J-NPC: Journal of Natural Products and Chemometrics
Vol. 1, No. 1, 15 May 2026, pp. 11–21

ANOVA-Based Evaluation of a Box-Behnken Design for Total Phenolic Content in *Morus alba* var. *shalun*

Taufik Hudha Nursyafaah[1], Muhammad Rizky[1], Fitri Aditri[1], M Miftahul Madya[1*]

[1]Lembaga Riset AI Creation (LRAC), Depok, Indonesia

*Corresponding author email: mmiftahulm29@gmail.com

(Received 31 March 2026 , Accepted 5 April 2026, Available Online 15 May 2026)

Abstract. This study optimized the extraction of total phenolic content (TPC) from *Morus alba* var. *shalun* leaves using Response Surface Methodology (RSM) with a Box–Behnken Design (BBD). Ethanol concentration, extraction temperature, and extraction time were evaluated as independent variables affecting TPC extraction. A total of 17 experimental runs were conducted, and the data were analyzed using Analysis of Variance (ANOVA) and regression modeling. Prior to analysis, the response data were transformed using natural logarithm transformation based on Box–Cox analysis to improve model adequacy. The developed model was highly significant ($p < 0.0001$) with a high coefficient of determination ($R^2 = 0.9953$), indicating excellent agreement between predicted and experimental values. Ethanol concentration was identified as the most influential factor, followed by extraction time and temperature. Diagnostic analyses confirmed good model stability and predictive capability. Overall, the developed RSM model effectively optimized TPC extraction from *Morus alba* var. *shalun*.

Keywords: *Morus alba* var. *shalun*; total phenolic content; extraction optimization; Response Surface Methodology; Box–Behnken Design

Abstrak. Penelitian ini bertujuan untuk mengoptimalkan ekstraksi total phenolic content (TPC) dari daun *Morus alba* var. *shalun* menggunakan Response Surface Methodology (RSM) dengan desain Box–Behnken (BBD). Konsentrasi etanol, suhu ekstraksi, dan waktu ekstraksi digunakan sebagai variabel independen yang mempengaruhi ekstraksi TPC. Sebanyak 17 percobaan dilakukan sesuai matriks BBD, kemudian data dianalisis menggunakan Analysis of Variance (ANOVA) dan pemodelan regresi. Sebelum analisis statistik dilakukan, data respon ditransformasikan menggunakan logaritma natural berdasarkan analisis Box–Cox untuk meningkatkan kecukupan model. Model yang dikembangkan menunjukkan hasil sangat signifikan ($p < 0{,}0001$) dengan nilai koefisien determinasi tinggi ($R^2 = 0{,}9953$), yang menunjukkan kesesuaian sangat baik antara nilai prediksi dan eksperimen. Konsentrasi etanol menjadi faktor paling berpengaruh terhadap ekstraksi TPC, diikuti waktu dan suhu ekstraksi. Analisis diagnostik menunjukkan model memiliki kestabilan dan kemampuan prediksi yang baik. Secara keseluruhan, model RSM berhasil mengoptimalkan ekstraksi TPC dari *Morus alba* var. *shalun*.

Kata kunci: *Morus alva* var. *shalun*; Total phenolic content; Optimasi ekstraksi; Response Surface Methodology; Box-Behnken Design

1. Introduction

The increasing demand for natural bioactive compounds has stimulated extensive research on plant-derived secondary metabolites, particularly phenolic compounds. Phenolic compounds are widely recognized for their important biological activities, including antioxidant, anti-inflammatory, antimicrobial, and therapeutic properties beneficial to human health. In recent years, natural phenolics have attracted considerable attention as potential alternatives to synthetic antioxidants in food, pharmaceutical, and nutraceutical applications.Among various medicinal plants, the genus Morus has been reported to contain diverse biologically active secondary metabolites, especially phenolic compounds. Previous phytochemical studies identified several major classes of compounds from Morus species, including 2-arylbenzofurans, stilbenes, flavonoids, and Diels–Alder adducts [1–4]. These metabolites contribute to numerous biological activities such as antioxidant, anti-inflammatory, antibacterial, antifungal, antimicrobial, antiviral, and cytotoxic effects [2,5–8]. Therefore, Morus species are considered promising sources of natural bioactive compounds for health-related applications.

Morus alba is one of the most extensively studied species due to its high phenolic content and remarkable pharmacological potential. The leaves of Morus alba have long been utilized in traditional medicine and functional food applications because they contain flavonoids, phenolics, and other secondary metabolites associated with antioxidant activity and glucose metabolism regulation [9,10]. One promising variety is *Morus alba* var. *shalun*, which has been reported to possess a rich phytochemical composition and significant antioxidant potential [4]. Despite its promising bioactivity, the recovery of phenolic compounds from plant materials strongly depends on extraction efficiency. Several extraction parameters, including solvent concentration, extraction temperature, and extraction time, significantly influence the yield of total phenolic compounds obtained [11]. Ethanol is commonly used as a green solvent in phenolic extraction because of its safety, effectiveness, low toxicity, and suitability for food and pharmaceutical applications.

Optimization of extraction conditions is essential to maximize phenolic recovery. Conventional optimization methods generally employ the one-factor-at-a-time (OFAT) approach, where one variable is changed while the others are kept constant. However, this approach is inefficient, requires a large number of experiments, and cannot adequately evaluate interactions among variables. To overcome these limitations, statistical and mathematical approaches such as Response Surface Methodology (RSM) have been widely applied in extraction optimization studies because they allow simultaneous evaluation of multiple variables and their interactions [12]. Among the available RSM designs, the Box–Behnken Design (BBD) is considered one of the most efficient experimental designs for process optimization. This design reduces the number of experimental runs while enabling the development of quadratic regression models without involving extreme experimental conditions [13]. Furthermore, BBD allows systematic evaluation of linear, interaction, and quadratic effects of independent variables on the response.

In addition, Analysis of Variance (ANOVA) is an important statistical tool for validating the developed regression model. ANOVA is used to determine the significance of individual variables and their interactions, as well as to evaluate the adequacy and predictive capability of the model. A statistically significant model with a high coefficient of determination (R^2) indicates that the variability of the response can be effectively explained by the selected variables [14]. Although several studies have reported the optimization of phenolic extraction from various plant materials using RSM, studies focusing specifically on the statistical optimization of total phenolic content extraction from *Morus alba* var. shalun remain limited. Most previous studies primarily emphasized phytochemical identification and biological activity evaluation, whereas systematic optimization of extraction parameters and interaction analysis for maximizing total phenolic content have not been comprehensively investigated. Therefore, this study aimed to optimize the extraction conditions for total phenolic content (TPC) from *Morus alba* var. *shalun* leaves using Response Surface Methodology with a Box–Behnken Design. The effects of ethanol concentration, extraction temperature, and extraction time, as well as their interactions, were systematically evaluated to develop a predictive quadratic model and determine the optimum extraction conditions for maximizing phenolic yield.

2. Method and Experimental

Experimental Design

The optimization of Total Phenolic Content (TPC) extraction from *Morus alba* var. *shalun* was performed using Response Surface Methodology (RSM) with a Box–Behnken Design (BBD) (Fig 1). This design was selected because of its efficiency in evaluating quadratic response surfaces with a relatively small number of experimental runs, making it suitable for multivariable optimization studies [6]. In this study, three independent variables were investigated, namely ethanol concentration (A), extraction temperature (B), and extraction time (C). Each variable was evaluated at three coded levels, represented as −1, 0, and +1, corresponding to low, middle, and high levels, respectively. Based on the Box–Behnken configuration, a total of 17 experimental runs were conducted, including several replicated center points to estimate experimental error and improve the reliability of the statistical model.

The extraction process was carried out using ethanol as the extraction solvent, with concentrations adjusted according to the experimental design matrix. *Morus alba* var. *shalun* plant materials were extracted under controlled conditions, where extraction temperature and extraction time were systematically varied according to the Box–Behnken

experimental combinations. After extraction, the samples were filtered and prepared for further analysis. Total phenolic content (TPC) was subsequently determined as the main response variable in this study.

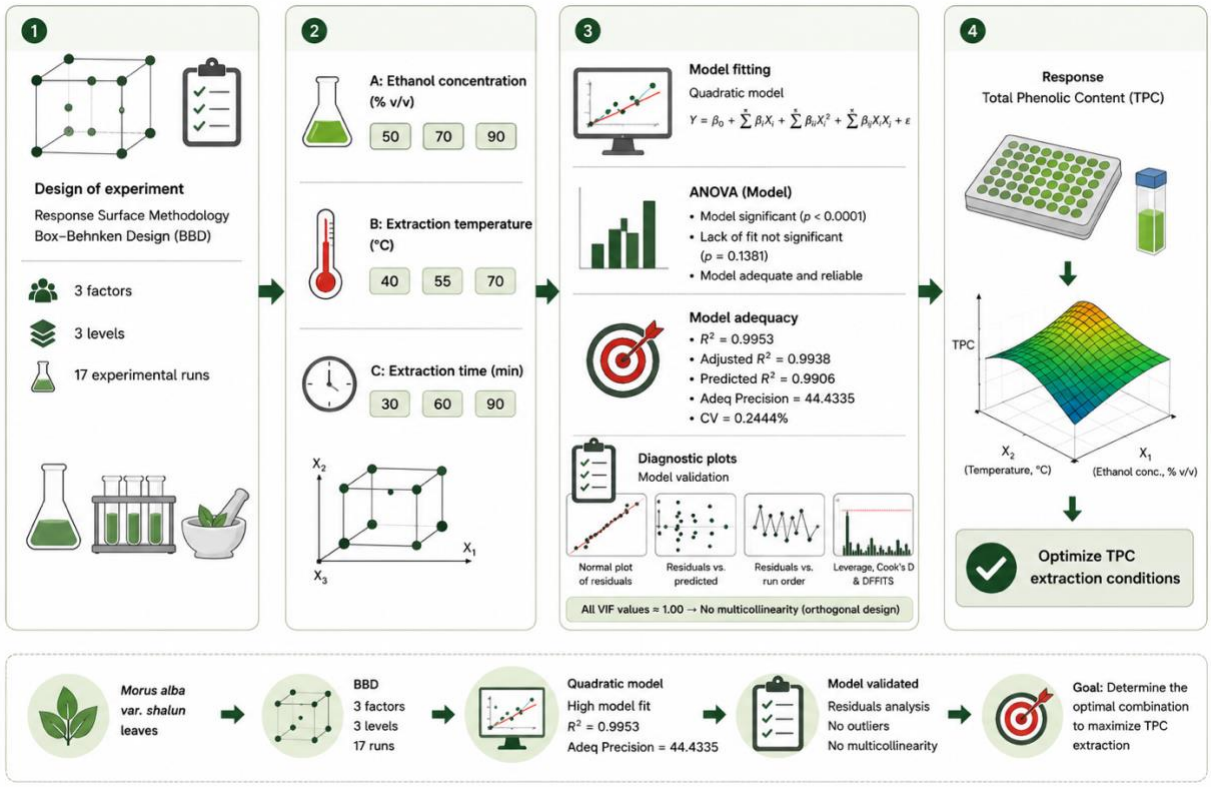

Fig 1. Experimental design, statistical modeling, and optimization workflow for total phenolic content (TPC) extraction from Morus alba var. shalun using Response Surface Methodology (RSM) and Box–Behnken Design (BBD).

Data Transformation

Prior to statistical analysis, the response data were evaluated using a Box–Cox transformation plot to determine the most appropriate transformation for improving the model assumptions. Based on the evaluation results, the response data were transformed using the natural logarithm (ln) transformation to improve variance stability, enhance the normal distribution of residuals, and increase the adequacy of the regression model, as commonly recommended in experimental design studies [7].

Statistical Modeling and ANOVA Analysis

Experimental data obtained from the Box–Behnken Design were analyzed using Analysis of Variance (ANOVA) to evaluate the significance and adequacy of the developed regression model. A second-order quadratic polynomial model was applied to describe the relationship between the independent variables and the response variable, namely total phenolic content (TPC). Prior to statistical analysis, the response data were transformed using the natural logarithm transformation based on the Box–Cox transformation recommendation to improve model adequacy and variance stability. The developed quadratic regression model included: linear effects of ethanol concentration (A), extraction temperature (B), and extraction time (C), interaction effects between variables (AB, AC, and BC), and quadratic effects (A^2, B^2, and C^2). The significance of the regression model and each model term was evaluated using

F-values and p-values at a significance level of $p < 0.05$. Statistical parameters used to assess model adequacy included the coefficient of determination (R^2), adjusted R^2, predicted R^2, coefficient of variation (CV), and Adeq Precision. The agreement between predicted R^2 and adjusted R^2 was also evaluated to determine the predictive capability of the model within the studied design space. The regression coefficients were further analyzed using coefficient estimates, standard errors, confidence intervals, and variance inflation factors (VIFs) to evaluate the contribution of each factor and to identify potential multicollinearity among variables. Model validation was conducted using diagnostic analyses generated by Design-Expert software, including Box–Cox transformation plots, Predicted vs Actual plots, residual analysis, leverage values, internally and externally studentized residuals, Cook's distance, and DFFITS values to evaluate model reliability and detect influential observations or potential outliers.

Model Validation and Diagnostic Analysis

Model validation was performed through graphical and diagnostic parameter analyses. The agreement between experimental and predicted response values was evaluated using a Predicted vs Actual plot, which demonstrates the predictive performance of the developed model. Residual analysis was conducted using Residuals vs Predicted plots to ensure that residuals were randomly distributed around zero without showing any systematic pattern, thereby satisfying the fundamental assumptions of regression analysis. In addition, several diagnostic parameters were applied to identify potential outliers or observations exerting excessive influence on the regression model, including: leverage, internally studentized residuals, externally studentized residuals, Cook's distance, and DFFITS (Difference in Fits). These parameters were analyzed according to the standard experimental order to ensure that no individual data point excessively affected the stability of the regression model.

Software and Data Analysis

All stages of the study, including experimental design, regression modeling, ANOVA analysis, response surface evaluation, and diagnostic analysis, were conducted using Design-Expert software version 12. The raw dataset used in this study consisted of three main factors: factor A representing ethanol concentration (%), factor B representing extraction temperature (°C), and factor C representing extraction time (min). The response values presented in Table 1 were obtained from laboratory experiments. To improve data stability and representativeness, a bootstrap sampling approach was applied using the bootstrapping concept, in which random samples were generated from the original dataset through sampling with replacement. This procedure generated multiple bootstrap datasets that allowed repeated observations. Several representative bootstrap data were subsequently selected and presented in Table 1 as the dataset used for regression modeling and statistical analysis.

Raw data

The raw experimental data presented in Table 1 were generated using the Box–Behnken Design (BBD) consisting of 17 experimental runs with three independent variables, namely ethanol concentration (%), extraction temperature (°C), and extraction time (min). Each factor was evaluated at three levels to investigate its effect on total phenolic content (TPC) expressed as mg GAE/g. The experimental TPC values ranged from 35.8 to 49.4 mg GAE/g, indicating variations in extraction efficiency under different experimental conditions. The obtained data were subsequently analyzed using Response Surface Methodology (RSM), ANOVA, and regression modeling in Design-Expert software version 12 to evaluate model adequacy and determine the optimum extraction conditions.

Table 1. Raw data factor and response

Run	Factor 1 Ethanol concentration	Factor 2 Temperature	Factor 3 Time	Response 1 Total phenolic content (mg GAE/g)
1	60	70	30	43.6
2	60	40	30	39.1
3	60	55	60	46.2
4	40	70	60	41.9
5	80	70	60	48.7
6	80	40	60	45.4
7	40	40	60	38.6
8	40	55	30	35.8
9	80	55	30	42
10	60	40	90	44.2
11	80	55	90	47.8
12	40	55	90	40.7
13	60	55	60	46.2
14	60	55	60	46.2
15	60	70	90	49.4
16	60	55	60	46.2
17	60	55	60	46.2

3. Results and Discussion

The experimental results obtained from the Box–Behnken Design (BBD) were analyzed using Analysis of Variance (ANOVA) to evaluate the significance and adequacy of the quadratic model for total phenolic content (TPC). The ANOVA results are summarized in Table 2.

Table 2 ANOVA for Quadratic Model of Total Phenolic Content

Source	Sum of Squares	df	Mean Square	F-value	p-value	Significance
Model	0.1270	9	0.0141	165.22	<0.0001	Significant
A (Ethanol concentration)	0.0501	1	0.0501	586.70	<0.0001	Significant
B (Temperature)	0.0173	1	0.0173	202.91	<0.0001	Significant
C (Time)	0.0319	1	0.0319	373.41	<0.0001	Significant
AB	0.0000	1	0.0000	0.4122	0.5413	Not significant
AC	2.893E-07	1	2.893E-07	0.0034	0.9552	Not significant
BC	1.312E-06	1	1.312E-06	0.0154	0.9048	Not significant
A^2	0.0154	1	0.0154	179.82	<0.0001	Significant
B^2	9.574E-08	1	9.574E-08	0.0011	0.9742	Not significant
C^2	0.0107	1	0.0107	125.04	<0.0001	Significant
Residual	0.0006	7	0.0001	—	—	—

The ANOVA results demonstrated that the developed quadratic model for total phenolic content (TPC) was highly significant, with an F-value of 165.22 and a p-value lower than 0.0001 (Table 1). These results indicate that the model adequately describes the relationship between the extraction variables and the response, with only a very small probability that the observed model significance occurred due to random error. The significance of the model confirms that the selected extraction parameters substantially contributed to the variation in TPC. Among the linear factors, ethanol concentration (A), extraction temperature (B), and extraction time (C) showed highly significant effects on TPC ($p < 0.0001$). Ethanol concentration exhibited the highest F-value (586.70), indicating that it was the most influential factor affecting phenolic extraction. Extraction time also showed a strong contribution with an F-value of 373.41, followed by extraction temperature with an F-value of 202.91. These findings suggest that changes in extraction conditions directly affected the efficiency of phenolic compound recovery. In contrast, the interaction terms AB, AC, and BC were not statistically significant ($p > 0.05$), indicating that the combined interactions among ethanol concentration, extraction temperature, and extraction time did not significantly influence the TPC response within the

studied experimental range. Similarly, the quadratic effect of temperature (B^2) was not significant ($p = 0.9742$), suggesting that temperature exhibited a relatively linear effect under the investigated conditions. However, the quadratic terms A^2 and C^2 were highly significant ($p < 0.0001$), indicating the presence of curvature effects in the response surface. This result suggests that optimum levels of ethanol concentration and extraction time existed for maximizing TPC, beyond which the extraction efficiency no longer increased proportionally.

Model Fitting and Statistical Validation

The adequacy and reliability of the developed model were further evaluated using several statistical parameters, as presented in Table 3. The coefficient of determination ($R^2 = 0.9953$) indicated that 99.53% of the variability in total phenolic content (TPC) could be explained by the developed model, demonstrating an excellent agreement between the experimental and predicted values.

Table 3 Model Fit Statistics

Parameter	Value
R^2	0.9953
Adjusted R^2	0.9893
Predicted R^2	0.9250
Std. Dev.	0.0092
Mean	3.78
C.V. (%)	0.2444
Adeq Precision	44.4335

The adjusted R^2 (0.9893) and predicted R^2 (0.9250) values were reasonably close to each other, with a difference lower than 0.2, indicating good predictive capability and satisfactory model consistency. This result suggests that the developed model possesses strong reliability for predicting TPC within the investigated experimental range. The low standard deviation value (0.0092) and coefficient of variation (CV = 0.2444%) further demonstrated the high precision and reproducibility of the experimental data. Generally, a low CV value indicates good experimental reliability and minimal variation among experimental runs. In addition, the Adeq Precision value of 44.4335 was substantially higher than the minimum desirable value of 4, indicating an adequate signal-to-noise ratio. This confirms that the model provided sufficient sensitivity and could be reliably used to navigate and optimize the experimental design space for maximizing total phenolic content extraction.

Regression Coefficients and Model Interpretation

The regression coefficients of the quadratic model are presented in Table 4. The positive coefficients of the linear terms, namely ethanol concentration (A), extraction temperature (B), and extraction time (C), indicate that increasing these variables contributed positively to the total phenolic content (TPC) response within the studied experimental range. Among the linear coefficients, ethanol concentration showed the highest coefficient value (0.0791), followed by extraction time (0.0631) and extraction temperature (0.0465), suggesting that ethanol concentration exerted the strongest influence on phenolic extraction efficiency.

Table 4 Regression Coefficients (Coded Factors)

Factor	Coefficient
Intercept	3.83
A (Ethanol concentration)	0.0791
B (Temperature)	0.0465
C (Time)	0.0631
AB	-0.0030
AC	0.0003
BC	0.0006
A^2	-0.0604
B^2	-0.0002
C^2	-0.0504

The interaction coefficients (AB, AC, and BC) exhibited relatively small values, indicating weak interaction effects among the investigated variables. This observation is consistent with the ANOVA results, where the interaction terms were statistically insignificant ($p > 0.05$). Therefore, the extraction variables primarily influenced the response independently rather than through combined interaction effects. The quadratic coefficients of ethanol concentration ($A^2 = -0.0604$) and extraction time ($C^2 = -0.0504$) were negative, indicating the presence of curvature effects in the response surface. These negative coefficients suggest that excessive levels of ethanol concentration and extraction time may reduce the extraction efficiency after reaching optimum conditions. In contrast, the quadratic coefficient of temperature ($B^2 = -0.0002$) was very small, indicating a relatively limited curvature effect of temperature within the investigated range. Furthermore, the variance inflation factor (VIF) (Table 5) values for all model terms were close to 1, indicating the absence of multicollinearity among the independent variables. This result confirms that the regression model was statistically stable and that the estimated coefficients could be interpreted reliably. The regression coefficients presented in Table 5 describe the contribution of each coded factor to the total phenolic content (TPC) response. Positive coefficients for the linear terms of ethanol concentration (A), extraction temperature (B), and extraction time (C) indicate that increasing these variables positively influenced TPC extraction within the investigated range. Among these variables, ethanol concentration exhibited the highest coefficient estimate (0.0791), suggesting that it was the most influential factor affecting phenolic recovery.

Table 5 Coefficients in Terms of Coded Factors

Factor	Coefficient Estimate	df	Standard Error	95% CI Low	95% CI High	VIF
Intercept	3.83	1	0.0041	3.82	3.84	
A-Konsentrasi Etanol	0.0791	1	0.0033	0.0714	0.0869	1.0000
B-Suhu ekstraksi	0.0465	1	0.0033	0.0388	0.0543	1.0000
C-Waktu ekstraksi	0.0631	1	0.0033	0.0554	0.0709	1.0000
AB	-0.0030	1	0.0046	-0.0139	0.0080	1.0000
AC	0.0003	1	0.0046	-0.0107	0.0112	1.0000
BC	0.0006	1	0.0046	-0.0104	0.0115	1.0000
A^2	-0.0604	1	0.0045	-0.0710	-0.0497	1.01
B^2	-0.0002	1	0.0045	-0.0108	0.0105	1.01
C^2	-0.0504	1	0.0045	-0.0610	-0.0397	1.01

The coefficient estimates represent the expected change in the response value for every unit change in the corresponding factor while the remaining variables are maintained constant. In an orthogonal experimental design, the intercept value corresponds to the overall average response obtained from all experimental runs, whereas the regression coefficients represent adjustments relative to this average according to the specific factor levels applied during the experiments. In contrast, the interaction terms (AB, AC, and BC) showed relatively small coefficient values, indicating weak interaction effects among the extraction variables. This observation is consistent with the ANOVA results, where the interaction terms were statistically insignificant. The negative coefficients of the quadratic terms A^2 and C^2 indicate the presence of curvature effects in the response surface, suggesting that excessive ethanol concentration and extraction time may reduce extraction efficiency beyond the optimum conditions. Furthermore, all variance inflation factor (VIF) values were approximately 1, indicating the absence of multicollinearity among the independent variables. Generally, VIF values greater than 1 indicate increasing correlations among factors, whereas values lower than 10 are considered acceptable for regression analysis. Therefore, the obtained VIF values confirmed the statistical stability and reliability of the developed regression model.

Box–Cox Transformation Analysis

The Box–Cox transformation analysis (Fig 2) was performed to evaluate the adequacy of the response data and determine the most appropriate transformation for improving the statistical assumptions of the regression model. This analysis is commonly applied in Response Surface Methodology (RSM) studies to stabilize variance, improve residual normality, and enhance the reliability of model prediction.

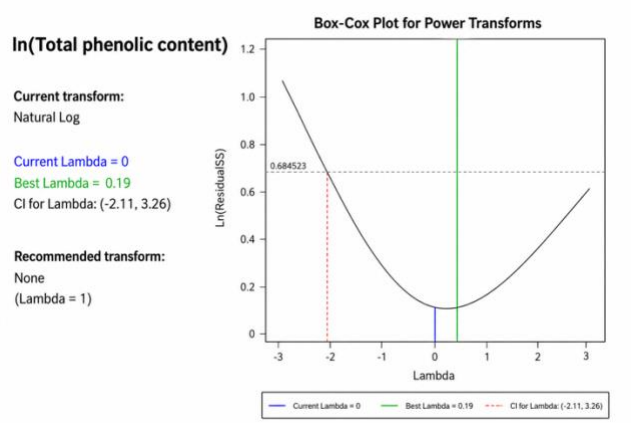

Fig 2. Box–Cox transformation plot for total phenolic content (TPC) response showing the optimal lambda value and suitability of natural logarithm transformation for regression model development.

Based on the Box–Cox transformation plot, the current lambda value was 0, indicating that the response data were transformed using the natural logarithm transformation. The best lambda value obtained from the analysis was 0.19, which is very close to zero, confirming that the natural logarithm transformation was appropriate for the experimental data. In addition, the confidence interval for lambda ranged from −2.11 to 3.26, which included the value of $\lambda = 1$. This result indicates that the original response data were also statistically acceptable without transformation. However, the logarithmic transformation was still applied to improve model stability and residual distribution. The Box–Cox plot also demonstrated a clear minimum point around the selected lambda region, indicating improved model fitting after transformation. The transformed response values showed better variance homogeneity and reduced the possibility of heteroscedasticity in the regression analysis. Therefore, the natural logarithm transformation was considered suitable for subsequent statistical modeling and ANOVA analysis.

Model Validation Using Predicted versus Actual Plot

The predictive capability and adequacy of the developed regression model were further evaluated using the Predicted versus Actual plot, as presented in fig 3. This graphical analysis is commonly employed in Response Surface Methodology (RSM) studies to assess the agreement between experimentally observed values and model-predicted responses. A reliable regression model is generally characterized by data points distributed closely around the diagonal line, indicating strong consistency between predicted and actual values.

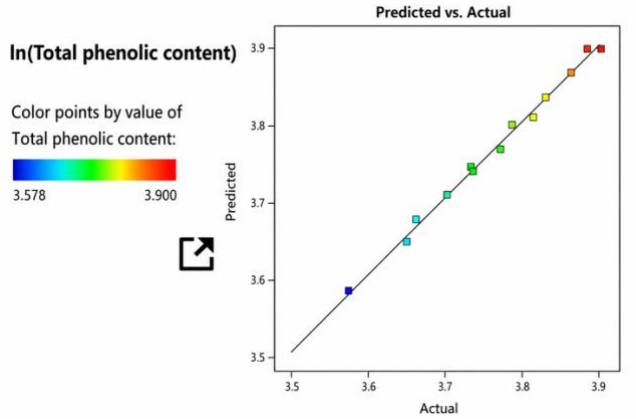

Fig 3. Predicted versus Actual plot showing the agreement between experimental and model-predicted values of total phenolic content (TPC) after natural logarithm transformation.

In the present study, the Predicted versus Actual plot demonstrated that nearly all experimental data points were located very close to the diagonal reference line. This result indicates an excellent agreement between the

predicted values generated by the quadratic model and the experimentally observed total phenolic content (TPC) values. The close clustering of the data points around the regression line suggests that the developed model possessed high predictive accuracy and minimal prediction error.The transformed response values ranged from approximately 3.578 to 3.900, corresponding to the natural logarithm transformation of TPC values. The application of the logarithmic transformation contributed to improving the distribution of the response data and enhancing the linearity between predicted and observed values. This finding is consistent with the Box–Cox transformation analysis, which indicated that the natural logarithm transformation was appropriate for stabilizing variance and improving model adequacy. In addition, the absence of substantial deviation or scattered outlier points from the diagonal line suggests that the residual errors were relatively small and randomly distributed.

This observation indicates that the regression model was capable of accurately capturing the relationship between the extraction variables and the TPC response within the investigated experimental region. The strong agreement observed in the plot further supports the high coefficient of determination ($R^2 = 0.9953$) obtained from the ANOVA analysis, confirming that the model successfully explained most of the variability in the experimental data. The color gradient displayed in the plot represents the magnitude of total phenolic content values, ranging from low values (blue region) to high values (red region). The gradual and consistent distribution of colors along the regression line further demonstrates the robustness and stability of the developed model across the studied response range. Overall, the Predicted versus Actual analysis confirmed that the developed quadratic model was statistically adequate, reliable, and highly suitable for predicting and optimizing total phenolic content extraction from *Morus alba* var. *shalun* under the investigated extraction conditions.

Diagnostic Analysis of Residuals and Influential Observations

The diagnostic statistics presented in Table 6 were used to evaluate the adequacy, stability, and reliability of the developed regression model. Several diagnostic parameters, including residuals, leverage values, internally studentized residuals, externally studentized residuals, Cook's distance, and DFFITS, were analyzed to identify potential outliers and influential observations that could adversely affect model performance.

Table 6. Diagnostic statistics of the regression model for total phenolic content (TPC), including residuals, leverage values, studentized residuals, Cook's distance, and DFFITS analysis.

Run Order	Actual Value	Predicted Value	Residual	Leverage	Internally Studentized Residuals	Externally Studentized Residuals	Cook's Distance	Influence on Fitted Value DFFITS	Standard Order
1	3.78	3.77	0.0098	0.750	2.113	3.249	1.339(1)	5.628(1)	10
2	3.67	3.67	-0.0072	0.750	-1.564	-1.795	0.734	-3.110(1)	9
3	3.83	3.83	0.0000	0.200	0.000	0.000	0.000	0.000	17
4	3.74	3.74	-0.0075	0.750	-1.626	-1.907	0.793	-3.304(1)	3
5	3.89	3.90	-0.0095	0.750	-2.051	-3.006	1.262(1)	-5.207(1)	4
6	3.82	3.81	0.0075	0.750	1.626	1.907	0.793	3.304(1)	2
7	3.65	3.64	0.0095	0.750	2.051	3.006	1.262(1)	5.207(1)	1
8	3.58	3.58	-0.0023	0.750	-0.487	-0.459	0.071	-0.795	5
9	3.74	3.74	-0.0003	0.750	-0.062	-0.057	0.001	-0.099	6
10	3.79	3.80	-0.0098	0.750	-2.113	-3.249	1.339(1)	-5.628(1)	11
11	3.87	3.86	0.0023	0.750	0.487	0.459	0.071	0.795	8
12	3.71	3.71	0.0003	0.750	0.062	0.057	0.001	0.099	7
13	3.83	3.83	0.0000	0.200	0.000	0.000	0.000	0.000	16
14	3.83	3.83	0.0000	0.200	0.000	0.000	0.000	0.000	15
15	3.90	3.89	0.0072	0.750	1.564	1.795	0.734	3.110(1)	12
16	3.83	3.83	0.0000	0.200	0.000	0.000	0.000	0.000	14
17	3.83	3.83	0.0000	0.200	0.000	0.000	0.000	0.000	13

The residual values obtained from the model were generally very small, ranging from −0.0098 to 0.0098. The small magnitude of the residuals indicates a strong agreement between the experimentally observed values and the predicted values generated by the regression model. Furthermore, the residuals were distributed around zero without showing extreme variation, suggesting that the model errors were relatively random and unbiased. Leverage values were also evaluated to determine the influence of each experimental point on the regression model. Most experimental runs exhibited leverage values of 0.750, whereas the replicated center points showed lower leverage values of 0.200. This behavior is expected in Box–Behnken experimental designs because center points are typically located near the geometric center of the design space and therefore exert lower influence on model fitting. The absence of excessively high leverage values indicates that no individual observation dominated the regression analysis.

The internally and externally studentized residuals were further examined to detect possible outliers. Most residual values were distributed within acceptable limits, although several runs exhibited relatively high studentized residual values. Specifically, runs 1, 5, 7, and 10 showed externally studentized residual values greater than ±3.0. However, these observations were not considered severe outliers because the residual magnitudes remained relatively small and no abnormal experimental trend was observed. In designed experimental studies, slightly elevated studentized residual values may occur due to model sensitivity near the boundary regions of the experimental design space. Cook's distance values were additionally analyzed to assess the influence of each observation on the estimated regression coefficients. Several experimental runs, including runs 1, 5, 7, and 10, showed relatively higher Cook's distance values compared with other runs. Nevertheless, the values remained within an acceptable range and did not indicate excessive influence capable of destabilizing the model. Similarly, DFFITS values identified several influential observations exceeding the recommended threshold limits. However, these observations corresponded to experimental points located near the edge regions of the Box–Behnken design space, where higher sensitivity is commonly observed.

Importantly, despite the presence of several influential observations, the overall diagnostic evaluation demonstrated that the developed regression model remained statistically stable and reliable. This conclusion is supported by the excellent agreement between predicted and experimental responses, the high coefficient of determination ($R^2 = 0.9953$), and the absence of extreme residual dispersion. Therefore, no experimental runs were removed from the dataset, and all observations were retained for subsequent optimization and response surface analyses. Overall, the diagnostic analysis confirmed that the developed quadratic model possessed good robustness, acceptable residual behavior, and satisfactory predictive reliability for describing and optimizing total phenolic content extraction from *Morus alba* var. *shalun*.

4. Conclusion

This study successfully optimized the extraction conditions for total phenolic content (TPC) from *Morus alba* var. *shalun* leaves using Response Surface Methodology (RSM) with a Box–Behnken Design (BBD). The statistical analysis demonstrated that the developed quadratic model was highly significant, with excellent predictive capability and model adequacy, as indicated by the high coefficient of determination ($R^2 = 0.9953$), low coefficient of variation (CV = 0.2444%), and high Adeq Precision value (44.4335). The close agreement between experimental and predicted values further confirmed the reliability and robustness of the developed model. Among the investigated extraction variables, ethanol concentration was identified as the most influential factor affecting TPC extraction, followed by extraction time and extraction temperature. The linear effects of all three variables were highly significant ($p < 0.0001$), while the interaction effects among variables were statistically insignificant. In addition, significant quadratic effects of ethanol concentration and extraction time indicated the existence of optimum extraction conditions for maximizing phenolic recovery.

The Box–Cox transformation analysis confirmed that natural logarithm transformation was appropriate for improving variance stability and model fitting. Diagnostic analyses, including residual analysis, leverage evaluation, Cook's distance, and DFFITS, demonstrated that the developed model exhibited acceptable statistical behavior without severe outliers or instability issues.

Overall, the developed RSM model proved to be an effective and reliable approach for predicting and optimizing total phenolic content extraction from *Morus alba* var. *shalun.* The findings of this study provide valuable information for

the development of efficient extraction processes and support the potential utilization of *Morus alba* var. *shalun* as a promising natural source of phenolic compounds for food, pharmaceutical, and nutraceutical applications.

References

[1] Nomura T. Phenolic compounds of the mulberry tree and related plants. Progress in the Chemistry of Organic Natural Products. 1988;53:87–201.

[2] Syah YM, Achmad SA, Ghisalberti EL, Hakim EH, Mujahidin D. Artoindonesianins X–Z, three new isoprenylated flavones from *Artocarpus lanceifolius*. Phytochemistry. 2000;54(7):715–717.

[3] Yang ZG, Matsuzaki K, Takamatsu S, Kitanaka S. Inhibitory effects of constituents from *Morus alba* var. *multicaulis* on differentiation of 3T3-L1 cells and nitric oxide production in RAW264.7 cells. Molecules. 2014;19(10):17109–17125.

[4] Fitriani, Nurlelasari, Mayanti T, Harneti D, Maharani R, Supratman U. Phenolic compounds from *Morus alba* var. *shalun* and their antioxidant activity. Natural Product Sciences. 2019;25(2):158–164.

[5] Dai SJ, Li WX, Wang GF, et al. Cytotoxic prenylated flavonoids from the root bark of *Morus alba*. Journal of Asian Natural Products Research. 2004;6(3):175–182.

[6] Naik R, Harmalkar DS, Xu X, Jang K, Lee K. Bioactive benzofuran derivatives: Moracins A–Z in medicinal chemistry. European Journal of Medicinal Chemistry. 2015;90:379–393.

[7] Qin J, He NJ, Wang ZY, et al. New bioactive compounds from the leaves of *Morus alba*. Chemistry & Biodiversity. 2015;12(9):1400–1412.

[8] Gong C, Tong Y, Zhang H, Wang L. Antioxidant and antimicrobial activities of phenolic compounds isolated from *Morus alba*. Food Chemistry. 2019;277:46–53.

[9] Butt MS, Nazir A, Sultan MT, Schroën K. *Morus alba* L. nature's functional tonic. Trends in Food Science & Technology. 2011;22(9):505–512.

[10] Katsube T, Imawaka N, Kawano Y, Yamazaki Y, Shiwaku K, Yamane Y. Antioxidant flavonol glycosides in mulberry (*Morus alba* L.) leaves isolated based on LDL antioxidant activity. Food Chemistry. 2006;97(1):25–31.

[11] Naczk M, Shahidi F. Extraction and analysis of phenolics in food. Journal of Chromatography A. 2004;1054(1–2):95–111.

[12] Myers RH, Montgomery DC, Anderson-Cook CM. Response Surface Methodology: Process and Product Optimization Using Designed Experiments. 4th ed. New York: Wiley; 2016.

[13] Box GEP, Behnken DW. Some new three-level designs for the study of quantitative variables. Technometrics. 1960;2(4):455–475.

[14] Montgomery DC. Design and Analysis of Experiments. 9th ed. New York: Wiley; 2017.