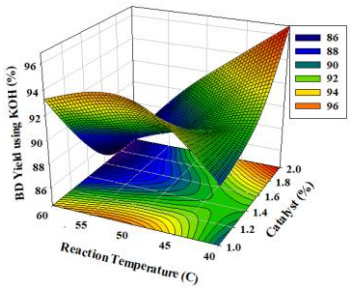
Table 3 Physical and chemical properties of Biodiesel

Properties	Units	ASTM-PD D6751 [6]	EN-BD: EN 14214 [19]	WVO	BD
Kinematic viscosity ± 0.01 (at 40 °C)	mm s^{-2}	1.9-6.0	3.5-5	19.24	5.86
Density ± 0.0001 (at 15 °C)	g cm^{-3}	-	0.86-0.9	0.920	0.877
Acid value	mg_{KOH} g^{-1}	0.8	0.5	1.36	0.22-0.45
Iodine value	g I_2/100g	-	120	140.14	25.36
Water content	% by weight	0.05	0.05	0.076	0.053
Peroxide value	mEq O_2 kg^{-1} oil	-	-	13.015	41.17
Flash point	°C	130	120	250	150
Ash content	% by weight.	0.02	<0.02	0.097	0.08
Saponification value	mg_{KOH} g^{-1}	-	-	220	170.1
Refractive Index	-	-	-	1.358	1.24
Heating value	MJ kg^{-1}	-	35 [20]	31.83	35.66

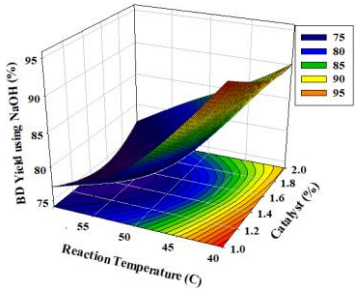
Furthermore, the biodiesel exhibits a low acid value (0.22-0.45 mg KOH/g), confirming minimal presence of free fatty acids and soaps. It also has a high heating value of 35.66 MJ/kg, a safe flash point of 150℃, and an iodine value of 25.36 g I_2/100g, which are well within specified limits and indicate good oxidative stability. In conclusion, the optimized process effectively transforms low-value waste oil into high-quality, specification-compliant biodiesel.

4.2 Oil-to-biodiesel yield

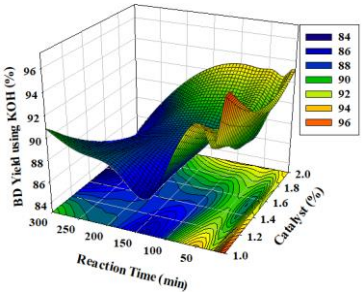
In this study, the transesterification of WFO to biodiesel is investigated through a comprehensive parametric analysis. While conventional 2D plots reveal the effects of individual variables, they often obscure nonlinear interactions among process parameters. To overcome this limitation, the results are visualized using 3D response surface plots (Fig. 1) that capture the complex interplay among reaction variables. The z-axis represents biodiesel yield (%), while the x- and y-axes represent two independent variables; the surfaces are generated by interpolation of the experimental data. These plots highlight regions of peak performance, plateaus, and declines caused by side reactions, such as saponification, providing an invaluable map for process optimization.



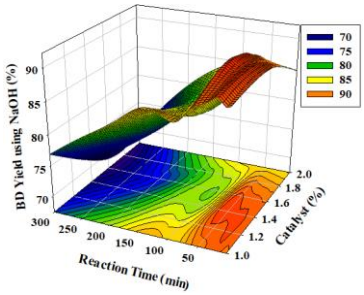
(a) Response to Catalyst/Temperature with KOH



(b) Response to Catalyst/Temperature with NaOH

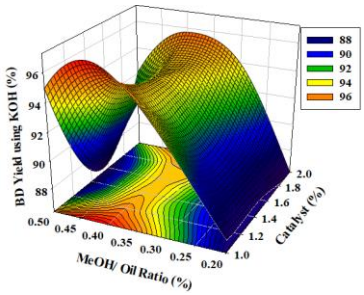


(c) Response to Catalyst/Time with KOH

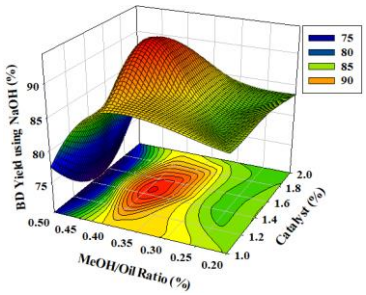


(d) Response to Catalyst/Time with NaOH

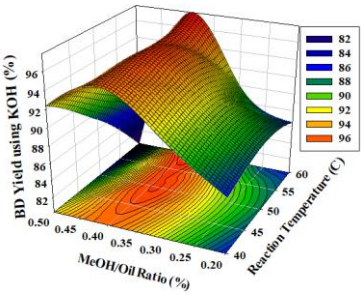
Fig. 1 3-D surface plots of biodiesel yield as a function of the rest of the input variables



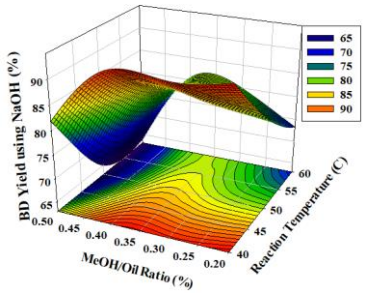
(e) Response to Catalyst/ MeOH Oil Ratio with KOH



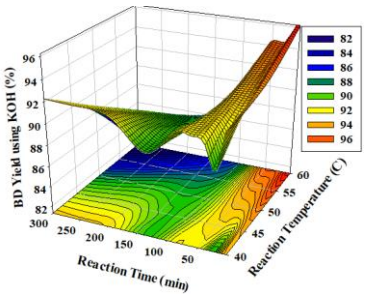
(f) Response to Catalyst/ MeOH Oil Ratio with NaOH



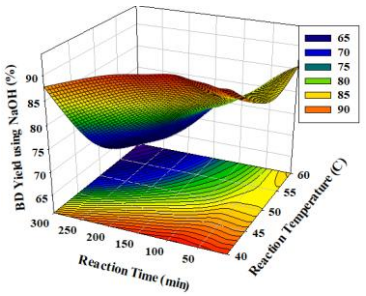
(g) Response to MeOH Oil Ratio /Temperature with KOH



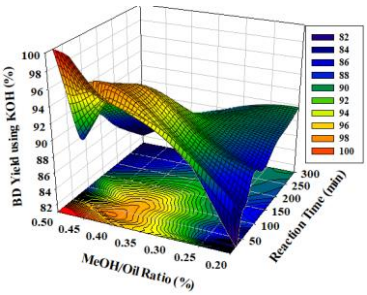
(h) Response to MeOH Oil Ratio /Temperature with NaOH



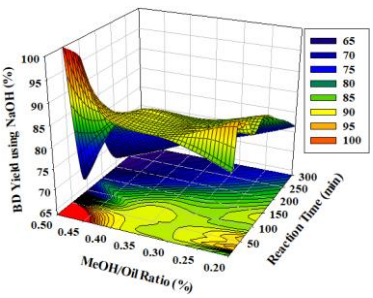
(i) Response to Time/Temperature with KOH



(j) Response to Time/Temperature with NaOH



(k) Response to MeOH Oil Ratio/Time with KOH



(l) Response to MeOH Oil Ratio/Time with NaOH

Fig. 1 3-D surface plots of biodiesel yield as a function of the rest of the input variables (continued)

4.3 Effect of catalyst loading and reaction temperature

3D surface plots of biodiesel yield as a function of catalyst loading and reaction temperature at a fixed reaction time, and MeOH/oil ratio (Figs. 1(a)-(b)) reveal a well-defined optimal region. At a constant MeOH/oil ratio, increasing the temperature from 40℃ to 50℃ significantly increases the yield across all catalyst concentrations, due to improved reaction kinetics.

The surface shows a distinct peak at approximately 50℃ and a catalyst loading of 1.5% w/v, where yields reach ≥ 98% (Fig. 1(a)). This represents the ideal balance, where the transesterification kinetics are maximized without triggering excessive saponification side reactions.

Beyond this optimum, a sharp decline is observed. At 60℃ (Fig. 1(b)), the yield surface collapses, particularly at higher catalyst concentrations (e.g., 2% w/v). This is attributed to the accelerated hydrolysis of esters and neutralization of free fatty acids (FFA, AV ~1.0 mg KOH/g) into soap, facilitated by the high-energy environment near methanol's boiling point (64.7℃).

The resulting soaps and gels increase emulsion stability, complicating phase separation and creating a pronounced "cliff" in the yield landscape. This identifies 50℃ as the critical temperature threshold for this system, beyond which side reactions dominate [21].

In contrast to the well-defined optimal region observed with KOH (Fig. 1(a)), the NaOH-catalyzed surfaces (Fig. 1(b)) exhibit a sharper yield collapse at 60℃, particularly at higher catalyst loadings (≥1.5% w/v), due to accelerated saponification and soap formation that destabilize the emulsion and hinder phase separation [22]. This highlights NaOH's greater sensitivity to thermal conditions compared to KOH.

4.4 Effect of catalyst concentration and time

biodiesel yield versus reaction time and KOH concentration (1–2% w/v) at a MeOH/oil ratio of 0.2 and 40℃ (Fig. 1(c)) shows a dome-shaped response surface. At 1% w/v, yield climbs steadily to 92% by 300 min. Increasing the catalyst to 1.5% w/v accelerates the reaction, reaching ~94% in 120 min with a broader high-yield region due to increased methoxide availability.

At 2% w/v, the yield decreases significantly, with values dropping to 85–90% across all reaction times. Excess catalyst promotes soap formation via FFA neutralization and ester hydrolysis, increasing mixture viscosity and impeding separation. The peak at 1.5% w/v (<120 min) represents the optimal balance between catalytic acceleration and the minimization of side reactions, consistent with the literature on alkali-catalyzed transesterification [22].

Unlike the broader high-yield plateau seen with KOH at 1.5% loading (Fig. 1(c)), NaOH performance (Fig. 1(d)) shows a narrower optimal window and significant yield decline at prolonged times (>120 min), reflecting its greater propensity for soap/gel formation and reverse transesterification under extended reaction durations.

4.5 Effect of reaction temperature and time

3D surface plots of biodiesel yield as a function of reaction time and temperature at a fixed MeOH/oil ratio and catalyst loading (Figs. 1(i)-(j)) show a pronounced peak in the mid-temperature range. At a MeOH/oil volumetric ratio of 0.4 (9.5 molar) and 1% w/v KOH, the surface at 40℃ rises gradually, achieving ~92% yield after 5 min and plateauing near 98% beyond 80 min. This indicates a kinetically limited process.

Increasing the temperature to 50℃ sharply elevates the surface, with yields ≥98% reached within 5–60 min, forming a broad, stable plateau (Fig. 1(i)). This suggests optimal kinetic enhancement. At 60℃, however, the surface initially peaks but declines after 60 min, dropping below 95% by 300 min (Fig. 1(j)). This decline confirms the temperature sensitivity observed in Figs. 1(a) and 1(b), where bubble formation and saponification hinder mass transfer and promote the reverse reaction. These observations align with Arrhenius kinetics, balancing the activation energies of transesterification and competing hydrolysis.

While KOH maintains a stable, high-yield plateau at 50℃ over a wide time range (Fig. 1(i)), NaOH surfaces (Fig. 1(j)) show a more pronounced decline after 60 min at elevated temperatures. The highlights NaOH's greater thermal sensitivity and its susceptibility to saponification-driven side reactions, which limit long-term stability.

4.6 Effect of MeOH/Oil ratio and time

The 3D surface showing biodiesel yield versus reaction time and MeOH/oil ratio (0.2–0.4 v/v; 4.5–9.5 molar) at 50℃ and 2% w/v KOH reveals a ridge along the intermediate ratio (Figs. 1(k)-(l)). At 0.2 v/v, the yield reaches only ~85% by 40 min, reflecting stoichiometric limitations. Increasing the ratio to 0.3 v/v raises the yield by ~13%, peaking at 98% around 40 min and maintaining a high-yield plateau, demonstrating Le Chatelier’s principle.

Further increasing the ratio to 0.4 v/v is detrimental over time. The surface dips after 10 min, declining below 90% by

350 min and forming a deep "valley" (Fig. 1(l)). This inversion results from the enhanced solubility of glycerol in the methanol-rich phase, which drives the reverse reaction, as well as the formation of increased foam and stable emulsions. Optimal yields (>95%) are confined to a narrow band around 0.3–0.35 v/v for times under 100 min.

In stark contrast to KOH, where intermediate methanol ratios sustain high yields over time (Fig. 1(k)), NaOH surfaces (Fig. 1(l)) exhibit a deeper, earlier-forming valley at higher methanol ratios (>0.35 v/v). This indicates greater intolerance to excess methanol, driven by enhanced glycerol solubility, emulsion stabilization, and promotion of the reverse reaction.

4.7 Effect of catalyst type (KOH vs. NaOH) and other interactions

Comparative 3D surfaces of biodiesel yields under varying conditions (Figs. 1(c)-(f)) reveal significant differences in the performance and stability of KOH- and NaOH-catalyzed reactions. These plots highlight how the methanol-to-oil ratio and catalyst loading influence reaction performance.

At a lower methanol-to-oil ratio (0.2 v/v), the difference between the catalysts is less pronounced but still notable. With a loading of 1% w/v (Figs. 1(c), (e)), NaOH exhibits faster initial kinetics. It reaches 87% yield at 5 min compared to 76% for KOH, and achieves a slightly higher final yield (~98% vs. ~95%) by 300 min. This suggests that NaOH's stronger basicity provides an initial kinetic advantage.

However, this advantage reverses at a higher methanol-to-oil ratio (0.4 v/v). At 1.5% w/v catalyst loading (Figs. 1(d), (f)), a critical divergence in stability is observed. While both catalysts achieve high yields (>97%) within the first 5 minutes, NaOH's performance plummets to 68% by 300 minutes. In stark contrast, KOH maintains a stable yield of ~95% throughout the reaction. This decline in NaOH is attributed to its greater tendency to form soap and gel via saponification side reactions. These side products complicate the separation of glycerol from biodiesel and promote the reverse reaction, an effect exacerbated by excess methanol. KOH’s ability to provide a more stable, high-yield (>90%) region under these conditions makes it the preferred catalyst for transesterifying waste feedstocks with moderate FFA content.

The interactions of other parameters, such as methanol-to-oil ratio and temperature, at a fixed catalyst loading (Fig. 1(g), h), further illustrate the system's complexity, confirming that the optimal operational window is highly dependent on the careful balance of all variables, with catalyst choice being a critical factor.

While the preceding sections have emphasized the superior yield stability and robustness of KOH (mean 90.7%, lower variability, fewer outliers), a parallel examination of NaOH-catalyzed transesterification reveals distinct performance characteristics and underlying limitations that underscore the catalyst-specific nature of the process.

Under identical CCRD conditions, NaOH achieved a lower mean biodiesel yield of 84.1% with markedly higher variability (standard deviation 8.902 vs. 5.046 for KOH) and nine low-yield outliers clustered in the 60% range (Fig. 2(a)), indicating inconsistent performance and a tendency toward failure modes.

The 3D response surfaces for NaOH (Figs. 1(b), (d), (f), (h), (j), (l)) exhibit narrower optimal regions and pronounced declines compared to KOH counterparts: yields peak sharply at ~50℃ and low-to-moderate methanol ratios (0.25–0.3 v/v) but collapse rapidly with increasing temperature (>50℃) or prolonged reaction time (>60 min), forming steep "cliffs" and "valleys" attributed to accelerated saponification and emulsion/gel formation. This thermal sensitivity is confirmed by random forest feature importance. Temperature emerges as the dominant factor for NaOH (~1.10 mean decrease in R^2). This value far exceeds the methanol ratio (~0.40). In stark contrast, the methanol ratio is the leading factor for KOH (~0.60).

ANOVA further supports this divergence: for NaOH, catalyst loading (A) and its quadratic term (A^2) are non-significant ($p > 0.44$), while temperature (B), time (C), methanol ratio (D), and their interactions (e.g., BC $F = 712.32$, $p < 0.0001$) dominate, reflecting irreversible side reactions triggered by excess thermal energy or time rather than insufficient catalyst

availability. PCA reinforces this mechanistic distinction, with PC2 (methanol efficacy axis) showing strong positive loading for KOH yield but negative for NaOH yield (+0.75 vs. –0.69), highlighting NaOH's intolerance to higher methanol levels due to enhanced glycerol solubility, reverse transesterification, and stable emulsions.

These observations align with earlier literature reporting that KOH is generally preferable to NaOH in waste-oil systems, primarily due to reduced soap formation, easier glycerol separation, and greater tolerance for moderate impurities or excess reagents. However, prior studies often relied on single-catalyst optimization or empirical comparisons without integrated multivariate ML tools to decode latent interactions and parameter dominance. The novelty of the present work lies in this systematic, head-to-head application of the same CCRD framework, combined with random forest and PCA. This approach not only confirms KOH's advantages but also elucidates the mechanistic reasons behind NaOH's inferior performance (e.g., heightened saponification propensity, leading to temperature/time dominance, and methanol antagonism).

4.8 Comparison with heterogeneous alkaline catalysts

The homogeneous alkali-catalyzed transesterification of WFO is the focus of the present study, whereas heterogeneous alkaline catalysts offer potential advantages, including reduced side reactions (e.g., saponification) and enhanced process sustainability through catalyst reusability and simplified separation. To contextualize the results, a comparison is made between the performance of the homogeneous KOH system and that of KOH-loaded zeolite tuff catalysts derived from Jordanian natural zeolite, as reported in prior investigations on similar waste sunflower-based WFO feedstocks [23]. These heterogeneous catalysts were prepared via a two-step impregnation method using acid-treated zeolite tuff (TZT) loaded with KOH (1–4 M KOH/TZT), achieving high activity without the need for high-temperature calcination.

The feedstocks in both investigations are comparable: the homogeneous system used WFO with an acid value of 1 mg KOH/g and an iodine value of 219 g I_2/100 g, while the heterogeneous studies employed waste sunflower oil with acid values of ~2 mg KOH/g and iodine values of ~217 g I_2/100 g. Both feedstocks exhibit similar fatty acid compositions, dominated by linoleic (~57%) and oleic (~28%) acids. Table 4 summarizes the key operating conditions and outcomes.

Table 4 Comparison between homogeneous KOH and heterogeneous KOH/zeolite tuff

Parameter	Homogeneous KOH (Present Study)	Heterogeneous KOH/Zeolite Tuff [23]
Optimum methanol-to-oil ratio	9.5:1 (0.4 v/v)	11.5:1 (0.5 v/v)
Catalyst loading	1.0% w/v (0.83 wt.%)	6.4 wt.%
Reaction temperature	50℃	50℃
Reaction time	20–40 min	2 h
Stirring speed	800 rpm	800 rpm
Catalyst particle size	Not applicable (homogeneous system)	335 µm
Biodiesel yield/conversion	98% yield	96.7% yield / ~100% conversion
Soap formation tendency	Notable at higher catalyst loadings or prolonged reaction times	Reduced (K^+ leaching ~2%)
Catalyst reusability	Not reusable	Reusable (up to 90% yield after regeneration)
Biodiesel viscosity (40℃)	5.86 $mm^2\ s^{-1}$	5.63 $mm^2\ s^{-1}$
Biodiesel density (15℃)	0.877 $g\ cm^{-3}$	0.884 $g\ cm^{-3}$
Biodiesel iodine value	25.36 g I_2 100 g^{-1}	16.55 g I_2 100 g^{-1}

The homogeneous KOH system yields slightly higher (98% vs. 96.7%) in significantly shorter reaction times (20–40 min vs. 2 h), attributable to superior mass transfer and the absence of diffusion limitations in the liquid phase. It also necessitates lower methanol ratios (9.5:1 vs. 11.5:1) and catalyst loadings (0.83% w/w vs. 6.4% w/w), rendering it more efficient under mild conditions for low-FFA feedstocks such as the WFO examined here. In contrast, the heterogeneous KOH/zeolite tuff requires higher methanol and catalyst amounts to overcome internal and external diffusion barriers. Parametric studies indicate optimal performance at 800 rpm stirring and a particle size of 335 µm to minimize mass-transfer resistance.

Despite these requirements, the heterogeneous system demonstrates superiority in mitigating side reactions. Soap formation is minimized, with K^+ leaching (~2%) being lower than in the homogeneous case, where excess KOH promotes saponification at extended times or higher loadings. This observation aligns with the literature on zeolite-based catalysts, where yields of 80–95% are commonly reported under optimized conditions, with enhanced reusability (e.g., 80–90% yield over 3 cycles post-regeneration) [23].

The biodiesels produced by both systems conform to EN 14214 and ASTM D6751. They exhibit similar viscosities (5.86 vs. 5.63 mm^2/s) and densities (0.877 vs. 0.884 g/cm^3), although the heterogeneous biodiesel has a lower iodine value (16.55 vs. 25.36 g I_2/100 g), suggesting improved oxidative stability. These comparisons indicate that homogeneous KOH offers kinetic advantages for batch processes, whereas heterogeneous KOH/zeolite tuff is better suited for continuous, eco-friendly operations. Direct side-by-side experiments under identical conditions are recommended for future investigations to quantify economic and environmental benefits.

4.9 Optimal conditions and implications

Integrating all 3D surface insights, the optimal conditions for this WFO feedstock are a MeOH/oil ratio of 0.4 v/v, 1% w/v KOH, 50℃, and a reaction time of 20–40 min. These conditions yield 98–99.5% while minimizing side reactions. This outcome represents a significant advancement in WFO-to-biodiesel yield, concurrently minimizing energy consumption, catalyst loading, and processing time, while eliminating the need for pre-treatment. For scientific rigor, the current results are most pertinently benchmarked against studies using low free fatty acid (FFA) feedstocks under single-step base-catalyzed transesterification, thereby excluding the confounding effects of acid pre-esterification or non-standard alcohols.

Reference [24] employed Afia waste frying oil from a commercial source (e.g., Afia brand sunflower oil as reported in [24]), compositionally comparable to the feedstock used in the current study, i.e., sunflower-derived WFO, reported a maximum yield of 97.76% with 0.4 wt.% KOH at 60℃ over 180 minutes. Although their yields are comparable, the methodology used in the present study reduced reaction time by 88.3% (from 180 to 20–40 minutes), despite a 2.5-fold increase in catalyst loading. This acceleration arises from three synergistic process attributes: (i) negligible residual FFA (<1 mg KOH/g) due to daily WFO replacement in restaurant settings, enabling direct base catalysis; (ii) enhanced mass transfer via vigorous stirring at 800 rpm; and (iii) precise water content regulation during pre-heating, preventing catalyst deactivation. These factors demonstrate that elevated catalyst concentrations can enhance efficiency when complemented by high feedstock quality and optimized mixing.

Moreover, Belkhanchi et al. [25] reported a 99.3% yield from sunflower WFO using 1% w/w KOH and a 6:1 molar ratio—chemically and compositionally comparable to feedstock used in this study—but required >60 min of reaction time, compared to 20–40 min as in the current study. Reference [26] achieved a 98% yield with 1% w/w KOH at 50℃ using goat tallow, validating the robustness of 1% KOH across diverse triglyceride sources; however, their process required 120 min, three times as long as the current work. They also reported that the properties of the produced biodiesel, such as flash point, cloud point, pour point, viscosity, and density, are within the standard range and can be used as a biofuel. Reference [27] observed a maximum yield of 87% from used vegetable oil under similar temperatures (55–60℃), yet still fell short of the plateau of the present work, indicating that even well-optimized systems outside the present work's parameter space fail to match the efficiency.

In contrast, Selvam et al. [17] used Sterculia foetida seed oil (low-FFA, non-edible) and achieved a 89.2% yield at 1.0 wt.% NaOH at 50℃ over 60 minutes, employing isopropanol to improve phase separation. Despite meeting ASTM D6751 specifications, their yield trailed that of the present research work by 8.8 percentage points. The reliance on conventional, cost-effective methanol used by the current work underscores the industrial viability of this approach, particularly where solvent

recovery and economic constraints are critical. Notably, the efficacy of 1.0 wt.% KOH in this study surpasses the 0.4 wt.% optimum reported by Hamed et al. [25], challenging the conventional hypothesis that minimal catalyst use is inherently optimal. Kinetic analysis indicates that saponification is the rate-limiting step only above 1.5 wt.% KOH, thus expanding the operational range and affirming 1.0 wt.% as an optimal compromise between yield and side-reaction suppression.

Conversely, studies by Jamil et al. [28], utilizing high-FFA feedstocks (>20%), required sequential acid esterification and base transesterification. Jamil et al. achieved a 96.2% yield after 120 minutes of esterification and 180 minutes of transesterification (total 300 minutes), reducing FFA from 22.14% to 0.84%. Similarly, Córdova et al. [29] explored ethanol substitution in bovine fat transesterification to reduce carbon footprint—a contribution to sustainability metrics—, but it is orthogonal to the focus of the current study on kinetic optimization and yield enhancement using methanol and WFO. The findings of the present work advance process intensification, offering a scalable framework for industrial biodiesel production.

The physicochemical properties of the resultant biodiesel further validate its quality: kinematic viscosity (4.92 mm^2/s) and iodine value (25.36 g I_2/100 g) comply with JUS EN14214, matching or exceeding values reported by Selvam et al. (4.8 cSt, acid value 0.3 mg KOH/g) and Kurniasih et al. (2.51 mm^2/s, 20.12 g I_2/100 g). The density (0.877 g/cm^3) also meets specifications, confirming high purity after washing and drying without requiring distillation.

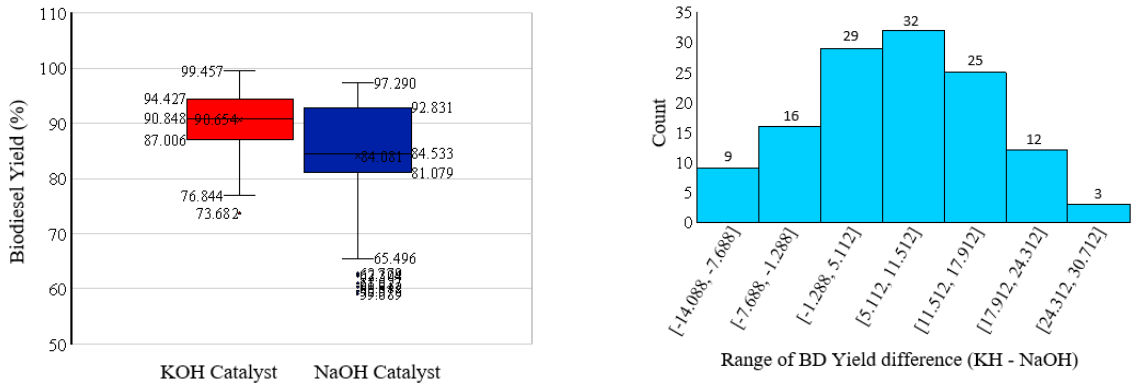
The reliance on conventional, cost-effective methanol (contributing ~$0.3–0.5 per gallon of biodiesel based on current market prices [30]) used by the current work underscores the industrial viability of this approach, particularly where solvent recovery and economic constraints are critical, thanks to methanol's low cost, high reactivity, and established industrial recovery infrastructure.

4.10 Statistical analysis and modeling of yield performance

To transform these observations into a predictive, statistically validated framework, RSM based on a CCD is employed. Subsequent analyses, including random forest feature importance, ANOVA, and PCA, are conducted to quantify variable significance, identify dominant interactions, and extract latent patterns governing yield performance. Through this integrated approach, the optimal operating conditions are identified, and deeper mechanistic insights into catalyst-specific behavior under identical operating conditions are obtained.

4.10.1 Range and distribution of biodiesel yields (*KOH vs. NaOH*)

Referring to Fig. 2(a), the box plot indicates that the KOH catalyst not only has a higher median yield (about 90.8% vs 84.5%), but also less variability (lower standard deviation and interquartile range of KOH 5.046/7.316 compared with 8.902/11.415 for NaOH) compared with NaOH. This provides a strong initial indication that KOH outperforms NaOH in both average yield and consistency.



(a) Biodiesel range for both catalysts (b) Frequency distribution of yield

Fig. 2 Range of biodiesel yield for both catalysts

Furthermore, KOH has one low outlier, while NaOH presents nine low outliers clustered in the low 60s, reinforcing its heavier lower tail. This disparity in outlier frequency and severity highlights NaOH's greater susceptibility to process variability and failure modes under the tested CCRD conditions. The clustered low-yield outliers for NaOH (around 60%) likely arise from localized triggering of irreversible side reactions—such as accelerated saponification, ester hydrolysis, and stable emulsion/gel formation—particularly when combinations of elevated temperature (>50℃), prolonged reaction time (>60 min), or higher methanol ratios push the system beyond its narrow stability window. This is evidenced by the steeper yield declines in NaOH response surfaces (Figs. 1(b), (d), (f), (j), (l)).

In contrast, KOH's single outlier and overall tighter distribution reflect its superior robustness, with better tolerance to excess methanol and thermal stress, reduced soap formation, and easier phase separation. These patterns are consistent with literature reports on alkali-catalyzed transesterification of waste oils containing moderate residual FFAs or moisture. These observations quantitatively support the random forest feature importance (Fig. 3), where temperature and time dominate NaOH performance (higher risk of side-reaction cascades), while methanol ratio drives KOH more predictably. Consequently, NaOH's heavier lower tail and outlier clustering indicate inconsistent reliability for industrial-scale application with waste frying oil, underscoring the need for tighter control of operating parameters or preference for KOH to minimize yield variability and process failures.

The figures above show that KOH yields are higher and more consistent. NaOH can achieve high yields, but its performance is less predictable, with several low-performing outliers that skew its distribution downward. The box plot in Fig. 2(a) clearly illustrates this contrast: KOH exhibits a higher median yield (~90.8%), narrower interquartile range, and markedly lower standard deviation (5.046 vs. 8.902 for NaOH), reflecting greater process robustness and repeatability across the 126 matched CCRD runs. In contrast, NaOH's distribution displays a heavier lower tail, driven by nine low-yield outliers clustered around 60%. This indicate sporadic but severe performance drops likely triggered by combinations of unfavorable parameter settings (e.g., elevated temperature, prolonged time, or excess methanol). Such conditions promote saponification, emulsion stabilization, and reverse transesterification—side reactions to which NaOH is more susceptible, as confirmed by the steeper yield collapses in its 3D response surfaces (Fig. 1).

The frequency distribution of yield differences (KOH – NaOH) in Fig. 2(b), centered positively at a mean difference of ~6.57%, further substantiates KOH's consistent superiority, with rare instances of NaOH outperforming KOH. These patterns align with the paired t-test results ($p = 6.881 \times 10^{-13}$). Random forest feature importance ranking temperature and time as dominant for NaOH (indicating higher sensitivity to side-reaction cascades), and PCA findings showing NaOH's confinement to a low-methanol, low-severity regime. Overall, the greater consistency and reduced outlier risk with KOH underscore its preference for reliable, industrial-scale biodiesel production from waste frying oil, where feedstock variability and operational fluctuations are common.

Fig. 2(b) shows the distribution of yield differences (BD yield; KOH - NaOH) for each experiment. The distribution is centered above zero, indicating that KOH generally produces higher yields than NaOH. On average, KOH yields are approximately 6.57 units higher than those of NaOH, with a median difference of 6.63 units. The range indicates that, in some cases, NaOH can outperform KOH, although this is less common.

In summary, KOH is generally more efficient than NaOH for BD yield, with statistically significantly higher and more consistent yields. The most critical factors for maximizing yield differ between the two catalysts, so optimization strategies should be tailored accordingly.

Next, a statistical comparison of the yields was conducted to compare them directly. A paired t-test is performed on the BD yields from KOH and NaOH under the same experimental conditions. The test results are presented in Table 5. A paired t-test is selected for this purpose because it accounts for the matched-pair design, in which each pair consists of yields from

the same set of conditions but differing only in catalyst type (KOH vs. NaOH), thereby controlling for variability due to feedstock quality, equipment, and environmental factors. The test assesses whether there is a statistically significant difference in mean BD yields between the two catalysts at a 95% confidence level ($p < 0.05$). The null hypothesis assuming no difference, whereas the alternative hypothesis indicating superior performance of one catalyst. Assumptions of normality and paired differences are verified prior to analysis. The results, including the calculated t-value, degrees of freedom, p-value, and mean difference, are summarized in Table 5 and provide quantitative evidence supporting conclusions on catalyst efficacy in the transesterification of waste oil.

Table 5 t-test results for both catalysts

	BD Yield using KOH	BD Yield using NaOH
Mean	90.653	84.080
Variance	25.463	98.058
Observations	126	126
Pearson Correlation	0.386	-
Hypothesized Mean Difference	0	-
df	125	-
t Stat	8.009	-
P(T<=t) one-tail	3.440E-13	-
t Critical one-tail	1.6571	-
P(T<=t) two-tail	6.881E-13	-
t Critical two-tail	1.979	-

The paired t-test is applied to directly compare biodiesel yields obtained using KOH versus NaOH under identical experimental conditions across 126 matched pairs. As summarized in Table 5, the mean yield with KOH is 90.653%, compared to 84.080% with NaOH, resulting in a mean difference of approximately 6.573%. The calculated t-statistic is 8.009 with 125 degrees of freedom. The two-tailed p-value was 6.881×10^{-13}, which is substantially lower than the conventional significance threshold of 0.05 (and even 0.001), providing strong evidence to reject the null hypothesis of no difference in mean yields between the two catalysts. This highly significant result indicates that KOH consistently produces higher biodiesel yields than NaOH under the tested conditions. The positive t-statistics further confirm the directionality of the difference, with KOH exhibiting superior catalytic performance. These findings are consistent with the literature, which reported higher basicity and transesterification activity of potassium-based catalysts compared to sodium-based counterparts in waste oil systems, particularly when free fatty acids are present.

4.10.2 Random forest analysis

To quantitatively assess the relative influence of each process parameter on biodiesel yield, a random forest regression analysis is employed. This ensemble machine learning algorithm operates by training multiple decision trees. For this regression task, the model predicts the continuous biodiesel yield value. The algorithm's output includes a metric known as "feature importance," which quantifies the mean decrease in the model's predictive accuracy (R^2) when a given variable's values are randomly permuted. A larger decrease indicates that the variable carries a stronger predictive signal and is a more critical factor in determining yield.

The results of this analysis are presented in Fig. 3, which demonstrates the influence of process variables on biodiesel yield for each catalyst. The bar height represents the mean decrease in model accuracy (R^2) when a variable is randomized, indicating its predictive power.

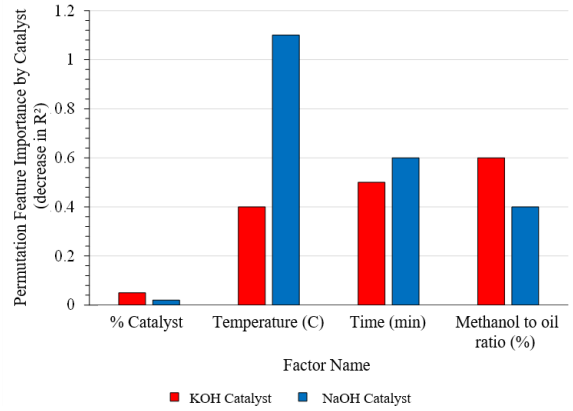


Fig. 3 Influence of factors for both catalysts using Random Forest regression analysis

The approximate mean decrease in R^2 for each variable is as follows:

- For KOH: Methanol-to-Oil Ratio (~0.60), Reaction Time (~0.50), Temperature (~0.40), Catalyst Loading (~0.05).
- For NaOH: Temperature (~1.10), Reaction Time (~0.60), Methanol-to-Oil Ratio (~0.40), Catalyst Loading (~0.02).

The feature importance profiles for the two catalysts are strikingly different, providing a statistical foundation for the trends observed in the experimental 3D yield surfaces (Fig. 1). For the KOH-catalyzed reaction, the Methanol-to-Oil ratio emerged as the most critical feature, followed closely by reaction time. This finding aligns perfectly with the experimental observations, where the yield landscape was susceptible to this ratio, forming a distinct optimal "ridge" at intermediate values (~0.3–0.35 v/v) where yield was penalized by both stoichiometric limitations at low values and reverse reactions and emulsions caused by excess methanol (Figs. 1(k)-(l)). In stark contrast, for the NaOH-catalyzed reaction, temperature is the dominant feature by a significant margin, with reaction time also being critical. This quantitatively confirms the high thermal sensitivity observed in the 3D plots, where NaOH yields are optimal within a narrow temperature range but decline with time at elevated temperatures. This is due to the catalyst's greater propensity for accelerated saponification and gel formation (Fig. 1(f)).

A critical insight from Fig. 3 is the relatively lower feature importance for Catalyst Loading in both systems. It is essential to interpret this correctly: this does not imply that catalyst concentration is unimportant. Rather, the low relative importance reflects that, within the selected experimental ranges, variations in catalyst loading produce comparatively smaller changes in biodiesel yield compared to the dominant effects of methanol-to-oil ratio, reaction temperature, and time.

This behavior is typical in base-catalyzed transesterification when the catalyst concentration is already sufficient to drive the reaction toward near-equilibrium conversion. Increasing the concentration further increases yield, diminishing returns or even slight declines due to mass transfer limitations, soap formation, or promotion of the reverse reaction. In contrast, temperature and alcohol ratio exert stronger nonlinear influences across the design space, leading to higher importance scores in the feature analysis (e.g., permutation importance or SHAP values derived from the fitted quadratic models).

Consequently, while catalyst loading remains a practically relevant parameter for cost, separation, and environmental considerations, its optimization margin appears narrower under the investigated conditions. This observation aligns with numerous literature reports on homogeneous alkali-catalyzed systems using waste oils, where catalyst dosage often operates in a plateau region beyond a threshold value.

The experimental data (Figs. 1(a)-(b)) unequivocally show that yield collapses catastrophically outside a narrow optimal range (~1.5% w/v) due to soap formation. Instead, the feature importance score suggests that within the tested range (1-2% w/v), the model found that variations in the other three parameters had a larger marginal effect on yield variance. Once a sufficient catalyst amount is present, fine-tuning temperature, time, and methanol ratio becomes more impactful for maximizing yield.

In summary, the Random Forest analysis provides the following practical guidance for optimization:

(1) If using KOH**:** Prioritize tuning the methanol-to-oil ratio and reaction time.
(2) If using NaOH**:** Meticulously control the reaction temperature and time.
(3) For both catalysts: Catalyst loading must be optimized to its specific value to avoid severe yield loss, even though its variable importance score is lower. The choice of catalyst fundamentally shifts the priority of the other process variables.

4.10.3 RSM model and its accuracy

As stated above in Eq. (1), the quadratic model is used for prediction and analysis. The general form is:

$$y = x_o + x_A A + x_B B + x_{AB} AB + x_{AA} A^2 + x_{BB} B^2 \tag{2}$$

where x represents the coefficient of the parameter, e.g., x_A represents the coefficient of the parameter A in the equation. The parameters A and B represent the model inputs, i.e., reaction time and oil-to-methanol ratio. x_{AB} represents the coefficient for the interaction term between A and B. The final equations in terms of actual factors are shown below:

$$\begin{aligned} Yield(KOH) = & -24.67549 + 11.75480A + 2.34311B + 0.204328C + 255.01378D \\ & - 0.490587AB - 0.006538AC - 18.03766AD - 0.002857BC - 1.29950BD \\ & - 0.18336CD + 5.80454A^2 - 0.010552B^2 - 4.48613C^2 - 186.98684D^2 \end{aligned} \tag{3}$$

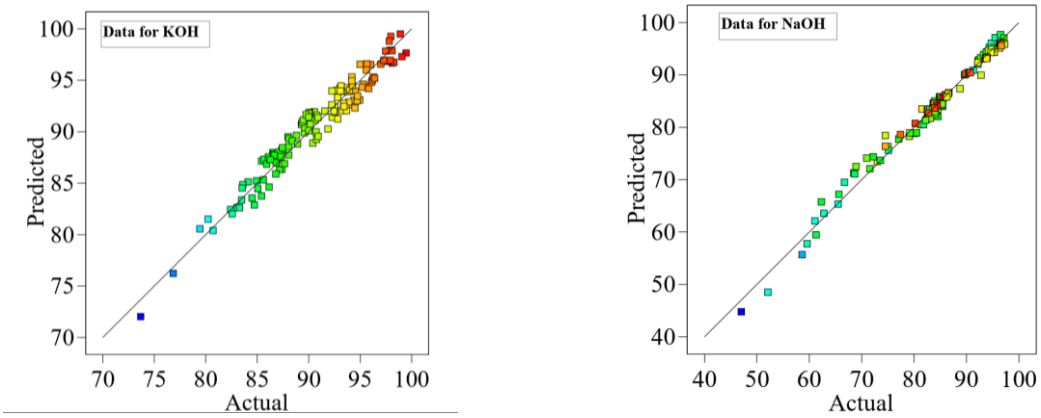
$$\begin{aligned} Yield(NaOH) = & 192.72746 - 17.34720A - 4.27773B + 0.296705C + 139.89173D \\ & + 0.346537AB + 0.011869AC - 3.21870AD - 0.004844BC - 1.21306BD \\ & - 0.265651CD - 0.430886A^2 + 0.038447B^2 - 0.000173C^2 - 115.58566D^2 \end{aligned} \tag{4}$$

These actual factor equations can be used to predict the response for a given set of factor levels. The levels should be expressed in the original units of each factor. Avoid using this equation to determine the relative contributions of each factor, as the coefficients are scaled to match each factor's units, and the intercept is not at the center of the design space. The accuracy of the equation is shown in Table 6. Taking the yield process as an example for clarification. There is around a 0.5% difference between the R^2 (Adj) and the R^2, indicating a fair agreement for both equations.

Table 6 Model accuracy details

	R^2	Adjusted R^2	Adequate Precision
KOH	0.9525	0.9465	68.2132
NaOH	0.9854	0.9836	115.9399

The signal-to-noise ratio is measured with adequate precision. Adequate precision is assessed by comparing the average prediction error to the range of expected values at the design points. The ratio should be greater than 4. With these models, a suitable signal-to-noise ratio of ~68 for KOH and ~115 for NaOH is achieved. This suggests that additional analysis can be done with those models. Additionally, Fig. 4 illustrates the correlation between the experimental and predicted biodiesel yields. The linear relationship between experimental and predicted yield values indicates strong agreement and high predictability, confirming the model's reliability and robustness for predicting biodiesel yield under varying process conditions.



(a) With KOH catalyst (b) With NaOH catalyst

Fig. 4 Correlation between the predicted and experimental values of biodiesel yield

4.10.4 ANOVA analysis of the model

Table **7** Analysis of Variance for biodiesel yield using KOH

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	3031.72	14	216.55	158.98	< 0.0001	sig
A- Catalyst (%)	18.12	1	18.12	13.30	0.0004	-
Source	Sum of Squares	df	Mean Square	F-value	p-value	
B-Temperature (°C)	518.39	1	518.39	380.57	< 0.0001	-
C-Time (min)	75.25	1	75.25	55.25	< 0.0001	-
D-Methanol to oil ratio (%)	19.62	1	19.62	14.41	0.0002	-
AB	122.67	1	122.67	90.05	< 0.0001	-
AC	1.61	1	1.61	1.18	0.2799	-
AD	26.01	1	26.01	19.09	< 0.0001	-
BC	476.31	1	476.31	349.67	< 0.0001	-
BD	125.55	1	125.55	92.17	< 0.0001	-
CD	286.53	1	286.53	210.35	< 0.0001	-
A^2	41.51	1	41.51	30.47	< 0.0001	-
B^2	24.97	1	24.97	18.33	< 0.0001	-
C^2	0.1171	1	0.1171	0.0860	0.7699	-
D^2	222.36	1	222.36	163.24	< 0.0001	-
Residual	151.20	111	1.36	-	-	-
Cor Total	3182.92	125	-	-	-	-

Table 8 Analysis of Variance for Biodiesel Yield Using NaOH

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	12929.24	14	923.52	531.67	< 0.0001	sig
A- Catalyst (%)	1.01	1	1.01	0.5791	0.4483	-
B-Temperature (°C)	4789.09	1	4789.09	2757.09	< 0.0001	-
C-Time (min)	2079.29	1	2079.29	1197.05	< 0.0001	-
D-Methanol to oil ratio (%)	2079.48	1	2079.48	1197.17	< 0.0001	-
AB	61.20	1	61.20	35.23	< 0.0001	-
AC	5.29	1	5.29	3.05	0.0837	-
AD	0.8281	1	0.8281	0.4767	0.4914	-
BC	1237.30	1	1237.30	712.32	< 0.0001	-
BD	102.24	1	102.24	58.86	< 0.0001	-
CD	560.17	1	560.17	322.49	< 0.0001	-
A^2	0.2274	1	0.2274	0.1309	0.7182	-
B^2	329.12	1	329.12	189.48	< 0.0001	-
C^2	166.97	1	166.97	96.12	< 0.0001	-
D^2	84.54	1	84.54	48.67	< 0.0001	-
Residual	191.07	110	1.74	-	-	-
Cor Total	13120.31	124	-	-	-	-

The ANOVA test results for the significance of the terms in the equation are presented in Tables 7 and 8. The model F-values of (~159 for KOH and ~532 for NaOH) indicate that the both models are statistically significant. There is only a 0.01% probability that such a F-value occurs due to noise.

P-values < 0.0500 indicate model terms are significant. In this case, both models are essential. Values > 0.1000 indicate the model terms are not significant. If many model terms are nonsignificant (other than those necessary to maintain the hierarchy), reducing the model may improve it. Regarding KOH, A, B, C, D, AB, AD, BC, BD, CD, A^2, B^2, D^2 are significant model terms, while for NaOH, B, C, D, AB, BC, BD, CD, B^2, C^2, D^2 are significant model terms.

Therefore, the ANOVA confirms that temperature (B) and its quadratic term (B^2) are among the most significant factors for both catalysts ($P < 0.0001$), with the highest F-values observed for NaOH ($F = 2757.09$). This underscores that temperature has the greatest single influence on yield, particularly for NaOH, whose performance is susceptible to thermal degradation.

The interaction between temperature and time (BC) is also highly significant for both catalysts ($P < 0.0001$), supporting the 3D observation that higher temperatures reduce the time required to reach peak yield—but only up to a point. Beyond 60°C, the quadratic decline (B^2) dominates, reflecting thermodynamic instability and the dominance of side reactions. Moreover, the analysis confirms that MeOH/oil ratio (D) and its quadratic term (D^2) are highly significant for both catalysts ($P < 0.0001$).

For KOH, the AD interaction (catalyst × ratio) is also significant ($P < 0.0001$), indicating that the effect of methanol ratio depends on catalyst level — consistent with the 3D plot showing that at 2% KOH, the 0.4 v/v valley is more pronounced. For NaOH, AD is not significant ($P = 0.4914$), indicating that catalyst concentration has a negligible effect on the methanol ratio's impact on yield and further evidence that NaOH operates under fundamentally different constraints.

In addition, for KOH, catalyst loading (A) and its quadratic term (A^2) are highly significant ($P < 0.0001$), confirming the existence of an optimum. For NaOH, however, A and A^2 are non-significant ($P > 0.44$), meaning that increasing NaOH from 1% to 2% has no measurable effect on yield, despite the 3D plots showing a dramatic drop in stability over time. This paradox is resolved by recognizing that NaOH's failure is not due to an insufficient catalyst, but rather to irreversible emulsification and reverse reactions triggered by specific temperature/time combinations. This phenomenon is invisible to simple concentration scans but is fully captured in interaction terms (AB, BC, BD, CD).

4.10.5 PCA: Decoupling multivariate interactions and catalytic landscapes

In summary, for KOH, the biodiesel yield depends on catalyst dose (A) and its quadratic term (A^2), both highly significant ($P < 0.0001$, $F = 13.30$ and 30.47, respectively; Table 7), indicating the existence of a distinct optimal loading (~1.0–1.5 wt.%) beyond which side reactions such as saponification reduce yield. This is further supported by the significant AD interaction (catalyst loading × methanol ratio, $P < 0.0001$), reflecting that methanol efficacy is modulated by catalyst concentration in the KOH system. At the same time, for NaOH, A and A^2 are non-significant ($P = 0.4483$ and 0.7182; Table 8). This means that variations in NaOH concentration within the tested range (1.0–2.0 wt.%) exert no measurable independent effect on yield, and the biodiesel yield is governed entirely by B (temperature), C (time), D (methanol ratio), and their interactions.

To transcend conventional bivariate analysis and elucidate the latent structure governing the transesterification process, PCA is applied to the whole experimental dataset (n = 126 runs). This multivariate technique projects correlated process variables and outcomes on a new orthogonal basis, revealing dominant underlying patterns that are not apparent in univariate or bivariate analysis. This approach provides a holistic framework to interpret the significant interactions identified by ANOVA and the optimal regions mapped by RSM. As shown in Fig. 5, the first four PCs are retained for interpretation, collectively accounting for 84.3% of the total systemic variance (PC1: 36.78%, PC2: 22.98%, PC3: 16.88%, PC4: 13.86%).

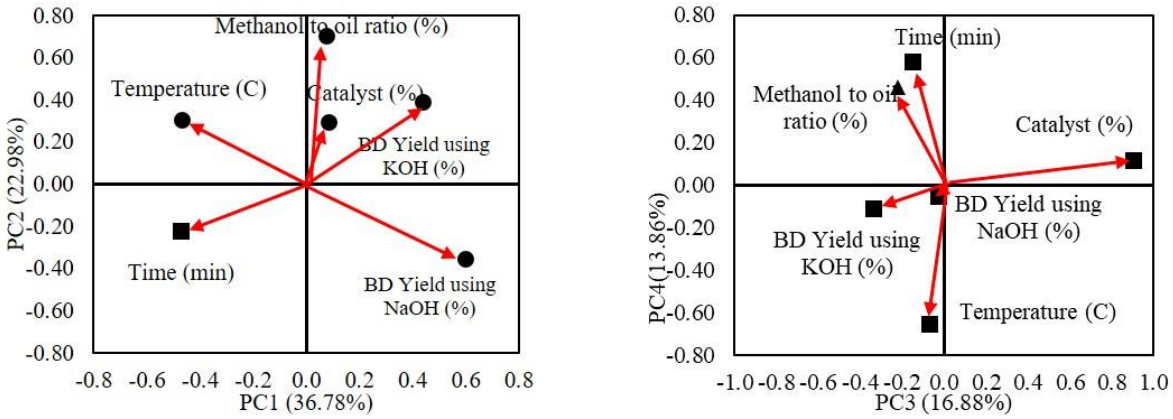


Fig. 5 PCA Loading plots for all PCs

PC1 (*36.78% variance*) – *The reaction severity axis*

PC1 is dominated by strong positive loadings for biodiesel yield with NaOH (+0.78) and KOH (+0.72), whereas reaction temperature (–0.65) and time (–0.61) show strong negative loadings. This inverse correlation clearly establishes that optimal yield is achieved under milder conditions, lower temperatures, and shorter durations rather than under harsh conditions. This statistical analysis validates the yield decline observed at 60℃ in the 3D response surfaces (Figs. 1(i)-(j)) and aligns with the highly significant BC (temperature × time) interaction term in the ANOVA (F-value = 349.67, P < 0.0001 for KOH). The preeminence of PC1 highlights that yield optimization is fundamentally an exercise in avoiding side reactions, rather than merely accelerating kinetics.

PC2 (*22.98% variance*) – *The catalyst-specific methanol efficacy axis:*

PC2 reveals a critical divergence in catalytic behavior. It is characterized by an extremely strong positive loading from the methanol-to-oil ratio (+0.82), which splits the yield responses: a strong positive loading for KOH yield (+0.75) and a strong negative loading for NaOH yield (–0.69). This axis unequivocally demonstrates that methanol stoichiometry exerts a catalyst-dependent, antagonistic effect. Increasing methanol concentration profoundly enhances the transesterification efficiency of KOH but concurrently suppresses the performance of NaOH. This finding provides the multivariate statistical basis for the ridge-and-valley structures in the RSM plots (Figs. 1(k)-(l)) and explains the stark contrast in the ANOVA results, where the AD (Catalyst × Ratio) interaction is highly significant for KOH (P < 0.0001) but insignificant for NaOH (P = 0.4914).

PC3 (*16.88% variance*) – *The catalyst loading isolate*

PC3 is almost exclusively defined by the catalyst concentration (loading = +0.96), with all other variables contributing negligibly. This component successfully isolates the unique effect of catalyst dose from other process parameters. Its intermediate contribution to variance, coupled with its statistical significance in the KOH ANOVA model (P < 0.001), indicates that while catalyst loading is a relevant factor. However, its effect is secondary and primarily manifests through interactions with other variables (e.g., soap formation from excess KOH at elevated temperature). Its minimal correlation with NaOH yield (ANOVA, P = 0.45) further supports its limited role as an independent variable for optimizing this catalyst system.

PC4 (*13.86% variance*) – *The kinetic compensation axis*

PC4 captures the classic kinetic trade-off, characterized by a strong positive loading on temperature (+0.72) and a strong negative loading on time (–0.68). It indicates that a higher reaction temperature can compensate for a shorter reaction time, and vice versa, to achieve a similar outcome. This trade-off is visually apparent in the 3D surfaces (Figs. 1(i)-(j)), where increasing the temperature reduces the time required to reach maximum yield. Furthermore, the moderate negative loading of the methanol ratio (–0.32) suggests a slight interplay with this balance. Crucially, PC4 explains the least variance and is not directly correlated with yield, indicating that this trade-off offers scheduling flexibility but does not, by itself, dictate the

ultimate yield efficiency. The latter is governed by the constraints defined in PC1 and PC2, which represent the fundamental limits of the catalytic system.

Synthesis and mechanistic implications

The PCA deconstruction yields a definitive hierarchy of parameter influence:

(1) Yield boundary (PC1): Ultimate yield is governed by maintaining operational parameters below the severity threshold that triggers extensive saponification.

(2) Catalyst selection dictates strategy (PC2): The optimal methanol stoichiometry is entirely contingent on the catalyst chosen, representing a fundamental process design choice.

(3) Secondary role of loading (PC3): Catalyst concentration is a fine-tuning parameter within a narrow optimal band, not a primary driver of yield.

(4) Process flexibility (PC4): Temperature and time can be traded within the constraints of the above hierarchy to accommodate practical reactor operation.

Most significantly, this analysis provides irrefutable multivariate evidence that KOH and NaOH operate on fundamentally distinct catalytic response landscapes. KOH performance is maximized in a high-methanol, moderate-severity quadrant (PC1+, PC2+), whereas NaOH is constrained to a low-methanol, low-severity window (PC1+, PC2–) to avoid its pronounced tendency for side reactions. This PCA-driven insight moves beyond qualitative observation to deliver a robust, quantitative framework for catalyst selection and holistic process optimization in alkali-catalyzed biodiesel synthesis.

4.10.6 Optimum conditions for the highest yield

To initiate the comparison, an optimization is performed to determine the conditions for maximum yield using both catalysts. This is conducted using the RSM method based on the fitted quadratic models (Eqs. 3 and 4). To determine the conditions that maximize biodiesel yield while remaining within the experimental design space. The desirability function approach is applied, assigning maximum weight to the biodiesel yield response and allowing all four independent variables (catalyst loading A, temperature B, time C, methanol-to-oil ratio D) to vary freely within their coded ranges (−2 to +2). The optimization results are presented in Table 9. It reveals clear catalyst-specific optimal windows: for KOH, the model predicts a maximum yield of 99.5% at 1.00 wt.% catalyst loading, 58.66℃, a very short reaction time of 5.20 min, and a high methanol-to-oil ratio of 0.424 v/v (~9.5:1 molar equivalent). In contrast, for NaOH, the predicted maximum is 97.32% at nearly identical catalyst loading (1.006 wt.%), but requiring a much lower methanol ratio (0.255 v/v, ~6:1 molar), a lower temperature (40.04℃), and a substantially longer reaction time (66.42 min).

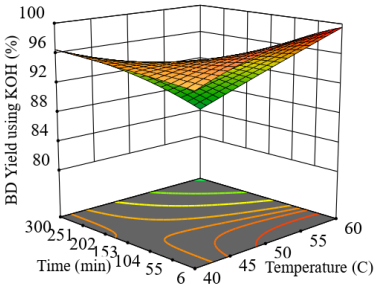
Table 9 Optimum conditions to maximize the biodiesel yield

	KOH	NaOH
Catalyst Loading (%)	1.00	1.006
Reaction Time (min)	5.20	66.42
Reaction Temperature (℃)	58.66	40.04
Methanol-Oil-Ratio	0.424	0.255
Maximum yield (%)	99.5	97.32

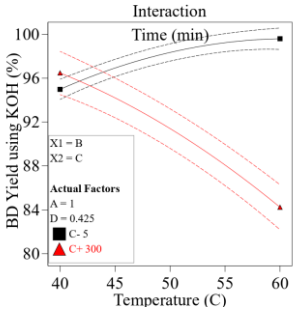
Under these conditions, using the principal components, the performance of both catalysts is discussed in terms of the relationships between the variables and PC1 and PC2. PC1 (Reaction Severity Axis, 36.78% variance) primarily captures the trade-off between high biodiesel yield and avoidance of excessive reaction severity (negative loadings for temperature and time). This indicates that optimal yields for both catalysts occur under milder conditions that minimize side reactions, such as saponification and reverse transesterification. PC2 (Catalyst-Specific Methanol Efficacy Axis, 22.98% variance) starkly

separates the catalysts: it shows a strong positive loading for the methanol-to-oil ratio (+0.82), coupled with a positive contribution to KOH yield (+0.75) but a negative contribution to NaOH yield (–0.69). This quantifies how excess methanol enhances KOH performance — likely via an equilibrium shift and improved miscibility — while detrimentally affecting NaOH through increased glycerol solubility, emulsion stabilization, and the promotion of the reverse reaction.

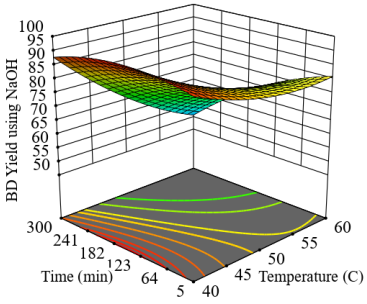
Fig. 6 illustrates the surface and interaction plots that correspond to the effects of PC1 loading variables (i.e., reaction temperature and time) and PC2 (i.e., catalyst and Methanol-to-oil ratio (MeOH/Oil ratio)) on biodiesel yields. It is clear that the interaction between reaction temperature and time significantly affects the biodiesel yields of both catalysts. For the KOH catalyst, at lower reaction temperatures, the time must be increased to achieve the best yield of approximately 97%. In contrast, for NaOH, there is little or no effect of reaction time.



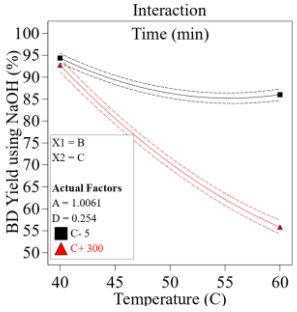
(a) Response to Time/Temperature with KOH



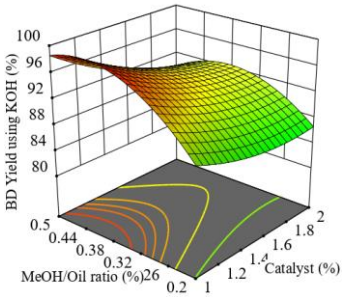
(b) Time/Temperature interaction with KOH



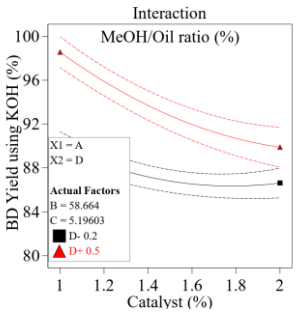
(c) Response to Time/Temperature with NaOH



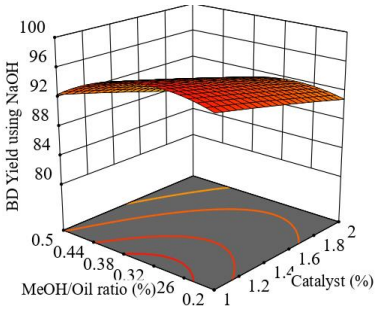
(d) Time/Temperature interaction with NaOH



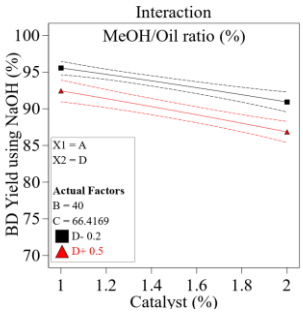
(e) Response to Catalyst/ MeOH Oil Ratio with KOH



(f) Catalyst/ MeOH Oil Ratio Interaction with KOH



(g) Response to Catalyst/ MeOH Oil Ratio NaOH



(h) Catalyst/ MeOH Oil Ratio Interaction with NaOH

Fig. 6 Effect of PC2 variables on PC1 loading

Both catalysts agree that at higher reaction temperatures, the time needed to achieve the best yield is reduced. Furthermore, when using KOH as a catalyst, the yield increases with reaction temperature at shorter reaction times. In contrast to NaOH, the trend is decreasing at both long and short reaction times across all temperatures.

From the surface plot, it can be observed that the KOH catalyst shows approximately a 10% increase in yield at lower reaction temperatures. In comparison, a 6% increase is observed with time at cooler reaction temperatures. Yield drops significantly at higher temperatures and for longer durations.

With NaOH, the yield decreases continuously with increasing reaction time across all temperatures. In contrast, at cooler temperatures, the yield peaks at around 95% after nearly 60 minutes. This pattern reflects NaOH's pronounced sensitivity to prolonged exposure, where extended reaction times promote secondary reactions—particularly saponification of residual free fatty acids, ester hydrolysis, and formation of stable emulsions/gels—that progressively consume catalyst, increase mixture viscosity, and drive the reversible transesterification backward. At lower temperatures (e.g., ~40℃), the kinetics of these detrimental side reactions are sufficiently suppressed, allowing the forward reaction to reach a maximum yield plateau (~95%) around 60 minutes before any significant decline sets in.

PC2 loading variables produce different effects on yield, as shown in Figs. 6(a) and (h). With KOH, there is a moderate interaction between the variables, as seen in the yield, whereas with NaOH, there is none. For the KOH catalyst, the best yield is achieved when the MeOH/IL ratio is maintained at a high level across all catalyst loadings, peaking at around 98% at 1% catalyst loading and decreasing to a maximum of 91% at higher catalyst loadings. This ratio should be kept low to achieve the best yield: 96% at 1% catalyst loading and 89% at higher loading.

From the surface plots, a general trend of yield reduction is observed at catalyst loadings exceeding 1% for both catalysts, with a slightly greater effect on KOH than on NaOH. The peak yield value is achieved at a 0.45% MeOH/Oil ratio for KOH, while it is 0.25% for NaOH.

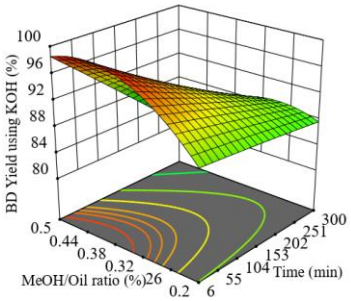
The effect of PC3 and PC4 on the biodiesel yield is illustrated in Fig. 7. Figs. 7(a) and (b) show a strong interaction between the variables and yield. Whereas the interaction is linear for KOH, it is clearly non-linear (quadratic) for the case of NaOH. For shorter-duration reactions, the MeOH/Oil ratio should be maximized to achieve a peak yield of 97%, whereas for longer reaction durations, it should be kept low. For NaOH, at shorter reaction times, the ratio must be maximized to achieve 92.6% yield, whereas there is a clear optimum time for the best yield at lower ratios.

Referring to the figures, the KOH-use ratio at which the yield is maximum varies with reaction time. The peak value is reached at shorter durations with a ratio of about 0.45%. Using NaOH, not much variation is noticed, except that the yield reaches its peak value at nearly 180 minutes for lower ratios and 0.45% ratio at a shorter time.

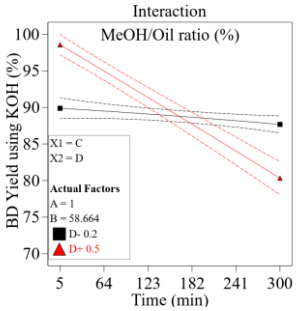
For the case of PC4 variables shown in Figs. 7(e) through (h), the effect on yield using the studied catalysts differs. Whereas a clear and strong interaction is observed with KOH, a weak interaction (within the studied range) is observed with NaOH. Further, the interaction is non-linear with KOH compared with NaOH.

It is also observed that, with higher catalyst loading, the yield decreases with increasing temperature for KOH, reaching a peak of 97% at 1% loading and higher reaction temperatures. For NaOH, the yield increases with catalyst loading at higher temperatures and decreases at lower ones. Furthermore, the yield decreases continuously with increasing reaction temperature across all catalyst loadings.

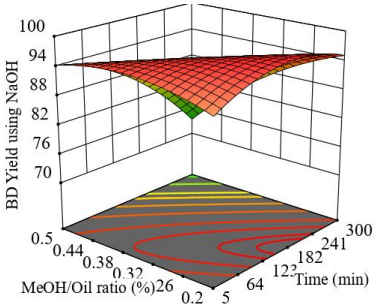
While the RSM model predicts a theoretical maximum yield at 58.7℃ for KOH, the practical operational optimum remains approximately 50℃. This is due to increased risks of saponification and methanol boiling near its boiling point (64.7℃) at temperatures above 60℃, as evidenced by the decline in yield in the longitudinal experimental data. The associated practical and safety challenges for scale-up outweigh the marginal predicted yield gain above 50℃.



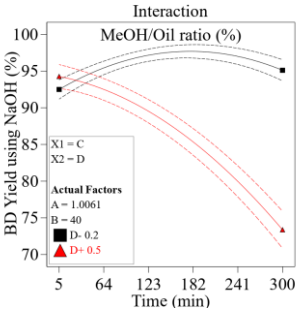
(a) Response to Time/ MeOH Oil Ratio with KOH



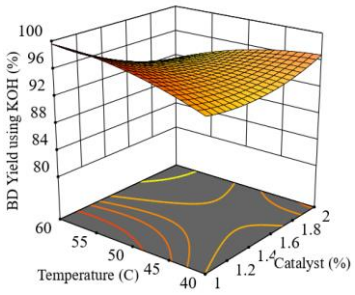
(b) Time/ MeOH Oil Ratio Interaction with KOH



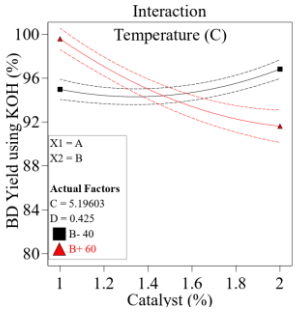
(c) Response to Time/ MeOH Oil Ratio with NaOH



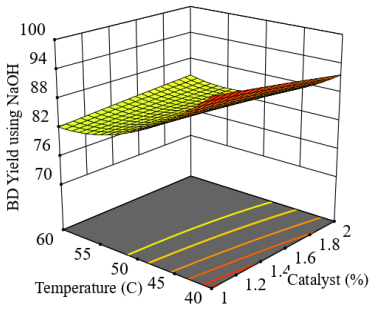
(d) Time/ MeOH Oil Ratio Interaction with NaOH



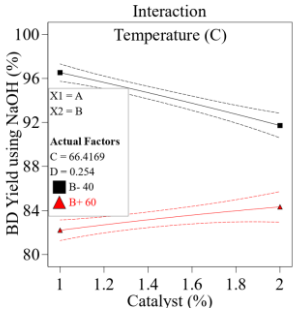
(e) Response to Catalyst/Temperature with KOH



(f) Catalyst/Temperature interaction with KOH



(g) Response to Catalyst/Temperature with NaOH



(h) Catalyst/Temperature interaction with NaOH

Fig. 7 Effect of PC3 and PC4 variables on PC1 loading

4.11 Comparative evaluation of RSM (*ANOVA*) *and machine learning approaches*

The present study employs two complementary modeling paradigms—RSM with ANOVA for quadratic polynomial fitting and ML *techniques*, including RFR and PCA—to analyze and predict biodiesel yield from waste frying oil under KOH- and NaOH-catalysis. A direct comparison of these approaches reveals distinct strengths, limitations, and complementary insights that together provide a more robust understanding than either method alone.

RSM, grounded in statistical design of experiments, offers high interpretability through explicit quadratic equations (Eqs. 3–4), ANOVA-derived significance tests (Tables 7-8), and response surface visualizations (Fig. 1). Its primary advantages include: (i) statistical rigor via p-values, F-values, and lack-of-fit tests, confirming model significance ($F \approx 159$ for KOH, 532 for NaOH; $p < 0.0001$); (ii) clear quantification of linear, quadratic, and two-factor interaction effects (e.g., highly significant

BC [temperature × time] and CD [time × methanol ratio] terms for both catalysts); (iii) low computational demand and direct usability for optimization without extensive tuning; and (iv) provision of prediction equations in actual factor units suitable for process control. However, RSM assumes a predefined quadratic form, which may inadequately capture highly non-linear or higher-order behaviors, and its predictive performance is constrained to the experimental design space (R^2 = 0.9525–0.9854; adjusted R^2 = 0.9465–0.9836).

In contrast, machine learning approaches—random forest for feature importance and PCA for latent structure extraction—excel at handling complex, non-linear, and multivariate relationships without assuming a specific functional form. Random forest identifies catalyst-specific parameter dominance (methanol-to-oil ratio as the leading positive driver for KOH [~0.60 mean decrease in R^2], versus temperature for NaOH [~1.10]), aligning with but extending beyond ANOVA by ranking relative global importance rather than local significance (Fig. 3). PCA decoupled four principal components explaining 84.3% variance, revealing fundamental mechanistic divergences (e.g., PC2 as the "catalyst-specific methanol efficacy axis" with opposing signs for KOH and NaOH yields).

ML advantages encompass: (i) superior detection of non-linear patterns and interactions invisible to quadratic models; (ii) robust handling of data noise and outliers; and (iii) quantitative prioritization of variables for process diagnosis. Limitations include reduced interpretability (the "black-box" nature of Random Forests), higher computational requirements, potential overfitting with small datasets, and the absence of formal statistical hypothesis testing.

In this work, RSM provides statistically validated optimization surfaces and explicit interaction coefficients, while ML offers deeper mechanistic insight into why KOH outperforms NaOH (e.g., greater tolerance to excess methanol and thermal stability). The integrated framework—RSM for precise, interpretable optimization and ML for exploratory decoding of underlying patterns—yields more comprehensive conclusions than a standalone application, demonstrating the value of hybrid statistical–data-driven methodologies for complex chemical processes such as heterogeneous-phase transesterification. Table 10 below summarizes the comparison of all methods used in the analysis.

Table 10 Comparative summary of RSM (ANOVA) and machine learning approaches in this study

Aspect	RSM + ANOVA	RFR + PCA	Synergistic Benefit
Model Form	Quadratic polynomial (assumed)	Non-parametric, no assumed form	Captures both structured and emergent patterns
Interpretability	High (coefficients, p-values, surfaces)	Moderate (feature importance, loadings)	Mechanistic + diagnostic insights
Key Strength	Statistical significance & optimization	Non-linear ranking & latent structure	Robust yield prediction + catalyst divergence
Main Limitation	Assumes quadratic form; design-space bound	Less direct mechanistic equations; tuning needed	Overcomes individual weaknesses
Performance (this study)	R^2 0.95–0.99; explicit interactions	Dominant factors identified; 84.3% variance	Catalyst-specific strategies derived

5. Conclusion

This study employed an integrated framework of RSM, RFR, and PCA to optimize and compare the catalytic performance of KOH and NaOH in the transesterification of waste frying oil (WFO) to biodiesel. This novel hybrid methodology advances mechanistic understanding beyond traditional empirical approaches. The multivariate analysis yielded the following key conclusions:

(1) KOH is the superior catalyst for WFO biodiesel production. It consistently achieved a significantly higher and more robust mean biodiesel yield (90.7%) than NaOH (84.1%), with lower variability and fewer low-yield outliers, as confirmed by a paired t-test ($p < 0.0001$).

(2) The optimal process conditions for maximum yield are catalyst-specific. For KOH, the RSM model identified optimal conditions of 1.0 wt.% catalyst, 58.7℃, 5.2 min reaction time, and a 0.42 methanol-to-oil ratio, predicting a 99.5% yield. For NaOH, the optimum required a lower methanol ratio (0.26) and a longer time (66.4 min) at 40.0℃, yielding a maximum of 97.3%.

(3) Machine learning reveals a fundamental mechanistic divergence. Random forest analysis showed that the methanol-to-oil ratio is the most important positive variable for KOH yield, whereas reaction temperature is the dominant (and sensitivity-inducing) factor for NaOH. This indicates that the catalysts operate on distinct chemical landscapes.

(4) PCA provides a multivariate explanation for catalyst performance. The analysis decoupled parameter interactions, demonstrating that KOH performance is enhanced in a high-methanol, moderate-severity regime. By contrast, NaOH is constrained to a low-methanol, low-temperature window in order to avoid its pronounced tendency toward saponification and emulsion formation.

(5) The recommended practical operating window for KOH prioritizes stability. While the model predicts a peak at 58.7℃, approximately 50℃ with a reaction time of 20–40 minutes is recommended for scalable operation. This maintains yields >98% while avoiding the risks of methanol boiling and accelerated side reactions near 60℃, offering a robust, efficient, and pretreatment-free pathway for biodiesel production from low-FFA waste oils.

(6) Comparative evaluation underscores the complementary roles of RSM and ML in advancing from empirical optimization to mechanistic understanding.

(7) NaOH exhibits inferior performance due to pronounced temperature-driven saponification and methanol intolerance, contrasting with KOH's robustness and confirming catalyst-specific optimization imperatives.

Acknowledgment

This work is a continuation of a project supported by the Deanship of Scientific Research at the University of Jordan and was carried out at the Chemical Engineering Department laboratories (The University of Jordan).

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] International Energy Agency, “Oil Market Report - October 2025,” https://www.iea.org/reports/oil-market-report-october-2025, 2025.

[2] United Nation Environment Programme, “Climate Risks in Transportation Sector,” https://www.unepfi.org/wordpress/wp-content/uploads/2024/05/Climate-Risks-in-the-Transportation-Sector-1.pdf, 2024.

[3] H. Hosseinzadeh-Bandbafha, A. S. Nizami, S. A. Kalogirou, V. K. Gupta, Y. K. Park, A. Fallahi, et al., "Environmental Life Cycle Assessment of Biodiesel Production from Waste Cooking Oil: A Systematic Review," Renewable and Sustainable Energy Reviews, vol. 161, article no. 112411, 2022.

[4] International Energy Agency, “Biofuels Can Provide up to 27% of World Transportation Fuel by 2050, IEA Report Says - IEA ‘Roadmap’ Shows How Biofuel Production Can be Expanded in a Sustainable Way, and Identifies Needed Technologies and Policy Actions.” https://www.iea.org/news/biofuels-can-provide-up-to-27-of-world-transportation-fuel-by-2050-iea-report-says-iea-roadmap-shows-how-biofuel-production-can-be-expanded-in-a-sustainable-way-and-identifies-needed-technologies-and-policy-actions, 2011.

[5] B. Azhar, M. I. Taipabu, C. Avian, K. Viswanathan, W. Wu, et al., "Artificial Intelligence-Driven Modeling of Biodiesel Production from Fats, Oils, and Grease (Fog) with Process Optimization Via Particle Swarm Optimization," Energy Conversion and Management: X, vol. 26, article no. 101000, 2025.

[6] D. Topi, "Transforming Waste Vegetable Oils to Biodiesel, Establishing of a Waste Oil Management System in Albania," SN Applied Sciences, vol. 2, no. 4, article no. 513, 2020.

[7] K. S. Mehra and G. Pant, "Production of Biofuel from Sesame Oil and Its Characterization as an Alternative Fuel for Diesel Engine," IOP Conference Series: Materials Science and Engineering, vol. 1116, article no. 012076, 2021.

[8] A. Sumayli, "Development of Advanced Machine Learning Models for Optimization of Methyl Ester Biofuel Production from Papaya Oil: Gaussian Process Regression (GPR), Multilayer Perceptron (MLP), and K-Nearest Neighbor (KNN) Regression Models," Arabian Journal of Chemistry, vol. 16, no. 2, article no. 104833, 2023.

[9] A. E.-F. Abomohra, M. Elsayed, S. Esakkimuthu, M. El-Sheekh, and D. Hanelt, "Potential of Fat, Oil and Grease (Fog) for Biodiesel Production: A Critical Review on the Recent Progress and Future Perspectives," Progress in Energy and Combustion Science, vol. 81, article no. 100868, 2020.

[10] M. Aghbashlo, M. Tabatabaei, H. Rastegari, H. S. Ghaziaskar, and T. R. Shojaei, "On the Exergetic Optimization of Solketalacetin Synthesis as a Green Fuel Additive through Ketalization of Glycerol-Derived Monoacetin with Acetone," Renewable Energy, vol. 126, pp. 242-253, 2018.

[11] W. R. Singh and H. N. Singh, "Maximizing Waste Cooking Oil Biodiesel Production Employing Novel Brotia Costula Derived Catalyst through Statistical and Machine Learning Optimization Techniques," Materials Today Communications, vol. 41, article no. 110838, 2024.

[12] H. S. Pali, A. Sharma, N. Kumar, and Y. Singh, "Biodiesel Yield and Properties Optimization from Kusum Oil by RSM," Fuel, vol. 291, article no. 120218, 2021.

[13] E. Betiku, A. S. Osunleke, V. Odude, A. Bamimore, B. Oladipo, A. Okeleye, et al., "Performance Evaluation of Adaptive Neuro-Fuzzy Inference System, Artificial Neural Network and Response Surface Methodology in Modeling Biodiesel Synthesis from Palm Kernel Oil by Transesterification," Biofuels, vol. 12, no. 3, pp. 339-354, 2021.

[14] M. Chhabra, B. S. Saini, and G. Dwivedi, "Optimization of the Dual Stage Procedure of Biodiesel Synthesis from Neem Oil Using RSM Based Box Behnken Design," Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, vol. 46, no. 1, pp. 8888-8911, 2024.

[15] W. Bajwa, A. Ikram, M. A. I. Malik, L. Razzaq, A. R. Khan, A. Latif, et al., "Optimization of Biodiesel Yield from Waste Cooking Oil and Sesame Oil Using RSM and ANN Techniques," Heliyon, vol. 10, no. 15, article no. e34804, 2024.

[16] J. Yamin, Z. Al-Hamamre, and A. Sandouqa, "Modelling and Optimisation of Biodiesel Production Using Waste Cooking Oil Using the Response Surface Methodology," International Journal of Sustainable Energy, vol. 43, no. 1, article no. 2355654, 2024.

[17] D. C. Selvam, G. Subbiah, N. Beemkumar, B. Juneja, A. Jena, Y. Devarajan, et al., "Optimization of Biodiesel Production from Sterculia Foetida Seed Oil: Leveraging Non-Edible Crops for Sustainable Alternative Fuel," Results in Chemistry, vol.16, article no. 102339, 2025.

[18] N. Budhraja, A. Pal, and R. Mishra, "Parameter Optimization for Enhanced Biodiesel Yield from Linum Usitatissimum Oil through Solar Energy Assistance," Biomass Conversion and Biorefinery, vol. 14, pp. 15335-15350, 2024.

[19] DIN EN 14214 , “Liquid Petroleum Products - Fatty Acid Methyl Esters (FAME) for Use in Diesel Engines and Heating Applications - Requirements and Test Methods,” German Version, DIN EN 14214:2014-06, 2014.

[20] Z. Al-Hamamre, M. Alnaief, S. Daameh, J. Yamin, A. Sandouqa, R. Alhammouri, et al., "Sustainable Biodiesel Production Using Lignin-Derived Sulfonated Carbon Aerogels Catalyst," International Journal of Sustainable Energy, vol. 43, no. 1, article no. 2400657, 2024.

[21] H. H. Naseef and R. H. Tulaimat, "Transesterification and Esterification for Biodiesel Production: A Comprehensive Review of Catalysts and Palm Oil Feedstocks," Energy Conversion and Management: X, vol. 26, article no. 100931, 2025.

[22] M. Dorado, E. Ballesteros, M. Mittelbach, and F. J. López, "Kinetic Parameters Affecting the Alkali-Catalyzed Transesterification Process of Used Olive Oil," Energy & Fuels, vol. 18, no. 5, pp. 1457-1462, 2004.

[23] N. Al-Jammal, Z. Al-Hamamre, and T. Juzsakova, “Parametric Study on the Production of Biodiesel from Waste Sunflower Oil Using Zeolitic Tuff Based Catalyst,” Energy Sources Part A: Recovery, Utilization, and Environmental Effect, vol. 45, no. 2, pp. 5310-5320, 2023.

[24] H. H. Hamed, A. E. Mohammed, O. A. Habeeb, O. M. Ali, O. S. Aljaf, and M. A. Abdulqader, "Biodiesel Production from Waste Cooking Oil Using Homogeneous Catalyst," Egyptian Journal of Chemistry, vol. 64, no. 6, pp. 2827-2832, 2021.

[25] H. Belkhanchi, M. Rouan, M. Hammi, Y. Ziat, and M. Chigr, "Synthesis of Biodiesel by Transesterification of Used Frying Oils (UFO) through Basic Homogeneous Catalysts (NaOH and KOH)," Biointerface Research in Applied Chemistry, vol. 11, no. 5, pp. 12858-12868, 2021.

[26] H. Esmaeili and R. Foroutan, "Optimization of Biodiesel Production from Goat Tallow Using Alkaline Catalysts and Combining Them with Diesel," Chemistry & Chemical Technology, vol. 12, no. 1, pp. 120-126, 2018.

[27] S. Krishnamoorthi, M. Prabhahar, M. S. Kumar, and S. Sendilvelan, "Yield Characteristic of Biodiesel Derived from Used Vegetable Oil Methyl Ester (Uvome) Blended with Diesel, in the Presence of Sodium Hydroxide (NaOH) and Potassium

Hydroxide (KOH) Catalyst, as Alternative Fuel for Diesel Engines," International Journal of Mechanical and Production Engineering Research and Development, vol. 8, no. 1, pp. 9-16, 2018.

[28] M. A. Jamil, "Production and Optimization Study of Biodiesel Produced from Non-Edible Seed Oil," Science and Technology for Energy Transition, vol. 79, article no. 38, 2024.

[29] M. Córdova, O. Cabrera, O. Ruíz, and E. Barreno-Avila, "Carbon Footprint in Bovine Fat Biodiesel Synthesis - Comparison Using Methanol or Ethanol," Materials Science Forum, vol. 1127, pp. 49-56, 2024.

[30] National Biodiesel Education Program, "Biodiesel Economics," https://biodieseleducation.org/Production/Biodiesel Economics.html, accessed 2026.