Experimental Study and Modeling Approach of Response Surface Methodology Coupled with Crow Search Algorithm for Optimizing the Extraction Conditions of Papaya Seed Waste Oil

S. M. Z. Hossain[1] · S. Taher[1] · A. Khan[1] · N. Sultana[2] · M. F. Irfan[1] · B. Haq[3] · S. A. Razzak[4]

Abstract

Papaya seed waste can be a reliable feedstock for producing valuable bioproducts (biodiesel, biolubricants, beauty products, etc.) due to its high oil content. This article focuses to explore the effects of Soxhlet extraction process conditions (extraction time and seed particle size) on the percent oil yield obtained from papaya seeds. Initially, two mathematical models were developed using response surface methodology (RSM) via central composite design and regression analysis (generalized linear model, GLM) to predict the oil yield. The prediction performance of RSM model was found to be superior than GLM. The extracted oil was characterized by gas chromatography–mass spectrometry (GC–MS) analysis. The analysis of variance results indicated that both factors were strongly significant. Later, crow search algorithm (nature-motivated metaheuristic algorithm) articulated with RSM was utilized for global optimal solution. The maximum yield of 29.96% was obtained at extraction time of 6.5 h and seed particle size of 0.85 mm. The similar results were obtained by desirability function-based optimization approach. The predicted optimal set was also validated further by experimental yield of 31.1% with the variation of < 5%.

Keywords Papaya seed waste oil · Solvent extraction · Optimization · Response surface methodology · Crow search algorithm

List of symbols

OFAT	One-factor-at-a-time
RSM	Response surface methodology
CCD	Central composite design
BBD	Box–Behnken design
GLM	Generalized linear model
A	Extraction time (T)
B	Seed particle size (S)
ANOVA	Analysis of variance
y	Predicted response
CSA	Crow search algorithm
AP	Awareness probability
fl	Flight length
N	Flock size
GC–MS	Gas chromatography–mass spectrometry
RE	Relative error
MAE	Mean absolute error
RMSE	Root mean squared error
DF	Desirability function

✉ S. M. Z. Hossain
zhossain@uob.edu.bh

[1] Department of Chemical Engineering, College of Engineering, University of Bahrain, Zallaq, Kingdom of Bahrain

[2] Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

[3] Department of Petroleum Engineering, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

[4] Department of Chemical Engineering, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

1 Introduction

The leading contribution to the greenhouse effect is due to the immense usage of fossil fuels in transportation and power generation. The petroleum reserves are not long-lasting, rather it has been stated that with the current production

rate oil reserves will diminish in less than half a century [1, 2]. Therefore, the renewable and environmentally friendly energy resources may take part important role to alleviate the probable problems related to the fossil fuel (health hazard, acid rain, global warming and reservoir exhaustion). In this regard, biofuels, (ester-based oxygenated fuel, sulfur-free, renewable, non-toxic and biodegradable) obtained from seed oil could be the alternate solution that may save the universe [3–9] Extracted oil from seeds are the major source of biofuel. In the last 10 years, more than 350 oil-containing seeds (palm, coconut, jatropha, sunflower, rapeseed, soybean, jojoba, karanja, neem, moringa, castor, cotton oil, etc.) have been identified for feedstocks of biofuel. However, these feedstocks are edible food for human [10, 11]. Therefore, non-edible feedstocks (e.g., papaya seeds, dates seeds, microalgae biomass and others) are recently attracting attention worldwide since those do not combat with human consumption [12–14]. Moreover, many studies were conducted and proven that seeds oil can be used for the production of a variety of other valuable products such as biolubricants, biosolvents, cosmetics and beauty products (e.g., demulcent skin care products, hair conditioners, bath oils and makeup) [15–17].

The papaya (Carica papaya) is the fourth trading tropical fruit worldwide. Almost 75% of the papaya produced in the world comes mainly from India, Bangladesh, Brazil, Indonesia, Nigeria and Mexico [13]. Among these, India alone is contributing 42% of world production (about 3 million tons per year) [18]. The papaya fruit consists of many nutrients like vitamin C and A, magnesium, folate, fiber and potassium [16]. The seed is black or dark brown, soft and round shape with a strong smell. The literature reveals that the dry papaya seed contains around 30% lipids, 28% proteins and 22% fibers [19]. The component of the unsaturated fatty acids (carbon ranging from C_{14} to C_{22}) in the seeds consists of mainly meristic acid, oleic acid, stearic acid, linoleic acid and palmitic acid [20]. The weight of papaya fruit is usually from 200 to 3000 g and approximately 15% of the wet weight of the fruit is seed. Since seeds are not eaten, 15% of the seeds or biomass is thrown as waste material, which can be used as the feedstock for valuable bioproducts syntheses [21, 22]. Nonetheless, to make papaya seeds more useful, it is imperative to examine and evaluate the production of papaya seed oil through extraction methodology.

Until now, several extraction techniques have been utilized for the extraction of useful oils from seeds of various plants. The method of extraction of oils from seeds is an important factor to produce high-quality oils. Solvent extraction is one of the choices, but it depends on the selectivity and needs extreme heat [23]. Cold press extraction is the conventional technique for oil extraction but have low yield [24]. Supercritical fluid extraction (SFE) and enzyme-assisted aqueous oil extraction are currently used techniques to extract plant oils and provide convenient features [25–29]. The Soxhlet extraction process with hexane is also mostly used for oil extraction from seeds [23, 30–34]. However, in order to find the high oil yield, the Soxhlet extraction process needs to be operated at the optimal conditions including extraction time, seed particle sizes, etc.

Most of the reports in the field of extraction of papaya seeds oil adopt either one-factor-at-a-time (OFAT) or response surface methodology (RSM) strategy via central composite design (CCD)/Box–Behnken design(BBD) to optimize the process variables [20, 35]. It is noteworthy that optimization is the technique of getting the point that minimizes or maximizes a response or output. It is well known that OFAT is a trial and error approach and it overlooks interactive effects within the process parameters and thereby, and it is unable to optimize a process perfectly. Though RSM is efficient, can solve problems with large numbers of design variables, considers interaction effects and requires little parameter tuning, it may give only local optimal solution [36–40]. To overcome this limitation, global optimum solution is preferable.

Several popular algorithms utilized for global optimization include genetic algorithm (GA) [41], simulated annealing (SA) [42, 43], ant colony optimization (ACO) [44], particle swarm optimization (PSO) [45], combinatorial optimization [46] and harmony search (HS) [47]. However, the usage of high number of tuning parameters with some disputes makes these processes tiresome. Recently, a new metaheuristic optimization algorithm entitled 'crow search algorithm (CSA)' has been documented, which can overcome the drawbacks [48]. It is a nature-inspired metaheuristic algorithm for solving global optimization problems in the areas of engineering research [48–57]. It has only two variables and as such it is remarkable to all other well-known algorithms.

RSM is a well-known effective method to explain the relationships between input and output parameters. However, there is possibility of this method to provide suboptimal solution. On the other hand, RSM coupled with CSA is a very efficient platform to optimize the independent variables for maximizing the response variables or outputs. This integrated platform always provides true optimal solution. A very limited number of past research investigations used this integrated approach for optimizing process variables. Water jet cutting process parameters are optimized by RSM coupled with CSA [54]. However, to our knowledge, this integrated platform has not been applied in any areas of extraction-based optimization problems. As such, it is very crucial for the scientists, engineers and researchers to optimize the extraction conditions of papaya seed oil using the RSM coupled with CSA.

This paper focuses on the optimization of the Soxhlet extraction parameters (extraction time and seed particle size) on papaya oil yield using RSM articulated with CSA. Ini-

tially, RSM integrated with CCD approach was adopted for development of the quadratic regression model to predict oil yield. The performance of RSM model is also compared with that of widely used generalized linear model (GLM) [58]. The GLM allows to fit regression models for univariate response data that follow a very general distribution called the exponential family. The exponential family includes the normal, binomial, Poisson, geometric, negative binomial, exponential, gamma and inverse normal distributions [59]. Later, the developed best regression model was used to investigate the optimum combination of the input variables using CSA to get global optimum solution and these results were compared with those of the desirability function-based optimization approach. The desirability function approach is used widely for factor optimization in engineering research [60–62]. In this function, the characteristics of each anticipated output are converted into a unitless value (*d*), which varies from 0 to 1. The desirability of output raises with the value of *d*. Finally, the predicted operating conditions have been verified by conducting triplicate experiments. The produced oil was then analyzed by using a GC–MS method.

2 Materials and Methods

2.1 Seed Preparation

The raw ripe papaya seeds were obtained from Al-Jazira supermarket in Bahrain. The seeds were dehydrated in the laboratory at 50 °C for 24 h in an oven. After drying, the seeds were ground into five different sizes using a coffee grinder. Four sieves of the following sizes were used to isolate the seed particles: 0.85 mm, 1.18 mm, 1.40 mm and 2.00 mm. The fifth size was the full seed size which is approximately 3.75 mm. The grounded seeds with each size was placed in small bottles and stored in a refrigerator until use.

2.2 Papaya Seed Oil Extraction

The seed oil was extracted using Soxhlet apparatus via hexane as a solvent at 80 °C. Hexane (purity ≥ 99%) was obtained from Sigma-Aldrich. A typical Soxhlet apparatus is composed of: (1) a percolator (boiler and reflux) that allows the flow of solvent, (2) a thimble (porous thick filter paper) that holds the solid to be extracted and (3) a siphon mechanism that systematically empties the thimble. Briefly, in working principle, initially the grounded seed of each size (2 g) is added in a thimble, which is placed inside the extraction chamber. 125 mL of hexane (in a 250 mL distillation flask) can evaporate at 80 °C. The hexane vapor travels up the distillation arm, condenses while passing through the condenser and drips on the grounded seeds in the thimble. The chamber containing the seed powder slowly fills with warm hexane, and the seed oil mixes in the hexane. As soon as the Soxhlet chamber is full, the chamber is devoided by the siphon. The oil–hexane mixture is then returned to the distillation flask. The thimble acts a filter, and thereby, no solid material moves to the still pot. This cycle is repeated many times.

2.3 Oil Separation

The oil–hexane mixture was poured in a conical flask and clamped in a water bath at 80°C. The entire separation process takes place in a fume hood. The mixture was evaporated until it resulted in a viscous yellow residue. The volume and weight of the extract (oil) were measured by using pipette and an electric balance, respectively. The percentage yield of oil extraction was determined by using Eq. (1).

$$\%\text{Yield} = \frac{\text{Mass of the oil}}{\text{Mass of the intial sample}} \times 100 \tag{1}$$

2.4 Response Surface Methodology (RSM)

The RSM is a bias-less statistical method to investigate the relationship between the output (response) and input variables, as well as to optimize the relevant processes [13, 63–66]. The central composite design (CCD), one of the RSMs, was employed in this study. The size of seeds and the extraction time were considered two independent (input) factors, while the percentage yield of oil extraction was taken as response. The behavior of the process is characterized by the second-order multiple regressions model as Eq. (2):

$$y = \beta_0 + \sum_{i=1}^{N} \beta_i x_i + \sum_{i=1}^{N} \beta_{ii} x_i^2 + \sum_{i<j} \sum \beta_{ij} x_i x_j + \varepsilon \tag{2}$$

where *y* denotes the predicted output, x_i denotes the coded factors, β_o denotes the intercept term, β_i designates the linear effect, β_{ii} denotes the squared effect, β_{ij} denotes the interaction effect and ε denotes the residual.

The relationship between natural and coded variables can be expressed as Eq. (3).

$$\text{Coded value} = \frac{\text{Natural value} - \text{Mean}}{\text{Range}/2} \tag{3}$$

The experimental design matrix (total number of runs, coded/un-coded variables, range and levels of two independent variables, combination of two variables, etc.) was generated using Minitab v.18 as shown in Table 1.

Table 1 CCD for optimizing size of seed and time of extraction. For each parameter, five coded levels of extremely low (− 1.414), low (− 1), center (0), high (+ 1) and extremely high (− 1.414) were considered

Experiment number	Time (T) Actual (coded)	Particle size (S) Actual (coded)
1	6.5 h (+ 1)	2 mm (+ 1)
2	5.5 h (0)	1.4 mm (0)
3	7 h (+ 1.414)	1.4 mm (0)
4	4.5 h (− 1)	2 mm (+ 1)
5	4.5 h (− 1)	1.18 mm (− 1)
6	5.5 h (0)	1.4 mm (0)
7	5.5 h (0)	1.4 mm (0)
8	4 h (− 1.414)	1.4 mm (0)
9	5.5 h (0)	0.85 mm (− 1.414)
10	5.5 h (0)	3.75 mm (+ 1.414)
11	6.5 h (+ 1)	1.18 mm (− 1)

2.5 Generalized Linear Model (GLM)

The GLM can be written as:

$$g(\mu_i) = g[E(y_i)] = x_i' \beta \qquad (4)$$

where $E(y)$ is the expected value (or expectation function) of the response y, x_i is a vector of covariates for the ith observation and β is the vector of regression coefficients. Every generalized linear model has three components: a response variable distribution, a linear predictor that involves the covariates and a link function g (identity, logit, log, etc.) that connects the linear predictor to the natural mean of the response variable [59]. Depending on the choice of the link function g, a GLM can include a nonlinear model.

2.6 Crow Search Algorithm (CSA)

The best developed regression model (obtained from RSM or GLM) was utilized to find the global optimum combination of the input factors via crow search algorithm (CSA) using MATLAB v.16. The optimization codes of CSA were based on the following steps [36, 49, 50]:

1. Initialization of parameters: The parameters include the flock size (N), flight length (fl), awareness probability (AP), iteration number (iter$_{max}$) and any other parameters that are considered constants.
2. Initialization of memory and position: The N crows are positioned randomly in the following matrix of dimension (d) where d denotes the number of decision parameters. Each crow in the matrix represents a viable solution to the problem.

$$\text{Crows} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_d^1 \\ \vdots & & \ddots & \vdots \\ x_1^N & x_2^N & \cdots & x_d^N \end{bmatrix} \qquad (5)$$

The memory of the crow is initialized. Since the crow has no experience, the food is assumed to be hidden in the initial position.

$$\text{Memory} = \begin{bmatrix} m_1^1 & m_2^1 & \cdots & m_d^1 \\ \vdots & & \ddots & \vdots \\ m_1^N & m_2^N & \cdots & m_d^N \end{bmatrix} \qquad (6)$$

3. Evaluation of the fitness function: The quality of the initialized position is computed in the fitness function using the decision variables for each crow.
4. Generation of new positions: A crow of the generated matrix (crow i, for example) selects a random crow in the flock to follow (crow j). Crow i either reaches to the position of the food of crow j, or crow j notices that crow i is following and generates new position according to (Eq. 7).

$$x^{i,\text{iter}+1} = \begin{cases} x^{i,\text{iter}} + r_i \times fl^{i,\text{iter}} \\ \times \left(m^{j,\text{iter}} - x^{i,\text{iter}} \right) & \text{when } r_j \geq \text{AP}^{j,\text{iter}} \\ \text{a random position} & \text{otherwise} \end{cases} \qquad (7)$$

5. Testing the viability of the new positions: The feasibility of the new positions of the crows is checked within the constrain or limit. If it is feasible, then the crow updates the position or otherwise stays in the old position.
6. Evaluation of the fitness function, given the new positions.
7. Memory update: The crow updates its memory as follows:

$$m^{i,\text{iter}+1} = \begin{cases} x^{i,\text{iter}+1} & \text{if } f\left(x^{i,\text{iter}+1}\right) \text{ is better than } f\left(m^{i,\text{iter}}\right) \\ m^{i,\text{iter}} & \text{otherwise} \end{cases} \qquad (8)$$

where $f(\cdot)$ is the objective function value.

8. Termination benchmark: The steps from 4 to 7 are reiterated until the maximum iteration (iter$_{max}$) is reached. The best positions of the memory give the best objective function and the solution of the optimization problem.

2.7 Oil Characterization

In order to characterize oil components, the fatty acid contents of the papaya oil were examined using GC–MS [67].

Table 2 Oil yield (%) obtained in the central composite design (CCD) in duplicate sets of experiments

Experiment number	Yield (%)					Relative error (RSM)	Relative error (GLM)
	Experimental (set 1)	Experimental (set 2)	Experimental (ave)	Model prediction (RSM)	Model prediction (GLM)		
1	17.10	12.50	14.80	14.86	15.55	0.131	0.091
2	21.50	13.50	17.50	16.27	17.98	0.243	0.164
3	15.00	12.85	13.93	14.58	21.85	0.028	0.457
4	17.00	13.70	15.35	13.83	13.28	0.186	0.219
5	21.00	17.15	19.08	15.47	17.22	0.263	0.180
6	13.50	15.00	14.25	16.27	17.98	0.205	0.332
7	17.00	17.15	17.08	16.27	17.98	0.043	0.058
8	4.20	5.00	4.60	7.49	14.12	0.783	2.361
9	27.00	21.00	24.00	27.25	23.03	0.009	0.147
10	17.00	21.00	19.00	19.30	12.94	0.135	0.239
11	27.50	25.50	26.50	24.47	25.88	0.110	0.059

The instrument is composed of a PerkinElmer Clarus® 600 GC coupled with an MSD and a capillary column (30 m long ×0.25 mm ID and thickness of 0.25 μm). The sample (1 μL) was introduced in split mode (40:1) at 240 °C with 1.2 mL/min of He flow. Prior to injection, the papaya oil was diluted 25 times using pure hexane. A temperature program: 110 °C (4 min)–10 °C/min–150–3.9 °C/min–230 °C (5 min) was used to separate fatty acids. The individual component of the oil was identified using in-built chemical library and quantified by normalizing of percent area.

2.8 Statistical Significance

Analysis of variance (ANOVA) was used to explain the significance. A p value less than 0.1 denotes significant at 10% level, while a p-value less than 0.05 denotes significant at 5% level. All experiments were conducted in duplicate. Minitab (version 18.0) and SPSS (version 17.0) softwares were applied for RSM and GLM models development, respectively, while MATLAB v.16 was utilized for CRA-based analysis.

3 Results and Discussion

3.1 Development of Regression Models

In order to find optimum combination of seed particle size and extraction time for maximizing yield of papaya seed oil, two predictive models using RSM with CCD and regression analysis techniques were implemented. Table 2 shows the highest values of oil yield obtained for each experimental run. The results indicate different values, which are obvious as the extraction modes were different for each run (except for Runs 2, 6 and 7). For RSM-based model development, the experimental data were analyzed by multiple regressions using Minitab v.18. The predictive model for oil yield with respect to coded variables was expressed by Eq (9), where y denotes the yield, T is the time and S depicts the seed particle size.

$$y = 16.27 + 2.508T - 2.812S - 2.619T^2 + 3.501S^2 - 1.99TS \tag{9}$$

Table 3 demonstrates the ANOVA data for the second-order response surface model. The F-value for the model was high ($F = 10.13$) with a very low probability ($p = 0.0000$), which indicates the very high significance of the model. The significance of each term was investigated by their respective p-values. In the model, all the linear terms (T and S) and quadratic terms (S^2 and T^2) were statistically significant. The effect of the interaction is not significant, due to the p-value being slightly > 0.1. The significance effects were also proven using normal plot of standardized effects as shown in Fig. 1. Generally, the terms located far from line are significant. It can be seen from the figure that the terms T, S, T^2 and S^2 are located far from line, and thereby, they are significant. From Fig. 2 which is a Pareto chart, it can be seen even clearer the significance effect of the variables with square of particle size being the best, followed by normal particle size and extraction time. The interaction effect between the particle size and extraction time falls even below the average which is 2.120 as indicated in the chart.

Similarly, for GLM development, the experimental data have been assessed by multiple linear regressions using SPSS

Table 3 Analysis of variance (ANOVA) for response surface second-order model

Source	DF	Adj SS	Adj MS	F value	P value
Model	5	561.86	112.371	10.13	0.000
Linear	2	227.15	113.573	10.24	0.001
T	1	100.62	100.622	9.07	0.008
S	1	126.53	126.525	11.41	0.004
Square	2	302.91	151.455	13.65	0.000
$T*T$	1	77.44	77.443	6.98	0.018
$S*S$	1	138.41	138.414	12.48	0.003
Two-way interaction	1	31.80	31.800	2.87	0.110
$T*S$	1	31.80	31.800	2.87	0.110
Error	16	177.50	11.093		
Lack-of-fit	3	77.81	25.936	3.38	0.051
Pure error	13	99.69	7.668		
Total	21	739.35			
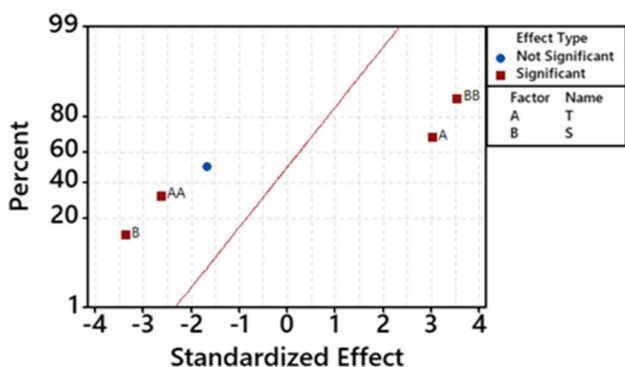
Fig. 1 Normal plot for the standardized effects of seed particle size and extraction time on percent oil yield. The significant terms (deep red in color) are located far from the diagonal line, while insignificant term (blue in color) is placed relatively closer to the line
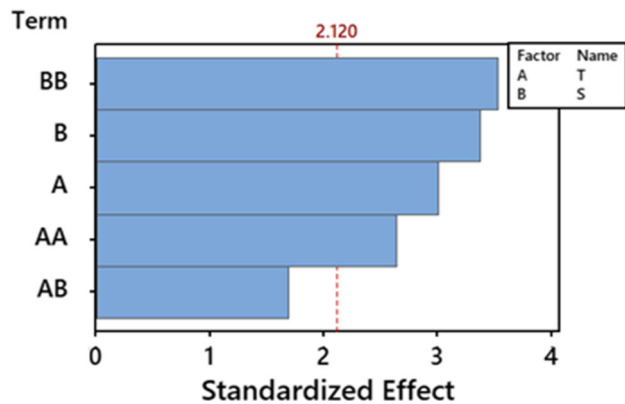

Fig. 2 Pareto chart for significant effects of seed particle size and extraction time on percent oil yield. Each bar denotes specific term. The bars for significant terms are crossed above the vertical red line (at an average value 2.120), while the bar for insignificant term falls below the average value as shown in the chart

Table 4 Performance comparison between RSM and GLM

Parameter criteria	RSM model	GLM model	Improvement of RSM wrt GLM (%)
Mean absolute error (MAE)	2.63	4.04	34.9
Root mean squared error (RMSE)	3.12	4.71	33.75

(version 17.0). The model for oil yield via coded parameters can be written as follows:

$$y = 17.982 + 2.734T - 3.568S - 1.600TS \tag{10}$$

where y, T, S and TS represent the oil yield, time, seed particle size and interaction between time and particle size, respectively.

3.2 Evaluation and Models Comparison

Both the response surface and generalized linear models are analyzed and compared with each other to find the best predictive model that leads better results with high accuracy. The data in Table 2 indicated that the lowest relative error, RE (%), was observed in RSM compared to GLM indicating better performance in the response model. To evaluate the models further, the results presented are compared based on two-performance grading criterions that include root mean squared error (RMSE) and mean absolute error (MAE). The values of RMSE and MAE are presented in Table 4. Based on RMSE, the RSM model performs better than GLM with performance enhancement of 33.75%. Similarly, using MAE to examine the model achievement, the RSM model outperforms the GLM model with performance enhancement of 34.9%.

Figure 3 shows the Parity plot, which is the relationship between anticipated and experimental data for oil yield. As shown in Fig. 3a (for RSM), the points are scattered around the line resulting in an $R^2 = 76\%$, indicating that the standard deviations between experimental and model predicted data are comparatively high. While it can be seen in Fig. 3b (for GLM), the R^2 value is comparatively low (41.3%), indicating that the model is unable to predict about 60% data. It has been reported that the R^2 value in the Parity plot for biological system should not be <0.75 [68]. Accordingly, though the results in RSM are satisfactory, a further investigation is required with wide ranges of experimental data to attain better predictions.
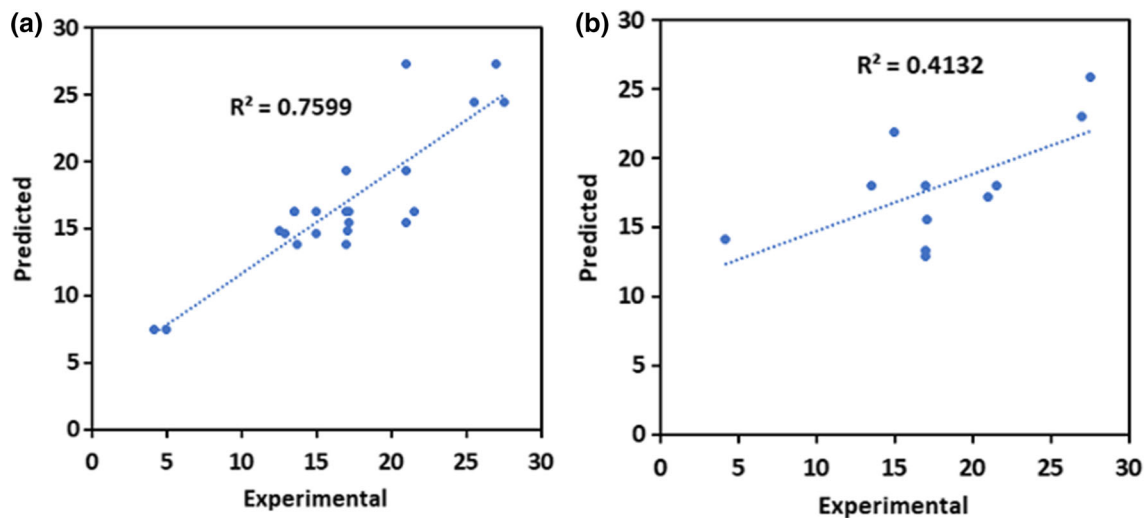
Fig. 3 Model predicted versus experimental percent oil yield via response surface methodology (RSM) (**a**) and generalized linear model (GLM) (**b**) approaches. The R^2 value in RSM is shown to be higher in compare to that in GLM
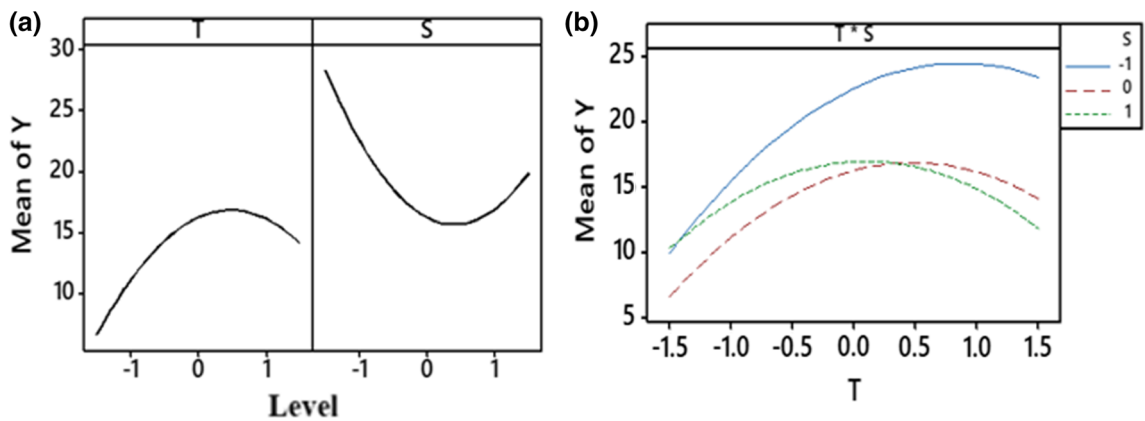

Fig. 4 Main effect (**a**) and interaction (**b**) plots of independent parameters (e.g., extraction time, seed particle size) on percent oil yield. Both parameters seem important since different levels of each factor affect the response (percent oil yield) differently. In interaction plot, factors influence each other. In this figure, the interaction effect among factors (e.g., T and S) is seemed to be low on the output

3.3 Effect of Environmental Conditions on Percent Oil Yield

To examine the effect of environmental conditions (e.g., extraction time and seed particle size) on response, percent oil yield by RSM, main effect and interaction plots were created as depicted in Fig. 4. Figure 4a shows that the response changes with the different levels of each factor, indicating that both parameters are important for papaya oil production. The data indicate that the percent oil yield increases with the extraction time and reaches to a plateau. This is expected since the contact time of solvent with the solute (seed particle) is longer, and diffusional mass transfer is favored, while in the case of particle size (S), at very small size (coded value of -1), the percent oil yield (y) is maximum. Small size means more surface area for the solvent and easier dif-

fusion. When the particle size increases, the solvent takes longer time to diffuse through the particle and as a result the percent yield decreases. As such, a minimum of extraction yield was observed at a maximum particles size. Contrarily, an interaction plot (Fig. 4b) is a visual image of the interaction of two input variables, extraction time and seed particle size on the percent oil yield. It shows three trends of different seed particle sizes -1, 0 and 1 which correspond to 1.18, 1.4 and 2 mm, respectively, with time and percent yield. From figure it is clear that the two-way interaction effects among factors (e.g., T and S) were low on the percent oil yield, and the blue line (corresponding to the size of 1.18 mm, smallest among the three) does not interact. The data agree well with the ANOVA results in Table 3 (p value $= 0.11$).

Fig. 5 Convergence rate of CSA for finding the best optimal solution. This plot assesses simultaneously the combined effects of all input factors and caters a stable output. After 54 iterations, a stable as well as an optimum point is achieved
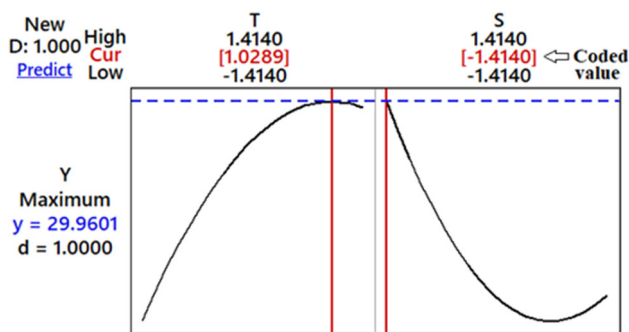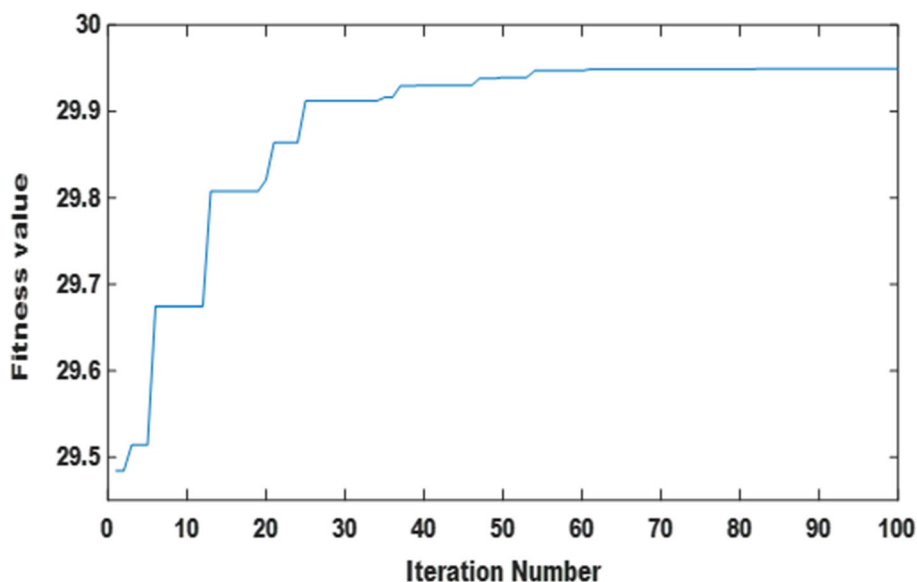
Fig. 6 Response optimizer plot for finding the optimum conditions to maximize percent oil yield. Optimum coded values for both parameters are shown in red color (above the figure), while predicted percent oil yield is presented in blue color (left side of the figure). The value desirability function (d) is also mentioned in the left side of the figure

3.4 Crow Search Algorithm (CSA) Coupled with RSM for Global Optimal Solution

In this study, the polynomial model obtained from RSM (Eq. 8) was used as the fitness function to get the global optimum solution since RSM model was found to be superior. The codes were run on MATLAB v.16 using the awareness probability (AP) of 0.1, flight length (fl) of 2, flock size (N) of 100 and maximum iteration of 150. A convergence plot of the fitness value and the iteration number were generated as shown in Fig. 5. It is well known that convergence plot explains the combined effects of all independent factors simultaneously and provides a stable or uniform output. The non-smooth line at the beginning is the result of random generation of parameters at the early iteration that might be far from each other. However, with further iterations the results get closer and after 54 iterations the optimum point is

Table 5 Statistical result obtained at optimized conditions by CSA (100 runs)

Best	Worst	Mean	Std.
29.949582	29.949257	29.949528	5×10^{-5}

achieved. The optimum coded combination of S and T is -1.414 and 1.013, which correspond 0.85 mm of particle size and 6.5 h of extraction time, respectively, to maximize yield of 29.95%.

The results obtained from integrated RSM–CSA-based approach have been compared with those of response surface methodology coupled with desirability function (RSM–DF)-based technique. In this regard, the transformation of oil yield is considered as a higher-the-better characteristic. For searching for a maximum, the desirability (d_i, denoted as the ith targeted output), can be written as follows:

$$d_i = \begin{cases} 0 & y_i < L \\ \left(\frac{y_i - L}{U - L}\right)^w & L \leq y_i \leq U \\ 1 & y_i > U \end{cases}, \qquad (11)$$

where w represents weight (assumed $w = 1$ in this study), L and U denote lower and upper values, respectively, and y_i denotes ith response. The response is entirely unsatisfactory when $d = 0$, while the response is ideal when $d = 1$. In order to get a single optimal point, response optimizer plot (which is the integrated RSM–DF-based approach) was constructed using Minitab 18, as presented in Fig. 6. As clearly shown, the particle size (S) and extraction time (T) obtained by RSM–DF approach give almost similar results as those found in RSM–CSA technique with very high desirability value ($d = 1$), indicating that the RSM–DF-based optimization was

Table 6 Predicted versus experimental oil yield data for: (a) RSM and (b) GLM

Optimum condition		Oil yield (%)		
Particle size (mm)	Extraction time (h)	Predicted by CSA	Experimental value (ave of triplicate data)	Error (%)
0.85	6.5	29.95	31.1	4.01

Table 7 Fatty acids found in papaya oil by GC–MS analysis

Fatty acid	Lipid number	Value (%)
Myristic acid	C14:0	2.96
Palmitic acid	C16:0	19.6
Oleic acid	C18:1 $n-9$	75
Linoleic acid	C18:2 $n-6$	1.85

extremely favorable and thereby supported the global optimal conditions. Since CSA is based on random generation, a little variation is observed for each run. Therefore, 100 runs were conducted separately and the standard deviation and mean of the results were taken as shown in Table 5. The very low standard deviation of 5×10^{-5} indicates that the optimal point obtained by CSA is robust and productive.

3.5 Validation in Optimal Conditions and Characterization of Extracted Oil

To validate the optimal settings for the extracted oil yield, the set of triplicate experiments were conducted with combination of seed particle size of 0.85 mm and extraction time of 6.5 h. The predicted and experimental yields were calculated to be 29.95 and 31.1%, respectively, with <5% error as shown in Table 6. Thus, the optimized conditions obtained by integrated RSM–CSA-based platform to predict the oil yield are robust and productive. The extracted oil was characterized by profiling fatty acids composition using GC–MS (Supplemental Figure S1) and quantified values presented in Table 7. The main fatty acids in the papaya oil were oleic (75.0%), followed by myristic (19.6%), palmitic (2.95%) and lenoleic (1.85%) acid. The data agreed well with those of previous reports [16, 17]. The high oleic oil makes it enough stability for food frying applications like those of high oleic vegetable oils in the market such as safflower (77%) and canola (75%). The intake of high oleic acid containing oils may have many nutritional and health benefits including reduction in oxidative stress in vivo [69]. In addition, for the production of high-quality biodiesel, biolubricants or other beauty products, the amount of fatty acids especially oleic acid present in oil play a very important key role [13].

4 Conclusions

The study demonstrated the effects of Soxhlet process variables on the percent yield of papaya seed waste oil. For this, comparison studies between RSM with CCD and GLM have been carried out to develop predictive models for predicting the oil yield. The performance of RSM model (based on RE, R^2, MAE, RMSE) was found to be superior compared to GLM. The ANOVA analysis (in RSM) indicates that both linear (T and S) and quadratic (T^2, S^2) terms are strongly associated (significant) with the percent yield of oil (all p values < 0.05), while the interaction effect within factors (T and S) was found to be insignificant (p value $= 0.11$). In order to find global optimal solution of these two parameters, RSM was then integrated with CSA. The optimal set of extraction time of 6.5 h and particle size of 0.85 mm provides the maximum yield of 29.95%. The similar results were obtained by desirability function-based optimization approach, indicating that the integrated RSM–CSA approach is robust. The predicted optimal conditions were validated experimentally with <5% error. Finally, the GC–MS results confirmed the oil composition. Overall, the application of CSA integrated with RSM has been utilized successfully for the first time in solving extraction-based factorial optimization problem. This integrated platform could easily be used in the future as an ideal pivotal tool for extraction of oil from other non-edible feedstocks such as dates seeds and microalgae biomass. Further, this platform might plausibly be utilized for other complex engineering processes where factors optimization is required in order to maximize/minimize either single or multiple objectives. Some of our current efforts are aligned in these directions.

References

1. Abas, N.; Kalair, A.; Khan, N.: Review of fossil fuels and future energy technologies. Futures **69**, 31–49 (2015)

2. Miller, R.G.; Sorrell, S.R.: The future of oil supply. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. (2014). https://doi.org/10.1098/rsta.2013.0179

3. Ganjehkaviri, A.; Mohd Jaafar, M.N.; Hosseini, S.E.; Musthafa, A.B.: Performance evaluation of palm oil-based biodiesel combustion in an oil burner. Energies 9, 1–10 (2016). https://doi.org/10.3390/en9020097

4. Mosarof, M.H.; Kalam, M.A.; Masjuki, H.H.; Alabdulkarem, A.; Habibullah, M.; Arslan, A.; Monirul, I.M.: Assessment of friction and wear characteristics of *Calophyllum inophyllum* and palm biodiesel. Ind. Crops Prod. 83, 470–483 (2016). https://doi.org/10.1016/J.INDCROP.2015.12.082

5. Mosarof, M.H.; Kalam, M.A.; Masjuki, H.H.; Alabdulkarem, A.; Ashraful, A.M.; Arslan, A.; Rashedul, H.K.; Monirul, I.M.: Optimization of performance, emission, friction and wear characteristics of palm and *Calophyllum inophyllum* biodiesel blends. Energy Convers. Manag. 118, 119–134 (2016). https://doi.org/10.1016/J.ENCONMAN.2016.03.081

6. Parthiban, K.S.; Perumalsamy, M.: Kinetic studies on oil extraction and biodiesel production from underutilized *Annona squamosa* seeds. Fuel 180, 211–217 (2016). https://doi.org/10.1016/j.fuel.2016.04.020

7. Mathiarasi, R.; Partha, N.: Optimization, kinetics and thermodynamic studies on oil extraction from Daturametel Linn oil seed for biodiesel production. Renew. Energy 96, 583–590 (2016). https://doi.org/10.1016/j.renene.2016.04.078

8. Anwar, M.; Rasul, M.G.; Ashwath, N.: The synergistic effects of oxygenated additives on papaya biodiesel binary and ternary blends. Fuel 256, 115980 (2019). https://doi.org/10.1016/j.fuel.2019.115980

9. Anwar, M.; Rasul, M.G.; Ashwath, N.; Nabi, M.D.N.: The potential of utilising papaya seed oil and stone fruit kernel oil as non-edible feedstock for biodiesel production in Australia—a review. Energy Rep. 5, 280–297 (2019). https://doi.org/10.1016/j.egyr.2019.02.007

10. Mofijur, M.; Masjuki, H.H.; Kalam, M.A.; Atabani, A.E.; Fattah, I.M.R.; Mobarak, H.M.: Comparative evaluation of performance and emission characteristics of Moringa oleifera and Palm oil based biodiesel in a diesel engine. Ind. Crops Prod. 53, 78–84 (2014). https://doi.org/10.1016/j.indcrop.2013.12.011

11. Ighose, B.O.; Adeleke, I.A.; Damos, M.; Junaid, H.A.; Okpalaeke, K.E.; Betiku, E.: Optimization of biodiesel production from Thevetia peruviana seed oil by adaptiveneuro-fuzzyinferencesystem coupledwithgeneticalgorithmandresponse surface methodology. Energy Convers. Manag. 132, 231–240 (2017)

12. Nayak, M.G.; Vyas, A.P.: Optimization of microwave-assisted biodiesel production from Papaya oil using response surface methodology. Renew. Energy 138, 18–28 (2019). https://doi.org/10.1016/J.RENENE.2019.01.054

13. Anwar, M.; Rasul, M.G.; Ashwath, N.: Production optimization and quality assessment of papaya (*Carica papaya*) biodiesel with response surface methodology. Energy Convers. Manag. 156, 103–112 (2018). https://doi.org/10.1016/J.ENCONMAN.2017.11.004

14. Sandouqa, A.; Al-Hamamre, Z.: Energy analysis of biodiesel production from jojoba seed oil. Renew. Energy 130, 831–842 (2019). https://doi.org/10.1016/j.renene.2018.07.015

15. Fernandes, K.V.; Papadaki, A.; da Silva, J.A.C.; Fernandez-Lafuente, R.; Koutinas, A.A.; Freire, D.M.G.: Enzymatic esterification of palm fatty-acid distillate for the production of polyol esters with biolubricant properties. Ind. Crops Prod. 116, 90–96 (2018). https://doi.org/10.1016/j.indcrop.2018.02.058

16. Malacrida, C.R.; Kimura, M.; Jorge, N.: Characterization of a high oleic oil extracted from papaya (*Carica papaya* L.) seeds. Ciência e Tecnol. Aliment. 31, 929–934 (2012). https://doi.org/10.1590/s0101-20612011000400016

17. Barroso, P.T.W.; de Carvalho, P.P.; Rocha, T.B.; Pessoa, F.L.P.; Azevedo, D.A.; Mendes, M.F.: Evaluation of the composition of *Carica papaya* L. seed oil extracted with supercritical CO_2. Biotechnol. Rep. 11, 110–116 (2016). https://doi.org/10.1016/J.BTRE.2016.08.004

18. Sharma, S.K., Mitra, S.K., Saran, S.: Papaya production in India—History, present status and future prospects. In: Acta Horticulturae, pp. 87–94. International Society for Horticultural Science (2016)

19. Puangsri, T.; Abdulkarim, S.M.; Ghazali, H.M.: Properties of *Carica papaya* L. (papaya) seed oil following extraction using solvent and aqueous enzymatic methods. J. Food Lipids. 12, 62–76 (2005). https://doi.org/10.1111/j.1745-4522.2005.00006.x

20. Singh, G.; Pai, R.S.; Devi, V.K.: Response surface methodology and process optimization of sustained release pellets using Taguchi orthogonal array design and central composite design. J. Adv. Pharm. Technol. Res. 3, 30–41 (2012). https://doi.org/10.4103/2231-4040.93565

21. Daryono, E.D.; Sinaga, E.J.: Rapid in situ transesterification of papaya seeds to biodiesel with the aid of co-solvent. J. Renew. Energy Res. 7, 379–385 (2017)

22. Chan, H.T.; Heu, R.A.; Tang, C.S.; Okazaki, E.N.; Ishizaki, S.M.: Composition of papaya seeds. J. Food Sci. 43, 255–261 (1978)

23. Kiralan, M.; Özkan, G.; Bayrak, A.; Ramadan, M.F.: Physicochemical properties and stability of black cumin (*Nigella sativa*) seed oil as affected by different extraction methods. Ind. Crops Prod. 57, 52–58 (2014). https://doi.org/10.1016/j.indcrop.2014.03.026

24. Soto, C.; Chamy, R.; Zúñiga, M.E.: Enzymatic hydrolysis and pressing conditions effect on borage oil extraction by cold pressing. Food Chem. 102, 834–840 (2007). https://doi.org/10.1016/j.foodchem.2006.06.014

25. Mohammed, N.K.; AbdManap, M.Y.; Tan, C.P.; Muhialdin, B.J.; Alhelli, A.M.; Hussin, A.S.M.: The effects of different extraction methods on antioxidant properties, chemical composition, and thermal behavior of black seed (*Nigella sativa* L.) oil. Evidence-based complement. Altern. Med. (2016). https://doi.org/10.1155/2016/6273817

26. Zhang, S.; Zu, Y.-G.; Fu, Y.-J.; Luo, M.; Liu, W.; Li, J.; Efferth, T.: Supercritical carbon dioxide extraction of seed oil from yellow horn (*Xanthoceras sorbifolia* Bunge.) and its anti-oxidant activity. Bioresour. Technol. 101, 2537–2544 (2009). https://doi.org/10.1016/j.biortech.2009.11.082

27. Shah, S.; Sharma, A.; Gupta, M.N.: Extraction of oil from *Jatropha curcas* L. seed kernels by combination of ultrasonication and aqueous enzymatic oil extraction. Bioresour. Technol. 96, 121–123 (2005). https://doi.org/10.1016/j.biortech.2004.02.026

28. Suryawanshi, B.; Mohanty, B.: Application of an artificial neural network model for the supercritical fluid extraction of seed oil from *Argemone mexicana* (L.) seeds. Ind. Crops Prod. 123, 64–74 (2018). https://doi.org/10.1016/j.indcrop.2018.06.057

29. Devi, V.; Khanam, S.: Development of generalized and simplified models for supercritical fluid extraction: case study of papaya (*Carica papaya*) seed oil. Chem. Eng. Res. Des. 150, 341–358 (2019). https://doi.org/10.1016/J.CHERD.2019.08.006

30. Ajiwe, V.I.E.; Umerie, S.C.; Okeke, C.A.; Oburota, V.N.: Extraction and utilisation of cassava seed oil. Bioresour. Technol. 47, 85–86 (1994). https://doi.org/10.1016/0960-8524(94)90033-7

31. Okieimen, F.E.; Eromosele, C.O.: Fatty acid composition of the seed oil of Khaya senegalensis. Bioresour. Technol. 69, 279–280 (1999). https://doi.org/10.1016/S0960-8524(98)00190-4

32. Sarin, R.; Sharma, M.; Khan, A.A.: *Terminalia belerica* Roxb seed oil: a potential biodiesel resource. Bioresour. Technol. 101, 1380–1384 (2010). https://doi.org/10.1016/j.biortech.2009.09.018

33. Ajiwe, V.I.E.; Okeke, C.A.; Agbo, H.U.: Extraction and utilization of breadfruit seed oil (*Treculia africana*). Bioresour. Technol. **53**, 183–184 (1995). https://doi.org/10.1016/0960-8524(95)00059-N

34. Samaram, S.; Mirhosseini, H.; Tan, C.P.; Ghazali, H.M.: Ultrasound-assisted extraction and solvent extraction of papaya seed oil: crystallization and thermal behavior, saturation degree, color and oxidative stability. Ind. Crops Prod. **52**, 702–708 (2014). https://doi.org/10.1016/j.indcrop.2013.11.047

35. Samaram, S.; Mirhosseini, H.; Tan, C.P.; Ghazali, H.M.; Bordbar, S.; Serjouie, A.: Optimisation of ultrasound-assisted extraction of oil from papaya seed by response surface methodology: oil recovery, radical scavenging antioxidant activity, and oxidation stability. Food Chem. **172**, 7–17 (2015). https://doi.org/10.1016/J.FOODCHEM.2014.08.068

36. Venter, G.: Review of optimization techniques. In: Blockley, R. (ed.) Encyclopedia of Aerospace Engineering. Wiley, Chichester (2010)

37. Nuchitprasittichai, A.; Cremaschi, S.: Optimization of CO_2 capture process with aqueous amines using response surface methodology. Comput. Chem. Eng. **35**, 1521–1531 (2011). https://doi.org/10.1016/J.COMPCHEMENG.2011.03.016

38. Bhran, A.A.; Shoaib, A.M.; Umana, B.: Optimization of crude oil hydrotreating process as a function of operating conditions: application of response surface methodology. Comput. Chem. Eng. **89**, 158–165 (2016). https://doi.org/10.1016/J.COMPCHEMENG.2016.03.026

39. Aniceto, J.P.S.; Cardoso, S.P.; Silva, C.M.: General optimization strategy of simulated moving bed units through design of experiments and response surface methodologies. Comput. Chem. Eng. **90**, 161–170 (2016). https://doi.org/10.1016/J.COMPCHEMENG.2016.04.028

40. Vargas, E.M.; Ospina, J.L.; Tarelho, L.A.C.; Nunes, M.I.: FAME production from residual materials: optimization of the process by Box–Behnken model. Energy Reports. **6**, 347–352 (2020). https://doi.org/10.1016/j.egyr.2019.08.071

41. Holland, D.G.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Boston (1989)

42. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P.: Optimization by simulated annealing (80-). Science **220**, 671–680 (1983). https://doi.org/10.1126/science.220.4598.671

43. Rabbouch, B.; Saâdaoui, F.; Mraihi, R.: Empirical-type simulated annealing for solving the capacitated vehicle routing problem. J. Exp. Theor. Artif. Intell. (2019). https://doi.org/10.1080/0952813X.2019.1652356

44. DORIGO, M.: Optimization, Learning and Natural Algorithms. Ph.D. Thesis, Politec, di Milano (1992)

45. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of the IEEE International Joint Conference on Neural Network,. pp. 1942–1948 (1995)

46. Rabbouch, B.; Mraihi, R.; Saâdaoui, F.: A recent brief survey for the multi depot heterogenous vehicle routing problem with time windows. In: Abraham, A., Muhuri, P., Muda, A., Gandhi, N. (eds.) Advances in Intelligent Systems and Computing, vol. 734, pp. 147–157. Springer, Cham (2018)

47. Geem, Z.W.; Kim, J.H.; Loganathan, G.V.: A new heuristic optimization algorithm: harmony search. Simulation. **76**, 60–68 (2001). https://doi.org/10.1177/003754970107600201

48. Askarzadeh, A.: A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm. Comput. Struct. **169**, 1–12 (2016). https://doi.org/10.1016/j.compstruc.2016.03.001

49. Samantha, B.S.; Trupthi, M.; Sairam, U.: A review on using crow search algorithm in solving the problems of constrained optimization. Int. J. Sci. Res. Sci. Tech. **4**, 1374–1387 (2018)

50. Javidi, A.; Salajegheh, E.; Salajegheh, J.: Enhanced crow search algorithm for optimum design of structures. Appl. Soft Comput. J. **77**, 274–289 (2019). https://doi.org/10.1016/j.asoc.2019.01.026

51. Gupta, D.; Sundaram, S.; Khanna, A.; Ella Hassanien, A.; de Albuquerque, V.H.C.: Improved diagnosis of Parkinson's disease using optimized crow search algorithm. Comput. Electr. Eng. **68**, 412–424 (2018). https://doi.org/10.1016/j.compeleceng.2018.04.014

52. Horng, S.C.; Lin, S.S.: Merging crow search into ordinal optimization for solving equality constrained simulation optimization problems. J. Comput. Sci. **23**, 44–57 (2017). https://doi.org/10.1016/j.jocs.2017.10.001

53. Nobahari, H., Bighashdel, A.: MOCSA: A multi-objective crow search algorithm for multi-objective optimization. In: 2017 2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC), pp. 60–65. IEEE (2017)

54. Tamilarasan, A., Renugambal, A., Manikanta, D., Sekhar Reddy, G.B.C., Sravankumar, K., Sreekar, B., Prasadreddy, G. V.: Application of crow search algorithm for the optimization of abrasive water jet cutting process parameters. In: IOP Conferernce on Series Material Science Engineering, vol. 390, (2018). https://doi.org/10.1088/1757-899X/390/1/012034

55. Sayed, G.I., Darwish, A., Hassanien, A.E.: Chaotic crow search algorithm for engineering and constrained problems. In: 2017 12th International Conference on Computer Engineering and Systems (ICCES), pp. 676–681. IEEE (2017)

56. Ela, A.A.A. El, El-Sehiemy, R.A., Shaheen, A.M., Shalaby, A.S.: Application of the crow search algorithm for economic environmental dispatch. In: 2017 Nineteenth International Middle East Power Systems Conference (MEPCON), pp. 78–83. IEEE (2017)

57. Zaki, D.A., Hasanien, H.M., El-Amary, N.H., Abdelaziz, A.Y.: Crow search algorithm for improving the performance of an inverter-based distributed generation system. In: 2017 Nineteenth International Middle East Power Systems Conference (MEPCON), pp. 656–663. IEEE (2017)

58. Kumar, U.A.: Comparison of neural networks and regression analysis: a new insight. Expert Syst. Appl. **29**, 424–430 (2005). https://doi.org/10.1016/j.eswa.2005.04.034

59. Myers, R.H.; Montgomery, D.C.; Vining, G.G.; Robinson, T.J.: Generalized Linear Models: with Applications in Engineering and the Sciences, 2nd edn. Wiley, New York (2010)

60. Moradi, M.; Arabi, H.; Shamsborhan, M.: Multi-objective optimization of high power diode laser surface hardening process of AISI 410 by means of RSM and desirability approach. Optik (Stuttg) (2019). https://doi.org/10.1016/j.ijleo.2019.163619

61. Palanikumar, K.; Muthukrishnan, N.; Hariprasad, K.S.: Surface roughness parameters optimization in machining A356/SiC/20p metal matrix composites by PCD tool using response surface methodology and desirability function. Mach. Sci. Technol. **12**, 529–545 (2008). https://doi.org/10.1080/10910340802518850

62. Chabbi, A.; Yallese, M.A.; Meddour, I.; Nouioua, M.; Mabrouki, T.; Girardin, F.: Predictive modeling and multi-response optimization of technological parameters in turning of Polyoxymethylene polymer (POM C) using RSM and desirability function. Meas. J. Int. Meas. Confed. **95**, 99–115 (2017). https://doi.org/10.1016/j.measurement.2016.09.043

63. Montgomery, D.C.; Runger, G.C.: Applied Statistics and probability for engineers. John Wiley & Sons, Ltd (2002)

64. Montgomery, D.C.: Design and Analysis of Experiments. Wiley, New York (2001)

65. Box, G.E.P.; Hunter, W.G.; Hunter, J.S.: Statistics for Experimenters. Wiley, New York (1978)

66. Aslan, N.; Cebeci, Y.: Application of Box–Behnken design and response surface methodology for modeling of some Turkish coals. Fuel **86**, 90–97 (2007). https://doi.org/10.1016/j.fuel.2006.06.010

67. Hajslova, J.; Cajka, T.: Gas chromatography–mass spectrometry (GC–MS). In: Yolanda, P. (ed.) Food Toxicants Analysis, pp. 419–473. Elsevier, Amsterdam (2007)

68. Hunter, E.A.; Haaland, P.D.: Experimental Design in Biotechnology. Biometrics **47**, 348 (2006). https://doi.org/10.2307/2532528

69. Cicero, A.F.G.; Nascetti, S.; López-Sabater, M.C.; Elosua, R.; Salonen, J.T.; Nyyssönen, K.; Poulsen, H.E.; Zunft, H.-J.F.; Kiesewetter, H.; de la Torre, K.; Covas, M.-I.; Kaikkonen, J.; Mursu, J.; Koenbick, C.; Bäumler, H.; Gaddi, A.V.: EUROLIVE Study Group: changes in LDL fatty acid composition as a response to olive oil treatment are inversely related to lipid oxidative damage: The EUROLIVE study. J. Am. Coll. Nutr. **27**, 314–320 (2008)