

Article

A Study of the Response Surface Methodology Model with Regression Analysis in Three Fields of Engineering

Hsuan-Yu Chen ¹ and Chiachung Chen ^{2,*} 

¹ Africa Industrial Research Center, National Chung Hsing University, Taichung 40227, Taiwan; wakaharu37@gmail.com

² Department of Bio-Industrial Mechatronics Engineering, National Chung Hsing University, Taichung 40227, Taiwan

* Correspondence: ccchen@dragon.nchu.edu.tw; Tel.: +886-4-22857562

Abstract

Researchers conduct experiments to discover factors influencing the experimental subjects, so the experimental design is essential. The response surface methodology (RSM) is a special experimental design used to evaluate factors significantly affecting a process and determine the optimal conditions for different factors. The relationship between response values and influencing factors is mainly established using regression analysis techniques. These equations are then used to generate contour and surface response plots to provide researchers with further insights. The impact of regression techniques on response surface methodology (RSM) model building has not been studied in detail. This study uses complete regression techniques to analyze sixteen datasets from the literature on semiconductor manufacturing, steel materials, and nanomaterials. Whether each variable significantly affected the response value was assessed using backward elimination and a *t*-test. The complete regression techniques used in this study included considering the significant influencing variables of the model, testing for normality and constant variance, using predictive performance criteria, and examining influential data points. The results of this study revealed some problems with model building in RSM studies in the literature from three engineering fields, including the direct use of complete equations without statistical testing, deletion of variables with *p*-values above a preset value without further examination, existence of non-normality and non-constant variance conditions of the dataset without testing, and presence of some influential data points without examination. Researchers should strengthen training in regression techniques to enhance the RSM model-building process.

Keywords: response surface methodology; regression analysis; backward elimination; semiconductor manufacturing; steel materials; nanomaterials



Academic Editor: Christos Douligeris

Received: 16 May 2025

Revised: 6 July 2025

Accepted: 16 July 2025

Published: 21 July 2025

Citation: Chen, H.-Y.; Chen, C. A Study of the Response Surface Methodology Model with Regression Analysis in Three Fields of Engineering. *Appl. Syst. Innov.* **2025**, *8*, 99. <https://doi.org/10.3390/asi8040099>

Copyright: © 2025 by the authors. Published by MDPI on behalf of the International Institute of Knowledge Innovation and Invention. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The results or responses of an engineering system or its process are often affected by many factors. Finding the optimal level of these factors is called optimization. In research evaluating the impact of multiple factors on one or more response values, the most common method to evaluate the most appropriate conditions of multiple factors is the response surface methodology (RSM) [1,2]. RSM is a special experimental design that can be used to establish process operations for new products, improve the utilization rate of operating processes, and reduce resource waste. Through the use of appropriate statistical techniques,

it is possible to evaluate the factors that have a significant impact on the process and thus determine the optimal conditions for these factors [1,3].

The special feature of RSM in experimental design is that this method can test multiple factors with a small number of samples. After completing the data collection, regression analysis is used to establish the mathematical relationship between the system response value and its influencing factors. These models are then used to create graphs that illustrate the impact of various factors on the response. Researchers can use these linear or curvilinear distribution graphs to observe this relationship. Due to this unique graphical observation capability, the response surface methodology has been widely used in industry [1–4].

Two excellent textbooks introduce the RSM concept, two- and three-level factorial design, experimental design for fitting the response surface, and some special topics of RSM [1,2]. Bas and Boyaci [4] proposed an essential concept about RSM. Asoo et al. [3] introduced the historical background of RSM. These review papers related to engineering applications involved processing in machinability [5], manufacturing [6], biofuel [7,8], energy [9,10], agro-industrial processes [11], and polymers [12].

In an experimental design, an experimental system includes the response of this system and is combined with different levels of the influencing factor. These factors serve as the independent variables for the regression analysis. Experimental runs represent a series of tests for the experiment. The response or output of the experiments is used as the dependent variable for further analysis.

The relationship between the response (y) and the input factors (x_1, x_2, \dots, x_k) of the RSM can be expressed as

$$y_i = f(x_1, x_2, \dots, x_k) + \varepsilon_i \quad (1)$$

where y is the response of the RSM model and is called the dependent variable; x_1, x_2, \dots , and x_k are the influencing variables and are called independent variables in regression; k is the number of factors; and ε_i is the model's error.

The mathematical model of the RSM model includes linear, interaction, and quadratic terms effects. If the process system involves three process variables, x_1, x_2 , and x_3 , the form of the RSM equation is

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + \varepsilon \quad (2)$$

If four factors are used, x_1, x_2, x_3 , and x_4 , the RSM equation is

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + b_{44}x_4^2 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4 + \varepsilon \quad (3)$$

If the RSM equation includes all variables, as in Equations (2) and (3), it is called a full RSM equation.

The relationship between the response variable, the independent variable, and the dependent variable can be established through regression analysis techniques. This regression model is an empirical equation, not a theoretical equation.

The major experimental methodologies used for the response surface methodology are the full factorial design (CFD), Box–Behnken design (BBD), and central composite design (CCD) [1,6,13–16]. The number of studies with each experimental methodology, as determined using Google Scholar, is listed in Table 1.

Table 1. The number of studies using each experimental methodology, determined using Google Scholar.

Year	Full Factorial Design (CFD)	Box–Behnken Design (BBD)	Central Composite Design (CCD)
2019	145,000	7250	218,000
2020	151,000	8450	235,000
2021	147,000	9860	224,000
2022	132,000	12,400	207,000
2023	114,000	13,600	159,000
2024	88,300	16,500	113,000

Table 1 presents the number of papers employing the three statistical experimental techniques from 2019 to 2024, as determined by a Google Scholar search. It can be seen that the number of papers using central composite design is the largest, followed by full factorial design, while Box–Behnken design is the least frequent. The number of papers using the central composite design is about 10 times that of paper using Box–Behnken design. However, upon examining the increasing trend, it can be seen that the number of papers using central composite design and full factorial design has been decreasing year by year. The number of papers using Box–Behnken design is increasing year by year.

The experimental matrices in the three methods for RSM design can be found in various publications and textbooks [1,2]. The number of required experiments for the experimental methods is 27, 22, and 13 or more for FFD, BBD, and CCD, respectively.

Many criteria have been proposed to evaluate the regression models' fitting agreement and predictive ability. The essential procedure in regression analysis is to find the significant effect of certain variables on the response. To evaluate the fitting agreement of a proposal model, the criteria include the F-value of ANOVA, coefficient of determination R^2 , adjusted R^2 (R^2_{adj}), lack-of-fit test, etc. To assess the significant effect of one variable on the response, the criterion is the t -value or p -value of the coefficients of this variable. The criteria to evaluate the predictive ability are the predicted residual error sum of squares (PRESS) and predicted R-squared (R^2_{pred}) [17–19].

After the evaluation, the appropriate equation (RSM model) establishes the relationship between the response value and the influencing factors as a mathematical function. Among the graphical methods, contour plots and three-dimensional surface response plots are particularly popular among researchers [1,3,20]. The studied techniques are described in detail in this paper [21].

The accuracy and predictive power of the RSM model determine whether it can effectively find the factors that actually affect the response value. However, engineers often do not care about the suitability of the regression equations they use. Many researchers directly use commercial software to calculate the RSM model. Reza et al. [22] emphasize the importance of the model's validity and accuracy in the context of RSM application, as well as the validity of the RSM model itself [22].

Compared with other experimental design methods, the most prominent feature of the response surface method is that it graphically represents the relationship between the influencing factors (variables) and the response value. Contour maps are used to present two-dimensional space. These contour maps illustrate the changes in response value under various influencing factors. Three-dimensional surface response maps represent information in three-dimensional space, with the response value as the z -coordinate and the x - and y -coordinates as the two influencing factors. Suppose the quadratic polynomial relationship does not exist. In that case, this type of contour plot will be drawn as a straight line, and the three-dimensional surface response plot will appear as a stable plane. If the

quadratic polynomial relationship is significant, both graphs will be represented by curves. Researchers often use these contour maps to intuitively observe the optimal level of the influencing factors that lead to the optimal response [1,3,4].

The purpose of performing a response surface methodology (RSM) experiment is to find the optimal operating conditions. Suppose the RSM model represents a linear relationship. In that case, the contour plot and the surface response plot indicate the direction of change in the response value concerning the original experimental design conditions. Assuming that the RSM model is a quadratic equation, the two plots will indicate the maximum, minimum, or saddle point conditions, respectively. Therefore, establishing a suitable RSM equation can ensure that the research results lead to the correct optimization conditions. Therefore, the regression technique for establishing the RSM needs careful consideration.

To the best of the authors' knowledge, regression techniques have not been fully considered when evaluating the accuracy and predictive ability of RSM equations. This study compiled 16 datasets from three engineering fields, drawn from the literature and supplemented with experimental datasets. These datasets were used to evaluate the adequacy of the RSM equations. The complete regression technique was illustrated. The adequate RSM models were proposed to highlight the importance of comprehensive regression analysis. The effect of the RSM equations on optimization was further discussed. Some suggestions were proposed for the use of RSM in engineering fields.

2. Materials and Methods

2.1. Data Sources for the Equations of the Response Surface Methodology

Table 2 presents 16 datasets from three engineering fields, used to evaluate the adequacy of RSM equations. These published papers present the original datasets.

The databases included ScienceDirect, Scopus, IEEE Xplore, SpringerLink, Google Scholar, and J-STAGE. Keywords were "RSM", "Regression analysis", "Model evaluation", and "Semiconductor", or "Steel materials", or "Nanomaterials".

The potential publication bias in the selected literature was a limitation of the data available. Some studies did not list their experimental data and presented their RSM results directly. Much helpful information in other non-selected literature was not considered. The publications years of 16 datasets from three engineering fields ranged from 2004 to 2024. The countries of the authors involved are Algeria, Brazil, China, Korea, Italy, Iran, Kazakhstan, Malaysia, Turkey, and the UK.

The parameters and criteria were estimated using Sigma Plot v.14.0 (SPSS Inc., Chicago, IL, USA).

2.2. Model Building for the Response Surface Methodology

A typical multiple regression model is expressed as

$$y_i = b_0 + b_1x_1 + \dots + b_kx_k + \varepsilon_i \quad (4)$$

where ε_i is a model error

The parameters of the regression equation are x_1, x_2, \dots, x_k , and the coefficients of the parameters are $b_0, b_1, b_2, \dots, b_k$. The parameters' coefficients are usually calculated using the least squares method. Many commercial software programs can help researchers perform this calculation easily and quickly.

Table 2. Published data in the literature on engineering for evaluating the adequate equation of the response surface methodology.

Study	Targets	Number of Data Points and Experimental Design	Software	Model Evaluation	Criteria for Parameter Selection	Reported Model	Optimization
I. Semiconductor							
Won et al. [23] y_{SR} = surface roughness x_1 = applied pressure x_2 = platen speed x_3 = slurry ratio	Si-wafer polishing	15 CCD	Not reported	Not reported	Not reported	Not reported	Contour plot, response surface plot
Lee et al. [24] y_{SR} = surface roughness x_1 = pressure x_2 = wheel speed x_3 = time	Final wafer polishing	15 BBD	MINITAB	R^2	Not reported	Full models	Contour plot, response surface plot
Zhang et al. [25] y_{CT} = CT y_{WTP} = WTP y_{TP} = TP x_1 = Ub x_2 = C ₁ x_3 = C ₂	Wafer fabricating	20 CCD	Design Expert, version not mentioned	ANOVA lack of fit, PRESS, R^2 , R^2 -adj	Not reported	Full models	Contour plot, response surface plot
Seo et al. [26] y_w (WAPR) y_{oxide} (Oxide MRR) x_1 = Free concentration x_2 = H ₂ O ₂ x_3 = SiO ₂		15 CCD	MINITAB	ANOVA R^2 , R^2 -adj	<i>t</i> -value <i>p</i> -value	Full models	Contour plot, response surface plot
Saleem and Soma [27] y_{PV} = pull-in voltage x_1 = TEL x_2 = TEN x_3 = TSL x_4 = TSW	MEMS	27 BBD	Not reported	R^2 , R^2 -adj	F-value <i>p</i> -value	Full models	Contour plot, response surface plot

Table 2. Cont.

Study	Targets	Number of Data Points and Experimental Design	Software	Model Evaluation	Criteria for Parameter Selection	Reported Model	Optimization
II. Steel materials							
Noordin et al. [28] y_{Ra} = surface roughness y_{Rc} = tangential force x_1 = cutting speed x_2 = SCEA	Coated carbide tools AISI 1045 steel	16 CCD	Design Expert Ver. 6.0	ANOVA R^2 , R^2_{adj} Lock of First, PRESS	p -value backward elimination	$y_{Ra} = f(x_1, x_2, x_3, x_2x_3, x_3^2)$ $y_{Rc} = f(x_1, x_2, x_3, x_2x_3, x_3^2)$	Contour plot, response surface plot
Bouacha et al. [29] y_{Ra} = Ra y_{Rt} = Rt y_{Rz} = Rz x_1 = VC x_2 = f x_3 = ap	Surface roughness, cutting forces AISI 52100 steel	27 Taguchi orthogonal array	Not reported	ANOVA R^2 R^2_{adj}	F-value p -value At once delete variance	$y_{Ra} = f(x_1, x_2, x_3, x_1x_2)$ $y_{Rt} = f(x_1, x_2, x_1x_2, x_2^2)$ $y_{Rz} = f(x_1, x_2, x_1x_2, x_2^2)$	Contour plot, response surface plot
Elbah et al. [30] y_{Fa} = Fa y_{Fr} = Fr y_{Ft} = Ft y_{Ra} = Ra x_1 = depth of cut x_2 = feed rate x_3 = cutting speed	Mixed ceramic tool, AISI 4140 steel	27 CCD	Design Expert 8.0.7	ANOVA R^2 R^2_{adj}	F-value Some variables are not significant in ANOVA	$y_{Fa} \sim y_{Ra}$ full model	Contour plot, response surface plot
Campos et al. [31] y_{Time} = Time y_{Ra} = Ra y_{Rt} = Rt x_1 = VC x_2 = f x_3 = ap	Machining of hardened steel	19 CCD	Design Expert, version not mentioned	ANOVA R^2 R^2_{adj}	Sequential model for some terms	$y_{Time} \sim y_{Rt}$ full model	Contour plot, response surface plot
Khalil et al. [32] y_{RS} = surface roughness x_1 = cutting spend x_2 = feed rate x_3 = depth of cut	Surface roughness AISI D3 steel	20 CCD	Design Expert, version not mentioned	Lack of fit	F-value p -value (cutoffs $p < 0.1$)	Full model	Contour plot, response surface plot

Table 2. Cont.

Study	Targets	Number of Data Points and Experimental Design	Software	Model Evaluation	Criteria for Parameter Selection	Reported Model	Optimization
III. Nanomaterials							
Pajae and Taghizadeh [33]							
$y_{Ethylene}$ = ethylene					F-value		
$y_{Propylene}$ = propylene	Yield	15 BBD	Design Expert ver. 6	R^2 R^2_{adj}	p -value ANOVA Table.	Full model	Contour plot, response surface plot
x_1 = MW aging time							
x_2 = US aging time							
x_3 = HT time							
Jourshabani et al. [34]							
$y_{phenol\ yield}$ = phenol yield	Benzene hydroxylation	20 CCD	Design Expert Ver. 7.1.3	R^2 R^2_{adj}	F-value p -value	Full model	Contour plot, response surface plot
x_1 = temperature							
x_2 = H ₂ O ₂ content							
x_3 = Catalyst							
Sheng et al. [35]							
y_{TC002} = TC002						Not reported	Contour plot, response surface plot
$y_{Aspect\ ratio}$ = Aspect ratio						Not reported	Contour plot, response surface plot
$y_D = D$		27 CCD	MOODE Ver. 10	ANOVA lack of fit	Not reported	$\log(Y_D + 0.5)$ of $f(x_1, x_2, x_3, x_4, x_1x_2, x_2x_3, x_2x_4, x_3x_4, x_3^2)$	Contour plot, response surface plot
x_1 = concentration							
x_2 = temperature							
x_3 = catalyst							
Rakhmanova et al. [36]							
$y_{Zinc\ Oxide}$ = zinc oxide synthesis						Not reported	Contour plot, response surface plot
x_1 = applied potential		15 BBD	Design Expert ver. 8.0.7.1	R^2 , R^2_{adj} Lack of fit, PRESS	Not reported	Not reported	Contour plot, response surface plot
x_2 = distance							
x_3 = temperature							

Table 2. Cont.

Study	Targets	Number of Data Points and Experimental Design	Software	Model Evaluation	Criteria for Parameter Selection	Reported Model	Optimization
Pakolpakcil et al. [37] y_{Aft} = Average fiber diameter x_1 = concentration x_2 = Rotational speed x_3 = Needle size	Poly nanofiber mats	15 BBD	Design Expert ver. 13	ANOVA sequential model, quadratic, and interaction	F-value p -value	Full model	Contour plot, response surface plot
Sreekumar et al. [38] y_{Nth} = Nth y_{nele} = nele $y_{nex,th}$ = nex,th $y_{nex,ele}$ = nex,ele x_1 = $\varphi\%$ x_2 = m x_3 = I x_4 = Π	Nanofluid-based PV/T	27 CCD	Design Expert, version not mentioned	ANOVA R^2 , R^2_{adj}	F-value p -value	Full model	Contour plot, Response surface plot

To establish the relationship between the variables and response of the RSM models, the significant effect of the variables on the response must be evaluated using statistical techniques. However, due to the convenience of using the commercial software programs, the evaluation of regression analysis is easily neglected.

2.3. Basic and Complete Regression Analyses

Regression analysis is a statistical technique used in experimental design to establish the relationship between response values and influencing factors. With the help of convenient commercial software programs, the estimated values of variable parameters can be easily calculated using the least squares method. These software programs also provide statistical tests to validate the regression models, such as analysis of variance tables. An analysis of variance table lists the t -value and p -value for each variable, facilitating further analysis. This process is called basic regression analysis.

Regression analysis does not simply accept all variables in the initial calculation equation. The next step is to screen out those variables that have no significant effect on the response. Due to the interaction and multicollinearity between these variables, it is impossible to remove the variables that have no significant effect on the response at the same time [17,18,39–43].

Several procedures are employed, including the sequential variable selection procedure, the forward procedure, stepwise regression, and backward elimination. In the establishment of the RSM models, the backward elimination procedure is recommended. For the regression analysis, two criteria are used to evaluate the accuracy (fitting agreement) and prediction [17,18,39–42].

Besides calculating the estimated values of the parameters and screening the effective variables, the detailed technique includes checking for violations of the assumptions and performing influence diagnostics. The influential data points are checked with other criteria [17,18,39–43].

In this study, the complete regression analysis includes performing a sequential variable selection procedure, evaluating RSM model accuracy and prediction using different criteria, checking for violations of assumptions, and performing influence diagnostics to identify influential data points.

2.4. Assumptions Involved in the Regression Analysis

The regression model includes the following assumptions: ε_i is uncorrelated across observations, the mean of ε_i is zero, the variance of ε_i is constant, and the distribution of ε_i follows a normal distribution. Based on this assumption, non-standard conditions involved in regression analysis include the balance between underfitting and overfitting, the non-normality of the dataset, the heterogeneity of the variance, and influential data points [17].

After completing the regression calculation using the least squares method, the first concern is to verify the significant effect of each parameter. Then, these assumptions must be verified, and any non-standard conditions must be addressed if they exist. Another concern of regression analysis is distinguishing between the model's performance in terms of model fitting and predictive ability. Two criteria are used [17,44–46]. Modern concepts in regression modeling have been introduced by Myers [17] and Marinou [46]. Checking the significant effect of each parameter and selecting a regression model is called classical regression. The concept of complete regression involves a trade-off between bias and variance when selecting a limited number of important variables, comparing predictive ability, and verifying and addressing the assumptions of regression.

2.5. Establishment of the Model

Once the regression equation is established, contour and 3D surface plots can be generated using commercial software programs, allowing researchers to easily visualize the optimization using these figures. Suppose the selection of the regression equation is inappropriate. In that case, these contour and 3D surface plots are inappropriate, so the conclusions and suggestions regarding the effect of the dependent variable (factors and levels) are meaningless.

Three methods for performing sequential variable selection are forward selection, stepwise regression, and backward elimination. These methods are illustrated in detail [17, 18,39–43,47].

For the forward selection procedure, all variables are selected as regressors and enter the model with a constant term. The variable that produces the largest R-squared value is selected first, and the resulting equation is referred to as the first equation. The other variables are considered as the second variable of the first equation. The variable that produces the largest R-squared value in the second analysis is selected as the second variable. The selection of variables continues in this way. For each procedure, the p -value of the selected variable is compared with the preset value. If the p -value of the selected variable is greater than the preset value, the procedure ends.

Stepwise regression is an improvement on forward selection. Variables that were deleted in the previous stage can be re-entered in the selection procedure. The selection procedure is the same as forward regression.

Backward elimination procedures are used to fit all variables for the regression equation and determine the p -value for each variable in the model. The variable with the lowest observed t -value and its corresponding p -value are compared with a preselected significance level, usually $p < 0.05$. If its p -value exceeds the preselected value, the variable is removed. The remaining variables are recalculated, and the variable with the lowest t -value and p -value is identified to compare its p -value with a value of $p < 0.05$. The above backward elimination procedures are repeated until no variable is dropped, and the procedure ends. The selection of the regression model consists of all remaining variables.

The problem with forward and stepwise methods is that the critical t -values they set are not appropriate in the early stages [17]. Since there are fewer variables in the early stages, the standard values of the estimates are usually overestimated, and the p -values of the variables may be too significant, thus preventing important variables from entering the model. Therefore, forward and stepwise models often underestimate. Mendenhall and Sincich [42] recommend the use of backward elimination because this method can select all possible explanatory variables as early as possible and eliminate those that are not important in explaining the response variation. Backward elimination is recommended because it considers the effects of all candidate variables [41,43,48].

In the backward elimination procedure, a t -test statistic for a variable is used to calculate its p -value, which is then used to assess the statistical interpretation of the variable in the regression model [14,48].

2.6. Criteria for the Evaluation of RSM Equations

Ten criteria are used to evaluate the RSM equations after calculating the coefficients of the variables. They are listed in Table 3.

Table 3. Criteria for the evaluation of the RSM equations.

Criterion	Description	Cutoffs
R^2	The coefficient of determination is used to determine the relationship between the response and the independent variable.	R^2 value near 1.0
Adjusted R^2	This value takes into account the impact of the number of independent variables on the R-squared value. The closer the adjusted R-squared (R^2_{adj}) is to 1.0, the better the descriptive ability of the regression equation.	R^2_{adj} value closer to 1.0
s	This represents the actual variability in the equation regarding the data distribution between the response and independent variables. It indicates the precision of those estimates. It reflects the variability in the estimates across different random samples from the same population.	A smaller s suggests a more precise estimate, meaning the estimated coefficient is likely closer to the actual population value.
t-value	The t-value is used to test the null hypothesis that the coefficients of the independent variables are significantly different from zero.	A large t-value for the independent variable indicates that the coefficient is statistically significant and not equal to zero.
p-value	The variable coefficient is calculated from its t-value and used to test the null hypothesis that the coefficients of the independent variables are significantly different from zero.	The p-value represents the probability of incorrectly determining whether the coefficient of the variable is not zero. A smaller p-value represents a greater probability of the validity of the variable.
PRESS, predicted residual error sum of squares	This evaluates the predictive ability of the regression model	The smaller, the better
Normality test	The normality test is used to evaluate whether the dataset is normally distributed. The normality test technique used in this study is the Kolmogorov–Smirnov method.	The p-value calculated with this method compares the preset value ($p = 0.05$).
Constant variance test	This assesses whether the dependent variable (response) has constant variance across its overall sources. The testing technique used in this study is the Spearman rank correlation method.	The cutoff value is $p = 0.05$.
t_i , externally studentized residuals	This computes the standard error of the residual of the estimated value, and this data is not used in model building.	Values of ± 2.0 are usually used to indicate the possibility of an outlier.
DFFITS _i	This evaluates the prediction effect for a data point. It is used to compare the estimated standard errors when the observed value is removed.	The cutoffs of DFFITS _i are ± 2.0 .

R^2 , the coefficient of determination, is affected by the number of variables, so it is not a reasonable criterion. The criteria of R^2_{adj} and s are considered measures of the effect of the number of variables; they serve as criteria for model accuracy. The PRESS value is used to compare the predictive performance [17,18,39–42].

The normality test technique employed is the Kolmogorov–Smirnov method, with a cutoff value of $p = 0.05$. The constant variance test uses the Spearman rank correlation method; the cutoff value is $p = 0.05$.

The externally studentized residuals, t_i , and DFFITS_{*i*} values are used to examine potentially influential data points. Both criteria are set at ± 2.0 . If a data point is identified as influential, it should not be removed from the dataset immediately. It may be due to experimental or instrumental error, or it may deviate from the trend predicted by the proposed model. Further observations should be made, and more relevant data should be collected under the same experimental conditions.

SPSS (Statistical Package for the Social Sciences) version V.14.0 (SPSS Inc., Chicago, IL, USA). The contour plots were produced by this software.

3. Results

3.1. Semiconductor Manufacturing

An experimental dataset was reported by Box and Draper [49] and was introduced by Myers et al. [1]. The process involved applying a coating material to a wafer. Several coating thicknesses at different locations on a wafer were measured. The mean y_1 and standard deviation y_2 of the thickness were calculated. The influencing variables were x_1 , speed; x_2 , pressure; and x_3 , distance. These datasets illustrate the backward elimination technique for an adequate equation and are listed in Appendix A.1.

For the y_1 mean of thickness [49], the complete regression procedure is

$$\begin{aligned}
 1. \quad y_1 = & 327.615 + 177.011x_1 + 109.422x_2 + 131.472x_3 + 32.022x_1^2 - 22.378x_2^2 \\
 & (<0.001) (<0.001) (<0.001) (0.317) (0.481) \\
 & - 29.061x_3^3 + 66.033x_1x_2 + 75.458x_1x_3 + 43.583x_2x_3 \\
 & (0.363) (0.008) (0.003) (0.064) \\
 & R^2 = 0.927, R^2_{\text{adj}} = 0.888, s = 76.111, \text{PRESS} = 337,737.94
 \end{aligned} \tag{5}$$

Delete x_2^2 and recalculate the equation.

$$\begin{aligned}
 2. \quad y_1 = & 312.696 + 177.011x_1 + 109.422x_2 + 131.475x_3 + 32.022x_1^2 - 29.061x_3^2 \\
 & (<0.001) (<0.001) (<0.001) (0.310) (0.356) \\
 & + 66.033x_1x_2 + 75.458x_1x_3 + 45.583x_2x_3 \\
 & (0.007) (0.003) (0.006)
 \end{aligned} \tag{6}$$

Delete x_3^2 and recalculate the equation.

$$\begin{aligned}
 3. \quad y_1 = & 293.322 + 177.011x_1 + 109.422x_2 + 131.475x_3 + 32.022x_1^2 - 66.033x_3^2 \\
 & (<0.001) (<0.001) (<0.001) (0.308) (0.007) \\
 & + 75.458x_1x_3 + 43.458x_2x_3 \\
 & (0.002) (0.058)
 \end{aligned} \tag{7}$$

Delete x_1^2 and recalculate the equation.

$$\begin{aligned}
 4. \quad y_1 = & 314.670 + 177.011x_1 + 109.422x_2 + 131.473x_3 + 66.033x_1x_2 + 75.458x_1x_3 \\
 & (<0.001) (<0.001) (<0.001) (0.006) (0.002) \\
 & + 47.583x_2x_3 \\
 & (0.0058) \\
 & R^2 = 0.916, R^2_{\text{adj}} = 0.891, s = 75.068, \text{PRESS} = 288,457.9
 \end{aligned} \tag{8}$$

The p -values of all variables are <0.05 ; this is an adequate equation.

The normality and constant variance tests are passed. Influential points include the 9th ($t_i = -3.316$, DFFITS_{*i*} = -2.858), 19th ($t_i = 2.255$, DFFITS_{*i*} = 2.056), and 25th ($t_i = -2.322$).

The adequate equation involves x_1 , x_2 , x_3 , x_1x_3 , and x_2x_3 . No quadratic terms exist in this equation. The contour plots of the response form a plateau, not a surface curve.

The adequate equation for the y_2 standard deviation and three variables is calculated as follows:

$$\begin{aligned}
 y_2 = & 34.904 + 11.522x_1 + 15.317x_2 + 29.183x_3 + 4.189x_1^2 - 1.328x_2^2 \\
 & (0.280) (0.156) (0.012) (0.818) (0.942) \\
 & -16.772x_3^3 + 7.717x_1x_2 + 5.117x_1x_3 + 14.075x_2x_3 \\
 & (0.362) (0.550) (0.691) (0.281) \\
 R^2 = & 0.454, R^2_{adj} = 0.615, s = 43.817, PRESS = 93,044.256
 \end{aligned}
 \tag{9}$$

After the execution of the backward elimination, the adequate equation for y_2 is

$$\begin{aligned}
 y_2 = & 47.993 + 29.183x_3 \\
 & (0.007) \\
 R^2 = & 0.256, R^2_{adj} = 0.227, s = 42.171, PRESS = 52,441.283
 \end{aligned}
 \tag{10}$$

The normality test is passed. The constant variance test is failed ($p < 0.001$).

The results of the evaluation of adequate RSM models for semiconductor manufacturing in five studies are listed in Table 4.

Table 4. Results of the evaluation of adequate RSM equations for semiconductor manufacturing.

Source	Reported Equations	Contour and Surface Response Plot	Adequate Equations	Normality Test	Constant Variance Test	Influential Data
Won et al. [23]	Not reported	Curve surface	$y_{SR} = f(x_1, x_3, x_1x_3, x_1^2, x_3^2)$	Passed	Passed	2, 15
Lee et al. [24]	Full models	Curve surface	$y_{SR} = f(x_1, x_2, x_3, x_1x_3, x_2^2, x_3^2)$	Passed	Passed	4, 6, 7, 12
Zhang et al. [25]	Full models	Curve surface	$y_{CT} = f(x_1, x_2, x_1^2, x_2^2)$ $y_{WTD} = f(x_1, x_2, x_3, x_1x_2, x_2^2, x_3^2)$ $y_{TP} = f(x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2)$	Passed Failed Failed	Passed Passed Failed	5, 13, 17 13 5, 14, 18
Seo et al. [26]	$y_W =$ full model $y_{Oxide} =$ full model	Curve surface	$y_W = f(x_2)$ $y_{Oxide} = f(x_1, x_3)$	Passed Passed	Passed Passed	12, 14 no
Saleem and Soma [27]	Full model	Curve surface	$y_{PV} = f(x_1, x_2, x_3, x_4, x_1x_3, x_3^2)$	Failed	Passed	27

Won et al. [23] employed the response surface method to optimize the final polishing of Si wafers. The experimental design was CCD, the response y_{SR} was surface roughness, and the variables were x_1 applied pressure, x_2 platen speed, and x_3 mixed slurry ratio. The RSM model was not reported in this study. Contour plot and response surface plot curves were produced by full models [23].

The full equation with these datasets is as follows:

$$\begin{aligned}
 y_{SR} = & 1.988 + 0.501x_1 - 0.0440x_2 - 0.172x_3 + 0.235x_1^2 + 0.0101x_2^2 \\
 & (<0.001) (0.464) (0.027) (0.084) (0.931) \\
 & + 0.223x_3^2 + 0.0711x_1x_2 - 0.128x_1x_2 + 0.00125x_2x_3 \\
 & (0.098) (0.302) (0.095) (0.985) \\
 R^2 = & 0.958, R^2_{adj} = 0.882, s = 0.172, PRESS = 1.299
 \end{aligned}
 \tag{11}$$

The adequate equation was evaluated with the regression technique of backward elimination:

$$y_{SR} = 1.991 + 0.501x_1 - 0.172x_3 + 0.238x_1^2 + 0.225x_3^2 - 0.128x_1x_2$$

(<0.001) (0.007) (0.030) (0.037) (0.044)

$$R^2 = 0.941, R^2_{adj} = 0.909, s = 0.154, PRESS = 0.646$$
(12)

The normality and constant variance tests are passed. Two influential data points are found, the 7th and 15th data points.

In this study [23], the authors used the full equation to produce the curved surface plots. However, the adequate equation indicated that the x_2 variable (platen speed) did not significantly affect the response, surface roughness. With the complete regression technique, the adequate RSM equation could help researchers to find the optimal condition.

Figure 1 indicates the contour plots for the complete and adequate equations. The difference between the two equations resulted in a difference in the distribution of curves between the two figures. The contour and response surface plots produced with the full equation were presented in the study [23]. The incorrect RSM equation could induce incorrect results of observation.

Lee et al. [24] investigated the polishing factors affecting surface roughness using a Box–Behnken design. The polishing factors were x_1 , pressure; x_2 , wheel speed; and x_3 , process time. The authors proposed the full equation as the best equation and found that the R^2 was 0.974 for this equation. This full equation produced contour and response surface plots.

The full equation proposed by the authors is

$$y_{SR} = 9.465 - 24.850x_1 - 0.163x_2 - 0.278x_3 + 46.375x_1^2 + 0.00309x_2^2$$

(<0.001) (0.006) (0.011) (0.001) (0.008)

$$+ 0.00815x_3^2 - 0.0325x_1x_2 + 0.425x_1x_3 + 0.000150x_2x_3$$

(0.038) (0.661) (0.029) (0.919)

$$R^2 = 0.974, R^2_{adj} = 0.928, s = 0.140, PRESS = 1.559$$
(13)

However, the p -values of the variables x_1x_2 and x_1x_3 are greater than 0.005.

The adequate equation evaluated with the complete regression technique in this study is

$$y_{SR} = 9.565 - 25.501x_1 - 0.168x_2 - 0.275x_3 + 46.375x_1^2 + 0.00309x_2^2$$

(<0.001) (<0.001) (0.002) (<0.001) (0.002)

$$+ 0.00815x_3^2 + 0.425x_1x_3$$

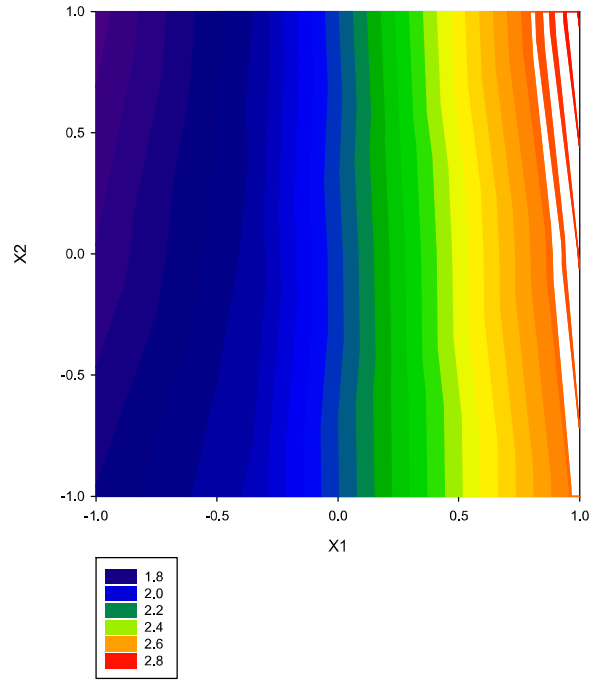
(0.014) (0.010)

$$R^2 = 0.973, R^2_{adj} = 0.946, s = 0.121, PRESS = 0.603$$
(14)

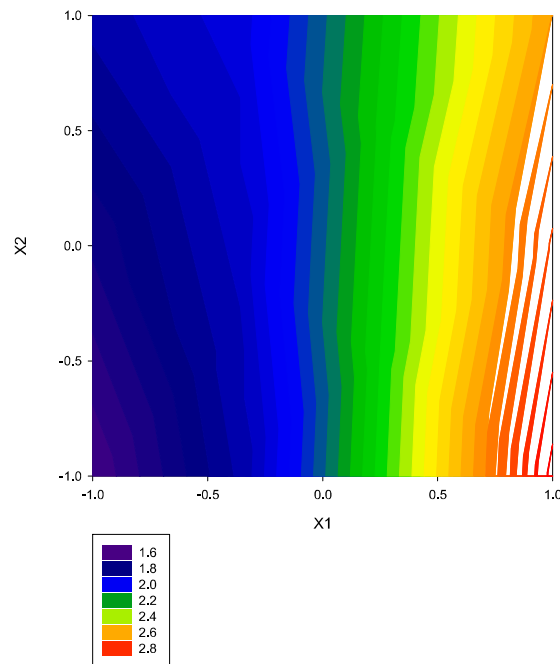
The normality and constant variance tests are passed. Four influential data points are found (4th, 6th, 7th, 12th).

Comparing the full and adequate equations, the adequacy equation has higher values of R^2_{adj} and s and a lower value of PRESS. This indicated that an adequate equation has better accuracy performance (R^2_{adj} , s) and prediction (PRESS).

Zhang et al. [25] investigated the optimization of dispatching rules for wafer manufacturing systems. The affecting variables were x_1 , the criterion of bottleneck, and x_2 and x_3 , which were two coefficients of work-in-progress (WIP) status. The responses included y_{CT} , the cycle time (CT); y_{WIP} , work-in-progress (WIP); and y_{TP} , throughput (TP). The authors proposed the full models [25], and contour and 3-D response surface plots were produced as curve surface plots.



(a)



(b)

Figure 1. Contour plots of the complete and adequate equations for the final polished surface roughness response of a silicon wafer. The contours present the effect of using the RSM equation on the relationship between the response variables and the influencing variables. (a) Full equation. (b) Adequate equation.

The results of the complete regression in our study are

$$y_{CT} = 932.466 - 10.392x_1 - 31.055x_2 + 0.0685x_1^2 + 3.849x_2^2 \quad (15)$$

$$R^2 = 0.813, R^2_{adj} = 0.763, s = 6.755, PRESS = 2307.945$$

The normality and constant variance tests were passed. Three influential data points were found.

$$y_{WIP} = 10514.817 - 166.362x_1 + 1878.219x_2 + 2653.558x_3 + 1.765x_1^2 - 164.955x_3^2 - 22.265x_1x_2 \quad (16)$$

$$R^2 = 0.922, R^2_{adj} = 0.886, s = 162.678, \text{PRESS} = 948,709.4$$

The normality test was failed ($p = 0.008$), and the constant variance test was passed. One data point was influential (13th).

$$y_{TP} = -1920.381 + 40.347x_1 + 124.394x_2 + 248.409x_3 - 0.184x_1^2 - 15.781x_2^2 - 8.908x_3^2 - 1.204x_1x_3 \quad (17)$$

$$R^2 = 0.768, R^2_{adj} = 0.633, s = 15.875, \text{PRESS} = 13,369.7$$

The normality test was failed ($p = 0.008$), and the constant variance test was failed ($p = 0.003$). Three influential data points were found (5th, 15th, 18th). For the y_{CT} response, x_3^2 did not significantly affect the response. For the y_{WIP} response, x_3^2 was not included in this RSM equation. The authors presented plots of the surface curve using the full equation. However, this full equation was inappropriate. The y_{WIP} response was under non-normality conditions. Both tests of normality and constant variance failed for the y_{TP} response. An advanced regression technique needs to be performed to remedy these conditions.

Seo et al. [26] optimized a tungsten chemical mechanical planarization (CMP) slurry for semiconductor manufacturing. The CCD experimental design was employed. The responses y_W and y_{Oxide} were the removal rates of the thickness of the W and oxide films. The full equations of the two responses were proposed [26] and used to produce the contour and response surface plots.

The adequate equations with datasets listed in the literature [26] were evaluated by complete regression analysis:

$$y_W = 25.601 + 843.010x_2 \quad (18)$$

$$R^2 = 0.721, R^2_{adj} = 0.698, s = 138.318, \text{PRESS} = 328.601$$

$$y_{Oxide} = 187.296 + 0.834x_1 + 10.682x_3 \quad (19)$$

$$R^2 = 0.688, R^2_{adj} = 0.636, s = 14.801, \text{PRESS} = 4.068$$

Both responses passed the tests of normality and constant variance. In this study, only the x_2 variable had a significant effect on y_W . The x_1 and x_3 variables had a linear relationship with y_{Oxide} . The surface plots were presented in the literature [26]. The effect of the adequate equation to present the appropriate contour and response surface plots is evident by comparing the adequate and proposed full equations in the literature [26].

Saleem and Soma [27] used the Box–Behnken design and RSM to study the optimization of MEMS devices. The influencing factors were x_1 , top electrode length (TEL); x_2 , top electrode width (TEW); x_3 , torsion spring length; and x_4 , torsion spring width (TSW). The response y was the pull-in voltage.

The full equation, which included $x_1, x_2, x_3, x_4, x_1^2, x_2^2, x_3^2, x_4^2, x_1x_2, x_1x_3, x_1x_4, x_2x_4,$ and x_3x_4 , was proposed and used to produce the surface plots.

The full equation calculated with the datasets in literature [27] is

$$\begin{aligned}
 y_{PV} = & 27.200 - 7.550x_1 - 2.717x_2 - 6.208x_3 + 3.292x_4 + 0.154x_1^2 - 0.696x_2^2 \\
 & (<0.001) (<0.001) (<0.001) (<0.001) (0.842) (0.376) \\
 & + 1.192x_3^2 - 0.258x_4^2 + 0.825x_1x_2 + 1.800x_1x_3 - 0.925x_1x_4 + 0.425x_2x_3 \\
 & (0.147) (0.731) (0.364) (0.062) (0.311) (0.635) \\
 & - 0.400x_2x_4 + 1.350x_3x_4 \\
 & (0.655) (0.148) \\
 R^2 = & 0.975, R^2_{adj} = 0.945, s = 1.748, PRESS = 211.08
 \end{aligned}
 \tag{20}$$

The adequate equation calculated by backward elimination regression is

$$\begin{aligned}
 y_{PV} = & 26.773 - 7.550x_1 - 2.717x_2 - 6.208x_3 + 3.292x_4 + 1.352x_3^2 + 1.801x_1x_3 \\
 R^2 = & 0.962, R^2_{adj} = 0.95, s = 1.659, PRESS = 100.206
 \end{aligned}
 \tag{21}$$

Comparing the R^2_{adj} and s values, the adequate equation has better accuracy than the full equation. The adequate equation also exhibits better predictive performance, as indicated by the PRESS criterion. The normality test is failed ($p = 0.0046$). Further analysis is needed.

3.2. Steel Materials

The results of evaluating adequate equations for steel materials are listed in Table 5.

Table 5. Results of the evaluation of adequate RSM equations for steel materials.

Source	Reported Equations	Contour and Surface Response Plot	Adequate Equations	Normality Test	Constant Variance Test	Influential Data Points
Noordin et al. [28]	$y_{SR} = f(x_2, x_3, x_2x_3, x_3^2)$ $y_{TF} = f(x_1, x_2, x_3, x_2x_3, x_3^2)$	Curve surface	$y_{SR} = f(x_2, x_3, x_1x_2, x_2x_3, x_3^2)$ $y_{TF} = f(x_1, x_2, x_3, x_2x_3, x_3^2)$	Passed	Passed	7, 12, 13
Bouacha et al. [29]	$y_{Ra} = f(x_1, x_2, x_3, x_1x_2)$ $y_{Rt} = f(x_1, x_2, x_1x_2, x_2^2)$ $y_{Rz} = f(x_1, x_2, x_1x_2, x_2^2)$	Curve surface Curve surface Curve surface	$y_{Ra} = f(x_1, x_2, x_1x_2, x_1^2, x_2^2)$ $y_{Rt} = f(x_1, x_2, x_1x_2, x_1^2, x_2^2)$ $y_{Rz} = f(x_1, x_2, x_3, x_1x_2, x_1x_3, x_1^2, x_2^2)$	Passed Passed Failed	Passed Passed Passed	6, 13 1, 4 5, 6
Elbah et al. [30]	Full models for y_{Fa} , y_{Fr} , y_{Ft} , y_{Ra}	Curve surface	$y_{Fa} = f(x_1, x_2, x_3, x_1x_3, x_2x_3)$ $y_{Fr} = f(x_1, x_2, x_3, x_1x_2, x_1x_3)$ $y_{Ft} = f(x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3)$ $y_{Ra} = f(x_1, x_2, x_1x_2)$	Passed Passed Passed Passed	Passed Passed Passed Passed	26 24 27 No
Campos et al. [31]	Full models	Curve surface	$y_{Time} = \text{full equation}$ $y_{Ra} = f(x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1^2, x_3^2)$ $y_{Rt} = f(x_1, x_1^2)$	Passed Passed Passed	Passed Passed Passed	1, 8, 16, 17 1, 3, 4, 6 18
Khalil et al. [32]	Full model	Curve surface	$y_{SR} = f(x_1, x_2, x_3, x_1x_2, x_1x_3, x_3^2)$	Passed	Passed	1, 7, 17, 20

Noordiu et al. [28] investigated the performance of coated carbide tools using a CCD design. The influencing variables were x_1 , cutting speed; x_2 , feed; and x_3 , side cutting edge angle (SCEA), and the responses were y_{Ra} , surface roughness (Ra), and y_{Fc} , tangential force (Fc). The regression procedure was introduced in detail, and the backward elimination procedure was used. The model assessment criteria included R^2 , R^2_{adj} , PRESS, and lack of fit [28].

The response y_{Ra} and y_{Fc} calculations in this study are presented in Appendices A.2 and A.3. The complete regression analysis produces the appropriate model, which is consistent with the report by the authors [28]. In our study, the tests of normality and constant variance are performed. Both responses, y_{Ra} and y_{Fc} , pass. For the y_{Ra} response, one influential data point (7th) is identified. Two influential data points (12th, 13th) exist for the y_{Fc} response.

Bouacha et al. [29] studied the physical properties in hard turning with a cubic boron nitride (CBN) tool. The factors affecting the response were x_1 , cutting speed; x_2 , feed rate; and x_3 , depth of cut. The responses were surface roughness y_{Ra} , arithmetic average of absolute roughness Ra; y_{Rt} , maximum height of the profile Rt; and y_{Rz} , average maximum height of the profile y_{Rz} .

The ANOVA tables of y_{Ra} , y_{Rt} , and y_{Rz} are presented in the literature [29]. The authors use $p < 0.05$ as a criterion and then delete all other variables for which the p -value is greater than 0.05. The coefficient of the parameters at this first calculation of variables with p -values < 0.05 is used to propose the final equation. The interaction of these variables is not considered. The remaining variables of their proposal equation, which consist of three responses, are presented in the published table [29].

The adequate equations evaluated and checked with the regression technique are as follows:

$$y_{Ra} = 0.285 - 0.00841x_1 + 14.410x_2 + 0.0000215x_1^2 - 33.681x_2^2 - 0.0128x_1x_2 \quad (22)$$

$$R^2 = 0.991, R^2_{adj} = 0.989, s = 0.018, \text{PRESS} = 0.0115$$

Both tests (normality and constant variance) are passed, and two influential data points are identified (6th and 13th).

$$y_{Rt} = 2.221 - 0.0548x_1 + 86.001x_2 + 0.000136x_1^2 - 208.333x_2^2 - 0.068x_1x_2 \quad (23)$$

$$R^2 = 0.986, R^2_{adj} = 0.982, s = 0.139, \text{PRESS} = 0.676$$

Both tests (normality and constant variance) are passed, and two influential data points are found (1st, 4th)

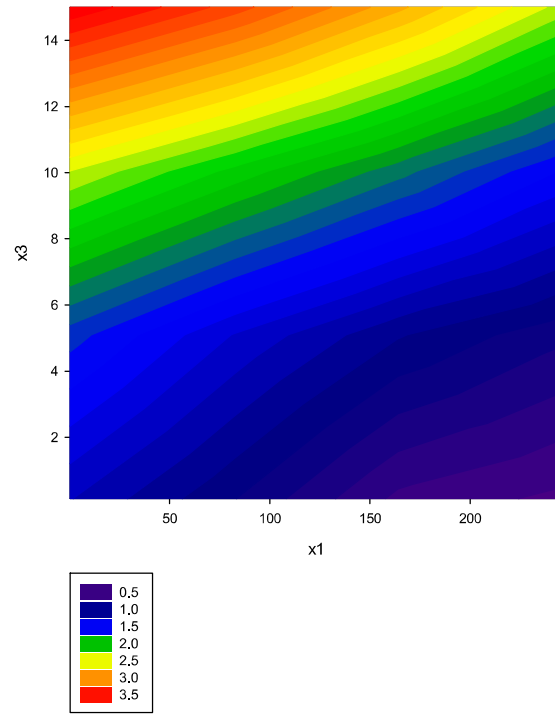
$$y_{Rz} = 2.994 - 0.0409x_1 + 40.071x_2 + 1.575x_3 + 0.000951x_1^2 - 78.472x_2^2 - 0.0357x_1x_2 - 0.00541x_1x_3 \quad (24)$$

$$R^2 = 0.993, R^2_{adj} = 0.990, s = 0.079, \text{PRESS} = 0.220$$

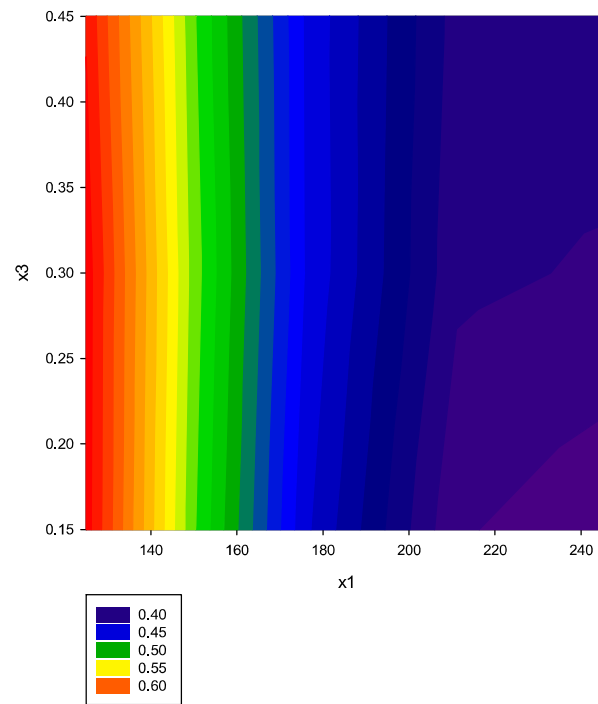
The normality test is failed ($p = 0.007$). The constant variance test is passed. Two influential data points are found (5th, 6th).

The x_3 variable did not significantly affect y_{Ra} and y_{Rt} . For the y_{Rz} response, the x_3 variable only had the linear effect, and the quadratic term was insignificant. In this study [29], curve surface plots were produced using full equations rather than the appropriate models.

Figure 2 shows the contour plots for the complete and adequate equations. The difference in the equations induces a difference in the distribution of curves between the two figures. The contour and response surface plots produced with the complete equation are presented in the study [29]. An inadequate RSM equation can induce incorrect results.



(a)



(b)

Figure 2. Contour plots of the complete and adequate equations for the average absolute roughness of AISI 52100 bearing steel. The contours present the effect of using the RSM equation on the relationship between the response variables and the influencing variables. (a) Full equation. (b) Adequate equation.

Elbah et al. [30] performed a mixed ceramic tool performance test. The factors affecting performance were x_1 , depth of cut; x_2 , feed rate; and x_3 , cutting speed. The response factors were y_{Fa} , axial force; y_{Fr} , thrust force; y_{Ft} , tangential force; and y_{Rs} , surface roughness.

In the literature [30], four ANOVA tables with four responses were listed, along with the p -values of each variable and a notation indicating whether the variable was significant

or not. However, this information was not used. The full equations of the four responses were proposed and used to produce the contour and response surface plots.

The forms of the four responses evaluated by the complete regression technique in our study are as follows:

$$y_{Fa} = f(x_1, x_2, x_3, x_1x_3, x_2x_3) \tag{25}$$

$$y_{Fr} = f(x_1, x_2, x_3, x_1x_2, x_1x_3) \tag{26}$$

$$y_{Ft} = f(x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3) \tag{27}$$

$$y_{Rs} = f(x_1, x_2, x_1x_2) \tag{28}$$

The quadratic terms had no significant effect on the responses. The curved surface plots were inappropriate and could easily induce incorrect results.

Campos et al. [31] observed the machining of hardened steels with CCD in an RSM study. The influencing factors were x_1 , cutting speed; x_2 , feed rate; and x_3 , cut depth. The responses were y_{Time} , time; y_{Ra} , average surface roughness Ra; and y_{Rt} , maximum height of the profile surface roughness Rt.

The sequential model test and the ANOVA results of the three responses are presented in the literature [31]. The significant effects are the linear + square model for y_{Time} , the linear + square model for y_{Ra} , and the linear + interaction model for y_{Rt} . In the three ANOVA tables, the p -values of some variables were >0.05 . However, the full models of the three responses were proposed. The curved surfaces of both plots were presented.

The forms of the adequate equations in our study are

$$y_{Time} = \text{full equation.} \tag{29}$$

$$y_{Ra} = f(x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1^2, x_3^2). \tag{30}$$

$$y_{Rt} = f(x_1, x_1^2). \tag{31}$$

The adequate equation for y_{Time} is the full equation, yielding the same results as in the literature [31], and four influential data points are identified. For the Ra response y_{Ra} , the x_2^2 variable was not included in the adequate equation, and four influential data points were found. The curved surface plots were inappropriate for this response. For the y_{Rt} Rt response, only the variables of x_1 and x_1^2 have a significant effect on y_{Rt} .

Khalil et al. [32] reported optimizing the effect of machining factors on surface roughness for machining AISI D3 steel. The influencing factors were x_1 , cutting speed; x_2 , feed rate; and x_3 , cut depth. The response y_{SR} was the surface roughness. The results of the ANOVA for the experiment are presented in the literature [32]. The p -values of variables x_1^2 and x_2^2 were >0.05 . However, the full equation was still adopted, and curved response surface plots were produced.

The form of the adequate equation evaluated by the complete regression technique is

$$y_{SR} = f(x_1, x_2, x_3, x_1x_2, x_1x_3, x_3^2). \tag{32}$$

For the quadratic terms, only the x_2^3 term significantly affects the response.

Using the full equation involving x_1^2 and x_2^2 to produce the curved surface plots was inappropriate and could easily cause incorrect results.

3.3. Nanomaterials

The results of evaluating adequate RSM equations for nanomaterials are presented in Table 6.

Table 6. Results of the evaluation of adequate RSM equations for nanomaterials.

Source	Reported Equations	Contour and Surface Response Plot	Adequate Equations	Normality Test	Constant Variance Test	Influential Data
Pakolpakcil et al. [37]	Full model	Curve surface	$y_{Afd} = f(x_1, x_2, x_3, x_2x_3, x_1^2, x_3^2)$	Passed	Failed	12
Pajaie and Taghizadeh [33]	Full model	Curve surface	$y_{Ethylene} = \text{full model}$ $y_{Propylene} = f(x_1, x_2, x_3, x_1x_3, x_1^2, x_2^2)$	Passed	Passed	3, 5, 8, 9, 11
Jourshabani et al. [34]	Full model	Curve surface	$y_{phenol\ yield} = f(x_1, x_2, x_3, x_1^2, x_2^2, x_3^2)$	Passed	Passed	9
Sheng et al. [35]	$\log(y_D + 0.5) = f(x_1, x_2, x_3, x_4, x_1x_2, x_2x_3, x_2x_4, x_3x_4, x_3^2)$	Curve surface	$y_D = f(x_1, x_2, x_3, x_4, x_1x_2, x_2x_4, x_3x_4, x_3^2)$	Passed	Passed	None
Rakhmanova et al. [36]	Not reported	Curve surface	$y_{Zinc\ oxide} = f(x_2, x_3)$	Passed	Passed	6, 14
Sareekumar et al. [38]	Full models for y_{nth} - $y_{nex,ele}$	Curve surface	$y_{nth} = f(x_1, x_2, x_4)$	Passed	Passed	18
		Curve surface	$y_{nele} = f(x_1, x_2, x_3, x_4, x_2x_3, x_2x_4)$	Passed	Passed	7, 24, 25
		Curve surface	$y_{nex,th} = f(x_1, x_2, x_3, x_2x_3, x_2x_4, x_3x_4)$	Passed	Passed	3, 25
		Curve surface	$y_{nex,ele} = f(x_1, x_2, x_3, x_4, x_1x_3, x_2x_3, x_2^2)$	Passed	Failed	3, 14

Pakolpakcil et al. [37] investigated the effect of processing parameters on the aerosol filter of poly nanofiber mats. A three-factorial BBD was used. The variables were x_1 concentration; x_2 , rotation speed; and x_3 , needle size. The response y_{Afd} was the average fiber diameter. The statistical results of the sequential models showed that the p -value of the quadratic term was <0.05 . The ANOVA table in this study [37] indicated that the p -values of x_1x_2 , x_1x_3 , and x_2^2 were >0.05 . That is, these variables did not have a significant effect on the response. However, the full equation was proposed and used to produce the contour and surface response plots [37].

The calculation of the adequate equation with the datasets in the literature is presented in Appendix A.4. The form of this adequate equation is

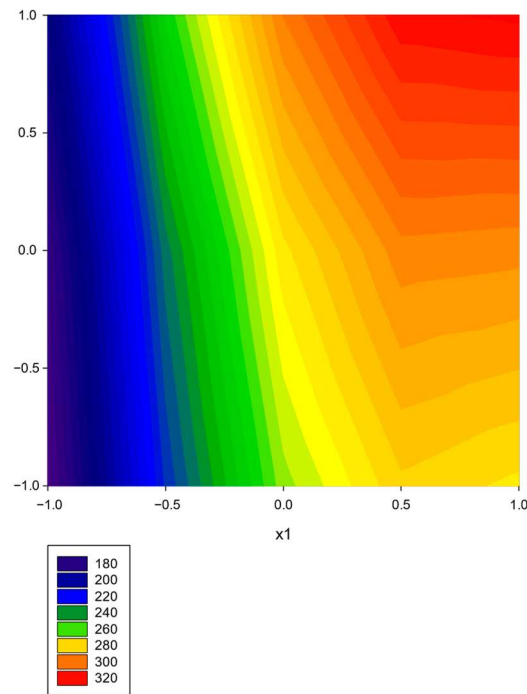
$$y_{Afd} = 277.462 + 60.375x_1 + 17.001x_2 + 15.875x_3 - 42.058x_1^2 + 34.442x_3^2 + 17.501x_1x_2 \tag{33}$$

The variable x_2^2 is excluded. That is, there is no quadratic relationship between x_2^2 and the response.

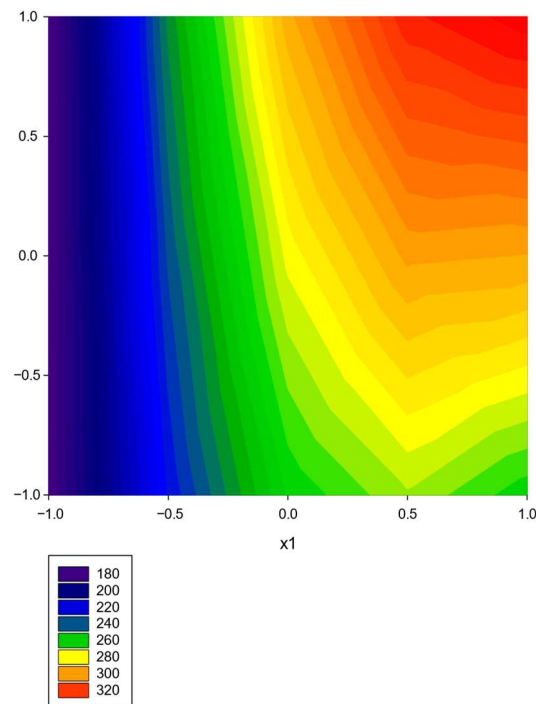
An adequate equation has a similar accuracy performance to that of a full equation. However, it has a better predictive ability (PRESS = 675.548) than the full equation (PRESS = 1,113,520).

Figure 3 reveals the contour plots for the complete and adequate equations. The difference in the equations induces a difference in the distribution of curves between the two figures. The contour and response surface plots produced with the full equation are

presented in the study [37]. An inadequate RSM equation can lead to incorrect observation results.



(a)



(b)

Figure 3. Contour plots of the complete and adequate equations for the average fiber diameter response. The contours present the effect of using the RSM equation on the relationship between the response variables and the influencing variables. (a) Full equation. (b) Adequate equation.

The normality test is passed. However, the constant variance test is failed. One suspicious data point (12th) is found. Further analysis needs to be performed.

Pajaie and Taghizadeh [33] reported optimizing the catalytic performance of synthesized catalysts for the methanol-to-olefin reaction. The BBD experimental design was used.

There were three variables: x_1 , the MW aging time; x_2 , the US aging time; and x_3 , the HT time. Two responses are *the yield of ethylene and the yield of propylene*.

Two ANOVA tables for $y_{ethylene}$ and $y_{propylene}$ are presented in the literature. For the ethylene yield, p -values of all variables are <0.05 . The full equation is an adequate representation of the ethylene yield. The R^2 value of ethylene is very close to 1.0 ($R^2 = 0.9995$).

In the ANOVA table of the propylene yield $y_{propylene}$, the p -values of x_2x_3 and x_3^2 were >0.05 . The two variables did not have a significant effect on response $y_{propylene}$. However, the authors used two full equations to produce the contour and response surface plots [33]. It is inappropriate to use the propylene yield as a measure. The x_3 HT time did not have the curved surface condition with $y_{propylene}$.

There are six suspicious data points for $y_{ethylene}$ and two suspicious data points for $y_{propylene}$. Further investigation needs to be performed.

Jourshabani et al. [34] investigated the factors influencing benzene hydroxylation to phenol using a V/SBA-16 nanoporous catalyst. The CCD was used. The variables were x_1 , reaction temperature; x_2 , H_2O_2 content; and x_3 , catalyst amount. The response was *the yield of phenol*.

In the ANOVA table for response, the p -value of the x_1x_3 variable is >0.05 . However, the authors proposed the full equation, which was used to produce the surface plots [34].

The adequate equation, evaluated using complete regression analysis, is $y_{phenol\ yield} = f(x_1, x_2, x_3, x_1^2, x_2^2, x_3^2)$. The interaction terms (x_1x_2, x_1x_3, x_2x_3) are excluded in this equation. That is, the full equation is inappropriate.

Sheng et al. [35] investigated the optimization of deposition variables to synthesize upright ZnO rod arrays with large diameters. There were four influencing factors: x_1 , the concentration of Zn^{+2} ; x_2 , reaction temperature; x_3 , reaction time; and x_4 , the molar ratio of Zn^{+2} . The response was y_D diameter. The authors used the logarithm transform for the response. The form of their proposal equation is

$$\text{Log}(y_D + 0.5) = f(x_1, x_2, x_3, x_4, x_3^2, x_1x_2, x_2x_3, x_2x_4, x_3x_4). \tag{34}$$

Both tests are passed for the y response using the normality and constant variance tests, so it is not necessary to transform the response y into a logarithmic form. The adequate equation calculated by complete regression analysis in our study is

$$\begin{aligned} y_D &= 1.182 - 57.722x_1 - 0.0289x_2 + 0.288x_3 + 0.00265x_4 - 0.00822x_3^2 + 0.778x_1x_2 \\ &\quad - 0.0000299x_2x_4 - 0.0000643x_3x_4 \end{aligned} \tag{35}$$

$$R^2 = 0.86, R^2_{adj} = 0.797, s = 0.169, \text{PRESS} = 1.18$$

Rakhmanova et al. [36] reported the optimization of nanosized zinc oxide synthesis conditions using electrospinning. Three influencing factors were applied: voltage; x_2 , distance; and x_3 , calcination temperature, and the response, y , was zinc oxide.

The authors did not propose the empirical equation; instead, surface plots of the curve were presented [36]. The full equation evaluated by complete regression with the datasets in the literature is

$$\begin{aligned} y_{Zinc\ oxide} &= 302.751 - 0.577x_1 + 1.625x_2 - 0.712x_3 + 0.000348x_1^2 - 0.00438x_2^2 \\ &\quad (0.629) (0.970) (0.985) (0.565) (0.997) \\ &\quad - 0.432x_3^2 - 0.00871x_1x_2 + 0.0113x_1x_3 + 0.0897x_2x_3 \\ &\quad (0.739) (0.733) (0.566) (0.932) \end{aligned} \tag{36}$$

$$R^2 = 0.768, R^2_{adj} = 0.351, s = 7.154, \text{PRESS} = 6124.736$$

In our study, the adequate equation evaluated by complete regression analysis is

$$y_{\text{Zinc oxide}} = 129.859 - 2.526x_2 - 3.663x_3$$

$$R^2 = 0.651, R^2_{\text{adj}} = 0.592, s = 5.668, \text{PRESS} = 639.412 \quad (37)$$

The normality and constant variance tests are passed, and two suspicious data points are found.

The PRESS values for the adequate and complete equation are 639.432 and 6124.736. The adequate equation offers a significant improvement in the predictive ability.

Sreekumar et al. [38] investigated the optimization of a photovoltaic/thermal system using a Mxene/water nanofluid as the heat transfer fluid via CCD. The four influencing variables were x_1 , nanofluid concentration; x_2 , nanofluid flow rate; x_3 , solar radiation; and x_4 , inlet temperature. The four responses were y_{nth} , thermal efficiency; y_{nele} , electrical efficiency; y_{nex} , thermal exergy efficiency; and $y_{nex,ele}$, electrical exergy efficiency.

The author proposed full equations for four responses, which were used to produce curve surface plots of contour and 3D response surface plots [38]. The coefficients of variables and their corresponding p -values for four responses (y_{nth} , y_{nele} , $y_{nex,th}$, and $y_{nex,ele}$) are presented in the table from the study [38]. In this table, many p -values of each parameter in the full equations are >0.05 . This indicates that these parameters did not significantly affect the response. However, the complete equations were adopted [38].

The complete regression technique evaluated the adequate equation of the four responses.

$$y_{nth} = 51.078 + 29.749x_1 + 0.257x_2 - 0.631x_4$$

$$R^2 = 0.814, R^2_{\text{adj}} = 0.789, s = 4.017, \text{PRESS} = 515.04 \quad (38)$$

The normality and constant variance tests are passed. One suspicious data point is found.

$$y_{nele} = 19.452 - 0.462x_1 + 0.00935x_2 - 0.00438x_3 - 0.0539x_4 + 0.0000238x_2x_3$$

$$-0.000406x_2x_4$$

$$R^2 = 0.984, R^2_{\text{adj}} = 0.979, s = 0.171, \text{PRESS} = 1.419 \quad (39)$$

The normality and constant variance tests are passed. Three suspicious data points are found.

$$y_{nex,th} = 0.601 - 3.364x_1 - 0.00775x_2 + 0.00353x_3 - 0.0271x_4 + 19.391x_1^2$$

$$-0.0000171x_2x_3 + 0.000328x_2x_4 - 0.0000258x_3x_4$$

$$R^2 = 0.979, R^2_{\text{adj}} = 0.970, s = 0.119, \text{PRESS} = 0.666 \quad (40)$$

The normality and constant variance tests are passed. Two suspicious data points are found.

$$y_{nex,ele} = 20.366 - 0.175x_1 + 0.0225x_2 - 0.00427x_3 - 0.0748x_4 - 0.000192x_2^2$$

$$-0.00158x_1x_3 + 0.0000196x_2x_3$$

$$R^2 = 0.998, R^2_{\text{adj}} = 0.997, s = 0.068, \text{PRESS} = 0.175 \quad (41)$$

The normality test is passed. However, the constant variance test is failed ($p = 0.047$). Two suspicious data points are found.

For y_{nth} and y_{nele} , only the linear relationship is valid. No quadratic terms (x_1^2 , x_2^2 , or x_4^2) significantly affect y_{nth} and y_{nele} . So, the curve surface plots were inappropriate. For $y_{nex,th}$, only x_2^2 has a significant effect on response. For $y_{nex,ele}$, only the x_2^2 term exists in the adequate equation. The constant variance test is failed for y_{nex} and y_{ele} . Advanced regression techniques should be performed to remedy these conditions.

4. Discussion

This study collected 16 papers related to the application of RSM models in three engineering fields. This literature dataset was used to evaluate the adequacy of RSM equations. The evaluation of the adequate RSM model was completed through a complete regression analysis. This analysis calculated the coefficients for all variables, tested the significant effect of each variable, verified the assumptions, and identified influential data points in the regression analysis. It was found that only one paper reported an equation that could fully and correctly express the relationship between the response and the influencing factors [28].

The common issues with the application of RSM in three engineering fields, as identified in the study, are listed in Table 7.

Table 7. Common issues with the applications of RSM in the literature.

Issue	Literature
1. The full model was used to generate contour plots and three-dimensional response surface plots, which were then used to optimize these variables. However, the coefficient values, <i>t</i> -values, and <i>p</i> -values for each variable in the ANOVA table were not used.	[24–27,30,32–34,38]
2. The contour plots and three-dimensional response surface plots with the full equation were proposed, but the RSM model was not reported.	[36]
3. All variables with <i>p</i> -values higher than the preselected value (usually $p < 0.05$) were deleted at once.	[29]
4. The ANOVA table of the sequential model was used and all variables were included in the linear or squared term directly, without conducting significance testing for each variable.	[31,37]
5. Datasets did not pass the normality test.	[25,27,29]
6. Datasets did not pass the constant variance test.	[25,37,38]
7. Influential data points were found.	[23–34,36–38]

Most papers adopted the full model and then used it to create contour plots and 3D response surface plots, allowing for the observation of the optimization of these variables. In the literature, ANOVA tables typically include the coefficient value of variable parameters, the *t*-value, and the *p*-value for each variable. However, this information was not used by some researchers [24–27,30,32–34,38] to evaluate the adequacy of equations.

One study did not report the RSM model [36]. The contour and response surface plots were produced with the full equation.

One study completely deleted unwanted variables after the first regression calculation [29]. When the ANOVA table of regression results indicated that the *p*-values of some variables were greater than 0.05, these variables were deleted simultaneously [29]. The equation was proposed using the coefficient values of the remaining variables. However, the full equations were still used to yield the contour and response surface plots for this study [29]. This method is unreasonable for model building.

Two studies employed sequential models to investigate the significant effects of linear, interaction, and quadratic terms on the response [31,37]. The *p*-values of some combinations (linear + interaction, linear + square terms) indicated that these combinations did not

significantly affect the response. However, full equations were still used to produce contour and response surface plots.

If the RSM models are not full equations, the curved contour and 3D surface response plots produced using the full equations are inappropriate. The performance of the optimization conditions of these variables by these plots is meaningless.

Some datasets failed the normality test [25,27,29], and others failed the constant variance test [25,37,38]. These datasets need to be transformed to align with the assumptions of regression analysis. Yang et al. [50] emphasized that departures from the homogenous assumption will induce seriously incorrect results and require remedying this violation. Sheng et al. [35] used the logarithmic transformation of the response to perform the regression analysis. However, the response y_D of their datasets did not violate the normality assumption, as determined by the Kolmogorov–Smirnov method in our study.

The implementation of remedial measures for heteroscedasticity in regression analysis includes transforming the dependent variable, such as log transformation, $\log(y)$; square root, $(y^{0.5})$, and Box–Cox transformation; using weighted least squares (WLS); considering segmenting the data; and using generalized least squares (GLS) [51–53].

Influential data points are usually found in the datasets of the examined literature. These data points may be the source of experimental error or indicate the need for further study of other forms of RSM models. The treatment of influential points in regression analysis involves identifying whether these points are due to a pure data entry error or a valid data point that is an extreme value [54,55].

If an influential point is due to a mistake (e.g., experimental error or sampling error), it can be corrected or removed. Valid data points with extreme values can be transformed to reduce their influence, or robust regression can be used [54,55].

Asoo et al. [3] reported the integration of computer technology in RSM. They described the high-performance computing used to calculate the models and visualize the results, which helped the researchers understand the relationships between influencing variables and responses. Graphical visualization helped the researchers interpret results and make decisions [3]. However, the effect of adequate RSM models on graphical visualization was not considered.

The challenges of RSM include limitations in modeling nonlinear systems, sensitivity to experimental error, model interpretability, and model validation [3]. The nonlinear systems of an RSM experiment can be evaluated with nonlinear regression. The criteria for evaluating linear regression can also be applied to a nonlinear system. Some advanced modeling techniques, such as machine learning, neural networks, and support vector machines, can be applied and incorporated into the regression analysis technique. The experimental error can be further studied and quantified by checking the criteria of influential data points. The model's interpretability problems can be addressed by incorporating the criteria of regression analysis. The PRESS criterion introduced in this study can be used as the model validation criterion to assess the predictive ability of other RSM models.

Sample size is an essential criterion for ensuring the power of statistical techniques. Researchers have proposed simple sample size equations to evaluate the required sample size (n) for multiple regression models. These criteria can be applied in the calculation of RSM models. Snee [56] proposed this equation: $n \geq 2p + 20$. Green's equation is $n \geq 8p + 50$ [57]. The calculation equation used by Khamis and Kepler is $n \geq 5p + 20$ [58]. In these equations, n is the required number of samples, and p is the number of parameters in the RSM models. With these equations, the sampling numbers are not great enough for the RSM equations in most studied. The solution method is to increase the replicates at each run.

The study used sixteen published studies to evaluate the adequacy of the RSM equation. The 16 papers used in this study employed the following commercial regression analysis software: Design Expert (10 papers), Minitab (2 papers), and MOODE (1 paper). Three papers did not report the software used. The three commercial software packages provide detailed analyses of regression calculations, including regression coefficients, t -values, and ANOVA tables. However, most users struggle to utilize the calculation results of these programs. Ten papers, accounting for 62.5% of the total literature analyzed, utilized Design Expert software. However, only one study [28] reported the appropriate equation.

Based on the results of this study, several suggestions are proposed for the application of RSM in experimental design within engineering fields.

1. For engineers using RSM, receiving complete regression analysis training will help them in their research work. Engineers not only need to be able to use commercial software programs to calculate the estimated values of parameter coefficients but also need to be familiar with screening influencing variables, checking the conditions of regression analysis assumptions, and examining all possible influencing data points. Training in complete regression techniques can enhance researchers' ability to establish appropriate RSM equations.
2. Ask a statistician for help with the experimental design and verify the validity of the regression calculation.
3. Many commercial software programs can calculate RSM models and create precision contour and response surface plots. The backward elimination technique is beneficial in finding an adequate equation. It is recommended to integrate this backward elimination method into commercial software to assist researchers in developing suitable RSM equations.

5. Conclusions

This study compiled sixteen datasets from the literature in three engineering fields to evaluate the adequacy of RSM equations with complete regression analysis. The results of this study raise some critical issues regarding the use of RSM models in engineering research, including the selection of the full equation without considering statistical validation, the removal of all variables with p -values above a preset value, the presence of non-normality and non-constant variance conditions in the data set, and the presence of influential data points.

These issues need to be considered in RSM modeling. The sample size should be increased to enhance statistical power. Some suggestions for engineering researchers include training them in the complete regression technique, seeking the assistance of a statistician in experimental design and data analysis, and incorporating the backward elimination technique into commercial software programs, especially RSM software.

Author Contributions: Conceptualization, H.-Y.C. and C.C.; methodology, H.-Y.C. and C.C.; software, C.C.; formal analysis, H.-Y.C.; investigation, H.-Y.C. and C.C.; data curation, H.-Y.C.; writing—original draft preparation, H.-Y.C. and C.C.; writing—review and editing, H.-Y.C. and C.C.; visualization, C.C. supervision, C.C.; project administration, C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data is unavailable because a statement is still required.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1. Data from the Coating Experiment [49]

Run	Speed	Pressure	Distance	Mean, y_1	Standard Deviation, y_2
1	-1	-1	-1	24	12.5
2	0	-1	-1	120.3	8.4
3	+1	-1	-1	213.7	42.8
4	-1	0	-1	86	3.5
5	0	0	-1	136.6	80.4
6	+1	0	-1	340.7	16.2
7	-1	+1	-1	112.3	27.6
8	0	+1	-1	256.3	4.6
9	+1	+1	-1	271.7	23.6
10	-1	-1	0	81	0.0
11	0	-1	0	101.7	17.7
12	+1	-1	0	357.0	32.9
13	-1	0	0	171.3	15.0
14	0	0	0	372.0	0.0
15	+1	0	0	501.7	92.5
16	-1	+1	0	264.0	63.5
17	0	+1	0	427.0	88.6
18	+1	+1	0	730.7	21.1
19	-1	-1	+1	220.7	133.8
20	0	-1	+1	239.7	23.5
21	+1	-1	+1	422.0	18.5
22	-1	0	+1	199.0	29.4
23	0	0	+1	485.3	44.7
24	+1	0	+1	673.7	158.2
25	-1	+1	+1	176.7	55.5
26	0	+1	+1	501.0	138.9
27	+1	+1	+1	1010.0	142.4

Appendix A.2. Evaluation of the RSM Models for y_{Ra} , Ra Surface Roughness [28]

$$\begin{aligned}
 1. \ y_{Ra} = & 0.210 - 0.0138x_1 + 15.133x_2 - 0.166x_3 + 0.0000308x_1^2 + 32.24x_2^2 \\
 & (0.403) (0.490) (0.541) (0.245) (0.468) \\
 & + 0.419x_3^2 - 0.0268x_1x_2 + 0.000110x_1x_3 + 1.375x_2x_3 \\
 & (0.059) 0.192) (0.773) (0.031) \\
 & R^2 = 0.982, R^2_{adj} = 0.954, s = 0.174, PRESS = 2.470
 \end{aligned} \tag{A1}$$

Delete x_1x_3 and recalculate the equation.

$$\begin{aligned}
 2. \ y_{Ra} = & 0.297 - 0.141x_1 + 15.102x_2 - 0.0827x_3 + 0.0000307x_1^2 + 33.310x_2^2 \\
 & (0.358) (0.457) (0.559) (0.210) (0.433) \\
 & + 0.0418x_2^2 - 0.0268x_1x_2 + 1.375x_2x_3 \\
 & (0.041) (0.159) (0.020)
 \end{aligned} \tag{A2}$$

Delete x_2^2 and recalculate the equation.

$$3. y_{Ra} = -0.863 - 0.0177x_1 + 30.425x_2 - 0.0606x_3 + 0.0000366x_1^2 + 0.0463x_3^2 - 0.0268x_1x_2 + 1.375x_2x_3$$

(0.222) (<0.001) (0.652) (0.115) (0.018)
(0.147) (0.015)

(A3)

Delete x_1x_2 and recalculate the equation.

$$4. y_{Ra} = 1.022 - 0.0239x_1 + 22.228x_2 - 0.0599x_3 + 0.0000366x_1^2 + 0.463x_3^2 + 1.372x_2x_3$$

(0.120) (<0.001) (0.680) (0.137) (0.023)
(0.020)

(A4)

Delete x_1^2 and recalculate the equation.

$$5. y_{Ra} = -2.231 - 0.00125x_1 + 22.228x_2 + 0.00395x_3 + 0.0590x_3^2 + 1.372x_1x_3$$

(0.181) (0.001) (0.979) (0.004) (0.020)

(A5)

Although the linear term x_3 is insignificant ($p > 0.05$), the quadratic term x_3^2 and interaction term x_1x_3 are significant in the model. The linear term x_3 is hierarchically retained in the equation.

Delete x_1 and recalculate the equation.

$$6. y_{Ra} = -2.714 + 22.288x_2 + 0.000289x_3 + 0.0583x_3^2 + 1.372x_1x_2$$

(<0.001) (0.0199) (0.005) (0.003)

$$R^2 = 0.958, R^2_{adj} = 0.942, s = 0.195, PRESS = 1.013.$$
(A6)

The normality and constant variance tests are passed. One suspicious data point is found, point 7 ($t_i = 2.350, DFFITS_i = 2.058$).

Appendix A.3. Evaluation of the RSM Models for y_{Rc} , Tangent Force [28]

$$1. y_{Rc} = 264.240 - 0.294x_1 + 199.40x_2 - 7.094x_3 + 0.000473x_1^2 + 3750.468x_2^2 + 2.668x_3^2 - 0.0707x_1x_2 + 0.00244x_1x_3 + 62.058x_2x_3$$

(0.725) (0.858) (0.481) (0.736) (0.143)
(0.029) (0.943) (0.244) (0.051)

$$R^2 = 0.994, R^2_{adj} = 0.985, s = 9.039, PRESS = 3864.088$$
(A7)

Delete x_1x_2 and recalculate the equation.

$$2. y_{Rc} = 269.233 - 0.311x_1 + 177.735x_2 - 7.092x_3 + 0.000473x_1^2 + 3750.468x_2^2 + 2.6688x_3^2 + 0.0277x_1x_3 + 62.050x_2x_3$$

(0.675) (0.857) (0.444) (0.715) (0.112)
(0.018) (0.205) (0.034)

(A8)

Delete x_1^2 and recalculate the equation.

$$3. y_{Rc} = 241.953 - 0.0410x_1 + 62.922x_2 - 6.563x_3 + 4000.060x_2^2 + 2.722x_3^2 + 0.0243x_1x_3 + 62.050x_2x_3$$

(0.490) (0.943) (0.445) (0.062) (0.007)
(0.178) (0.024)

(A9)

Delete x_1x_3 and recalculate the equation.

$$4. y_{Rc} = 261.851 - 0.104x_1 + 58.033x_2 + 0.842x_3 + 4010.659x_1^2 + 2.764x_3^2 + 62.050x_2x_3$$

(0.027) (0.950) (0.905) (0.072) (0.008) (0.027)

(A10)

Delete x_2^2 and recalculate the equation.

$$5. y_{Rc} = 52.737 - 0.103x_1 + 1902.950x_2 + 4.737x_3 + 3.543x_3^2 + 62.505x_2x_3$$

(0.046) (<0.001) (0.544) (0.002) (0.046)

$$R^2 = 0.988, R^2_{adj} = 0.982, s = 9.654, PRESS = 2279.002$$
(A11)

The normality and constant variance tests are passed. Two suspicious data points are found, the 12th ($t_i = 2.322$) 13th ($t_i = -2.326$).

Appendix A.4. Evaluation of the RSM Models for y_{Afd} , Average Fiber Diameter [37]

$$1. y_{Afd} = 274.667 + 60.375x_1 + 17.000x_2 + 15.875x_3 - 41.708x_1^2 + 4.542x_2^2 + 34.792x_3^2 + 6.000x_1x_2 + 9.750x_1x_3 - 17.501x_1x_3$$

(<0.001) (<0.014) (0.019) (0.002) (0.535) (0.004) (0.402) (0.197) (0.044)

$$R^2 = 0.982, R^2_{adj} = 0.95, s = 13.099, PRES S = 1,113,250$$
(A12)

Delete x_2^2 and recalculate the equation.

$$2. y_{Afd} = 277.462 + 60.375x_1 + 17.000x_2 + 15.875x_3 - 42.058x_1^2 + 34.44x_3^2 + 6.001x_1x_2 + 9.750x_1x_3 - 17.501x_2x_3$$

(<0.001) (<0.014) (0.019) (0.002) (0.535) (0.373) (0.169) (0.031)

(A13)

Delete x_1x_2 and recalculate the equation.

$$3. y_{Afd} = 277.462 + 60.375x_1 + 17.000x_2 + 15.875x_3 - 42.0058x_1^2 + 34.442x_3^2 + 9.750x_1x_3 - 17.501x_2x_3$$

(<0.001) (<0.006) (0.009) (<0.001) (0.001) (0.160) (0.026)

(A14)

Delete x_1x_3 and recalculate the equation.

$$4. y_{Afd} = 277.462 + 60.375x_1 + 17.001x_2 + 15.875x_3 - 42.058x_1^2 + 34.442x_3^2 + 17.501x_2x_3$$

$$R^2 = 0.97, R^2_{adj} = 0.947, s = 13.502, PRESS = 675,548$$
(A15)

The normality test is passed. The constant variance test is failed ($p \geq 0.038$). One suspicious data point is found, point 12 (DFFITS_i = 2.981)

References

1. Myers, R.H.; Montgomery, D.C.; Anderson-Cook, C.M. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*; John Wiley & Sons: Hoboken, NJ, USA, 2016.
2. Anderson, M.J.; Whitcomb, P.J. *RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments*; Productivity Press: University Park, IL, USA, 2016.
3. Asoo, H.R.; Alakali, J.S.; Ikya, J.K.; Yusufu, M.I. Historical background of RSM. In *Response Surface Methods-Theory, Applications and Optimization Techniques*; IntechOpen: London, UK, 2024.

4. Baş, D.; Boyacı, İ.H. Modeling and optimization I: Usability of response surface methodology. *J. Food Eng.* **2007**, *78*, 836–845. [[CrossRef](#)]
5. Chelladurai, S.J.S.; Murugan, K.; Ray, A.P.; Upadhyaya, M.; Narasimharaj, V.; Gnanasekaran, S. Optimization of process parameters using response surface methodology: A review. *Mater. Today Proc.* **2021**, *37*, 1301–1304. [[CrossRef](#)]
6. De Oliveira, L.G.; de Paiva, A.P.; Balestrassi, P.P.; Ferreira, J.R.; da Costa, S.C.; da Silva Campos, P.H. Response surface methodology for advanced manufacturing technology optimization: Theoretical fundamentals, practical guidelines, and survey literature review. *Int. J. Adv. Manuf. Technol.* **2019**, *104*, 1785–1837. [[CrossRef](#)]
7. Veza, I.; Spraggon, M.; Fattah, I.R.; Idris, M. Response surface methodology (RSM) for optimizing engine performance and emissions fueled with biofuel: Review of RSM for sustainability energy transition. *Results Eng.* **2023**, *18*, 101213. [[CrossRef](#)]
8. Boshagh, F.; Rostami, K. A review of application of experimental design techniques related to dark fermentative hydrogen production. *J. Renew. Energy Environ.* **2020**, *7*, 27–42.
9. Mäkelä, M. Experimental design and response surface methodology in energy applications: A tutorial review. *Energy Convers. Manag.* **2017**, *151*, 630–640. [[CrossRef](#)]
10. Mishra, P.; Mohapatra, T.; Sahoo, S.S.; Padhi, B.N.; Giri, N.C.; Emará, A.; AboRas, K.M. Experimental assessment and optimization of the performance of a biodiesel engine using response surface methodology. *Energy Sustain. Soc.* **2024**, *14*, 28. [[CrossRef](#)]
11. Pais-Chanfrau, J.M.; Núñez-Pérez, J.; del Carmen Espin-Valladares, R.; Lara-Fiallos, M.V.; Trujillo-Toledo, L.E. Uses of the response surface methodology for the optimization of agro-industrial processes. In *Response Surface Methodology in Engineering Science*; IntechOpen: London, UK, 2021.
12. Boublia, A.; Lebouchera, S.E.I.; Haddaoui, N.; Guexxout, Z.; Ghriga, A.A.; Hasanzadeh, M.; Benguerba, Y.; Drouiche, N. State-of-the-art review on recent advances in polymer engineering: Modeling and optimization through response surface methodology approach. *Polym. Bull.* **2023**, *80*, 5999–6031. [[CrossRef](#)]
13. Bezerra, M.A.; Ferreira, S.L.C.; Novaes, C.G.; Dos Santos, A.M.P.; Valasques, G.S.; da Mata Cerqueira, U.M.F.; dos Santos Alves, J.P. Simultaneous optimization of multiple responses and its application in Analytical Chemistry—A review. *Talanta* **2019**, *194*, 941–959. [[CrossRef](#)]
14. Dejaegher, B.; Vander Heyden, Y. Experimental designs and their recent advances in set-up, data interpretation, and analytical applications. *J. Pharm. Biomed. Anal.* **2011**, *56*, 141–158. [[CrossRef](#)]
15. Szpisják-Gulyás, N.; Al-Tayawi, A.N.; Horváth, Z.H.; László, Z.; Kertész, S.; Hodúr, C. Methods for experimental design, central composite design and the Box–Behnken design, to optimise operational parameters: A review. *Acta Aliment.* **2023**, *52*, 521–537. [[CrossRef](#)]
16. Olabinjo, O.O. Response surface techniques as an inevitable tool in optimization process. In *Response Surface Methods—Theory, Applications and Optimization Techniques*; IntechOpen: London, UK, 2024.
17. Myers, R.H. *Classical and Modern Regression with Applications*, 2nd ed.; Duxbury Press: Monterey, CA, USA, 1990.
18. Berger, D.E. Introduction to Multiple Regression. Master’s Thesis, Claremont Graduate University, Claremont, CA, USA, 2008.
19. Meloun, M.; Militký, J. Detection of single influential points in OLS regression model building. *Anal. Chim. Acta.* **2001**, *439*, 169–191. [[CrossRef](#)]
20. Bhattacharya, S. Central composite design for response surface methodology and its application in pharmacy. In *Response Surface Methodology in Engineering Science*; IntechOpen: London, UK, 2021.
21. Rodrigues, A.C. Response surface analysis: A tutorial for examining linear and curvilinear effects. *Rev. Adm. Contemp.* **2021**, *25*, e200293. [[CrossRef](#)]
22. Reza, A.; Chen, L.; Mao, X. Response surface methodology for process optimization in livestock wastewater treatment: A review. *Heliyon* **2024**, *10*, e30326. [[CrossRef](#)]
23. Won, J.K.; Lee, J.H.; Lee, J.T.; Lee, E.S. The selection on the optimal condition of Si-wafer final polishing by combined Taguchi method and respond surface method. *Trans. Korean Soc. Eng. A.* **2008**, *17*, 21–28.
24. Lee, E.S.; Hwang, S.C.; Lee, J.T.; Won, J.K. A study on the characteristic of parameters by the response surface method in final wafer polishing. *Int. J. Precis. Eng. Manuf.* **2009**, *10*, 25–30. [[CrossRef](#)]
25. Zhang, H.; Jiang, Z.; Guo, C. Simulation-based optimization of dispatching rules for semiconductor wafer fabrication system scheduling by the response surface methodology. *Int. J. Adv. Manuf. Technol.* **2009**, *41*, 110–121. [[CrossRef](#)]
26. Seo, J.; Kim, J.H.; Lee, M.; You, K.; Moon, J.; Lee, D.H.; Paik, U. Multi-objective optimization of tungsten CMP slurry for advanced semiconductor manufacturing using a response surface methodology. *Mater. Des.* **2017**, *117*, 131–138. [[CrossRef](#)]
27. Saleem, M.M.; Somá, A. Design of experiments based factorial design and response surface methodology for MEMS optimization. *Microsyst. Technol.* **2015**, *21*, 263–276. [[CrossRef](#)]
28. Noordin, M.Y.; Venkatesh, V.C.; Sharif, S.; Elting, S.; Abdullah, A. Application of response surface methodology in describing the performance of coated carbide tools when turning AISI 1045 steel. *J. Mater. Process. Technol.* **2004**, *145*, 46–58. [[CrossRef](#)]

29. Bouacha, K.; Yallese, M.A.; Mabrouki, T.; Rigal, J.F. Statistical analysis of surface roughness and cutting forces using response surface methodology in hard turning of AISI 52100 bearing steel with CBN tool. *Int. J. Refract. Met. Hard Mater.* **2010**, *28*, 349–361. [[CrossRef](#)]
30. Elbah, M.; Aouici, H.; Meddour, I.; Yallese, M.A.; Boulanouar, L. Application of response surface methodology in describing the performance of mixed ceramic tool when turning AISI 4140 steel. *Mech. Ind.* **2016**, *17*, 309. [[CrossRef](#)]
31. Campos, d.S.P.H.; de Carvalho Paes, V.; de Carvalho Gonçalves, E.D.; Ferreira, J.R.; Balestrassi, P.P.; Davim, J.P. Optimizing production in machining of hardened steels using response surface methodology. *Acta Sci. Technol.* **2019**, *41*, e38091. [[CrossRef](#)]
32. Khalil, K.; Mohd, A.; Mohamad, C.O.C.; Faizul, Y.; Ariffin, S.Z. The optimization of machining parameters on surface roughness for AISI D3 steel. *J. Phys. Conf. Ser.* **2021**, *1874*, 012063. [[CrossRef](#)]
33. Pajaie, H.S.; Taghizadeh, M. Optimization of nano-sized SAPO-34 synthesis in methanol-to-olefin reaction by response surface methodology. *J. Indust. Eng. Chem.* **2015**, *24*, 59–70. [[CrossRef](#)]
34. Jourshabani, M.; Badiei, A.; Lashgari, N.; Mohammadi Ziarani, G. Application of response surface methodology as an efficient approach for optimization of operational variables in benzene hydroxylation to phenol by V/SBA-16 nanoporous catalyst. *J. Nanostructures* **2016**, *6*, 107–115.
35. Sheng, X.; Cheng, Y.; Yao, Y.; Zhao, Z. Optimization of synthesizing upright ZnO rod arrays with large diameters through response surface methodology. *Processes* **2020**, *8*, 655. [[CrossRef](#)]
36. Rakhmanova, A.; Kalybekkyzy, S.; Soltabayev, B.; Bissenbay, A.; Kassenova, N.; Bakenov, Z.; Mentbayeva, A. Application of response surface methodology for optimization of nanosized zinc oxide synthesis conditions by electrospinning technique. *Nanomaterials* **2022**, *12*, 1733. [[CrossRef](#)]
37. Pakolpakçıl, A.; Kılıç, A.; Draczynski, Z. Optimization of the centrifugal spinning parameters to prepare poly (butylene succinate) nanofibers mats for aerosol filter applications. *Nanomaterials* **2023**, *13*, 3150. [[CrossRef](#)]
38. Sreekumar, S.; Chakrabarti, S.; Hewitt, N.; Mondol, J.D.; Shah, N. Performance prediction and optimization of nanofluid-based PV/T using numerical simulation and response surface methodology. *Nanomaterials* **2024**, *14*, 774. [[CrossRef](#)]
39. Rawlings, J.O.; Pantula, S.G.; Dickey, D. *Applied Regression Analysis*; Springer: New York, NY, USA, 1998.
40. Allen, M.P. *Understanding Regression Analysis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2004.
41. Dielman, T.E. *Applied Regression Analysis for Business and Economics*, 4th ed.; Duxbury/Thomson Learning: Pacific Grove, CA, USA, 2005.
42. Mendenhall, W.; Sincich, T. *Regression Analysis. A Second Course in Statistics*, 12th ed.; Pearson: London, UK, 2012.
43. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2021.
44. Ryan, T.P. *Modern Regression Methods*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
45. Wilcox, R.R.; Keselman, H.J. Modern regression methods that can substantially increase power and provide a more accurate understanding of associations. *Eur. J. Pers.* **2012**, *26*, 165–174. [[CrossRef](#)]
46. Marinoiu, C. Classic and modern in regression modelling. *Econom. Insights Trends Chall.* **2017**, *69*, 41–50.
47. Rowley, E.K. Comparison of Variable Selection Methods. Ph.D. Thesis, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 2019.
48. Chowdhury, M.Z.I.; Turin, T.C. Variable selection strategies and its importance in clinical prediction modelling. *Fam. Med. Community Health* **2020**, *8*, e000262. [[CrossRef](#)]
49. Box, G.E.P.; Draper, N.R. *Response Surface, Mixtures, and Ridge Analysis*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2007.
50. Yang, K.; Tu, J.; Chen, T. Homoscedasticity: An overlooked critical assumption for linear regression. *Gen. Psychiatry* **2019**, *32*, e100148. [[CrossRef](#)]
51. Wang, G.C.; Akabay, C.K. Heteroscedasticity: How to handle in regression modeling. *J. Bus. Forecast.* **1994**, *13*, 11.
52. Agunbiade, D.A.; Adeboye, N.O. Estimation of heteroscedasticity effects in a classical linear regression model of a cross-sectional data. *J. Pro. Appl. Math.* **2012**, *4*, 18–28.
53. Kumar, N.K. Autocorrelation and heteroscedasticity in regression analysis. *J. Business Soc. Sci.* **2023**, *5*, 9–20. [[CrossRef](#)]
54. Stevens, J.P. Outliers and influential data points in regression analysis. *Psychol. Bull.* **1984**, *95*, 334. [[CrossRef](#)]
55. Chatterjee, S.; Hadi, A.S. *Regression Analysis by Example*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
56. Snee, R.D. Validation of regression models: Methods and examples. *Technometrics* **1977**, *19*, 415–428. [[CrossRef](#)]
57. Green, S.B. How many subjects does it take to do a regression analysis? *Multivar. Behav. Res.* **1991**, *26*, 499–510. [[CrossRef](#)]
58. Khamis, H.J.; Kepler, M. Sample size in multiple regression: $20 + 5k$. *J. Appl. Stat. Sci.* **2010**, *17*, 505–517.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.