JOURNAL OF ENGINEERING AND MANAGEMENT IN INDUSTRIAL SYSTEM VOL. 12 NO. 2 YEAR 2024

e-ISSN 2477-6025
https://doi.org/10.21776/ub.jemis.2024.012.02.3

A FRAMEWORK FOR OBSERVATIONAL DATA-BASED RESPONSE SURFACE METHODOLOGY

Mochammad Arbi Hadiyat[1)*], Bertha Maya Sopha[2)], Budhi Sholeh Wibowo[3)]
Industrial Engineering program, Universitas Surabaya (Ubaya), Surabaya[1)]
Industrial Engineering program, Department of Mechanical and Industrial Engineering, Universitas Gadjah Mada, Yogyakarta[2)]

Abstract Response Surface Methodology (RSM) is an integrated tool for optimization purposes based on an experiment. The designed experiment ensures the researcher fully controls all factors that potentially influence the response and simultaneously fulfills the orthogonal assumption among factors. On the other side, conducting DoE for a continuous production process raises difficulties since it should be interrupted during experiment runs. Meanwhile, the rapid development of production data acquisition systems provides stored records or observational data with potentially useful information for supporting process optimization. This paper proposes an alternative framework for adopting observational data for RSM analysis to reduce the need for real experimentation and ongoing production disruption. Referring to three stages of classic RSM and adopting the instance selection concept in the data mining context, the proposed framework aimed to achieve an observational data condition similar to an orthogonal D-optimal DoE based on criteria of Variance Inflation Factor (VIF) and determinant of matrix containing factor levels. It starts by applying a genetic algorithm for iteratively selecting an orthogonal subset of observational data and generating new actual experiment points to satisfy an orthogonality criterion. Then, a linear RSM model is fitted and continued by adding new experiment points. Then a standard numerical optimization method is applied to search among factor levels that optimize the response. A simulated data-based case study was taken in this paper, aiming to maximize the response of a production process with some pre-determined factors. The proposed framework has been implemented successfully, orthogonality of the data subset is achieved, and an optimal solution is found. Both criteria show acceptable results and raise some improvement opportunities.

Keywords: Response surface methodology; Observational data; Orthogonality; Optimization

1. Introduction

Response surface methodology (RSM) develops the design of experiment (DoE) for optimization purposes by involving it with additional tools. Using the basic concept of DoE for data acquiring, the RSM fits a mathematical model to fit the data and continues by performing optimization referring to the model. Therefore, the RSM is actually an integrated and sequential analysis of three tools, i.e., the designed experiment (DoE), mathematical modeling, and optimization technique [1]. For more than ten decades since first introduced by [2] in 1951, the RSM has played an essential role in optimizing kinds of processes in industrial or laboratory scope. Moreover, various research fields involving optimization also apply RSM, and over 48.000 SCOPUS-indexed papers employ this methodology.

RSM works by first experimenting; the researchers must define potential influencing factors to the response, determine each factor level, and accommodate them into a designed experiment (DoE). Once experiment data is obtained, a linear mathematical model that captures causality between factor and response is fitted and evaluated by performing statistical inference and at the same time, correcting the model until specific criteria are reached. The final fitted model becomes the reference in searching for an optimal factor level setting that optimizes the response, and then the final aim of RSM is obtained. Full explanations of RSM have been published in the primary RSM reference, such as in [3] and [4].

The ideal implementation of RSM is suitable for laboratory experimentation scope, where almost all factors are fully controlled to minimize experimental noises. As mentioned by [5] in an editorial, if someone needs to study the influence of a factor in a type of process, then the engineer should change the setting/level of it based on a designed experiment; this statement will work for such a laboratory-based experiment. However, this approach will not be

* Corresponding author. Email : arbi@staff.ubaya.ac.id
Published online at http://Jemis.ub.ac.id

Cite this Article As Hadiyat, et al. (2024). A Framework For Observational Data-Based Response Surface Methodology. Journal of Engineering and Management in Industrial System, 12(2), p87-97
Paper Submitted : Aug, 13th 2024
Paper Published : Sept, 16th 2024

e-ISSN 2477-6025
https://doi.org/10.21776/ub.jemis.2024.012.02.3

fully accepted for a type of continuous process or production. As the RSM needs to conduct a designed-experiment, then such an already running process should be stopped in order to change the setting to accommodate the designed experiment, and it may produce some waste and raise additional costs (see [6] and [7]). Even if scheduled maintenance pauses the process, changes to a fixed factor setting is not always acceptable since it will revise a long-term tacit knowledge they believe.

On the other side, the development of a data recording system that is installed on such a continuous process provides a large dataset involving the characteristics of the process and product; some examples of manufacturing data acquisition are provided in [8] who records data using Human-Machine Interface dan Machine Execution System. Instead of experimenting, some research proposed to use observational or historical datasets as an alternative. They believe that the dataset contains useful information that can lead to optimization purposes, as explained by [9] with implementation for semiconductor manufacturing. Some methodological development of observational data-based RSM (RSM-OD) started by [10] where the framework treats the observational data to have similar characteristics as factorial design. Another approach also proposed by [6] and [7], where DoE matching procedure produces a subset by finding a Taguchi orthogonal array within the dataset and then performing analysis as if it is a DoE. A similar approach also implemented by [11], an extensive dataset from a continuous process is investigated for some shifting, and a DoE subset match is then performed to find a potential orthogonal design within the data. A different RSM-OD approach was also proposed in [12] and [13]; all the observational data becomes RSM model input instead of finding a subset from the dataset. Actually, this approach is different from real-time data-driven predictive modeling proposed by [14] and [15]; the RSM-OD emphasizes the final aim of optimization while that approach focuses on real-time prediction purposes.

The aim of this research is to propose an alternative framework for adopting observational data in RSM-OD. An approach of instance selection in the data mining concept (see [16]) becomes a reference in selecting an observation subset that similarly fulfills orthogonal criteria as if it were a designed experiment. Once an orthogonal observation is found, the standard procedure of RSM analysis will be suitable for finding the optimal level/setting to optimize the response. Additionally, as there will be difficulties in reaching such a perfect orthogonal observations subset, a new procedure is also proposed to provide additional new experiment points to increase the orthogonality of the founded subset. This paper structure starts with conveying strong rationales for adopting observational data for RSM analysis, followed by delivering the concept of classic RSM completed by some examples of successful RSM-OD implementations. A proposed modified RSM to accommodate observational data is then explained in detail, including some pseudo algorithms to execute the framework. A case study for implementing the proposed framework is selected to perform an evaluation and comparison to classic RSM analysis.

2. Involving observational data for RSM analysis

As mentioned by [1], classic RSM consists of three sequential tools, i.e., the DoE, mathematical modeling, and optimization technique. The DoE stage ensures that all factors involved are in orthogonal conditions so that the effect of each factor is independent among them. As in standard regression analysis (see [17]), inter factors independence allows the researchers to study the influence of factors individually without being affected by other factors, except for some predefined interactions between two or more factors.

Once the factors are independent, the modeling stage will follow its ideality to fit a type of linear model, and there will be no concerns about the presence of multicollinearity that violates the statistical assumptions in the inference. Moreover, a straightforward interpretation of the model will perform well, involving the effects of each factor and interactions and their significance. Standard criteria for evaluating the model's goodness of fit were also calculated, for example, the coefficient of determination (R^2) and the MSE. A linear polynomial model containing second-order quadratic and interaction components is preferred to get an optimum point on its response surface. When there are no such optimum solutions in the model, then a procedure of steepest ascent will direct to move the factor

e-ISSN 2477-6025
https://doi.org/10.21776/ub.jemis.2024.012.02.3

levels at certain experimentation area until the optimum is reached (see [3] and [4]).

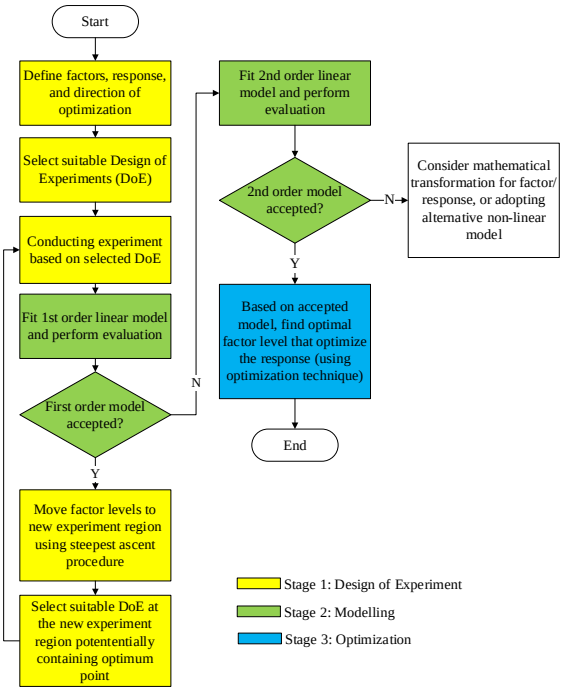

Fig 1. Classic RSM analysis procedure, adopted from [1]

The complete procedure in classic RSM analysis shown in Fig 1 consists of three stages, i.e., design of experiment, modeling, and optimization. Briefly, the procedures start with designing the experiment that meets orthogonality among factors. A first-order linear model as equation (1) should not be suitable to model the data because of the lack of an optimum point, whereas the experiment region should be ensured to contain such a point. A suitable second-order model as equation (2) in the region will lead the researcher to find optimal factor levels. However, in specific conditions with operational constraints, this model can sometimes not be fitted, and the optimization is forced to find the optimum solution using ridge analysis (see [4]). For an ideal second-order model, a standard mathematical optimization will provide an optimal solution; additionally, some numerical approach is also considered as an alternative, such as the metaheuristics method (see [18])

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \varepsilon \quad (1)$$

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \sum_{j=1}^{k} \beta_{ij} x_i x_j + \varepsilon \quad (2)$$

As mentioned in section 1, some rationales allowed the researcher to use observational/historical data for RSM analysis for optimization purposes instead of conducting designed experiments. Referring to *Fig 1*, the stage of DoE is modified to accommodate such observational data. Some selected references on adopting observational data for RSM analysis are shown in ***Table 1***.

Table 1. Selected references for RSM modifications in adopting observational data

Modification of RSM stage			**References**
Stage 1 (replace the DoE)	**Stage 2 (modelling)**	**Stage 3 (optimization)**	
Subset of observations as RSM input	Linear model	Local serch	[11], [7], [9]
Subset of observations as RSM input	Machine learning	Metaheuristics	[19], [20]
All observations as RSM input	Machine learning	Metaheuristics	[21], [22]
All observations as RSM input	Linear model	Local serch	[13], [23]
All observations as RSM input	Linear model	Metaheuristics	[12], [24]

e-ISSN 2477-6025
https://doi.org/10.21776/ub.jemis.2024.012.02.3

These modifications of RSM provide various approaches in each stage of RSM. For stage 1, the main problem is treating the observational data as if it is a designed experiment, using all observations or a subset of them. Some evaluation criteria are applied, such as orthogonality criteria and outlier detection, as explained in [25] using happenstance data. In stage 2, as the observational data is adopted, the standard linear model often fails to fulfill some assumptions during the statistical inference. Therefore, some non-linear machine-learning model is considered, such as neural networks; [23] has implemented this method for a pollutant removal process. The consequence is modifying the optimization technique with the numerical approach to increase optimal point search, such as metaheuristics. All these modifications are considered to develop a new approach as proposed in this paper.

3. Method

In order to answer this research's aim, sequential steps direct the research activities in an integrated research methodology, as shown in **Fig 2**. The final output of implementing these steps leads to the proposed framework in adopting observational data for RSM analysis. A systematic literature review on this topic has been published in [26]; the gaps are obtained with a focus on evaluating and adopting observational data. As the data will contain high factor level variation and violate the concept of DoE orthogonality, a criterion of data evaluation is then proposed. Based on these criteria, a procedure to treat the observational data is then developed to reach orthogonality among factors. As the orthogonality criteria from the data will not be perfectly obtained, additional steps are developed to increase it. Finally, a case study will illustrate the practical implementation of the proposed RSM-OD framework completed by its performance evaluation.

4. The proposed RSM-OD framework

This section is divided into the following steps in research methodology as in **Fig 2**. Modification from classic RSM, as in **Table 1**, becomes the bases for developing the proposed RSM-OD framework.

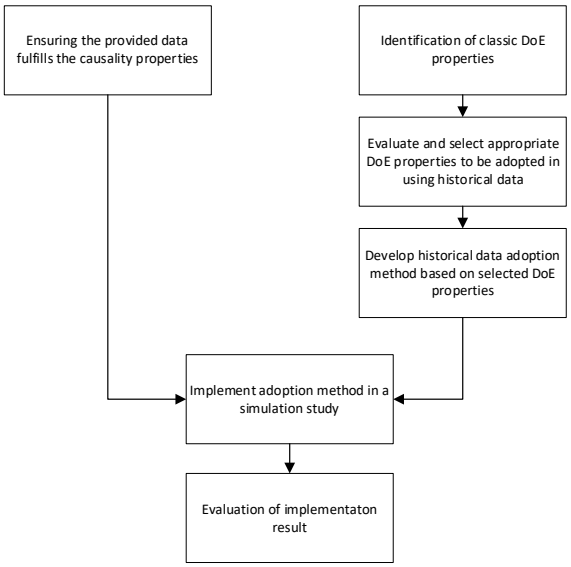

Fig 2. Research methodology

4.1. Orthogonality criteria

The concept of orthogonality in DoE ensures the independence of involved factors in RSM analysis. Therefore, to accept observational data as if it is similar to a designed experiment, the orthogonality criteria are adopted in developing this proposed framework. Following the standard RSM model in [4], an ordinary linear RSM model forms equation (3).

$$\boldsymbol{y} = \boldsymbol{X\beta} + \varepsilon \quad (3)$$

$$\boldsymbol{\beta} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y} \quad (4)$$

where **y** represents the response or dependent variable, **X** contains factor level or independent variable, and **β** represents the regression coefficient estimated using the least-squares method as in (4). The **X** matrix should be orthogonally arranged by implementing the DoE; the typical evaluation of this condition is calculating the determinant of $(\mathbf{X'X})^{-1}$ that should be minimized (or maximizing the determinant of (**X'X**)) to reach orthogonality. Some DoE designs follow these criteria to generate a designed-experiment matrix, such as D-optimal and A-optimal designs (see [27]). Another approach to evaluating orthogonal criteria in calculating the Variance Inflation Factor (VIF) appears in [17]; this criterion is commonly used in regression analysis to detect the existence of multicollinearity among factors, and it is the indication of non-orthogonality in the **X** matrix.

Having a perfect orthogonal matrix from all observational data is difficult since the

e-ISSN 2477-6025
https://doi.org/10.21776/ub.jemis.2024.012.02.3

researcher did not control each factor level as in the designed experiment. Therefore, the determinant of matrix $(\mathbf{X'X})^{-1}$ and VIF become alternatives to evaluate the orthogonal condition of the data; the acceptable boundary of both criteria is shown in **Table 2**.

Table 2. evaluation of orthogonality

Criteria	Acceptable boundary for orthogonality	Reference
Determinant of $(\mathbf{X'X})^{-1}$	Minimized	[27], [3]
VIF	Less than 5	[28], [29]

4.2. Procedure to select a subset from observational data

According to references in **Table 1**, instead of using all observational data with less orthogonal conditions, a subset is selected to obtain sub-observation with higher orthogonality. Some examples are shown in [11] and [7] by manually selecting a subset that matches an ideal DoE, such as factorial and Taguchi design; an effort to obtain such a subset will increase as the observational data becomes large in number. A new approach in this paper considers **Table 2** criteria and, at the same time, selects a subset that fulfills both by optimizing the determinant, as in Equation (5). This subset should consider the sufficient degrees of freedom of the RSM model to accommodate the predefined term.

$$\begin{aligned} &\text{Maximize} && w_1 \det(\mathbf{X'X})^{-1} + w_2 \text{ VIF} \\ &\text{Subject to} && w_1 + w_2 = 1, \\ &\text{X} \in \text{available design space} && \end{aligned} \quad (5)$$

Once a subset is selected, the criteria are evaluated and then followed by updating the selected subset with other observations to get better ones. This optimization is done iteratively by adopting a binary genetic algorithm with both criteria as the fitness function. The complete pseudo-code for subset selection is shown in **Algorithm 1**. This subset selection process is similar to the procedure of instance selection (see [16]) in data mining applications; however, in this case, such a procedure is implemented in RSM analysis with modification of the fitness functions.

4.3. Procedure to increase orthogonality

When orthogonality in a subset is not satisfied by implementing Algorithm 1 (Fig 3), it means the factors are highly correlated with each other, and there are consequences in the RSM modeling stage, i.e., assumption violation in statistical inference. Following the concept of D-Optimal design, conducting the additional new experiment (a new factor level combination) should increase the level of orthogonality as long as the researcher can fully control the factor level at a specific region that maximizes the criteria. Therefore, a selected less-orthogonal subset will still be brought to the RSM modeling stage, considering a new experiment point generated based on the subset information. After the experiment point is generated and the researcher conducts it, an evaluation or orthogonality is taken. This step works iteratively; one-by-one new experiment point is generated until reaching satisfied orthogonality. A genetic algorithm also gives the capability to resolve this procedure, as shown in Algorithm 2 (Fig 4), by maximizing Equation (5).

```
DEFINE
    Data X : contains factors and levels,  N : represents number of all observation
    n : number of selected observations for the subset, n<N (considering RSM model degrees of freedom)
    specify RSM model term (linear, quadratic, interaction)
INITIALIZE
    Genetic algorithm properties (number of population, parents, offspring, mutation rate)
    Gene code 1: selected for subset, code 0: not selected for subset
    Generate initial population chromosomes represents selected observations for subset
    Involve the RSM model term in the subset
WHILE termination criteria is not satisfied
    SELECT parents chromosomes from population
    CROSSOVER pairs of parents chromosomes to produce offsprings
    GENERATE MUTATION chromosomes from population
    COMBINE the offsprings with mutated chromosomes
    EVALUATING the fitness function (refer to criteria in Table 2)
    SELECT best chromosomes for next parents generation
ENDWHILE
```

Fig. 3 Algorithm 1. Genetic algorithm pseudo code for subset selection

e-ISSN 2477-6025
https://doi.org/10.21776/ub.jemis.2024.012.02.3

DEFINE
Data **X :** selected subset**,** contains factors and levels
Boundary of new experiment point (level setting at each factor)
Convert the boundary (decimal) into binary chromosome
specify RSM model term (linear, quadratic, interaction)
INITIALIZE
Genetic algorithm properties (number of population, parents, offspring, mutation rate)
Generate initial population chromosomes represents factor levels (within the boundary)
Involve the RSM model term in the subset
WHILE termination criteria is not satisfied
SELECT parents chromosomes from population
CROSSOVER pairs of parents chromosomes to produce offsprings
GENERATE MUTATION chromosomes from population
COMBINE the offsprings with mutated chromosomes
CONVERT the chromosomes into decimal
EVALUATING the fitness function (refer to criteria in Table 2)
SELECT best chromosomes for next parents generation
ENDWHILE
GENERATED a new experiment point is the best chromosome

Fig. 4 *Algorithm* 2. Genetic algorithm pseudo code to generate a new experiment point

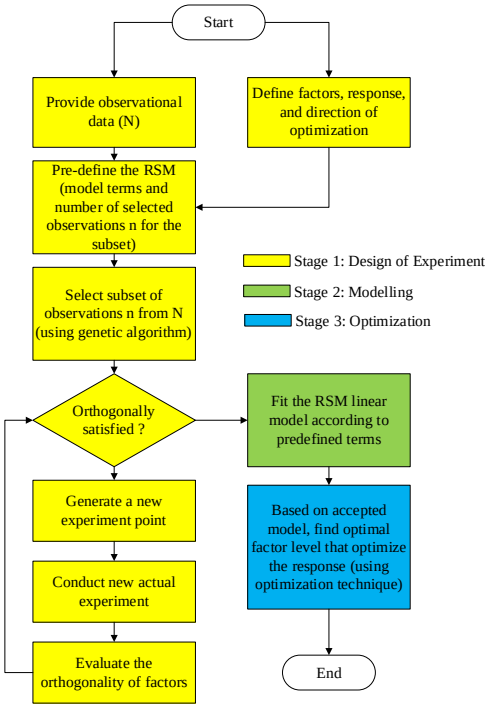

Fig 5. proposed RSM-OD framework

As Algorithm 2 finishes generating each new experiment point, the next step is conducting the actual experiment, referring to generated points (factor level setting). It will loop until the desired number of points meets a certain level of orthogonality. This procedure is similar to a steepest ascent in classic RSM (see Fig 1), replacing the optimization target with a level of orthogonality.

4.4. Complete RSM-OD framework

The proposed framework of RSM-OD in this paper is compiled according to the three stages of classic RSM. As mentioned before, the level combinations among factors in observational data are not designed as ideal DoE, the fitted RSM model will use the data as it is, and the procedure of moving the experiment region to one which contains optimal setting should not be conducted as in classic RSM.

Referring to Algorithm 1, the framework in Fig 5 is started by providing observational data and predefining the term involved in the RSM model. For huge data, there are possibilities to contain many features (as potential factors) and responses, and it is needed to select and define interesting ones for the RSM analysis.

e-ISSN 2477-6025
https://doi.org/10.21776/ub.jemis.2024.012.02.3

Predefined factors and model terms direct the subset selection process to find the best observation subset with the calculated level of orthogonality based on mentioned criteria above. Recommendation of new actual experiment point should be taken in case of low orthogonality in the subset. When the best subset that includes the new experiment is obtained, the standard RSM model accommodating predefined terms is then fitted, and finally, the optimization process works based on this model.

5. Simulation study and implementation result

A number of 100 observations dataset with two factors (X_1, X_2) and a response (Y) is generated for a case study; the generating process adopts a second-order regression model term, i.e., linear, quadratic, and interaction, as equation (5) with additional error component ε as if it were an actually observed causality dataset.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_3 X_1 X_2 + \varepsilon \quad (6)$$

In order to practice the proposed RSM-OD frameworks, a condition of multicollinearity between both factors is also involved, it means that this dataset represents the condition of non-orthogonal observational data needed to evaluate the framework. As mentioned in [17], non-orthogonal conditions in the dataset raise the possibility of biased inference; with the existence of multicollinearity, then the dataset cannot be further analyzed using RSM because of assumption violation. Therefore, the subsetting algorithm will produce an orthogonal one.

5.1. Selecting subset and generating new actual experiment point

A subset of 120 observations that fulfills orthogonality among factors is selected from 150 ones using **Algorithm 1**, considering less wasted observations and sufficiency of degrees of freedom in the linear model inference.

The main criteria to be minimized in this case study is total VIF from the terms, with the assumption that less value of VIF will provide more orthogonal conditions in the subset. With a maximum iteration of 2500, the convergence of the genetic algorithm is reached (Fig 6), and its shows that subset finding is successfully found.

Table 3. Initial setting of Algorithm 1

Predefined condition	Setting
Maximum number of iterations in genetic algorithm	2500
Number of populations generated in genetic algorithm	20
RSM term involved	Linear, quadratic, interaction (see equation 5)
Factors in RSM	X_1, X_2
Response in RSM	Y
Subset number	120
Dataset number	150

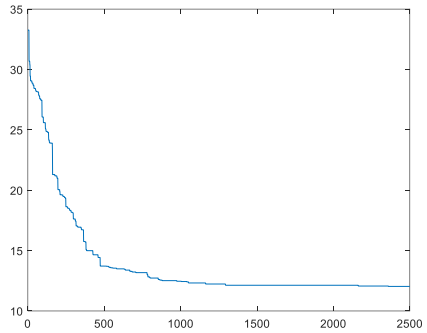

Fig 6. Convergence of Algorithm 1

Table 4. Comparation of complete dataset and the subset

Orthogonality criteria			Initial condition from complete dataset	Condition of subset
VIF	Linear term	X_1	5.49	3.06
		X_2	3.97	2.18
	Quadratic term	X_1^2	8.58	2.66
		X_2^2	6.04	1.62
	Interaction	X_1X_2	15.61	2.55
Total VIF from the terms			39.69	12.07
Determinant of $(X'X)^{-1}$			0.0173944	2.60342

The result of Algorithm 1 is summarized in Table 4; it shows that the subset with 120 observations has higher orthogonality rather than the original dataset. Moreover, the target in Table 2 is reached with all VIF<5, although the determinant of $(X'X)^{-1}$ failed to reduce. In this case, the orthogonal condition is fulfilled, indicating that

e-ISSN 2477-6025
https://doi.org/10.21776/ub.jemis.2024.012.02.3

the RSM model is ready to be fitted based on the subset.

However, to have a lower VIF in the subset, then Algorithm 2 will give recommendations for new actual experiment points that complete the observation and increase orthogonality. Only three new actual experiment points are generated, considering that the RSM-OD focuses on observational data and less on actual experimenting. Convergence of the algorithm is reached and shown in Fig 7 with less than 500 iterations, and the generated points are summarized in **Table 5**. Each point successfully reduces the total VIF and its corresponding model terms; the determinant of $(\mathbf{X'X})^{-1}$ is also decreasing.

5.2. Fitting RSM model and optimization

Since the subset with additional experiment points has fulfilled orthogonality, the RSM model starts to fit. Referring to the terms in equation (5), the fitted model is shown in equation (6) with ANOVA analysis for its inference (see Table 6). All the calculation and graph is performed by MINITAB ®.

$$Y = 79.838 + 3.298\,X_1 - 2.27\,X_2 - 3.005\,X_1^2 + 7.77\,X_2^2 - 1.43\,X_1X_2 + \varepsilon \quad (7)$$

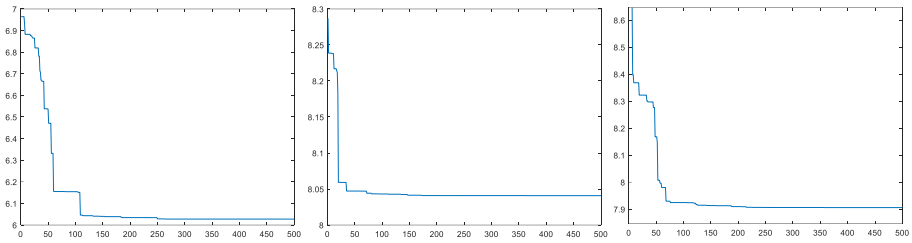

Fig 7. Convergence of Algorithm 2 for each of 3 additional point (cut at 500 iterations)

Table 5. Recommendation of new actual experiment point using **Algorithm 2**

Additional new actual experiment point (factor setting)				VIF for each term					Total VIF from the terms	Determinant of $(X'X)^{-1}$
The i-th point	X_1	X_2	Simulated experiment response Y	X_1	X_2	X_1^2	X_2^2	X_1X_2		
				Intial VIF and determinant (from **Table 4**)						
				3.06	2.18	2.66	1.62	2.55	12.07	2.603
1	-1.0000	0.9042	78.3348	1.8672	1.1646	2.6705	3.0477	2.4605	11.2105	0.323
2	1.0000	-0.1574	79.76588	2.064	1.1307	2.4701	2.8468	2.4394	10.951	1.123×10^{-7}
3	-0.8592	0.218	75.58634	1.4451	1.1188	1.9027	2.2231	2.4957	9.185	5.361×10^{-8}

Table 6. ANOVA for RSM model

Analysis of Variance					
Source	**DF**	**Adj SS**	**Adj MS**	**F-Value**	**P-Value**
Model	5	65.349	13.0698	13.18	0.000
Linear	2	59.793	29.8967	30.15	0.000
X1	1	50.116	50.1158	50.55	0.000
X2	1	5.566	5.5656	5.61	0.019
Square	2	15.595	7.7977	7.86	0.001
X1*X1	1	13.108	13.1075	13.22	0.000
X2*X2	1	3.965	3.9646	4	0.048
2-Way Interaction	1	0.387	0.3872	0.39	0.533
X1*X2	1	0.387	0.3872	0.39	0.533
Error	117	116.006	0.9915		
Total	122	181.355			

e-ISSN 2477-6025
https://doi.org/10.21776/ub.jemis.2024.012.02.3

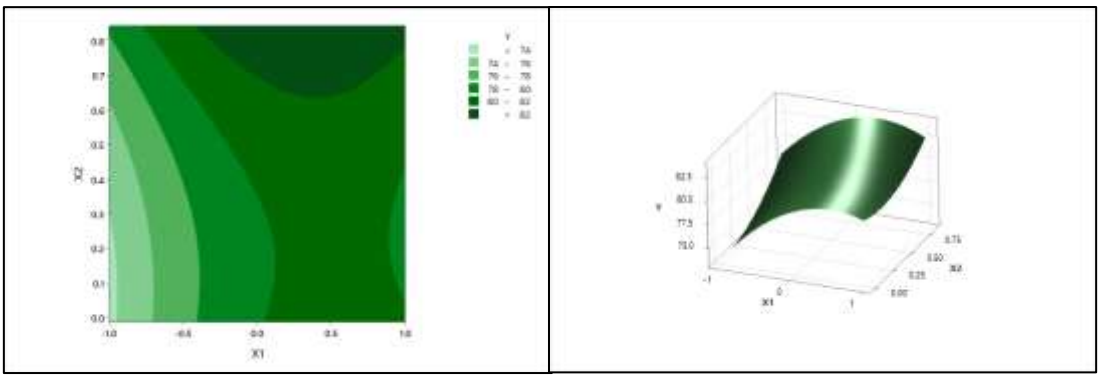

Fig 8. Response surface from RSM model in (6), processed by MINITAB®

The linear and quadratic inference significantly affects the model while the interaction does not. This fitted model becomes the bases for finding the optimal setting of X_1 and X_2 that maximizes the response Y in an optimization. Since the fitted RSM model adopts a commonly linear polynomial equation, then the optimization procedure uses the standard desirability function approach. MINITAB ® provides this optimization process by referring to [4], as follows

$$\text{Maximize } Y = 79.838 + 3.298\,X_1 - 2.27\,X_2 - 3.005\,X_1^2 + 7.77\,X_2^2 - 1.43\,X_1X_2$$
subject to
$X_1, X_2, \in$ available design space (8)

Solving the optimization of equation (8), Fig 9 shows the Optimal response by setting the X_1=0.3535 and X_2 = 0.8416, with the prediction of response Y=83.7991. This solution is the final aim of implementing the RSM-OD based on the subset that is selected from the dataset. Since the orthogonal between factors is reached, then the unbiased estimation of the RSM model in (6) provided meaningful interpretation.

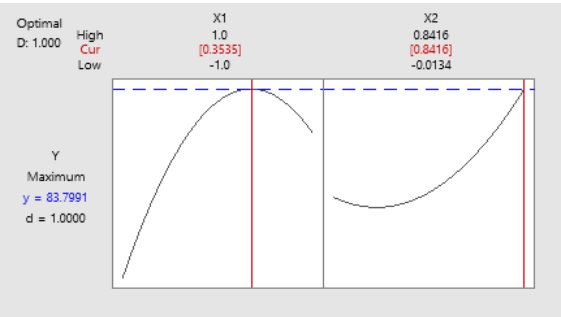

Fig 9. RSM optimization result

5.3. Evaluation of the proposed method

Successful implementation of the proposed framework in **Fig 3** for the case study gives the opportunity for the RSM-OD analysis to develop and be adopted as a formal optimization procedure based on the concept of classic RSM. The first weakness of this framework emphasizes the criteria in selecting observation from the dataset to become its subset with better orthogonality and, at the same time, minimize the determinant of $(X'X)^{-1}$ as if it is a D-optimal designed experiment. The VIF criteria successfully helped Algorithm 1 to find the subset, but the determinant remains increased. Alternatively, for the next improvement, both criteria should join to become a single measurable one for supporting the algorithm. Meanwhile, Algorithm 2 has a similar issue but gives better criteria achievement; the VIF and the determinant show decreased trends as each new point is generated. This result indicates potential effectiveness for the additional new experiment to reach orthogonality and opens the chance to develop such a better procedure. As initial research in developing the alternative framework for RSM-OD, some improvement for further investigation involves; (a) developing better criteria for selecting the subset, (b) comparison of criteria between actual observational data and simulated one, and (c) consideration to improve the RSM model and optimization technique to accommodate nonlinearity on the dataset with remains provide clear interpretation.

6. Conclusion

RSM-OD gives an alternative to classic RSM with the accommodation of observational data instead of experimenting. This approach is suitable for types of a continuous process where some interruptions for experimenting are preferred to avoid. With an installed data-acquiring system, recorded data should give potential optimization information. Unlike

e-ISSN 2477-6025
https://doi.org/10.21776/ub.jemis.2024.012.02.3

actual experiment data, the observational data provides non-ideal conditions, such as non-orthogonality and some outliers. Thus, a procedure to adopt such data in RSM becomes the focus of this paper

The proposed RSM-OD framework in this paper recommends selecting observations as a subset from the complete dataset with considering orthogonality criteria. Three stages in classic RSM remain the reference to developing such a procedure. A simulated data is generated as a case study for implementing the frameworks. A genetic algorithm helps to find the subset and has been proven to increase inter factors orthogonality. Moreover, to increase the orthogonality, a similar algorithm was also adopted in the framework by generating new actual experiment points to complete the selected subset.

The procedure successfully works, but some issues were raised during the RSM-OD analysis. The result shows that VIF criteria help the algorithm find the desired subset, but the calculated determinant $(\mathbf{X'X})^{-1}$ remains worse. On the other side, both criteria provide a better trend in generating new experiment points. Perfecting the proposed framework is needed in order to develop a better approach to RSM-OD analysis. The potential applications of proposed framework require a provided data recording system and it is usually fulfilled by a type of smart-manufacturing system. Without conducting real experiments that disrupt ongoing processes, the provided observational data that record process parameter changes will help the optimization of the desired process.

REFERENCES

[1] de Oliveira LG, de Paiva AP, Balestrassi PP, et al (2019) Response surface methodology for advanced manufacturing technology optimization: theoretical fundamentals, practical guidelines, and survey literature review. Int J Adv Manuf Technol 104:1785–1837. https://doi.org/10.1007/s00170-019-03809-9

[2] Box AGEP, Wilson KB (1951) On the Experimental Attainment of Optimum Conditions. J R Stat Soc Ser B 13:1–45

[3] Montgomery DC (2017) Design and Analysis of Experiments, 9th ed. Wiley

[4] Myers RH, Montgomery DC, Anderson-Cook CM (2016) Response Surface Methodology: Process and Product Optimization Using Designed Experiments. Wiley

[5] Montgomery D (2017) Exploring observational data. Qual Reliab Eng Int 33:1639–1640. https://doi.org/10.1002/qre.2243

[6] Khoei AR, Masters I, Gethin DT (2002) Design optimisation of aluminium recycling processes using Taguchi technique. J Mater Process Technol 127:96–106. https://doi.org/10.1016/S0924-0136(02)00273-X

[7] Sukthomya W, Tannock JDT (2005) Taguchi experimental design for manufacturing process optimisation using historical data and a neural network process model. Int J Qual Reliab Manag 22:485–502. https://doi.org/10.1108/02656710510598393

[8] Lee SW, Nam SJ, Lee JK (2012) Real-time data acquisition system and HMI for MES. J Mech Sci Technol 26:2381–2388. https://doi.org/10.1007/s12206-012-0615-0

[9] Chien CF, Wang WC, Cheng JC (2007) Data mining for yield enhancement in semiconductor manufacturing and an empirical study. Expert Syst Appl 33:192–198. https://doi.org/10.1016/j.eswa.2006.04.014

[10] Loy C, Goh TN, Xie M (2002) Retrospective factorial fitting and reverse design of experiments. Total Qual Manag 13:589–602. https://doi.org/10.1080/09544120220000002009

[11] Chien CF, Chang KH, Wang WC (2014) An empirical study of design-of-experiment data mining for yield-loss diagnosis for semiconductor manufacturing. J Intell Manuf 25:961–972. https://doi.org/10.1007/s10845-013-0791-5

[12] Sadati N, Chinnam RB, Nezhad MZ (2018) Observational data-driven

e-ISSN 2477-6025
https://doi.org/10.21776/ub.jemis.2024.012.02.3

modeling and optimization of manufacturing processes. Expert Syst Appl 93:456–464. https://doi.org/10.1016/j.eswa.2017.10.028

[13] Nookaraju BC, Sohail M (2020) Experimental investigation and optimization of process parameters of hybrid wick heat pipe using with RSM historical data design. Mater Today Proc. https://doi.org/10.1016/j.matpr.2020.05.634

[14] Wang ML, Qu T, Zhong RY, et al (2012) A radio frequency identification-enabled real-time manufacturing execution system for one-of-a-kind production manufacturing: A case study in mould industry. Int J Comput Integr Manuf 25:20–34. https://doi.org/10.1080/0951192X.2011.575183

[15] Wu D, Wei Y, Terpenny J (2019) Predictive modelling of surface roughness in fused deposition modelling using data fusion. Int J Prod Res 57:3992–4006. https://doi.org/10.1080/00207543.2018.1505058

[16] Liu H, Motoda H (2001) Instance Selection and Construction for Data Mining. Springer Science

[17] Draper NR, Smith H (1998) Applied Regression Analysis, 3rd ed. John Wiley & Sons

[18] Nagaraju S, Vasantharaja P, Chandrasekhar N, et al (2016) Optimization of welding process parameters for 9Cr-1Mo steel using RSM and GA. Mater Manuf Process 31:319–327. https://doi.org/10.1080/10426914.2015.1025974

[19] Liu YC, Yeh IC (2017) Using mixture design and neural networks to build stock selection decision support systems. Neural Comput Appl 28:521–535. https://doi.org/10.1007/s00521-015-2090-x

[20] Vlassides S, Ferrier JG, Block DE (2001) Using historical data for bioprocess optimization: Modeling wine characteristics using artificial neural networks and archived process information. Biotechnol Bioeng 73:55–68. https://doi.org/10.1002/1097-0290(20010405)73:1<55::AID-BIT1036>3.0.CO;2-5

[21] Chi HM, Ersoy OK, Moskowitz H, Altinkemer K (2007) Toward automated intelligent manufacturing systems (AIMS). INFORMS J Comput 19:302–312. https://doi.org/10.1287/ijoc.1050.0171

[22] Shin SJ, Woo J, Rachuri S, Meilanitasari P (2018) Standard data-based predictive modeling for power consumption in turning machining. Sustain 10:1–19. https://doi.org/10.3390/su10030598

[23] Mahmoodi NM, Taghizadeh M, Taghizadeh A (2019) Activated carbon/metal-organic framework composite as a bio-based novel green adsorbent: Preparation and mathematical pollutant removal modeling. J Mol Liq 277:310–322. https://doi.org/10.1016/j.molliq.2018.12.050

[24] Šibalija T, Majstorovic V, Sokovic M (2011) Taguchi-based and intelligent optimisation of a multi-response process using historical data. Stroj Vestnik/Journal Mech Eng 57:357–365. https://doi.org/10.5545/sv-jme.2010.061

[25] Anderson MJ, Whitcomb PJ (2017) RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments, 2nd Ed. CRC Press, Boca Raton, Florida

[26] Hadiyat MA, Sopha BM, Wibowo BS (2022) Response Surface Methodology Using Observational Data: A Systematic Literature Review. Appl Sci 12:. https://doi.org/10.3390/app122010663

[27] Goos P, Jones B (2011) Optimal Design of Experiments. Wiley, West Sussex

[28] Gujarati DN, Porter DC (2009) Basic Econometrics, Fifth Ed. McGraw Hill, New York

[29] Hair JF, Black WC, Babin BJ, Anderson RE (2009) Multivariate Data Analysis, 7th ed. Pearson