A new regression model that combines the LLR estimates of both the raw response and its residuals

Eguasa Omo[1,a], Julian I. Mbegbu[2] and E. Edionwe[3]

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

The modeling phase of response surface methodology (RSM) involves the use of regression models to estimate the functional relationship between the response and the explanatory variables using data obtained from a suitable experimental design. In RSM, the Ordinary Least Squares (OLS) is traditionally used to model the data via user-specified low-order polynomials. The OLS model is found to perform poorly if for instance the constant variance assumption is violated. Recently, nonparametric regression models, such as the Local Linear Regression (LLR), have been proposed to address the model inadequacy issue associated with the use of the OLS model. The LLR model is flexible, hence, can capture local trend and structure in the data that are missed by an inadequate OLS model. The successful application of the LLR model has been limited to studies with three unique features, namely: a single explanatory variable, fairly large sample sizes and space-filling designs. Therefore, the LLR model is scantily used in RSM which general underpinning include economy of data points (small sample size), typically sparse data, and oftentimes, more than one explanatory variables. In this paper, we propose a new nonparametric regression model that incorporates the smoothing of residuals to provide a second opportunity of fitting part of the data that is not captured by the LLR model. Using an example from the literature, it is observed that the goodness-of-fits of the proposed model are considerably better when compared with those of the OLS and the LLR models.

Keywords: Genetic algorithm, Local linear regression, Locally Adaptive bandwidths, Mixing parameter, Ordinary least squares, Response surface studies.

1. Introduction

Response surface methodology (RSM) is sequential statistical tool employed by statistician and engineers for empirical model building, such that the response variable is optimized. RSM consist of three main phases namely, Experimental Design Phase, Modeling phase and the Optimization phase of the fitted regression models.

The peculiarity of RSM data which include, small sample size, sparse data and curse of dimensionality have reduced the performance of nonparametric regression models in terms of goodness of fit statistics and optimization result.

1.1 Existing Modeling Procedures for RSM

We presented existing modeling procedures for RSM. Hence, we briefly discussed and highlight the limitations OLS and LLR models and present the variable bandwidths proposed by Edionwe *et al.* (2016).

[a] *Corresponding author's email*: oeguasa@biu.edu.ng

1.2 The Parametric Regression Model (OLS)

The general parametric regression model can be written as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}_{(n\times 1)}$ is a vector of response, $\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}_{(n\times(k+1))}$ is the

model matrix, where $\boldsymbol{X} = \boldsymbol{X}^{(\boldsymbol{OLS})}$, $\boldsymbol{\varepsilon}$ is a vector of error term.

The common technique for the estimated responses for the i^{th} data points can be written as:

$$\hat{\boldsymbol{y}}_i^{(OLS)} = \boldsymbol{x}_{\boldsymbol{i}}'^{(OLS)}\left(\boldsymbol{X}'^{(OLS)}\boldsymbol{X}^{(OLS)}\right)^{-1}\boldsymbol{X}'^{(OLS)}\boldsymbol{y}, \quad i = 1,2,\dots,n. \tag{2}$$

In matrix form, equation (2) is expressed as:

$$\hat{\boldsymbol{y}}^{(OLS)} = \boldsymbol{H}^{(OLS)}\boldsymbol{y} = \begin{bmatrix} \boldsymbol{h}_{\boldsymbol{1}}^{(OLS)} \\ \boldsymbol{h}_{\boldsymbol{2}}^{(OLS)} \\ \vdots \\ \boldsymbol{h}_{\boldsymbol{n}}^{(OLS)} \end{bmatrix}\boldsymbol{y}, \tag{3}$$

where the $1 \times n$ vector $\boldsymbol{h}_{\boldsymbol{i}}^{(OLS)}$ is the i^{th} row of the $n \times n$ OLS Hat matrix.

The drawback of the parametric regression model is that, if misspecified, the estimates are usually biased (Swamy *et al*., 2008 and Fathi *et al*., 2011).

1.3 The Local Linear Regression Model (LLR)

From typical weighted least squares theory, the **LLR** estimator $\hat{\boldsymbol{y}}_i^{(LLR)}$ of $\boldsymbol{y}_i$ is given as:

$$\hat{\boldsymbol{y}}_i^{(LLR)} = \boldsymbol{x}_{\boldsymbol{i}}'^{(LLR)}\left(\boldsymbol{X}'^{(LLR)}\boldsymbol{W}_{\boldsymbol{i}}\boldsymbol{X}^{(LLR)}\right)^{-\boldsymbol{1}}\boldsymbol{X}'^{(LLR)}\boldsymbol{W}_{\boldsymbol{i}}\boldsymbol{y} \tag{4}$$

where $\boldsymbol{X}^{(LLR)}$ is the LLR model matrix which depends solely on the number of explanatory variable utilized in the experiment, $\boldsymbol{W}_{\boldsymbol{i}}$**,** is the diagonal matrix of kernel (Gaussian) weight function and $\boldsymbol{x}_{\boldsymbol{i}}'^{(LLR)}$ is the i^{th} row of the LLR model matrix.

In terms of locations of the LLR Hat matrix, equation (4) can be expressed as:

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}}^{(\boldsymbol{LLR})} = \boldsymbol{h}_{\boldsymbol{i}}'^{(LLR)}\boldsymbol{y}, \quad i = 1,\ 2\ ,\ \dots,\ n \tag{5}$$

The drawback of LLR model is that, it suffers high bias in regions where the data exhibits curvature (Hastie, *et al*., 2009 and Rivers, 2009).

1.4 Review of Bandwidths selectors

The choice of bandwidth for a nonparametric regression models is a critical criterion and challenging in regression analysis (Kai, 2009, Osemwenkhae and Ogbeide, 2010 and Aydin *et al.*, 2013). Bandwidth selection was designed to minimize bias and variance of the estimate (Rivers, 2009).

A bandwidth, b, is said to be fixed if its value is constant for all the locations in a given regression technique, otherwise it is referred to as locally adaptive bandwidths (Prewitt and Lohr, 2006).

Hence, the kernel function, $K(.)$ employed in RSM is the simplified Gaussian kernel given in Wan and Birch (2011) as:

$$K\left(\frac{x_i-x_0}{b}\right) = K\left(\frac{x_0-x_i}{b}\right) = e^{-(\frac{x_i-x_0}{b})^2}, i = 1, 2, \; . \; . \; . \; , n. \quad (6)$$

where the kernel weights w_{i0} in the kernel weight matrix is given as:

$$w_{i0} = \frac{K(\frac{x_i-x_0}{b})}{\sum_{j=1}^{n} K(\frac{x_j-x_0}{b})}, \quad i = 1, \; 2 \; , \; . \; . \; . \; , n \, . \quad (7)$$

According to Wan (2007), $K\left(\frac{x_i-x_0}{b}\right)$ in equation (6) is referred to as kernel function which regulates the shape of the kernel weights (e.g. Gaussian kernel),x_0 is a dummy known as target point, b, is called bandwidth.

A situation where more than one explanatory variable is applied in the model matrix $\boldsymbol{X}^{(LLR)}$, the kernel weights w_{i0} is a product from simplified Gaussian kernel given as:

$$w_{i0} = \prod_{j=1}^{k} K\left(\frac{x_{ij}-x_0}{b}\right) / \sum_{i=1}^{n} \prod_{j=1}^{k} K\left(\frac{x_{ij}-x_0}{b}\right), \; i = 1, \; . \; . \; . \; , n \, . \quad (8)$$

(Mays *et al*., 2001 and Pickle, 2006).

For data which originated from RSM, the vector of optimal bandwidths $\Phi = [b_1^*, \; b_1^*, \dots, b_n^*]$ is obtained based on the minimization of the Penalized Prediction Error Sum of Squares ($PRESS^{**}$) (Wan and Birch, 2011). The form of the $PRESS^{**}$ criterion for selecting the bandwidths is given as:

$$PRESS^{**}(\Phi) = \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_{i,-i}(.)\right)^2}{n - trace\left(H^{(.)}(\Phi)\right) + (n-k-1)\frac{SSE_{max}-SSE_{\Phi}}{SSE_{max}}}, \quad (9)$$

where SSE_{max} is the maximum sum of squared errors obtained as the $b_1, b_2, \dots, b_n$ approaches infinity, SSE_{Φ} is the sum of squared errors associated with a set of bandwidths $b_1, b_2, \dots, b_n$, $tr\left(H^{(.)}\Phi\right)$ is the trace of the Hat matrix and $\hat{y}_{i,-i}$ is the leave-one-out cross-validation estimated value of y_i with the i^{th} observation left out (Mays *et al*., 2001 and Wan and Birch, 2011).

1.5 Locally Adaptive Bandwidths

Edionwe *et al*. (2016) presented locally adaptive bandwidths selector given as:

$$b_i = \frac{b^* N\left(C\sum_{i=1}^{n} y_i - y_i\right)}{(Cn-1)\sum_{i=1}^{n} y_i}, \qquad i = 1,2,\dots,n, \tag{10}$$

where b^* is a fixed optimal bandwidth,y_i, $i = 1,2,\dots,n$, could be taken as y_i or any statistics that mirrors the insufficiencies in the OLS estimates of the responses, $N > 0$, and $c \geq 0$, is a parameter introduced to address the problem of clustering within a small range between [0, 1].

2. Methodology

The nonparametric regression model is not restricted to a user specified form as in the parametric counterpart. In spite of its flexibility, nonparametric regression models are challenged in a study such as RSM due to three important aspects in RSM namely,

- Sparseness of RSM data
- Cost efficient design (small sample sizes)
- The study utilizes more than one explanatory variable (a term referred to as curse of dimensionality).

2.1 The Proposed Nonparametric Regression Model (PNRM)

The model is referred to as proposed Nonparametric Regression Model (PNRM**)** for ease of reference. PNRM can be viewed in terms of location of the estimated response, *i*th row of the model matrix, model matrix and diagonal matrix of kernel weight such that the *i*$^{\text{th}}$ location of the fitted response is given as:

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}}^{PNRM} = \boldsymbol{x}_{\boldsymbol{i}}^{\prime(\boldsymbol{LLR})}\left(\boldsymbol{X}^{\prime(\boldsymbol{LLR})}\boldsymbol{W}_{\boldsymbol{i}}\boldsymbol{X}^{(\boldsymbol{LLR})}\right)^{-\mathbf{1}}\boldsymbol{X}^{\prime(\boldsymbol{LLR})}\boldsymbol{W}_{\boldsymbol{i}}\boldsymbol{y} + \lambda\, \boldsymbol{x}_{\boldsymbol{i}}^{\prime(\boldsymbol{LLR})}\left(\boldsymbol{X}^{\prime(\boldsymbol{LLR})}\boldsymbol{W}_{\boldsymbol{i}}\boldsymbol{X}^{(\boldsymbol{LLR})}\right)^{-\mathbf{1}}\boldsymbol{X}^{\prime(\boldsymbol{LLR})}\boldsymbol{W}_{\boldsymbol{i}}\,(\boldsymbol{y} - \boldsymbol{x}_{\boldsymbol{i}}^{\prime(\boldsymbol{LLR})}\left(\boldsymbol{X}^{\prime(\boldsymbol{LLR})}\boldsymbol{W}_{\boldsymbol{i}}\boldsymbol{X}^{(\boldsymbol{LLR})}\right)^{-\mathbf{1}}\boldsymbol{X}^{\prime(\boldsymbol{LLR})}\boldsymbol{W}_{\boldsymbol{i}}\boldsymbol{y}) \tag{11}$$

Using matrix notation, the PNRM can be expressed as:

$$\hat{\boldsymbol{y}}^{(PNRM)} = \begin{bmatrix} \boldsymbol{h}_{\mathbf{1}}^{(LLR)}\boldsymbol{y} + \boldsymbol{\lambda}\boldsymbol{h}_{\mathbf{1}}^{(LLR)}(\boldsymbol{y} - (\boldsymbol{h}_{\mathbf{1}}^{(LLR)}\boldsymbol{y}) \\ \boldsymbol{h}_{\mathbf{2}}^{(LLR)}\boldsymbol{y} + \boldsymbol{\lambda}\boldsymbol{h}_{\mathbf{2}}^{(LLR)}(\boldsymbol{y} - (\boldsymbol{h}_{\mathbf{2}}^{(LLR)}\boldsymbol{y}) \\ \vdots \\ \boldsymbol{h}_{\boldsymbol{n}}^{(LLR)}\boldsymbol{y} + \boldsymbol{\lambda}\boldsymbol{h}_{\boldsymbol{n}}^{(LLR)}(\boldsymbol{y} - (\boldsymbol{h}_{\boldsymbol{n}}^{(LLR)}\boldsymbol{y}) \end{bmatrix}, \tag{12}$$

$$\hat{\boldsymbol{y}}^{(PNRM)} = \begin{bmatrix} \boldsymbol{h}_1^{(LLR)} + \boldsymbol{\lambda h}_1^{(LLR)}(\boldsymbol{I} - (\boldsymbol{h}_1^{(LLR)}) \\ \boldsymbol{h}_2^{(LLR)} + \boldsymbol{\lambda h}_2^{(LLR)}(\boldsymbol{I} - (\boldsymbol{h}_2^{(LLR)}) \\ \vdots \\ \boldsymbol{h}_n^{(LLR)} + \boldsymbol{\lambda h}_n^{(LLR)}(\boldsymbol{I} - (\boldsymbol{h}_n^{(LLR)}) \end{bmatrix} \mathbf{y}, \tag{13}$$

$$\hat{\boldsymbol{y}}^{(PNRM)} = \boldsymbol{H}^{(PNRM)}\boldsymbol{y},$$

where $\boldsymbol{I}$ is the $n \times n$ identity matrix, the $1\times n$ vector, $\boldsymbol{h}_i^{(LLR)} + \boldsymbol{\lambda h}_i^{(LLR)}(\boldsymbol{I} - (\boldsymbol{h}_i^{(LLR)}),$ is the i^{th} row of the PNRM Hat matrix, $\boldsymbol{H}^{(PNRM)}$. The i^{th} row of the model matrix $\boldsymbol{X}^{(\boldsymbol{LLR})}$ is given as:

$$\boldsymbol{x}_i'^{(\boldsymbol{LLR})} = (1 \quad x_{i1} \quad x_{i2} \quad . \quad . \quad . \quad x_{ik}) \tag{14}$$

The model matrix is given as:

$$\boldsymbol{X}^{(\boldsymbol{LLR})} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}_{(n\times(k+1))} \tag{15}$$

$\boldsymbol{W_i}$ is an $n \times n$ diagonal matrix of the kernel weights for estimating the ith response (Wan and Birch, 2011). The matrix $\mathbf{W}_i$ is given as:

$$\boldsymbol{W_i} = \begin{bmatrix} w_{1i} & 0 & \cdots & 0 \\ 0 & w_{2i} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{ni} \end{bmatrix}_{n\times n}, i = 1, \ 2, \ . \ . \ . \ . , n \tag{16}$$

The parameter $\boldsymbol{\lambda}$, is known as the mixing parameter whose optimal value λ^* of λ, may be selected based on the minimization of the form of the $PRESS^{**}$ criterion given as:

$$PRESS^{**}(\lambda) = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{i,-i}^{(.)}(\Phi,\lambda)\right)^2}{n - tr\left(H^{(.)}(\Phi,\lambda)\right) + (n-k-1)\frac{SSE_{max} - SSE_{\Phi}}{SSE_{max}}}, \tag{17}$$

where $\Phi = [b_1^*, b_2^*, \dots, b_n^*]$ is the vector of optimal bandwidths, SSE_Φ is the Sum of Squared Errors associated with the set of the optimal bandwidths, $[b_1^*, b_2^*, \dots, b_n^*]$, $tr\left(H^{(.)}(\Phi,\lambda)\right)$ is the trace of Hat matrix, and $\hat{y}_{i,-i}^{(.)}(\Phi,\lambda)$ is the leave-one-out cross-validation estimate of y_i (Mays *et al.*, 2001 and Wan and Birch, 2011).

2.2 Estimation of $\hat{y}_1^{PNRM}$

To estimate, $\hat{\boldsymbol{y}}_1^{PNRM}$, diagonal matrix of kernel weight, $\boldsymbol{W_1}$, and the target points, $x_{11}, x_{12}, \dots, x_{1k}$, are written as:

$$\hat{\boldsymbol{y}}_{\boldsymbol{1}}^{PNRM} = \boldsymbol{x}_{\boldsymbol{1}}^{\prime(\boldsymbol{LLR})}\left(\boldsymbol{X}^{\prime(\boldsymbol{LLR})}\boldsymbol{W}_{\boldsymbol{1}}\boldsymbol{X}^{(\boldsymbol{LLR})}\right)^{-\boldsymbol{1}}\boldsymbol{X}^{\prime(\boldsymbol{LLR})}\boldsymbol{W}_{\boldsymbol{1}}\boldsymbol{y} + \lambda\, \boldsymbol{x}_{\boldsymbol{1}}^{\prime(\boldsymbol{LLR})}\left(\boldsymbol{X}^{\prime(\boldsymbol{LLR})}\boldsymbol{W}_{\boldsymbol{1}}\boldsymbol{X}^{(\boldsymbol{LLR})}\right)^{-\boldsymbol{1}}\boldsymbol{X}^{\prime(\boldsymbol{LLR})}\boldsymbol{W}_{\boldsymbol{1}}\,(\boldsymbol{y} - \boldsymbol{x}_{\boldsymbol{1}}^{\prime(\boldsymbol{LLR})}\left(\boldsymbol{X}^{\prime(\boldsymbol{LLR})}\boldsymbol{W}_{\boldsymbol{1}}\boldsymbol{X}^{(\boldsymbol{LLR})}\right)^{-\boldsymbol{1}}\boldsymbol{X}^{\prime(\boldsymbol{LLR})}\boldsymbol{W}_{\boldsymbol{1}}\boldsymbol{y}) \quad (18)$$

$$\boldsymbol{W}_{\boldsymbol{1}} = \begin{bmatrix} w_{11} & 0 & \cdots & 0 \\ 0 & w_{21} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{n1} \end{bmatrix}_{(n\times n)} \quad (19)$$

The entries from equation (19) and the locally adaptive bandwidths of Edionwe *et al*., (2016) are translated as:

$$w_{11} = \frac{\prod_{j=1}^{k} K\left(\frac{x_{1j}-x_{1j}}{b_i}\right)}{\sum_{i=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{ij}-x_{1j}}{b_i}\right)}, \quad i = 1, \ 2\ , \ . \ . \ . \ , n; j = 1,\ 2, \dots, k. \quad (20)$$

$$w_{11} = \frac{e^{-(\frac{x_{11}-x_{11}}{b_1})^2} e^{-(\frac{x_{12}-x_{12}}{b_2})^2} \dots e^{-(\frac{x_{1k}-x_{1k}}{b_n})^2}}{\left[e^{-(\frac{x_{11}-x_{11}}{b_1})^2} e^{-(\frac{x_{12}-x_{12}}{b_2})^2} \dots e^{-\left(\frac{x_{1k}-x_{1k}}{b_n}\right)^2} + e^{-\left(\frac{x_{21}-x_{11}}{b_1}\right)^2} e^{-\left(\frac{x_{22}-x_{12}}{b_2}\right)^2} \dots e^{-\left(\frac{x_{2k}-x_{1k}}{b_n}\right)^2} + \dots + e^{-\left(\frac{x_{n1}-x_{11}}{b_1}\right)^2} e^{-\left(\frac{x_{n2}-x_{12}}{b_2}\right)^2} \dots e^{-(\frac{x_{nk}-x_{1k}}{b_n})^2}\right]}$$

$$w_{21} = \frac{\prod_{j=1}^{k} K\left(\frac{x_{2j}-x_{1j}}{b_i}\right)}{\sum_{i=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{ij}-x_{1j}}{b_i}\right)}, \quad i = 1, \ 2\ , \ . \ . \ . \ , n; j = 1,\ 2, \dots, k. \quad (21)$$

$$w_{21} = \frac{e^{-(\frac{x_{21}-x_{11}}{b_1})^2} e^{-(\frac{x_{22}-x_{12}}{b_2})^2} \dots e^{-(\frac{x_{2k}-x_{1k}}{b_n})^2}}{\left[e^{-(\frac{x_{11}-x_{11}}{b_1})^2} e^{-(\frac{x_{12}-x_{12}}{b_2})^2} \dots e^{-\left(\frac{x_{1k}-x_{1k}}{b_n}\right)^2} + e^{-\left(\frac{x_{21}-x_{11}}{b_1}\right)^2} e^{-\left(\frac{x_{22}-x_{12}}{b_2}\right)^2} \dots e^{-\left(\frac{x_{2k}-x_{1k}}{b_n}\right)^2} + \dots + e^{-\left(\frac{x_{n1}-x_{11}}{b_1}\right)^2} e^{-\left(\frac{x_{n2}-x_{12}}{b_2}\right)^2} \dots e^{-(\frac{x_{nk}-x_{1k}}{b_n})^2}\right]}$$

$$\vdots$$

$$w_{n1} = \frac{\prod_{j=1}^{k} K\left(\frac{x_{nj}-x_{1j}}{b_i}\right)}{\sum_{i=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{ij}-x_{1j}}{b_i}\right)}, \quad i = 1, \ 2\ , \ . \ . \ . \ , n; j = 1,\ 2, \dots, k. \quad (22)$$

$$w_{n1} = \frac{e^{-\left(\frac{x_{n1}-x_{11}}{b_1}\right)^2} e^{-\left(\frac{x_{n2}-x_{12}}{b_2}\right)^2} \dots e^{-\left(\frac{x_{nk}-x_{1k}}{b_n}\right)^2}}{\left[e^{-(\frac{x_{11}-x_{11}}{b_1})^2} e^{-(\frac{x_{12}-x_{12}}{b_2})^2} \dots e^{-\left(\frac{x_{1k}-x_{1k}}{b_n}\right)^2} + e^{-\left(\frac{x_{21}-x_{11}}{b_1}\right)^2} e^{-\left(\frac{x_{22}-x_{12}}{b_2}\right)^2} \dots e^{-\left(\frac{x_{2k}-x_{1k}}{b_n}\right)^2} + \dots + e^{-\left(\frac{x_{n1}-x_{11}}{b_1}\right)^2} e^{-\left(\frac{x_{n2}-x_{12}}{b_2}\right)^2} \dots e^{-(\frac{x_{nk}-x_{1k}}{b_n})^2}\right]}$$

To estimate, $\hat{\boldsymbol{y}}_{\boldsymbol{2}}^{\boldsymbol{PNRM}}, \hat{\boldsymbol{y}}_{\boldsymbol{3}}^{\boldsymbol{PNRM}}, \dots, \hat{\boldsymbol{y}}_{\boldsymbol{n}}^{\boldsymbol{PNRM}}$ with respective diagonal matrices of kernel weights,

$\boldsymbol{W}_{\boldsymbol{2}}, \boldsymbol{W}_{\boldsymbol{3}}, \dots, \boldsymbol{W}_{\boldsymbol{n}}$ follows pattern from equations 20, 21 and 22.

3. Application

A multiple response problem is used in order to compare the statistical performance of the proposed model with the existing OLS and LLR models.

3.1 The Multiple Response Chemical Process Data

This problem was analyzed by He *et al*., (2009) and He *et al*., (2012). The aim of the study was to get the setting of the explanatory variables x_1 and x_2 (representing reaction time and temperature, respectively) that would simultaneously optimize three quality measures of a chemical solution y_1, y_2 and y_3 (representing yield, viscosity, and molecular weight, respectively). The process requirements for each response are as follows:

Maximize y_1 with lower limit $L = 78.5$, and target value $\emptyset = 80$;

y_2 should take a value in the range $L = 62$ and $U = 68$ with $\emptyset = 65$;

Minimize y_3 with upper limit $U = 3300$ and target value $\emptyset = 3100$.

As it is with the procedure when nonparametric regression is involved, the real values of the explanatory variables are coded to lie between 0 and 1. The data collected via a Central Composite Design is presented in Table 1. A full second-order polynomial model was found adequate for fitting each response using the OLS model (Wan and Birch, 2011).

3.2 Desirability Function

Desirability function is applied in Multi-Response Optimization (MRO) where responses are classified as larger the better (LTB) for maximizing the response, smaller the better (STB) for minimizing the response, and nominal the better (NTB) is a two sided transformation of the response (Pickle, 2006 and He *et al.*, 2009, 2012).

For larger-the-better (LTB) response, $d_1(\hat{y}_1(\boldsymbol{x}))$ is given as:

$$d_1(\hat{y}_1(\boldsymbol{x})) = \begin{cases} 0, & \hat{y}_1(\boldsymbol{x}) < 78.5 \\ \left\{\frac{\hat{y}_1(\boldsymbol{x})-78.5}{80-78.5}\right\}^{t_1}, & 78.5 \le \hat{y}_1(\boldsymbol{x}) \le 80, \\ 1, & \hat{y}_1(\boldsymbol{x}) > 80, \end{cases}$$

$$s.t\ \boldsymbol{x} \in [0,1], \tag{23}$$

where the desirability function $d_1(\hat{y}_1(\boldsymbol{x})) = d_1$ is a scalar measure, $T = 80$ and $L = 78.5$ are the maximum acceptable value and lower limit, respectively, t_1 is taken to be 1. The objective is to maximize the $\hat{y}_1(\boldsymbol{x})$ response.

For the nominal-the-better (NTB) response, $d_2(\hat{y}_2(\boldsymbol{x}))$ is a two sided transformation given as:

$$d_2(\hat{y}_2(\boldsymbol{x})) = \begin{cases} \left\{\frac{\hat{y}_2(\boldsymbol{x})-62}{65-62}\right\}^{t_1}, & 62 \le \hat{y}_2(\boldsymbol{x}) < 65, \\ \left\{\frac{68-\hat{y}_2(\boldsymbol{x})}{68-65}\right\}^{t_2}, & 65 \le \hat{y}_2(\boldsymbol{x}) \le 68, \\ 0, & otherwise \end{cases}$$

$$s.t\ \boldsymbol{x} \in [0,1], \tag{24}$$

where $d_2(\hat{y}_2(\boldsymbol{x})) = d_2, L = 62, U = 68,\ \rho = 65$ is the target value of the $\hat{y}_2(\boldsymbol{x})$ response. However, for RSM data, the parameters values of $t_1\ and\ t_2$ are taken to be 1 (Castillo, 2007, Wan, 2007 and He *et al.*, 2012).

When the response is of the smaller-the-better (STB) type, $d_3(\hat{y}_3(\boldsymbol{x}))$ is given as:

$$d_3(\hat{y}_3(\boldsymbol{x})) = \begin{cases} 1, & \hat{y}_3(\boldsymbol{x}) < 3100, \\ \left\{\dfrac{3300 - \hat{y}_3(\boldsymbol{x})}{3300 - 3100}\right\}^{t_2}, & 3100 \le \hat{y}_3(\boldsymbol{x}) \le 3300, \\ 0, & \hat{y}_3(\boldsymbol{x}) > 3300, \end{cases}$$

$s.t\ \boldsymbol{x} \epsilon\ [0, 1]$, (25)

where $d_3(\hat{y}_3(\boldsymbol{x})) = d_3, T = 3100$ and $U = 3300$ are the minimum acceptable value and upper limit, respectively. The objective is to minimize the $\hat{y}_3(\boldsymbol{x})$ response.

3.3 The Overall Desirability

The objective of desirability function is to assigns values in the interval [0, 1] to each estimated response $\hat{y}_m(\boldsymbol{x})$ so as to maximize the overall desirability, D with respect to $\boldsymbol{x}$ into a single scalar value, which is the geometric mean of the individual desirability functions. Overall desirability $D(\boldsymbol{x})$, is:

$$D(\boldsymbol{x}) = \sqrt[3]{(d_1(\hat{y}_1(\boldsymbol{x})) \times d_2(\hat{y}_2(\boldsymbol{x})) \times d_3(\hat{y}_3(\boldsymbol{x})))} \quad (26)$$

(Ramakrishnan and Arumugam, 2012).

Table 1: 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

The optimal values of the parameters of both the proposed model and the LLR for each response are shown in Table 2.

Table 2: Optimal values of the tuning parameters and mixing parameter of the proposed model and the LLR model for the multiple response chemical process data

	Proposed Model					LLR	
Response	$N^*(W^{Raw})$	$C^*(W^{Raw})$	$N^*(W^{Resid})$	$C^*(W^{Resid})$	λ^*	N^*	C^*
y_1	3.6241	1.2876	3.0413	0.0798	0.9457	3.0971	0.0797
y_2	6.5583	0.1246	1.2854	0.0952	1.0000	1.2297	0.0952
y_3	1.9999	0.0664	1.2050	0.0935	1.0000	4.8181	0.0896

The locally adaptive optimal bandwidths of each regression model are presented in Table 3 and Table 4, respectively.

Table 3: Locally Adaptive Optimal Bandwidths for the Proposed Model

	Optimal Bandwidths for W^{Raw}			Optimal Bandwidths for W^{Resid}		
i	y_1	y_2	y_3	y_1	y_2	y_3
1	0.2792	0.5475	0.0057	0.3972	0.1209	0.1501
2	0.2789	0.4978	0.2513	0.2733	0.0955	0.0549
3	0.2791	0.5724	0.1816	0.3559	0.1336	0.0819
4	0.2785	0.5848	0.3210	0.1495	0.1400	0.0279
5	0.2794	0.4356	0.0322	0.4715	0.0637	0.1398
6	0.2788	0.4729	0.1451	0.2403	0.0827	0.0961
7	0.2791	0.6097	0.0754	0.3559	0.1527	0.1231
8	0.2788	0.5973	0.2347	0.2320	0.1463	0.0613
9	0.2785	0.4232	0.1849	0.1164	0.0573	0.0806
10	0.2784	0.4605	0.0920	0.0834	0.0764	0.1166
11	0.2784	0.4729	0.1617	0.1082	0.0827	0.0896
12	0.2785	0.4481	0.1218	0.1329	0.0700	0.1051
13	0.2785	0.4356	0.1916	0.1247	0.0637	0.0781

Table 4: Locally Adaptive Optimal Bandwidths for LLR Model

i	y_1	y_2	y_3
1	0.4045	0.1156	0.6669
2	0.2783	0.0913	0.1755
3	0.3624	0.1278	0.3149
4	0.1522	0.1339	0.0360
5	0.4802	0.0609	0.6138
6	0.2447	0.0792	0.3880
7	0.3624	0.1461	0.5275

8	0.2363	0.1400	0.2087
9	0.1186	0.0548	0.3083
10	0.0849	0.0731	0.4943
11	0.1102	0.0792	0.3548
12	0.1354	0.0670	0.4345
13	0.1270	0.0609	0.2950

The performance statistics (goodness-of-fit) of the models for the chemical process data are shown in Table 5.

Table 5: Goodness-of-fit of the models for the Chemical Process Data

Response	Model	DF	$PRESS^{**}$	$PRESS$	SSE	MSE	R^2	R^2_{Adj}
	OLS	7.0000	0.3361	**2.3525**	0.4962	0.0709	0.9827	0.9704
y_1	LLR	4.7810	0.2063	3.0148	0.3113	0.0651	0.9892	0.9728
	Proposed	4.7093	**0.2046**	2.9781	**0.2909**	**0.0618**	**0.9899**	**0.9742**
	OLS	7.0000	28.8726	202.1082	36.2242	5.1749	0.8997	0.8281
y_2	LLR	4.0000	9.4343	129.4141	**10.0000**	**2.5000**	**0.9723**	**0.9170**
	Proposed	4.0000	**9.0889**	**124.6763**	**10.0000**	**2.5000**	**0.9723**	**0.9170**
	OLS	7.0000	159080	1113600	207870	29696	0.7590	0.5868
y_3	LLR	5.8380	**40779**	**508170**	92621	**15865**	0.8926	**0.7795**
	Proposed	4.0000	44326	514370	**65720**	16430	**0.9238**	0.7714

The results as shown in Table 5 obviously explain the goodness-of-fit statistics for multiple response chemical process data. Clearly, the proposed model is superior in terms of minimum values for the $PRESS**$ within two responses (y_1 and y_2), minimum SSE across two responses (y_1 and y_3) and a tie with LLR in y_2 and minimum $\boldsymbol{R^2}$ across the two responses (y_1 and y_3) and a tie with LLR iny_2. In general, it implies that the proposed model produces a more practical and trustworthy results in eight cells, and with a joint performance in other four cells, which obviously promise for a better model.

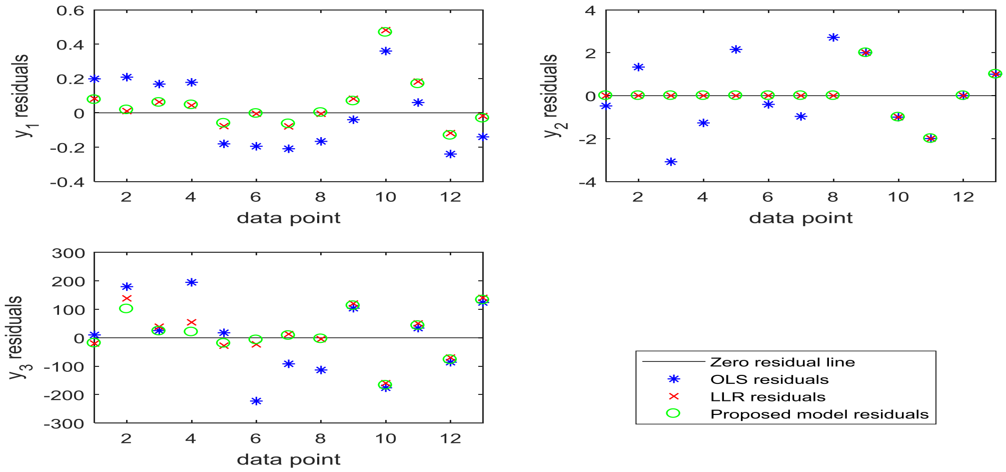

Figure 1: Plots of Model residuals for the multiple response chemical process data

There is relationship between Table 5 and Figure 1 in terms of the three residual plots for all the data points. The plot for y_1 residual has a somewhat better explained variation for the proposed model over LLR, but clearly outperforms the OLS which is also confirmed in Table 6. The data points for y_2 residual coincide between the proposed model and the LLR but differ with the OLS by way of enhanced explained variation. The data point for the proposed model in y_3 residual has higher explained variation over LLR and OLS. Apparently, these observations specify that the proposed model offers more precise fits over LLR and OLS.

Table 6: Model optimal solution based on the Desirability function for the multiple response chemical process data

Model	x_1	x_2	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	d_1	d_2	d_3	$D(\%)$
OLS	0.4449	0.2226	78.7616	66.4827	3229.9	0.1744	0.5058	0.3504	31.5800
LLR	0.5155	0.3467	78.6965	65.0328	3285.9	0.1310	0.9891	0.0703	20.8837
Proposed Model	0.4845	0.3641	78.8072	65.7368	3251.2	0.2048	0.7544	0.2441	**33.5343**

The overall goal of the desirability function as given in Table 6 is to determine setting of the explanatory variables that would simultaneously optimize the responses. The operating settings for the proposed model optimize the responses with a greater desirability as compared with LLR and OLS. Consequently, it is established that the contribution from the proposed model satisfies the production requirements over LLR and OLS.

3.4 Discussion of Results

The results as shown in Table 6, shows that PNRM, either completely or conjointly, provides the best results in terms of all the statistics for y_1 and y_2. For the y_3, the PNRM provides the best results in two out of the six statistics for comparison. Remarkably, PNRM gives the best $PRESS^{**}$ in y_1 and y_2. Figure 1 is a reflection of the results offered in Table 6, where the interest is to give a pictorial display of the measure of variability not explained in the data by OLS, LLR and PNRM models for a multiple response problem. The obvious from Figure 1, is that PNRM has less variability compared with OLS and LLR models.

Lastly, Table 6, gives a clear explanation of the production or process requirements for each response, such that $\hat{y}_1$ must not be less 78.5, otherwise, the desirability, $d_1(\hat{y}_1(\boldsymbol{x}))$, becomes zero. The $\hat{y}_2$ must lie between the values 62 and 68 inclusive, else the desirability, $d_2(\hat{y}_2(\boldsymbol{x}))$, takes the value zero. Also, $\hat{y}_3$ must not exceed 3300, otherwise the desirability, $d_3(\hat{y}_3(\boldsymbol{x}))$, is assign a zero. The model with the highest overall desirability, D (%), has the optimal settings of the explanatory variable that will optimize the responses. Hence, PNRM provides the best settings that optimize the response for the multiple response chemical process data as compared with OLS and LLR models.

4. Conclusion

In this paper, we considered two existing regression models, the OLS and LLR and proposed a nonparametric regression model (PNRM) that uses the locally adaptive bandwidths of Edionwe *et al*. (2016) for adequate smoothing RSM data.

The goodness-of-fit statistics obtained from an empirical data and optimal solutions show that the PNRM regression model performs better than OLS and the LLR that uses the locally adaptive bandwidths of Edionwe *et al*. (2016).

Consequently, worthy to refer is the remarkable low values of the $PRESS^{**}$ criterion and SSE of the PNRM. This promises high accuracy in predicting yield, viscosity, and molecular weight, for multiple response problems.

Conclusively, the PNRM in Table 6, display higher level of desirability over OLS and LLR models and as such provided a setting for the explanatory variables that optimized the response for multiple response chemical process data.

References

Aydin, D., Memmedli, M. and Omay, R. E. (2013). Smoothing parameter selection for nonparametric regression using smoothing spline. *European Journal of Pure and Applied Mathematics*, 6(2), 222 - 238.

Castillo, D. E. (2007). *Process Optimization: A Statistical Method*, Springer International Series in Operations Research and Management Science, New York.

Edionwe, E., Mbegbu, J. I. and Chinwe, R. (2016). A New Function for Generating Local Bandwidths for Semi-parametric MRR2 Model in Response Surface Methodology. *Journal of Quality Technology*, 48(4), pp. 388 – 404.

Fathi, H. K., Moghadam, M. B., Ahmadi, N., Fayaz, M. and Navaee, M. (2011). Application of Nonparametric Optimization Methods for Dyeing of Wool. *Middle-East Journal of Scientific Research*, **9**(2), pp. 270 - 278.

Hastie, T., Tibshirani, R. and Friedman, T. A. (2009). *The Elements of Statistical Learning, Data Mining, Inference and Prediction*. Springer Series in Statistics, Second Edition, Springer, New York.

He, Z., Wang, J., Oh, J. and Park, S. H. (2009). Robust optimization for multiple responses using response surface methodology. *Applied Stochastic Models in Business and Industry*, 26**,** 157 – 171.

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.

Kai, B., Li, R. and Zou, H. (2010). Local Composite Quantile Regression Smoothing: an efficient and safe alternative to the local polynomial regression. *Journal of the Royal Statistical Society*, 72, 49 – 69.

Mays, J., Birch, J. B. and Starnes, B. A., (2001). Model robust regression: combining parametric, nonparametric and semiparametric methods. *Journals of Nonparametric Statistics*, 13, 245 - 277.

Pickle, S. M. (2006). Semiparametric Techniques for Response Surface Methodology. Unpublished Ph.D. Dissertation, Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA.

Prewitt, K. and Lohr, S. (2006). Bandwidth selection in local polynomial regression using eigenvalues. *Journal of Royal Statistical Society, Series B*, 68(1), 135 – 154.

Ramakrishnan, R. and Arumugam, R. (2012). Application to Response Surface Methodology (RSM) for Optimization of Operating Parameters and Performance Evaluation of Cooling Tower Cold Water Temperature. *An International Journal of Optimization and Control, Theories and Applications*, **1**, 39 - 50.

Rivers, D. L. (2009). A Graphical Analysis of Simultaneously Choosing the Bandwidth and Mixing Parameter for Semiparametric Regression Techniques. Unpublished M.Sc. Dissertation, Virginia Commonwealth University, Richmond, VA 23284, USA.

Swamy, P. A. V. B., Tavlas, G. S., Hall, S. T. and Hondroyiannis, G. (2008). Estimation of Parameters in the Presence of Model Misspecification and Measurement Error, University of Leicester, No. 8-27.

Wan, W. (2007). Semi-parametric techniques for multi-response optimization. Unpublished Ph.D. Dissertation submitted to the Department of Statistics, Virginia Polytechnic Institute, and State University, Blacksburg, Virginia, USA.

Wan, W. and Birch, J. B. (2011). A semi-parametric technique for multi-response optimization. *Quality and Reliability Engineering International*, 27(1), 47 - 59.