BENIN JOURNAL OF STATISTICS
ISSN 2682-5767
Vol. 5, pp. 137– 158 (2022)

An Adaptive Averaging Regression Model with Application to Response Surface Methodology

O. Eguasa[1 *], **E. Edionwe**[2] **and J. I. Mbegbu**[3]

[1]*Department of Physical Sciences, Benson Idahosa University, Benin City, Edo State, Nigeria*

[2]*Department of Mathematics, Ambrose Alli University, Ekpoma, Edo State, Nigeria*

[3]*Department of Statistics, University of Benin, Benin City, Edo State, Nigeria.*

(*Received: 18 April 2022; Accepted: 19 May 2022*)

Abstract. Response Surface Methodology (RSM) is a sequential statistical technique with the goal to find settings of the explanatory variables that would optimize the response. In literature, the nonparametric regression models are affected by the idiosyncrasies of RSM data, such as dimensionality problem, sparseness of the data and small sample size. In this paper, we proposed an adaptive averaging regression model that combines local linear regression (LLR) and the kernel regression models via convex combination, which utilized the locally adaptive bandwidths from literature. The proposed averaging regression model applied to RSM data showed improved goodness-of-fit statistics and process requirements over Ordinary Least Squares (OLS), LLR with fixed bandwidths and LLR that uses existing bandwidths in a variety of data considered. Furthermore, simulation study was carried out on the multi-response data and the results show that the proposed adaptive averaging regression model that employed the locally adaptive bandwidths gives the smallest Average Sum of Squares Error.

Keywords: Adaptive averaging regression model, convex combination, locally adaptive bandwidths, Response surface methodology, Curse of dimensionality.

Published by: Department of Statistics, University of Benin, Nigeria

1. Introduction

Nair *et al.* (2014) and Yeniay, (2014) defined RSM as statistical procedure applied by engineers and industrial statistician for empirical model building, with the aim of optimizing the response variables which are influenced by several explanatory variables.

*Corresponding author. Email: oeguasa@biu.edu.ng

RSM is appropriate for optimizing the response variable y as a function of several explanatory variables $(x_{i1}, x_{i2}, \cdots, x_{ik})$ which can be modeled as:

$$y_i = f(x_{i1}, x_{i2}, \cdots, x_{ik}) + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{1}$$

where ε_i is the error term and assumed to have a normal distribution with mean zero and variance σ^2. The surface represented by $f(x_{i1}, x_{i2}, \cdots, x_{ik})$ is termed a response surface (Wan and Birch, 2011).
The true response function f is usually unknown which must be estimated. Applying RSM, we seek to identify the functional relationship between the responses y and associated explanatory variables $(x_{i1}, x_{i2}, \cdots, x_{ik})$.
The general parametric regression model in matrix notation can be written as:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \tag{2}$$

where $\mathbf{y}$ is a vector of response, $\mathbf{X} = \mathbf{X}^{(OLS)}$ is the OLS model matrix, β is the unknown parameter vector and ε is the vector of error term assumed to be normally distributed with zero mean and constant variance property.
The common approach for estimating the parameter vector in (2) is usually based on the Method of OLS. The parameter vector estimates $\hat{\beta}$ in (2) is given as:

$$\hat{\beta}^{(OLS)} = \left(\mathbf{X}'^{(OLS)}\mathbf{X}^{(OLS)}\right)^{-1}\mathbf{X}'^{(OLS)}\mathbf{y}, \quad \mathbf{X} = \mathbf{X}^{(OLS)} \tag{3}$$

The estimated responses for the i^{th} location can be written as :

$$\hat{\mathbf{y}}_i^{(OLS)} = \mathbf{x_i}'^{(OLS)}\hat{\beta}^{(OLS)} = \mathbf{x_i}'^{(OLS)}\left(\mathbf{X}'^{(OLS)}\mathbf{X}^{(OLS)}\right)^{-1}\mathbf{X}'^{(OLS)}\mathbf{y}, \quad i = 1, 2, \cdots, n \tag{4}$$

where $\mathbf{x_i}'^{(OLS)}$is the i^{th} row of matrix $\mathbf{X}^{(OLS)}, n \times (k+1)$ vector.

$\mathbf{H_i}'^{(OLS)} = \mathbf{x_i}'^{(OLS)}\left(\mathbf{X}'^{(OLS)}\mathbf{X}^{(OLS)}\right)^{-1}\mathbf{X}'^{(OLS)}$ is the i^{th} row of the OLS "HAT" matrix of dimension $n \times n$, $\mathbf{H}^{(OLS)}$. The estimated response in the i^{th} location is given by:

$$\hat{\mathbf{y}}^{(OLS)} = \mathbf{H}^{(OLS)}\mathbf{y} \tag{5}$$

where the matrix $\mathbf{H}^{(OLS)}$ is given as:

$$\mathbf{H}^{(OLS)} = \begin{bmatrix} \mathbf{H}_1^{(OLS)} \\ \mathbf{H}_2^{(OLS)} \\ \vdots \\ \mathbf{H}_n^{(OLS)} \end{bmatrix}, \tag{6}$$

(Carley, *et al*., (2004); River, (2009))

The LLR model is a weighted form of the least squares derived from Local Polynomial Regression of order one ($d = 1$) which is an improvement over the kernel regression model because it can adjust to bias both at the boundaries and unequal spacing of the explanatory variables (Ruppert and Wand, 1994; Walker *et al.*, 2002).
The LLR model is derived from standard least squares theory. The LLR estimator $\hat{\mathbf{y}}_i^{(LLR)}$ of $\mathbf{y}_i$ is given as:

$$\hat{\mathbf{y}}_i^{(LLR)}=\mathbf{x_i}^{'(LLR)}(\mathbf{X}^{'(LLR)}\mathbf{W_i}\mathbf{X}^{(LLR)})^{-\mathbf{1}}\mathbf{X}^{'(LLR)}\mathbf{W_i}\mathbf{y} =\mathbf{H_i}^{(LLR)}\mathbf{y}, \tag{7}$$

where $\mathbf{y} = (y_1, \ldots y_n)^{'}$, $\mathbf{x_i}^{'(LLR)}= (1x_{i1} \ldots x_{ik})$ is the i^{th} row of the LLR model matrix, $\mathbf{X}^{(LLR)}$ given as:

$$\mathbf{X}^{(LLR)} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \tag{8}$$

We define ***W***, an $n \times n$ diagonal matrix of kernel weights for estimating the response as

$$\mathbf{W} = \mathbf{c_i}\delta_{\mathbf{ip}}, \quad i = 1, 2, ..., n; p = 1, 2, ..., n$$

where c_i are kernels weight at *ith* location and δ_{ip} is the Kronecker delta function given as

$$\delta_{ip} = \begin{cases} 1, if i = p \\ \\ 0, otherwise \end{cases} \qquad i = 1, 2, ..., n; p = 1, 2, ..., n \tag{9}$$

Thus,

$$\mathbf{W} = \begin{bmatrix} c_1\delta_{11} & c_1\delta_{12} & \cdots & c_1\delta_{1n} \\ c_2\delta_{21} & c_2\delta_{22} & \cdots & c_2\delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_n\delta_{n1} & c_n\delta_{n2} & \cdots & c_n\delta_{nn} \end{bmatrix} \tag{10}$$

$$\mathbf{W} = \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_n \end{bmatrix}$$

where $c_1 = w_{i1}$, $c_2 = w_{i2}$, . . ., $c_n = w_{in}$. In terms of location, $\mathbf{W} =\mathbf{W_i}$

$$\mathbf{W_i} = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix}, i = 1, 2, ..., n. \tag{11}$$

(Wan and Birch, 2011; Eguasa, 2019).
For a single explanatory variable problem used in the diagonal weight matrix $\mathbf{W_i}$, the kernel function $K\left(\frac{x_{ij}-x_{1j}}{b_{ij}}\right)$ is a simplified Gaussian kernel when one explanatory variable problem is used in the model matrix ***X*,** given as:

$$w_{i1} = K\left(\frac{x_{ij} - x_{1j}}{b_{ij}}\right) = e^{-\left(\frac{x_{ij}-x_{1j}}{b_{ij}}\right)^2} \tag{12}$$

In a situation where more than one explanatory variable are used in the model matrix ***X*,** the kernel weight w_{i1}, is a product kernel given as:

$$\frac{w_{i1} = \prod_{j=1}^{k} K\left(\frac{x_{ij}-x_{1j}}{b_{ij}}\right)}{\sum_{p=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{pj}-x_{1j}}{b_{pj}}\right)}, p = 1, 2, \ldots, n, j = 1, 2, \ldots, k, \tag{13}$$

Wan and Birch (2011); Eguasa *et al.* (2022).
In RSM, the matrix comprising the vector of optimal bandwidths $b_{11}^*, b_{12}, \ldots, b_{nk}^*$ is obtained from the minimization of the Penalized Prediction Error Sum of Squares ($PRESS^{**}$):

$$Minimize PRESS^{**}\{b_{11}, b_{12}, \ldots, b_{nk}\} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{i,-i}^{(LLR)}\right)^2}{n - trace\left(H^{(LLR)}(\omega)\right) + (n-k-1)\frac{SSE_{\max}-SSE_{\omega}}{SSE_{\max}}} \tag{14}$$

For $i = 1$ in (11), we have:

$$\mathbf{W_1} = \begin{bmatrix} w_{11} & 0 & \cdots & 0 \\ 0 & w_{12} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{1n} \end{bmatrix}_{(n \times n)} \tag{15}$$

$$w_{11} = \frac{\prod_{j=1}^{k} K\left(\frac{x_{1j}-x_{1j}}{b_{ij}}\right)}{\sum_{p=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{pj}-x_{1j}}{b_{pj}}\right)}, p = 1, 2, ..., n; j = 1, 2, \ldots, k. \tag{16}$$

$w_{11} = \frac{S}{[S+T+...+U]}, S = e^{-(\frac{x_{11}-x_{11}}{b_{11}})^2} e^{-(\frac{x_{12}-x_{12}}{b_{12}})^2} \dots e^{-(\frac{x_{1k}-x_{1k}}{b_{1k}})^2},$

$T = e^{-\left(\frac{x_{21}-x_{11}}{b_{21}}\right)^2} e^{-\left(\frac{x_{22}-x_{12}}{b_{22}}\right)^2} \dots e^{-\left(\frac{x_{2k}-x_{1k}}{b_{2k}}\right)^2}$

and $U = e^{-\left(\frac{x_{n1}-x_{11}}{b_{n1}}\right)^2} e^{-\left(\frac{x_{n2}-x_{12}}{b_{n2}}\right)^2} \dots e^{-(\frac{x_{nk}-x_{1k}}{b_{nk}})^2}$

$$w_{12} = \frac{\prod_{j=1}^{k} K\left(\frac{x_{2j}-x_{1j}}{b_{ij}}\right)}{\sum_{p=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{pj}-x_{1j}}{b_{pj}}\right)}, p = 1, 2, ..., n; j = 1, 2, \dots, k. \tag{17}$$

$w_{12} = \frac{V}{[W+V+...+Z]}, V = e^{-(\frac{x_{21}-x_{11}}{b_{21}})^2} e^{-(\frac{x_{22}-x_{12}}{b_{22}})^2} \dots e^{-(\frac{x_{2k}-x_{1k}}{b_{2k}})^2},$

$W = e^{-(\frac{x_{11}-x_{11}}{b_{11}})^2} e^{-(\frac{x_{12}-x_{12}}{b_{12}})^2} \dots e^{-\left(\frac{x_{1k}-x_{1k}}{b_{1k}}\right)^2}$

and $Z = e^{-\left(\frac{x_{n1}-x_{11}}{b_{n1}}\right)^2} e^{-\left(\frac{x_{n2}-x_{12}}{b_{n2}}\right)^2} \dots e^{-(\frac{x_{nk}-x_{1k}}{b_{nk}})^2}$

$$w_{1n} = \frac{\prod_{j=1}^{k} K\left(\frac{x_{nj}-x_{1j}}{b_{ij}}\right)}{\sum_{p=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{pj}-x_{1j}}{b_{pj}}\right)}, p = 1, 2, ..., n; j = 1, 2, \dots, k. \tag{18}$$

$$w_{1n} = \frac{R}{[M + H + \dots + R]} \tag{19}$$

$R = e^{-\left(\frac{x_{n1}-x_{11}}{b_{n1}}\right)^2} e^{-\left(\frac{x_{n2}-x_{12}}{b_{n2}}\right)^2} \dots e^{-\left(\frac{x_{nk}-x_{1k}}{b_{nk}}\right)^2},$

$H = e^{-\left(\frac{x_{21}-x_{11}}{b_{21}}\right)^2} e^{-\left(\frac{x_{22}-x_{12}}{b_{22}}\right)^2} \dots e^{-\left(\frac{x_{2k}-x_{1k}}{b_{2k}}\right)^2}$

and $M = e^{-(\frac{x_{11}-x_{11}}{b_{11}})^2} e^{-(\frac{x_{12}-x_{12}}{b_{12}})^2} \dots e^{-\left(\frac{x_{1k}-x_{1k}}{b_{1k}}\right)^2}$

(11) translates to $\mathbf{W_i}=dia(w_{i1}, w_{i2},\dots, w_{in})$ for each $i = 1, 2, ..., n$.

$$\hat{\mathbf{y}}^{(LLR)}=\mathbf{H}^{(LLR)}\mathbf{y}, \tag{20}$$

(7) can be written in terms of hat matrix as:

$$\mathbf{H}^{(LLR)} = \begin{bmatrix} \mathbf{x_1}^{'(LLR)} \left(\mathbf{X}^{'(LLR)}\mathbf{W_1}\mathbf{X}^{(LLR)}\right)^{-1} \mathbf{X}^{'(LLR)}\mathbf{W_1} \\ \mathbf{x_2}^{'(LLR)} \left(\mathbf{X}^{'(LLR)}\mathbf{W_2}\mathbf{X}^{(LLR)}\right)^{-1} \mathbf{X}^{'(LLR)}\mathbf{W_2} \\ \vdots \\ \mathbf{x_n}^{'(LLR)} \left(\mathbf{X}^{'(LLR)}\mathbf{W_n}\mathbf{X}^{(LLR)}\right)^{-1} \mathbf{X}^{'(LLR)}\mathbf{W_n} \end{bmatrix} \tag{21}$$

where the $n \times n$ matrix, $\mathbf{H}^{(LLR)}$ is the LLR hat matrix.
The drawback of LLR model is high bias in regions where there is curvature because the structure of model matrix of the LLR model lacks the quadratic terms (Hastie, *et al.*, 2009; Rivers, 2009).

1.1 *The Kernel Regression Model*

The kernel regression was proposed by Nadaraya (1964), and the mathematical expression for the kernel regression model is given as:

$$\hat{\mathbf{y}}_i^{(KER)} = \frac{\sum_{j=1}^{k} K(\frac{x_{ij-} x_o}{b})}{\sum_{l=1}^{n} K(\frac{x_{l-} x_o}{b})}\mathbf{y}; \quad i = 1, 2, \ldots, n \tag{22}$$

where x_{ij} is the design points, x_0 is the target points, b = fixed bandwidth and $K(.)$ is kernel function.

1.2 *Locally Adaptive Bandwidths*

The bandwidth is the most important parameter in nonparametric regression estimation because of its smoothing properties (Kohler *et al.*, (2014)).

Eguasa *et al.* (2022) presented data-driven locally adaptive bandwidths:

$$b_{ij} = T_{1j}(\frac{1}{2} - \frac{x_{ij}}{T_{2j}})^2, i = 1, 2, \ldots, n; j = 1, 2, \ldots, k \tag{23}$$

where, $0 < b_{ij} \leq 1, , T_{1j} > 0, T_{2j} > 0.$

The $b*_{ij}$ of the locally adaptive optimal bandwidths from (23) is obtained at an optimally selected values of T_{1j}, T_{2j}, (hereafter referred to as T^*_{1j} andT^*_{2j}, respectively), $j = 1, 2, \ldots, k$, based on the minimization of the $PRESS^{**}$ criterion in (14).

Where $b_{ij} = b$, is called a fixed bandwidth, otherwise b_{ij}, $i = 1, 2, \ldots, n; j = 1, 2, \ldots, k$ are locally adaptive bandwidths.

2. Materials and Method

The philosophy behind the averaging regression model stem from the features of the component models such that the kernel regression which is utilized as a result of its simplicity in terms of application of the model, whereas the local linear regression is attractive because it can adapt favorably in addressing boundary bias problem inherent in the kernel regression (Fan, 1993; Choi and Hall, 1998).

2.1 *Adaptive averaging regression model*

The averaging estimator proposed by Liu (2011) is an affine combination applied to density estimation. In this paper, we propose an adaptive averaging regression (AVGR) model combining the local linear regression and the kernel regression via the mixing parameter λ, with an adaptive bandwidths (smoothing parameter) b_{ij} as given in Eguasa *et al.* (2022). This is achieved through the combination of models in a convex manner as:

$$\hat{\mathbf{y}}_i^{(AVGR)} = \lambda \hat{\mathbf{y}}_i^{(LLR)} + (1-\lambda)\,\hat{\mathbf{y}}_i^{(KER)},\;\; 0 \leq \lambda \leq 1, i = 1, 2, \ldots, n, j = 1, 2, \ldots, k \tag{24}$$

$$\hat{\mathbf{y}}_i^{(AVGR)} = \lambda \mathbf{x}_{\mathbf{i}}^{'(LLR)} (\mathbf{X}^{'(LLR)} \mathbf{W_i} \mathbf{X}^{(LLR)})^{-\mathbf{1}} \mathbf{X}^{'(LLR)} \mathbf{W_i} \mathbf{y} + (1-\lambda)\, \frac{\sum_{j=1}^{k} K(\frac{x_{ij}-x_o}{b_{ij}})}{\sum_{l=1}^{n} K(\frac{x_l - x_o}{b_{ij}})} \mathbf{y} \tag{25}$$

$$\begin{aligned}\hat{\mathbf{y}}_i^{(AVGR)} = {} & \lambda \begin{bmatrix} \mathbf{x}_{\mathbf{1}}^{'} \left(\mathbf{X}^{'}\mathbf{W_1}\mathbf{X}\right)^{-1} \mathbf{X}^{'}\mathbf{W_1} \\ \mathbf{x}_{\mathbf{2}}^{'} \left(\mathbf{X}^{'}\mathbf{W_2}\mathbf{X}\right)^{-1} \mathbf{X}^{'}\mathbf{W_2} \\ \vdots \\ \mathbf{x}_{\mathbf{n}}^{'} \left(\mathbf{X}^{'}\mathbf{W_n}\mathbf{X}\right)^{-1} \mathbf{X}^{'}\mathbf{W_n} \end{bmatrix} \mathbf{y} \\ & + (1-\lambda) \begin{pmatrix} \frac{\mathbf{K}(\frac{\mathbf{x_{11}-x_1}}{\mathbf{b_{11}}})}{\sum_{\mathbf{l=1}}^{\mathbf{n}} \mathbf{K}(\frac{\mathbf{x_l-x_1}}{\mathbf{b_{11}}})} & \frac{\mathbf{K}(\frac{\mathbf{x_{12}-x_1}}{\mathbf{b_{12}}})}{\sum_{\mathbf{l=1}}^{\mathbf{n}} \mathbf{K}(\frac{\mathbf{x_l-x_1}}{\mathbf{b_{12}}})} & \cdots & \frac{\mathbf{K}(\frac{\mathbf{x_{1k}-x_1}}{\mathbf{b_{1k}}})}{\sum_{\mathbf{l=1}}^{\mathbf{n}} \mathbf{K}(\frac{\mathbf{x_l-x_1}}{\mathbf{b_{1k}}})} \\ \frac{\mathbf{K}(\frac{\mathbf{x_{21}-x_2}}{\mathbf{b_{21}}})}{\sum_{\mathbf{l=1}}^{\mathbf{n}} \mathbf{K}(\frac{\mathbf{x_l-x_2}}{\mathbf{b_{21}}})} & \frac{\mathbf{K}(\frac{\mathbf{x_{22}-x_2}}{\mathbf{b_{22}}})}{\sum_{\mathbf{l=1}}^{\mathbf{n}} \mathbf{K}(\frac{\mathbf{x_l-x_2}}{\mathbf{b_{22}}})} & \cdots & \frac{\mathbf{K}(\frac{\mathbf{x_{2k}-x_2}}{\mathbf{b_{2k}}})}{\sum_{\mathbf{l=1}}^{\mathbf{n}} \mathbf{K}(\frac{\mathbf{x_l-x_2}}{\mathbf{b_{2k}}})} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathbf{K}(\frac{\mathbf{x_{n1}-x_n}}{\mathbf{b_{n1}}})}{\sum_{\mathbf{l=1}}^{\mathbf{n}} \mathbf{K}(\frac{\mathbf{x_l-x_n}}{\mathbf{b_{n1}}})} & \frac{\mathbf{K}(\frac{\mathbf{x_{n2}-x_n}}{\mathbf{b_{n2}}})}{\sum_{\mathbf{l=1}}^{\mathbf{n}} \mathbf{K}(\frac{\mathbf{x_l-x_n}}{\mathbf{b_{n2}}})} & \cdots & \frac{\mathbf{K}(\frac{\mathbf{x_{nk}-x_n}}{\mathbf{b_{nk}}})}{\sum_{\mathbf{l=1}}^{\mathbf{n}} \mathbf{K}(\frac{\mathbf{x_l-x_n}}{\mathbf{b_{nk}}})} \end{pmatrix} \mathbf{y} \end{aligned} \tag{26}$$

where $\mathbf{W_i} = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix}, i = 1, 2, ..., n.$

$$\frac{w_{i1}=\prod_{j=1}^{k} K\left(\frac{x_{ij}-x_{1j}}{b_{ij}}\right)}{\sum_{p=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{pj}-x_{1j}}{b_{pj}}\right)}, p=1,2,\ldots,n, j=1,2,\ldots,k,$$

$$\begin{pmatrix} \hat{y}_1^{(AVGR)} \\ \hat{y}_2^{(AVGR)} \\ . \\ . \\ . \\ \hat{y}_n^{(AVGR)} \end{pmatrix} = \begin{pmatrix} \lambda(\mathbf{x_1'}\left(\mathbf{X'W_1X}\right)^{-1}\mathbf{X'W_1})\mathbf{y} \\ \lambda(\mathbf{x_2'}\left(\mathbf{X'W_2X}\right)^{-1}\mathbf{X'W_2})\mathbf{y} \\ . \\ . \\ . \\ \lambda(\mathbf{x_n'}\left(\mathbf{X'W_nX}\right)^{-1}\mathbf{X'W_n})\mathbf{y} \end{pmatrix}$$

$$+(\mathbf{1}-\lambda)\begin{pmatrix} \frac{\mathbf{K}(\frac{\mathbf{x_{11}-x_1}}{\mathbf{b_{11}}})\mathbf{y_1}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_1}}{\mathbf{b_{11}}})} + \frac{\mathbf{K}(\frac{\mathbf{x_{12}-x_1}}{\mathbf{b_{12}}})\mathbf{y_2}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_1}}{\mathbf{b_{12}}})} + \cdots + \frac{\mathbf{K}(\frac{\mathbf{x_{1k}-x_1}}{\mathbf{b_{1k}}})\mathbf{y_n}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_1}}{\mathbf{b_{1k}}})} \\ \frac{\mathbf{K}(\frac{\mathbf{x_{21}-x_2}}{\mathbf{b_{21}}})\mathbf{y_1}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_2}}{\mathbf{b_{21}}})} + \frac{\mathbf{K}(\frac{\mathbf{x_{22}-x_2}}{\mathbf{b_{22}}})\mathbf{y_2}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_2}}{\mathbf{b_{22}}})} + \cdots + \frac{\mathbf{K}(\frac{\mathbf{x_{2k}-x_2}}{\mathbf{b_{2k}}})\mathbf{y_n}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_2}}{\mathbf{b_{2k}}})} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \ldots \vdots \quad \vdots \\ \frac{\mathbf{K}(\frac{\mathbf{x_{n1}-x_n}}{\mathbf{b_{n1}}})\mathbf{y_1}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_n}}{\mathbf{b_{n1}}})} + \frac{\mathbf{K}(\frac{\mathbf{x_{n2}-x_n}}{\mathbf{b_{n2}}})\mathbf{y_2}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_n}}{\mathbf{b_{n2}}})} + \cdots + \frac{\mathbf{K}(\frac{\mathbf{x_{nk}-x_n}}{\mathbf{n_k}})\mathbf{y_n}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_n}}{\mathbf{b_{nkk}}})} \end{pmatrix}$$

$$Z_1 = \left(\mathbf{x_1'}\left(\mathbf{X'W_1X}\right)^{-\mathbf{1}}\mathbf{X'W_1}\right)\mathbf{y}, \quad \mathbf{Z_2} = \left(\mathbf{x_2'}\left(\mathbf{X'W_2X}\right)^{-\mathbf{1}}\mathbf{X'W_2}\right)\mathbf{y},$$

$$\ldots, \mathbf{Z_n} = \left(\mathbf{x_n'}\left(\mathbf{X'W_nX}\right)^{-\mathbf{1}}\mathbf{X'W_n}\right)\mathbf{y}$$

and $\xi_1 = \frac{\mathbf{K}\left(\frac{\mathbf{x_{11}-x_1}}{\mathbf{b_{11}}}\right)\mathbf{y_1}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}\left(\frac{\mathbf{x_l-x_1}}{\mathbf{b_{11}}}\right)} + \frac{\mathbf{K}\left(\frac{\mathbf{x_{12}-x_1}}{\mathbf{b_{12}}}\right)\mathbf{y_2}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}\left(\frac{\mathbf{x_l-x_1}}{\mathbf{b_{12}}}\right)} + \ldots + \frac{\mathbf{K}(\frac{\mathbf{x_{1k}-x_1}}{\mathbf{b_{1k}}})\mathbf{y_n}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_1}}{\mathbf{b_{1k}}})}$,

similarly,

$$\xi_2 = \frac{\mathbf{K}\left(\frac{\mathbf{x_{21}-x_2}}{\mathbf{b_{21}}}\right)\mathbf{y_1}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}\left(\frac{\mathbf{x_l-x_2}}{\mathbf{b_{21}}}\right)} + \frac{\mathbf{K}\left(\frac{\mathbf{x_{22}-x_2}}{\mathbf{b_{22}}}\right)\mathbf{y_2}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}\left(\frac{\mathbf{x_l-x_2}}{\mathbf{b_{22}}}\right)} + \ldots + \frac{\mathbf{K}(\frac{\mathbf{x_{2k}-x_2}}{\mathbf{b_{2k}}})\mathbf{y_n}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_2}}{\mathbf{b_{2k}}})}$$

and

$$\xi_n = \frac{\mathbf{K}(\frac{\mathbf{x_{n1}-x_n}}{\mathbf{b_{n1}}})\mathbf{y_1}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_n}}{\mathbf{b_{n1}}})} + \frac{\mathbf{K}(\frac{\mathbf{x_{n2}-x_n}}{\mathbf{b_{n2}}})\mathbf{y_2}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_n}}{\mathbf{b_{n2}}})} + \ldots + \frac{\mathbf{K}(\frac{\mathbf{x_{nk}-x_n}}{\mathbf{n_k}})\mathbf{y_n}}{\sum_{\mathbf{l=1}}^{\mathbf{n}}\mathbf{K}(\frac{\mathbf{x_l-x_n}}{\mathbf{b_{nkk}}})}$$

$$\begin{pmatrix} \hat{y}_1^{(AVGR)} \\ \hat{y}_2^{(AVGR)} \\ . \\ . \\ . \\ \hat{y}_n^{(AVGR)} \end{pmatrix} = \begin{pmatrix} \lambda^* Z_1 \\ \lambda^* Z_2 \\ . \\ . \\ . \\ \lambda^* Z_n \end{pmatrix} + (1-\lambda^*)\begin{pmatrix} \xi_1 \\ \xi_2 \\ . \\ . \\ . \\ \xi_n \end{pmatrix}$$

where, $\varrho_1 = \lambda^* Z_1 + (1-\lambda^*)\,\xi_1, \varrho_2 = \lambda^* Z_2 + (1-\lambda^*)\,\xi_2, \ldots, \varrho_n = \lambda^* Z_n + (1-\lambda^*)\,\xi_n$

$$\begin{pmatrix} \hat{y}_1^{(AVGR)} \\ \hat{y}_2^{(AVGR)} \\ . \\ . \\ . \\ \hat{y}_n^{(AVGR)} \end{pmatrix} = \begin{pmatrix} \varrho_1 \\ \varrho_2 \\ . \\ . \\ . \\ \varrho_n \end{pmatrix}$$

In a location where $\lambda = 0$,the adaptive averaging regression model is reduced to a kernel regression and for $\lambda = 1$,the adaptive averaging regression model is a LLR. But if locations exist where some portion of the component estimates of KER and LLR needed to be added to obtain the estimate of AVGR, then proportion of the mixing parameter, λ is required to be in the interval (0, 1). The optimal value λ^* of λ in (25) is chosen based on the minimization of the cross-validation criterion.

$$PRESS^{**}(\Omega, \lambda) = \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_{i,-i}^{(\cdot)}(\Omega, \lambda) \right)^2}{n - trace\left(H^{(\cdot)}(\Omega, \lambda) \right) + (n-k-1)\frac{SSE_{\max} - SSE_{\Omega}}{SSE_{\max}}} \quad (27)$$

where Ω=$[b_{1j}^*, b_{2j}^*, \ldots, b_{nj}^*]$ is the vector of optimal bandwidths, $SSE_{\mathbf{\Omega}}$ is the Sum of Squared Errors associated with the set of the optimal bandwidths,$[b_{1j}^*, b_{2j}^*, \ldots, b_{nj}^*]$, $trace\left(H^{(\cdot)}(\Omega, \lambda) \right)$ is the trace of the Hat matrix, and $\hat{y}_{i,-i}^{(\cdot)}(\Omega, \lambda)$ is the leave-one-out cross-validation estimate of y_i (Mays *et al.* 2001; Wan and Birch 2011).

3. Results and Discussion

We shall investigate the performance of the adaptive $AVGR_{AB}$ over OLS, LLR_{FB} and LLR_{EAB} in terms of the goodness-of-fit statistics and the optimal settings of the explanatory variables that optimize the response using the two RSM data.

3.1 *Single response optimization problem*

In this paper that involves a single response, the optimization of the fitted response is aimed at the identification of the settings of the explanatory variables that will maximize or minimize the fitted response (Pickle, 2006). Therefore, the optimization criterion is based on the constrained minimization of the estimated Squared Distance from Target, (SDT), given as:

$$Minimize \quad S\hat{D}T = (\hat{y}(\mathbf{x}) - T)^2 \qquad s.t \quad x\epsilon\varphi, \tag{28}$$

where φ is the design space for the study, T denotes the target value set by the researcher, $\hat{y}(\mathbf{x})$ is the estimated response at the settings x of the explanatory variables (Pickle, 2006; Najafi *et al.*, 2011; Eguasa *et al*. 2022).

3.2 *Multi-Response optimization problem*

This involves the optimization of two or more responses simultaneously with the associated explanatory variables $(x_{i1}, x_{i2}, ..., x_{ik})$.

Based on the type of response, the desirability function transforms the estimated response, $\hat{y}_p(\mathbf{x})$ to different individual scalar measure, $d_p(\hat{y}_p(\mathbf{x}))$ namely:

For larger-the-better (LTB) response, $d_p(\hat{y}_p(\mathbf{x}))$ is given as:

$$d_p(\hat{y}_p(\mathbf{x})) = \begin{cases} 0, & \hat{y}_p(\mathbf{x}) < L \\ \left\{\frac{\hat{y}_p(\mathbf{x}) - L}{T - L}\right\}^{t_1}, & L \leq \hat{y}_p(\mathbf{x}) \leq T, \\ 1, & \hat{y}_p(\mathbf{x}) > T, \end{cases} \qquad s.t \quad \mathbf{x}\epsilon\varphi, \tag{29}$$

where$T$and L are the maximum acceptable value and lower limit, respectively, of the p^{th} response.
For a smaller-the-better (STB) response, $d_p(\hat{y}_p(\mathbf{x}))$ is given as:

$$d_p(\hat{y}_p(\mathbf{x})) = \begin{cases} 1, & \hat{y}_p(\mathbf{x}) < T, \\ \left\{\frac{U - \hat{y}_p(\mathbf{x})}{U - T}\right\}^{t_2}, & T \leq \hat{y}_p(\mathbf{x}) \leq U, \\ 0, & \hat{y}_p(\mathbf{x}) > U, \end{cases} \qquad s.t \quad \mathbf{x}\epsilon\varphi \tag{30}$$

where T and U are the minimum acceptable value and upper limit, respectively, of the p^{th} response.
For the nominal-the-better (NTB) response, $d_p(\hat{y}_p(\mathbf{x}))$ is a two sided transformation given as:

$$d_p\left(\hat{y}_p\left(\mathbf{x}\right)\right)=\begin{cases}\left\{\frac{\hat{y}_p(\mathbf{x})-L}{\rho-L}\right\}^{t_1}, L\leq\hat{y}_p(\mathbf{x})<\rho,\\ \left\{\frac{U-\hat{y}_p(\mathbf{x})}{U-\rho}\right\}^{t_2}, \rho\leq\hat{y}_p\left(\mathbf{x}\right)\leq U,\\ \quad 0, \quad otherwise\end{cases} \qquad s.t \quad \mathbf{x}\epsilon\varphi, \tag{31}$$

where ρ is the target value of the p^{th} response. However, for RSM data, the parameters values of t_1 and t_2 are weights taken to be 1 for linearity (Castillo, 2007; Wan, 2007; He *et al*., 2009; He *et al.,* 2012).

3.3 *The overall desirability function*

The objective of desirability function is to maximize the overall desirability,D, which is the geometric mean of the individual desirability functions. Overall desirability *D* is given as:

$$D=\sqrt[p]{\left(d_1(\hat{y}_1\left(\mathbf{x}\right)).d_2(\hat{y}_2\left(\mathbf{x}\right))\ldots d_v(\hat{y}_v\left(\mathbf{x}\right)\right)} \tag{32}$$

where v is the number of response variables, $d_1\hat{y}_1(\mathbf{x})$, $d_2\hat{y}_2(\mathbf{x})$,. . ., $d_v\hat{y}_v(\mathbf{x})$ are the individual desirability (He *et al*., 2012; Ramakrishnan and Arumugam, 2012; Granato and Calado, 2014). The desirability function $d_{\mathbf{m}}\left(\hat{y}_{\mathbf{m}}\left(\mathbf{x}\right)\right), \quad m = 1, 2, \ldots, v$ allocate values between 0 and 1 centered on the process requirements such that the most undesirable and desirable values are $d_{\mathbf{v}}\left(\hat{y}_{\mathbf{v}}\left(\mathbf{x}\right)\right) = 0$ and $d_{\mathbf{v}}\left(\hat{y}_{\mathbf{v}}\left(\mathbf{x}\right)\right) = 1$ respectively.

3.4 *Genetic Algorithm*

Genetic algorithms are iterative optimization techniques that repetitively apply GA operations (such as selection, crossover and mutation) to a set of solutions until some criterion of convergence is reached (Wan, 2007).

3.5 *Application I: Single response chemical process data*

The problem of the study as given in Myers and Montgomery (2002), Pickle (2006) and Edionwe *et al*. (2016) was to relate chemical yield (y) to temperature (x_1) and time (x_2) with the intention to maximize the chemical yield. The data is obtained using the Central Composite Design is given in Table 1.

Table 1: Single Response Chemical Process Data generated from the CCD

i	x_1	x_2	y
1	-1	-1	88.55
2	1	-1	85.80
3	-1	1	86.29
4	1	1	80.44
5	-1.414	0	85.50
6	1.414	0	85.39
7	0	-1.414	86.22
8	0	1.414	85.70
9	0	0	90.21
10	0	0	90.85
11	0	0	91.31

Source: Myers and Montgomery (2002)

3.6 *Data transformation using Central Composite Design (CCD)*

Following nonparametric regression procedures in RSM, the values of the explanatory variables are coded between 0 and 1. The data collected via a CCD is transformed by a mathematical relation:

$$x_{NEW} = \frac{Min\left(x_{OLD}\right) - x_0}{\left(Min\left(x_{OLD}\right) - Max\left(x_{OLD}\right)\right)} \tag{33}$$

where x_{NEW} is the transformed value, x_0 is the target value that needed to be transformed in the vector containing the old coded value, represented as x_{OLD}, $Min(x_{OLD})$ and $Max\left(x_{OLD}\right)$ are the minimum and maximum values in the vector x_{OLD} respectively, (He *et al*., (2012)).
The natural or coded variables in Table 1 can be transformed to explanatory variables in Table 2 using (33)
Target points needed to be transformed for location 6 under the coded variables are given below:
Target points x_0 : $1.414, 0$; $Min\left(x_{OLD}\right)$: $-1.414, -1.414$; $Max\left(x_{OLD}\right)$: $1.414, 1.414$

$$x_{NEW} = \frac{Min\left(x_{OLD}\right) - x_0}{\left(Min\left(x_{OLD}\right) - Max\left(x_{OLD}\right)\right)}$$

$$Explanatory \quad variable \quad x_1 : x_{61} = \frac{-1.414 - (1.414)}{((-1.414) - (1.414))} = 1.0000$$

$$Explanatory \quad variable \quad x_2 : x_{62} = \frac{-1.414 - (0)}{((-1.414) - (1.414))} = 0.5000$$

Table 2: Single Response Chemical Process Data

i	x_1	x_2	y
1	0.1464	0.1464	88.55
2	0.8536	0.1464	85.80
3	0.1464	0.8536	86.29
4	0.8536	0.8536	80.44
5	0.0000	0.5000	85.50
6	1.0000	0.5000	85.39
7	0.5000	0.0000	86.22
8	0.5000	1.0000	85.70
9	0.5000	0.5000	90.21
10	0.5000	0.5000	90.85
11	0.5000	0.5000	91.31

A second-order model was specified for the parametric technique (Pickle, 2006). The R^2_{adj} from the OLS method using the full second-order model gives 67.77%. Here, the interest is to determine if the amount of variability not explained by the specified model can be reduced by the application of the LLR method.

Table 3: Locally Adaptive Optimal Bandwidths for $AVGR_{AB}$ in the Single Response Chemical Process Data

i	$\mathbf{AVGR_{AB}}$ LLR portion of Raw Bandwidths		$\mathbf{AVGR_{AB}}$ KER portion of Raw Bandwidths	
	x_1	x_2	x_1	x_2
	b_{i1} $T^*_{11} = 1.3151 \times 10^{16}$ $T^*_{21} = 2.974 \times 10^{16}$	b_{i2} $T^*_{12} = 1.4134 \times 10^{16}$ $T^*_{22} = 1.0412 \times 10^{16}$	b_{i1} T^*_{11} =0.8279287983947552 T^*_{21} =0.7273434352880813	b_{i2} T^*_{12} =0.7727697768632593 T^*_{22} =0.8600799666900565
1	0.2672	0.1826	0.0739	0.0840
2	0.0597	0.1826	0.3756	0.0840
3	0.2672	0.1446	0.0739	0.1874
4	0.0597	0.1446	0.3756	0.1874
5	0.3288	0.0006	0.2070	0.0051
6	0.0353	0.0006	0.6337	0.0051
7	0.1448	0.3534	0.0291	0.1932
8	0.1448	0.2996	0.0291	0.3394
9	0.1448	0.0006	0.0291	0.0051
10	0.1448	0.0006	0.0291	0.0051
11	0.1448	0.0006	0.0291	0.0051

Table 4: Mixing Parameters of different models for Single Response Chemical Process Data

Response	**Model**	λ
y	OLS	NOT APPLICABLE
	LLR_{FB}	NOT APPLICABLE
	LLR_{EAB}	NOT APPLICABLE
	$AVGR_{AB}$	0.9858066436933622

Table 5: Comparison of the goodness-of-fit statistics of each method for the chemical process data

METHOD	b^*	DF_{error}	MSE	SSE	R^2	R^2_{adj}	$PRESS$	$PRESS^*$	$PRESS^{**}$
OLS	-	5.000	3.1600	15.8182	0.8388	0.6777	109.5179	21.9036	21.9036
LLR_{FB}	0.5200	5.6509	5.7000	32.2355	0.6717	0.4190	93.2835	16.5076	8.9508
LLR_{EAB}	$*$	2.9261	0.5974	1.7481	0.9822	0.9391	**46.0765**	**15.7467**	**4.2858**
$AVGR_{AB}$	$*$	2.0607	**0.4014**	**0.8272**	**0.9916**	**0.9591**	46.1257	22.3838	4.5848

Table 6. Comparison of optimization results (process requirement) for the Chemical Process Data

Approach	x_1	x_2	$\hat{y}$
OLS	0.43930	0.43610	90.9780
LLR_{FB}	0.40140	0.39438	88.3509
LLR_{EAB}	0.40771	0.42312	91.1278
$AVGR_{AB}$	**0.7271761099247455**	**0.5000008537785228**	**92.5767**

From Table 6, $AVGR_{AB}$provides the best chemical yield over OLS, LLR_{FB} and LLR_{EAB} and the two settings of the explanatory variables give the best process satisfaction.

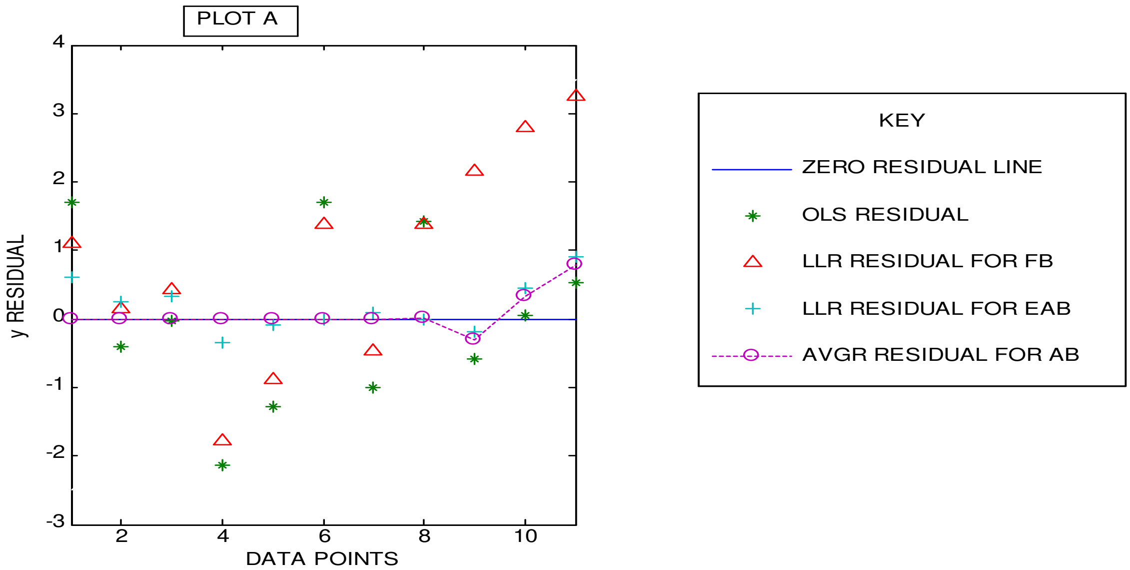

Figure 1: Model Residuals for single Response Chemical Process Data

As a follow up of the optimization result as given above, the residual result show that AVGR model have the closest residual line to the zero residual line. This also suggests that AVGR model that utilizes the adaptive bandwidth has a better fit to OLS, LLR with fixed bandwidths and LLR with existing bandwidths.

3.7 *Application II: Multi-Response Chemical Process Data*

The following problem as given in He *et al*. (2009, 2012) was to obtain the settings 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, with target value = 80; y_2 should take a value in the range L = 62 and U = 68 with target value = 65; minimize y_3 with upper limit U = 3300 and target value = 3100.
Based on the process requirements a Central Composite Design (CCD) was conducted to establish the design experiment and observed responses as presented in Table 7.

Table 7: Designed experiment and response values (He *et al*. 2009, 2012)

i	Experimental variables		Responses		
	x_1	x_2	y_1	y_2	y_3
1	-1	-1	76.5	62	2940
2	1	-1	78.0	66	3680
3	-1	1	77.0	60	3470
4	1	1	79.5	59	3890
5	-1.414	0	75.6	71	3020
6	1.414	0	78.4	68	3360
7	0	-1.414	77.0	57	3150
8	0	1.414	78.5	58	3630
9	0	0	79.9	72	3480
10	0	0	80.3	69	3200
11	0	0	80.0	68	3410
12	0	0	79.7	70	3290
13	0	0	79.8	71	3500

Table 8 is the transformed data from CCD to RSM data using the mathematical relation in (33) that needed to lie between zero and one inclusively.

Table 8: Multiple Response 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

Tables 9, 10 and 11 are the LLR portion of Raw Bandwidths and KER portion of Raw Bandwidths, and their respective tuning parameters $T^*_{ij}, i = 1, 2, \ldots, n; j =$

$1, 2, \ldots, k$.

Table 9: Optimal values of tuning parameters and Proposed Bandwidths for $y_1$using $AVGR_{AB}$

i	$\mathbf{AVGR_{AB}}$ LLR portion of Raw Bandwidths		$\mathbf{AVGR_{AB}}$ KER portion of Raw Bandwidths	
	x_1	x_2	x_1	x_2
	b_{i1} $T^*_{11} = 1.0031 \times 10^{16}$ $T^*_{21} = 5.9998 \times 10^{16}$	b_{i2} $T^*_{12} = 1.2694 \times 10^{16}$ $T^*_{22} = 1.0541 \times 10^{16}$	b_{i1} $T^*_{11} = 0.43958982658026424$ $T^*_{21} = 0.5433341698401398$	b_{i2} $T^*_{12} = 0.24313039334432818$ $T^*_{22} = 0.5007258971363912$
1	0.2269	0.1655	0.0234	0.0105
2	0.1284	0.1655	0.5043	0.0105
3	0.2269	0.1218	0.0234	0.3529
4	0.1284	0.1218	0.5043	0.3529
5	0.2508	0.0008	0.1099	0.0604
6	0.1115	0.0008	0.7899	0.0604
7	0.1741	0.3174	0.0776	0.0608
8	0.1741	0.2555	0.0776	0.5449
9	0.1741	0.0008	0.0776	0.0604
10	0.1741	0.0008	0.0776	0.0604
11	0.1741	0.0008	0.0776	0.0604
12	0.1741	0.0008	0.0776	0.0604
13	0.1741	0.0008	0.0776	0.0604

Table 10: Optimal values of tuning parameters and locally adaptive bandwidths for y_2 using $AVGR_{AB}$

i	$\mathbf{AVGR_{AB}}$ LLR portion of Raw Bandwidths		$\mathbf{AVGR_{AB}}$ KER portion of Raw Bandwidths	
	x_1	x_2	x_1	x_2
	b_{i1} T^*_{11} =1.524×10^{16} T^*_{21} =4.4808×10^{16}	b_{i2} T^*_{12} =0.4149×10^{16} T^*_{22} =7.6721×10^{16}	b_{i1} $T^*_{11} = 0.1629894899642587$ $T^*_{21} = 0.9121204184044590$	b_{i2} $T^*_{12} = 0.3168700509293532$ $T^*_{22} = 0.8144733668083716$
1	0.3328	0.0960	0.0188	0.0325
2	0.1460	0.0960	0.0310	0.0325
3	0.3328	0.0627	0.0188	0.0952
4	0.1460	0.0627	0.0310	0.0952
5	0.3810	0.0784	0.0407	0.0041
6	0.1168	0.0784	0.0580	0.0041
7	0.2299	0.1037	0.0004	0.0792
8	0.2299	0.0567	0.0004	0.1678
9	0.2299	0.0784	0.0004	0.0041
10	0.2299	0.0784	0.0004	0.0041
11	0.2299	0.0784	0.0004	0.0041
12	0.2299	0.0784	0.0004	0.0041
13	0.2299	0.0784	0.0004	0.0041

Table 11: Optimal values of tuning parameters and locally adaptive bandwidths for $y_3$using $AVGR_{AB}$

i	$\mathbf{AVGR_{AB}}$ LLR portion of Raw Bandwidths		$\mathbf{AVGR_{AB}}$ KER portion of Raw Bandwidths	
	x_1	x_2	x_1	x_2
	b_{i1} $T^*_{11} = 0.9987 \times 10^{16}$ $T^*_{21} = 3.2763 \times 10^{16}$	b_{i2} $T^*_{12} = 0.6603 \times 10^{16}$ $T^*_{22} = 3.6044 \times 10^{16}$	b_{i1} $T^*_{11} = 0.1953366132287132$ $T^*_{21} = 0.9273616121051431$	b_{i2} $T^*_{12} = 0.0841374339872218$ $T^*_{22} = 0.4158832406589298$
1	0.2070	0.1393	0.0229	0.0018
2	0.0573	0.1393	0.0345	0.0018
3	0.2070	0.0457	0.0229	0.2028
4	0.0573	0.0457	0.0345	0.2028
5	0.2497	0.0862	0.0488	0.0415
6	0.0379	0.0862	0.0653	0.0415
7	0.1205	0.1651	0.0003	0.0210
8	0.1205	0.0327	0.0003	0.3052
9	0.1205	0.0862	0.0003	0.0415
10	0.1205	0.0862	0.0003	0.0415
11	0.1205	0.0862	0.0003	0.0415
12	0.1205	0.0862	0.0003	0.0415
13	0.1205	0.0862	0.0003	0.0415

Table 12: Mixing Parameters of different models for Multiple Chemical Process Data

Response	**Model**	λ
y_1	OLS	NOT APPLICABLE
	LLR_{FB}	NOT APPLICABLE
	LLR_{EAB}	NOT APPLICABLE
	$AVGR_{AB}$	1.0000000000000000
y_2	OLS	NOT APPLICABLE
	LLR_{FB}	NOT APPLICABLE
	LLR_{EAB}	NOT APPLICABLE
	$AVGR_{AB}$	0.8229136094582358
y_3	OLS	NOT APPLICABLE
	LLR_{FB}	NOT APPLICABLE
	LLR_{EAB}	NOT APPLICABLE
	$AVGR_{AB}$	1.0000000000000000

Table 12 is the mixing parameters only applicable to $AVGR_{AB}$ model for the multi-response chemical process data.

Table 13: Model goodness of fits statistics for Multiple Response Chemical Process Data

Response	Model	DF	PRESS**	PRESS	SSE	MSE	$R^2(\%)$	$R^2_{Adj}(\%)$
y_1	OLS	7.0000	0.3361	2.3525	0.4962	0.0709	98.27	97.04
	LLR_{FB}	7.4717	0.5686	8.4888	4.7536	0.6362	83.46	73.44
	LLR_{EAB}	4.7777	0.2063	3.0144	0.3103	0.0649	98.92	97.29
	$AVGR_{AB}$	4.0144	**0.0480**	**0.6687**	**0.2165**	**0.0539**	**99.25**	**97.75**
y_2	OLS	7.0000	28.8726	202.1082	36.2242	5.1749	89.98	82.81
	LLR_{FB}	7.2576	22.0691	330.8149	80.2383	11.0558	77.79	63.27
	LLR_{EAB}	4.0015	9.7580	133.8681	10.0013	**2.4994**	**97.23**	**91.70**
	$AVGR_{AB}$	4.0008	**9.1371**	**127.9099**	**10.0000**	2.4995	**97.23**	**91.70**
y_3	OLS	7.0000	159080	1113600	207870	29696	75.90	58.68
	LLR_{FB}	9.2798	56513	588010	243460	26235	71.77	63.50
	LLR_{EAB}	5.8380	40779	508170	92621	**15865**	89.26	**77.93**
	$AVGR_{AB}$	4.0000	**22024**	**308240**	**65720**	16430	**92.3804**	77.14

The results obtained from Table 13, clearly shows that $AVGR_{AB}$ gave the better performance statistic as compared with OLS, LLR_{FB} and LLR_{EAB} for the multi-response problem.

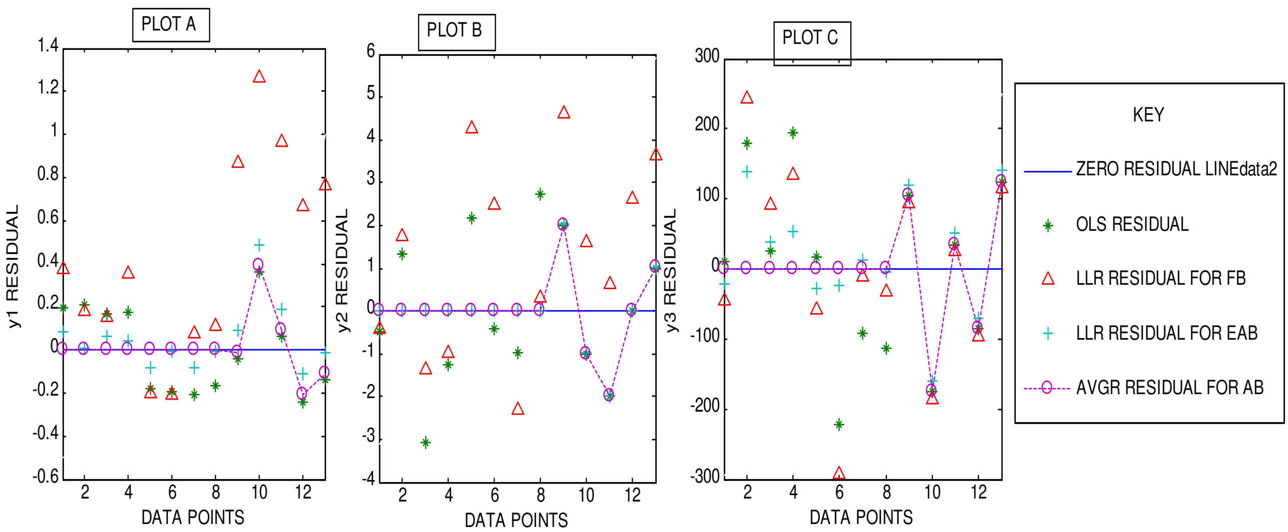

Figure 2: Plot A: maximize chemical yield; Plot B: is a two sided transformation of viscosity; Plot C: minimize molecular weight

In Figure 2, a critical look at residual plots for chemical yield, viscosity and molecular weight shows that $AVGR_{AB}$ estimated f better than OLS, LLR_{FB} and LLR_{EAB} as given in (1).

Table 14: Model optimal solution based on the Desirability function for Multiple 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	31.3800
LLR_{FB}	0.4481	0.3709	78.5537	66.7908	3290.8	0.0358	0.4031	0.0461	8.7200
LLR_{EAB}	0.5155	0.3467	78.6965	65.0328	3285.9	0.1310	0.9891	0.0703	20.8837
$AVGR_{AB}$	**0.9966702516947067**	**0.6354728955289123**	79.5582	63.3972	3224.0000	0.7054	0.4657	0.3802	**49.9899**

From Table 14, $AVGR_{AB}$provides the best chemical yield, viscosity and molecular weight over OLS, LLR_{FB} and LLR_{EAB} and the two settings of the explanatory variables give the best process satisfaction.

3.8 *Simulation study*

In this section, we compare the performances of the respective regression models, OLS, LLR_{FB}, LLR_{EAB} and $AVGE_{AB}$using simulated data. Each simulation comprises of 500 data sets based on the following underlying polynomial models:

$$\textbf{Model 1}: y_i = 20 - 10x_{1i} - 25x_{2i} - 15x_{1i}x_{2i} + 20x_{1i}^2 + 50x_{2i}^2 + \gamma\left\{2\sin\left(4\pi x_{1i}\right) + 2\cos\left(4\pi x_{2i}\right) - 2\sin\left(4\pi x_{1i}x_{2i}\right)\right\} + \varepsilon_i;$$

$$\textbf{Model 2}: y_i = 66 + 22x_{1i} + 10x_{2i} + 13x_{1i}x_{2i} - 23x_{1i}^2 - 25x_{2i}^2 + \gamma\left\{2\sin\left(3\pi x_{1i}\right) - 2\cos\left(3\pi x_{2i}\right) + 2\sin\left(2\pi x_{1i}x_{2i}\right)\right\} + \varepsilon_i;$$

$$\textbf{Model 3}: y_i = 45 + 27x_{1i} + 9x_{2i} + 19x_{3i} - 22x_{1i}x_{2i} - 17x_{2i}x_{3i} - 8x_{1i}x_{3i} + 10x_{1i}^2 + 13x_{2i}^2 + 13x_{3i}^2 + \gamma\left(2\sin\left(3\pi x_{1i}\right) - 2\cos\left(3\pi x_{2i}\right) - 3\cos\left(4\pi x_{3i}\right) + 2\sin\left(3\pi x_{1i}x_{2i}\right) + 2\cos\left(3\pi x_{2i}x_{3i}\right) + 2\sin\left(3\pi x_{1i}x_{3i}\right) + \varepsilon_i$$

where the x_{1i}, x_{2i} and x_{3i} are the explanatory variables, ε_i,$i = 1, 2, \ldots, n$, are the error terms which are normally distributed with mean zero and variance 1, and represents a misspecification parameter. The values of the explanatory variables are presented in Table 15.

3.8.1 ***Simulation Study 1:*** *Multi-Response Chemical Process Problem*

This problem is analyzed by He *et al.*, (2012) with the aim to get the setting of the explanatory variables x_1 and x_2 (representing reaction time and temperature, respectively).

Table 15: The CCD for the Simulating Data for Models 1-3

i	x_1	x_2
1	0.8536	0.8536
2	0.1464	0.8536
3	0.8536	0.1464
4	0.1464	0.1464
5	1.0000	0.5000
6	0.0000	0.5000
7	0.5000	1.0000
8	0.5000	0.0000
9	0.5000	0.5000
10	0.5000	0.5000
11	0.5000	0.5000
12	0.5000	0.5000
13	0.5000	0.5000

The goal of the simulation study is to validate the polynomial regression models when applied to study that involve two explanatory variables. The model Average Sum of Squares (AVESSE) for each degree of model misspecification is presented in Table 15.

Table 16: Comparison of the AVESSE of each method for each model in the simulation studies

Model	γ	OLS	LLR_{FB}	LLR_{EAB}	$AVGR_{AB}$
(1)	0.00	6.9849	68.9816	**6.3277**	6.6450
	0.50	18.0887	61.6146	**14.4455**	15.6081
	1.00	51.0910	99.0211	**15.1152**	34.8047
(2)	0.00	7.0210	34.0919	13.6632	**6.6107**
	0.50	13.7667	41.8323	20.9044	**13.4577**
	1.00	39.1912	72.1624	38.9560	**34.8232**
(3)	0.00	7.0113	28.9237	**6.2117**	6.3072
	0.50	125.2006	254.4773	12.5466	**9.8355**
	1.00	479.6291	747.5212	71.8911	**12.1333**

Clearly, from the simulated results as given in Table 16, $AVGR_{AB}$ gave smaller AVESSE over other regression models considered even with the misspecification parameters ranges from zero to one.

4. Conclusion

The quality of a process or product is one of the most crucial indicators that enlighten a consumer's choice from one product in the midst of several contending products. Therefore, refining the quality of a product is a vital approach that embraces growth in business, improved competitiveness and enormous revenues to investment also see (Pickle, (2006); Castillo, (2007)).
In RSM, since the stages are sequential, a new product is subjected to experimental design phase, modeling phase of the fitted regression model and the optimization phase with the goal to find setting of the explanatory variables that optimize responses as it relates to the quality of the new product. This sequential procedure is referred to as product qualification in the manufacturing industries, see (Najafi *et al.*, (2011); Nair *et al.*, (2014)).
Thereafter, large amount of this product is produced upon receipt of the quality and reliability of the optimal setting of the explanatory variables which hinge on how well the regression model fits the data, also see (Castillo, (2007); He *et al.,* (2012)).
In this paper, we proposed an adaptive averaging regression ($AVGR_{AB}$) model that combines local linear regression (LLR) and the kernel regression models via convex combination for RSM data. The proposed $AVGR_{AB}$ model utilized the locally adaptive bandwidths from literature for its fitting procedures and simulation was carried-out for the multi-response data. The results of the goodness of fits and optimal solutions obtained indicated that the $AVGR_{AB}$ regression model utilizing the proposed bandwidths selector (Eguasa *et al.*, (2022)) performs better than OLS, the LLR with fixed bandwidth, and the LLR that utilizes the existing bandwidths.

The limitations of local linear regression models stemmed from the peculiarities in RSM data such as small sample size, utilizing more than one explanatory variables and data sparsity. In spite of this, the $AVGR_{AB}$ model in RSM, shown to be the appropriate choice which tremendously improved the goodness-of-fit statistics and optimization results for single and multi-response problems considered. More so, the multi-response applications considered, $AVGR_{PAB}$ outperformed existing models in terms of the goodness-of-fit statistics, and process requirements.
Lastly, simulation study was carried out on OLS, LLR with fixed bandwidth, $LLR_{EAB}$and $AVGR_{AB}$to investigate the effect of the misspecification parameter as it increases from zero to one. It was observed that the$AVGR_{AB}$was considerably stable over OLS, LLR with fixed bandwidth, LLR_{EAB} in terms of AVESSE as the misspecification parameter increases from zero to one in all the problems considered.

Acknowledgment

I sincerely wish to thank the reviewers for the due diligence with the corrections made on the manuscript.

References

Carley, M. K., Kamneva, Y. N. and Reminga, J. (2004). Response surface Methodology. Technical report written to Center for Computational Analysis of Social and Organizational Systems (CASOS), Institute for Software Research International - Carnegie Mellon University, School of Computer Science, Pittsburgh, PA, 15213.

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

Choi, E. and Hall, P. (1998). On Bias Reduction in Local Linear Smoothing. Biometrika, 85(2), 333-345.

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), 388-404.

Eguasa, O., Mbegbu, J. I. and Edionwe, E. (2019). On the Use of Nonparametric Regression Model for Response Surface Methodology (RSM). Benin Journal of Statistics, 2(5), 61-75.

Eguasa, O. Edionwe, E and Mbegbu, J. I. (2022): Local Linear Regression and the problem of dimensionality: A remedial strategy via a new locally adaptive bandwidths selector, Journal of Applied Statistics, DOI:10.1080/02664763.2022.2026895

Fan, J. (1993). Local Linear Regression Smoothers and Their Minimax Efficiencies. The Annals of Statistics, 21(1), 196-216.

Gramato, D. and Calado, V. (2014). The use and importance of design of experiments (DOE) in Process modelling in food science and technology. Mathematical and Statistical Methods in Food Science and Technology, John Wiley and Sons, Ltd. First Edition.

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.*

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.

Kohler, M. A., Schindler, and Sperlich, S (2014). A review and comparison of bandwidth selection methods for kernel regression. International Statistical Review, 82, 243-274.

Liu, C. (2011). Averaging Estimators for Kernel Regressions. Working Paper, University of Wisconsin. chuanliu.weebly.com/../42.pdf.

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.

Myers, R. H. and Montgomery, D. C. (2002). Response Surface Methodology: Product and Process Optimization Using Designed Experiments. Second Edition, John Wiley & Sons, New York.

Nadaraya, E.A. (1964). On estimating regression. Journal of Theory of Probability and its Application, 9(1), 141-142.

Nair, A. T., Makwana, A. R. and Ahammed, M. M. (2014). The use of Response Surface Methodology for modelling and analysis of water and waste – water treatment processes: A Review. Water Science and Technology, 69(3), 464-478.

Najafi, S., Salmasnia, A. and Kazemzadeh, R.B. (2011). Optimization of robust design for multiple response problem. Australian Journal of Basic and Applied Sciences, 5(9), 1566-1577.

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

Ramakrishnan, R. and Arumugam, R. (2011). 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, M. Sc. Thesis and Dissertations, Virginia Commonwealth University.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. The Annals of Statistics, 22, 1346-1370.

Walker, E. L., Starnes, B. A., Birch, J. B. and Mays, J. E., (2002). Model Robust Calibration: Methods and Application to Electrically-Scanned Pressure Transducers, Technical Report, Dept. of Mathematics, Virginia Commonwealth University, Richmond, VA, USA.

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

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(1), 93-104.