FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 6 No. 5, October, 2022, pp 41 - 49
DOI: https://doi.org/10.33003/fjs-2022-0605-1095

AN ADAPTIVE LOCAL LINEAR REGRESSION METHOD FOR MOBILE SIGNAL STRENGTH WITH APPLICATION TO RESPONSE SURFACE METHODOLOGY

[*1]Akhideno I. O. and [2]Eguasa O.

[1]Department of Electrical/Electronic Engineering, Benson Idahosa University, Benin City, Edo State
[2]Department of Physical Sciences, Benson Idahosa University, Benin City, Edo State

*Corresponding authors' email: iakhideno@biu.edu.ng

ABSTRACT

The received signal strength is vital in telecommunication network or technology as it is affected by varying environmental factors such as temperature (T^0C), relative humidity (H%), air quality index (m) and distance from base station (m). In this paper, we seek to find a regression model via a circumscribed central composite design that can adequately represent the functional relationship between the received signal strength and the respective factors applied to response surface methodology with the goal to obtain settings of these factors that would simultaneously optimize the received signal strength (Long Term Evolution (LTE), Third Generation network (3G) and second Generation network (2G)) technologies. The frequently utilized regression model is the parametric regression model (second-order model), though superior but lack credibility in terms of model misspecification and as a result, the optimum setting of the factors are miscalculated. In addressing the pitfall of the parametric regression model (PRM), we introduce a flexible adaptive local linear regression model that can capture local trend in the data, which ordinarily a misspecified PRM could not address. In the application, two regression models were used and the results show that the adaptive local linear regression model outperformed the parametric counterpart in terms of goodness-of-fit statistics, residual plot and optimization of the received signal strength.

Keywords: Circumscribed central composite design, Local linear regression model, Parametric regression model, Received signal strength, Response surface methodology

INTRODUCTION

In this study, the data collected was motivated to capture the variation in received signal strength for Long Term Evolution (LTE), Third generation network (3G) and Second generation network (2G) wireless communication technologies. The technologies are mostly affected by varying environmental factors which were observed between the time duration from January 2019 to March 2020 (Choudhary *et al*., 2021). The Influence of four factors; that is the varying Temperature (12°C to 48°C), Relative Humidity (25% to75%), Distance from Base Station (98 m to 300 m) and Air Quality index (AQI) for PM 2.5 (50 m to 500 m) were observed on Received Signal Strength of LTE (RSSLTE), 3G (RSS3G) and 2G (RSS2G) (Choudhary *et al*., 2021). These measurements on the factors were coded using appropriate experimental design technique and thereafter be transformed to Response Surface Methodology (RSM) data by a mathematical relation that needed to lie in the interval of zero and one inclusively. The objective of this study is to obtain a better goodness-of-fit statistics, residual plot and optimization results.

RSM is a collection of mathematical and statistical techniques employed by Industrial Statistician and Engineers for empirical model building. In the modeling and analysis of data, the response is influenced by one or more explanatory variables (Eguasa *et al*., 2022). There are three main stages in RSM, namely, the Experimental Design Phase, the Modeling Phase, and the Optimization Phase (Castillo, 2007).

In the Modeling phase of RSM, a fundamental assumption is that the relationship between the response variable y and k explanatory variables $x_1, x_2, \dots, x_k$, can be represented as:

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

where the mean function f denotes the true but unknown relationship between the response variable and the k explanatory variables, ε_i, $i = 1,2,\dots,n$, are random error terms assumed to have a normal distribution with mean zero and constant variance and n is the sample size (Myers *et al*., 2009; Wan and Birch, 2011).

Parametric Regression Model

The parametric regression models are superior if the user can specify a parametric form for the data, otherwise misspecified. 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 idiosyncrasies of RSM data namely;

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

Ordinary least Squares

The Ordinary Least Squares (OLS) is an existing regression models applied in the estimation of the unknown function f in Equation (1) (Anderson-Cook and Prewitt, 2005; Edionwe *et al*., 2016). The OLS model is applied in the estimation of the unknown parameters (coefficients) in the parametric (polynomial) model that the experimenter assumes adequate to approximate f in Equation (1) (Wan and Birch, 2011). The OLS estimate $\hat{y}_i^{(OLS)}$ response in the i^{th} data point is given as:

$$\hat{y}_i^{(OLS)} = \boldsymbol{x_i}(\boldsymbol{X^T X})^{-1}\boldsymbol{X^T y} \tag{2}$$

where $\boldsymbol{y}$ is a $n \times 1$ vector of response, $\mathbf{X}$ is a $\mathrm{n} \times \mathrm{p}$ model matrix, p is the number of model parameters (coefficients), $\boldsymbol{X^T}$ is the transpose of the matrix $\mathbf{X}$, and $\boldsymbol{x_i}$ is the i^{th} row vector of the matrix $\boldsymbol{X}$ (Pickle *et al*., 2008).

In matrix notation, the vector of OLS estimated response is expressed as:

$$\hat{\boldsymbol{y}}^{(\boldsymbol{OLS})} = \begin{bmatrix} \boldsymbol{h}_1^{(\boldsymbol{OLS})} \\ \boldsymbol{h}_2^{(\boldsymbol{OLS})} \\ \vdots \\ \boldsymbol{h}_n^{(\boldsymbol{OLS})} \end{bmatrix} \boldsymbol{y} = \boldsymbol{H}^{(\boldsymbol{OLS})}\boldsymbol{y}, \qquad (3)$$

where the vector $\boldsymbol{h}_i^{(\boldsymbol{OLS})} = \boldsymbol{x_i}(\boldsymbol{X^T X})^{-\boldsymbol{1}}\boldsymbol{X^T}$ is the i^{th} row of the $n \times n$ OLS Hat matrix $\boldsymbol{H}^{(\boldsymbol{OLS})}$.

The OLS model requires several assumptions to be met for valid interpretation of its parameter estimates. Furthermore, it performs poorly if the assumed polynomial model is inadequate for the data (Wan and Birch, 2011).

A second-order linear regression model is given as:

$$y_i = \beta_0 + \sum_{j=1}^{k}\beta_j x_{ij} + \sum_{j=1}^{k}\beta_{jj}x_{ij}^2 + \sum_{j=1}^{k-1}\sum_{r=j+1}^{k}\beta_{jr}x_{ij}x_{ir} + \varepsilon_i,\ i = 1,2,\dots,n;\ r = j+1, j+2, \dots, k \qquad (4)$$

where x_{ij}, x_{ir} are the explanatory variables; β_0 is a constant coefficient; the varying coefficients β_j, β_{jj} and β_{jr} are the coefficients of linear, quadratic and interaction terms respectively.

MATERIALS AND METHODS

The philosophy behind the local linear regression model is because it is flexible and can adapt favourably in addressing boundary bias problem and is not restricted user specified form for the data (Eguasa *et al.*, 2022).

The Local Linear Regression (LLR)

The LLR model is a nonparametric regression version of the weighted least squares model (Fan and Gijbels, 1995; Hardle *et al*., 2005; Kohler *et al*., 2014).

The LLR estimate, $\hat{y}_i^{(LLR)}$ of y_i, is given as:

$$\hat{y}_i^{(LLR)} = \tilde{\boldsymbol{x}}_i(\tilde{\boldsymbol{X}}^T\boldsymbol{W_i}\tilde{\boldsymbol{X}})^{-\boldsymbol{1}}\tilde{\boldsymbol{X}}^T\boldsymbol{W_i}\boldsymbol{y} = \boldsymbol{h}_i^{(\boldsymbol{LLR})}\boldsymbol{y}, \qquad (5)$$

where $\tilde{\boldsymbol{x}}_i$ is the i^{th} row of the LLR model matrix $\tilde{\boldsymbol{X}}$ given as:

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

where x_{ij}, $i = 1,2,\dots,n, j = 1,2,\dots,k$, denotes the value of the j^{th} explanatory variable in the i^{th} data point, $\boldsymbol{W_i}$ is a $n \times n$ diagonal weights matrix given as:

$$\boldsymbol{W_i} = \begin{bmatrix} w_{1i} & 0 & \cdots & 0 \\ 0 & w_{2i} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & w_{ni} \end{bmatrix}_{n\times n} \qquad (6)$$

For instance, w_{1i}, $i = 1$, is obtained from the product kernel as:

$$w_{11} = \prod_{j=1}^{k} K\left(\frac{x_{ij}-x_{1j}}{b}\right) \Big/ \sum_{i=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{ij}-x_{1j}}{b}\right), i = 1,2,\dots,n, j = 1,2,\dots,k, \qquad (7)$$

where $K\left(\frac{x_{ij}-x_{1j}}{b}\right) = e^{-\left(\frac{x_{ij}-x_{1j}}{b}\right)^2}$ is the simplified Gaussian kernel function and b_i, $0 < b \le 1$, $i = 1,2,\dots,n$, $j = 1,2,\dots,k$, is the fixed bandwidth (smoothing parameter) (Myers *et al*., 2009; Eguasa, 2020).

Thus,

For $i = 1$ in equation (7), we have:

$$\boldsymbol{W_1} = \begin{bmatrix} w_{11} & 0 & \cdots & 0 \\ 0 & w_{12} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{1n} \end{bmatrix}_{(n\times n)} \qquad (8)$$

$$w_{11} = \frac{\prod_{j=1}^{k} K\left(\frac{x_{1j}-x_{1j}}{b}\right)}{\sum_{p=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{pj}-x_{1j}}{b}\right)}, \quad p = 1, 2, ... , n; j = 1, 2, \dots, k. \qquad (9)$$

$$w_{11} = \frac{S}{[S+T+\cdots+U]}, S = e^{-(\frac{x_{11}-x_{11}}{b})^2} e^{-(\frac{x_{12}-x_{12}}{b})^2} \dots e^{-(\frac{x_{1k}-x_{1k}}{b})^2}, T = e^{-\left(\frac{x_{21}-x_{11}}{b}\right)^2} e^{-\left(\frac{x_{22}-x_{12}}{b}\right)^2} \dots e^{-\left(\frac{x_{2k}-x_{1k}}{b}\right)^2}$$

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

$$w_{12} = \frac{\prod_{j=1}^{k} K\left(\frac{x_{2j}-x_{1j}}{b}\right)}{\sum_{p=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{pj}-x_{1j}}{b}\right)}, \quad p = 1, 2, ... , n; j = 1, 2, \dots, k. \qquad (10)$$

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

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

$$\vdots$$

$$w_{1n} = \frac{\prod_{j=1}^{k} K\left(\frac{x_{nj}-x_{1j}}{b}\right)}{\sum_{p=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{pj}-x_{1j}}{b}\right)}, \quad p = 1, \ 2, \ . \ . \ . \ , n; j = 1, 2, \dots, k. \tag{11}$$

$$w_{1n} = \frac{R}{[M+H+\cdots+R]} \tag{12}$$

$$R = e^{-\left(\frac{x_{n1}-x_{11}}{b}\right)^2} e^{-\left(\frac{x_{n2}-x_{12}}{b}\right)^2} \dots e^{-\left(\frac{x_{nk}-x_{1k}}{b}\right)^2}, H = e^{-\left(\frac{x_{21}-x_{11}}{b}\right)^2} e^{-\left(\frac{x_{22}-x_{12}}{b}\right)^2} \dots e^{-\left(\frac{x_{2k}-x_{1k}}{b}\right)^2}$$

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

Experimental design

In RSM, the number of factors is usually more than one. Hence, if the number of factors is too large, it may directly affect the response (Received signal strength) of interest, and since not all factors are desirable to be included in the experimental design for reason due to cost implication, it required the use of factor screening approach or two-level full factorial design to identify the variables with main effects (Montgomery, 2009; Nair *et al*., (2014); Eguasa *et al*., 2022). Choice of adequate levels to be studied for the explanatory variables is also important as it can affect model accuracy.

The Experimental Design phase permits an appropriate design that can provide adequate and considerable estimation relationship between the response and one or more factors. Usually applied DOEs in RSM include: 2^k full factorial design, 3^k full factorial design, and the Central Composite Design (CCD).

Table 1: Coded stages and range for the design of experiments (Choudhary *et al*., 2021)

Factors or Input parameters	Symbol	Coded Levels				
		-2(-α)	**-1(Low)**	**0(Medium)**	**1(High)**	**2(+α)**
Temperature (^{0}C)	Temp	12	23	30	40	48
Relative Humidity (%)	RH	25	37	50	63	75
Distance from Base Station (m)	DFBS	98	150	200	250	300
Air Quality Index (m)	AQI	50	150	250	370	500

The central composite design

A Central Composite Design allows for the building of the second-order regression model in a given response that is frequently used for process optimization (Sivarao *et al*., 2010; Eguasa, 2020). The three types of CCD are based on the locations of the factorial and star points in the design space namely; Circumscribed CCD (CCCD), Faced-Centered CCD and the Inscribed CCD.

The circumscribed central composite design

The most common CCD utilized in RSM is the circumscribed CCD because it allows for the estimation of curvature and the values of star points maintain rotatability which in turn depends on the factorial point of the design (Dutka *et al*., 2015).

The circumscribed CCD involves three types of trials namely; two levels (2^k) full factorial designs, $2k$ axial (star) points which are located at distance $\alpha = \sqrt[4]{2^k}$ from the center point and k_c, k^{th} central points (Bezerra *et al.*, 2008). The Circumscribed CCD can express geometrically as:

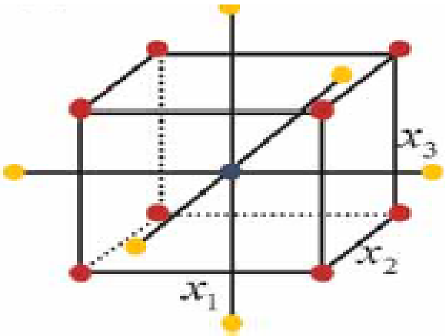

Figure 1: Circumscribed CCD (17 points, when k=3) with factorial design points (8 points), axial points (6 points) and with at least k^{th} central point (3 points).
Sources: Nair *et al*., (2014); Peasura (2015)

In this study, the CCCD has been utilized because it is cost efficient, maintain rotatability and accommodates small number of experimental runs in the design.

The mathematical expression for the CCCD is given as:

$$CCCD = 2^k + 2k + k_c \tag{13}$$

where 2^k is the factorial portion, $2k$ is the axial or star points and k_c is at least kth central points utilized in the design. In this design $k = 4$ and $k_c = 7$ which from equation (13) sum up to 31 experimental runs.

Table 2: Experimental coded level for RSM data (Choudhary *et al*., 2021)

Exptal. Run	Temp. (^{0}C)	RH (%)	DFBS (m)	AQI (m)	RSSLTE (dBm)	RSS3G (dBm)	RSS2G (dBm)
1	-1(23)	-1(37)	-1(150)	-1(150)	-123.60	-97.59	-86.68
2	1(40)	-1(37)	-1(150)	-1(150)	-107.00	-84.00	-75.00
3	-1(23)	1(63)	-1(150)	-1(150)	-107.00	-84.00	-75.00
4	1(40)	1(63)	-1(150)	-1(150)	-123.00	-97.00	-86.60
5	-1(23)	-1(37)	1(250)	-1(150)	-116.55	-92.35	-82.63
6	1(40)	-1(37)	1(250)	-1(150)	-108.75	-86.50	-78.13
7	-1(23)	1(63)	1(250)	-1(150)	-120.15	-95.10	-84.72
8	1(40)	1(63)	1(250)	-1(150)	-97.20	-77.19	-69.29
9	-1(23)	-1(37)	-1(150)	1(370)	-108.75	-86.50	-78.13
10	1(40)	-1(37)	-1(150)	1(370)	-103.86	-81.22	-74.54
11	-1(23)	1(63)	-1(150)	1(370)	-107.00	-84.00	-75.00
12	1(40)	1(63)	-1(150)	1(370)	-89.00	-68.00	-62.80
13	-1(23)	-1(37)	1(250)	1(370)	-100.80	-79.94	-71.38
14	1(40)	-1(37)	1(250)	1(370)	-107.00	-84.00	-75.00
15	-1(23)	1(63)	1(250)	1(370)	-110.90	-87.00	-77.55
16	1(40)	1(63)	1(250)	1(370)	-105.15	-83.30	-73.92
17	-2(12)	0(50)	0(200)	0(250)	-115.80	-91.74	-82.18
18	2(48)	0(50)	0(200)	0(250)	-102.70	-81.60	-73.50
19	0(30)	-2(25)	0(200)	0(250)	-112.35	-89.10	-80.22
20	0(30)	2(75)	0(200)	0(250)	-104.00	-81.55	-73.02
21	0(30)	0(50)	-2(98)	0(250)	-111.50	-87.60	-77.85
22	0(30)	0(50)	2(300)	0(250)	-112.20	-88.99	-81.09
23	0(30)	0(50)	0(200)	0(250)	-107.00	-84.00	-73.00
24	0(30)	0(50)	0(200)	0(250)	-107.00	-84.00	-73.00
25	0(30)	0(50)	0(200)	-2(50)	-105.00	-83.04	-73.79
26	0(30)	0(50)	0(200)	2(500)	-108.60	-85.79	-75.88
27	0(30)	0(50)	0(200)	0(250)	-120.00	-94.84	-84.59
28	0(30)	0(50)	0(200)	0(250)	-97.35	-77.45	-69.42
29	0(30)	0(50)	0(200)	0(250)	-107.00	-84.00	-75.00
30	0(30)	0(50)	0(200)	0(250)	-101.55	-80.55	-71.83
31	0(30)	0(50)	0(200)	0(250)	-119.00	-94.50	-82.80

Hereafter, circumscribed CCD shall be referred to as CCCD for easy reference. A CCCD has an advantage over 3^k full factorial design because it reduces the number of experimental runs (e.g. 31points in CCCD as against 81 points in 3^k design for *k*= 4).

Data transformation using central composite design (CCD) to RSM data

The values of the explanatory variables are coded between 0 and 1. The data collected via a CCD is transformed by a mathematical relation:

$$x_{NEW} = \frac{Min(x_{OLD})-x_0}{(Min(x_{OLD})-Max(x_{OLD}))} \tag{14}$$

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} , $\text{Min}(x_{OLD})$ and $Max(x_{OLD})$ are the minimum and maximum values in the vector x_{OLD} respectively, (Eguasa *et al*., 2022).

The natural or coded variables in Table 1 can be transformed to explanatory variables in Table 2 using Equation (14)

Target points needed to be transformed for location 2 under the coded variables are given below:

Target points x_0: $1, -1, -1, -1$; $Min(x_{OLD})$: $-2, -2, -2, -2$; $Max(x_{OLD})$: $2, 2, 2, 2$

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

$$Explanatory\ variable\ x_1 :\ x_{21} = \frac{-2-(1)}{((-2)-(2))} = 0.7500$$

$$Explanatory\ variable\ x_2 :\ x_{22} = \frac{-2-(-1)}{((-2)-(2))} = 0.2500$$

$$Explanatory\ variable\ x_3 :\ x_{23} = \frac{-2-(-1)}{((-2)-(2))} = 0.2500$$

$$Explanatory\ variable\ x_4 :\ x_{24} = \frac{-2-(-1)}{((-2)-(2))} = 0.2500$$

where $x_1 = \text{Temp.}\,(0\text{C}), x_2 = \text{RH}\ (\%),\ x_3 = \text{DFBS}\ (\text{m}),\ x_4 = \text{AQI}$

Table 3: Experimental CCCD for the transformed RSM data that are coded btw 0 and 1

Exp. Run	Temp. (^{0}C)	RH (%)	DFBS (m)	AQI (m)	RSSLTE (dBm)	RSS3G (dBm)	RSS2G (dBm)
1	0.2500(23)	0.2500(37)	0.2500(150)	0.2500(150)	-123.60	-97.59	-86.68
2	**0.7500(40)**	**0.2500(37)**	**0.2500(150)**	**0.2500(150)**	-107.00	-84.00	-75.00
3	0.2500(23)	0.7500(63)	0.2500(150)	0.2500(150)	-107.00	-84.00	-75.00
4	0.7500(40)	0.7500(63)	0.2500(150)	0.2500(150)	-123.00	-97.00	-86.60
5	0.2500(23)	0.2500(37)	0.7500(250)	0.2500(150)	-116.55	-92.35	-82.63
6	0.7500(40)	0.2500(37)	0.7500(250)	0.2500(150)	-108.75	-86.50	-78.13
7	0.2500(23)	0.7500(63)	0.7500(250)	0.2500(150)	-120.15	-95.10	-84.72
8	0.7500(40)	0.7500(63)	0.7500(250)	0.2500(150)	-97.20	-77.19	-69.29
9	0.2500(23)	0.2500(37)	0.2500(150)	0.7500(370)	-108.75	-86.50	-78.13
10	0.7500(40)	0.2500(37)	0.2500(150)	0.7500(370)	-103.86	-81.22	-74.54
11	0.2500(23)	0.7500(63)	0.2500(150)	0.7500(370)	-107.00	-84.00	-75.00
12	0.7500(40)	0.7500(63)	0.2500(150)	0.7500(370)	-89.00	-68.00	-62.80
13	0.2500(23)	0.2500(37)	0.7500(250)	0.7500(370)	-100.80	-79.94	-71.38
14	0.7500(40)	0.2500(37)	0.7500(250)	0.7500(370)	-107.00	-84.00	-75.00
15	0.2500(23)	0.7500(63)	0.7500(250)	0.7500(370)	-110.90	-87.00	-77.55
16	0.7500(40)	0.7500(63)	0.7500(250)	0.7500(370)	-105.15	-83.30	-73.92
17	0.0000(12)	0.5000(50)	0.5000(200)	0.5000(250)	-115.80	-91.74	-82.18
18	1.0000(48)	0.5000(50)	0.5000(200)	0.5000(250)	-102.70	-81.60	-73.50
19	0.5000(30)	0.0000(25)	0.5000(200)	0.5000(250)	-112.35	-89.10	-80.22
20	0.5000(30)	1.0000(75)	0.5000(200)	0.5000(250)	-104.00	-81.55	-73.02
21	0.5000(30)	0.5000(50)	0.0000(98)	0.5000(250)	-111.50	-87.60	-77.85
22	0.5000(30)	0.5000(50)	1.0000(300)	0.5000(250)	-112.20	-88.99	-81.09
23	0.5000(30)	0.5000(50)	0.5000(200)	0.5000(250)	-107.00	-84.00	-73.00
24	0.5000(30)	0.5000(50)	0.5000(200)	0.5000(250)	-107.00	-84.00	-73.00
25	0.5000(30)	0.5000(50)	0.5000(200)	0.0000(50)	-105.00	-83.04	-73.79
26	0.5000(30)	0.5000(50)	0.5000(200)	1.0000(500)	-108.60	-85.79	-75.88
27	0.5000(30)	0.5000(50)	0.5000(200)	0.5000(250)	-120.00	-94.84	-84.59
28	0.5000(30)	0.5000(50)	0.5000(200)	0.5000(250)	-97.35	-77.45	-69.42
29	0.5000(30)	0.5000(50)	0.5000(200)	0.5000(250)	-107.00	-84.00	-75.00
30	0.5000(30)	0.5000(50)	0.5000(200)	0.5000(250)	-101.55	-80.55	-71.83
31	0.5000(30)	0.5000(50)	0.5000(200)	0.5000(250)	-119.00	-94.50	-82.80

Table 4: Experimental CCCD for the transformed RSM data that are coded between 0 and 1

Exp. Run	x_1	x_2	x_3	x_4	$y_1\ (RSSLTE)$ **(Nano Watt)**	$y_2(RSS3G)$ **(Nano Watt)**	$y_3(RSS2G)$ **(Nano Watt)**
1	0.2500	0.2500	0.2500	0.2500	**0**	0.17	2.15
2	0.7500	0.2500	0.2500	0.2500	0.02	3.98	31.16
3	0.2500	0.7500	0.2500	0.2500	0.02	3.98	31.16
4	0.7500	0.7500	0.2500	0.2500	**0**	0.19	2.19
5	0.2500	0.2500	0.7500	0.2500	0.002	0.58	5.46
6	0.7500	0.2500	0.7500	0.2500	0.01	2.24	15.4
7	0.2500	0.7500	0.7500	0.2500	**0**	0.31	**3.37**
8	0.7500	0.7500	0.7500	0.2500	0.19	19.09	117
9	0.2500	0.2500	0.2500	0.7500	0.013	2.24	15.38
10	0.7500	0.2500	0.2500	0.7500	0.04	7.55	35.2
11	0.2500	0.7500	0.2500	0.7500	0.02	3.98	31.6
12	0.7500	0.7500	0.2500	0.7500	1.26	1.58	525
13	0.2500	0.2500	0.7500	0.7500	0.08	10.14	72.77
14	0.7500	0.2500	0.7500	0.7500	0.02	3.98	31.6
15	0.2500	0.7500	0.7500	0.7500	0.009	1.99	17.58
16	0.7500	0.7500	0.7500	0.7500	0.03	4.67	40.55
17	0.0000	0.5000	0.5000	0.5000	0.003	0.67	6.05
18	1.0000	0.5000	0.5000	0.5000	0.05	**6.92**	44.7
19	0.5000	0.0000	0.5000	0.5000	0.006	1.23	9.5
20	0.5000	1.0000	0.5000	0.5000	0.04	6.99	49.8
21	0.5000	0.5000	0.0000	0.5000	0.007	1.74	16.4
22	0.5000	0.5000	1.0000	0.5000	0.006	1.29	7.78
23	0.5000	0.5000	0.5000	0.5000	0.02	3.98	31.6
24	0.5000	0.5000	0.5000	0.5000	0.02	3.98	31.6
25	0.5000	0.5000	0.5000	0.0000	0.03	4.96	41.8
26	0.5000	0.5000	0.5000	1.0000	0.02	2.64	25.8
27	0.5000	0.5000	0.5000	0.5000	0.001	0.33	3.48
28	0.5000	0.5000	0.5000	0.5000	0.18	1.79	114.28

29	0.5000	0.5000	0.5000	0.5000	0.02	3.98	31.6
30	0.5000	0.5000	0.5000	0.5000	0.07	10	65.5
31	0.5000	0.5000	0.5000	0.5000	0.001	0.35	52.5

Genetic algorithm
Once the data has been modeled, the resulting fitted curve is used for determining the setting of the explanatory variables that optimizes the response based on the signal strength requirement. This task summarizes the aim of the optimization phase of RSM (Mays *et al*., 2001; Johnson and Montgomery, 2009). In this paper, we perform all the optimization tasks using the Genetic Algorithm (GA) optimization toolbox available in Matlab software.

The individual desirability
For multiple response studies that involve m responses, $m > 1$, it is essential to obtain an optimal setting of the explanatory variables that simultaneously optimize all the responses with respect to their individual signal strength requirements (He *et al*., 2012; Sestelo *et al*., 2017; Wan and Birch 2011). The most popular criterion applied in the optimization of multiple responses is the Desirability function.

Based on the production requirement of a response, the desirability function transforms the estimated response, $\hat{y}_p(\boldsymbol{x})$ into a scalar measure, $d_p\left(\hat{y}_p(\boldsymbol{x})\right)$.
For larger-the-better (LTB) response, $d_1(\hat{y}_1(\boldsymbol{x}))$ is given as:

$$d_1(\hat{y}_1(\boldsymbol{x})) = \begin{cases} 0, & \hat{y}_1(\boldsymbol{x}) < L \\ \left\{\frac{\hat{y}_1(x)-L}{T-L}\right\}^{t_1}, & L \le \hat{y}_1(\boldsymbol{x}) \le T, \\ 1, & \hat{y}_1(\boldsymbol{x}) > T, \end{cases} \qquad s.t\ \boldsymbol{x} \epsilon\ \varphi\ , \tag{15}$$

For larger-the-better (LTB) response, $d_2(\hat{y}_2(\boldsymbol{x}))$ is given as:

$$d_2(\hat{y}_2(\boldsymbol{x})) = \begin{cases} 0, & \hat{y}_2(\boldsymbol{x}) < L \\ \left\{\frac{\hat{y}_2(x)-L}{T-L}\right\}^{t_1}, & L \le \hat{y}_2(\boldsymbol{x}) \le T, \\ 1, & \hat{y}_2(\boldsymbol{x}) > T, \end{cases} \qquad s.t\ \boldsymbol{x} \epsilon\ \varphi\ , \tag{16}$$

For larger-the-better (LTB) response, $d_3(\hat{y}_3(\boldsymbol{x}))$ is given as:

$$d_3(\hat{y}_3(\boldsymbol{x})) = \begin{cases} 0, & \hat{y}_3(\boldsymbol{x}) < L \\ \left\{\frac{\hat{y}_3(x)-L}{T-L}\right\}^{t_1}, & L \le \hat{y}_3(\boldsymbol{x}) \le T, \\ 1, & \hat{y}_3(\boldsymbol{x}) > T, \end{cases} \qquad s.t\ \boldsymbol{x} \epsilon\ \varphi\ , \tag{17}$$

In all cases, t_1 is the parameter that controls the shape of the desirability function, enabling the user to accommodate nonlinear desirability functions. However, for RSM data, the values of t_1 is taken to be 1 (Castillo, 2007; He *et al*., 2012).

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

$$D = maximize \sqrt[3]{d_1(\hat{y}_1(\boldsymbol{x}))d_2(\hat{y}_2(\boldsymbol{x}))d_3(\hat{y}_3(\boldsymbol{x}))} \tag{18}$$

RESULTS AND DISCUSSION
In Table 5, the fixed bandwidths for y_1 $(RSSLTE_{FB})$, y_2 $(RSS3G_{FB}$ and y_3 $(RSS2G_{FB})$ were obtained via genetic algorithm tool in Matlab and it is only applicable to local linear regression model, since it accommodates the diagonal weight matrix as given in equation (6).

Table 5: Fixed bandwidths for responses; $\boldsymbol{y_1}$ $(\boldsymbol{RSSLTE_{FB}})$, $\boldsymbol{y_2}$ $(\boldsymbol{RSS3G_{FB}})$ and $\boldsymbol{y_3}$ $(\boldsymbol{RSS2G2_{FB}})$

Response	Model	$\boldsymbol{FIXED\ BANDWIDTH\ (b)}$
y_1 $(RSSLTE_{FB})$	OLS_{LTE}	NOT APPLICABLE
	LLR_{LTEFB}	b =0.1000
y_2 $(RSS3G_{FB})$	OLS_{3G}	NOT APPLICABLE
	LLR_{3GFB}	b =0.2600
y_3 $(RSS2G2_{FB})$	OLS_{2G}	NOT APPLICABLE
	LLR_{2G}	b =0.1000

Table 6: Model goodness-of-fits statistics for received signal strength data

Response	Model	$\boldsymbol{DF}$	$\boldsymbol{PRESS^{**}}$	$\boldsymbol{PRESS}$	$\boldsymbol{SSE}$	$\boldsymbol{MSE}$	$\boldsymbol{R^2}$**(%)**	$\boldsymbol{R^2_{Adj}}$**(%)**
y_1 $(RSSLTE_{FB})$	OLS	16.0000	0.2480	3.9686	0.7078	0.0442	53.5626	12.9299
	LLR_{FB}	7.0000	**0.0858**	**2.7876**	**0.0252**	**0.0036**	**98.3500**	**92.9200**
y_2 $(RSS3G_{FB})$	OLS	16.0000	81.0873	1297.40	275.5207	17.2200	40.9796	-10.6632
	LLR_{FB}	9.9907	**21.0072**	**649.9338**	**77.2716**	**7.7344**	**83.4500**	**50.3000**

$y_3\ (RSS2G_{FB})$	OLS	16.0000	43070	689110	125370	7835.50	51.6268	9.3036
	LLR_{FB}	7.0000	**15144**	**490530**	**7784.50**	**1112.10**	**97.0000**	**87.1300**

The results obtained from Table 6, clearly shows that LLR_{FB} from the respective received signal strength gave the better performance statistic as compared with OLS, for the multi-response problem. For $y_1\ (RSSLTE_{FB})$, $y_2\ (RSS3G_{FB})$, $y_3\ (RSS2G_{FB})$ the LLR for fixed bandwidth outperformed the OLS in terms PRESS**, PRESS, SSE, MSE, R^2 and R^2Adj and gives a better predictive power over OLS.

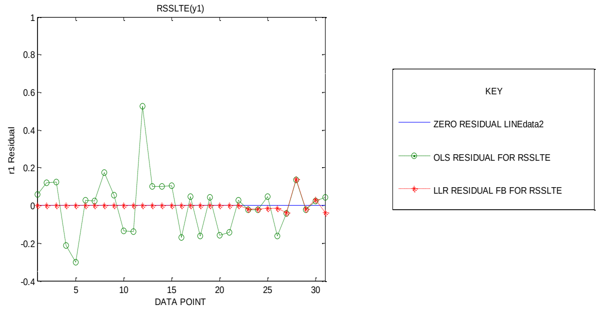

Figure 2: Residual plot for the two regression models LLR FB and OLS for $y_1\ (RSSLTE_{FB})$

In Figure 2, the local linear regression model with fixed bandwidth gave a smaller residual (red line) over OLS with residual line spread away more from the zero residual line. This is a clear indication that LLR with a FB is a better regression model over the OLS.

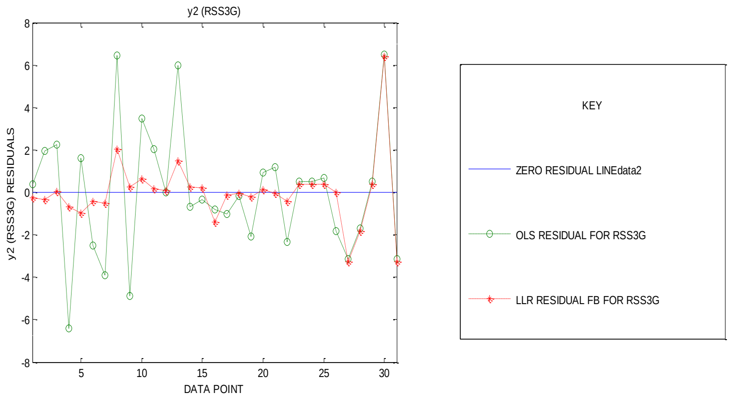

Figure 3: Residual plot for the two regression models LLR FB and OLS for $y_2\ (RSS3g_{FB})$

In Figure 3, the local linear regression model with fixed bandwidth gave a smaller residual (red line) over OLS with residual line spread away more from the zero residual line. This is a justification of result obtained from the goodness-of-fit statistics that LLR with a FB is a better regression model over the OLS.

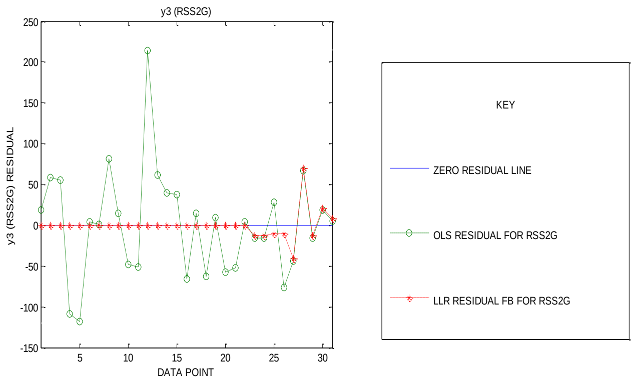

Figure 4: Residual plot for the two regression models LLR FB and OLS for $y_3\ (RSS2g_{FB})$

In Figure 4, the LLR with a local linear regression model with fixed bandwidth (LLR with a FB) gave a smaller residual (red line) over OLS with residual line spread away more from the zero residual line. This is a justification of result obtained from the goodness-of-fit statistics that LLR with a FB is a better regression model over the OLS.

Table 7: Model optimal solution based on the Desirability function for $y_1\,(RSSLTE_{FB})$, $y_2\,(RSS3G_{FB})$, $y_3\,(RSS2G_{FB})$

Model	x_1	x_2	x_3	x_4	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	$d_1(\hat{y}_1)$	$d_2(\hat{y}_2)$	$d_3(\hat{y}_3)$	$D(\%)$
OLS	0.3764	1.0000	0.7155	0.5000	-0.0426	6.2472	2.7043	0.0000	1.0000	1.0000	0.0000
LLR_{FB}	**0.8394**	**0.5549**	**0.8791**	**0.9946**	0.0054	4.0983	47.2905	1.0000	1.0000	1.0000	**100.00**

From Table 7, LLR_{FB} provides the best received signal strength for LTE, 3G and 2G technologies with optimum setting of the factors; atmosphere temperature (TEMP), Relative Humidity (RH), Air Quality Index (AQI) and Distance from Base Station (DFBS) over OLS and the four settings of the factors were found 0.8394 (40°C), 0.5549 (37%), 0.8791 (150 m) and 0.9946 (150 m) respectively to give the best process satisfaction for the received signal strength. Whereas, the optimum received signal strength were found to be $\hat{y}_1 = 0.0054$ and its equivalent are 0.02 (Nano Watt) and -107(dBm) for $y_1\,(RSSLTE_{FB})$; $\hat{y}_2$ =4.0983 and its equivalent are 3.98 (Nano Watt) and -84(dBm) for $y_2\,(RSS3G_{FB})$ and $\hat{y}_3 = 47.2905$ and its equivalent are 31.16 (Nano Watt) and -75(dBm) for $y_3\,(RSS2G_{FB})$ communication technologies respectively.

CONCLUSION

In this study, we presented a CCCD in other to capture rotatability and curvature in the data, a LLR_{FB} for adequate fitting of the data, and lastly, to find optimum settings of the factors that optimizes $y_1\,(RSSLTE_{FB})$, $y_2\,(RSS3G_{FB})$ and $y_3\,(RSS2G_{FB})$ respectively. The performance statistics carried out is a clear indication that the LLR_{FB} outperformed the OLS for LTE with (PRESS**= 0.0858, PRESS = 2.7876, SSE = 0.0252, MSE = 0.0036, R^2= 98.35% and R^2Adj = 92.92%); 3G with (PRESS**= 21.0072, PRESS = 649.9338, SSE = 77.2716, MSE = 7.7344, R^2= 83.45% and R^2Adj = 50.30%) and 2G with (PRESS**= 15144, PRESS = 490530, SSE = 7784.50, MSE = 1112.10, R^2= 97.00% and R^2Adj = 87.13%) for the three communication technologies and also provided minimum residual plots for their respective network. The optimization results show that the optimum received signal strength were found to be 0.02 (Nano Watt) and (**-107** against existing -77.92) (dBm) for $y_1\,(RSSLTE_{FB})$; 3.98 (Nano Watt) and (**-84** against existing -60.03) (dBm) for $y_2\,(RSS3G_{FB})$ and 31.16 (Nano Watt) and (**-75** against existing 58.13) (dBm) for $y_3\,(RSS2G_{FB})$ communication technologies respectively.

REFERENCES

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.*

Bezerra, M. A., Santelli, R. E., Oliveira, E. P., Villar, L. S. and Escaleira, L. A. (2008). Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta,* ***76****, 965-977*.

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

Choudhary, S., Sharma, A. and Srivastava, K. (2021). Modeling and Optimization of Mobile Signal Strength in Challenging Atmospheric Conditions. Research Square, *DOI:https://doi.org/10.21203/rs.3.rs-295295/01*

Dutka, M., Ditaranto, M. and Lovas, T. (2015). Application of a Central Composite Design for the study of NO_X Emission Performance of Low NO_X Burner. *Energies,* ***8****, 3606-3627*.

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****, 388 – 404.*

Eguasa O. (2020). Adaptive Nonparametric Regression Models for Response Surface Methodology (RSM). PhD Thesis, Department of Mathematics, University of Benin, Benin City, Nigeria.

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. and Gijbels I. (1995). Data-driven bandwidth selection in local polynomial fitting: A variable bandwidth and spatial adaptation. Journal of the Royal Statistical Society, Series b, *371 – 394.*

Hardle W., Muller M., Sperlich S., and Werwatz A. (2005). Nonparametric and Semiparametric Models: An Introduction. Berlin: Springer-Verlag.

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

Johnson R. T., and Montgomery D. C. (2009). Choice of second-order response surface designs for logistics and Poisson regression models, Int. Journal of Experimental Design and Process Optimization, ***1****, 2 – 23.*

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

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., (2009). Introduction to statistical quality control. 7th edition, John Wiley & Sons, New York.

Myers R., Montgomery D. C., and Anderson-Cook C. M. (2009). Response Surface Methodology: Process and Product Optimization Using Designed Experiments Wiley.

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.*

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

Peasura, P. (2015). Application of response surface methodology for modeling of posteld heat treatment process in a pressure vessel steel. *The Scientific World Journal, http.//dx.doi.org/10.1155/2015/318475.*

Sestelo M., Villanueva N. M., Meira-Machado L., and Roca-Pardinas J.,(2017). An R package for Nonparametric estimation and inference in life sciences, Journal of Statistical Software, ***82****, 1- 27.*

Sivarao, I., Anand, T. J. S. and Shukor, A. (2010). RSM Based Modeling for Surface Roughness Prediction in Laser Machining. *International Journal of Engineering and Technology IJET-IJENS,* ***10*** *(4), 26 – 32*.

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.*