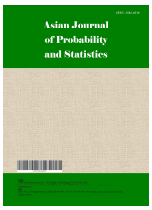

Asian Journal of Probability and Statistics

11(4): 60-85, 2021; Article no.AJPAS.59570
ISSN: 2582-0230

Multivariate-Based Technique for Solving Multi-Response Surface Optimization (MRSO) Problems: The Case of a Maximization Problem

John E. Usen[1*], Stephen S. Akpan[2], Thomas A. Ugbe[2], Ikpang N. Ikpang[3], Joy O. Uket[2] and Bright O. Obeten[1]

[1]Department of Statistics, Cross River University of Technology, P. O. Box 1123, Calabar, Nigeria.
[2]Department of Statistics, University of Calabar, P.O. Box 1115, Calabar, Nigeria.
[3]Department of Statistics, Akwa-Ibom State University, P.O. Box 1167, Mkpat Enin, Akwa-Ibom State, Nigeria.

This work was done in collaboration among all the six authors. Author JEU designed the study, performed the analysis and wrote the first draft of the manuscript. Authors SSA, TAU, INI, JOU and BOO supervised the study and analysis of the data. All the authors managed the literature search, as well as the writing of the final manuscript. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2021/v11i430275
Editor(s):
(1) Dr. Manuel Alberto M. Ferreira, Lisbon University, Portugal.
Reviewers:
(1) Paolo Renna, University of Basilicata, Italy.
(2) Ahmad Jafari Samimi, University of Mazandaran, Iran.
Complete Peer review History: http://www.sdiarticle4.com/review-history/59570

Original Research Article

Received: 20 June 2020
Accepted: 24 August 2020
Published: 07 April 2021

Abstract

Multi-response surface optimization (MRSO) is a problem that is peculiar to an industrial setting, where the aim of a process engineer is to set his process at operating conditions that simultaneously optimize a set of process responses. In Statistics, several methods have been proffered for tackling problems of this nature. Some of such methods are that of: overlapping contour plots, constrained optimization problem, loss function approach, process capability approach, distance function approach, game theory approach, and the desirability function approach. These, methods are however, not without teething flaws as they are either too problem specific, or require very complex and inflexible routines; little wonder, the method of desirability function has gained popularity especially because it overcomes the latter limitation. In this article, we have proposed and implemented a multivariate-based technique for solving MRSO problems. The technique fused the ideas of response surface methodology (RSM), multivariate multiple regression and Pareto optimality. In our technique, RSM was implemented on an all-maximization problem as a case-study process; in which case, first-order models (FOMs) for the responses were fitted using 2^k factorial designs until the FOMs proved to be inadequate, while uniform precision rotatable central

**Corresponding author: E-mail: usenjohn@crutech.edu.ng;*

composite design was used to obtain second-order models (SOMs) for the respective responses in the event of model inadequacy of the FOMs. With the implementation of the proposed technique to the case study, optimal operating conditions were obtained, with observations stemming thereof summarized as axioms. The first, second and third axioms respectively stated that: (1) the mid-point of all optimal operating conditions obtained via the proposed technique is Pareto optimal, (2) the mid-point of all optimal responses at the Pareto optimal operating condition is Pareto optimal, and (3) the region bounded by each of the optimal operating conditions from each second-order model (SOM) is a Pareto front.

Keywords: *Multi-Response Surface Optimization (MRSO); Response surface Methodology (RSM); pareto optimality; pareto front.*

1 Introduction

The production of goods via industrial processes is based on the combination of contributory factors at certain operating conditions that are deemed favorable to a process response [1-4]. In most cases, the interest of the process engineer is to obtain a stable operating condition of the factors which optimizes the process response, in contrast to monitoring the quality of the process [5-8]. The latter objective is the subject matter of statistical quality control. However, in 1951, circumstances of this nature were studied by G. E. P. Box and K. B. Wilson, and from their study stemmed the methodology of response surface which basically tackled the problem of optimizing processes involving just one process response with little or no attention given to the scenario of optimizing processes involving more than one process response [9,7,10,11,2]. And although a variety of definitions have emerged in recent years, this methodology of response surface, popularly regarded as response surface methodology (RSM) was defined as a collection of mathematical and statistical techniques useful for modeling and analyzing problems in which a response of interest is influenced by several factors, and the objective is to optimize the response [9,12,13]. Little wonder, over the years, the technique has been embraced in areas other than the field of Statistics for improving and optimizing processes [5,7,14,15,8,16].

Notwithstanding, in recent years, optimizing multi-response variables has become the subject matter for Statisticians with a view to tackling trending industrial problems which involve the simultaneous optimization of several quality characteristics [7,8,13,17,11], among other authors have all argued that the simultaneous consideration of multi-response variables should commence with the development of appropriate response surface models for each of the response, first, after which attempts to find a set of operating conditions that optimize all the process responses simultaneously, or which at least keep them in desired ranges, can be made. Over the years, beginning with [18-20], scholarly research articles have strictly adhered to the first part of the routine in tackling the problem of multi-response optimization, with variations in their techniques emerging from the second part of the routine.

Deringer and Suich [21] developed and presented a method to construct an overall desirability using a desirability function – a popular method which necessarily requires the decision maker's preference information, and fails otherwise. [22] developed the distance function method, while the method of process capability index was developed by [23]; both methods were either too problem-specific or situation-specific [24]. presented a priority-based approach for MRSO which considers the highest importance as the objective function and the rest of the functions are considered as constraints – a method earlier suggested independently by [25], and then by [26]; but this method ultimately depends on the knowledge of what response is of highest importance among the lot. [27] proffered the use of the squared error loss function which was too problem-specific requiring the use of a cost function. [28] suggested overlays of contour plots for each response as an approach for optimizing several responses; but the application of this technique was restricted to a small set of design variables since it becomes awkward for more than three design variables. [3] developed the method of game theory; but this technique ends up masking the details of the experimental process. [17] used a posterior preference approach to perform robust multi-response optimization; but, this method is yet only efficient when the decision maker's preference information is available, and fails otherwise [29,30,20].

Notwithstanding, depending on the nature of the problem, all the afore-mentioned methods have produced near-satisfactory results at best. Hence, the search for more flexible and robust methods is still on, and which, without any loss of information, optimizes multiple response variables. This anchor point of this article is, thus, to attempt to bridge the said gap. Like all previous attempts, this article embraces the first part of the routine; but, for the second part, it approaches the problem in a manner different from all previous attempts via a multivariate perspective.

2 Materials and Methods

2.1 Materials

2.1.1 The procedure of RSM for single response

[5,6,1] summarizes the existing procedure of the RSM as follows.

2.1.1.1 Plan and run a factorial design around the current operating condition

This process is usually called a screening experiment, particularly because at this stage, all the influencing factors are screened based on how significant their influence is on the response variable of interest.

2.1.1.2 Fit a linear model (with no interaction or quadratic terms) to the data

In most RSM problems, the form of the relationship between the response and the independent variables is unknown. Thus, after the screening of all factors in RSM comes finding a suitable approximation for the true functional relationship between the response variable of interest y and the set of strongly influencing factors. Such a model is mostly fitted using the coded variables $x_1, x_2, x_3, \ldots, x_k$ instead of the natural independent variables $\pi_1, \pi_2, \pi_3, \ldots, \pi_k$. Frequently, a low-order polynomial in some region of these strongly influencing factors is employed. And if the response is well-modeled by a linear function of such factors, then the approximating function is the first-order model (FOM).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon = \beta_0 + \sum_{i=1} \beta_i x_i + \varepsilon$$

2.1.1.3 Determine the PSA (maximization) or PSD (minimization)

To find the PSA (Path of Steepest Ascent) or PSD (Path of Steepest Descent) requires the use of the method of steepest ascent (MSA) or method of steepest descent (MSD) respectively. In particular, the MSA is a procedure for moving sequentially in the direction of the maximum increase in the response. Of course, if minimization is desired, then we employ the MSD. The fitted first-order model is

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i \qquad (1)$$

The PSA is proportional to the sign and magnitude of the regression coefficients in the fitted FOM. A general algorithm for determining the coordinates of a point on the PSA may be stated as follows, assuming that the point $x_1 = x_2 = x_3 = \cdots = x_k = 0$ is the base or origin.

(i) Choose the step size in one of the process variables, say Δx_j . Usually, we would select the variable we know the most about, or we would select the variable that has the largest absolute regression coefficient $\left|\hat{\beta}_j\right|$.

(ii) The step size in other variables is $\Delta x_i = \dfrac{\hat{\beta}_i}{\hat{\beta}_j / \Delta x_j} \quad i = 1, 2, 3, \ldots, k\,;\, i \neq j$.

(iii) Convert the Δx_j from coded variables to the natural variables.

2.1.1.4 Run tests on the PSA until response no longer improves

Here, more experiments are carried out along the PSA until the "lack of fit test" proves the FOM to be an inadequate fit for the data. In such a case, a further increase in the experimental run may lead to a decrease in response (for a maximization problem) or an increase in response (for a minimization problem).

2.1.1.5 If curvature of surface is large, proceed to step 6. Else, return to 2.1.1.1

Here, when the "lack of fit test" proves the FOM to be inadequate for the data, an SOM must be introduced to account for the presence of curvature in the system. This curvature is always indicative of the fact that we are the neighborhood of the optimum.

2.1.1.6 In the neighborhood of the optimum, design, run and fit an SOM using LST

When the experimenter is relatively usually close to the optimum, a model that incorporates curvature is usually required to approximate the response. In most cases, the second-order model (SOM) is adequate. That is,

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

2.1.1.7 Based on the SOM, locate optimal setting of the strongly influencing factors

Here, the optimal setting is the point $x_{1,0}, x_{2,0}, x_{3,0}, \cdots, x_{k,0}$ satisfying the equation

$$\partial \hat{\mathbf{y}} / \partial x_1 = \partial \hat{\mathbf{y}} / \partial x_2 = \partial \hat{\mathbf{y}} / \partial x_3 = \cdots = \partial \hat{\mathbf{y}} / \partial x_k = 0 \tag{3}$$

provided it exists. In other words, the said point is the stationary point, which may represent a point of maximum response, a point of minimum response, or a saddle point. In order to obtain such a point, the procedure of RSM requires the writing the SOM in matrix notation:

$$\hat{y} = \hat{\beta}_0 + \mathbf{x}^T \mathbf{b} + \mathbf{x}^T \mathbf{b} \mathbf{x} \tag{4}$$

In which case,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_k \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/2 & \cdots & \hat{\beta}_{1k}/2 \\ \hat{\beta}_{21}/2 & \hat{\beta}_{22} & \cdots & \hat{\beta}_{2k}/2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\beta}_{k1}/2 & \hat{\beta}_{k2}/2 & \cdots & \hat{\beta}_{kk} \end{pmatrix} \tag{5}$$

In other words, $\mathbf{b}_{k\times 1}$ is a vector of the first-order regression coefficients, whereas $\mathbf{B}_{k\times k}$ is a symmetric matrix having its main diagonal elements as the pure quadratic coefficients $\left\{\hat{\beta}_{ii}\right\}$ for $i=1,2,3,\ldots,k$, but having its off-diagonal elements as one-half of the mixed quadratic coefficients $\left\{\hat{\beta}_{ij}, i\neq j\right\}$. The partial derivatives of $\hat{\mathbf{y}}$ with respect to each element in $\mathbf{x}_{k\times 1}$ when equated to zero gives

$$\frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{x}} = \mathbf{b} + 2\mathbf{B}\mathbf{x} = 0 \tag{6}$$

The stationary point is the solution to (6), in other words:

$$\mathbf{x}_0 = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b} \tag{7}$$

By substituting (7) into (4) we can now obtain the expected response at the stationary point to be:

$$\hat{y}_0 = \hat{\beta}_0 + \frac{1}{2}\mathbf{x}_0^T\mathbf{b} \tag{8}$$

2.1.2 Multivariate multiple regression

Multivariate multiple regression (MMR) considers the problem of modeling the relationship between m responses $y_1, y_2, \ldots, y_m$ and a single set of predictor variables $z_1, z_2, \ldots, z_r$ [31-33]. According to [33], each response is assumed to follow its own regression model, so that:

$$\begin{array}{ccccccccccc} Y_1 & = & \beta_{01} & + & \beta_{11}z_1 & + & \cdots & + & \beta_{r1}z_r & + & \varepsilon_1 \\ Y_2 & = & \beta_{02} & + & \beta_{12}z_1 & + & \cdots & + & \beta_{r2}z_r & + & \varepsilon_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots \\ Y_m & = & \beta_{0m} & + & \beta_{1m}z_1 & + & \cdots & + & \beta_{rm}z_r & + & \varepsilon_m \end{array} \tag{9}$$

The error term $\boldsymbol{\varepsilon}^T = \left(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m\right)$ has $E\left(\boldsymbol{\varepsilon}\right) = \mathbf{0}$ and $\operatorname{var}\left(\boldsymbol{\varepsilon}\right) = \sum$. Thus, the error terms associated with different responses may be correlated [31,32]. To establish notation conforming to the classical linear regression model, let $\mathbf{Z} = \left(z_{j0}, z_{j1}, \ldots, z_{jr}\right)$ denote the values of the predictor variables for the j^{th} trial, let $\mathbf{Y}_J^T = \left(Y_{j1}, Y_{j2}, \ldots, Y_{jm}\right)$ be the responses, and let $\boldsymbol{\varepsilon}^T = \left(\varepsilon_{j1}, \varepsilon_{j2}, \ldots, \varepsilon_{jm}\right)$ be the errors [31,32]. In matrix notation, the design matrix

$$\underset{(n\times(r+1))}{\mathbf{Z}} = \begin{pmatrix} z_{10} & z_{11} & \cdots & z_{1r} \\ z_{20} & z_{21} & \cdots & z_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n0} & z_{n1} & \cdots & z_{nr} \end{pmatrix} \tag{10}$$

is the same as that for the single-response regression model [33]. The other matrix quantities have multivariate counterparts [33]. Set

$$\underset{(n\times m)}{\mathbf{Y}}=\begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ Y_{21} & Y_{22} & \cdots & Y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nm} \end{pmatrix}=\left(\mathbf{Y}_{(1)}\vdots\mathbf{Y}_{(2)}\vdots\cdots\vdots\mathbf{Y}_{(m)}\right) \tag{11}$$

$$\underset{((r+1)\times m)}{\boldsymbol{\beta}}=\begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0m} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{r1} & \beta_{r2} & \cdots & \beta_{rm} \end{pmatrix}=\left(\boldsymbol{\beta}_{(1)}\vdots\boldsymbol{\beta}_{(2)}\vdots\cdots\vdots\boldsymbol{\beta}_{(m)}\right) \tag{12}$$

$$\underset{(n\times m)}{\boldsymbol{\varepsilon}}=\begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nm} \end{pmatrix}=\left(\boldsymbol{\varepsilon}_{(1)}\vdots\boldsymbol{\varepsilon}_{(2)}\vdots\cdots\vdots\boldsymbol{\varepsilon}_{(m)}\right)=\begin{pmatrix} \boldsymbol{\varepsilon}_{(1)}^{T} \\ \cdots \\ \boldsymbol{\varepsilon}_{(2)}^{T} \\ \cdots \\ \vdots \\ \cdots \\ \boldsymbol{\varepsilon}_{(n)}^{T} \end{pmatrix} \tag{13}$$

The multivariate linear regression model is

$$\underset{(n\times m)}{\mathbf{Y}}=\underset{(n\times(r+1))}{\mathbf{Z}}\underset{((r+1)\times m)}{\boldsymbol{\beta}}+\underset{(n\times m)}{\boldsymbol{\varepsilon}} \tag{14}$$

with $E\left(\boldsymbol{\varepsilon}_{(i)}\right)=\mathbf{0}$ and $\operatorname{cov}\left(\boldsymbol{\varepsilon}_{(i)},\boldsymbol{\varepsilon}_{(k)}\right)=\sigma_{ik}\mathbf{I}$ for $i,k=1,2,\ldots,m$ [31,32].

The m observations on the j^{th} trial have covariance matrix $\Sigma=\{\sigma_{ik}\}$, but observations from different trials are uncorrelated. Here $\boldsymbol{\beta}$ and σ_{ik} are unknown parameters; the design matrix $\mathbf{Z}$ has j^{th} row $\left(z_{j0},z_{j1},\ldots,z_{jr}\right)$ [31,32]. Simply stated, the i^{th} response $\mathbf{Y}_{(i)}$ follows the linear regression model

$$\mathbf{Y}_{(i)}=\mathbf{Z}\boldsymbol{\beta}_{(i)}+\boldsymbol{\varepsilon}_{(i)},\qquad i=1,2,\ldots,m \tag{15}$$

with $\operatorname{cov}\left(\boldsymbol{\varepsilon}_{(i)}\right)=\sigma_{ii}\mathbf{I}$. However, the errors for different responses on the same trial can be correlated [31,32]. Given the outcomes $\mathbf{Y}$ and the values of the predictor variables $\mathbf{Z}$ with full column rank, we determine the least squares estimates $\hat{\boldsymbol{\beta}}_{(i)}$ exclusively from the observations $\mathbf{Y}_{(i)}$ on the i^{th} response [33]. In conformity with the single-response solution, we take

$$\hat{\boldsymbol{\beta}}_{(i)}=\left(\mathbf{Z}^{T}\mathbf{Z}\right)^{-1}\mathbf{Z}^{T}\mathbf{Y}_{(i)} \tag{16}$$

Collecting these univariate least squares estimates, we obtain

$$\hat{\boldsymbol{\beta}}=\left(\hat{\boldsymbol{\beta}}_{(1)}\vdots\hat{\boldsymbol{\beta}}_{(2)}\vdots\cdots\vdots\hat{\boldsymbol{\beta}}_{(m)}\right)=\left(\mathbf{Z}^{T}\mathbf{Z}\right)^{-1}\mathbf{Z}^{T}\left(\mathbf{Y}_{(1)}\vdots\mathbf{Y}_{(2)}\vdots\cdots\vdots\mathbf{Y}_{(m)}\right)$$

or

$$\hat{\boldsymbol{\beta}}_{(i)} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{Y} \tag{17}$$

For any choice of parameters $\mathbf{B} = \left(\mathbf{b}_{(1)} \vdots \mathbf{b}_{(2)} \vdots \cdots \vdots \mathbf{b}_{(m)}\right)$, the matrix of errors is $\mathbf{Y} - \mathbf{ZB}$. The error sum of squares and cross products matrix is

$$\begin{aligned}&\left(\mathbf{Y}-\mathbf{ZB}\right)^T\left(\mathbf{Y}-\mathbf{ZB}\right)\\ &=\begin{pmatrix} \left(\mathbf{Y}_{(1)}-\mathbf{Zb}_{(1)}\right)^T\left(\mathbf{Y}_{(1)}-\mathbf{Zb}_{(1)}\right) & \cdots & \left(\mathbf{Y}_{(1)}-\mathbf{Zb}_{(1)}\right)^T\left(\mathbf{Y}_{(m)}-\mathbf{Zb}_{(m)}\right) \\ \vdots & \ddots & \vdots \\ \left(\mathbf{Y}_{(m)}-\mathbf{Zb}_{(m)}\right)^T\left(\mathbf{Y}_{(1)}-\mathbf{Zb}_{(1)}\right) & \cdots & \left(\mathbf{Y}_{(m)}-\mathbf{Zb}_{(m)}\right)^T\left(\mathbf{Y}_{(m)}-\mathbf{Zb}_{(m)}\right) \end{pmatrix}\end{aligned} \tag{18}$$

The selection $\mathbf{b}_{(i)} = \hat{\boldsymbol{\beta}}_{(i)}$ minimizes the i^{th} diagonal sum of squares $\left(\mathbf{Y}_{(i)}-\mathbf{Zb}_{(i)}\right)^T\left(\mathbf{Y}_{(i)}-\mathbf{Zb}_{(i)}\right)$

Consequently, $tr\left(\left(\mathbf{Y}-\mathbf{ZB}\right)^T\left(\mathbf{Y}-\mathbf{ZB}\right)\right)$ is minimized by the choice $\mathbf{B} = \hat{\boldsymbol{\beta}}$. Also, the generalized variance $\left|\left(\mathbf{Y}-\mathbf{ZB}\right)^T\left(\mathbf{Y}-\mathbf{ZB}\right)\right|$ is minimized by the least squares estimates $\hat{\boldsymbol{\beta}}$.

Using the least squares estimates $\hat{\boldsymbol{\beta}}$ we can form the matrices of:

Predicted values: $\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} = \mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{Y}$

Residuals: $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = \left(\mathbf{I} - \mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\right)\mathbf{Y} = \mathbf{0}$ (19)

The orthogonality conditions among the residuals, predicted values, and columns of $\mathbf{Z}$, which hold in classical linear regression, hold in multivariate multiple regression. They follow from $\mathbf{Z}^T\left(\mathbf{I} - \mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\right) = \mathbf{Z}^T - \mathbf{Z}^T = \mathbf{0}$. Specifically,

$$\mathbf{Z}^T\hat{\boldsymbol{\varepsilon}} = \mathbf{Z}^T\left(\mathbf{I} - \mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\right)\mathbf{Y} = \mathbf{0} \tag{20}$$

so the residuals $\hat{\boldsymbol{\varepsilon}}_{(i)}$ are perpendicular to the columns of $\mathbf{Z}$. Also,

$$\hat{\mathbf{Y}}^T\hat{\boldsymbol{\varepsilon}} = \hat{\boldsymbol{\beta}}^T\mathbf{Z}^T\left(\mathbf{I} - \mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\right)\mathbf{Y} = \mathbf{0} \tag{21}$$

confirming that the predicted values $\hat{\mathbf{Y}}_{(i)}$ are perpendicular to all residual vectors $\hat{\boldsymbol{\varepsilon}}_{(k)}$.

Because $\mathbf{Y} = \hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}$,

$$\mathbf{Y}^T\mathbf{Y} = \left(\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}\right)^T\left(\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}\right) = \hat{\mathbf{Y}}^T\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} + \mathbf{0} + \mathbf{0}^T$$

or

$$\underset{\left(\begin{array}{l}\text{total sum of squares}\\ \text{and cross products}\end{array}\right)}{\mathbf{Y}^T\mathbf{Y}} = \underset{\left(\begin{array}{l}\text{predicted sum of squares}\\ \text{and cross products}\end{array}\right)}{\hat{\mathbf{Y}}^T\hat{\mathbf{Y}}} + \underset{\left(\begin{array}{l}\text{residual error sum}\\ \text{of squares and cross}\\ \text{product}\end{array}\right)}{\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}} \quad (22)$$

The residual sum squares and cross products can also be written as

$$\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} = \mathbf{Y}^T\mathbf{Y} - \hat{\mathbf{Y}}^T\hat{\mathbf{Y}} = \mathbf{Y}^T\mathbf{Y} - \hat{\boldsymbol{\beta}}^T\mathbf{Z}^T\mathbf{Z}\hat{\boldsymbol{\beta}} \quad (23)$$

2.1.3 Pareto optimality

Pareto optimality is a situation that cannot be modified so as to make any one individual or preference criterion better off without making at least one individual or preference criterion worse [3,8]. The concept is named after Vilfredo Pareto (1848 – 1923), an Italian engineer and economist, who used the concept in his studies of economic efficiency and income distribution. The following are four definitions or closely-related concepts on Pareto optimality.

2.1.3.1 Definition 1 – Pareto improvement

Given an initial situation, a Pareto improvement is a new situation which is weakly preferred by all agents, and strictly preferred by at least one agent [8,16]. In a sense, it is a unanimously-agreed improvement such that a movement to the new situation would imply gain to some agents, and no agent will lose.

2.1.3.2 Definition 2 – Pareto dominated

A situation is called Pareto dominated if it has a Pareto improvement [3,16].

2.1.3.3 Definition 3 – Pareto optimal

A situation is Pareto optimal or Pareto efficient if it is not Pareto dominated [3,8].

2.1.3.4 Definition 4 – Pareto frontier

Pareto frontier is the set of all Pareto efficient allocations. It also variously known as Pareto front or Pareto set [3,8,16].

2.2 Methods

2.2.1 Already existing techniques for solving MRSO problems

From its initial development till date, a variety of methods have been introduced for multiple-response optimization, all of which have been categorized at different times by several authors. For instance, [27] categorized the existing methods into three; [34] categorized the existing methods into four. In this article, we have considered the most recent classifications being that of [7], and [35]. These categories are the approaches of: overlapping contour plots, constrained optimization problem, loss function, process capability, distance function, game theory and desirability function.

2.2.1.1 Overlapping contour plots

[28] suggested that an approach for optimizing several responses is to overlay the contour plots for each response. Here, the experimenter can visually examine the contour plot to discover the appropriate operating conditions. [7] emphasized that this technique is mainly suitable when there are few design variables, since it

becomes awkward for more than three design variables. More so, in this approach [7] explains that there is no need for the decision maker's information especially as contour plots play the main role.

2.2.1.2 Constrained optimization problem

The formulation and solving of the multi-response problem like a constrained optimization problem was described by [36] as a popular approach. [24] classified it as a priority based approach. The priority-based approach which is similar to a method-bounded objective in the multi-objective decision-making problem chooses the response with the highest importance as the objective function and the rest of the functions are considered as constraints, although it is not always much straightforward. This idea was first suggested by [25]. In their study, the responses were assumed and referred to as a "primary response" and a "constraint response". The objective was to find conditions on a set of designed variables which maximizes the primary response function subject to the constraint response function. Subsequently, [26] considered multiple process responses by extending the study and formulation of [25].

2.2.1.3 Loss function approach

The squared error loss function was first suggested by [27] as:

$$L\left(y\left(x\right)\right)=\left(y\left(x\right)-\Phi\right)^{T}C\left(y\left(x\right)-\Phi\right) \tag{24}$$

where $y\left(x\right)$ is the response vector, $\Phi\left(x\right)$ is the target vector, and C is the cost matrix and is used to determine the relative importance of the response variables.

2.2.1.4 Process capability approach

Process capability index is used to evaluate whether a process is able to meet current specification limits. [23] presented the index C_{pm} as:

$$C_{pm}=\frac{USL-LSL}{6\sqrt{\sigma^{2}+\left(\mu-T\right)^{2}}} \tag{25}$$

where USL and LSL are specification limits, and μ, σ^{2} and T respectively denote the mean, variance, and target in the above equation. Subsequently, in an independent work, [37] further extended this index; hence, this index could now be applied in multi-response optimization. The maximization of process capability as a criterion for multi-response optimization was further considered by [19].

2.2.1.5 Distance function approach

The distance function approach was proposed by [22]. The distance function is

$$\text{Distance}\left\{\hat{y}\left(x\right),T\right\}=\left\{\hat{y}\left(x-T\right)^{T}\sum\nolimits_{\hat{y}\left(x\right)}^{-1}\left\{\hat{y}\left(x\right)-T\right\}\right\} \tag{26}$$

where T represents the target value, $\hat{y}\left(x\right)$ is the predicted response, and $\Sigma_{\hat{y}\left(x\right)}$ is the variance-covariance matrix of the predicted responses. The optimal operating condition is achieved if the distance function gets minimized.

2.2.1.6 Game theory approach

[3] proposed a game theoretic-based approach for multi-response optimization by viewing each response as a player and each factor as strategies of each player. Their approach could determine the best predictor factor sets in order to obtain the best joint desirability of responses. This was achieved this, the signal to noise ratio (SN) index for each response will be calculated by considering the joint values of strategies, following which the obtained SN ratios for each strategy are modeled in the game theory table. To end the procedure, via Nash Equilibrium, the best strategy which is the best values of predictor factors is then determined.

2.2.1.7 The desirability function approach

The desirability function approach is one of the most widely used methods in industry for the optimization of multiple response processes. It is based on the idea that the "quality" of a product or process that has multiple quality characteristics, with one of them outside of some "desired limits", is completely unacceptable. The method finds operating conditions $\mathbf{x}$ that provide the "most desirable" response values.

For each response y_i, a desirability function $d_i(y_i)$ assigns numbers between 0 and 1 to the possible values of y_i with $d_i(y_i)=0$ representing a completely undesirable value of y_i and $d_i(y_i)=1$ representing a completely desirable or ideal response value. The individual desirabilities are then combined using the geometric mean, which gives the overall desirability $D=\{d_1(y_1)\times d_2(y_2)\times d_3(y_3)\times\cdots\times d_k(y_k)\}^{\frac{1}{k}}$ with k denoting the number of responses. Notice that if any response is completely undesirable (that is, if $d_i(y_i)=0$), then the overall desirability is zero. In practice, fitted response values $\hat{y}_i$ are used in place of the y_i.

Depending on whether a particular response y_i is to be maximized, minimized, or assigned a target value, different desirability functions $\hat{d}_i(y_i)$ can be used. A useful class of desirability functions was proposed by [21]. Let L_i, T_i and U_i be the lower, upper and target values, respectively, that are desired for response y_i, with $L_i \le T_i \le U_i$.

If a response is of the "target is best" kind, then its individual desirability function is:

$$d_i(\hat{y}_i)=\begin{cases} 0 & if\ \hat{y}_i < L_i \\ \left\{\dfrac{\hat{y}_i - L_i}{T_i - L_i}\right\}^s & if\ L_i \le \hat{y}_i \le T_i \\ \left\{\dfrac{\hat{y}_i - U_i}{T_i - U_i}\right\}^t & if\ T_i \le \hat{y}_i \le U_i \\ 0 & if\ \hat{y}_i > U_i \end{cases}$$

with the exponent s and t determining how important it is to hit the target value. For $s=t=1$, the desirability function increases linearly towards T_i; for $s<1$, $t<1$, the function is convex; for $s>1$, $t>1$, the function is concave (see the example below for an illustration).

If a response is to be maximized instead, the individual desirability is defined as:

$$d_i(\hat{y}_i) = \begin{cases} 0 & \text{if } \hat{y}_i < L_i \\ \left\{ \dfrac{\hat{y}_i - L_i}{T_i - L_i} \right\}^s & \text{if } L_i \le \hat{y}_i \le T_i \\ 1.0 & \text{if } \hat{y}_i > T_i \end{cases}$$

with T_i in this case interpreted as a large enough value for the response.

Finally, if we want to minimize a response, we could use:

$$d_i(\hat{y}_i) = \begin{cases} 1.0 & \text{if } \hat{y}_i < T_i \\ \left\{ \dfrac{\hat{y}_i - U_i}{T_i - U_i} \right\}^s & \text{if } T_i \le \hat{y}_i \le U_i \\ 0 & \text{if } \hat{y}_i > U_i \end{cases}$$

with T_i denoting a small enough value for the response.

The desirability approach consists of the following steps:

1. Conduct experiments and fit response models for all k responses;
2. Define individual desirability functions for each response;
3. Maximize the overall desirability D with respect to the controllable factors.

However, the most important advantage of this approach is that the Decision Maker's preference information can be easily applied in the model. In addition, it is easy to use, and is popular among available methods.

2.2.2 Our proposed technique for solving MRSO problems

Our proposed alternative to multi-response surface optimization can be achieved via the following steps.

Step 1: Implement the RSM procedure on each response to obtain their respective SOMs.
Step 2: For each SOM, obtain the respective optimal operating conditions and responses.
Step 3: Average the optimal solutions in step 2 to get the Pareto optimal operating condition.
Step 4: Average the optimal responses in step 2 to get the Pareto optimal response.

3 Results and Discussion

3.1 Implementation

As a case study for our implementation, we considered a hypothetical scenario in which the target of a process engineer was to determine, for his process, the operating condition which maximized the volume y_1 of his 50 cl carbonated beverage packaged per bottle, and the number y_2 of bottles filled every 10 minutes, of his process. Three factors which influenced both responses were volume of carbon-dioxide (π_1), operating pressure (π_2), and line speed (π_3). As it was unlikely that the region about the current operating condition housed the optimum, a first order model (FOM) was fit and the method of steepest ascent (MSA) applied. To fit the FOM, the engineer had decided that the region of exploration for fitting the FOM should be $(0.5, 1.5)$ gl^{-1} of CO_2, $(30, 40)$ psi of pressure, and $(15, 25)$ cms^{-1} of line speed. A 2^3 factorial experiment

with two replicates was conducted while observing both responses. The order in which the observations were taken was determined randomly.

To simplify the calculations, the controllable variables were coded to a $(-1, 1)$ interval. Thus, if π_1 denotes the natural variable percentage carbonation, π_2 denotes the natural variable operating pressure in the filter, and π_3 denotes the natural variable line speed, then the coded variables are:

$$x_1 = \frac{\pi_1 - 1}{0.5} \qquad x_2 = \frac{\pi_2 - 35}{5} \qquad x_3 = \frac{\pi_3 - 20}{5}$$

The coded data that resulted from this experiment are shown in Table 1.

For the multiple regression analysis, the null hypothesis states that "the multiple linear regression model is adequately fitted", while the alternative hypothesis states that "there is lack of fit in the multiple linear regression model". If the p-value is less than the 0.05 significance level, we reject the null hypothesis in favor of the alternative, hence concluding that "there is sufficient evidence at the 0.05 significance level to conclude that there is a lack of fit in the multiple linear regression model. However, if the p-value is greater than the significance level 0.05, we do not reject the null hypothesis; hence, we conclude that "there is insufficient evidence at the 0.05 level of significance to conclude that there is a lack of fit in the multiple linear regression model".

Table 1. Coded data for the first experiment in the case-study

Run	Natural Variables			Coded Variables			Responses	
	π_1	π_2	π_3	x_1	x_2	x_3	y_1	y_2
1.	1	35	20	0	0	0	43	54
2.	0.5	30	15	1	-1	-1	47	73
3.	1.5	30	25	-1	-1	1	43	53
4.	1	35	20	0	0	0	42	68
5.	1	35	20	0	0	0	44	67
6.	0.5	40	15	-1	1	-1	45	84
7.	0.5	40	25	-1	1	1	46	76
8.	1.5	30	25	1	-1	1	47	92
9.	1.5	40	25	1	1	1	46	57
10.	1.5	40	15	1	1	-1	44	89
11.	0.5	30	15	-1	-1	-1	43	62
12.	1	35	20	0	0	0	42	75
13.	1	35	20	0	0	0	45	83
14.	1	35	20	0	0	0	43	91

Uniform precision 2^3 factorial design at a 0.05 level of significance

Regression Analysis: y1 versus x1, x2, x3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	7.3750	2.4583	0.78	0.531
x1	1	6.1250	6.1250	1.95	0.193
x2	1	0.1250	0.1250	0.04	0.846
x3	1	1.1250	1.1250	0.36	0.563
Error	10	31.4821	3.1482		
Lack-of-Fit	5	24.6488	4.9298	3.61	0.093
Pure Error	5	6.8333	1.3667		

Total 13 38.8571

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.77432	18.98%	0.00%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	44.286	0.474	93.39	0.000	
x1	0.875	0.627	1.39	0.193	1.00
x2	0.125	0.627	0.20	0.846	1.00
x3	0.375	0.627	0.60	0.563	1.00

Regression Equation

y1 = 44.286 + 0.875 x1 + 0.125 x2 + 0.375 x3

Regression Analysis: y2 versus x1, x2, x3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	359.00	119.67	0.59	0.633
x1	1	162.00	162.00	0.80	0.391
x2	1	84.50	84.50	0.42	0.532
x3	1	112.50	112.50	0.56	0.472
Error	10	2014.71	201.47		
Lack-of-Fit	5	1164.71	232.94	1.37	0.369
Pure Error	5	850.00	170.00		
Total	13	2373.71			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
14.1941	15.12%	0.00%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	73.14	3.79	19.28	0.000	
x1	4.50	5.02	0.90	0.391	1.00
x2	3.25	5.02	0.65	0.532	1.00
x3	-3.75	5.02	-0.75	0.472	1.00

Regression Equation

y2 = 73.14 + 4.50 x1 + 3.25 x2 - 3.75 x3

Fits and Diagnostics for Unusual Observations

Obs	y2	Fit	Resid	Std Resid	
8	92.00	70.64	21.36	2.02	R

R Large residual

Since the p-value (0.093) in the ANOVA for the first regression model is larger than the significance level 0.05, we fail to reject the null hypothesis in favour of the alternative. We conclude that “there is insufficient

evidence at the 0.05 level of significance to conclude that there is a lack of fit in the multiple linear regression model".

Since the p-value (0.369) in the ANOVA for the second regression model is larger than the significance level 0.05, we fail to reject the null hypothesis in favour of the alternative. We conclude that "there is insufficient evidence at the 0.05 level of significance to conclude that there is a lack of fit in the multiple linear regression model".

To move away from the design center $(x_1 = 0, x_2 = 0, x_3 = 0)$ along the PSA with regards to the first response y_1 (volume of carbonated beverage packaged per bottle), we moved 0.875 units in the x_1 direction for every 0.125 units in the x_2 direction and 0.375 units in the direction x_3 direction. Thus, in this case, the PSA passed through the points $(x_1 = 0, x_2 = 0, x_3 = 0)$ and had slopes $x_2/x_1 = 0.125/0.875$ and $x_3/x_1 = 0.375/0.875$. The engineer decided to use 0.5cl of carbon-dioxide as the basic step size. Using the relationship between π_1 and x_1 we observed that 0.5cl of carbon-dioxide is equivalent to a step in the coded variable x_1 of $\Delta x_1 = 1$. Therefore, the steps along the PSA are $\Delta x_1 = 1.0000$, $\Delta x_2 = \left(0.125/0.875\right)\Delta x_1 \approx 0.1429$ and $\Delta x_3 = \left(0.375/0.875\right)\Delta x_1 \approx 0.4286$. We computed points along this path and observed values for the first response y_1 volume of carbonated beverage packaged per bottle) until decrease in response were noted. Specifically, for the first response y_1, increase in response was observed through the second step; but the third step produced a decrease in response y_1. Therefore, an FOM should have been fitted in the general vicinity of the point $-(\pi_1 = 2, \pi_2 = 36.4290, \pi_3 = 24.2860)$. The result is shown in Table 2. The steps are shown in both the coded and natural variables. However, while the coded variables were easier to manipulate mathematically, the natural variables were used in running the process.

To move away from the design center $(x_1 = 0, x_2 = 0, x_3 = 0)$ along the PSA with regards to the second response y_2 (number of bottles filled every 10 minutes), we moved 4.50 units in the x_1 – direction for every 3.25 units in the x_2 – direction and -3.75 units in the x_3 – direction. Thus, in this case, the PSA passed through the points $(x_1 = 0, x_2 = 0, x_3 = 0)$ and had slopes $-\ x_2/x_1 = 3.25/4.50$ and $x_3/x_1 = 3.75/4.50$. The engineer decided to use 0.5cl of carbon-dioxide as the basic step size. Using the relationship between π_1 and x_1 we observed that 0.5cl of carbon-dioxide is equivalent to a step in the coded variable x_1 of $\Delta x_1 = 1$. Therefore, the steps along the PSA are, $\Delta x_1 = 1.0000$, $\Delta x_2 = \left(3.25/4.50\right)\Delta x_1 \approx 0.7222$ and $\Delta x_3 = \left(3.75/4.50\right)\Delta x_1 \approx -0.8333$. We computed points along this path and observed values for the second response y_2 (number of bottles filled every 10 minutes) until decrease in response were noted. Specifically, for the second response y_2, increase in response was observed was through the fourth step; but the fifth step produced a decrease in response y_2. Therefore, another FOM should have been fitted in the general vicinity of the point $-(\pi_1 = 3, \pi_2 = 49.440, \pi_3 = 3.3340)$. The result is shown in Table 3. The steps are shown in both the coded and natural variables. Once again, while the coded variables were easier to manipulate mathematically, the natural variables were used in running the process. In order to find a region of compromise, we obtained the mid-point of the two coordinates, $(\pi_1 = 3, \pi_2 = 43, \pi_3 = 14)$, and fit a new FOM in its general vicinity. In order to simplify the calculations, the controllable variables were again coded to a $(-1, 1)$ interval giving:

$$x_1 = \frac{\pi_1 - 3}{0.5} \qquad x_1 = \frac{\pi_2 - 43}{5} \qquad x_1 = \frac{\pi_3 - 14}{5}$$

The coded data that resulted from this experiment are shown in Table 4.

Table 2. Steepest ascent experiment in the first response for the case-study

Run No.	Natural variables			Coded variables			Response
	For the first response type						
	π_1	π_2	π_3	x_1	x_2	x_3	y_1
Origin	1.0	35	20	0	0	0	
Δ				1.0000	0.1429	0.4286	
Origin+Δ	1.5	35.7145	22.1430	1.0000	0.1429	0.4286	43
Origin+2Δ	2.0	36.4290	24.2860	2.0000	0.2858	0.8572	**47**
Origin+3Δ	2.5	37.1435	26.4290	3.0000	0.4287	1.2858	45

Table 3. Steepest ascent experiment in the second response for the case-study

Run No.	Natural variables			Coded variables			Response
For the first response type							
	π_1	π_2	π_3	x_1	x_2	x_3	y_1
Origin	1.0	35	20	0	0	0	
Δ				1.0000	0.7222	-0.8333	
Origin+Δ	1.5	38.6110	15.8335	1.0000	0.7222	-0.8333	73
Origin+2Δ	2.0	42.2220	11.6670	2.0000	1.4444	-1.6666	75
Origin+3Δ	2.5	45.8330	7.5005	3.0000	2.1666	-2.4999	76
Origin+4Δ	3.0	49.4440	3.3340	4.0000	2.8888	-3.3332	**80**
Origin+5Δ	3.5	53.0550	-0.8325	5.0000	3.6110	-4.1665	77

Table 4. Coded data for the second experiment in the case-study

Run	Natural Variables			Coded Variables			Responses	
	π_1	π_2	π_3	x_1	x_2	x_3	y_1	y_2
1.	3	43	14	0	0	0	47	80
2.	2.5	38	9	1	-1	-1	49	83
3.	3.5	38	19	-1	-1	1	49	62
4.	3	43	14	0	0	0	47	80
5.	3	43	14	0	0	0	47	80
6.	2.5	48	9	-1	1	-1	50	82
7.	2.5	48	14	-1	1	1	52	84
8.	3.5	38	14	1	-1	1	51	91
9.	3.5	48	14	1	1	1	52	83
10.	3.5	48	9	1	1	-1	54	89
11.	2.5	38	9	-1	-1	-1	53	83
12.	3	43	14	0	0	0	47	80
13.	3	43	14	0	0	0	47	80
14.	3	43	14	0	0	0	47	80

Uniform precision 2^3 factorial design at a 0.05 level of significance

Regression Analysis: y1 versus x1, x2, x3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	5.5000	1.8333	0.23	0.874
x1	1	0.5000	0.5000	0.06	0.808
x2	1	4.5000	4.5000	0.56	0.470
x3	1	0.5000	0.5000	0.06	0.808
Error	10	79.9286	7.9929		
Lack-of-Fit	5	79.9286	15.9857	*	*
Pure Error	5	0.0000	0.0000		
Total	13	85.4286			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.82716	6.44%	0.00%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	49.429	0.756	65.42	0.000	
x1	0.25	1.00	0.25	0.808	1.00
x2	0.75	1.00	0.75	0.470	1.00
x3	-0.25	1.00	-0.25	0.808	1.00

Regression Equation

y1 = 49.429 + 0.25 x1 + 0.75 x2 - 0.25 x3

Fits and Diagnostics for Unusual Observations

Obs	y1	Fit	Resid	Std Resid	
11	53.00	48.68	4.32	2.05	R

R Large residual

Regression Analysis: y2 versus x1, x2, x3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	234.375	78.125	2.46	0.123
x1	1	153.125	153.125	4.82	0.053
x2	1	45.125	45.125	1.42	0.261
x3	1	36.125	36.125	1.14	0.312
Error	10	317.982	31.798		
Lack-of-Fit	5	317.982	63.596	*	*
Pure Error	5	0.000	0.000		
Total	13	552.357			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5.63899	42.43%	25.16%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	81.21	1.51	53.89	0.000	

x1 4.38 1.99 2.19 0.053 1.00
x2 2.38 1.99 1.19 0.261 1.00
x3 -2.13 1.99 -1.07 0.312 1.00

Regression Equation
y2 = 81.21 + 4.38 x1 + 2.38 x2 - 2.13 x3

Fits and Diagnostics for Unusual Observations
Obs y2 Fit Resid Std Resid
3 62.00 72.34 -10.34 -2.46 R
8 91.00 81.09 9.91 2.36 R

R Large residual

Since the p-value (0.000) in the ANOVA for the first regression model is smaller than the 0.05 significance level, we reject the null hypothesis in favor of the alternative. We conclude that "there is sufficient evidence at the 0.05 level of significance to conclude that there is a lack of fit in the multiple linear regression model".

Since the p-value (0.000) in the ANOVA for the second regression model is smaller than the 0.05 significance level, we reject the null hypothesis in favor of the alternative. We conclude that "there is sufficient evidence at the 0.05 significance level to conclude that there is a lack of fit in the multiple linear regression model".

In order to cater for curvature in both systems, two (2) SOMs had to be fitted using a uniform precision design. The coded data that resulted from this experiment are shown in Table 5.

Table 5. Coded data for fitting the SOMs for the case study

Run	Natural Variables			Coded Variables			Responses	
	π_1	π_2	π_3	x_1	x_2	x_3	y_1	y_2
1.	3	43	14	0	0	0	47	80
2.	2.5	38	9	1	-1	-1	49	83
3.	3.5	38	19	-1	-1	1	49	62
4.	3	43	14	0	0	0	47	80
5.	3	43	14	0	0	0	47	80
6.	2.5	48	9	-1	1	-1	50	82
7.	2.5	48	14	-1	1	1	52	84
8.	3.5	38	14	1	-1	1	51	91
9.	3.5	48	14	1	1	1	52	83
10.	3.5	48	9	1	1	-1	54	89
11.	2.5	38	9	-1	-1	-1	53	83
12.	3	43	14	0	0	0	47	80
13.	3	43	14	0	0	0	47	80
14.	3	43	14	0	0	0	47	80
15.	3.841	43	14	1.682	0	0	58	86
16.	2.159	43	14	-1.682	0	0	56	87
17.	3	51.410	14	0	1.682	0	54	89
18.	3	34.590	14	0	-1.682	0	53	86
19.	3	43	22.410	0	0	1.682	57	85
20.	3	43	5.590	0	0	-1.682	58	83

Uniform precision 2^3 factorial design at a 0.05 level of significance

Regression Analysis: y1 versus x1, x2, x3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	185.697	20.6330	2.00	0.147
x1	1	2.107	2.1066	0.20	0.661
x2	1	4.321	4.3207	0.42	0.532
x3	1	0.993	0.9926	0.10	0.763
x1*x1	1	84.292	84.2923	8.18	0.017
x2*x2	1	20.104	20.1044	1.95	0.193
x3*x3	1	97.064	97.0644	9.41	0.012
x1*x2	1	4.500	4.5000	0.44	0.524
x1*x3	1	0.500	0.5000	0.05	0.830
x2*x3	1	0.500	0.5000	0.05	0.830
Error	10	103.103	10.3103		
Lack-of-Fit	5	103.103	20.6206	*	*
Pure Error	5	0.000	0.0000		
Total	19	288.800			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.21097	64.30%	32.17%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	47.17	1.31	36.02	0.000	
x1	0.393	0.869	0.45	0.661	1.00
x2	0.562	0.869	0.65	0.532	1.00
x3	-0.270	0.869	-0.31	0.763	1.00
x1*x1	2.418	0.846	2.86	0.017	1.02
x2*x2	1.181	0.846	1.40	0.193	1.02
x3*x3	2.595	0.846	3.07	0.012	1.02
x1*x2	0.75	1.14	0.66	0.524	1.00
x1*x3	0.25	1.14	0.22	0.830	1.00
x2*x3	0.25	1.14	0.22	0.830	1.00

Regression Equation

y1 = 47.17 + 0.393 x1 + 0.562 x2 - 0.270 x3 + 2.418 x1*x1 + 1.181 x2*x2 + 2.595 x3*x3 + 0.75 x1*x2 + 0.25 x1*x3 + 0.25 x2*x3

Fits and Diagnostics for Unusual Observations

Obs	y1	Fit	Resid	Std Resid	
2	49.00	52.71	-3.71	-2.01	R

R Large residual

Regression Analysis: y2 versus x1, x2, x3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	335.911	37.3235	1.12	0.427
x1	1	81.276	81.2761	2.44	0.149
x2	1	42.334	42.3341	1.27	0.286
x3	1	13.614	13.6138	0.41	0.537

x1*x1	1	28.406	28.4060	0.85	0.377
x2*x2	1	44.513	44.5130	1.34	0.274
x3*x3	1	3.899	3.8987	0.12	0.739
x1*x2	1	66.125	66.1250	1.99	0.189
x1*x3	1	55.125	55.1250	1.66	0.227
x2*x3	1	10.125	10.1250	0.30	0.593
Error	10	332.639	33.2639		
Lack-of-Fit	5	332.639	66.5278	*	*
Pure Error	5	0.000	0.0000		
Total	19	668.550			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5.76749	50.24%	5.46%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	80.14	2.35	34.07	0.000	
x1	2.44	1.56	1.56	0.149	1.00
x2	1.76	1.56	1.13	0.286	1.00
x3	-1.00	1.56	-0.64	0.537	1.00
x1*x1	1.40	1.52	0.92	0.377	1.02
x2*x2	1.76	1.52	1.16	0.274	1.02
x3*x3	0.52	1.52	0.34	0.739	1.02
x1*x2	-2.88	2.04	-1.41	0.189	1.00
x1*x3	2.62	2.04	1.29	0.227	1.00
x2*x3	1.13	2.04	0.55	0.593	1.00

Regression Equation

y2 = 80.14 + 2.44 x1 + 1.76 x2 - 1.00 x3 + 1.40 x1*x1 + 1.76 x2*x2 + 0.52 x3*x3 - 2.88 x1*x2 + 2.62 x1*x3 + 1.13 x2*x3

Fits and Diagnostics for Unusual Observations

Obs	y2	Fit	Resid	Std Resid	
3	62.00	71.99	-9.99	-3.02	R

R Large residual

Now, we obtain the optimum operating conditions of the process with respect to each response. For the first SOM, we have:

$$\mathbf{b}_1 = \begin{pmatrix} 0.393 \\ 0.562 \\ -0.270 \end{pmatrix} \qquad \mathbf{B}_1 = \begin{pmatrix} 2.418 & 0.375 & 0.125 \\ 0.375 & 1.181 & 0.125 \\ 0.125 & 0.125 & 2.595 \end{pmatrix}$$

In this case the stationary point is

$$\mathbf{x}_0^1 = -\frac{1}{2}\mathbf{B}_1^{-1}\mathbf{b}_1$$

But

$$\mathbf{B}_1^{-1} = \begin{pmatrix} 0.43552 & -0.13677 & -0.01439 \\ -0.13677 & 0.894028 & -0.03648 \\ -0.01439 & -0.03648 & 0.387807 \end{pmatrix}$$

$$\Rightarrow \mathbf{x}_0^1 = -\frac{1}{2}\begin{pmatrix} 0.43552 & -0.13677 & -0.01439 \\ -0.13677 & 0.894028 & -0.03648 \\ -0.01439 & -0.03648 & 0.387807 \end{pmatrix}\begin{pmatrix} 0.393 \\ 0.562 \\ -0.270 \end{pmatrix}$$

$$\Rightarrow \mathbf{x}_0^1 = -\frac{1}{2}\begin{pmatrix} 0.09818 \\ 0.458543 \\ -0.13086 \end{pmatrix}$$

$$\Rightarrow \mathbf{x}_0^1 = \begin{pmatrix} -0.04909 \\ -0.2292715 \\ 0.06543 \end{pmatrix}$$

That is, $\mathbf{x}_{0,1}^1 = -0.04909$, $\mathbf{x}_{0,2}^1 = -0.2292715$ and $\mathbf{x}_{0,3}^1 = 0.06543$. In terms of the natural variables, the stationary point is

$$-0.04909 = \frac{\pi_1^1 - 3}{0.5} \qquad -0.2292715 = \frac{\pi_2^1 - 43}{5} \qquad 0.06543 = \frac{\pi_3^1 - 14}{5}$$

which yields $\pi_1^1 = 2.975455$, $\pi_2^1 = 41.8536425$, and $\pi_3^1 = 14.32715$. The predicted response at the stationary point is $\hat{y}_0^1 = 47.087095$ and is gotten as follow.

$$\hat{y}_0^1 = 47.17 + \frac{1}{2}\begin{pmatrix} -0.04909 & -0.2292715 & 0.06543 \end{pmatrix}\begin{pmatrix} 0.393 \\ 0.562 \\ -0.270 \end{pmatrix}$$

$$\Rightarrow \hat{y}_0^1 = 47.17 + \frac{1}{2}(-0.16581)$$

$$\Rightarrow \hat{y}_0^1 = 47.087095$$

However, for the second SOM we have:

$$\mathbf{b}_2 = \begin{pmatrix} 2.44 \\ 1.76 \\ -1.00 \end{pmatrix} \qquad \mathbf{B}_2 = \begin{pmatrix} 1.400 & -1.440 & 1.310 \\ -1.440 & 1.760 & 0.565 \\ 1.310 & 0.565 & 0.520 \end{pmatrix}$$

In this case the stationary point is

$$\mathbf{x}_0^2 = -\frac{1}{2}\mathbf{B}_2^{-1}\mathbf{b}_2$$

But

$$\mathbf{B}_2^{-1} = \begin{pmatrix} -0.11045 & -0.27594 & 0.578071 \\ -0.27594 & 0.183121 & 0.496194 \\ 0.578071 & 0.496194 & -0.07235 \end{pmatrix}$$

$$\Rightarrow \mathbf{x}_0^2 = -\frac{1}{2}\begin{pmatrix} -0.11045 & -0.27594 & 0.578071 \\ -0.27594 & 0.183121 & 0.496194 \\ 0.578071 & 0.496194 & -0.07235 \end{pmatrix}\begin{pmatrix} 2.44 \\ 1.76 \\ -1.00 \end{pmatrix}$$

$$\Rightarrow \mathbf{x}_0^2 = -\frac{1}{2}\begin{pmatrix} -1.33323 \\ -0.8472 \\ 2.356147 \end{pmatrix}$$

$$\Rightarrow \mathbf{x}_0^2 = \begin{pmatrix} 0.666615 \\ 0.4236 \\ -1.1780735 \end{pmatrix}$$

That is, $\mathbf{x}_{0,1}^2 = 0.666615$, $\mathbf{x}_{0,2}^2 = 0.4236$ and $\mathbf{x}_{0,3}^2 = -1.1780735$. In terms of the natural variables, the stationary point is

$$-0.666615 = \frac{\pi_1^2 - 3}{0.5} \qquad 0.4236 = \frac{\pi_2^2 - 43}{5} \qquad -1.1780735 = \frac{\pi_3^2 - 14}{5}$$

which yields $\pi_1^2 = 3.3333075$, $\pi_2^2 = 45.118$, and $\pi_3^2 = 8.1096325$. The predicted response at the stationary point is $\hat{y}_0^2 = 81.915075$ and is gotten as follow.

$$\hat{y}_0^2 = 80.14 + \frac{1}{2}\begin{pmatrix} 0.666615 & 0.4236 & -1.1780735 \end{pmatrix}\begin{pmatrix} 2.44 \\ 1.76 \\ -1.00 \end{pmatrix}$$

$$\Rightarrow \hat{y}_0^2 = 80.14 + \frac{1}{2}(3.55015)$$

$$\Rightarrow \hat{y}_0^2 = 81.915075$$

Now, we obtain the mid-point of $\mathbf{x}_0^1$ and $\mathbf{x}_0^2$ to get $\bar{\mathbf{x}}_0$ as follows:

$$\bar{\mathbf{x}}_0 = \frac{1}{2}\left[\begin{pmatrix} -0.4909 \\ -0.2292715 \\ 0.06543 \end{pmatrix} + \begin{pmatrix} 0.666615 \\ 0.4236 \\ -1.1780735 \end{pmatrix}\right] = \begin{pmatrix} 0.3087625 \\ 0.09716425 \\ -0.55632175 \end{pmatrix}$$

In terms of the natural variables, the new stationary point $\bar{\mathbf{x}}_0$ is

$$-0.3087625 = \frac{\bar{\pi}_1 - 3}{0.5} \qquad 0.09716425 = \frac{\bar{\pi}_2 - 43}{5} \qquad -0.55632175 = \frac{\bar{\pi}_3 - 14}{5}$$

which yields $\bar{\pi}_1 = 3.15438125$, $\bar{\pi}_2 = 43.48582185$ and $\bar{\pi}_3 = 11.21839125$. But, in order to obtain the predicted response at the new stationary point $\bar{\mathbf{x}}_0$, we substitute for $\bar{\mathbf{x}}_0$ approximately to get optimal responses $\bar{\hat{y}}_0^1$ and $\bar{\hat{y}}_0^2$ which is averaged to get $\bar{\hat{y}}_0$ as follows.

$$\bar{\hat{y}}_0^1 = 47.17 + \frac{1}{2}\begin{pmatrix}0.3087625 & 0.09716425 & -0.55632175\end{pmatrix}\begin{pmatrix}0.393\\0.562\\-0.270\end{pmatrix}$$

$$\Rightarrow \bar{\hat{y}}_0^1 = 47.17 + \frac{1}{2}\left(0.326157\right)$$

$$\Rightarrow \bar{\hat{y}}_0^1 = 47.3330785$$

Also,

$$\bar{\hat{y}}_0^2 = 80.14 + \frac{1}{2}\begin{pmatrix}0.3087625 & 0.09716425 & -0.55632175\end{pmatrix}\begin{pmatrix}2.44\\1.76\\-1.00\end{pmatrix}$$

$$\Rightarrow \bar{\hat{y}}_0^2 = 80.14 + \frac{1}{2}\left(1.480711\right)$$

$$\Rightarrow \bar{\hat{y}}_0^2 = 80.8803555$$

Therefore,

$$\bar{\hat{y}}_0 = \frac{\bar{\hat{y}}_0^1 + \bar{\hat{y}}_0^2}{2} = \frac{47.3330785 + 80.8803555}{2} = 64.106717$$

3.2 Discussion

Based on our implementation of the proposed technique, we have made certain findings which have been summarized as axioms.

3.2.1 Axiom 1 – The optimal operating condition obtained via the proposed technique

Given that $y_i = f_i\left(\mathbf{x}\right) = \hat{\beta}_{0i} + \mathbf{x}^T\mathbf{b}_i + \mathbf{x}^T\mathbf{B}_i\mathbf{x}$ for $i = 1, 2, 3, \ldots, m$ SOMs having respective optimal responses $\hat{y}_0^i$ at each $\mathbf{x}_0^i$ optimal operating condition, then the mid-point $\bar{\mathbf{x}}_0$ of all $\mathbf{x}_0^i$ is a Pareto optimal operating condition.

3.2.2 Axiom 2 – The optimal yield obtained via the proposed technique

Given that $y_i = f_i(\mathbf{x}) = \hat{\beta}_{0i} + \mathbf{x}^T\mathbf{b}_i + \mathbf{x}^T\mathbf{B}_i\mathbf{x}$ for $i = 1, 2, 3, \ldots, m$ SOMs having respective optimal responses $\hat{y}_0^i$ at each $\mathbf{x}_0^i$ optimal operating condition, then the mid-point $\bar{y}_0$ of all $\bar{\hat{y}}_0^i$ (obtained at the mid-point $\bar{\mathbf{x}}_0$ of all $\mathbf{x}_0^i$) is a Pareto optimal yield.

3.2.3 Axiom 3 – Pareto front obtained via the proposed technique

The region bounded by each of the $\mathbf{x}_0^i$ optimal conditions based on each of the SOMs $y_i = f_i(\mathbf{x}) = \hat{\beta}_{0i} + \mathbf{x}^T\mathbf{b}_i + \mathbf{x}^T\mathbf{B}_i\mathbf{x}$ for $i = 1, 2, 3, \ldots, m$ is a Pareto front.

4 Conclusion

The findings in this article have shown that the proposed multivariate-based technique for solving MRSO problems is suitable for use under circumstances in which the decision maker's preference information is absent; hence, making it a more flexible and robust alternative for optimizing multiple response variables. More so, our results have shown that the optimal operating condition obtained via the proposed technique is Pareto optimal, having as its Pareto front, the region bounded by respective optimal operating conditions obtained from each SOM. The optimal response obtained is also Pareto optimal.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Alonzo AG, Hiroyuki N. Application of response surface methodology in the formulation of nutrient broth systems with predetermined PH and water activities. African Journal of Microbiology Research. 2009;45(2):634-639.

[2] Khuri AI, Mukhopadhyay S. Response surface methodology. Computational Statistics. 2010;2(2):23-34.

[3] Navidi HR, Amiri AH, Kamranrad R. Multi-responses optimization through game theory approach. International Journal of Industrial Engineering & Production Research. 2014;25(3):215-224.

[4] Ruben LL, Maria AMC, Carlos BL, Pedro JRF. Using the multi-response method with desirability functions to optimization the zinc electroplating of steel screws. Metals. 2018;8(71):1-20.

[5] Adinarayana K, Ellaiah O. Response surface optimization of the critical medium components for the production of alkaline phosphatase by newly isolated *Bacilus Sp.* J. Pharm. Pharmaceut. Sci. 2002;5(3):272-278.

[6] Akpan SS, Ugbe TA, Usen JE. An alternative procedure for solving second order response surface design problems. International Journal of Scientific and Engineering Research. 2013;4(9):2233-2245.

[7] Amineh Z, Kazem N. Considering preference parameters in multi-response surface optimization approaches. International Journal of Modeling and Optimization. 2011;1(2):158-162.

[8] Raissi S, Farsani R. Statistical process optimization through multi-response surface methodology. World Academy of Science, Engineering and Technology. 2009;51(3):267-271.

[9] Amayo T. Response surface methodology and its application to automative suspension designs. London: Oxford Press; 2010.

[10] Hakan A. Semet AA. Multi-response optimization application on a manufacturing factory. Mathematical and Computational Applications. 2013;18(3):531-538.

[11] Khuri AI. Response surface methodology and its applications in agricultural and food sciences. Biometrics and Biostatistics. 2017;5(5):3-11.

[12] Bradley DN. The response surface methodology. Indiana University of South Bend: Nikon Publishers; 2007.

[13] Montgomery DC. Design and analysis of experiment (8^{th} ed.). USA: John Wiley & Sons, Inc.; 2013.

[14] Andrew MC. Beyond rationality: Rigor without mortis in game theory. Behavioral and Brain Science. 2003;26(2):180-198.

[15] Gupta S, Manohar CS. An improved response surface method for the determination of failure probability and using instance measures. India: Elsevier Limited. 2003;123-139.

[16] Taha HH, Mahdi B, Jose AD, Kazem N. Optimization of probabilistic multiple response surfaces. Applied Mathematical Modeling. 2012;36:1275-1285.

[17] Mahdi B, Amir M, Seyed, TAN. Robust multi-response optimization: A posterior preferences approach. International Transactions in Operational Research. 2017:1-30.

[18] Harrington E. The desirability function. Industrial Quality Control. 1965;21(10):494-498.

[19] Plante RD. Process capability: A criterion for optimizing multiple response product and process design. IIE Transactions. 2001;33(2):497-509.

[20] Akteke-Öztürk B, Weber GW, Köksal G. Optimization of generalized desirability functions under model uncertainty. Optimization. 2017;66(12):2157-2169.

[21] Deringer G, Suich R. Simultaneous optimization of several response variables. Journal of Quality Technology. 1980;12(4):214-219.

[22] Khuri AI, Conlon JA. Simultaneous optimization of multiple responses represented by polynomial regression functions. Technometric. 1981;23:363-375.

[23] Hsiang TC, Taguchi G. A tutorial on quality control and assurance – the Taguchi methods. In ASA Annual Meeting. USA: Las Vegas; 1985.

[24] Kim KJ, Byun, JH, Min D, Jeong IJ. Multi-response surface optimization: Concept, methods and future directions. Korean Society for Quality Management. 2001;34(6):29-67.

[25] Myers RH, Carter WH Jr. Response surface techniques for dual response systems. Technometrics. 1973;15(2):301-317.

[26] Biles W. A response surface method for experimental optimization of multi-response processes. Industrial and Engineering Chemistry, Process Design and Deployment, 1975;14:152-158.

[27] Pignatiello, JJ Jr. Strategies for robust multi-response quality engineering. IIE Trans. 1993;25(5):5-15.

[28] Myers RH, Montgomery DC. Response surface methodology: Process and product optimization using designed experiments (2nd ed.). USA: Wiley & Sons, Inc.; 2002.

[29] Jeong IJ, Lee DH. A pairwise comparison-based interactive procedure for the process capability approach to multiple-response surface optimization. Engineering Optimization. 2019;1-18.

[30] Lee DH, Kim SH, Byun JH. A method of steepest ascent for multiresponse surface optimization using a desirability function method. Quality and Reliability Engineering International; 2020.

[31] Alvin CR. Methods of multivariate analysis (2nd ed.). USA: Wiley & Sons Inc.; 2002.

[32] Avril C. A little book of R for multivariate analysis. United Kingdom: Cambridge; 2017.

[33] Karl LW. An introduction to multivariate statistics. USA: Wiley & Sons Inc.; 2019.

[34] Tajbakhsh SD, Noorossana R. Important issues in multiple response optimization. Paper presented at the 4th International Management Conference, Canada; 2006. Available:http://www.canadian.managementconference.

[35] Taha HH, Mirmehdi SS, Majid R. New hybrid multivariate analysis approach to optimize multiple response surfaces considering correlations in both inputs and outputs. Acta Scientiarum Technology. 2014;36(3):469-477.

[36] Flavio SF. A survey of techniques for optimizing multi-response experiments. ENGEP. 1998;2(4):1-8.

[37] Chan LK, Cheng SW, Spiring, FA. A new measure of process capability: Cpm. Journal of Quality Technology. 1988;20(3):162-175.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/59570