Local Quadratic Regression: Maximizing Performance via a Modified PRESS** for Bandwidths Selection

E. Edionwe[1,2] and O. Eguasa[3]

[2]*Department of Mathematical Sciences, Edwin Clark University, Kiagbodo, Delta State, Nigeria*
[3]*Department of Physical Sciences, Benson Idahosa University, Benin City, Edo State, Nigeria*

ABSTRACT

In the application of nonparametric regression model, it is a well-established fact that the bandwidth - also called smoothing parameter- is the single most crucial parameter that determines the quality of the estimated responses that are obtained from the regression procedure, and that its choice (how small or large the size) is hugely influenced by the criterion that is applied for its selection. Under small-sample settings, which is typical of response surface studies, the penalized Prediction Error Sum of Squares (PRESS**) criterion is recommended for selecting this all-important parameter. However, for the purpose of selecting bandwidths of improved statistical properties, we propose a modified version of the PRESS** criterion specifically for Local Quadratic Regression (LQR) model. Results from simulated data as well as those from two popular problems from the literature show that LQR procedure that utilizes the bandwidths selected via the proposed modified criterion performs outstandingly better than its counterpart that utilizes bandwidths selected via PRESS** criterion.

Keywords: *Desirability Function, Hat matrix, Penalized Prediction Error Sum of Squares, Response Surface Methodology.*

1. INTRODUCTION:

Although, response surface methodology (RSM) was developed by Box and Wilson in 1951 in the field of agricultural science, the methodology has now gained prominence as an important collection of statistical tools that is applied in process and product optimization in the fields of science, engineering and technology (Hill and Hunter, 1996; Rajewski and Dobrzynski-Inger, 2021).

In the modeling phase of RSM, it is assumed that the relationship between a response variable y and k explanatory variables $x_1, x_2, \dots, x_k$, takes the form:

$$y_i = f\left(x_{i1}, x_{i2}, \dots, x_{ik}\right) + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n \tag{1}$$

where y_i is the output at the i^{th} data point, $x_{ij}, j = 1,2, \dots, k,$ is the value of the j^{th} explanatory variable at the i^{th} data point, f represents the true but unknown function that depicts the exact mathematical relationship between the variables, $\boldsymbol{\varepsilon}_i$ is a random error term assumed to be independent, identically distributed with mean zero and constant variance σ^2, and n is the sample size (Montgomery, 2005; Castillo, 2007).

[1] Corresponding Author: trustnelson24@yahoo.com

In order to get an estimate of the relationship defined in (1) for a given study, a low-order polynomial is assumed for f and data from a designed experiment is used in fitting the assumed polynomial. The ultimate goal of RSM is to determine the values (setting) of the explanatory variables that optimize the fitted polynomial model (Myers, et al., 2009; He, et al., 2012; Yeniay, 2014).

A high-performing fitted model of a study does not only provide the researcher an opportunity to critically study the mathematical relationship among the variables but also enables the prediction of outputs at respective settings of the explanatory variables with high accuracy. Hence, a premium value is placed on a proficiently fitted predictive model (Pickle, et al., 2008; Wan and Birch, 2011).

The traditional regression model for estimating f in (1) is the ordinary least squares (OLS). However, if the data consists of salient patterns and trends that might be overlooked by OLS, nonparametric regression models could provide better alternatives (Hardle, et al., 2005; Wan and Birch, 2011). Essentially, nonparametric regression models such as LQR are applied if, according to Myers (1999), the following are important research targets:

1. The researcher is interested in optimizing a response.
2. The researcher is less interested in an interpretive function (i.e., interpreting the estimated regression coefficients) and more interested in studying the shape of the response surface.
3. The functional form of the relationship between the explanatory variables and the response is not well behaved.

Mathematically, LQR estimate $y_i^{(LQR)}$ of y_i takes the form:

$$\hat{y}_i^{(LQR)} = \boldsymbol{x}_i'(\boldsymbol{X}_q'\boldsymbol{W}_i\boldsymbol{X}_q)^{-1}\boldsymbol{X}_q'\boldsymbol{W}_i\boldsymbol{y}, \quad (2)$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of response, $\boldsymbol{W_i}$ is an $n \times n$ diagonal matrix of weights for estimating y_i, $\boldsymbol{x}_i'$ is the i^{th} row vector of the LQR model matrix $\boldsymbol{X_q}$, $\boldsymbol{X}_q'$ is the transposed LQR model matrix whose general form is given by:

$$\boldsymbol{X_q} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} & x_{11}^2 & x_{12}^2 & \cdots & x_{1k}^2 \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} & x_{21}^2 & x_{22}^2 & \cdots & x_{2k}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} & x_{n1}^2 & x_{n2}^2 & \cdots & x_{nk}^2 \end{bmatrix},$$

In matrix form, the vector of LQR estimates presented in (2) is expressed as:

$$\begin{bmatrix} \hat{\boldsymbol{y}}_1^{(LQR)} \\ \hat{\boldsymbol{y}}_2^{(LQR)} \\ \vdots \\ \hat{\boldsymbol{y}}_n^{(LQR)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1(\boldsymbol{X}_q'\boldsymbol{W}_1\boldsymbol{X}_q)^{-1}\boldsymbol{X}_q'\boldsymbol{W}_1 \\ \boldsymbol{x}_2(\boldsymbol{X}_q'\boldsymbol{W}_2\boldsymbol{X}_q)^{-1}\boldsymbol{X}_q'\boldsymbol{W}_2 \\ \vdots \\ \boldsymbol{x}_n(\boldsymbol{X}_q'\boldsymbol{W}_n\boldsymbol{X}_q)^{-1}\boldsymbol{X}_q'\boldsymbol{W}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_n \end{bmatrix}, \quad (3)$$

$$= \begin{bmatrix} \boldsymbol{h}_1^{(LQR)'}(\boldsymbol{b}) \\ \boldsymbol{h}_2^{(LQR)'}(\boldsymbol{b}) \\ \vdots \\ \boldsymbol{h}_n^{(LQR)'}(\boldsymbol{b}) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad (4)$$

$$= \boldsymbol{H}^{(LQR)}(\boldsymbol{b})\boldsymbol{y}, \quad (5)$$

where $\boldsymbol{h}_i^{(LQR)'}(\boldsymbol{b}) = \left(\boldsymbol{h}_{i1}^{(LQR)}\boldsymbol{h}_{i2}^{(LQR)} \dots \boldsymbol{h}_{in}^{(LQR)}\right)$ is the i^{th} row vector of the $n \times n$ LQR Hat matrix, $\boldsymbol{H}^{(LQR)}(\boldsymbol{b})$.

If all the quadratic terms in the LQR model matrix $\boldsymbol{X_q}$ are deleted, LQR reduces to local linear regression (Anderson-Cook and Prewitt, 2005).

The rth entry, say w_{rr} of the weight matrix $\boldsymbol{W_o}$ for estimating y_o is obtained from the product kernel given as:

$$w_{or} = \prod_{j=1}^{k} K\left(\frac{x_{oj}-x_{rj}}{b_r}\right) / \sum_{i=1}^{n} \prod_{j=1}^{k} K\left(\frac{x_{oj}-x_{ij}}{b_i}\right), \quad i = 1,2,\dots,n, \quad (6)$$

where $K\left(\frac{x_{oj}-x_{rj}}{b_r}\right) = e^{-\left(\frac{x_{oj}-x_{ij}}{b_r}\right)^2}$ is the simplified Gaussian function which assigns relatively heavier weights to the observations close to x_{oj} than those far from x_{oj}, and b_i, $i = 1,2,\dots,n$, are referred to as the local or locally adaptive bandwidths which reduce to a fixed or global bandwidth b in situations where we have $b_1 = b_2 \dots = b_n = b$ (Edionwe and Mbegbu, 2014).

In nonparametric regression procedure, the values, x_{ij} of the explanatory variables are transformed such that $0 \leq x_{ij} \leq 1$, and consequently, b_i, $i = 1,2,\dots,n$, are constrained to lie in the interval $0 < b_i \leq 1$ (Edionwe, et al., 2016).

A model for selecting local bandwidths presented in Edionwe, et al. (2017) and Edionwe, et al. (2018) is given by:

$$b_i = \frac{N(TC-y_i)}{T(Cn-1)}, \quad (7)$$

where $C \geq 0$ and $N > 0$ are data-driven tuning parameters and $T = \sum_i^n y_i$.

The bandwidth is adjudged the most crucial parameter as far as nonparametric regression procedure are concerned and several criteria for its selection based on factors such as data type (single and equally-spaced explanatory variable, econometric data, e.tc) and for a specific estimation task (density estimation, regression function estimation, estimation of the derivative of regression function, e.tc) (Fan and Gijbels, 1996; Loader, 1999).

For small-sample studies such as RSM, the PRESS** criterion developed by Mays *et al* (2001) for selecting bandwidths in the application of both nonparametric and semiparametric regression models is given by:

$$PRESS^{**}(\boldsymbol{b}) = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{i,-i}^{(LQR)}(\boldsymbol{b})\right)^2}{\left[n - tr\left(\boldsymbol{H}^{(LQR)}(\boldsymbol{b})\right)\right] + \left[(n-k-1)\left(\frac{SSE_{max} - SSE(\boldsymbol{b})}{SSE_{max}}\right)\right]},$$

$$= \frac{PRESS}{DF + \left[(n-k-1)\left(\frac{SSE_{max} - SSE(\boldsymbol{b})}{SSE_{max}}\right)\right]}, \quad (8)$$

where $SSE_{max} = \sum_{i=1}^{n}\left(y_i - \hat{y}_i^{(LQR)}(\boldsymbol{b})\right)^2$ is the maximum Sum of Squared Errors obtained as b tends to infinity, $SSE(\boldsymbol{b})$ is the Sum of Squared Errors for a given vector of bandwidths, $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)$, Degree of Freedom (DF)= $n - tr\{\boldsymbol{H}^{(LQR)}(\boldsymbol{b})\}$, $tr\{\boldsymbol{H}^{(LQR)}(\boldsymbol{b})\}$ is the sum of the diagonal elements of the LQR Hat matrix for a given vector of bandwidths, $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)$, and $\hat{y}_{i,-i}^{(LQR)}(\boldsymbol{b})$ is the leave-one-out estimate of y_i with the i^{th} observation left out.

In applying (7) to generate bandwidths for a given data, we search for the optimal values, $C^*$and N^*, of C and N, respectively, that give the optimal local bandwidths for minimizing the $PRESS^{**}(\boldsymbol{b})$ criterion. The phase succeeding the modelling phase in RSM is the optimization phase, where the setting of the explanatory variables that optimize the fitted regression model according to the process specifications (or production requirements) is sought.

In studies that involve say m responses, $m > 1$, the goal is to obtain the setting of the explanatory variables which simultaneously optimize the m fitted models with respect to their individual process specifications (Harrington, 1965; Derringer and Suich, 1980).

A few criteria for carrying out multiple response optimization exist amongst which the desirability measure (function) stands out. The desirability function, with respect to the process specification of individual response, transforms the fitted model, $\hat{y}_p(\boldsymbol{x})$, into a scalar measure, $d_p\left(\hat{y}_p(\boldsymbol{x})\right)$, $p = 1,2,\ldots,m$, after which the setting of each of the explanatory variables that maximize the geometric mean of the m transformed scalar measures is subsequently sought (Wan and Birch, 2011). The classifications of $d_p\left(\hat{y}_p(\boldsymbol{x})\right)$ based on the process specification of the responses is presented in Appendix A.

In this paper, the optimization procedure including the minimization of the $PRESS^{**}(\boldsymbol{b})$ criterion in (8), the proposed modified $PRESS^{**}(\boldsymbol{b})$ criterion in (10), and the maximization of the desirability measure in (4*) in the Appendix for the determination of the optimal setting of the explanatory variables are all carried out using the Genetic Algorithm (GA) optimization toolbox in Matlab.

II. METHODOLOGY

The Penalizing Factor of the $PRESS^{**}(\boldsymbol{b})$ vis-à-vis the Flexibility of the LQR Hat Matrix

In general, the $n \times n$ diagonal matrix weights $\boldsymbol{W_o}$ derived from (6) for estimating y_o can be expressed as:

$$\boldsymbol{W_o} = \begin{bmatrix} w_{01} & 0 & \cdots & 0 \\ 0 & w_{02} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & w_{0n} \end{bmatrix},$$

$$= \begin{bmatrix} \left(\frac{\prod_{j=1}^{k} e^{-\left(\frac{x_{oj}-x_{1j}}{b_1}\right)^2}}{\sum_{i=1}^{n} \prod_{j=1}^{k} e^{-\left(\frac{x_{oj}-x_{ij}}{b_i}\right)^2}} \right) & 0 & \cdots & 0 \\ 0 & \left(\frac{\prod_{j=1}^{k} e^{-\left(\frac{x_{oj}-x_{2j}}{b_2}\right)^2}}{\sum_{i=1}^{n} \prod_{j=1}^{k} e^{-\left(\frac{x_{oj}-x_{ij}}{b_i}\right)^2}} \right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\frac{\prod_{j=1}^{k} e^{-\left(\frac{x_{oj}-x_{nj}}{b_n}\right)^2}}{\sum_{i=1}^{n} \prod_{j=1}^{k} e^{-\left(\frac{x_{oj}-x_{ij}}{b_i}\right)^2}} \right) \end{bmatrix},$$

According to Cleveland (1979), Fan and Gijbels (1996), Mays and Birch (1998), the advantage of LQR over OLS is its flexibility which is a function of the choice of bandwidths that are selected for the procedure.

In situations where $b_1, b_2, \ldots, b_n$ all tend to 1 and above in (3), the elements of $\boldsymbol{W_1}, \boldsymbol{W_2}, \ldots, \boldsymbol{W_n}$ which respectively form part of the row vectors $\boldsymbol{h}_1^{(LQR)'}(\boldsymbol{b}), \boldsymbol{h}_2^{(LQR)'}(\boldsymbol{b}), \ldots, \boldsymbol{h}_n^{(LQR)'}(\boldsymbol{b})$, all would return the same value, say w. That is $w_{11} = w_{22} = \cdots = w_{nn} = w$, and consequently, $\boldsymbol{W_1} = \boldsymbol{W_2} = \cdots = \boldsymbol{W_n}$. In this case, assuming that the same model matrix $\boldsymbol{X_q}$ is used for OLS procedure, we will get $\hat{\boldsymbol{y}}_1^{(LQR)} = \hat{\boldsymbol{y}}_1^{(OLS)}, \hat{\boldsymbol{y}}_2^{(LQR)} = \hat{\boldsymbol{y}}_2^{(OLS)}, \ldots, \hat{\boldsymbol{y}}_n^{(LQR)} = \hat{\boldsymbol{y}}_n^{(OLS)}$, meaning that LQR returns the same vector of estimated responses as that of the OLS. On the other hand, the more distinct or dissimilar the elements of $\boldsymbol{W_1}, \boldsymbol{W_2}, \ldots, \boldsymbol{W_n}$ in the vectors $\boldsymbol{h}_1^{(LQR)'}(\boldsymbol{b}), \boldsymbol{h}_2^{(LQR)'}(\boldsymbol{b}), \ldots, \boldsymbol{h}_n^{(LQR)'}(\boldsymbol{b})$, respectively, the higher the flexible of the resulting LQR over that of the OLS.

The advantage of nonparametric regression models over their parametric counterpart is flexibility and, therefore, one of the ways to upgrade the flexibility of LQR is to ensure that the bandwidths selected allow the $1 \times n$ elements of each of the vectors $\boldsymbol{h}_i^{(LQR)'}(\boldsymbol{b}) = \left(\boldsymbol{h}_{i1}^{(LQR)} \boldsymbol{h}_{i2}^{(LQR)} \ldots \boldsymbol{h}_{in}^{(LQR)}\right)$, $i = 1,2, \ldots, n$, to be as distinct from one another as possible. In other words, since distinctiveness implies variability, the variability of the n elements of each of the rows of the Hat matrix should be as high as possible.

There are two shortcomings of the DF that need to be highlighted as it relates to its computation in the $PRESS^{**}(\boldsymbol{b})$ criterion: (1) the DF, as it is computed, is not a function of the distinctiveness (variability) of the elements of the row vectors of the LQR Hat matrix and so does not in any way enhance the flexibility of the LQR. (2) Two, $DF = n - tr\{\boldsymbol{H}^{(LQR)}(\boldsymbol{b})\}$, where $tr\{\boldsymbol{H}^{(LQR)}(\boldsymbol{b})\}$ is the sum of the diagonal elements of the LQR Hat matrix, makes use of only n of the n times $n = n^2$ elements of the Hat matrix, neglecting the remaining $n(n-1) = n^2 - n$ elements. This negligence has negative consequences on the regression procedure since the unused $n^2 - n$ elements contain important information about the data under study. Thus, there is a need for the inclusion of an appropriate statistic in the penalizing factor of the $PRESS^{**}(\boldsymbol{b})$ in order to address these shortcomings.

Methodology for the modification of the penalizing factor in the $PRESS^{**}(\boldsymbol{b})$ criterion

The proposed modification of the $PRESS^{**}(\boldsymbol{b})$ criterion is motivated by the fact that the impact of a statistical measure that allows the elements in each of the rows of the LQR Hat matrix to be as distinct (dissimilar or variable) as possible would trump that of the DF that considers only the sum of the diagonal elements of the same matrix.

In modifying the $PRESS^{**}(\boldsymbol{b})$ criterion, the following objectives are targeted to be achieved:

i. to provide for the inclusion of a penalizing factor that utilizes every bit of information which the entire n^2 elements of the LQR Hat matrix can provide.
ii. to provide for the inclusion of a penalizing factor that allows the flexibility of LQR to be over and above that of the OLS.

If the proposed modified $PRESS^{**}(\boldsymbol{b})$ criterion is designated by $PRESS^{sv}(\boldsymbol{b})$, then $PRESS^{sv}(\boldsymbol{b})$ may be given by:

$$PRESS^{sv}(\boldsymbol{b}) = \frac{PRESS}{\left[(n-k-1)\left(\frac{SSE_{max}-SSE(\boldsymbol{b})}{SSE_{max}}\right)\right]Q}, \tag{9}$$

where Q is a function of a statistical measure that encapsulates objectives (i) and (ii) above.

From previous section, it is shown that the flexibility of the LQR derives from the dissimilarity or variability of the n elements in each of the rows of LQR Hat matrix. Hence, we compute the variance of the n elements in each vector $\boldsymbol{h}_i^{(LQR)}(\boldsymbol{b}) = \left(\boldsymbol{h}_{i1}^{(LQR)} \boldsymbol{h}_{i2}^{(LQR)} \ldots \boldsymbol{h}_{in}^{(LQR)}\right)$, $i = 1,2,\ldots,n$, of the LQR Hat matrix. This gives a $n \times 1$ vector of variances, say:

$$\boldsymbol{V} = \begin{pmatrix} Var\left(\boldsymbol{h}_1^{(LQR)}(\boldsymbol{b})\right) \\ Var\left(\boldsymbol{h}_2^{(LQR)}(\boldsymbol{b})\right) \\ \vdots \\ Var\left(\boldsymbol{h}_n^{(LQR)}(\boldsymbol{b})\right) \end{pmatrix}, \text{ for a given vector of bandwidths, } \boldsymbol{b}.$$

Next, the sum of the variances in $\boldsymbol{V}$ is obtained as $\sum \boldsymbol{V} = \sum_{i=1}^{n} Var(\boldsymbol{h}_i^{(LQR)}(\boldsymbol{b}))$.

Finally, in order to ensure that the proposed modified $PRESS^{**}(\boldsymbol{b})$ selects bandwidths according to objective (ii) above, we will have Q given by:

$$Q = \sum_{i=1}^{n} Var\left(\boldsymbol{h}_{\boldsymbol{i}}^{(\boldsymbol{LQR})}(\boldsymbol{b})\right) - \sum_{i=1}^{n} Var\left(\boldsymbol{h}_{\boldsymbol{i}}^{(\boldsymbol{OLS})}\right).$$

where $Var\left(\boldsymbol{h}_{\boldsymbol{i}}^{(\boldsymbol{OLS})}\right)$, $i = 1,2,\dots,n,$ is the variance of the elements in the i^{th} row of the n by n OLS Hat matrix, $\boldsymbol{H}^{(\boldsymbol{OLS})}$.

Therefore, the proposed bandwidths selection criterion comes out as:

$$PRESS^{sv}(\boldsymbol{b}) = \frac{PRESS}{\left[(n-k-1)\left(\frac{SSEmax-SSE(\boldsymbol{b})}{SSEmax}\right)\right]\left[\sum_{i=1}^{n} Var\left(\boldsymbol{h}_{\boldsymbol{i}}^{(\boldsymbol{LQR})}(\boldsymbol{b})\right)-\sum_{i=1}^{n} Var\left(\boldsymbol{h}_{\boldsymbol{i}}^{(\boldsymbol{OLS})}(\boldsymbol{b})\right)\right]}, \quad (10)$$

Clearly, the statistic, $\sum \boldsymbol{V} = \sum_{i=1}^{n} Var(\boldsymbol{h}_{\boldsymbol{i}}^{(\boldsymbol{LQR})}(\boldsymbol{b}))$ is computed from the entire n^2 elements of the LQR Hat matrix. Further, $PRESS^{sv}(\boldsymbol{b})$ is minimized at a given vector of bandwidths, $\boldsymbol{b}$ at which the difference $\sum_{i=1}^{n} Var\left(\boldsymbol{h}_{\boldsymbol{i}}^{(\boldsymbol{LQR})}(\boldsymbol{b})\right) - \sum_{i=1}^{n} Var\left(\boldsymbol{h}_{\boldsymbol{i}}^{(\boldsymbol{OLS})}(\boldsymbol{b})\right)$ (that is the flexibility between the LQR and OLS) is as large as possible.

An algorithm written in Matlab codes for implementing (10) is provided in Appendix B. It utilizes the genetic toolbox in Matlab by linking the function name with a @ handle.

III. APPLICATION

LQR that utilizes the proposed $PRESS^{sv}(\mathbf{b})$ criterion for bandwidths selection (herein designated $LQR*$ for ease of reference) is applied to two multiple response problems from RSM literature and two sets of simulated data and results compared with those from OLS and LQR that utilizes the $PRESS^{**}(\boldsymbol{b})$ criterion.

The performance statistics for comparison include SSE and Coefficient of Determination, (R^2), which respectively indicate the nearness of the fitted responses to the observed values and a measure of variability in the data that is explained or captured by each regression model.

For each model, the values of the sum of the variances, $\sum_{i=1}^{n} Var\left(\boldsymbol{h}_{\boldsymbol{i}}^{(.)}\right)$ of the elements in each row of the Hat matrix is presented under the column labelled SRV in Tables 4 and 9.

For the comparison of optimization results, the values of desirability measures in (4*) in the Appendix were used.

The best value for each performance statistics (goodness-of-fit and optimization solution) are shown in bold.

1. The chemical process data

This problem originates from Montgomery (2005) were it was analyzed using OLS. It involves three response variables, namely the y_1 (yield), y_2 (viscosity), and y_3 (molecular weight). Two inputs (factors) were found to influence these responses: reaction time (x_1) and temperature (x_2). A full second-order polynomial was found to be adequate for each of the response variables.

The process specifications for each response are as follows:

Maximize y_1 with lower limit $L = 78.5$, and target value, $\emptyset = 80$;
y_2 is to take a value in the range of $L = 62$ and $U = 68$ with target value, $\emptyset = 65$;
Minimize y_3 with upper limit $U = 3300$ and target value, $\emptyset = 3100$.

The data, collected via a Central Composite Design (CCD), is presented in Table 1. The optimal tuning parameters for the nonparametric models based on $PRESS^{sv}(\boldsymbol{b})$ and $PRESS^{**}(\boldsymbol{b})$ for $LQR*$ and LQR, respectively, are given in Table 2, and the corresponding locally adaptive bandwidths are shown in Table 3. The goodness-of-fit and optimization results for each of the regression models are presented in Tables 4 and 5, respectively.

Table 1: The chemical process data

i	x_1	x_2	y_1	y_2	y_3
1	0.1464	0.1464	76.5	62	2940
2	0.8536	0.1464	78.0	66	3680
3	0.1464	0.8536	77.0	60	3470
4	0.8536	0.8536	79.5	59	3890
5	0.0000	0.5000	75.6	71	3020
6	1.0000	0.5000	78.4	68	3360
7	0.5000	0.0000	77.0	57	3150
8	0.5000	1.0000	78.5	58	3630
9	0.5000	0.5000	79.9	72	3480
10	0.5000	0.5000	80.3	69	3200
11	0.5000	0.5000	80.0	68	3410
12	0.5000	0.5000	79.7	70	3290
13	0.5000	0.5000	79.8	71	3500

Table 2: Optimal tuning parameters for the chemical process data.

Response	Models	N^*	C^*
y_1	LQR	6.3536	0.0797
	LQR*	3.2291	0.0852
y_2	LQR	5.3234	0.0228
	LQR*	3.4926	0.0181
y_3	LQR	5.9081	0.0884
	LQR*	2.4262	0.0884

Table 3: Bandwidths obtained from optimal tuning parameters in Table 2.

i	y_1		y_2		y_3	
	LQR	LQR*	LQR	LQR*	LQR	LQR*
1	0.8298	0.3065	0.3787	0.2501	0.8558	0.3514
2	0.5710	0.2624	0.4143	0.2716	0.1901	0.0781
3	0.7435	0.2918	0.3609	0.2393	0.3790	0.1557
4	0.3122	0.2183	0.3520	0.2340	0.0012	0.0005
5	0.9851	0.3330	0.4587	0.2984	0.7838	0.3219
6	0.5020	0.2507	0.4321	0.2823	0.4780	0.1963
7	0.7435	0.2918	0.3343	0.2232	0.6669	0.2739
8	0.4848	0.2477	0.3432	0.2286	0.2351	0.0966
9	0.2432	0.2065	0.4676	0.3038	0.3700	0.1520
10	0.1742	0.1948	0.4410	0.2877	0.6219	0.2554
11	0.2260	0.2036	0.4321	0.2823	0.4330	0.1778
12	0.2777	0.2124	0.4498	0.2930	0.5410	0.2222
13	0.2605	0.2095	0.4587	0.2984	0.3521	0.1446

In Table 4, it can be noticed that LQR^* gives better SSE and R^2 across the three responses, signifying a model of better fit than the OLS and LQR. Further, LQR^* has largest SRV across the three responses as well. Having the largest SRV implies that the proposed bandwidths selection criterion, $PRESS^{sv}$, guarantees higher flexibility of the LQR^* over the OLS and the LQR.

Table 4: Goodness-of-fit of the regression models for the chemical process data.

Response	Model	$PRESS^{**}$	$PRESS^{sv}$	DF	SRV	SSE	R^2
y_1	OLS	-	-	7.0000	0.4167	0.4962	98.2735
	LQR	0.2526	-	6.5802	0.3920	0.4468	98.4456
	LQR*	-	1.4831	4.0351	0.6609	**0.2122**	**99.2618**
y_2	OLS	-	-	7.0000	0.4167	36.2242	89.9720
	LQR	13.0953	-	4.9192	0.5392	12.4398	96.5563
	LQR*	-	63.9644	4.0149	0.6642	**10.0008**	**97.2315**
y_3	OLS	-	-	7.0000	0.4167	207870	75.8990
	LQR	82840	-	5.0093	0.5325	77067	91.0648
	LQR*	-	393130	4.0006	0.6666	**65720**	**92.3804**

Figure 1 shows the plots of residuals of $\hat{y}_1$ (Top Left), $\hat{y}_2$ (Top Right), and $\hat{y}_3$ (Bottom Left) show that those from the LQR* are seen to lie relatively closest to the zero residual lines, indicative of relatively better fit of the LQR* to the given data.

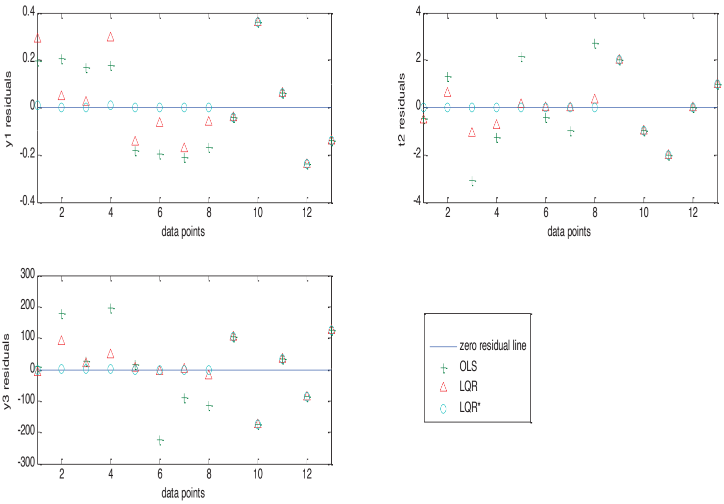

Figure 1: Graphical comparison of the residuals plots of response estimates.

The results presented in Table 5 show that LQR^* found a better setting of the explanatory variables that simultaneously optimizes the three responses with a desirability measure of 72.9%, indicating a product that meets approximately 73% of the production requirements as compared with the 41% for the LQR. The enhanced flexibility of LQR enables its exploration of the solution space for better optimal results.

Table 5: Optimization results based on the desirability measure for the chemical process data.

Models	x_1	x_2	Max($\hat{y}_1$)	$\emptyset(\hat{y}_2)$	min ($\hat{y}_3$)	d($\hat{y}_1$)	d($\hat{y}_2$)	d($\hat{y}_3$)	D(%)
OLS	0.4449	0.2226	78.7616	66.4827	3229.9	0.1744	0.5058	0.3504	31.3800
LQR	0.4892	0.2093	78.7993	66.1764	3188.5	0.1996	0.6079	0.5576	40.7450
LQR*	0.4473	0.2180	79.0816	65.0000	3088.1	0.3877	1.0000	1.0000	**72.9183**

2. Tire thread data

This data is taken from a study carried out in Derringer and Suich (1980) for the development of a material for the manufacture of tire thread. The inputs that were found to affect the quality of the tire thread include hydrated silica level (x_1), silane coupling agent (x_2), and sulphur level (x_3). Four outputs (responses) representing the different aspects in the quality of the tire thread are PICO Abrasive index (y_1), 200% modulus (y_2), Elongation at break (y_3), and Hardness (y_4).

The product specifications are as follows:

Maximize y_1, with L= 120 and target value $\emptyset$ =170;
Maximize y_2, with L =1000 and $\emptyset$ =2000;
$400 < y_3 < 600, \emptyset = 500; 60 < y_4 < 75, \emptyset = 67.5$.

For the OLS, apart from y_3 for which a first-order polynomial was specified, a full second-order polynomial was specified for the fitting of y_1, y_2, y_4. The data collected via a CCD is presented in Table 6. The optimal tuning parameters for the locally adaptive bandwidths for the respective regression models are shown in Table 7, and the corresponding adaptive bandwidths are presented in Table 8. Tables 9 and 10 present the regression statistics and optimization results, respectively.

Table 6: The Tire thread data.

i	x_1	x_2	x_3	y_1	y_2	y_3	y_4
1	0.1933	0.1933	0.1933	102	900	470	67.5
2	0.8067	0.1933	0.1933	120	860	410	65.0
3	0.1933	0.8067	0.1933	117	800	570	77.5
4	0.8067	0.8067	0.1933	198	2294	240	74.5
5	0.1933	0.1933	0.8067	103	490	640	62.5
6	0.8067	0.1933	0.8067	132	1289	270	67.0
7	0.1933	0.8067	0.8067	132	1270	410	78.0
8	0.8067	0.8067	0.8067	139	1090	380	70.0
9	0.0000	0.5000	0.5000	102	770	590	76.0
10	1.0000	0.5000	0.5000	154	1690	260	70.0
11	0.5000	0.0000	0.5000	96	700	520	63.0
12	0.5000	1.0000	0.5000	163	1540	380	75.0
13	0.5000	0.5000	0.0000	116	2184	520	65.0
14	0.5000	0.5000	1.0000	153	1784	290	71.0
15	0.5000	0.5000	0.5000	133	1300	380	70.0
16	0.5000	0.5000	0.5000	133	1300	380	68.5
17	0.5000	0.5000	0.5000	140	1145	430	68.0
18	0.5000	0.5000	0.5000	142	1090	430	68.0
19	0.5000	0.5000	0.5000	145	1260	390	69.0
20	0.5000	0.5000	0.5000	142	1344	390	70.0

Table 7: Optimal tuning parameters for the for the tire thread data.

Response	Model	N^*	C^*
y_1	LQR	7.1522	0.1907
	LQR*	6.3585	0.2191
y_2	LQR	8.7784	0.0157
	LQR*	4.1621	0.0195
y_3	LQR	3.7683	0.0174
	LQR*	3.5231	0.7037
y_4	LQR	10.6275	0.0394
	LQR*	5.5450	0.0431

Table 8: Bandwidths obtained from optimal tuning parameters in Table 7.

i	y_1		y_2		y_3		y_4	
	LQR	**LQR***	**LQR**	**LQR***	**LQR**	**LQR***	**LQR**	**LQR***
1	0.3873	0.3399	0.2579	0.1116	0.2248	0.1745	0.4497	0.2117
2	0.3701	0.3272	0.2375	0.1007	0.1832	0.1764	0.3598	0.1398
3	0.3730	0.3293	0.2070	0.0844	0.2940	0.1712	0.8089	0.4997
4	0.2956	0.2721	0.9686	0.4905	0.0656	0.1819	0.7011	0.4133
5	0.3863	0.3392	0.0489	0.0001	0.3424	0.1690	0.2700	0.0678
6	0.3587	0.3187	0.4563	0.2173	0.0863	0.1809	0.4317	0.1973
7	0.3587	0.3187	0.4466	0.2122	0.1832	0.1764	0.8268	0.5141
8	0.3520	0.3138	0.3548	0.1633	0.1625	0.1774	0.5395	0.2837
9	0.3873	0.3399	0.1917	0.0763	0.3078	0.1706	0.7550	0.4565
10	0.3377	0.3032	0.6607	0.3264	0.0794	0.1812	0.5395	0.2837
11	0.3930	0.3441	0.1560	0.0572	0.2594	0.1728	0.2880	0.0822
12	0.3291	0.2968	0.5842	0.2856	0.1625	0.1774	0.7191	0.4277
13	0.3739	0.3300	0.9125	0.4606	0.2594	0.1728	0.3598	0.1398
14	0.3386	0.3039	0.7086	0.3519	0.1002	0.1803	0.5754	0.3125
15	0.3577	0.3180	0.4619	0.2203	0.1625	0.1774	0.5395	0.2837
16	0.3577	0.3180	0.4619	0.2203	0.1625	0.1774	0.4856	0.2405
17	0.3510	0.3131	0.3828	0.1782	0.1971	0.1758	0.4676	0.2261
18	0.3491	0.3116	0.3548	0.1633	0.1971	0.1758	0.4676	0.2261
19	0.3462	0.3095	0.4415	0.2095	0.1694	0.1770	0.5035	0.2549
20	0.3491	0.3116	0.4843	0.2323	0.1694	0.1770	0.5395	0.2837

Table 9 shows that $LQR*$ gives the best SSE and R^2 for y_1, y_3, and y_4, with competitive respective results for y_2. Further, it gives the largest SRV for the four responses, where the largest SRV indicates a higher flexibility of LQR* in comparison to that of OLS and LQR models.

Table 9: Regression statistics of each model and response for the tire thread data.

Response	Model	**PRESS****	$PRESS^{sv}$	**DF**	SRV	**SSE**	R^2
y_1	OLS	-	-	10.0000	0.4737	1841.1380	83.6544
	LQR	238.9268	-	5.6592	0.6724	148.9582	98.6775
	LQR*	-	129.8541	5.2560	0.7109	**130.2447**	**98.8437**
y_2	OLS	-	-	10.0000	0.4737	1204800	71.2487
	LQR	171260	-	6.1151	0.6415	148440	96.4575
	LQR*	-	564280	5.0377	0.7331	**51164**	**98.7790**
y_3	OLS	-	-	16.0000	0.1579	72914	68.0449
	LQR	4905.6	-	5.0414	0.7325	2803.1	98.7715
	LQR*	-	2319.6	5.0084	0.7360	**2800**	**98.7729**
y_4	OLS	-	-	10.0000	0.4737	50.8573	86.7296
	LQR	6.7645	-	7.7850	0.5091	11.9695	96.8767
	LQR*	-	4.1705	5.1290	0.7239	**4.2704**	**98.8857**

On the average, the plots of residuals of $\hat{y}1$ (Top Left), $\hat{y}2$ (Top Right), $\hat{y}3$ (Bottom Left), and $\hat{y}4$ (Bottom Right) from Figure 2 show that those from the LQR* are seen to lie comparatively closest to the zero residual lines. This reveals that the LQR* does a better job in terms of the estimation to the given data.

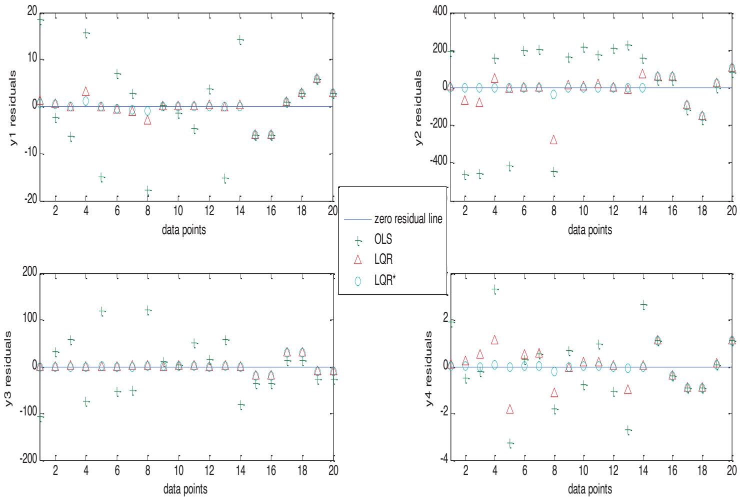

Figure 2: Graphical comparisons of the residual plots for the tire thread data.

In Table 10, it can be seen that the comparatively better regression statistics of $LQR*$ results in a better set of optimal values of the explanatory variables which corresponds to a desirability measure of 88.012%. This solution indicates a tire thread that meets approximately 88% of the production specifications.

Table 10: Optimization results based on the desirability measure for the tire thread data.

Model	x_1	x_2	x_3	$(\hat{y}_1)$	$(\hat{y}_2)$	$(\hat{y}_3)$	$(\hat{y}_4)$	d$(\hat{y}_1)$	d$(\hat{y}_2)$	d$(\hat{y}_3)$	d$(\hat{y}_4)$	**D(%)**
OLS	0.5043	0.5892	0.0000	140.0	2070	451.8	69.02	0.4002	1.0000	0.5181	0.7966	63.7500
LQR	0.8433	0.4373	1.0000	160.3	1840	463.6	69.6	0.8059	0.8404	0.6361	0.7150	74.4994
LQR*	1.0000	0.5529	1.0000	167.7	2118	499.9	70.3	0.9532	1.0000	1.0000	0.6295	**88.0120**

3. *Simulation study*

The simulation study focuses on how each of the regression models performs when random errors with different variances, σ^2 are added to the outputs generated from three different polynomial models involving one, two and three explanatory variables and given as follows:

Polynomial model I:

$$y_i = 29 + 6x_{1i} - 10x_{1i}^2 + \varepsilon_i,\ \ i = 1, 2, \dots, n, \tag{11}$$

Polynomial model II:

$$y_i = 54 + 10x_{1i} + 55x_{2i} - 5x_{1i}^2 - 53x_{2i}^2 - 9x_{1i}x_{2i} + \varepsilon_i,\ \ i = 1, 2, \dots, n, \tag{12}$$

Polynomial model III:

$$y_i = 63 - 14x_{1i} + 23x_{2i} + 3x_{3i} + 17x_{1i}^2 + 0.5x_{2i}^2 - 4x_{3i}^2 - 17x_{1i}x_{2i} + 5x_{1i}x_{3i} - 2x_{2i}x_{3i} + \varepsilon_i,\quad i = 1, 2, \dots, n, \tag{13}$$

where x_{i1}, x_{i2}, and x_{i3}, $i = 1,2,\dots,n$, are the respective values of the explanatory variables x_1, x_2 and x_3 from the coded values of the explanatory variables in Table 1 for the Polynomial models I and II, and Table 6 for the Polynomial model III, ε_i, $i = 1,2,\dots,n$, is the i^{th} value of the normally distributed random error, $\boldsymbol{\varepsilon}$ with mean 0 and variance $= \sigma^2$.

The random error variance varied from 1, 16, 49 and 100 for each of the respective polynomials in order to add more complex trends and patterns to the simulated data. Data from Table 1 and Table 6 are used as in Equations (11), (12) and (13) since they are standard CCD designs involving two (x_1 and x_2) and three (x_1, x_2 and x_3) input (explanatory) variables, respectively. (Wan and Birch, 2011; He *et al.,* 2012.).

For each of the 500 data sets from each of polynomials, σ^2 varies from 1, 16, 49, and 100. As the variance increases from 1 through 100, the discrepancies between the output generated from Equations (11), (12), and (13) and the true (underlying) output given by $y_i = 29 + 6x_{1i} - 10x_{1i}^2$ in Equation (11), $y_i = 54 + 10x_{1i} + 55x_{2i} - 5x_{1i}^2 - 53x_{2i}^2 - 9x_{1i}x_{2i}$ from Equation (12) and $y_i = 63 - 14x_{1i} + 23x_{2i} + 3x_{3i} + 17x_{1i}^2 + 0.5x_{2i}^2 - 4x_{3i}^2 - 17x_{1i}x_{2i} + 5x_{1i}x_{3i} - 2x_{2i}x_{3i}$ from Equation (13) become larger, and the goal is to find out how each of the models is able to deal with the increasing difficulty of getting estimates of the output that are as close to the true output as possible. It is also of importance to see how each model handles these discrepancies as the dimension or the number of explanatory variables increases from one through three in Equations (11) to (13).

For the 500 data each for variance of random error ranging from 1 through 100, the average of the optimal bandwidths are 0.6299, 0.4543, and 0.3379 for Polynomial model I, Polynomial model II, and Polynomial model III, respectively, when LQR model is applied. For LQR*, the respective averages are 0.1552, 0.1558, and 0.2667.

A measure of model performance used here is the Average Sum of Squares of Error (AVESSE) given as:

$$\text{AVESSE} = \frac{\sum_{l=1}^{500}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{500}, \tag{14}$$

where y_i and $\hat{y}_i$, $i = 1,2,\dots,n$, are the raw (simulated) response and the model estimated response, respectively. AVESSE measures how close or accurate the estimated responses are to the raw response and the computer AVESSE computed from each regression model using 500 data sets is presented in Table 11.

Table 11: Comparison of the AVESSE for the simulated data.

Polynomial	$\sigma^2(\varepsilon)$	OLS	LQR	LQR^*
I	1.0	0.8302	0.7213	**0.6513**
	16.0	13.3046	10.6168	**9.0211**
	49.0	40.7766	35.3550	**33.0712**
	100.0	83.9231	70.3257	**68.8800**
II	1.0	0.5862	0.3365	**0.3152**
	16.0	9.3652	5.6501	**5.6075**
	49.0	28.8809	18.0361	**16.0720**
	100.0	58.4639	37.6078	**32.3240**
III	1.0	0.8286	0.8255	**0.5738**
	16.0	13.3679	9.8815	**7.4196**
	49.0	40.8621	34.0595	**22.4840**
	100.0	82.8702	71.5926	**42.6964**

Results in Table 11 show that LQR* gives the best AVESSE across the three polynomials irrespective of the value of the random error variance. The results from LQR^* are seen to be progressively better than those from its competitors as the variance of the error term increases, meaning that bandwidths selected by the proposed criterion offer higher flexibility as well as robustness against the deviation of the responses from the true polynomial models. This is an important feature for a regression model since it is an arduous task to determine a model that can match a given data with 100% precision (Box, 1976).

Although, there is no basis for comparing the values of PRESS** and $PRESS^{sv}$ criterion since their respective penalizing term differs, Table 12 presents their averages for the sake of completeness.

Table 12: The AVESSE of PRESS** and $\boldsymbol{PRESS^{sv}}$ for the simulated data

Polynomial model	$\sigma^2(\varepsilon)$	AVE$\boldsymbol{PRESS}^{**}$	AVE$\boldsymbol{PRESS^{sv}}$
I	1	0.1557	61,829
	16	2.4213	106,700
	49	7.3439	152,800
	100	15.0140	1,139,400
II	1	1.3422	2,901
	16	4.9199	10,987
	49	13.3287	26,876
	100	26.7416	44,930
III	1	2.7708	3,899
	16	5.5789	10,077
	49	13.0771	56,570
	100	23.0506	50717

From results in Table 12, across the three polynomials, it is observed that the average of each of the two versions of PRESS increases as the variance of the random error increases from 1 through 100, except the one for $PRESS^{sv}$ at variance equal to 100.

IV. CONCLUSION

In this paper, a modification of the PRESS** criterion to suit the selection of bandwidths for LQR in the response surface settings was proposed. The modified PRESS** criterion, denoted as $PRESS^{sv}$ criterion, involves the replacement of the DF term in the denominator of the PRESS** criterion with the difference in the sum of the variances of the rows of the Hat matrix of the LQR and that of the OLS models. This replacement was done in order to improve on the flexibility of the LQR model which happens to be the most appealing feature of nonparametric regression models in general.

From the data analyzed, LQR utilizing $PRESS^{sv}$ (designated $LQR*$) gives the best SSE and R2 in seven out of the seven responses in the two problems from literature as well as the best SSE in the entire twelve simulated responses. Much more significantly, $LQR*$ gives superior optimization results in the two problems taken from the literature, trumping those from OLS and the LQR. Specifically, better optimization result (Tables 5 and 10) translates to better use of scarce resources (raw material, time, e.tc). It enhances products' conformity to standards as well as costumers' satisfaction.

V. LITERATURE CITED

Anderson-Cook, C.M and Prewitt, K. 2005, Some guidelines for using nonparametric models for modeling data from response surface designs, *Journal of Modern Applied Statistical Models,* 4, 106-119.

Box, G.E.P. and Wilson B. 1951, On the experimental attainment of optimum conditions, *Journal of the Royal Statistical Society,* Series B, 13, 1- 45.

__________. 1976, Science and statistics, *Journal of the American Statistical Association,* 71(356), 791-799.

Castillo, D.E. 2007, *Process optimization: A statistical method.* New York: Springer International Series in Operations Research and Management Science.

Cleveland, W.S. 1979, Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association,* 368(74), 829 – 836.

Derringer, G., and Suich, R. 1980, Simultaneous optimization of several response variables, *Journal of Quality Technology,* 12(4), 214 – 219.

Edionwe, E. and Mbegbu, J.I. 2014, Local bandwidths for improving the performance statistics of model robust regression 2, *Journal of Modern Applied Statistical Methods,* 13(2), 506-527.

________________________, and Chinwe, R. 2016, A new function for generating local bandwidths for semiparametric MRR2 model in response surface methodology, *Journal of Quality Technology,* 48(4), 388 –404.

________________________, and Iguodala, W.A. 2017, Improving the performance of model robust regression 2 (MRR2) method using new adaptive mixing parameters and a modified penalized error sum of squares, *Journal of the Nigerian Association of Mathematical Physics,* 41, 229-240.

________________________, Ekhosuehi, N., and Obiora-Ilouno, H.O. 2018, An improved robust regression model for response surface methodology, *Croatian Operational Research Review,* 9, 317 – 330.

Fan, J., and Gijbels, I. 1996, *Local polynomial modeling and its applications.* Chapman and Hall, London.

Hill, W.J, and Hunter, W.G. 1996, A review of response surface methodology: A literature review, Technometrics, 8, 571 – 590.

Hardle, W., Muller, M., Sperlich, S., and Werwatz, A. 2005, *Nonparametric and semiparametric models: An introduction.* Berlin: Springer-Verlag.

Harrington, E.C. 1965, The desirability function, *Industrial Quality Control*, 21(10), 494 – 498.

He, Z., Zhu, P.F., and Park, S.H. 2012, A robust desirability function for multi-response surface optimization, *European Journal of Operational Research,* 221, 241-247.

Loader, C.R. 1999, Bandwidth selection: Classical or plug-in?, The Annals of Statistics, *2, 415-438.*

Mays, J.E. and Birch, J.B. 1998, Smoothing considerations in nonparametric and semiparametric regression, *Technical Report Number 98-2,* Department of Statistics, Virginia Polytechnic Institute and State University.

Mays, J.E., Birch, J.B, and Starnes, B.A. 2001, Model-robust regression: combining parametric, nonparametric, and semi-parametric models, *Journal of Nonparametric Statistics,* 13, 245-277.

Montgomery, D.C. 2005, *Design and analysis of experiments.* Sixth ed. Wiley, New York.

Myers, R.H., 1999. Response surface methodology - current status and future directions, Journal of Quality Technology 31, 30-44.

__________, Montgomery, D.C., and Anderson-Cook, C.M. 2009, *Response surface methodology: process and product optimization using designed experiments.* 3rd Edition, Wiley, ISBN: 978-0-470-17446-3.

Pickle, S.M., Robinson, T.J., Birch, J.B., and Anderson-Cook, C.M. 2008, A Semi-parametric model approach to robust parameter design, *Journal of Statistical Planning and Inference,* 138, 114-131.

Rajewski, J., and Dobrzynska-Inger, A. 2021, Application of Response Surface Methodology (RSM) for the Optimization of Chromium(III) Synergistic Extraction by Supported Liquid Membrane, 11, 854, 1-17.

Wan, W. and Birch, J.B. 2011, A semi-parametric technique for multi-response optimization, *Journal of Quality and Reliability Engineering International,* **27**, 47-59.

Yeniay, O. 2014, Comparative study of algorithm for response surface optimization, *Journal of Mathematical and Computational Applications,* 19, 93 -104.

Appendix A:

If the response is of nominal-the-better (NTB) type where the p^{th} response acceptable value lies between an upper limit, U and a lower limit, L, $d_p\left(\hat{y}_p(\boldsymbol{x})\right)$ is given as:

$$d_p\left(\hat{y}_p(\boldsymbol{x})\right) = \begin{cases} 0 & \hat{y}_p(\boldsymbol{x}) < L \\ \left\{\frac{\hat{y}_p(x)-L}{\emptyset-L}\right\} & L \leq \hat{y}_p(\boldsymbol{x}) < \emptyset, \\ \left\{\frac{U-\hat{y}_p(x)}{U-\emptyset}\right\} & \emptyset \leq \hat{y}_p(\boldsymbol{x}) \leq U, \\ 0 & \hat{y}_p(\boldsymbol{x}) > U, \end{cases} \quad (1^*)$$

where $\emptyset$ is the target value of the p^{th} response.

If the goal is to maximize the p^{th} response, $d_p\left(\hat{y}_p(\boldsymbol{x})\right)$ is given by a one-sided transformation as:

$$d_p\left(\hat{y}_p(\boldsymbol{x})\right) = \begin{cases} 0 & \hat{y}_p(\boldsymbol{x}) < L, \\ \left\{\frac{\hat{y}_p(x)-L}{\emptyset-L}\right\} & L \leq \hat{y}_p(\boldsymbol{x}) \leq \emptyset, \\ 1 & \hat{y}_p(\boldsymbol{x}) > \emptyset, \end{cases} \quad (2^*)$$

where $\emptyset$ is interpreted as a large enough value of the p^{th} response.

If the goal is to minimize the p^{th} response, $d_p\left(\hat{y}_p(\boldsymbol{x})\right)$ is given by a one-sided transformation as:

$$d_p\left(\hat{y}_p(\boldsymbol{x})\right) = \begin{cases} 1 & \hat{y}_p(\boldsymbol{x}) < \emptyset, \\ \left\{\frac{U-\hat{y}_p(x)}{U-\emptyset}\right\} & \emptyset \leq \hat{y}_p(\boldsymbol{x}) \leq U, \\ 0 & \hat{y}_p(\boldsymbol{x}) > U, \end{cases} \quad (3^*)$$

where $\emptyset$ is a small enough value of the p^{th} response.

The overall objective of the desirability criterion is getting the values of the explanatory variables that maximize the geometric mean (D) of all the individual desirability measures given as:

$$D = maximize\left(\left(\prod_{p=1}^{m} d_p\left(\hat{y}_p(\boldsymbol{x})\right)\right)^{1/m}\right) \times 100\%, \quad (4^*)$$

Appendix B: Algorithm for implementing the LQR procedure using the modified PRESS**

```
1.  function pressdd=my_lqr_modified_press(D)
2.  x1; vector of values of explanatory variable
3.  x2; vector of values of explanatory variable
4.  ⋮
5.  Xj; vector of values of explanatory variable
6.  y;  vector of values of response variable
7.  e=2.7183;
8.  k=j; % No of explanatory variables
9.  n=length(x1); % No of data points
10. const=ones(n,1); % vector of ones
11. X=[const x1 x2 … xj x1.^2 x2.^2 … xj.^2]; % LQR model matrix
12. % use #13 - #18 to preallocate dimensions for vectors/arrays
13. yLQR_max=zeros(n,1); % empty vector to store y estimates for large bandwidth
14. yLQRcv=zeros(n,1);   % empty vector to store leave-one-out estimates of y
15. a=zeros(n,1);        % empty vector to store diag elements of LQR Hat matrix
16. yLQR=zeros(n,1);     % empty vector to store y estimates for bandwidths
17. varHatmin=zeros(n,1);% empty vect to store var of elements in row of HLQR_max
18. varHatmax=zeros(n,1);% empty vect to store variance of elements in row of HLQR
19. % use #20 - #31 to get sum of minimum variance(Vmin) and maximum SSE (SSEmax)
20. bmax=9999999999;
21. for i=1:n;
22. w1max=((1/e).^(((x1(i)-x1)/bmax).^2)).*((1/e).^(((x2(i)-x2)/bmax).^2))...
                                    .*((1/e).^(((xj(i)-xj)/bmax).^2));
23. WWmax=sum(w1max);
24. kerweight_max=w1max./WWmax;
25. Wmax=diag(kerweight_max); % n by n diagonal weight matrix
26. HatLQR_max=X(i,:)*((X'*Wmax*X)\(X'*Wmax)); % ith row of the LQR Hat matrix
27. yLQR_max(i)=HatLQR_max*y; % LQR estimate of y for large bandwidth;
28. varHatmin(i)=var(HatLQR_max); % variance of the elements in each row of HLQR_max
29. Vmin=sum(varHatmin); % sum of variance of the elements in each row of HLQR_max
30. end
31. SSEmax=sum((y-yLQR_max).^2);
32. % use #33 - #41 to get the kernel weights
33. D;% vect that contains values of N as D(1) and C as D(2), as required by GATool;
34. T=sum(y);
35. bloc=D(1)*(D(2)*T-y)/(T*((D(2)*n)-1)); % vector of locally adaptive bandwidths
36. %%%%%%%%%%%%%%%%%%%
37. for i=1:n;
38. w1=((1/e).^(((x1(i)-x1)./bloc).^2)).*((1/e).^(((x2(i)-x2)./bloc).^2))...
39.                                    .*((1/e).^(((xj(i)-xj)./bloc).^2))
40. WW=sum(w1);
41. kerweight=w1./WW;
42. % use #43 - #50 to obtain leave-one-out estimates of y
43. kerweight(i,:)=[];y(i,:)=[];X(i,:)=[];
44. W=diag(kerweight); % diagonal weight matrix for leave-one-out estimate of y
45. a2=(X'*W*X)\X'*W*y;
46. % use #47 - #48 to restore original dimensions of arrays
47. X=[const x1 x2 ... xj x1.^2 x2.^2...xj.^2];
```

```
48. y=[76.5 78 77 79.5 75.6 78.4 77 78.5 79.9 80.3 80 79.7 79.8]';
49. kerweight=w1./WW;
50. yLQRcv(i)=X(i,:)*a2;
51. % use #52 -#58 get LQR estimates of responses & maximum sum of variance(Vmax)
52. W=diag(kerweight);  % n by n diagonal weight matrix for estimate of y
53. HLQR=X(i,:)*((X'*W*X)\X'*W);
54. a(i)=HLQR(1,i);
55. yLQR(i)=HLQR*y;
56. varHatmax(i)=var(HLQR);% variance of the elements in each row of HLQR
57. Vmax=sum(varHatmax);% sum of variance of the elements in each row of HLQR
58. end
59. df=n-sum(a),
60. % use #61 - #65 Get the values of goodness-of-fit
61. PRESS=sum((y-yLQRcv).^2);
62. SSE=sum((y-yLQR).^2);
63. ymean=mean(y);
64. ySSM=sum((y-ymean).^2);
65. Rsqr=100*(1-(SSE/ySSM));
66. % use #67 - #72 to ensure bandwidths and PRESS_SV are within acceptable ranges
67. PRESS_SV=PRESS/(((Vmax-Vmin))*((n-k-1)*((SSEmax-SSE)/SSEmax)));
68. if PRESS_SV<0; pdstar=9215918691;
69. elseif min(bloc)<0; pdstar=7777777777;
70. elseif max(bloc)>1;pdstar=88888882929;
71. else pdstar=PRESS/(((Vmax-Vmin))*((n-k-1)*((SSEmax-SSE)/SSEmax)));
72. end
73. pressdd=pdstar,
74. df,
75. Vmax,
76. SSE,
77. Rsqr,
78. N=D(1),
79. C=D(2),
80. locally_adaptive_bandwidths=bloc,
```