Sains Malaysiana 53(4)(2024): 921-934
http://doi.org/10.17576/jsm-2024-5304-15

Dual Response Surface Optimization Based on Skill Scores
(Pengoptimuman Permukaan Tindak Balas Dwi Berdasarkan Skor Kemahiran)

AGAH KOZAN, MELIS ZEYBEK & ELIF KOZAN*

Department of Statistics, Faculty of Science, Ege University, Turkey

Received: 12 April 2023/Accepted: 14 March 2024

ABSTRACT

The popular formulations of dual-response optimization are constructed on minimizing a function of bias and system variability. This study provides an opportunity to evaluate the dual response surface (DRS) problem from a different perspective by adapting two new terms such that internal and external quality forecasts. The background of the proposed approach focuses on the relationship between internal and external quality forecasts and discusses the DRS problem in regards of skill scores by defining a model quality criterion. Skill is the relative accuracy of the forecast and defines a correspondence between forecast of interest and reference forecasts. The reference forecast does not require any knowledge or modelling; thus, it is an unskilled forecast. In this context, skill score is a measure of this relative improvement and widely used in evaluating the performance of operational and experimental forecasts. An alternative version of mean square error (MSE) which is reconstructed by skill scores and model quality criterion is proposed as an objective function for the DRS problem. Integrating the relationship between internal and external quality forecasts into such a response function can improve the effectiveness and cooperation of the applied technique. The proposed approach has a flexible structure and provides decision makers alternative solutions for different values of the model quality criterion. The proposed procedure is discussed by conducted a simulation study and demonstrated in an engineering process.

Keywords: Dual response optimization; mean square error; model quality criterion; robust parameter design; skill scores

ABSTRAK

Formulasi popular pengoptimuman gerak balas dual dibina untuk meminimumkan fungsi bias dan kebolehubahan sistem. Kajian ini memberi peluang untuk menilai masalah permukaan gerak balas dual (DRS) dari perspektif yang berbeza dengan menyesuaikan dua istilah baharu seperti ramalan kualiti dalaman dan luaran. Latar belakang pendekatan yang dicadangkan memfokuskan pada hubungan antara ramalan kualiti dalaman dan luaran dan membincangkan masalah DRS dalam hal skor kemahiran dengan mentakrifkan kriteria kualiti model. Kemahiran ialah ketepatan relatif ramalan dan mentakrifkan perpadanan antara ramalan kepentingan dan ramalan rujukan. Ramalan rujukan tidak memerlukan sebarang pengetahuan atau pemodelan; oleh itu, ia adalah ramalan yang tidak mahir. Dalam konteks ini, skor kemahiran adalah ukuran peningkatan relatif ini dan digunakan secara meluas dalam menilai prestasi ramalan operasi dan uji kaji. Versi alternatif bagi ralat min kuasa dua (MSE) yang dibina semula oleh skor kemahiran dan kriteria kualiti model dicadangkan sebagai fungsi objektif untuk masalah DRS. Mengintegrasikan hubungan antara ramalan kualiti dalaman dan luaran ke dalam fungsi tindak balas sedemikian boleh meningkatkan keberkesanan dan kerjasama teknik yang digunakan. Pendekatan yang dicadangkan mempunyai struktur yang fleksibel dan menyediakan penyelesaian alternatif pembuat keputusan untuk nilai yang berbeza bagi kriteria kualiti model. Prosedur yang dicadangkan dibincangkan dengan menjalankan kajian simulasi dan ditunjukkan dalam proses kejuruteraan.

Kata kunci: Kriteria kualiti model; pengoptimuman gerak balas dual; ralat min kuasa dua; reka bentuk parameter teguh; skor kemahiran

INTRODUCTION

Robust parameter design (RPD), introduced by Taguchi (1986), is a methodology that involves the use of efficient experimental designs and modeling of factors that affect the quality of a system. RPD, along with Taguchi's philosophy, has received considerable attention for more than thirty years in different industrial fields. However, his experimental methodology and analysis techniques have been exposed to a lot of criticism from the statistical community – e.g., Box (1985), and Vining and Myers (1990). Consequently, new methodologies have been proposed to address these drawbacks.

Response surface methodology (RSM), first developed by Box and Wilson (1951), was revisited in the early 1990s as an effective approach to RPD and has since received much attention from researchers. RSM defines a relationship between a quality characteristic and design factors, and the resulting relationship is then exploited to find optimal operating conditions. DRS approach, proposed by Vining and Myers (1990), brought a new perspective to off-line quality and many researchers focused on improving alternative optimization tools for the DRS problem. Most of the existing approaches are focused on providing a regular framework which optimal solution of the DRS problem can be determined. This attempt is typically done by minimizing a function of bias and system variability. These novel optimization approaches have become sound and widely quoted in the current literature.

This study provides an opportunity to evaluate the DRS problem from a different perspective by adapting two new terms such that *internal* and *external quality forecasts*. The background of the proposed approach focuses on the relationship between internal and external quality forecasts and discusses the DRS problem in regards of *skill scores* by defining a *model quality criterion*. From the perspective of robust design optimization, this relationship determines how much one can reach the process requirements with the fitted response models. Therefore, the model quality criterion can be defined as a degree of the sensitive of the fitted response surfaces to the process requirements and can be used to adjust the measure of the model performance for the DRS problem.

The proposed approach presents an alternative version of the MSE criterion which is reconstructed by integrating skill scores. The proposed approach establishes a degree of the closeness of internal and external quality forecasts and determines the best operating conditions with a restriction on the model quality criterion. Therefore, the proposed approach can set how much of the quality requirements can be met by adjusting the restriction on the bias. In this context, the decision-maker should decide which value of the model quality criterion fit the capability of his/her technology. Consequently, it generates more alternative solutions and provides adjustability to the decision-maker. With this respect, it differs from the existing approaches in the current literature.

The remainder of this manuscript is structured as follows: In the next section, the DRS problem is overviewed, followed by a brief review of the skill scores. In the following section, the proposed approach is presented. All findings are illustrated based on a simulation study and on the well-known printing process study, before the manuscript finally ends with a conclusion.

AN OVERVIEW ON THE DRS PROBLEM

Vining and Myers (1990), who are the pioneers of this field, conducted one of the earliest research attempts to develop an alternative tool for off-line quality. They discussed a procedure, called DRS, constructed on combining RSM and some effective properties of Taguchi's RPD. The DRS approach is configured by separately fitting the system mean and variance response surfaces.

Let $\hat{\mu}(x)$ and $\hat{\sigma}(x)$ are the fitted system mean and standard deviation responses, respectively. Generally, $\hat{\mu}(x)$ and $\hat{\sigma}(x)$ can be modelled as follows:

$$\hat{\mu}(x) = \hat{\gamma}_0 + \sum_{i=1}^{k} \hat{\gamma}_i x_i + \sum_{i=1}^{k} \hat{\gamma}_{ii} x_i^2 + \sum_{i<t}^{k} \sum \hat{\gamma}_{it} x_i x_t \quad (1)$$

and

$$\hat{\sigma}(x) = \hat{\delta}_0 + \sum_{i=1}^{k} \hat{\delta}_i x_i + \sum_{i=1}^{k} \hat{\delta}_{ii} x_i^2 + \sum_{i<t}^{k} \sum \hat{\delta}_{it} x_i x_t \quad (2)$$

where $\hat{\gamma} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{w}_{\mu}$ and $\hat{\delta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{w}_{\sigma}$. The vectors of the sample mean and standard deviation, $\mathbf{w}_{\mu} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n)'$ and $\mathbf{w}_{\sigma} = (s_1, s_2, \dots, s_n)'$, are obtained by using sample point estimators for each design point. $\mathbf{X}$ denotes the design matrix.

In the dual-response optimization, one response is chosen as a primary response to be optimized subject to a pre-defined value of the other. i.e., secondary response.

DRS meet the fundamental situations of RPD such as larger-the-better (LTB), smaller-the-better (STB), and nominal-the-best (NTB). The optimization procedure for the NTB case is constructed on minimizing $\hat{\sigma}(x)$ while keeping $\hat{\mu}(x)$ at the target value. In the case of LTB/STB, one seeks to maximize/minimize $\hat{\mu}(x)$ while controlling $\hat{\sigma}(x)$ at a pre-defined value. The regular DRS optimization assigns the best operating conditions subject to an additional constraint ($\boldsymbol{x}^* \in R$) which defines the experimental region, i.e., $-1 \leq x_i \leq 1$, $i = 1 \ldots . k$ for cuboidal designs and $\boldsymbol{x'x} \leq \rho^2$ for spherical designs, where ρ is the design radius.

Further improvement for the DRS problem was carried out by Del Castillo and Montgomery (1993) which presents an alternative approach using generalized reduced gradient algorithm. In this procedure, the system responses do not have to be modelled by second order polynomials, and the experimental region could be either cuboidal or spherical. Due to its flexibility, the GRD solution sometimes performs better than Vining and Myers (1990)'s approach. Kim and Lin (1998) discussed the DRS problem from a different perspective and stated that the classical approaches can be misleading in nonlinear and asymmetric cases, since RSM give the linear change of the degree of satisfaction. They proposed a fuzzy modeling approach to tackle the same problem. This method offers a balance between the distance from the target and variability due to allow the modeling a decision maker's preferences on the estimated responses. On the other hand, Köksoy and Doganaksoy (2003) claimed that the traditional techniques might be very useful for finding 'one-shot' optimum solutions, however failing to interpret the trade-offs between the mean and variance. They proposed an alternative formulation based on joint optimization of the system mean and standard deviation responses by generating Pareto optimal solutions under no constraints/minimally constrained.

The idea of relaxing zero bias assumption, which is configured under the assumption that estimated system mean can be located far from the target to achieve a smaller variance, is handled by Copeland and Nelson (1996) and Lin and Tu (1995). Lin and Tu (1995) presented the MSE criterion for the DRS problem. Their approach considers the distance of the estimated mean response from the target, i.e., bias, along with the variability and aims minimizing the MSE criterion to determine the optimal factors setting of a given system. The MSE criterion-based optimization approach has some advantages such as it is not necessarily limited on using full second order models and can be conducted under no constraints on the secondary response. However, this approach has some difficulties in application. For example, minimizing MSE does not offer any restriction on the bias. To overcome the possible difficulties often encountered related to minimizing MSE, Copeland and Nelson (1996) discussed a modified version of the MSE based optimization procedure. Their technique focuses on a restriction on how far the mean response could be located from its target and uses direct function optimization based on simplex procedure.

Considering MSE criterion, many methods have been improved considering the trade-off between the bias and variance. Ding, Lin and Wei (2004) suggested the data driven weighted MSE approach. Their approach is configured on the idea that the optimal solution lied on the efficient curve. Subsequently, Steenackers and Guillaume (2008) integrated MSE to the bias-specified model of Shin and Cho (2005) to obtain the optimal solution of the DRS problem. Further improvements based on the MSE model are related to the generalized linear mixed models and the inverse problem model (Robinson et al. 2006; Truong & Shin 2012). Additionally, Köksoy (2006) and Köksoy and Yalcinoz (2006) proposed the MSE model as a criterion to solve the multiple-response quality problems by assigning the weights to the individual MSE functions of each response. Following these articles, some studies examining the DRS problem is listed as follows: Baba et al. (2022), Del Castillo, Colosimo and Alshraideh (2012), Kim and Cho (2002), Köksoy and Fan (2012), Lizotte, Greiner and Schuurmans (2012), Midi and Aziz (2019), Shaibu and Cho (2009), Tang and Xu (2002), Zeybek (2020), Zeybek and Köksoy (2020, 2016), and Zeybek, Köksoy and Robinson (2020).

AN OVERVIEW ON SKILL SCORES

The term forecast is defined as the prediction of a system, while forecast verification is the operation of identifying the quality of a forecast. Murphy (1993) focused on three types of goodness of a forecast: consistency, quality and value. While consistency is defined as the degree of correspondence between the forecasts and the decision maker's judgements based upon his/her prior experience, quality is a degree to which the forecast corresponds to the reality. Finally, value relates to the benefits achieved by a decision maker who use the forecasts.

Traditionally, forecast verification emphasizes accuracy and skill which are two features that conduce to the quality of a forecast. Accuracy is the average degree of resemblance between forecast and observations, i.e., the reality. Since the difference between a forecast and observation is defined as error. MSE is a basic measure of accuracy:

$$MSE(f,y) = \frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2 \quad (3)$$

where f_i and y_i represent the ith forecast and ith observation, $i = 1, \ldots, n$. Note that, $f_i = y_i$ (for all i) indicate MSE(f,y) = 0 which signs the maximum accuracy, i.e., the perfect forecast.

Skill is the relative accuracy of the forecast and defines a correspondence between forecast of interest and reference forecasts. The reference forecast does not require any knowledge or modelling; thus, it is an unskilled forecast. In this context, skill score is a measure of this relative improvement and widely used in evaluating the performance of operational and experimental forecasts. A skill score can be expressed based on the MSE measure of accuracy as follows:

$$SS(f,\pi,y) = 1 - \frac{MSE(f,y)}{MSE(\pi,y)} \quad (4)$$

Here, $SS(f,\pi,y)$ is a function of forecast of interest (f), reference forecast (π) and observations (y) (Gupta et al. 2009; Murphy 1988; Weglarczyk 1998; Wheatcroft 2019).

A decomposition of the MSE measure of accuracy given in Equation (3) can be formed as follows:

$$MSE(f,y) = \left(\bar{f} - \bar{y}\right)^2 + s_f^2 + s_y^2 - 2s_f s_y r \quad (5)$$

where $\bar{f}$, $\bar{y}$, and s_f, s_y are the sample means and standard deviations of the forecasts of interest (simulated/ computed values) and the relevant observations, respectively. r is the linear correlation coefficient between forecasts of interest and observations. In the context of forecast verification, r is a measure of forecasting performance and is assessed as a model quality criterion. The practitioner must identify exactly the following two aspects: 1) what the criterion is sensitive to the process requirements, and 2) how the value assigned to the criterion is to be interpreted.

Assuming $\bar{y}$ represents the reference forecast, MSE($\pi.y$) is formed as:

$$MSE(\bar{y},y) = s_y^2 \quad (6)$$

and the skill score defined by Equation (4) can be expressed as follows:

$$SS(f,\bar{y},y) = r^2 - \left(r - \frac{s_f}{s_y}\right)^2 - \left(\frac{\bar{f}-\bar{y}}{s_y}\right)^2 \quad (7)$$

Here, the terms r^2, $\left(r - \frac{s_f}{s_y}\right)^2$, and $\left(\frac{\bar{f}-\bar{y}}{s_y}\right)^2$ represent the strength of the linear relationship between the reference forecast and observations, the conditional bias and the unconditional bias, respectively (Murphy 1988; Weglarzyk 1998).

One of the most popular transformations of MSE is Nash and Sutcliffe's (1970) efficiency (NSE). In fact, given by Equation (7) is a decomposition of the NSE criterion. The NSE criterion is commonly used as a measure for comparing model performance. It is a strong alternative to MSE due to the better reflection to desirable and undesirable features of the model assessed. In fact, $SS(\bar{f}, \bar{y}, y)$ the NSE approach was introduced to the field of quality improvement with the study of Zeybek (2018). This study presents a reference in which the NSE model performance criterion is used for the first time in the field of quality and modeled with the response surface approach. The approach in Zeybek (2018) is structured on optimizing the NSE response surface for the 'target is best' situation. The approach in this study has been referenced in many fields (Edamo et al. 2022a, 2022b; Iqbal et al. 2022; Li et al. 2022; Mushore, Mutanga & Odindi 2022; Nosratpour, Rahimzadegan & Beikahmadi 2022; Şen 2021).

An alternative decomposition of MSE is presented by Gupta et al. (2009). They proposed using MSE as a model calibration criterion and defined MSE measure of accuracy based on NSE. This model calibration criterion is defined as follows:

$$MSE(f,y) = 2s_f s_y(1-r) + \left(s_f - s_y\right)^2 + \left(\bar{f} - \bar{y}\right)^2 \quad (8)$$

which shows the related decomposition of MSE by focusing on the ability of the model (r), variability error, i.e., $\left(s_f - s_y\right)^2$ and bias error, i.e., $\left(\bar{f} - \bar{y}\right)^2$. Therefore. optimizing Equation (8) is a search for a balanced solution among these components.

THE PROPOSED PROCEDURE

The proposed approach is constructed on a new criterion - *model quality criterion* - by integrating skill scores into MSE to solve the DRS problem. In this regard, two new terms are defined for quality improvement studies: *internal quality forecast* and *external quality forecast*. Reference forecast is assumed to be based solely on a relevant design of experiments which is conducted to find the best operating conditions that meet the system requirements. Therefore, reference forecast depends on a set of observations and the statistical moments of the interested responses obtained from a design of experiment. Thus, they are so-called internal quality forecast for the DRS problem. For example, response surface models created for features such as system average and system variance based on the data obtained by an experimental design in a DRS problem are considered as internal quality forecast. On the other hand, system requirements which are determined based on customer requirements, production technology and process capability -i.e., the process target and desired standard deviation are taken as forecast of interest. Thus, they are so-called external quality forecasts for the DRS problem. The targets determined in a DRS problem are considered as external quality forecast.

In line with the definitions of the internal and external quality forecasts, the skill score in Equation (4), can be modelled by response surface approach for the DRS problem as follows:

$$\widehat{SS}_{\tau,\bar{y},y}(x) = 1 - \frac{\widehat{MSE}_{\tau,y}(x)}{\widehat{MSE}_{\bar{y},y}(x)} \quad (9)$$

Here, $\widehat{MSE}_{\tau,y}(x)$ and $\widehat{MSE}_{\bar{y},y}(x)$ are the fitted response surface of measure of accuracy related with external and internal quality forecasts, respectively, and take the following forms:

$$\widehat{MSE}_{\tau,y}(x) = \left(\tau - \hat{\mu}(x)\right)^2 + \sigma_d^2 + \hat{\sigma}^2(x) - 2\sigma_d\hat{\sigma}(x)r \quad (10)$$

and

$$\widehat{MSE}_{\bar{y},y} = \hat{\sigma}^2(x) \quad (11)$$

which are obtained based on Equations (5) and (6). Here, τ and σ_d are the process requirements about the target and desired standard deviation of the system, respectively. r is the model quality criterion that measures the correspondence between external and internal quality forecasts, $\hat{\mu}(x)$ and $\hat{\sigma}(x)$ are the fitted mean and standard deviation response surfaces given by Equations (1) and (2).

Thus, the fitted model calibration criterion, which is modelled by response surface approach based on Equations (8) and (10) - (11), is proposed in the following form.

$$\widehat{MSE}_{s.s.}(x) = 2\sigma_d\hat{\sigma}(x)(1-r) + (\sigma_d - \hat{\sigma}(x))^2 + (\tau - \hat{\mu}(x))^2 \quad (12)$$

This proposed response surface of the decomposition of MSE in terms of skill scores focuses on three components: *variability error response*, $(\sigma_d - \hat{\sigma}(x))^2$, *bias error response*, $(\tau - \hat{\mu}(x))^2$, and *model quality criterion*, *r*. Therefore, it is suggested to use proposed $MSE_{s.s.}$ as an objective function, since optimizing Equation (12) is a search for a balanced solution among these components.

Finally, the optimization scheme takes the following form for the NTB case:

$$\text{Min } \widehat{MSE}_{s.s.}(x) = 2\sigma_d\hat{\sigma}(x)(1-r) + (\sigma_d - \hat{\sigma}(x))^2 + (\tau - \hat{\mu}(x))^2 \quad (13)$$
$$s.t. \; x \in \mathrm{R}$$

under an additional constraint of experiment region such as cuboidal ($-1 \le x_i \le 1$, $i = 1.2. \ldots .k$) or spherical regions ($\boldsymbol{x'x} \le \rho^2$). The proposed approach is based on minimizing $\widehat{MSE}_{s.s.}(x)$ which uses a restriction in the terms of the forecasting performance - i.e., model quality criterion. In fact, the proposed approach uses a degree of closeness of the internal and external quality forecasts and determines the optimal factor settings with a restriction on this measure. The term r is assessed as only a model performance criterion in the proposed optimization approach and is used to adjust the measure of the model performance for the DRS problem. In this regard, r is defined as a pre-specified constraint for the proposed approach, and can range from zero to one, $0 \le r \le 1$, while $r = 0$ indicates no restriction, $r = 1$ displays the strict restriction. The value of 1 indicates the restriction on the degree of closeness of the external and internal quality forecasts is adjusted as the perfect similarity. In other words, it means that the system features hit the target, the system mean, and variance are achieved at the determined targets, and the production will be made by meeting the desired targets in this way, and production will be made with the desired targets without any deviation in the system mean and variance. Thus, the best operating conditions to be determined by the proposed approach are set to provide mean and standard deviation response estimations with perfect proximity to the system requirements without any biases. The value of zero implies that the optimization is run without a restriction about the closeness of the external and internal quality forecasts. When $r = 0$, it means that no constraints can be used. This scenario corresponds to the method suggested in Lin and Tu (1995). They proposed the idea that by adopting MSE optimization, the target value can be deviated to reduce the variance. While the value of r decreases, the estimated mean response moves away from the target, but a smaller variance is obtained. On the other hand, r can take any value between 0 and 1. r approaching 1 means that the estimated mean response estimate is approaching the target, and r approaching 0 means that this estimate is moving away from the target. The variance response estimate will increase as r approaches 1 and decrease as r approaches 0. Since in DRS problems there is a restriction for the variance to be below a certain value, this value shows how much we will sacrifice in the mean response estimate. In this regard, the different values of r yields a string of solutions and facilities, and the decision-maker should decide which values of r fits the capability of technology considering the structure of the DRS problem.

EXAMPLES

This section presents two examples: a simulation study and an industrial application of the proposed approach for the printing process.

A SIMULATION STUDY

In this section, the simulation results are analyzed for purposes the effects of the model quality criterion on the optimal settings. For the simulation study, the case that three controllable factors (x_1, x_2, x_3) affect the response (y) is considered. Specifically, let y_{ij}~N($\mu_i(x)$, $\sigma i^2(x)$) for the $i = 1, 2, 3, \ldots, 27$ design locations and let j denote the number of replicates at each design location (i.e., $j = 3$ in this study), where

$$\mu_i(\mathbf{x}) = 50 + 5(x_{i1}^2 + x_{i2}^2 + x_{i3}^2) \tag{14}$$

$$\sigma_i^2(\mathbf{x}) = 100 + 5((x_{i1} - 0.5)^2 + x_{i2}^2 + x_{i3}^2) \tag{15}$$

Note that, the models for the mean and variance given in Equations (14) - (15) are Park and Cho (2003)'s simulation models. Three responses are generated from the Normal distribution with $\mu_i(\mathbf{x})$ and $\sigma_i^2(\mathbf{x})$ at each design factor settings (x_{i1}, x_{i2}, x_{i3}). The total number of iterations is 1000, each having 27 design locations and 81 responses. The simulation is conducted by using MATLAB.

For the external quality forecasts, the assumptions are the pre-defined target is $\tau = 60$ and desired standard deviation is $\sigma_d = 10$. And the internal forecasts, $\hat{\mu}(x)$ and $\hat{\sigma}(x)$, are estimated from the simulated design of experiment for each 1000 iteration. In the optimization phase, for each 1000 iteration, the proposed $\widehat{MSE}_{s.s.}(x)$, given by Equation (13), is minimized to determine the optimal factor setting for the values under the cuboidal experiment region ($-1 \le x_i \le 1$, $i = 1, 2, \ldots, 27$). For the purposes to determine the possible effects of the value of the model quality criterion, the simulation study is performed under $r = 0.00$, 0.25, 0.50, 0.75 and 1.00.

Table 1 gives the results of simulation study for $r = 0.00$, 0.25, 0.50, 0.75 and 1.00. In Table 1, $\hat{\mu}(x)$ and $\hat{\sigma}(x)$ represent the average of estimated mean and estimated standard deviation across the 1000 simulations. Additionally, while $\widehat{MSE}_{s.s.}(x)$ represent the average of the optimization results of Equation (13) across the 1000 simulations, the average of the estimates of variability error response, $\sum_{i=1}^{1000}(\sigma_d - \hat{\boldsymbol{\sigma}}_i)^2/1000$, and bias error response, $\sum_{i=1}^{1000}(\tau - \hat{\mu}_i)^2/1000$, are also given in Table 1.

Considering Table 1, in general, as the model quality criterion, r, increases from 0 to 1, the bias between internal quality forecasts and external quality forecasts decreases. In other words, while the process mean is estimated closer to the target value, the standard deviation increases, but the optimized MSE value in terms of skill scores, $\widehat{MSE}_{s.s.}(x)$, decreases. For example, when $r = 0.25$, $\hat{\mu}(x) = 60.11$, $\hat{\sigma}(x) = 6.88$, $\widehat{MSE}_{s.s.}(x) = 115.27$ is obtained, while $r = 0.75$, these values are estimated as $\hat{\mu}(x) = 59.97$, $\hat{\sigma}(x) = 7.96$, $\widehat{MSE}_{s.s.}(x) = 44.55$.

Thus, as the model quality criterion increases, the variability error response and bias error response estimates decrease. In addition, to comparing the effects of the model quality criterion on the performance of the proposed method, we also use the estimated mean and standard deviation at the optimal conditions. Figures 1 and 2 show the kernel densities of the optimal estimated mean and standard deviation, respectively, obtained in the simulations under $r = 0.00$, 0.25, 0.50, 0.75, and 1.00. Note in both figures that the kernel densities are more tightly distributed for the predictions obtained as the value of the model quality characteristic increases. Table 2 provides the lower 2.5th, median, and upper 97.5th

percentiles of the estimated mean and standard deviation responses at the optimal operating conditions under r = 0.00, 0.25, 0.50, 0.75 and 1.00. Note that, as the value of the model quality characteristic increases, process mean, and standard deviation estimates are characterized by lower uncertainty.

TABLE 1. Results of simulation study for r = 0, 0.25, 0.50, 0.75, 1.00

r	$\hat{\mu}(x)$	$\hat{\sigma}(x)$	$\widehat{MSE}_{s.s.}(x)$	$\sum_{i=1}^{1000}(\sigma_d-\hat{\sigma}_i)^2/1000$	$\sum_{i=1}^{1000}(\tau-\hat{\mu}_i)^2/1000$
0.00	60.20	6.71	147.85	1.22	12.46
0.25	60.11	6.88	115.27	0.64	11.40
0.50	60.02	7.08	81.10	0.24	10.02
0.75	59.97	7.96	44.55	0.06	4.72
1.00	59.98	9.99	0.08	0.01	0.07

TABLE 2. Lower 2.5th percentile, median, and upper 97.5th percentile of the estimated mean and standard deviation responses from the simulation study

	$\hat{\mu}(x)$				$\hat{\sigma}(x)$		
r	2.5th percentile	Median	97.5th percentile	**r**	2.5th percentile	Median	97.5th percentile
0.00	56.55	60.70	64.84	0.00	1.78	6.11	10.44
0.25	56.07	59.95	63.83	0.25	2.24	6.66	11.09
0.50	56.93	59.89	62.85	0.50	4.59	8.00	11.41
0.75	57.49	59.68	61.87	0.75	6.47	9.72	12.97
1.00	58.39	59.65	60.91	1.00	7.46	10.18	12.90

THE PRINTING PROCESS STUDY

In this section, the proposed approach is illustrated by a well-known printing process study example. This experiment is an exercise in Box and Draper (1987) and many authors have been used this study to illustrate their methods (Copeland & Nelson 1996; Del Castillo & Montgomery 2003; Lin & Tu 1995; Köksoy & Doganaksoy 2003; Vining & Myers 1990). factorial design with three replicates (Table 3), is performed to examine the effect of speed (x_1), pressure (x_2) and distance (x_3) on the ability of a printing machine (y) to apply colored inks to the package labels. The fitted mean and standard deviation responses for the printing ink data are obtained by Vining and Myers (1990) as follows:

$$\begin{aligned}\hat{\mu}(x) &= 327.6 + 177.0x_1 + 109.4x_2 + 131.5x_3 + 32.0{x_1}^2 \\ &-22.4{x_2}^2 - 29.1{x_3}^2 + 66.0x_1x_2 \\ &+75.5x_1x_3 + 43.6x_2x_3\end{aligned} \tag{16}$$

and

$$\begin{aligned}\hat{\sigma}(x) &= 34.9 + 11.5x_1 + 15.3x_2 + 29.2x_3 \\ &+ 4.2{x_1}^2 - 1.3{x_2}^2 \\ &+16.8{x_3}^2 + 7.7x_1x_2 + 5.1x_1x_3 \\ &+14.1x_2x_3.\end{aligned} \tag{17}$$

The printing process study requires a situation where the target is 500 and the desired standard deviation is 60. Note that these values were used in previous studies. For the printing process study, while the internal quality forecasts are the fitted response surfaces given in Equations (15)-(17), the external quality forecasts are assumed as the system requirements such that τ = 500 and the desired standard deviation, i.e., 60. Therefore, the proposed DRS problem formulation for the printing process problem under the cuboidal constraint is formed as:

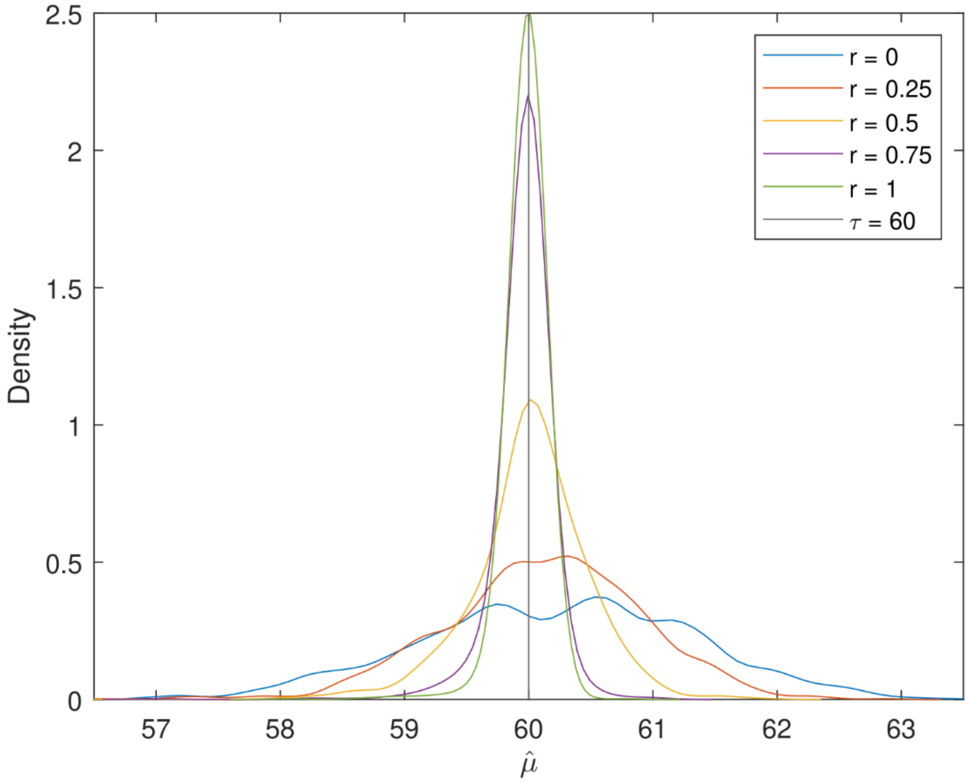

FIGURE 1. Simulated distributions of the estimated mean response under $r = 0.00, 0.25, 0.50, 0.75$ and 1.00

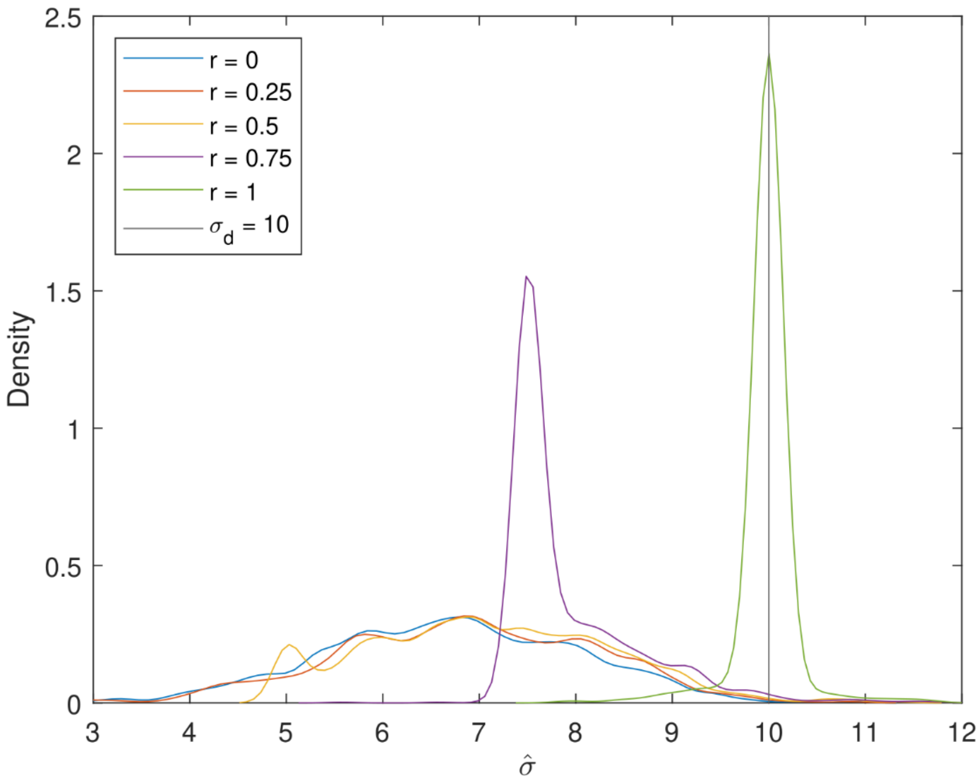

FIGURE 2. Simulated distributions of the estimated standard deviation response under $r = 0.00, 0.25, 0.50, 0.75$ and 1.00

$$\text{Min } \widehat{MSE}_{s.s.}(x) = 2(60)(1-r)\hat{\sigma}(x) + (60-\hat{\sigma}(x))^2 + (500-\hat{\mu}(x))^2 \quad (18)$$
$$s.t. \quad -1 \le x_i \le 1.\ i = 1.2.3$$

where $\hat{\mu}(x)$ and $\hat{\sigma}(x)$ are defined in Equations (14) and (15).

Figure 3 displays the surface plots of the proposed $\widehat{MSE}_{s.s.}(x)$ for the various values of the model quality criterion. r = 1.00, 0.50, 0.00. From Figure 3(a), it is clear that if a perfect relevance between the internal and external quality forecasts is desired, then the minimum value of the $\widehat{MSE}_{s.s.}(x)$ takes the value of zero, and the best operating conditions are obtained under this perfect

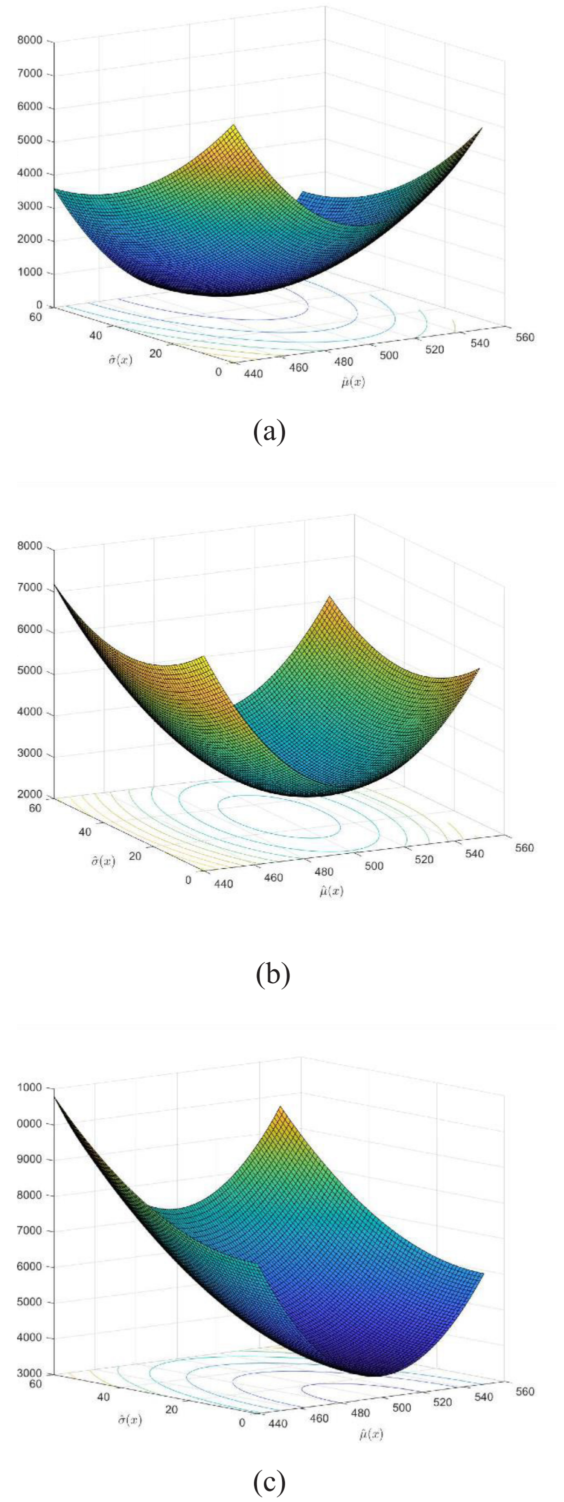

FIGURE 3. Surface plots of the proposed $\widehat{MSE}_{s.s.}(x)$ for the various values of the model quality criterion: (a) r =1, (b) r = 0.5, (c) r = 0

calibration condition. Therefore, the mean response hits the target, and the standard deviation response meets system requirement about the desired variability. While the degree of the model quality criterion gets weaker, the minimum value of $\widehat{MSE}_{s.s.}(x)$ increases, see Figure 3(b)-3(c). The mean response is located far from the target and the standard deviation response gets smaller values than the pre-defined system variability.

While Table 4 summarizes the results based on minimizing $\widehat{MSE}_{s.s.}(x)$ under various values of r under NTB case. Table 5 also illustrates a comparative study of the proposed formulation with the past approaches. In Table 4, it is clear that when $r = 1$, $\widehat{MSE}_{s.s.}(x)$ has the smallest value. The best operating conditions turns out to be $x^* = (0.291, 0.700, 0.220)$ where the estimated mean response hits the target. $\hat{\mu}(x) = 500$ with $\hat{\sigma}(x) = 60$. and the resulting $\widehat{MSE}_{s.s.}(x) = 0$. However, when the degree

TABLE 3. The printing ink data set

i	x_1	x_2	x_3	y_1	y_2	y_3	$\bar{y}$	s
1	-1	-1	-1	34	10	28	24.0	12.49
2	0	-1	-1	115	116	130	120.3	8.39
3	1	-1	-1	192	186	263	213.7	42.80
4	-1	0	-1	82	88	88	86.0	3.46
5	0	0	-1	44	178	188	136.7	80.41
6	1	0	-1	322	350	350	340.7	16.17
7	-1	1	-1	141	110	86	112.3	27.57
8	0	1	-1	259	251	259	256.3	4.62
9	1	1	-1	290	280	245	271.7	23.63
10	-1	-1	0	81	81	81	81.0	0.00
11	0	-1	0	90	122	93	101.7	17.67
12	1	-1	0	319	376	376	357.0	32.91
13	-1	0	0	180	180	154	171.3	15.01
14	0	0	0	372	372	372	372.0	0.00
15	1	0	0	541	568	396	501.7	92.50
16	-1	1	0	288	192	312	264.0	63.50
17	0	1	0	432	336	513	427.0	88.61
18	1	1	0	713	725	754	730.7	21.08
19	-1	-1	1	364	99	199	220.7	133.80
20	0	-1	1	232	221	266	239.7	23.46
21	1	-1	1	408	415	443	422.0	18.52
22	-1	0	1	182	233	182	199.0	29.45
23	0	0	1	507	515	434	485.3	44.64
24	1	0	1	846	535	640	673.7	158.20
25	-1	1	1	236	126	168	176.7	55.51
26	0	1	1	660	440	403	501.0	138.90
27	1	1	1	878	991	1161	1010.0	142.50

of model quality criterion gets weaker, then $\widehat{MSE}_{s.s.}(x)$ tends to increase and this provides flexibility to meet the system requirements. Generally, the DRS problem is constructed on hitting the target with as small variance as possible, and bias can be tolerated to achieve small variance estimation. In this regard, Table 4 illustrates the best operating solutions based on the r values ranged zero to one. For example, if r = 0.75, the proposed approach yields the solution of $\hat{\mu}(x)$ = 499.99 and $\hat{\sigma}(x)$ = 45.10 with $\widehat{MSE}_{s.s.}(x)$ =1575.0. However, when this value decreases, e.g., r = 0.25. then $\hat{\mu}(x)$ = 496.25 (with a larger bias compared with r = 0.75), $\hat{\sigma}(x)$ = 44.67 (with a smaller estimated standard deviation compared with r = 0.75), and a larger $\widehat{MSE}_{s.s.}(x)$ = 4268.07 are achieved. As a result, if the problem is constructed on relaxing zero bias assumption, then, decision-maker should make a reduction of the value of model quality criterion. Thus, while the value of this restriction decreases. the estimated mean response moves away from the target. but a smaller variance is obtained.

TABLE 4. The optimal settings for the printing ink data set under the NTB case

Proposed approach				
r	x^*	$\hat{\mu}(x)$	$\hat{\sigma}(x)$	$\widehat{MSE}_{s.s.}(x)$
1.00	(0.291. 0.700. 0.220)	500.00	60.00	0.00
0.95	(0.363. 0.657. 0.142)	500.00	57.00	351.00
0.90	(0.653. -0.038. 0.287)	500.00	54.00	684.00
0.85	(0.589. 0.418. -0.005)	500.00	51.00	999.00
0.80	(0.949. -0.315. 0.174)	500.00	48.00	1296.00
0.75	(1.000. 0.121. -0.261)	499.99	45.10	1575.01
0.70	(1.000. 0.117. -0.260)	499.63	45.05	1845.46
0.65	(1.000. 0.113. -0.259)	499.28	45.01	2115.66
0.60	(1.000. 0.107. -0.257)	498.93	44.97	2385.60
0.55	(1.000. 0.107. -0.258)	498.57	44.93	2655.28
0.50	(1.000. 0.102. -0.256)	498.22	44.89	2924.72
0.45	(1.000. 0.101. -0.257)	497.86	44.84	3193.89
0.40	(1.000. 0.098. -0.256)	497.51	44.80	3462.82
0.35	(1.000. 0.095. -0.256)	497.16	44.76	3731.49
0.30	(1.000. 0.092. -0.255)	496.80	44.71	3999.90
0.25	(1.000. 0.089. -0.255)	496.45	44.67	4268.07
0.20	(1.000. 0.086. -0.254)	496.10	44.63	4535.98
0.15	(1.000. 0.083. -0.254)	495.74	44.59	4803.63
0.10	(1.000. 0.080. -0.253)	495.39	44.55	5071.03
0.05	(1.000. 0.077. -0.252)	495.04	44.50	5338.18
0.00	(1.000. 0.074. -0.252)	494.69	44.43	5605.08

Vining and Myers (1990) reported the point $\boldsymbol{x}^*$ = (0.614,0.0228,0.1000) with $\hat{\sigma}(x)$ = 51.78 and $\hat{\mu}(x)$ = 500 as the optimal solution of the printing process study. This stationary point is almost obtained with the proposed approach under r = 0.86 (Table 5). Additionally, the best operating conditions found by Del Castillo and Montgomery (1993) and Köksoy and Doganaksoy (2003) also can be achieved with approximately r = 0.75 and r = 0.18, respectively, using the proposed formulation. On the other hand, r = 0 means that there is no restriction on the model quality criterion. Thus, the obtained stationary point under r = 0 is same as the solution of Lin and Tu (1995). The fact that the solution set obtained from some values of r between 0 and 1 gives results close to the solution sets of some approaches already existing in the literature shows that the proposed method is actually a valid method. Because the proposed method is in harmony with the existing and fundamental approaches in the literature in terms of its philosophy and the criteria it advocates.

TABLE 5. Comparison of the optimal settings for the printing process data under NTB case

Existing approaches			Proposed approach	
	$\hat{\mu}(x)$	$\hat{\sigma}(x)$	r	$\widehat{MSE}_{s.s.}(x)$
Vining and Myers (1990)	500.00	51.78	0.86	919.08
Del Castillo and Montgomery (1993)	500.00	45.10	0.75	1567.15
Köksoy and Doganaksoy (2003)	495.99	44.62	0.18	4617.58
Lin and Tu (1995)	494.68	44.43	0.00	5605.08

CONCLUSION

Various approaches have been proposed in the current literature to solve the DRS problem. As previously discussed, most of them focus on minimizing a response function of the bias and variability. This study is an important alternative to the earlier formulations of the MSE criterion that seek a unique optimum solution in a single shot. The proposed approach discusses MSE in regards of skill scores and defines a new decomposition of the MSE criterion for the DRS problem. In fact, the proposed method adopts the relaxing zero bias assumption, but unlike the existing approaches, it imposes a restriction on the relationship between internal and external quality forecasts. Additionally, it allows the decision-maker to adjust the model performance. With different values of the model quality criterion, a different set of optimal parameters will have different result. For the larger values of , the optimal set of process parameters predicted would be closer to the system requirements. While the value of decreases, the estimated mean response moves away from the target, but a smaller variance is obtained. Consequently, the proposed approach provides a flexible formulation of the problem and generates more alternative solutions. This study therefore offers a useful reference for practitioners and benefit of providing alternatives for assessing process performance. While current approaches focus on a single result, the proposed method offers many alternative results to the practitioner. Depending on the production performance and features, the practitioner decides how much to sacrifice the goals and chooses the solution.

REFERENCES

Baba, I., Midi, H., Ibragımov, G. & Rana, S. 2022. A new optimization scheme for robust design modeling with unbalanced data. *Sains Malaysiana* 51(5): 1577-1586.

Box, G.E.P. 1985. Discussion of "Off-line quality control, parameter design and the Taguchi method". *Journal of Quality Technology* 17: 198-206.

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

Box, G.E.P. & Wilson, K.B. 1951. On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society: Series B (Methodological)* 13(1): 1-38.

Copeland, K.A. & Nelson, P.R. 1996. Dual response optimization via direct function minimization. *Journal of Quality Technology* 28(1): 331-336.

Del Castillo, E. & Montgomery, D.C. 1993. A nonlinear programming solution to the dual response problem. *Journal of Quality Technology* 25: 199-204.

Del Castillo, E., Colosimo, B.M. & Alshraideh, H. 2012. Bayesian modeling and optimization of functional responses affected by noise factors. *Journal of Quality Technology* 44(2): 117-135.

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

Edamo, M.L., Ukumo, T.Y., Lohani, T.K., Miranic, K.B. & Ayele, M.A. 2022(a). Flood inundation mapping under climate change scenarios in the Boyo watershed of Southern Ethiopia. *Journal of Water and Climate Change* 13: 3170-3178.

Edamo, M.L., Bushira, K.M., Ukumo, T.Y., Ayele, M.A., Alaro, M.A. & Borko, H.B. 2022(b). Effect of climate change on water availability in Bilate catchment, Southern Ethiopia. *Water Cycle* 3: 86-99.

Gupta, H.V., Kling, H., Yilmaz, K.K. & Martinez, G.F. 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* 377: 80-91.

Iqbal, A., Mondal, M.S., Khan, M.S.A., Hakvoort, H. & Veerbeek, W. 2022. Flood propagation processes in the Jamuna River Floodplain in Sirajganj. In *Water Management: A View from Multidisciplinary Perspectives*, edited by Tarekul Islam, G.M., Shampa, S. & Chowdhury, A.I.A. Springer. pp. 45-67.

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

Kim, Y.J. & Cho, B.R. 2002. Development of priority-based robust design. *Quality Engineering* 14: 355-363.

Köksoy, O. 2006. Multiresponse robust design: Mean square error (MSE criterion). *Applied Statistics and Computation* 175: 1716-1729.

Köksoy, O. & Doganaksoy, N. 2003. Joint optimization of mean and standard deviation in response surface experimentation. *Journal of Quality Technology* 35(3): 239-252.

Köksoy, O. & Fan, S.S. 2012. An upside-down normal loss function-based method for quality improvement. *Engineering Optimization* 44(8): 935-945.

Köksoy, O. & Yalcinoz, T. 2006. Mean square error criteria to multi-response process optimization by a new genetic algorithm. *Applied Mathematics and Computation* 175: 1657-1674.

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

Li, H., Jiang, C., Choy, S., Wang, X., Zhang, K. & Zhu, D.A. 2022. Comprehensive study on factors affecting the calibration of potential evapotranspiration derived from the Thornthwaite Model. *Remote Sensing* 14(18): 4644.

Lizotte, D.J., Greiner, R. & Schuurmans, D. 2012. An experimental methodology for response surface optimization methods. *Journal of Global Optimization* 52(4): 698-736.

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

Murphy, A.H. 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review* 116: 2417-2424.

Murphy, A.H. 1993. What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting* 8: 281-293.

Mushore, T.D., Mutanga, O. & Odindi, J. 2022. Estimating urban LST using multiple remotely sensed spectral indices and elevation retrievals. *Sustainable Cities and Society* 78: 103623.

Nash, J.E. & Sutcliffe, J.V. 1970. River flow forecasting through conceptual models. Part I: A discussion of principles. *Journal of Hydrology* 10: 282-290.

Nosratpour, R., Rahimzadegan, M. & Beikahmadi, N. 2022. Introducing a merged precipitation satellite model using satellite precipitation products, land surface temperature, and precipitable water vapor. *Geocarto International* 37(26): 11782-11812.

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

Robinson, T.J., Wulff, S.S., Montgomery, D.S. & Khuri, A.I. 2006. Robust parameter design using generalized linear mixed models. *Journal of Quality Technology* 38: 65-75.

Shaibu, A.B. & Cho, B.R. 2009. Another view of dual response surface modeling and optimization in robust parameter design. *The International Journal of Advanced Manufacturing Technology* 41: 631-641.

Shin, S. & Cho, B.R. 2005. Bias-specified robust design optimization and analytical solutions. *Computers & Industrial Engineering* 48: 129-148.

Steenackers, G. & Guillaume, P. 2008. Bias-specified robust design optimization: A generalized mean squared error approach. *Computers & Industrial Engineering* 54: 259-268.

Şen, Z. 2021. Model efficiency performance assessment through a standard triangular diagram (STD). *Modeling Earth Systems and Environment* 7: 1193-1205.

Taguchi, G. 1986. *Introduction to Quality Engineering: Designing Quality into Products and Processes*. New York: Kraus International Publications.

Tang, L.C. & Xu, K. 2002. A unified approach for dual response surface optimization. *Journal of Quality Technology* 34(4): 437-447.

Truong, N.K.V. & Shin, S. 2012. Development of a new robust design method based on Bayesian perspectives. *International Journal of Quality Engineering and Technology* 3(1): 50-78.

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.

Weglarczyk, S. 1998. The interdependence and application of some statistical quality measures for hydrological models. *Journal of Hydrology* 206: 98-103.

Wheatcroft, E. 2019. Interpreting the skill score form of forecast performance metrics. *International Journal of Forecasting* 35(2): 573-579.

Zeybek, M. 2020. Interval robust design under contaminated and non-normal data, *Communications in Statistics - Theory and Methods* 49: 5406-5418.

Zeybek, M. & Köksoy, O. 2020. The effects of gamma noise on quality improvement. *Communications in Statistics – Simulation and Computation* 49(7): 1783-1797.

Zeybek, M. 2018. Nash-Sutcliffe efficiency approach for quality improvement. *Journal of Applied Mathematics and Computation* 2(11): 496-503.

Zeybek, M. & Köksoy, O. 2016. Optimization of correlated multi-response quality engineering by the upside-down normal loss function. *Engineering Optimization* 48(8): 1419-1431.

Zeybek, M., Köksoy, O. & Robinson, T.J. 2020. A dual-response surface modeling approach for gamma robust design. *Quality and Reliability Engineering International* 36(3): 315-327.

*Corresponding author; email: elif.kozan@ege.edu.tr