Sains Malaysiana 51(5)(2022): 1577-1586
http://doi.org/10.17576/jsm-2022-5105-25

A New Optimization Scheme for Robust Design Modeling with Unbalanced Data

(Skema Pengoptimuman Baru bagi Pemodelan Reka Bentuk Teguh dengan Data Tak Seimbang)

ISHAQ BABA[1,2], HABSHAH MIDI[1,*], GAFURJAN IBRAGIMOV[1] & SOHEL RANA[3]

[1]Department of Mathematics and Statistics, Faculty of Science and Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia
[2]Department of Mathematical Sciences, Faculty of Science, Taraba State University Jalingo P.M.B. 1164 Jalingo, Taraba State, Nigeria
[3]Department of mathematical and Physical Sciences, East West University, Dhaka-1212, Bangladesh

Received: 5 February 2021/Accepted: 20 September 2021

ABSTRACT

The Lin and Tu (LT) optimization scheme which is based on mean squared error (MSE) objective function is the commonly used optimization scheme for estimating the optimal mean response in robust dual response surface optimization. The ordinary least squares (OLS) method is often used to estimate the parameters of the process location and process scale models of the responses. However, the OLS is not efficient for the unbalanced design data since this kind of data make the errors of a model become heteroscedastic, which produces large standard errors of the estimates. To remedy this problem, a weighted least squares (WLS) method is put forward. Since the LT optimization scheme produces a large difference between the estimates of the mean response and the experimenter actual target value, we propose a new optimization scheme. The OLS and the WLS are integrated in the proposed scheme to determine the optimal solution of the estimated responses. The results of the simulation study and real example indicate that the WLS is superior when compared with the OLS method irrespective of the optimization scheme used. However, the combination of WLS and the proposed optimization scheme (PFO) signify more efficient results when compared to the WLS combined with the LT optimization scheme.

Keywords: Optimization; robust design; unbalanced data; weighted least squares

ABSTRAK

Skema pengoptimuman Lin dan Tu (LT) yang berdasarkan fungsi objektif min kuasadua ralat (MSE) sering digunakan dalam skema pengoptimuman bagi menganggarkan min gerak balas optimum dalam pengoptimuman permukaan berganda teguh. Kaedah kuasadua terkecil biasa (OLS) sering digunakan untuk menganggarkan parameter model proses lokasi dan model proses skala bagi gerak balas. Walau bagaimanapun, kaedah OLS tidak cekap bagi data reka bentuk yang tak seimbang kerana data yang begini membuatkan ralat model menjadi heteroskedastik dan menghasilkan penganggar ralat piawai besar. Untuk mengatasi masalah ini, kaedah kuasadua terkecil berpemberat (WLS) dicadangkan. Kami mencadangkan skema pengoptimuman baru disebabkan skema pengoptimuman LT menghasilkan perbezaan yang besar antara penganggar min gerak balas dan nilai sebenar sasaran penyelidik. Kaedah OLS dan WLS digabungkan dalam skema yang dicadangkan bagi menentukan penyelesaian optimum bagi gerak balas yang dianggarkan. Keputusan kajian simulasi dan contoh sebenar menunjukkan bahawa kaedah WLS mengatasi kaedah OLS tanpa mengira skema pengoptimuman yang digunakan. Walau bagaimanapun, gabungan WLS dan skema pengoptimuman yang dicadang (PFO) menunjukkan keputusan yang lebih cekap apabila dibandingkan dengan WLS yang digabungkan dengan skema pengoptimuman LT.

Kata kunci: Data tak seimbang; kuasadua terkecil berpemberat; pengoptimuman; reka bentuk teguh

INTRODUCTION

Quality characteristics of customer interest must be well defined to design a qualitative product that satisfies customer requirement. These characteristics may include several variables such as production cost, product quality, production time, and production dimension and or any other important product characteristics specified by a customer. The application of robust design optimization is used to solve real life industrial problems. This attracts the attention of design expert to seek the best approach that can achieve the target of the experimenter with small response variation. A Japanese engineer known as Genich Taguchi in 1980's first introduced this method (Taguchi & Wu 1980). The purpose of robust design is to secure excellence performance and promotes product quality and procedures in industries such as engineering, biotechnology, chemical, and agriculture among others. The primary objective of robust design optimization is to find the best solution that can reduce response variation while achieving a specific output. Despite the application of Taguchi method in industries, the approach encountered a load of criticism from practitioners and researchers about its statistical and optimization techniques. For example, a panel of practitioners and researchers edited by Nair et al. (1992) discussed the implementation of Taguchi method. Other researchers including Kackar (1985), Vining and Myers (1990) and Myers and Montgomery (1995) highlighted a number of shortcomings attached to Taguchi method in robust design method. In view of the weaknesses of Taguchi method, Box and Draper (1987) developed response surface methodology (RSM) to find the optimal factor settings for the input variables which can either maximize or minimize the given response functions. The traditional method emphasizes on optimizing the mean response by assuming the variance of the response is constant. However, in real situation, this assumption may not be achieved. In this situation, both the mean and the standard deviation of the response should be considered when determining the optimum conditions for the input variables. Continuous research in this area leads to the development of dual response surface optimization (DRSO) proposed by Vining and Myers (1990), which attempts to optimize both the mean and standard deviation of the response. Lin and Tu (1995) pointed out that the Vining and Myers (1990) approach (VM) does not guarantee global optimal due to restriction of the constraint to a specific value. In this respect, Lin and Tu (1995) proposed LT optimization scheme, which is based on mean squared error (MSE) objective function that allows a small bias. The objective function consists of the bias and variance. The LT optimization scheme is the most popular scheme used in solving dual response and multiple response problems (Boylan & Choo 2013; Geothels & Cho 2011; Park & Cho 2003; Park & Leeds 2016). Park et al. (2017) used the LT optimization scheme with robust location and scale estimators of the response variables. Chelladurai et al. (2021) suggested using the RSM and the analysis of variance technique to optimize the process parameters in various manufacturing processes. However, as already mentioned, the RSM has its weakness whereby it only focused on optimizing the mean response and assuming that the variance of the response is constant which is not valid in practice. Kim and Lin (1998) pointed out that the objective function of LT does not take into account the measure of the violation of the constraint, and it places no restriction on how large the estimated mean value might deviate from the target value. The modification of MSE method called the WMSE was introduced by Ding et al. (2004) to further reduce the influence of variance by introducing some weight on the MSE optimization scheme. The idea behind this approach is to determine the weight in such a way that it can reduce the effect of the bias and the variance in determining the optimal setting conditions. The major shortcoming of this approach as mentioned in Jeong et al. (2005) is that it does not consider the interest of the decision maker in the determination of the weight functions. Hence, the estimated optimal mean response based on the WMSE will be affected particularly for unbalanced data. It is now evident that achieving equal number of replicates of the response variables at each designed point may be difficult when something goes incorrect during the experimental process. The problem may lead to collection of an unbalanced designed data at the end of the experiment, which normally violates the usual assumption of balanced design data. In such case, the ordinary least squares (OLS) method which is the commonly used method to estimate the parameters of multiple linear regression may not be a good estimation method due to the presence of unequal variation in the response observations that may give rise to heteroscedasticity problem (Midi et al. 2021; Rana et al. 2012). The problem affects the estimates of standard error (making it large), and inference problems such as test of hypothesis. To overcome this shortcoming, Cho and Park (2005) applied the weighted least squares (WLS) to estimate the parameters of the mean and variance model. Subsequently, Cho and Park (2005) also used LT optimization scheme to find the optimal setting condition that can achieve the target with small variance and unequal replications in the response. Although the LT optimization work well for experimental data, it can be a burden to process engineers when dealing with large number of experiments. Lee et al. (2018) noted that the LT optimization is not practical for large volume of operational data such as data from manufacturing lines. Then, Lee et al. (2018) proposed an optimization scheme based on data mining approach for operational data. However, in this paper, our study is limited to experimental data and also focus on LT optimization scheme in determining the optimal setting conditions as studied by many previous researchers since it is the widely used method. Nonetheless, the LT optimization scheme is suspected not to be that efficient since it does not achieve the target with minimal bias in the estimated mean response due to some little bias introduced in its formulation. The weakness of LT method has motivated us to propose a new optimization scheme based on penalty function and call it penalty function optimization (PFO) scheme. In this paper, we combine the PFO with the WLS method. The WLS is employed to rectify the problems of heteroscedasticity for unbalanced data and the PFO is used to offer substantial improvements over the LT method.

This paper is organized as follows. The development of the mean and the variance functions for multiple response problem is presented in the next section. The derivation of the weighted least squares method for balanced and unbalanced design data is described in the following section. Subsequently, the mathematical proof that the biases and variances of the proposed PFO is less than the LT optimization scheme is discussed. The simulation procedures and numerical example are presented in the subsequent section. The last section summarizes the conclusion of the study.

MATERIALS AND METHODS

DEVELOPMENT OF MEAN AND VARIANCE MODEL FOR MULTIPLE RESPONSE PROBLEMS

Suppose that an experimenter is interested in identifying influential control setting or study the effect of some control factors for a given experimental design problem. Consider a design system, where the characteristics of interest y depend upon a set of control variables $X = (x_1, \ldots, x_p)$ where p is the number of predictor variables. Let y_{ij} denotes the response variable at the ith design point ($i = 1,2,\ldots n$) and the j^{th} replicate, where $j = 1,2,\ldots m.$. Table 1 which is known as balanced data, gives the summary of the experimental format for multiple responses, where $X, \bar{y}_i$ and s_i^2 represent a set of control factor settings, sample mean and sample variance of the responses, respectively.

TABLE 1. Experimental format for multiple responses

Run	X	Replications			$\bar{y}_i$	s_i^2
1		y_{11}	$\cdots$	y_{1m}	$\bar{y}_1$	s_1^2
2		y_{21}	$\cdots$	y_{2m}	$\bar{y}_2$	s_2^2
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\bar{y}_3$	s_3
i		y_{i1}	$\cdots$	y_{im}	$\bar{y}_i$	s_i^2
$\vdots$		$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
n		y_{n1}	$\cdots$	y_{mn}	$\bar{y}_n$	s_n^2

From Table 1, the mean and variance estimators for each design point for $i = 1, 2, \ldots, n$ can be estimated using the following formulas:

$$\bar{y}_i = \frac{1}{m}\sum_{j=1}^{m} y_{ij} \tag{1}$$

and

$$s_i^2 = \frac{1}{m-1}\sum_{j=1}^{m}\left(y_{ij} - \bar{y}_i\right)^2 \tag{2}$$

The estimate of the mean and variance functions are then formulated based on the estimated means and estimated variances of the response variables. The OLS method is then applied to the mean and the variance functions. The general steps to obtain the optimal response are summarized as follows:

Step1: Given the set of control factors $X = (x_1, \ldots, x_p)$
Step 2: For each run, there are m responses
Step 3: Compute the mean and the variance s_i^2 of the response variables y_{ij} for $i = 1, 2, \ldots$,n, j = 1, 2, …, m

Step 4: Determine the fitted mean and the fitted variance response functions by regressing y_i against the control factors X and regressing s^2_i against the control factors X using the OLS method.
Step 5: Determine the optimal setting condition $X = (x^*_{i1}, \ldots, x^*_{ip})$
Step 6: Determine the estimated optimal mean response by substituting the optimal settings $X = (x^*_{i1}, \ldots, x^*_{ip})$ into the obtained fitted mean and variance functions in Step 4.

PROPOSED OPTIMIZATION SCHEME FOR DUAL RESPONSE FUNCTION

In this section, a dual response surface optimization technique based on the penalty function approach is derived, specifically for our proposed PFO method. The penalty function method convert a series of constrained optimization into unconstrained optimization problem whose optimum solution are also true solution of the formulated function and the original objective function. The unconstrained objective function is formulated by adding a penalty parameter to the real objective function which includes the penalty term multiply by the measure of violation of the constraints equation (Dong 2006; Shin et al. 1990; Wan et al. 2009). To derive the objective function, we first consider the general form of the constrained optimization problem defined by

Minimize $f(x)$

$$\text{Subject to } g_j(x) \le 0, j = 1, 2, \ldots, m \qquad (3)$$
$$h_i(x) = 0, i = 1, 2, \ldots, n$$

By applying the method of penalty function, one can determine the solution of Equation (3) from the following objective function

$$F(x) = f(x) + \sum_{j=1}^{m} \mu\left(g_j(x)\right) + \sum_{i=1}^{n} \mu\left(h_i(x)\right) \quad (4)$$

where $f(x)$ is the original objective function to be minimized, h_i and g_j are set of the inequality and equality constraints function, respectively. This article, specifically focuses on a quadratic penalty function defined by;

$$F(x, \mu) = f(x) + \frac{\mu}{2} \sum_{i=1}^{n} \left(h_i(x)\right)^2 \qquad (5)$$

where μ is called the penalty parameter that penalizes the equality constraints when the constraints relations are not satisfied. For simplicity, we substitute $h_i = [\widehat{m}(x) - T]$ and $f(x) = \hat{v}(x)$ in Equation (5) and then write the following quadratic unconstrained minimization problem as;

$$\min F(x\ \mu,) = \hat{v}(x) + \frac{\mu}{2}[\widehat{m}(x) - T]^{2} \qquad (6)$$

where $\widehat{m}(x)$ is the estimated mean response, $\hat{v}(x)$ is the estimated variance response surface function and T is the desired target selected by the decision maker. If $\mu = \infty$ the solution of (6) is exact. For the quadratic function in Equation (6), we may use nonlinear optimization package in any programming software to determine the optimal setting condition of the estimated mean response. In this paper, a statistical software package in R programming Language (Rsolnp) introduced in Ghalanos and Theussl (2012) and Ye (1989) are used to compute the approximate optimum solution for the newly formulated and existing optimization schemes. Our goal is to find the optimum value so that the estimated mean response will be equal to or varies to the target value T, while making the variance small. The optimization scheme proposed is superior over the existing approaches. Firstly, the proposed approach takes into account the measure of violation of the constrained function, while LT optimization functions are minimized without given concern to the effect of the violation of the constrained. Secondly, the penalty term in Equation (6) pushes $\widehat{m}(x) - T$ to be equal or close to zero so that it can achieve the target of the decision maker. These two properties make the proposed method more efficient and reliable to estimate the mean and the variance functions for approximation of the optimum settings conditions. In the next section, we will prove that the PFO method produces smaller variance than the LT method.

THE WEIGHTED LEAST SQUARES AND OPTIMIZATION PROCEDURES

In the presence of heteroscedasticity in the response observations, the usual assumption of constant variance in linear regression model may not hold in many real life problems. The OLS regression assumed that the observed response Y come from a normal distribution function with mean $X\beta$ and variance $\sigma^2 I$ where X is a data matrix of containing the predictor variables, β is a vector of estimated coefficients, and I is an identity matrix. The summary of the OLS assumption can be express as:

$$y \sim N(X\beta, \sigma^2 I)$$

Suppose that the distribution of the errors have non constant variance whose constant of proportionality k_i are known, then the variance of regression errors $\varepsilon_1, \ldots, \varepsilon_n$ is given as:

Var(ε) = $k_i\,\sigma^2$, for i = 1,..., n, in matrix form is given as; $y \sim N(X\beta,\sigma^2 V)$ where V = diag[c_1, ...c_n]. Therefore we can express

$$\varepsilon = y - X\beta \sim N(0, \sigma^2 V) \tag{7}$$

Applying the symmetric diagonal matrix, $W^{\frac{1}{2}} =$ daig$\left[1/\sqrt{c_1}, \ldots, 1/\sqrt{c_n}\right]$, note that $W = W^{\frac{1}{2}}W^{\frac{1}{2}} = V^{-1}$, multiplying both side of equation (7) by $W^{\frac{1}{2}}$ we have

$$W^{\frac{1}{2}}\varepsilon = W^{\frac{1}{2}}(y - X\beta) \sim N(0, \sigma^2 V) \tag{8}$$

For convenience, we rewrite equation (8) as:

$$W^{\frac{1}{2}}y = W^{\frac{1}{2}}X\beta + W^{\frac{1}{2}}\varepsilon$$

This can be express as:

$$y_{wsl} = X_{wsl}\,\beta_{wsl} + \varepsilon_{wsl} \tag{9}$$

Equation (8) is equivalent to the standard regression model $y = X\beta + \varepsilon$. Applying the ordinary least squares (OLS) method, the transformed model in Equation (8) becomes:
$\beta_{wsl} = (X'_{wls}X_{wsl})^{-1}X_{wsl}y_{wsl}$, where $y_{wsl} = W^{\frac{1}{2}}, X_{wsl} = W^{\frac{1}{2}}X$,

$$\hat{\beta}_{wls} = \left[(W^{\frac{1}{2}}X)'W^{\frac{1}{2}}X\right]^{-1}(W^{\frac{1}{2}}X)'W^{\frac{1}{2}}y$$

$$= \left(X'W^{\frac{1}{2}}W^{\frac{1}{2}}X\right)^{-1}X'W^{\frac{1}{2}}W^{\frac{1}{2}}y \tag{10}$$

$$= (X'WX)^{-1}X'Wy$$

Incorporating the WLS in Equation (9) into robust design approach, we developed the fitted response functions for the process mean and variance as:

$$\widehat{m}(x) = X\hat{\beta}_\mu \tag{11}$$

$$\hat{V}(x) = X\hat{\beta}_\sigma\,, \tag{12}$$

where $\hat{\beta}_\mu = (X'WX)^{-1}X'W\bar{y}_i$

and $\hat{\beta}_\sigma = (X'WX)^{-1}X'Ws_i^2$

Furthermore, this approach work effectively when the number of replications is complete for each design value, but it may not be suitable for unbalanced data when the sample sizes are not equal. Hence, an alternative technique which choose weight based on the sample sizes at each design point were suggested by Goethals and Cho (2011) and Cho and Park (2005). They defined weights for mean and variance as W_m = diag[m_1, ..., m_n], W_v = diag[m_1-1, ..., m_n-1] where (m_1, ..., m_n) is the number of observation at each design point. More so, applying the formulated models for the mean and variance, the MSE objective function used in the LT optimization scheme is defined as:

$$\text{MSE} = \left\{[m(x_n) - T]^2 + v(x_n)\right\} \tag{13}$$

such that $x_i \in [L_j, U_j]$ for $j = 1, \ldots, k,$

where T is the specified target value, usually selected according to the quality characteristic of interest of the experimenter and x_j is the experimental region of the factorial designs with k levels. However, the mean squared error model introduced some bias and it does not take into account how large the estimated mean response should deviate from the actual specify target output (Copeland & Nelson 1996). This shortcoming often results to large difference between the estimated mean response and the target output that may lead to misleading conclusion. Therefore, we employ the weighted least squares (WLS) combined with the newly proposed PFO optimization scheme to obtain the optimal setting of the estimated mean response. For convenient, we re-write the newly propose optimization scheme (PFO) as follows:

$$\min\left\{\tfrac{\mu}{2}[m(x_n) - T]^2 + v(x_n)\right\} \tag{14}$$

Equations (13) and (14) can be used to find the optimal setting condition that optimizes the fitted mean and variance functions while achieving the target output. The PFO optimization scheme is based on the following fact:

Let $x_n \to x_0$ as $n \to \infty$.
If $m(x_n) \to T, v(x_n) \to v_0$ as $n \to \infty$
where v_0 is the minimum of $v(x_n)$
Then, clearly $\left\{[m(x_n) - T]^2 + v(x_n)\right\} \to v_0$ as $x_n \to \infty$.

Also, $\left\{\tfrac{\mu}{2}[m(x_n) - T]^2 + v(x_n)\right\} \to v_0$ as $x_n \to \infty$ since the first term approaches zero, where μ is any positive number. The parameter μ plays a key role to manage the convergence rate of limit. Therefore, we consider the function (14) instead of (13) because it will tend to produce less bias and less variability.

SIMULATION STUDY

In this section, we consider two simulation studies. The first objective of the simulation study is to show that our proposed PFO method is more efficient than some existing methods in this study. As per Midi and Aziz (2019), Lee et al. (2007) and Park and Cho (2003) simulation studies, at each control factor setting $x_i = (x_{i1}, x_{i2}, x_{i3})$, $i = 1, \ldots, 27$, five responses (y_{i1}, y_{i2}, y_{i5}) are randomly generated from a normal distribution with mean, $m(x_i) = 500 + (x_1 + x_2 + x_3)^2 + x_1 + x_2 + x_3$ and variance, $v(x_i) = 100 + (x_1 + x_2 + x_3)^2 + x_1 + x_2 + x_3$ where the target value, T equals 500. The VM, LT, WMSE, and PFO were then applied to the data and we considered 1000 simulation runs. The estimated mean of the optimal mean response, computed over m iterations are given by $\bar{\hat{\mu}} = \sum_{i=1}^{m}(\hat{\mu}/m)$, Bias $= \bar{\hat{\mu}} - 500$, and $var(\hat{\mu}) = \sum_{i=1}^{m}(\hat{\mu} - \bar{\hat{\mu}})^2/m$. The mean squared error (MSE) is written as MSE $(\mu) = (\text{Bias})^2 + \text{var}(\mu)$. The results are exhibited in Table 2. It is interesting to observe that the proposed PFO is the most efficient method as it has the smallest bias, SE, and RMSE, followed by the LT, WMSE and VM methods. Hence, we will consider the PFO and the LT in the next objective.

TABLE 2. Bias, standard error (SE) and RMSE of the optimal mean response

Method	Bias	SE	RMSE
VM	1.30	4.41	4.52
LT	1.46	3.73	4.01
WMSE	1.47	3.78	4.06
PFO	0.07	0.23	0.24

The second objective of the simulation study is to show that the PFO combined with the WLS methods is more efficient than the LT combined with the WLS methods for unbalanced data. The VM and the WMSE were not considered in this study because both methods do not perform very well based on the results of the first objective. Moreover, as already mentioned in the introduction section, these methods have some drawbacks. Following the simulation technique developed by Cho and Park (2005) and Park and Cho (2003), 3^2 factorial design with two factors and 3 levels represented by the digits (-1, 0, +1) were considered. The responses $(y_{i1}, y_{i2}, \ldots, y_{im})$ are generated from a normal distribution with mean $m(x_i) = 50 + 10(x_1^2 + x_2^2)$ and variance $v(x_i) = 100 + 25(x_1^2 + x_2^2)$. For each factor settings $x_i = (x_{i1}, x_{i2})$ for $i = 1,2,\ldots,9$, the response variables are replicated for a specific number represented in a circle as shown in the simulation scheme given in Figure 1. The customer target value is assumed to be 50, i.e. T = 50 and a total of 1000 iteration were considered.

For each of the four simulation scheme in Figure 1, we can compute the bias, variance and mean squares error (MSE) of the optimum mean response in order to evaluate the performance of the OLS and WLS based on LT optimization function and compared the results based on our proposed PFO optimization scheme given in Equation (14). Tables 3 and 4 reported the estimated bias, variance and MSE of the optimum mean response, based on LT and PFO optimization function, respectively. The results of Tables 3 and 4 show that the WLS method is superior to the OLS method irrespective of the optimization scheme used; LT or PFO. However, it is interesting to see that by comparing the results of Tables 3 and 4, the WLS based on our proposed (PFO) method is better than the WLS based on the LT method evident by having smaller values of bias, SE, and RMSE. Due to space constrained, in this paper we only consider 3^3 (first simulation) and 3^2 (second simulation) factorial designs. It is important to note that factorial design with different factors and levels may be considered. However, the results are consistent.

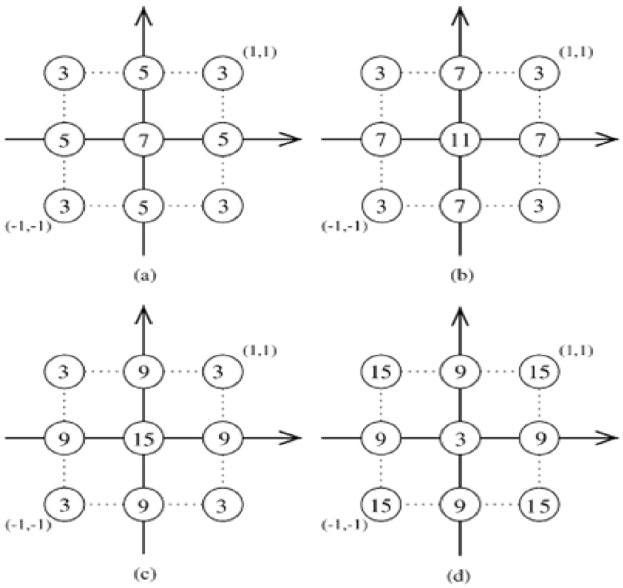

FIGURE 1. Simulation Scheme of Cho and Park (2005)

TABLE 3. Estimated bias, variance and MSE of the optimal mean response for the various simulation scheme with LT

Scheme	OLS			WLS		
	Bias	SE	RMSE	Bias	SE	RMSE
(a)	3.24	5.55	16.94	3.20	4.86	15.13
(b)	3.01	4.02	13.14	2.97	3.80	12.59
(c)	2.17	3.35	11.62	2.53	2.76	9.18
(d)	2.39	5.92	11.61	2.36	4.97	10.55

TABLE 4. Estimated bias, variance and MSE of the optimal mean response for the various simulation scheme with PFO

Scheme	OLS			WLS		
	Bias	SE	RMSE	Bias	SE	RMSE
(a)	1.19	2.85	4.26	1.17	2.63	4.01
(b)	1.08	2.41	3.58	0.96	1.76	2.70
(c)	0.94	1.70	2.58	0.86	1.43	2.16
(d)	1.55	4.35	6.77	1.20	1.20	4.89

NUMERICAL EXAMPLE

Considering the case study data reported in Cho and Park (2005), suppose that an injection modeling company is given the responsibility of producing silicon wafers for Motor Corporation. Tables 5 reported the experimental data obtained from the development of silicon wafers, where $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ are factors variable representing the mold temperature and injection flow rate and the response $\boldsymbol{y_{ij}}$ represent the coating thickness of the wafers with target value of T = 50, respectively. The weight for the mean and variance function obtained from the data based on $\boldsymbol{m_i}$ and $\boldsymbol{m_i}$**-1** is given by $\boldsymbol{W_m}$ = diag[3, 5, 3, 5 7, 5, 3, 5, 3] and $\boldsymbol{W_v}$ = diag[2, 4, 2, 4, 6, 2, 4, 2]. The estimated coefficients of the fitted mean response function, $\hat{m}(x)$ and the fitted variance response function, $\hat{v}(x)$ using OLS and WLS methods are exhibited in Table 6. The OLS and WLS with LT optimization scheme and the newly proposed PFO optimization scheme were applied to the data. The results of the OLS method based on LT optimization and WLS method based on our proposed PFO optimization scheme are presented in Table 7. It can be observed from Table 7 that the $\boldsymbol{WLS_{LT}}$ is more efficient compared to $\boldsymbol{OLS_{LT}}$ which have smaller bias, variance, and RMSE. Nevertheless, it is interesting to observe that our proposed $\boldsymbol{WLS_{PFO}}$ is superior and more reliable compared to $\boldsymbol{WLS_{LT}}$.

TABLE 5. Case study data from Cho and Park (2005)

index	x_1	x_2	y_{ij}							$\bar{y}$	s_i^2
1	1	-1	84.30	57.00	56.50					65.93	253.06
2	0	-1	75.70	87.10	71.80	43.80	51.60			66.00	318.28
3	1	-1	65.59	47.90	63.30					59.03	94.65
4	1	0	51.00	60.10	69.70	84.80	74.74			68.06	170.35
5	0	0	53.10	36.20	61.80	68.60	63.40	48.60	42.5	53.46	139.89
6	1	0	46.50	65.90	51.80	48.40	64.40			55.40	83.11
7	1	1	65.70	79.80	79.10					74.87	63.14
8	0	1	54.40	63.80	56.2	48.00	64.50			57.38	47.54
9	1	1	50.70	68.30	62.90					60.63	81.29

TABLE 6. Estimated coefficients of the fitted mean response function, and the fitted variance response function, using OLS and WLS methods

Coefficients	OLS		WLS	
	$\hat{m}(x)$	$\hat{v}(x)$	$\hat{m}(x)$	$\hat{v}(x)$
$\hat{\beta}_0$	55.61	160.65	55.08	154.26
$\hat{\beta}_1$	-5.63	-37.92	-5.76	-39.34
$\hat{\beta}_2$	0.32	-79.00	-0.52	-93.09
$\hat{\beta}_{11}$	5.04	-44.30	5.51	-38.31
$\hat{\beta}_{22}$	5.00	11.88	5.47	17.87
$\hat{\beta}_{12}$	-1.84	44.14	-1.84	44.14

TABLE 7. Estimated optimum settings, Bias, Variance and RMSE of the optimum mean response with OLS and WLS methods

Method	Optimal Settings	Bias	Var	RMSE
OLS_{LT}	(1.00, 1.00)	8.45	55.45	11.30
OLS_{PFO}	(1.00,0.94)	7.98	56.12	10.96
WLS_{LT}	(1.00, 1.00)	7.93	45.66	10.42
WLS_{PFO}	(1.00,0.99)	7.89	45.59	10.39

CONCLUSION

The problem of unbalanced design data lead to non-constant error variances in the response observation. The method of least squares combined with LT optimization scheme ($\boldsymbol{OLS_{LT}}$) gives less efficient results in solving these problems, while the integration of WLS concepts into robust design with LT optimization ($\boldsymbol{WLS_{LT}}$) has improved the estimated mean response. However, the overall results show that incorporating the WLS with our proposed PFO optimization scheme ($\boldsymbol{WLS_{PFO}}$) provide the most efficient results as it has the smallest bias, SE, and RMSE. Hence, the newly proposed scheme can be a good alternative method in dealing with robust design optimization problem with unbalanced data points.

REFERENCES

Box, G.E. & Draper, N.R. 1987. *Empirical Model-Building and Response Surfaces*. New York: John Wiley & Sons.

Boylan, G.L. & Cho, B.R. 2013. Comparative studies on the high-variability embedded robust parameter design from the perspective of estimators. *Computers & Industrial Engineering* 64(1): 442-452.

Chelladurai, S.J.S., Murugan, K., Ray, A.P., Upadhyaya, M., Narasimharaj, V. & Gnanasekaran, S. 2021. Optimization of process parameters using response surface methodology: A review. *Materials Today: Proceedings* 37: 1301-1304.

Cho, B.R. & Park, C. 2005. Robust design modeling and optimization with unbalanced data. *Computers & Industrial Engineering* 48(2): 173-180.

Copeland, K.A. & Nelson, P.R. 1996. Dual response optimization via direct function minimization. *Journal of Quality Technology* 28(3): 331-336.

Ding, R., Lin, D.K. & Wei, D. 2004. Dual-response surface optimization: A weighted MSE approach. *Quality Engineering* 16(3): 377-385.

Dong, S. 2006. *Methods for Constrained Optimization*. Massachusetts: Massachusetts Institute of Technology.

Ghalanos, A. & Theussl, S. 2012. Rsolnp: General non-linear optimization using augmented Lagrange multiplier method. *R package version*, 1.

Goethals, P.L. & Cho, B.R. 2011. Solving the optimal process target problem using response surface designs in heteroscedastic conditions. *International Journal of Production Research* 49(12): 3455-3478.

Jeong, I.J., Kim, K.J. & Chang, S.Y. 2005. Optimal weighting of bias and variance in dual response surface optimization. *Journal of Quality Technology* 37(3): 236-247.

Kackar, R.N. 1985. Off-line quality control, parameter design, and the Taguchi method. *Journal of Quality Technology* 17(4): 176-188.

Kim, K.J. & Lin, D.K.J. 1998. Dual response surface optimization: A fuzzy modeling approach. *Journal of Quality Technology* 30(1): 1-10.

Lee, S.B. & Park, C. 2006. Development of robust design optimization using incomplete data. *Computers & Industrial Engineering* 50(3): 345-356.

Lee, D.H., Yang, J.K. & Kim, K.J. 2018. Dual-response optimization using a patient rule induction method. *Quality Engineering* 30(4): 610-620.

Lin, D.K. & Tu, W. 1995. Dual response surface optimization. *Journal of Quality Technology* 27(1): 34-39.

Myers, R.H. & Montgomery, D.C. 1995. *Response Surface Methodology: Process and Product Optimization using Designed Experiments*. https://doi.org/10.2307/1270613

Midi, H. & Aziz, N.A. 2019. High breakdown estimator for dual response optimization in the presence of outliers. *Sains Malaysiana* 48(8): 1771-1776.

Midi, H., Ismaeel, S.S., Arasan, J. & Mohammed, M.A. 2021. Simple and fast generalized-M (GM) estimator and its application to real data set. *Sains Malaysiana* 50(3): 859-867.

Nair, V.N., Abraham, B., MacKay, J., Box, G., Kacker, R.N., Lorenzen, T.J., Lucas, J.M., Myers, R.H., Vining, G.G., Nelder, J.A. & Phadke, M.S. 1992. Taguchi's parameter design: A panel discussion. *Technometrics* 34(2): 127-161.

Park, C. & Cho, B.R. 2003. Development of robust design under contaminated and non-normal data. *Quality Engineering* 15(3): 463-469.

Park, C. & Leeds, M. 2016. A highly efficient robust design under data contamination. *Computers & Industrial Engineering* 93: 131-142.

Park, C., Ouyang, L., Byun, J.H. & Leeds, M. 2017. Robust design under normal model departure. *Computers & Industrial Engineering* 113: 206-220.

Rana, S., Midi, H. & Imon, A.R. 2012. Robust wild bootstrap for stabilizing the variance of parameter estimates in heteroscedastic regression models in the presence of outliers. *Mathematical Problems in Engineering* 2012: 730328.

Shin, D.K., Gürdal, Z. & Griffin, O.H.J. 1990. A penalty approach for nonlinear optimization with discrete design variables. *Engineering Optimization+ A35* 16(1): 29-42.

Taguchi, G. & Wu, Y.I. 1980. *Introduction to Off-Line Quality Control Systems*. Nagoyo: Central Japan Quality Control Association.

Vining, G.G. & Myers, R.H. 1990. Combining taguchi and response surface philosophies: A dual response approach. *Journal of Quality Technology* 22(1): 38-45.

Wan, Z., Zhang, S.J. & Wang, Y.L. 2009. Penalty algorithm based on conjugate gradient method for solving portfolio management problem. *Journal of Inequalities and Applications* 2009: Article ID. 970723.

Ye, Y. 1989: Solnp users' guide -a nonlinear optimization program in matlab. Department of Management Sciences, College of Business Administration, University of Iowa.

*Corresponding author; email: habshahmidi@gmail.com