Review Article

Indian Journal of Science and Research

Eguasa O. & E Eguasa M. E., Ind. J. Sci. Res. 2023, 3(3), 120-128

ISSN 2583 – 2913

AN ADAPTIVE LOCAL LINEAR REGRESSION MODEL APPLIED IN THE MULTI-RESPONSE OPTIMIZATION OF OIL-IN-WATER EMULSION FOR RESPONSE SURFACE METHODOLOGY

Eguasa O. [*1] and E Eguasa M. E.[2]

Abstract: Response surface methodology (RSM) data was analyzed using the ultrafiltration of oil-in-water emulsion process via the design of experiment (DOE) phase, modeling phase with the choice of appropriate regression model and the optimization phase such that the process is optimized. In the literature, the quadratic regression model (QM) performed better in the region where the data exhibits quadratic trend. However, the QM becomes infamous at the boundaries of the data due to boundary bias problem. The drawback of the QM prompted the use of the ordinary least squares (OLS) regression model and an adaptive local linear regression (LLR_{AB}) model with the intention to address the drawback of QM and hence improve on the goodness-of-fit statistics, zero residual line and the optimization results with a clear justification using the 3D surface plots. In the application, the LLR_{AB} outperformed the OLS and QM in terms of goodness-of-fit statistics, minimized residual errors and optimization requirement with over desirability of 62% as against 56% and 55% for OLS and QM respectively.

Keywords: Adaptive local linear regression model; Chemical oxygen demand; Permeate flux (PF); Transmembrane pressure; Turbidity removal; Velocity

Introduction*:* Oil in water is a product of wastewaters from many industrial facilities such petrochemical industries, natural gas processing plants, nuclear power plants, food processing and other agricultural facilities. These wastewaters are a multifaceted mixture of water, oil, and condiments, such as emulsifiers, corrosion inhibitors, antifoaming agents, and biocides. To recycle this wastewater or to return it to natural waterways, this wastewater must be treated (E. L. Walker *et al*, 2002; C. M. Anderson-Cook *et al*, 2005; M. Sadeghian *et al*, 2014).

According to (C. M. Anderson-Cook *et al*, 2005; R. Myers *et al*, 2009), explained that there are some published articles on RSM, for which the origin of this statistical tool was actually laid by (G. E. P. Box *et al*, 1951). Afterwards, other noteworthy contributors for the improvement of RSM include (G. E. P. Box *et al*, 1957; G. E. P. Box *et al*, 1959) and later paper published by (W. J. Hill *et al*, 1966) discusses the practical applications of RSM in the area of chemical and processing industries. More so, the applications of RSM cut across various fields of studies (C. M. Anderson-Cook *et al*, 2005; R. Myers *et al*, 2009).

The goal of this paper is to find and improve the operating factors that would simultaneously optimize the multi-response problems via RSM data for permeate flux (L/m^{2}h), turbidity removal (%) and chemical oxygen demand (COD) removal (%) with the data from the literature (M. Sadeghian *et al*, 2014).

Response Surface Methodology: According to (R. Myers *et al*, 2009; D. Gramato *et al*, 2014) [9], defined RSM as statistical procedure applied by engineers and industrial statistician for empirical model building, with the intention of optimizing the response variables which are influenced by several explanatory variables or design

*Corresponding author

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

E-mail: oeguasa@biu.edu.ng,

Published on Web 30/07/2023, www.ijsronline.org

variables with a limited number of experiments (E. L. Walker *et al*, 2002; O. Eguasa *et al*, 2022).

RSM is suitable for optimizing the response variable $\boldsymbol{y}$ as a function of more than one explanatory variable $(x_{i1}, x_{i2}, \ldots, x_{ik})$ which can be modeled as:

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

where ε_i is the error term and assumed to have a normal distribution with mean zero and variance $\boldsymbol{\sigma}^2$. The surface represented by $f(x_{i1}, x_{i2}, \ldots, x_{ik})$ is called a response surface (W. Wan *et al*, 2011).

The true response function f is unknown which needs to be estimated. Applying RSM, we try to identify the functional relationship between the responses y and related explanatory variables $(x_{i1}, x_{i2}, \ldots, x_{ik})$.

The Parametric Regression Model

The conventional technique used in modeling the relationship between the *kth* explanatory variables and the *ith* response is to assume that the basic functional form can efficiently be expressed parametrically. The parametric regression model may be superior, if the user can identify adequately a parametric form for the data.

Hence, the general parametric regression model in matrix notation can be written as:

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

where y is a vector of response, $X = X^{(OLS)}$ 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 Quadratic Regression (QM) Model

The Quadratic model is given as:

$$y_i = \beta_0 + \beta_1 A + \beta_2 B + \beta_{11} A^2 + \beta_{22} B^2 + \beta_{12} AB \tag{3}$$

where $A = x_1$, $B = x_2$ are the explanatory variables; β_0 is a constant coefficient; the varying coefficients β_1, β_2 and β_{11}, β_{22} are the coefficients of linear, quadratic and interaction terms respectively (M. Sadeghian *et al*, 2014).

The ordinary least squares

The common approach for estimating the parameter vector in Equation (2) is usually based on the Method of OLS. The parameter vector estimates $\widehat{\boldsymbol{\beta}}$ in (2) is given as:

$$\widehat{\boldsymbol{\beta}}^{(OLS)} = \left(\boldsymbol{X}'^{(OLS)}\boldsymbol{X}^{(OLS)}\right)^{-1}\boldsymbol{X}'^{(OLS)}\boldsymbol{y}, \quad \boldsymbol{X} = \boldsymbol{X}^{(OLS)} \tag{4}$$

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

$$\widehat{\boldsymbol{y}}_i^{(OLS)} = \boldsymbol{x}_i'^{(OLS)}\widehat{\boldsymbol{\beta}}^{(OLS)} = \boldsymbol{x}_i'^{(OLS)}\left(\boldsymbol{X}'^{(OLS)}\boldsymbol{X}^{(OLS)}\right)^{-1}\boldsymbol{X}'^{(OLS)}\boldsymbol{y}, \quad i = 1,2,\ldots,n \tag{5}$$

where $\boldsymbol{x}_i'^{(OLS)}$ is the i^{th} row of matrix $\boldsymbol{X}^{(OLS)}$, $n \times (k+1)$ vector. (M. K. Carley *et al*, 2004; D. L. Rivers, 2009).

Materials and Methods

As given in the literature, the two operating factors are trans-membrane pressure (TMP) and velocity with the multi-response variables; permeate flux (PF), turbidity removal and the chemical oxygen demand (COD) removal. The goal is to obtain an optimum setting of the factors that would simultaneously maximize PF, turbidity removal and COD removal (M. Sadeghian *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 (D. Gramato *et al*, 2014; O. Eguasa *et al*, 2019).

The Local Linear Regression (LLR) Model

An alternative technique when the researcher lack information or there is partial knowledge of the functional form of the model, then the use of nonparametric regression model is most appropriate model and as such, the LLR model gain emphasis in this situation. The LLR model is a weighted form of the least squares obtained from Local Polynomial Regression of order one, which is an improvement over the kernel regression model because it can adjust to bias both at the boundaries of the explanatory variables (D. Ruppert *et al*, 1994; E. L. Walker *et al*, 2002).

The LLR model is derived from ordinary least squares theory. The LLR estimator $\widehat{\boldsymbol{y}}_i^{(LLR)}$ of $\boldsymbol{y}_i$ is given as:

$$\widehat{\boldsymbol{y}}_i^{(LLR)} = \boldsymbol{x}_i'^{(LLR)}(\boldsymbol{X}'^{(LLR)}\boldsymbol{W}_i\boldsymbol{X}^{(LLR)})^{-1}\boldsymbol{X}'^{(LLR)}\boldsymbol{W}_i\boldsymbol{y} = \boldsymbol{H}_i^{(LLR)}\boldsymbol{y}, \tag{6}$$

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

$$\boldsymbol{X}^{(LLR)} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \tag{7}$$

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

$$\boldsymbol{W} = v_i\delta_{ir}, \quad i = 1, 2, ..., n; \quad r = 1, 2, ..., n$$

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

$$\delta_{ir} = \begin{cases} 1, & if\ i = r \\ 0, & otherwise \end{cases}$$

$$i = 1, 2, ..., n; r = 1, 2, ..., n \tag{8}$$

Thus,

$$\boldsymbol{W} = \begin{bmatrix} v_1\delta_{11} & v_1\delta_{12} & \cdots & v_1\delta_{1n} \\ v_2\delta_{21} & v_2\delta_{22} & \cdots & v_2\delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_n\delta_{n1} & v_n\delta_{n2} & \cdots & v_n\delta_{nn} \end{bmatrix} \tag{9}$$

$$\boldsymbol{W} = \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_n \end{bmatrix}$$

where $v_1 = w_{i1}, v_2 = w_{i2}, ..., v_n = w_{in}$. In terms of location, $\boldsymbol{W} = \boldsymbol{W_i}$

$$\boldsymbol{W_i} = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix},$$

$$i = 1, 2, ..., n. \tag{10}$$

(W. Wan *et al*, 2011; O. Eguasa *et al*, 2019).

Equation (10) can be written as:

$$\boldsymbol{W_i} = dia(w_{i1}, w_{i2}, ..., w_{in}) \quad \text{for each } i = 1, 2, ..., n.$$

We can rewrite Equation (6) in terms of hat matrix as:

$$\hat{\boldsymbol{y}}^{(LLR)} = \boldsymbol{H}^{(LLR)}\boldsymbol{y}, \tag{11}$$

where the $n \times n$ matrix, $\boldsymbol{H}^{(LLR)}$ is the LLR hat matrix written as:

$$\boldsymbol{H}^{(LLR)} = \begin{bmatrix} \boldsymbol{x}_1'^{(LLR)}\left(\boldsymbol{X}'^{(LLR)}\boldsymbol{W}_1\boldsymbol{X}^{(LLR)}\right)^{-1}\boldsymbol{X}'^{(LLR)}\boldsymbol{W}_1 \\ \boldsymbol{x}_2'^{(LLR)}\left(\boldsymbol{X}'^{(LLR)}\boldsymbol{W}_2\boldsymbol{X}^{(LLR)}\right)^{-1}\boldsymbol{X}'^{(LLR)}\boldsymbol{W}_2 \\ \vdots \\ \boldsymbol{x}_n'^{(LLR)}\left(\boldsymbol{X}'^{(LLR)}\boldsymbol{W}_n\boldsymbol{X}^{(LLR)}\right)^{-1}\boldsymbol{X}'^{(LLR)}\boldsymbol{W}_n \end{bmatrix} \tag{12}$$

The weakness of LLR model is high bias in areas where there is curvature because the setup of model matrix of the LLR model do not incorporates quadratic terms (Z. He *et al*, 2009).

Locally Adaptive Bandwidths

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

$$b_{ij} = f\left(x_{ij}, \frac{1}{2}, T_{1j}, T_{2j}\right) = T_{1j}\left(\frac{1}{2} - \frac{x_{ij}}{T_{2j}}\right)^2, \quad i = 1,2, ..., n; j = 1,2, ..., k \tag{13}$$

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

Experimental Design

In RSM application, the factors are more than one. Hence, the choice of suitable levels to be studied for the explanatory variables is also vital as it can affect model correctness. The Experimental Design phase permits an appropriate design that can adequately and substantially 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).

Therefore, the factors and coded levels as given in the literature are shown below in Table 1:

Table 1: Coded stages and range for the design of experiments (M. Sadeghian *et al*, 2014)

Variables	*Factors or Input parameters*	*Coded Levels*				
		-1.414(-α)	-1(Low)	0(Medium)	1(High)	+1.414(+α)
TMP (bar)	$A = x_1$	0.7	1.0	1.7	2.5	2.8
Velocity (m/s)	$B = x_2$	1.0	1.5	2.6	3.7	4.2

In Table 2, given below is the CCD as given in the literature for two experimental factors and three responses.

Table 2: Coded stages and range for the design of experiments (M. Sadeghian *et al*, 2014)

Exptal. Run	*Exptal. Design*			*Coded Levels*	
	A: TMP (code)	B: Velocity (code)	PF (L/m^{2}h)	Turbidity Removal (%)	COD Removal (%)
1	0.7 (-1.414)	2.6 (0)	40	82	78
2	1.7 (0)	1.0 (-1.414)	32	85	82
3	1.7 (0)	2.6 (0)	48	74	79
4	2.5 (+1)	3.7 (+1)	64	72	68
5	2.8 (+1.414)	2.6 (0)	52	70	69
6	2.5 (+1)	1.5 (-1)	42	68	78
7	1.7 (0)	2.6 (0)	50	73	81
8	1.7(0)	2.6 (0)	47	75	78
9	1.7(0)	2.6 (0)	49	77	80
10	1.7(0)	4.2 (+1414)	68	69	68
11	1.0 (-1)	3.7 (+1)	58	80	73
12	1.0 (-1)	1.5 (-1)	33	88	82
13	1.7 (0)	2.6 (0)	49	76	79

Data transformation to RSM data in the interval of zero and one

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

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

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}, Min(x_{OLD}) and $Max(x_{OLD})$ are the minimum and maximum values in the vector x_{OLD} respectively, (O. 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 $x_0: -1.414, 0$; $Min(x_{OLD}): -1.414, -1.414$; $Max(x_{OLD}): +1.414, +1.414$

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

$$Explanatory\ variable\ x_1 : x_{11} = \frac{-1.414 - (-1.414)}{((-1.414) - (-1.414))} = 0.0000$$

$$Explanatory\ variable\ x_2 : x_{12} = \frac{-1.414 - (0)}{((-1.414) - (+1.414))} = 0.5000$$

where $x_1 = $ A: TMP, $x_2 = B: Velocity$

Hence, we now generate values of the operating factors that lie between zero and one for RSM data as given in Table 3. We shall fit the coded values to the quadratic model of Equation (3) and thereafter fit the OLS and the adaptive LLR model to the RSM data that are coded to lie in the interval of 0 and 1 inclusively as given in Table 3.

Indian Journal of Science and Research
Eguasa O. & E Eguasa M. E., Ind. J. Sci. Res. 2023, 3(3), 120-128

Table 3: Transformed RSM data in the interval of zero and one

Exptal. Run	*Exptal. Design*		*Coded Levels*		
	A: TMP (code)	B: Velocity (code)	PF (L/m^2h)	Turbidity Removal (%)	COD Removal (%)
1	0.0000	0.5000	40	82	78
2	0.5000	0.0000	32	85	82
3	0.5000	0.5000	48	74	79
4	0.8536	0.8536	64	72	68
5	1.0000	0.5000	52	70	69
6	0.8536	0.1464	42	68	78
7	0.5000	0.5000	50	73	81
8	0.5000	0.5000	47	75	78
9	0.5000	0.5000	49	77	80
10	0.5000	1.0000	68	69	68
11	0.1464	0.8536	58	80	73
12	0.1464	0.1464	33	88	82
13	0.5000	0.5000	49	76	79

Multi-Response Optimization Problem

This involves the optimization of two or more responses simultaneously with the associated factors $(x_{i1}, x_{i2}, \ldots, x_{ik})$. The optimization criteria and the desired goal for the multi-response problem as given in (M. Sadeghian *et al*, 2014) as shown in Table 4 below:

Table 4: Optimization criteria at the desired goal (M. Sadeghian *et al*, 2014)

Criteria	*Goal*	*Lower Limit*	*Upper Limit*
TMP (bar)	In the range	1	2.5
Velocity (m/s)	In the range	1.5	3.7
PF (L/m^2 h)	Maximize	32	68.4
Turbidity Removal (%)	Maximize	68	88
COD Removal (%)	Maximize	68	82

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

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

$$d_p\left(\hat{y}_p(\boldsymbol{x})\right) = \begin{cases} 0, & \hat{y}_p(\boldsymbol{x}) < L \\ \left\{\frac{\hat{y}_p(x)-L}{T-L}\right\}^{t_1}, & L \leq \hat{y}_p(\boldsymbol{x}) \leq T, \\ 1, & \hat{y}_p(\boldsymbol{x}) > T, \end{cases}$$

$$s.t\ \boldsymbol{x} \epsilon\ \varphi\ , \qquad (15)$$

where T and L are the maximum acceptable value and lower limit, respectively, of the p^{th} response.

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 (D. E. Castillo, 2007; W. Wan, 2007; Z. He *et al*, 2009; Z. He *et al*, 2012).

In the application, we give the individual desirability of the process requirement for the three responses.

For PF (L/m2 h); $d_p\left(\hat{y}_p(\boldsymbol{x})\right)$ is given as:

Review Article

Indian Journal of Science and Research

Eguasa O. & E Eguasa M. E., Ind. J. Sci. Res. 2023, 3(3), 120-128

$$d_1(\hat{y}_1(\boldsymbol{x})) = \begin{cases} 0, & \hat{y}_1(\boldsymbol{x}) < 32 \\ \left\{\frac{\hat{y}_1(\boldsymbol{x})-32}{68.4-32}\right\}^{t_1}, & 32 \le \hat{y}_1(\boldsymbol{x}) \le 68.4, \\ 1, & \hat{y}_1(\boldsymbol{x}) > 68.4, \end{cases}$$
$$s.t\ \boldsymbol{x} \epsilon\ \varphi\ , \qquad (16)$$

where $T = 68.4,\ L = 32$ are the maximum acceptable value and lower limit, respectively, of the 1^{st} response.

For Turbidity Removal (%); $d_2(\hat{y}_2(\boldsymbol{x}))$ is given as:

$$d_2(\hat{y}_2(\boldsymbol{x})) = \begin{cases} 0, & \hat{y}_2(\boldsymbol{x}) < 68 \\ \left\{\frac{\hat{y}_2(\boldsymbol{x})-68}{88-68}\right\}^{t_1}, & 68 \le \hat{y}_p(\boldsymbol{x}) \le 88, \\ 1, & \hat{y}_2(\boldsymbol{x}) > 88, \end{cases}$$
$$s.t\ \boldsymbol{x} \epsilon\ \varphi\ , \qquad (17)$$

where $T = 88,\ L = 68$ are the maximum acceptable value and lower limit, respectively, of the 2^{nd} response.

For COD Removal (%); $d_3(\hat{y}_3(\boldsymbol{x}))$ is given as:

$$d_3(\hat{y}_3(\boldsymbol{x})) = \begin{cases} 0, & \hat{y}_3(\boldsymbol{x}) < 68 \\ \left\{\frac{\hat{y}_3(\boldsymbol{x})-68}{82-68}\right\}^{t_1}, & 68 \le \hat{y}_3(\boldsymbol{x}) \le 82, \\ 1, & \hat{y}_3(\boldsymbol{x}) > 82, \end{cases}$$
$$s.t\ \boldsymbol{x} \epsilon\ \varphi\ , \qquad (18)$$

where $T = 82,\ L = 68$ are the maximum acceptable value and lower limit, respectively, of the 3^{rd} response.

The overall desirability function

The goal 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:

$$D = \sqrt[p]{(d_1(\hat{y}_1(\boldsymbol{x})).\ d_2(\hat{y}_2(\boldsymbol{x})) \ldots\ d_v(\hat{y}_v(\boldsymbol{x}))} \qquad (19)$$

where $v = 3$ is the number of response variables, $d_1\hat{y}_1(\boldsymbol{x}),\ d_2\hat{y}_2(\boldsymbol{x}),\ldots,\ d_v\hat{y}_v(\boldsymbol{x})$ are the individual desirability (R. Ramakrishnan *et al*, 2011; Z. He *et al*, 2012; D. Gramato *et al*, 2014). The desirability function $d_m(\hat{y}_m(\boldsymbol{x}))\ , m = 1,2,\ldots,v$ allocate values between 0 and 1 centered on the process requirements such that the most undesirable and desirable values are $d_v(\hat{y}_v(\boldsymbol{x})) = 0$ and $d_v(\hat{y}_v(\boldsymbol{x})) = 1$ respectively.

Results and Discussion

Table 5: Model Goodness-of-fits statistics for OLS, QM and LLR_{AB}

Response	*Model*	*DF*	*PRESS***	*PRESS*	*SSE*	*MSE*	R^2*(%)*	R^2_{Adj}*(%)*
y_1	OLS	7.0000	6.7789	47.4521	10.7304	1.5329	99.2081	98.6425
	QM	-	-	-	-	-	99.1700	98.5800
	LLR_{AB}	4.0610	**5.3970**	**66.1589**	**5.2245**	**1.2865**	**99.6100**	**98.8600**
y_2	OLS	7.0000	62.2476	435.7333	69.0777	9.8682	84.8820	74.0834
	QM	-	-	-	-	-	97.7000	**94.5000**
	LLR_{AB}	4.0002	**10.8406**	**142.3846**	**10.0000**	**2.4999**	**97.8114**	93.4347
y_3	OLS	7.0000	5.0451	35.3159	9.0237	**1.2891**	97.2254	95.2436
	QM	-	-	-	-	-	97.2500	94.5000
	LLR_{AB}	4.0307	**1.0194**	**13.6117**	**5.2902**	1.3125	**98.3734**	**95.1574**

The results obtained from Table 5 clearly shows that LLR_{AB} from the respective responses (Permeate flux (L/m^{2}h), Turbidity removal (%) and COD removal (%)) gave the better performance statistics as compared with drug dosage over $OLS\ and\ QM$ in twenty two cells as against eight cells for the multi-response problem supporting a reasonable correlation between the experimental and predicted values . The bolded cells indicate a better performance over cells that are not bold.

Indian Journal of Science and Research

Eguasa O. & E Eguasa M. E., Ind. J. Sci. Res. 2023, 3(3), 120-128

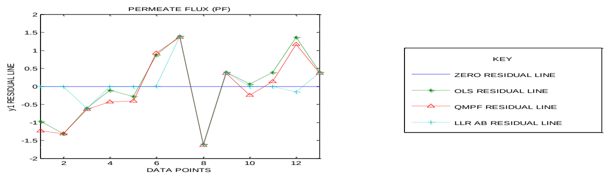

Figure 1: Model Residuals for y1 (Permeate flux) response plotted against the data points for $OLS, QM\ and\ LLR_{AB}$ models.

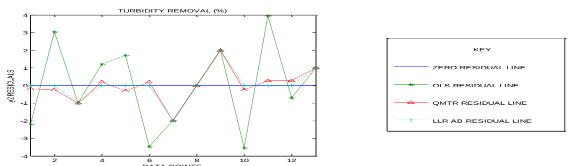

Figure 2: Model Residuals for y2 (Turbidity removal) response plotted against the data points for $OLS, QM\ and\ LLR_{AB}$ models.

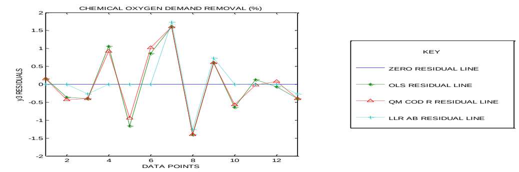

Figure 3: Model Residuals for y3 (COD removal) response plotted against the data points for $OLS, QM\ and\ LLR_{AB}$ models.

In Figures 1, 2 and 3, it is obvious that the LLR_{AB} model has the minimum residual line over the $OLS\ and\ QM$ models as it relates to the zero residual line for Permeate flux (L/m^{2}h), Turbidity removal (%) and COD removal (%).

Table 6: Model optimal solution based on the multi-response desirability function

Model	x_1	x_2	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	$d_1(\hat{y}_1)$	$d_2(\hat{y}_2)$	$d_3(\hat{y}_3)$	$D(\%)$
OLS	0.2066	0.5396	46.0553	79.4585	79.2505	0.3861	0.5729	0.8036	56.23
QM	0.2727 (1)	0.6250 (3)	50.6600	78.8300	76.6300	0.5126	0.5415	0.6164	55.51
LLR_{AB}	**0.1818**	**0.1642**	41.4898	87.0076	81.5456	0.2607	0.9504	0.9675	**62.12**

From Table 6, LLR_{AB} provides a better multi-response in the maximization of the ultrafiltration of oil-in-water emulsion over OLS and QM in terms of overall desirability for the respective factors $x_1 =$ Transmembrane pressure (TMP), $x_2 =$ Velocity.

Obviously, LLR_{AB} gave a better process requirement with **62%** desirability and with operating factors $x_1 = \mathbf{0.1818}$, $x_2 = \mathbf{0.1642}$ with the best choice based on oil-in-water emulsion.

Table 7: Optimization and process requirement for the actual, predicted factors and responses

Criteria	Goal	Lower limit	Target point	LLR_{AB}	QM	OLS
TMP (bar)	Range	1	2.5	0.1818	0.2727	0.2066
Velocity (m/s)	Range	1.5	3.7	0.1642	0.6250	0.5396
PF (L/m^2 h)	Maximize	32	68.4	42	**51**	46
Turbidity R (%)	Maximize	68	88	**87**	79	80
COD R (%)	Maximize	68	82	**82**	77	79

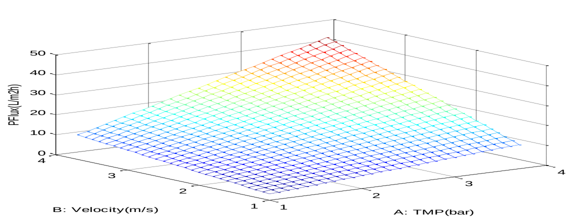

Figure 4: LLR_{AB} surface plot for maximum PF of 42% showing the effect of TMP and Velocity

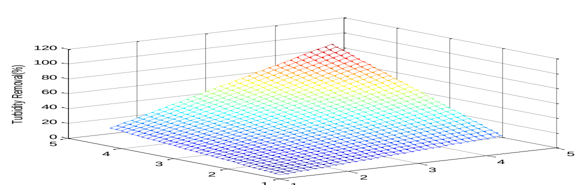

Figure 5: LLR_{AB} surface plot for maximum Turbidity removal of 87% showing the effect of TMP and

Review Article

Indian Journal of Science and Research

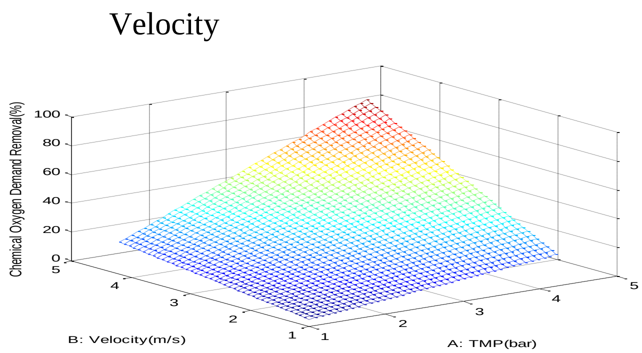

Figure 6: LLR_{AB} surface plot for maximum COD of 82% showing the effect of TMP and Velocity

The maximization of the individual desirability functions are given in Figure 4, 5 and 6 shows the shapes of the different colour variation in the surface plots representing individual desirability of the optimization criteria for Permeate flux (L/m^{2}h), Turbidity removal (%) and COD removal (%) respectively. In Figure 4, the individual desirability function for LLR_{AB} is 42% as against the optimization results for OLS and QM with respective individual desirability functions of 46% and 51% for Permeate flux (L/m^{2}h). Whereas, in Figure 5 the individual desirability function for LLR_{AB} is 87% as against the optimization results for OLS and QM with respective individual desirability functions of 80% and 79% for Turbidity removal (%) and in Figure 6, the individual desirability function for LLR_{AB} is 82% as against the optimization results for OLS and QM with respective individual desirability functions of 79% and 77% for COD removal (%).

Conclusions: In this work, we examined results from three regression models OLS, QM and LLR_{AB} via the CCD in terms of three factors, five levels and three responses for RSM data. We presented a more robust LLR_{AB} and the factors in the experimental matrix were transformed by a mathematical relation that needed to lie between zero and one for RSM data. We statistically analyzed the experimental data using LLR_{AB} model for Permeate flux (L/m^{2}h), Turbidity removal (%) and COD removal (%) in terms of performance statistics, residual plots and the surface plots over the existing Quadratic model and OLS. The results obtained shows that the LLR_{AB} is expected to assist further research work on ultrafiltration of oil-in-water emulsion.

Acknowledgment: I sincerely want to thank God Almighty the giver of life, strength and wisdom to enable us accomplished this task.

To my PhD supervisor, Prof. J. I. Mbegbu, I am indeed grateful for his tutelage and patience towards my postgraduate study.

Funding

This research work received no external funding.

Conflict of interest

There was no conflict of interest as regards authorship or publication of this article.

References

1. Anderson-Cook, C. M. & Prewitt, K.. Some guidelines for using nonparametric models for modeling data from Response Surface Designs. Journal of Modern Applied Statistical Models, 4, pp. 106-119. (2005)
2. Box, G. E. P. & Wilson, K. B. On the experimental attainment of optimum conditions. Journal of the Royal Statistical Society Series B, 13, pp. 1- 45. (1951)
3. Box, G. E. P. & Hunter, J. S. Multifactor experimental designs for exploring response surfaces, Ann Math Statist, 28, pp.195-241. (1957)
4. Box, G. E. P. & Draper, N. R. A basis for the selection of a response surface design. Journal American Stat. Association, Vol. 54, pp. 622-654. (1959)
5. Carley, M. K., Kamneva, Y. N. & Reminga, J. 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. (2004)
6. Castillo, D. E. Process Optimization: A statistical Method. Springer International Series in Operations Research and Management Science: New York. (2007)

7. Eguasa, O., Mbegbu, J. I. & Edionwe, E. On the Use of Nonparametric Regression Model for Response Surface Methodology (RSM), Benin Journal of Statistics, 2, pp. 61-75. (2019)
8. Eguasa, O. Edionwe, E. & Mbegbu, J. I. 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 (2022)
9. Gramato, D. & Calado, V. 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. (2014)
10. He, Z., Wang, J., Oh, J. & Park, S. H. Robust optimization for multiple responses using response surface methodology, Applied Stochastic Models in Business and Industry, 26, pp.157 – 171. (2009)
11. He, Z., Zhu, P. F. & Park, S. H. A robust desirability function for multi-response surface optimization. European Journal of Operational Research, 221, pp. 241-247. (2012)
12. Hill, W. J. & Hunter, W. G. A review of response surface methodology: A literature survey. *Technimetrics*, 8, pp. 571-579. (1966)
13. Myers, R. Montgomery, D. C. & Anderson-Cook, M. C. Response Surface Methodology: Process and Product Optimization Using Designed Experiments Wiley. (2009)
14. Nair, A. T., Makwana, A. R. & Ahammed, M. M. The use of Response Surface Methodology for modelling and analysis of water and waste – water treatment processes: A Review. Water Science and Technology, 69, pp. 464 – 478. (2014)
15. Ramakrishnan, R. & Arumugam, R. 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, pp. 39 – 50. (2011)
16. Rivers, D. L. A Graphical Analysis of Simultaneously Choosing the Bandwidth and Mixing Parameter for Semiparametric Regression Techniques, M. Sc. Thesis and Dissertations, Virginia Commonwealth University. (2009)
17. Ruppert, D. & Wand, M. P. Multivariate locally weighted least squares regression. The Annals of Statistics, 22, pp. 1346-1370. (1994)
18. Sadeghian, M., Sadeghi, M., Hesampour, M. & Moheb, A. Application of response surface methodology (RSM) to optimize operating conditions during ultrafiltration of oil in water emulsion, Desalination and water treatment, DOI:10.1080/19443994.2014.919607 (2014)
19. Walker, E. L., Starnes, B. A., Birch, J. B. & Mays, J. E. Model Robust Calibration: Methods and Application to Electrically-Scanned Pressure Transducers, Technical Report, Dept. of Mathematics, Virginia Commonwealth University, Richmond, VA, USA. (2002)
20. Wan, W. Semi-parametric techniques for multi-response optimization. Ph.D Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University, Blacksburg, Virginia. (2007)
21. Wan, W. & Birch, J. B. A semi-parametric technique for multi-response optimization. Journal of Quality and Reliability Engineering International, 27, pp. 47-59. (2011)
22. Yeniay, O. Comparative study of algorithm for response surface optimization. Journal of Mathematical and Computational Applications, 19, pp. 93 – 104. (2014)