FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 6 No. 6, December, 2022, pp 262 - 270
DOI: https://doi.org/10.33003/fjs-2022-0606-1180

MULTI-RESPONSE OPTIMIZATION OF METFORMIN -HYDROCHLORIDE DRUG VIA EXPERIMENTAL DESIGN TECHNIQUES APPLIED ON THE ADAPTIVE LOCAL LINEAR REGRESSION MODEL FOR RESPONSE SURFACE METHODOLOGY

***Eguasa O. and Eguasa M. E.**

Department of Physical Sciences, Benson Idahosa University, Benin City, Edo State

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

ABSTRACT

The challenge encountered by industrial statisticians and engineers during statistical process control in other to build-up and formulates pharmaceutical optimization of drugs have prompted several choices of experimental design tool and regression models. The most commonly applied regression model is the second-order polynomial which may perform poorly due to model misspecification. In this paper, we present two experimental design methods namely; the Full Factorial Design (FFD) and the Circumscribed Central Composite Design (CCCD) applied to an existing Adaptive Local Linear Regression (LLR_{AB}) model to ameliorate the problem of boundary bias for a multi-response problem. The FFD do not make reference to the star points and as such could not address variability in the data, hence we also applied the CCCD to accommodate the star points in order to maintain rotatability and curvature in the data. In the application, we minimized Metformin Hydrochloride (Met-HCL) drug usage via FFD on LLR_{AB} and CCCD on LLR_{AB} and the results from the goodness-of-fit statistics, residual plots and optimization were obtained and analyzed. The LLR_{AB} that utilized CCCD outperformed LLR_{AB} that uses FFD in terms of the goodness-of-fit statistics, minimum residual plots as it relates to the zero residual line and optimization of Met-HCL for Response Surface Methodology (RSM) data.

Keywords: Adaptive local linear regression model, Circumscribed central composite design, Full factorial design, Metformin hydrochloride

INTRODUCTION

Recently, experimental design or design of experiments (DOE) and statistical process control have been modified by pharmaceutical industries, electric power generation companies, nuclear power plants, manufacturing industries, petroleum refineries, petrochemical plants, natural gas processing plants, food processing plants and several other industrial facilities with the aim to minimize cost, time and to optimized processes and products during process, product design and development. This suggest and gives a clear guild to the industries in other to scale-up for visibility studies alongside with cost reduction and better choice of management policies during production process. Response Surface Methodology (RSM) is an essential industrial statistical and engineering tool employed in the process and product development through three main stages namely; experimental design stage, modeling stage of the fitted regression model and in the optimization stage. It gives enlightenment to the industries in several ways such as pharmaceutical product design, process development, quality, manufacturing engineering and operations by improving their performances, design and formulations, and also in the optimization of the optimum control factors as well as the final products (Shubhasis *et al*., 2014; Eguasa *et al.*, 2022).

In other to resolve the challenge encountered during the continuous release dosage procedure, it is of necessity to find appropriate experimental design, model building and optimization of pharmaceutical formulation with a suitable conclusion rate in a record time and minimum number of experimental runs. Most commonly utilized model in this scenario is the second order polynomial model which may be inadequate due to model misspecification as given in (Emami *et al.*, 2008; Shubhasis *et al*., 2014; Eguasa, 2020). The optimization technique guarantee the use of genetic algorithm that employs a systematic design to minimize the number of trials, analysis of the response surfaces in order to aid understanding of the effect of the operating factors in obtaining suitable operating settings for the purpose of achieving the target goals either by a constrained optimization, two sided optimization, smaller the better optimization and lager the better optimization for a single and a multi-response optimization problem in practical preparation (Dan and Pal, 2013; Eguasa *et al.*, 2022; Akhideno and Eguasa, 2022).

The purpose of this study is to minimized Metformin Hydrochloride (Met-HCL) drug usage as the multi-response constrained problem with two operating factors such as Hydroxypropyl Methyl Cellulose K-15M (HPMC K-15M) and Polyvinyl Pyrrolidone K-30 (PVP K-30). In other to achieve this, two experimental design tool such as the Full Factorial Design (FFD) and the Circumscribed Central Composite Design (CCCD) were used to obtain the coded levels and thereafter be transformed to RSM data. Lastly, the data generated from FFD and CCCD were used to fit the Adaptive Local Linear Regression (LLR_{AB}) model for the multi-response problems.

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 constrained to a user specified form unlike the parametric counterpart. In spite of its flexibility, nonparametric regression models are confronted in a study such as RSM due to the peculiarities 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; Eguasa *et al*., 2020). 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:

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

where $\boldsymbol{y}$ is a $n \times 1$ vector of response, $\mathbf{X}$ is a n × 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:

$$\hat{\boldsymbol{y}}^{(OLS)} = \begin{bmatrix} \boldsymbol{h_1^{(OLS)}} \\ \boldsymbol{h_2^{(OLS)}} \\ \vdots \\ \boldsymbol{h_n^{(OLS)}} \end{bmatrix} \boldsymbol{y} = \boldsymbol{H^{(OLS)}y}, \quad (2)$$

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

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:

$$y_i = \beta_0 + \sum_{j=1}^{k}\beta_j x_{ij} + \sum_{j=1}^{k}\beta_{jj}x_{ij}^2 + \sum_{j=1}^{k-1}\sum_{r=j+1}^{k}\beta_{jr}x_{ij}x_{ir} + \varepsilon_i, i = 1,2,\dots,n;\ r = j+1, j+2,\dots,k \quad (3)$$

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

As presented in the literature, the two explanatory variables are Hydroxypropyl Methyl Cellulose K-15M (HPMC K-15M) and Polyvinyl Pyrrolidone K-30 (PVP K-30) and the response variable is the Met-HCL. All the experimental runs for these chemicals used were gotten from the same batch (Shubhasis *et al*., 2014).

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

The Local Linear Regression (LLR)

The LLR model is a nonparametric regression form 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:

$$\hat{y}_i^{(LLR)} = \tilde{\boldsymbol{x}}_i(\tilde{\boldsymbol{X}}^T\boldsymbol{W}_i\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^T\boldsymbol{W}_i\boldsymbol{y} = \boldsymbol{h_i^{(LLR)}y}, \quad (4)$$

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

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

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:

$$\boldsymbol{W_i} = \begin{bmatrix} w_{1i} & 0 & \cdots & 0 \\ 0 & w_{2i} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & w_{ni} \end{bmatrix}_{n\times n} \quad (5)$$

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

$$w_{11} = \prod_{j=1}^{k} K\left(\frac{x_{ij}-x_{1j}}{b}\right) \Big/ \sum_{i=1}^{n}\prod_{j=1}^{k} K\left(\frac{x_{ij}-x_{1j}}{b}\right),$$
$$i = 1,2,\dots,n, j = 1,2,\dots,k, \quad (6)$$

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, $0 < b \leq 1$, $i = 1,2,\dots,n$, $j = 1,2,\dots,k$, is the fixed bandwidth (smoothing parameter) (Myers *et al*., 2009; Eguasa, 2020).

Experimental design

In RSM, the factors are usually more than one. Hence, if the number of factors is too large, it may directly affect the response 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).

The choice of suitable levels to be studied for the explanatory variables is also vital as it can affect model correctness.

The Experimental Design stage permits an appropriate design that can offer adequate and substantial 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 (Dan *et al*., 2014; Shubhasis *et al*., 2014)

Factors or Input parameters	-1(Low)	0(Medium)	1(High)
X1= HPMC K 15M	100	300	500
X2= PVP K30	50	75	100

Table 2: Experimental range for constrained multi-responses (Dan *et al.*, 2014)

Response(s)	Experimental range for responses	% by range for responses (Met-HCL)
y_1	Not more than 30 percentage	$y_1 \leq 30$
y_2	Not more than 40 percentage	$y_2 \leq 40$
y_3	Not more than 60 percentage	$y_3 \leq 60$
y_4	Not more than 70 percentage	$y_4 \leq 70$
y_5	Not more than 80 percentage	$y_5 \leq 80$

Table 3: Coded stages and range for the design of experiments

Factors or Input parameters	Coded Levels				
	-2(-α)	**-1(Low)**	**0(Medium)**	**1(High)**	**2(+α)**
X1= HPMC K 15M	50	100	300	500	550
X2= PVP K30	25	50	75	100	125

The central composite design

A Central Composite Design permits for modeling of the second-order regression model in a given response that is often used for process optimization (Sivarao *et al*., 2010; Eguasa, 2020; Akhideno and Eguasa, 2022). The three types of CCD are centered 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 permits for the estimation of curvature and the values of star points maintain rotatability which in turn hinge on the factorial point of the design (Dutka *et al*., 2015).

The circumscribed CCD comprises of 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).

In this study, the CCCD has been employed 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:

$$CCCD = 2^k + 2k + k_c \quad (7)$$

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 $k = 2$ and $k_c = 1$ which from equation (7) sum to 9 experimental runs.

Table 4: Factor combinations as per the chosen experimental design (Dan *et al*., 2014)

Exptal. Run	X1= HPMC K 15M	X2= PVP K30
1	-1	-1
2	-1	0
3	-1	1
4	0	-1
5	0	0
6	0	1
7	1	-1
8	1	0
9	1	1

Table 5: Experimental coded level for RSM data

Exptal. Run	x_1= HPMC K 15M	x_2= PVP K30
1	-1	-1
2	1	-1
3	-1	1
4	1	1
5	-1.4142	0
6	1.4142	0
7	0	-1.4142
8	0	1.4142
9	0	0

Hereafter, the 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. 31 points 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:

$$x_{NEW} = \frac{Min(x_{OLD}) - x_0}{(Min(x_{OLD}) - Max(x_{OLD}))} \quad (8)$$

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 (8).

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

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

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

$$Explanatory\ variable\ x_1 :\ x_{21} = \frac{-2 - (1)}{((-2) - (2))} = 0.0000$$

$$Explanatory\ variable\ x_2 :\ x_{22} = \frac{-2 - (-1)}{((-2) - (2))} = 0.5000$$

where $x_1 =$ HPMC K 15M, $x_2 =$ PVP K30

Table 6: Factor combinations as per the chosen experimental design (Dan *et al.*, 2014)

Exp. Run	x_1	x_2	y_1	y_2	y_3	y_4	y_5
1	0.0000	0.0000	30	35	55	70	75
2	0.0000	0.5000	25	40	60	65	76
3	0.0000	1.0000	27	32	60	68	80
4	0.5000	0.0000	28	36	45	60	80
5	0.5000	0.5000	30	38	58	69	74
6	0.5000	1.0000	26	34	52	70	78
7	1.0000	0.0000	28	39	48	62	72
8	1.0000	0.5000	22	40	50	64	79
9	1.0000	1.0000	24	35	57	66	80

Table 7: Experimental CCCD for the transformed RSM data that are coded btw 0 and 1

Exp. Run	x_1	x_2	y_1	y_2	y_3	y_4	y_5
1	0.1464	0.1464	30	35	55	70	75
2	0.8536	0.1464	25	40	60	65	76
3	0.1464	0.8536	27	32	60	68	80
4	0.8536	0.8536	28	36	45	60	80
5	0.0000	0.5000	30	38	58	69	74
6	1.0000	0.5000	26	34	52	70	78
7	0.5000	0.0000	28	39	48	62	72
8	0.5000	1.0000	22	40	50	64	79
9	0.5000	0.5000	24	35	57	66	80

Genetic algorithm

Once the data has been modeled, the resulting fitted curve is used in obtaining the setting of the explanatory variables that optimizes the response based on the Met-HCL requirement. This apparently summarizes the aim of the optimization stage of RSM (Mays *et al*., 2001; Johnson and Montgomery, 2009). In this paper, we accomplished all the optimization tasks using the Genetic Algorithm (GA) optimization toolbox available in Matlab software.

The individual desirability

In the multiple-response studies that include m responses, $m > 1$, it is vital to obtain an optimal setting of the explanatory variables that would simultaneously optimize all the responses with respect to their individual Met-HCL requirements (Wan and Birch 2011; He *et al*., 2012; Shubhasis *et al*., 2014; Sestelo *et al*., 2017). The most common criterion applied in the optimization of multiple responses is the Desirability function.

Based on the process 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 a smaller-the-better (STB) response, $d_\varrho\left(\hat{y}_\varrho(\boldsymbol{x})\right), \varrho = 3$ is given as:

$$d_1\left(\hat{y}_1(\boldsymbol{x})\right) = \begin{cases} 1, & \hat{y}_1(\boldsymbol{x}) < 28, \\ \left\{\frac{U-\hat{y}_1(x)}{29-28}\right\}^{t_2}, & 28 \le \hat{y}_1(\boldsymbol{x}) \le 29, \quad s.t\ \boldsymbol{x}\epsilon\ \varphi\ , \\ 0, & \hat{y}_1(\boldsymbol{x}) > 29, \end{cases} \tag{9}$$

where $\rho = 28$ and $U = 29$ are the minimum acceptable value and upper limit, respectively, of the ϱ^{th} response. However, for RSM data, the parameters values of $t_1\ and\ t_2$ are weights taken to be 1 for linearity (Eguasa *et al*., 2022).

For a smaller-the-better (STB) response, $d_\varrho\left(\hat{y}_\varrho(\boldsymbol{x})\right), \varrho = 3$ is given as:

$$d_2\left(\hat{y}_2(\boldsymbol{x})\right) = \begin{cases} 1, & \hat{y}_2(\boldsymbol{x}) < 36.5, \\ \left\{\frac{U-\hat{y}_2(x)}{37-36.5}\right\}^{t_2}, & 36.5 \le \hat{y}_2(\boldsymbol{x}) \le 37, \quad s.t\ \boldsymbol{x}\epsilon\ \varphi\ , \\ 0, & \hat{y}_2(\boldsymbol{x}) > 37, \end{cases} \tag{10}$$

where $\rho = 36.5$ and $U = 37$ are the minimum acceptable value and upper limit, respectively, of the ϱ^{th} response.

For a smaller-the-better (STB) response, $d_\varrho\left(\hat{y}_\varrho(\boldsymbol{x})\right), \varrho = 3$ is given as:

$$d_3\left(\hat{y}_3(\boldsymbol{x})\right) = \begin{cases} 1, & \hat{y}_3(\boldsymbol{x}) < 58, \\ \left\{\frac{U-\hat{y}_3(x)}{59-58}\right\}^{t_2}, & 58 \le \hat{y}_3(\boldsymbol{x}) \le 59, \quad s.t\ \boldsymbol{x}\epsilon\ \varphi\ , \\ 0, & \hat{y}_3(\boldsymbol{x}) > 59, \end{cases} \tag{11}$$

where $\rho = 58$ and $U = 59$ are the minimum acceptable value and upper limit, respectively, of the ϱ^{th} response.

For a smaller-the-better (STB) response, $d_\varrho\left(\hat{y}_\varrho(\boldsymbol{x})\right), \varrho = 3$ is given as:

$$d_4(\hat{y}_4(\boldsymbol{x})) = \begin{cases} 1, & \hat{y}_4(\boldsymbol{x}) < 65, \\ \left\{\frac{U-\hat{y}_4(\boldsymbol{x})}{67-65}\right\}^{t_2}, & 65 \le \hat{y}_4(\boldsymbol{x}) \le 67, \quad s.t\ \boldsymbol{x}\epsilon\ \varphi\ , \\ 0, & \hat{y}_4(\boldsymbol{x}) > 67, \end{cases} \tag{12}$$

where $\rho = 65$ and $U = 67$ are the minimum acceptable value and upper limit, respectively, of the ϱ^{th} response.

For a smaller-the-better (STB) response, $d_\varrho\left(\hat{y}_\varrho(\boldsymbol{x})\right), \varrho = 3$ is given as:

$$d_5(\hat{y}_5(\boldsymbol{x})) = \begin{cases} 1, & \hat{y}_5(\boldsymbol{x}) < 79.5, \\ \left\{\frac{U-\hat{y}_5(\boldsymbol{x})}{79.9-79.5}\right\}^{t_2}, & 79.5 \le \hat{y}_5(\boldsymbol{x}) \le 79.9, \quad s.t\ \boldsymbol{x}\epsilon\ \varphi\ , \\ 0, & \hat{y}_5(\boldsymbol{x}) > 79.9, \end{cases} \tag{13}$$

where $\rho = 79.5$ and $U = 79.9$ are the minimum acceptable value and upper limit, respectively, of the ϱ^{th} response.

In all cases, t_2 is the parameter that controls the shape of the desirability function, allowing the user to accommodate nonlinear desirability functions. However, for RSM data, the values of t_2 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:

$$D = minimize\sqrt[5]{d_1(\hat{y}_1(\boldsymbol{x}))d_2(\hat{y}_2(\boldsymbol{x}))d_3(\hat{y}_3(\boldsymbol{x}))d_4(\hat{y}_4(\boldsymbol{x}))d_5(\hat{y}_5(\boldsymbol{x}))} \tag{14}$$

RESULTS AND DISCUSSION

In Tables 8 to 12 gives the constraints of Met-HCL drug dosage and the adaptive bandwidths for y_1 (≤ 30), y_2 (≤ 40), $y_3(\le 60)$, $y_4(\le 70)$ and y_5 (≤ 80) 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 (7).

Table 8: Adaptive bandwidths for the $FFD\ LLR_{AB}$ and $CCCD\ LLR_{AB}$

i	$CCCD\ LLR_{AB}\ for\ y_1$		FFD LLR_{AB} for y_1	
	b_{i1}	b_{i2}	b_{i1}	b_{i2}
1	0.1314	0.8130	0.2028	0.3441
2	0.0611	0.8130	0.2028	0.0001
3	0.1314	0.2581	0.2028	0.3665
4	0.0611	0.2581	0.0008	0.3441
5	0.1493	0.4968	0.0008	0.0001
6	0.0498	0.4968	0.0008	0.3665
7	0.0929	0.9666	0.2562	0.3441
8	0.0929	0.1819	0.2562	0.0001
9	0.0929	0.4968	0.2562	0.3665

Table 9: Adaptive bandwidths for the $FFD\ LLR_{AB}$ and $CCCD\ LLR_{AB}$

i	$CCCD\ LLR_{AB}\ for\ y_2$		FFD $LLR_{AB}\ for\ y_2$	
	b_{i1}	b_{i2}	b_{i1}	b_{i2}
1	0.9310	0.1663	0.3527	0.1824
2	0.6612	0.1663	0.3527	0.0174
3	0.9310	0.0281	0.3527	0.4777
4	0.6612	0.0281	0.0379	0.1824
5	0.9926	0.0828	0.0379	0.0174
6	0.6112	0.0828	0.0379	0.4777
7	0.7904	0.2093	0.9669	0.1824
8	0.7904	0.0139	0.9669	0.0174
9	0.7904	0.0828	0.9669	0.4777

Table 10: Adaptive bandwidths for the $FFD\ LLR_{AB}$ and $CCCD\ LLR_{AB}$

i	$CCCD\ LLR_{AB}\ for\ y_3$		FFD $LLR_{AB}\ for\ y_3$	
	b_{i1}	b_{i2}	b_{i1}	b_{i2}
1	0.1151	0.0610	0.5770	0.2228
2	0.0908	0.0610	0.5770	0.0114
3	0.1151	0.5738	0.5770	0.4700
4	0.0908	0.5738	0.0851	0.2228
5	0.2226	0.0651	0.0851	0.0114
6	0.1884	0.0651	0.0851	0.4700
7	0.0004	0.2070	0.0311	0.2228

8	0.0004	0.9321	0.0311	0.0114
9	0.0004	0.0651	0.0311	0.4700

Table 11: Adaptive bandwidths for the $FFD\ LLR_{AB}$ and $CCCD\ LLR_{AB}$

i	$CCCD\ LLR_{AB}\ for\ y_4$		FFD LLR_{AB} for y_4	
	b_{i1}	b_{i2}	b_{i1}	b_{i2}
1	0.3043	0.1535	0.3639	0.9749
2	0.0065	0.1535	0.3639	0.7320
3	0.3043	0.4661	0.3639	0.5239
4	0.0065	0.4661	0.0387	0.9749
5	0.4214	0.0212	0.0387	0.7320
6	0.0003	0.0212	0.0387	0.5239
7	0.0999	0.3773	0.9933	0.9749
8	0.0999	0.8193	0.9933	0.7320
9	0.0999	0.0212	0.9933	0.5239

Table 12: Adaptive bandwidths for the $FFD\ LLR_{AB}$ and $CCCD\ LLR_{AB}$

i	$CCCD\ LLR_{AB}\ for\ y_5$		FFD LLR_{AB} for y_5	
	b_{i1}	b_{i2}	b_{i1}	b_{i2}
1	0.0568	0.1315	0.6430	0.1751
2	0.5422	0.1315	0.6430	0.0354
3	0.0568	0.1869	0.6430	0.6315
4	0.5422	0.1869	0.0235	0.1751
5	0.1936	0.0012	0.0235	0.0354
6	0.8800	0.0012	0.0235	0.6315
7	0.0620	0.2780	0.2450	0.1751
8	0.0620	0.3563	0.2450	0.0354
9	0.0620	0.0012	0.2450	0.6315

Table 13: Model goodness-of-fits statistics for $FFD\ LLR_{AB}$ and $CCCD\ LLR_{AB}$

Response	Model	DF	$PRESS^{**}$	$PRESS$	SSE	MSE	$R^2(\%)$	$R^2_{Adj}(\%)$
y_1	$FFD\ LLR_{AB}$	0.6651	6.7861	43.1104	1.7615	2.6485	96.96	63.47
	$CCCD\ LLR_{AB}$	0.0694	**2.2450**	**13.6193**	**0.0179**	**0.2582**	**99.97**	**96.44**
y_2	$FFD\ LLR_{AB}$	1.0543	**3.6278**	**25.3250**	0.5215	**0.4946**	99.19	**93.84**
	$CCCD\ LLR_{AB}$	0.7071	12.8319	85.5988	**0.3568**	0.5046	**99.44**	93.71
y_3	$FFD\ \ LLR_{AB}$	0.1603	**18.3169**	**112.3068**	0.4574	2.8536	99.81	90.28
	$CCCD\ LLR_{AB}$	0.0145	55.7651	335.3824	**0.0096**	**0.6620**	**99.99**	**97.75**
y_4	$FFD\ LLR_{AB}$	3.0229	16.6531	119.2933	17.9236	5.9292	82.43	53.50
	$CCCD\ LLR_{AB}$	0.3614	**12.8116**	**80.4416**	**1.1435**	**3.1638**	**98.88**	**75.19**
y_5	$FFD\ LLR_{AB}$	0.3366	12.7527	80.3087	0.3311	**0.9839**	99.53	**88.90**
	$CCCD\ LLR_{AB}$	0.1137	**7.1742**	**43.3214**	**0.2594**	2.2814	**99.63**	74.25

The results obtained from Table 13 clearly shows that $CCCD\ LLR_{AB}$ from the respective Met-HCL gave the better performance statistics as compared with drug dosage over $FFD\ LLR_{AB}$ in twenty two cells as against eight cells for the multi-response Met-HCL problem. The bolded cells indicates a better performance over cells that are not bold and obviously gives a better predictive power over the $FFD\ LLR_{AB}$.

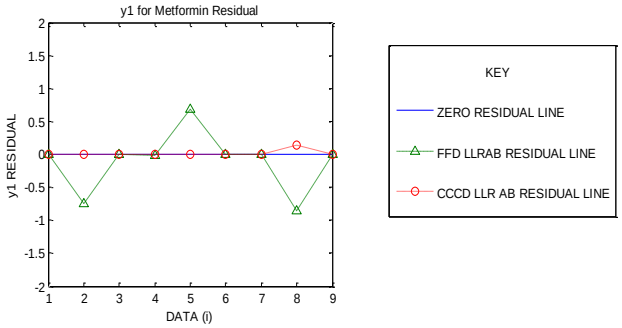

Figure 1: Residual plot for the two regression models $FFD\ LLR_{AB}$ and $CCCD\ LLR_{AB}$ y_1

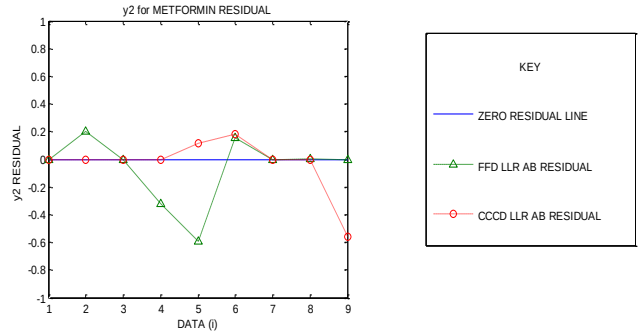

Figure 2: Residual plot for the two regression models $FFD\ LLR_{AB}$ and $CCCD\ LLR_{AB}\ y_2$

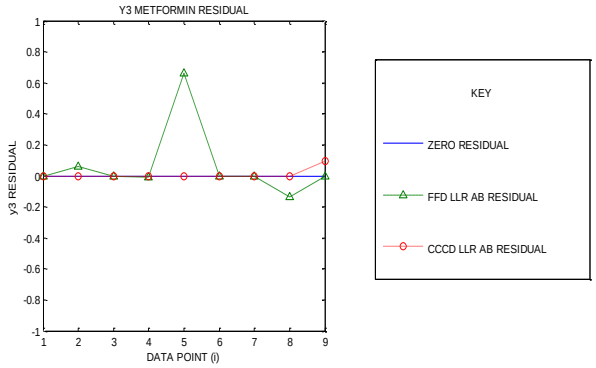

Figure 3: Residual plot for the two regression models $FFD\ LLR_{AB}$ and $CCCD\ LLR_{AB}\ y_3$

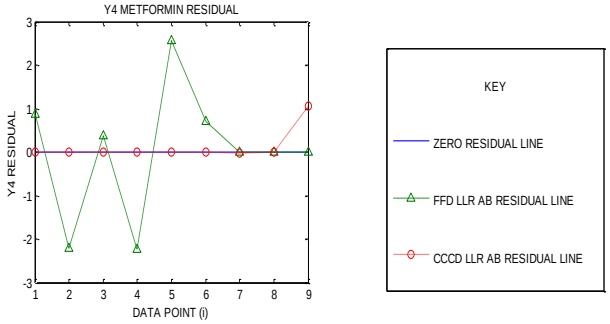

Figure 4: Residual plot for the two regression models $FFD\ LLR_{AB}$ and $CCCD\ LLR_{AB}\ y_4$

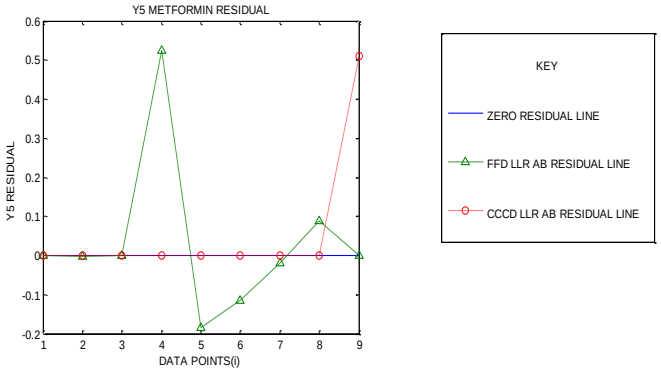

Figure 5: Residual plot for the two regression models $FFD\ LLR_{AB}$ and $CCCD\ LLR_{AB}\ y_5$

In Figures 1-5, the $CCCD\ LLR_{AB}$ gave a smaller residual (red line) over $FFD\ LLR_{AB}$ (green line) 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 $CCCD\ LLR_{AB}$ is a better regression model over the $FFD\ LLR_{AB}$.

Table 14: Model optimal solution based on the multi-response Desirability function

Model	x_1	x_2	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	y_4	y_5	$d_1(\hat{y}_1)$	$d_2(\hat{y}_2)$	$d_3(\hat{y}_3)$	$d_4(\hat{y}_4)$	$d_5(\hat{y}_5)$	$D(\%)$
FFD LLR_{AB}	0.6574	0.4359	27.0505	36.2059	46.3372	65.0155	76.5808	1.0000	1.0000	1.0000	0.9922	1.0000	99.84
$CCCD$ LLR_{AB}	0.5574	0.4359	27.5138	35.7426	55.5432	63.1012	79.1100	1.0000	1.0000	1.0000	1.0000	1.0000	**100.00**

From Table 14, $CCCD\ LLR_{AB}$ provides a better multi-response Met-HCL over $FFD\ LLR_{AB}$ in terms of overall desirability for the respective factors $x_1 =$ HPMC K 15M, $x_2 =$ PVP K30 . Obviously, $CCCD\ LLR_{AB}$ gave a better process requirement with 100% desirability and with operating factors $x_1(\text{HPMC K 15M}) = 0.5574$, $x_2(\text{ PVP K30}) = 0.4359$ with the best choice of Met-HCL drug dosage.

CONCLUSION

The minimization of Metformin Hydrochloride (Met-HCL) drug usage for individual and the overall desirability functions were analyzed for a multi-response constrained problem with two operating factors such as Hydroxypropyl Methyl Cellulose K-15M (HPMC K-15M) and Polyvinyl Pyrrolidone K-30 (PVP K-30). The results show that $CCCD\ LLR_{AB}$ from the respective Met-HCL performed better in terms of goodness statistics as compared with drug dosage over $FFD\ LLR_{AB}$ in twenty two cells as against eight cells for the multi-response Met-HCL problem. The bolded cells indicates a better performance over cells that are not in bold and apparently gives a better predictive power over the $FFD\ LLR_{AB}$. Thus, for the residual plots, it is obvious that the $CCCD\ LLR_{AB}$ offer a smaller residual (red line) over $FFD\ LLR_{AB}$ (green line) with residual line spread away more from the zero residual line. Lastly, the $CCCD\ LLR_{AB}$ provides a better multi-response Met-HCL over $FFD\ LLR_{AB}$ in terms of overall desirability for the respective factors $x_1 =$ HPMC K 15M, $x_2 =$ PVP K30 . Obviously, $CCCD\ LLR_{AB}$ gave a better process requirement with 100% desirability with settings of the factors $x_1(\text{HPMC K 15M}) = 0.5574$ and $x_2(\text{ PVP K30}) = 0.4359$.

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.

Dan S., Dan N. and Pal T. K. (2014). Application of response surface methodology (RSM) in statistical optimization and pharmaceutical characterization of a matrix tablets formulation using Metformin HCL as a model drug. Innoriginal International Journal of Sciences, **1**, 1-6.

Dan S., Pal T. K. (2013). Sustained Release Matrix Tablet of Diltiazem Hydrochloride: Formulation Development, in-vitro Evaluation and Statistical Optimization by Response Surface Methodology (RSM). Electronic Journal of Biosciences, **1**, 06-22.

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.

Eguasa O. (2020). Adaptive Nonparametric Regression Models for Response Surface Methodology (RSM). Dissertation, 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. https//doi*:10.1080/02664763.2022.2026895*

Akhideno I. O. and Eguasa O. (2022). An Adaptive Local Linear Regression Method for Mobile Signal Strength with Application to Response Surface Methodology. FUDMA Journal of Sciences (FJS), **6** (5), 41 – 49.

Emami J., Taj A. M. and Ahmadi F. (2008). Preparation and In Vitro Evaluation of Sustained-Release Matrix Tablets of Flutamide Using Synthetic and Naturally Occurring Polymers. Iranian Journal of Pharmaceutical Research, 247-257.

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. A. (2014). 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.

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.