AIP Conference Proceedings

RESEARCH ARTICLE | MARCH 17 2017

Estimating multivariate response surface model with data outliers, case study in enhancing surface layer properties of an aircraft aluminium alloy FREE

Edy Widodo; Kariyam

Check for updates

AIP Conf. Proc. 1823, 020054 (2017)
https://doi.org/10.1063/1.4978127

View Online
Export Citation
CrossMark

28 February 2024 01:41:34

AIP Advances
Why Publish With Us?
25 DAYS average time to 1st decision
740+ DOWNLOADS average per article
INCLUSIVE scope
Learn More
AIP Publishing

AIP Publishing

Estimating Multivariate Response Surface Model with Data Outliers, Case Study in Enhancing Surface Layer Properties of an Aircraft Aluminium Alloy

EdyWidodo[a)] and Kariyam [b)]

[1]*Department of Statistics Islamic University of Indonesia*

[a)] Corresponding author: edywidodo@uii.ac.id
[b)]kariyam@uii.ac.id

Abstract. To determine the input variable settings that create the optimal compromise in response variable used Response Surface Methodology (RSM).There are three primary steps in the RSM problem, namely data collection, modelling, and optimization. In this study focused on the establishment of response surface models, using the assumption that the data produced is correct. Usually the response surface model parameters are estimated by OLS. However, this method is highly sensitive to outliers. Outliers can generate substantial residual and often affect the estimator models. Estimator models produced can be biased and could lead to errors in the determination of the optimal point of fact, that the main purpose of RSM is not reached. Meanwhile, in real life, the collected data often contain some response variable and a set of independent variables. Treat each response separately and apply a single response procedures can result in the wrong interpretation. So we need a development model for the multi-response case. Therefore, it takes a multivariate model of the response surface that is resistant to outliers. As an alternative, in this study discussed on M-estimation as a parameter estimator in multivariate response surface models containing outliers. As an illustration presented a case study on the experimental results to the enhancement of the surface layer of aluminium alloy air by shot peening..

28 February 2024 01:41:34

INTRODUCTION

Response Surface Methodology (RSM), first introduced by Box and Wilson [1]. This method is often used to search for variable conditions of design that results in optimum response (maximum, minimum, or more broadly, looking for the conditions around a stationary point containing ridge). In general, there are three main stages in RSM, namely: (i) data collection, (ii) estimation model/data modelling, and (iii) optimization. The three stages are interrelated, where each stage will affect the next stage. Therefore the breadth of the field work of RSM, then this paper will be focused on the estimation of regression models for RSM, assuming that the data collection is satisfactory and the data have been collected.

To approach the relationship between the response variable and independent variables in the operating area is limited, in RSM, commonly used first-degree polynomial function of one or two order polynomial (quadratic model). Polynomial second order function is hereinafter referred to as second order models. In Myers et al. [5], a second order response surface model with independent variables $x_1, x_2, ..., x_p$ and the response variable y is written as follows:

$$y = \beta_o + \sum_{j=1}^{p} \beta_j x_j + \sum_{j=1}^{p} \beta_{jj} x_j^2 + \sum_{j<j'=2}\sum^{p} \beta_{jj'} x_j x_{j'} + \varepsilon \quad (1)$$

In the response surface models, estimation of the linear regression coefficient in equation (1), usually used Ordinary Least Square (OLS). However, the OLS method has a weakness, that is very sensitive to any irregularities

International Conference on Chemistry, Chemical Process and Engineering (IC3PE) 2017
AIP Conf. Proc. 1823, 020054-1–020054-13; doi: 10.1063/1.4978127
Published by AIP Publishing. 978-0-7354-1491-4/$30.00

or violations of the assumptions on the data. One form of the offence is a violation of the normality assumption. The normality assumption is often violated when the data contain outliers.

Outlier very likely produce a large residual and often affect the regression model was produced, so that the estimation model becomes biased and could result in errors in the actual determination of the optimal point. An example of the associated model error determination of the optimal point of the OLS method, is presented in the Fig. 1.

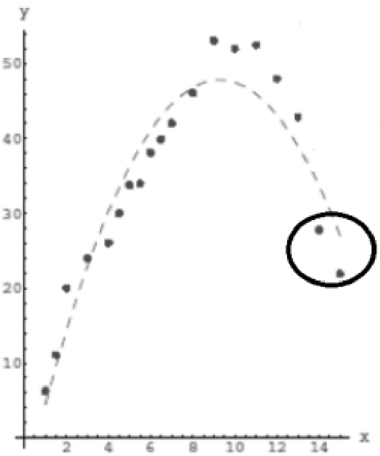

FIGURE 1. Plot of Raw Data and OLS Tensile Strength Paper (source: Wen [7])

Fig. 1 shows the plot between the tensile strength of paper (in psi) as the response variable (y), and the percentage of wood in batches of paper pulp (x) taken from the Wen [7]. Points (…) symbolize the raw data and the line (- - -) denote a model plot obtained with the OLS estimates. The figures show that the plot of raw data has a higher peak than the estimated plot by OLS. This indicates that this may have been an error in the determination of the model. One possible cause of these errors is the presence of data for outliers. It can be viewed in two recent data or data to the 14th and 15th, which has a value tended to be lower than the data patterns in general. As a result, if the model is forced to use it would be probable errors in determining the actual value of the optimal point. Therefore, the outliers data can influence the determination of the optimal point, while the purpose of the RSM is to determine the optimum point we need a response surface model is robust to outliers.

Some of the ways commonly used by researchers to treat the outer data, namely: (i) eliminates the outliers and (ii) maintain an outlier of data or use the entire data. At a certain moment, this outlier data should not just be abolished, because sometimes these outliers of data can provide information that cannot be provided by another observation point. For example, because the combination of unusual circumstances and there should need to be further investigated. An outlier may be eliminated if. After traced these observations are the product of measurement errors or errors in the measurement setup. Meanwhile, outlier data may also be influential. Outlier which is not an influential observation is an outlier that does not possess a strong influence on the model. Conversely, if the outlier is the data that effect, it will have an effect on the model, so as to affect the data, the data remain an outlier used to give little weight to data outliers. Furthermore, this method is called the robust regression method.

Robust usually connoted as the inconsistencies of an estimator to deviations from assumptions. Huber and Ronchetti [3] stated that more robust methods brought closer to the average parameters and variance-covariance of a particular estimator. The approach needed to be done to standardize the parameter estimator average and variance-covariance estimator can be generated so that a consistent against these parameters. Standardization is done with a form of restriction on the estimated values for the parameters. Robust with the expected results of the estimation will not stray too far.

Robust Regression is a regression method that is used when there are some outliers in the model. This method is an important tool to discuss data that is affected by outliers so that the resulting models are robust or stocky against outliers. A robust estimator is relatively unaffected by major changes on a small part of data or small changes in the bulk of the data. One method is a robust regression estimation with the M-estimation (Maximum Likelihood type). M-estimation method is the simplest and most widely used and is granted a high value of efficiency.

Meanwhile, in real life is very sparse data collected contains only one response variable, but frequent data collected include multiple response variable and a set of independent variables. Treat each response separately and

28 February 2024 01:41:34

apply a single response procedures can cause mistaken interpretations in the result. So we need a development model for multirespon case.

Development of the case multirespon of the response surface model was made by Wen [7] by using an approach Robust Regression Model 2 (MRR2), by combining parametric methods for matching OLS and Local Linear Regression (LLR) for nonparametric matching. As is well known that the OLS sensitive to outliers that MRR2 also has a weakness against the existence of data outliers, so it takes a robust method, one alternative is m-regression to develop the case multirespon.

Based on the above description, it appears that the presence of outliers can influence the response surface models, generated and may result in errors in the actual determination of the optimal point. Meanwhile, not all outliers can be removed from the data, so it takes a response surface model robust against outliers. One method is a robust regression estimation with the M-estimation. On the other hand, often in real life data collected includes multiple response variable and a set of independent data. So we need a development model for multirespon case. Therefore, the problem that arises is how-M estimation method for multivariate response surface model with data outliers. As illustrated in this paper presented a case based on the results of the experiments reported in Nam et al. [6], doing research on the enhancement of the surface layer of aluminium alloy air by shot peening.

ESTIMATOR PARAMETERS IN MULTIVARIATE RESPONSE SURFACE MODEL WITH DATA OUTLIERS

Consider the model

$$Y_i = X_i^T \tilde{\beta} + E_i, i = 1,2,3,...n \tag{2}$$

where

X_i express design matrix sized

$\tilde{\beta}$ express vector regression coefficients of unknown sized $(p \times 1)$

Y_i vector responses related to the value of X_i sized $(q \times 1)$

E_i is a random error that cannot be observed sized $(q \times 1)$

Parameter $\tilde{\beta}$ is usually estimated by using OLS, but the estimator obtained is very influenced by the presence of outliers, then this section will discuss the parameter estimator in a multivariate response surface model is robust to outliers. Robust estimator of regression coefficient in a multivariate response problem is an important issue. Treat each response separately and applying a single robust response procedure may cause erroneous interpretations of results. So, it is necessary to take into consideration all responses simultaneously and estimate the variance-covariance matrix.

The difference between the multivariate response surface and the univariate response surface is a measure of the distance involved. In response to the problem of single-use Euclidean distance residue while multivariate response problems using Mahalanobis distance which takes into consideration the correlation between response. In the approach of MRSM, lower weights gave to residual size greater distance. In each iteration, the proposed weighting functions down-weights residuals by considering all the responses simultaneously.

One method to estimate the parameters in robust regression is to use the M-estimation. Parameter estimator $\tilde{\beta}$ is obtained by minimizing the function $\sum_{i=1}^{n} \rho(E_i)$ or

$$\sum_{i=1}^{n} \rho\left(Y_i - X_i^T \tilde{\beta}\right) \tag{3}$$

to $\tilde{\beta}$ or

$$\hat{\tilde{\beta}} = \arg \min_{\tilde{\beta}} \sum_{i=1}^{n} \rho\left(Y_i - X_i^T \tilde{\beta}\right) \tag{4}$$

28 February 2024 01:41:34

that is

$$\frac{\partial \sum_{i=1}^{n} \rho(E_i)}{\partial \tilde{\beta}} = \frac{\partial \sum_{i=1}^{n} \rho\left(Y_i - X_i^T \tilde{\beta}\right)}{\partial \tilde{\beta}} = 0 \tag{5}$$

so obtained

$$\sum_{i=1}^{n} x_{ij} \psi\left(y_i - X_i^T \tilde{\beta}\right) = 0; j = 0,1,2,3,\ldots,p \tag{6}$$

with ρ is a convex function and $\psi = \rho', x_{ij}$ represent the value of observation *i-th* on the independent variable *j-th*, and $x_{i0} = 1$.

Suppose $E_i = Y_i - X_i^T \hat{\tilde{\beta}}$ and defined a function weighting $w_i = \frac{\psi(E_i)}{E_i} = \frac{\psi\left(Y_i - X_i^T \hat{\tilde{\beta}}\right)}{Y_i - X_i^T \hat{\tilde{\beta}}}$ to obtain $\psi\left(Y_i - X_i^T \hat{\tilde{\beta}}\right) = w_i \left(Y_i - X_i^T \hat{\tilde{\beta}}\right)$ equation (6) be

$$\sum_{i=1}^{n} x_{ij} w_i \left(Y_i - X_i^T \hat{\tilde{\beta}}\right) = 0, j = 0,1,2,3,\ldots,p \tag{7}$$

From equation (7) obtained

$$\begin{aligned} \sum_{i=1}^{n} x_{ij} w_i Y_i - \sum_{i=1}^{n} x_{ij} w_i X_i^T \hat{\tilde{\beta}} &= 0 \\ \sum_{i=1}^{n} x_{ij} w_i Y &= \sum_{i=1}^{n} x_{ij} w_i X_i^T \hat{\tilde{\beta}} \end{aligned} \tag{8}$$

Equation (8) to $j = 0,1,\ldots,p$ can be written:n matrix form as follows :

$$X^T W X \hat{\tilde{\beta}} = X^T W Y \tag{9}$$

so obtained

$$\hat{\tilde{\beta}} = \left(X^T W X\right)^{-1} X^T W Y \tag{10}$$

where W = diagonal matrix size $n \times n$ with the diagonal elements in the form of weight $w_1, w_2, \ldots, w_n$

If the function ψ is not linear, the equation solved by the iteration method. The method used is iteratively reweighted least squares with the procedure:

i. Choose the value of the initial estimator $\hat{\tilde{\beta}}^{(0)}$, so it can be calculated $w_i^{(0)} = \frac{\psi\left(Y_i - X_i^T \hat{\tilde{\beta}}^{(0)}\right)}{Y_i - X_i^T \hat{\tilde{\beta}}^{(0)}}$

ii. At each iteration t, calculate the residual $E_i^{(t-1)} = Y_i - X_i^T \hat{\tilde{\beta}}^{(t-1)}$ and weight $w_i^{(t-1)} = w\left(E_i^{(t-1)}\right)$ from the previous iteration

iii. Compute the weighted least squares estimator new $\hat{\tilde{\beta}}^{(t)} = \left(X^T W^{(t-1)} X\right)^{-1} X^T W^{(t-1)} Y$ where $W^{(t-1)}$ is the weight of the $(t-1)$ iteration and the X matrix size

28 February 2024 01:41:34

iv. Step *-ii*and *-iii* repeatedly until the parameter estimates obtained convergence, in other words, if $\left\| \hat{\tilde{\beta}}^{(t)} - \hat{\tilde{\beta}}^{(t-1)} \right\|$ is quite small.

There are several kinds of functions ρ, one of the functions ρ is commonly used Huber in Claudio A at. Al [2]. If the parameter estimates of robust regression model found using the Huber function, and so the weight of w can be found utilizing the following steps:

i. Define the variable r_{ik}, for residual associated with the observation to-*i* response to-*k* response with $r_{ik} = Y_{ik} - \hat{Y}_{ik}, i = 1,2,3,...,n; k = 1,2,3,...,q$

ii. Calculate the residual for each response Y_k, for the first time residual obtained by using the initiation of a response estimate $\hat{Y}_k$.

iii. Calculate scaled residuals (sr_{ik}) for each response, using

$$sr_{ik} = \frac{r_{ik} - \overline{r}_k}{s_{r_k}} \tag{11}$$

with $\overline{r}_k$ is the average residual samples and s_{r_k} standard deviation of residual samples. Because the factors that can be controlled are assumed constant (not random), the correlation between the response, then estimated using scaled residuals. The result of this estimation is used to obtain covariance matrices $\hat{\Sigma}$.

iv. Count of the Mahalanobis distance by using the next equation

$$d\left(r(i)\right) = \sqrt{\left(r(i)\right)^T \hat{\Sigma}^{-1} r(i)} \tag{12}$$

with $r(i) = [sr_{i1}, sr_{i2}, ..., sr_{iq}]; i = 1,2,3,...,n$ is scaled residual matrix response to the observation $i-th$ and response to $q-th$. As described earlier, that the difference between the surface and the surface of a single response multivariate response is a measure of the distance involved. In response to the problem of single-use Euclidean distance while the residual problems of multivariate response using Mahalanobis distance which takes into consideration the correlation between response. In the approach of MRSM, lower weights gave to residual size greater distance. In each iteration, the proposed weighting functions down-weights residuals by considering all the responses simultaneously.

v. Determining the weights. To determine the weights w used the following stages:

a. Consider the Huber function

$$\rho\left(d\left(r(i)\right)\right) = \begin{cases} \frac{1}{2}\left(d\left(r(i)\right)\right)^2 & if \quad \left|d\left(r(i)\right)\right| \le c \\ c\left|d\left(r(i)\right)\right| - \frac{1}{2}c^2 & if \quad \left|d\left(r(i)\right)\right| > c \end{cases} \tag{13}$$

b. Derivatives of Huber function are:

$$\psi\left(d\left(r(i)\right)\right) = \begin{cases} -c & if \quad d\left(r(i)\right) < -c \\ d\left(r(i)\right) & if \quad \left|d\left(r(i)\right)\right| \le c \\ c & if \quad d\left(r(i)\right) > c \end{cases} \tag{14}$$

because $d\left(r(i)\right) \ge 0$, then

28 February 2024 01:41:34

$$\psi\left(d\left(r(i)\right)\right)=\begin{cases} d\left(r(i)\right) & if \quad d\left(r(i)\right)\le c \\ c & if \quad d\left(r(i)\right)> c \end{cases} \tag{15}$$

c. By using the definition of the weight of *Huber and Ronchetti [3]* then obtained $w_i=\dfrac{\psi\left(d\left(r(i)\right)\right)}{d\left(r(i)\right)}$ so

$$w_i=\begin{cases} 1 & if \quad d\left(r(i)\right)\le c \\ \dfrac{c}{d\left(r(i)\right)} & if \quad d\left(r(i)\right)> c \end{cases} \tag{16}$$

According to Montgomery et al. [4], the distribution of Mahalanobis distance squared approached by Chi square with q degrees of freedom. The critical point of this distribution in the confidence level of α or $\left(\chi^2_{q,\alpha}\right)$ is used as assign the weight Widodo et al [8]. In other words, if the squared Mahalanobis distance is less than or equal to $\left(\chi^2_{q,\alpha}\right)$ are given a weight of 1, while the other weights derived from the proportion of the distance. So the equation (16) to

$$w_i=\begin{cases} 1 & if \quad (d\left(r(i)\right))^2\le \chi^2_{q,\alpha} \\ \dfrac{\chi^2_{q,\alpha}}{(d\left(r(i)\right))^2} & other \end{cases} \tag{17}$$

vi. Estimates the models using w obtained in step (v)

vii. Step ii-vi to be repeated until the parameter estimates obtained convergence, in other words, if $\left\|\hat{\hat{\beta}}^{(t)}-\hat{\hat{\beta}}^{(t-1)}\right\|$ is quite small

CASE STUDY

Case studies presented in this section are based on the experiments reported in Nam et al. [6], which conducts research on enhancement of the surface layer of aluminium alloy air by shot peening. Shot peening leads to the characteristic surface topography, increased dislocation density, the development of residual compression stresses, and the work surface hardened layer, which serves to inhibit the initial cracks and inhibit the growth of a fatigue crack.

Shot peening caused four major effects on the surface: surface dimpling, strain hardening, structural changes, and residual stress. Strain hardening and residual stress are an important factor in fatigue strength. Strain hardening can enhance the effects of plastic deformation, which results in slowing the crack propagation. At the same time, the residual stress will distribute stress crack closure, reducing the driving force for crack propagation, and is effective in delaying the start of a fatigue crack.

Experiments performed on aluminium alloy Al 2124-T851 are used extensively in the aerospace industry because of its static high specific strength. Test plate used in this study has the dimensions of $(20\times 40\times 150)$ *mm*. Fig. 2 shows a schematic of the experimental facility shot peening. Shot peening machines are installed in the space and storage tanks containing a ball shot and compressed air. Shot peening machine consisting of pike and body shot peening nozzle. Shot the ball storage tank is connected to the body shot peening with flexible hose and air mixture that is included in the body through a tube shot peening. Specimens for shot peening treatment is placed on the table and the position of the specimen adjusted using verification tools and supply blocks.

28 February 2024 01:41:34

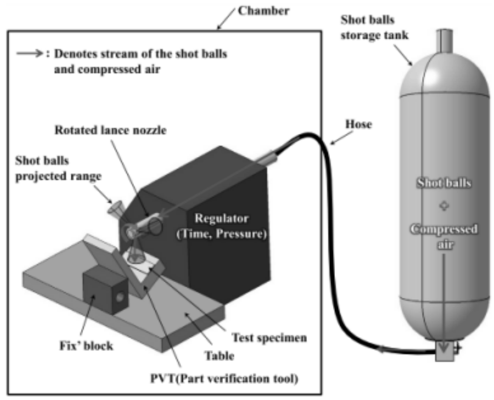

FIGURE 2. Schematic Diagram of The Shot Peening Facilities. Source *Nam et al. [6]*

The parameters used in the experiments are the exposure time shot peening, nozzle spacing, type of shot, shot size, pressure (flow rate, speed, speed), the size of the nozzle and nozzle angle. Nozzle size, the type of shot and the shot size predetermined is 12 *mm*, ASH-230 cast steel, and from 0.56 to 0.70*mm*. To ensure the minimum coverage level, nozzle spacing should be less than 150*mm*, the pressure is more than $2kg / cm^2$, the impact angle is greater than 30^0, and the exposure time greater than 60 seconds. If the distance of the nozzle is less than 50*mm,* the pressure is greater than $4kg / cm^2$, the angle of impact is greater than 60^0, the exposure time longer than 180 seconds, there will be cracks. Distance nozzle (ξ_1), pressure (ξ_2), the impact angle (ξ_3), and the exposure time (ξ_4) investigated parameters.

Box-Behnken design is used for fitting a second order response surface models. However, due to the limited resources of the facility, only 10 runs are possible per day. A total of 30 runs was divided into three days, with 10 runs (including the central points) were performed per day.

Table 1 shows orthogonal blocking of the Box-Behnken design for four parameters in three blocks with matrix design and experimental results. Blocking Orthogonal is a desired property when the tribulation should be arranged in blocks because of special experiments that cannot be run under homogeneous conditions. Because blocking orthogonal design, a second order model can be expanded to include the effects of the block and second order models can be estimated without being influenced by the effect block.

Furthermore, to change the level of the actual parameters to the code level is used the equation $x_i = \frac{\xi_i - \xi_{i0}}{\Delta\xi_i}$ with x_i is a variable code or standard, ξ_i is a variable uncoded, ξ_{i0} is the value of uncoded at a central point in the interval investigated, and $\Delta\xi_i$ is a step change from encoded variable according to the code values 0 and 1. In this study, experiments were conducted at random to reduce errors arising from the experimental process. Furthermore, the relationship between the response variable y (Microhardness (y_1) or Residual Stress (y_2)) with the corresponding parameters stated in a second order polynomial equation as follows:

$$y = \beta_0 + \sum_{j=1}^{4} \beta_j x_j + \sum_{j=1}^{4} \beta_{jj} x_j^2 + \sum_{j=1}^{4} \sum_{j<j'} \beta_{jj'} x_j x_{j'} + \varepsilon \tag{18}$$

where y is the value of the response observed, x_j and $x_{j'}$ is the code values, β_0 are constants, and β_j, β_{jj} and $\beta_{jj'}$, respectively coefficient of linear, quadratic and interaction.

28 February 2024 01:41:34

TABLE 1. Table Matrix Design and Data Box-Behnken experimental results shot peening. Source Nam et al. [6]

Std Order	Run Order	Pt Type	Blocks	Distance (mm)	Pressure (kg/cm3)	Angle (^{0})	Time (s)	Microhardness (Hv)	Residual stress (MPa)
1	14	2	1	50	2	45	120	170	-285.2
2	11	2	1	150	2	45	120	161.6	-228.5
3	13	2	1	50	4	45	120	173	-299.2
4	19	2	1	150	4	45	120	167.4	-286
5	18	2	1	100	3	30	60	161.8	-222.7
6	16	2	1	100	3	60	60	170.2	-289.2
7	15	2	1	100	3	30	180	164	-265.6
8	20	2	1	100	3	60	180	173.4	-290.1
9	17	0	1	100	3	45	120	168	-281.2
10	12	0	1	100	3	45	120	170.2	-270.9
11	22	2	2	50	3	45	60	170	-299.9
12	24	2	2	150	3	45	60	163	-254.6
13	28	2	2	50	3	45	180	174.6	-294.8
14	30	2	2	150	3	45	180	166.8	-273.2
15	26	2	2	100	2	30	120	162	-217.9
16	21	2	2	100	4	30	120	166	-275.4
17	23	2	2	100	2	60	120	166.8	-276.8
18	29	2	2	100	4	60	120	174.8	-319.8
19	27	0	2	100	3	45	120	168.2	-280.5
20	25	0	2	100	3	45	120	169.4	-297.2
21	7	2	3	50	3	30	120	171.3	-284
22	5	2	3	150	3	30	120	160.6	-218.9
23	1	2	3	50	3	60	120	174	-300.7
24	2	2	3	150	3	60	120	167.8	-280.2
25	3	2	3	100	2	45	60	162	-231.9
26	10	2	3	100	4	45	60	167.4	-283.2
27	8	2	3	100	2	45	180	163.6	-254.5
28	9	2	3	100	4	45	180	171.8	-299.1
29	4	0	3	100	3	45	120	168.6	-283.2
30	6	0	3	100	3	45	120	171	-279.6

As mentioned at the beginning that the focus of this paper is to develop a response surface model, then this section the researcher will restrict the analysis only on the model parameter estimation. Based on experimental data in Table 1, estimate the parameters of the model using the OLS method, as has been done by Nam et al. [6]. The estimation results can be seen in Table 2. The model has been used as a tool to determine the optimum conditions of the response, however, Nam et al. [6] has not conducted an analysis of the presence of outliers. As is well known that the OLS method is very sensitive to outliers. The existence of outliers can lead to errors in the estimation of the model, which in turn can provide a fitting conclusion to the determination of the optimal point of the response. Therefore, researchers interested in studying the outliers.

28 February 2024 01:41:34

TABLE 2. Table of Calculation Results of Regression Coefficients Microhardness (y_1) and Residual Stress (y_2) with OLS.

Variable	Coefficients	
	y_1	y_2
Constanta	169.233	282.1
x_1	-3.85	-18.533
x_2	2.867	22.325
x_3	3.4	22.692
x_4	1.65	7.983
x_1^2	0.275	-0.042
x_2^2	-1.5	-6.004
x_3^2	-0.65	-8.029
x_4^2	-1.225	-5.842
x_1x_2	0.7	10.875
x_1x_3	1.25	11.15
x_1x_4	-0.2	5.925
x_2x_3	1	-3.625
x_2x_4	0.7	-1.675
x_3x_4	0.25	-10.5

28 February 2024 01:41:34

To check for the presence of outliers The researchers used a standardized residual TS plot using a confidence level of 95% or cut off at ± 2. TS plot in Fig. 3 serving standardized residuals of the model y_1 and y_2. From pictures y_1 seen that there is one observation, the observation of 24th with a value of -2.31075 is out-of-bounds. Because of the observed values exceed the cutoff values that has been determined. The observations can be characterized as an outlier. Likewise for y_2, of the image can be seen that there are two observations are out-of-bounds, the observation of 12th and the observation 20th with a value respectively 2.06156 and 2.03343. Because of the observed values exceed the cutoff values that has been determined. The observations can be characterized as an outlier. As is generally known that the OLS method is very sensitive to outliers, so we need a method of estimating the model that is resistant to outliers.

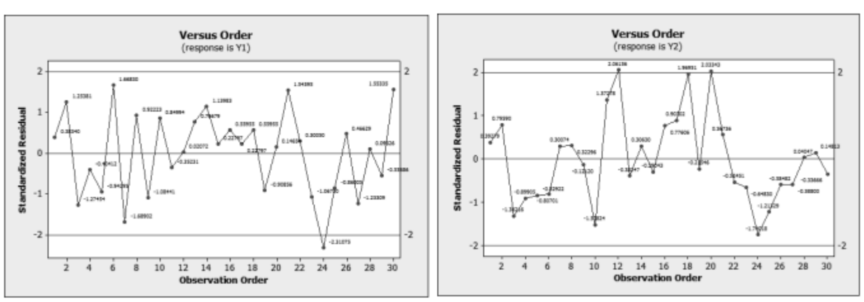

FIGURE 3. TS Plots standardized residuals of y_1 and y_2 with OLS.

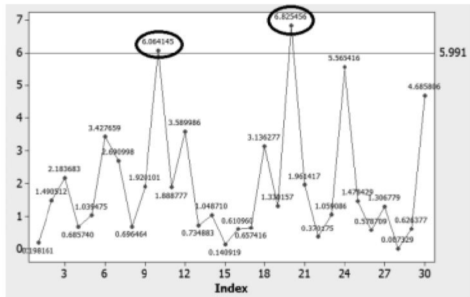

FIGURE 4. TS Plot Mahalanobis Distance Squared

Meanwhile, in the development of the model, Nam et al. [6] yet to consider the relationship between the response variable. In theory, it is known that there is a strong relationship between the variables Microhardness(y_1) or Residual Stress (y_2). The relationship can be described as follows Strain hardening can increase the effects of plastic deformation, which results in slowing the crack propagation. At the same time, compression residual stress will distribute stress crack closure. Thereby reducing the driving force for crack propagation, and can effectively delay the time fatigue crack. This information is supported by the calculation of correlation between two variables, namely 0.902. The existence of a strong relationship between the variables Microhardness (y_1) or Residual Stress (y_2) as well as the magnitude of the correlation between two variablemodellingesponses encourage researchers to perform modeling using the multivariate response surface model. Next is checking the existence of outlier together. To investigate the presence of outliers in multivariate, researchers use of TS plot Mahalanobis distance squared with 95% or cut off in 5.991, as presented in the Fig. 4. The figures show that there is two observation points are out-of-bounds. The observation of the 10th and 20th with Mahalanobis distance squared values, respectively 6.06415 and 6.82546. Because of the observed values exceed the predetermined cut off, then both of these observations are outliers. As is well known that the OLS method is very sensitive to outliers, so we need a method of estimating the model that is resistant to outliers. Based on the arguments above, in this section we propose estimating multivariate response surface model by using one of the methods of estimation that is resistant to outliers M-estimation method.

Estimates the multivariate response surface model begins with an initial estimate for the model using OLS. The results of calculations are presented in the table in table 2. Having obtained the initial model estimation followed by calculation of residual and residual scaled using equation (11) for each response. The results of these calculations, then used as the basis for calculating the estimated variance covariance matrix $\hat{\sigma}$ and Mahalanobis distance squared using equation (12). The next step is to determine the tuning constant, tuning constant used in this study was $\chi^2_{2,0.95}$ or 5.991. This value is used as the basis for giving weight to each observation. Observe, in this case the distance Mahalanobis squares, will be given a score of 1 if the value of Mahalanobis distance squared is less than or equal to the value of tuning constant, otherwise if the value of Mahalanobis distance is more than the value of tuning constant will be given a weight of less than one, as presented in equation (17). The result of the calculation of the residual, scaled residual (sr), Mahalanobis distance squared (dr2) and the initial weighting is presented in the table 3. The table shows that there are two observations that have squared Mahalanobis distance is more than tuning constant, the observation of the 10th and 20th observation. As a result of both of these observations during the weighting given weight of less than 1, respectively 0.988 for the observation of 10th and 0.8778 for the observation of the 20th.

The next step is to estimate the models using weights that have been obtained. New estimation results are then compared with the results of previous estimates. If the result of the new estimates provides significant changes from previous estimates, the calculation is repeated. The calculation is repeated from the residual calculation as described above. In this case, researchers used a norm to determine changes in the new parameter estimates with previously. This study used a limit of less than or equal to 10^{-13}, which means that the iteration process stops if a change of model parameters y_1 and y_2 less than or equal to 10^{-13}. The final results of simulation are presented in the Table 4.

28 February 2024 01:41:34

TABLE 3. Table of Residual, Scaled Residual (sr), Mahalanobis Distance Squared (dr2) and The Initial Weighting

Observation	Residual y_1	Residual y_2	sr1	sr2	dr2	w1
1	0.30833	2.0625	0.34411	0.35254	0.19816	1
2	1.00833	4.1792	1.12533	0.71434	1.49051	1
3	-1.025	-6.8375	1.14393	1.16872	2.18368	1
4	-0.325	-4.7208	0.36271	0.80693	0.68574	1
5	0.75833	-4.3542	0.84632	0.74425	1.03947	1
6	1.34167	-4.2375	1.49734	0.72431	3.42766	1
7	1.35833	1.5792	1.51594	0.26992	2.691	1
8	0.74167	1.6958	0.82772	0.28987	0.69646	1
9	1.23333	-0.9	1.37644	0.15384	1.9201	1
10	0.96667	-11.2	1.07883	-1.9144	6.06415	0.988
11	0.28333	7.2083	0.31621	1.23211	1.88878	1
12	0.01667	10.825	0.0186	1.8503	3.58999	1
13	0.61667	-2.0083	0.68822	0.34328	0.73488	1
14	0.91667	1.6083	1.02303	0.27491	1.04871	1
15	0.18333	-1.525	0.20461	0.26067	0.14092	1
16	0.45	4.075	0.50221	0.69653	0.61096	1
17	0.18333	4.7417	0.20461	0.81049	0.65742	1
18	0.45	10.3417	0.50221	1.76769	3.13628	1
19	1.03333	-1.6	1.15323	0.27349	1.33016	1
20	0.16667	15.1	0.18601	2.58102	6.82546	0.8778
21	1.24167	2.9792	1.38574	0.50923	1.96142	1
22	0.24167	-2.7542	0.26971	0.47077	0.37017	1
23	-0.85833	-3.4042	0.95793	0.58187	1.05909	1
24	1.85833	-9.1375	2.07396	1.56186	5.56542	1
25	-0.69167	-6.3708	0.77192	1.08896	1.47843	1
26	0.375	-3.0708	0.41851	0.52489	0.57871	1
27	0.99167	-3.0875	1.10673	0.52774	1.30678	1
28	0.075	0.2125	0.0837	0.03632	0.00733	1
29	0.63333	1.1	0.70682	0.18802	0.62638	1
30	1.76667	-2.5	1.97166	0.42732	4.68581	1

Table 4 presents a comparison table of results of OLS estimates and M-estimation for multivariate response surface model. It was clear from the table that the presence of outliers affects the estimation results. The first effect can be seen from the changes in the estimated value of the parameter when used estimation methods-M when compared with the OLS method, it can be seen from some of the values given mark boxes or thick circle. Substantial changes occur in the model y_2. The influence of the latter can be seen in the value of MSE. The M-estimation method provides the MSE is smaller when compared with the OLS method. This indicates that the M-estimation method for multivariate response surface models can provide better results when compared with the OLS method in terms of data contains outliers.

28 February 2024 01:41:34

TABLE 4. Comparison Table of Results of OLS Estimates and M-estimation of Multivariate ResponseSurface Model

Variable	y_1			y_2		
	OLS	**M-estimation**	**(OLS -(M-Estimation))**	**OLS**	**M-estimation**	**(OLS -(M-Estimation))**
Constanta	169.233	169.229	0.0045182	282.1	281.691	0.40935
x_1	-3.85	-3.85	0	-18.533	-18.533	0
x_2	2.867	2.867	0	22.325	22.325	0
x_3	3.4	3.4	0	22.692	22.692	0
x_4	1.65	1.65	0	7.983	7.983	0
x_1^2	0.275	0.277	-0.0022591	-0.042	0.163	-0.20467
x_2^2	-1.5	-1.498	-0.0022591	-6.004	-5.799	-0.20467
x_3^2	-0.65	-0.648	-0.0022591	-8.029	-7.824	-0.20467
x_4^2	-1.225	-1.223	-0.0022591	-5.842	-5.637	-0.20457
x_1x_2	0.7	0.7	0	10.875	10.875	0
x_1x_3	1.25	1.25	0	11.15	1.15	0
x_1x_4	-0.2	-0.2	0	5.925	5.925	0
x_2x_3	1	1	0	-3.625	-3.625	0
x_2x_4	0.7	0.7	0	-1.675	-1.675	0
x_3x_4	0.25	0.25	0	-10.5	-10.5	0
MSE	1.552	1.552	0.0003	66.172	63.7	2.4725

28 February 2024 01:41:34

CONCLUSION

From the discussions, it can be concluded that the M-estimation method for multivariate response surface model with the data outliers (MRSM) is

$$\hat{\hat{\beta}}^{(t)} = \left(X^T W^{(t-1)} X\right)^{-1} X^T W^{(t-1)} Y$$

where W = diagonal matrix of size $n \times n$ with its diagonal elements in the form of weight $w_1, w_2, \ldots, w_n$

$$w_i^{(t-1)} = \begin{cases} 1 & if \quad d\left(r\left(i\right)\right) < \chi^2_{q,\alpha} \\ \dfrac{\chi^2_{q,\alpha}}{\sum_{i=1}^{l} d\left(r\left(i\right)\right)} & other \end{cases}$$

From the case study on the enhancement of the surface layer of aluminium alloy air by shot peening, could be seen that MRSM have a better efficiency than the robust procedures individually and OLS. By using MRSM possibility of the influence of outliers can be minimized, so as to minimize the chances of error in determining the optimum point as a result of the model is not correct.

REFERENCES

1. G. E. P. Box and K. B. Wilson, Journal of the Royal Statistical Society Series B**13**, 1-45 (1951).
2. A. Azzalini, "Flexible Distributions as an Approach to Robustness: The Skew-t Case," in *Recent Advances in Robust Statistics: Theory and Applications*, edited by A. Claudio*et al*.(Springer India, New Delhi, 2016), pp. 1-15.
3. P. J. Huber and E. M. Ronchetti, *Robust statistics*(John Wiley & Sons, New Jersey, 2009), pp. 1-6.
4. D. C. Montgomery, *Introduction to Statistical Quality Control*(John Wiley & Sons, New York, 2009),pp. 486-520.
5. R. H. Myers, D. C. Montgomery and C. M. Anderson-Cook, *Response surface Methodology, Process and Product Optimization Using Designed Experiments(Wiley series in probability and statistics)*(John Wiley & Sons, New Jersey, 2016), pp. 40-146.
6. Y. S. Nam, Y. I. Jeong, B. C. Shin and J. H. Byun, Materials & Design **83**, 566-576 (2015).
7. W. Wan, "Semi-parametric Techniques for multi-response Optimization,"Ph. D Dissertation, Virginia Polytechnic Institute & State University, 2007.
8. E. Widodo, S.Guritno and S. Haryatmi, "Estimation of Multivariate Response Surface Model with Data Outliers,"Dissertation, GadjahMada University, 2015.

28 February 2024 01:41:34