Research Article / Review Article FUPRE JOURNAL 9(1):410-428(2025)

FUPRE Journal

of

Scientific and Industrial Research

ISSN: 2579-1184(Print) ISSN: 2578-1129 (Online)

http://fupre.edu.ng/journal

Response Surface Methodology for Optimizing and Simulating the Extraction of Anti-Diabetic Compounds from *Cucumis Sativus*

***EGUASA, O.*[1,*], *OKUNGBOWA, A. I.*[2]**

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

ARTICLE INFO

Received: 10/01/2025
Accepted: 14/04/2025

Keywords

β-glucosidase, Cucumis Sativus, Response surface methodology, Simulation study, Quadratic regression model

ABSTRACT

The growing prevalence of diabetes mellitus has intensified the search for effective, plant-based therapeutic agents with minimal side effects. *Cucumis Sativus* (commonly known as cucumber) has been recognized for its potential anti-diabetic properties, attributed to its rich phytochemical profile. This study aims to optimize and simulate the extraction process of anti-diabetic compounds from *Cucumis Sativus* using Response Surface Methodology (RSM), a statistical and mathematical tool effective for modeling and analyzing problems where multiple variables influence the desired outcome. In the literature, experimental data were fitted to a second-order polynomial regression model, and the model's adequacy was confirmed through with a high R^2 value indicating a strong predictive reliability. A Central Composite Design (CCD) was employed to systematically evaluate the influence of four independent variables; extraction temperature (°C), extraction incubation-time (minutes), agitation speed (rpm), and volume of solvent (mL) on the yield of bioactive compounds exhibiting $\beta - glucosidase$ inhibitory activities. The response was fitted to a second-order polynomial model, and model adequacy was confirmed by goodness-of-fit statistics and statistically significant with minimal residual. The optimization results for Squared - Distance from target revealed that moderate extraction temperature (°C), extraction incubation-time (minutes), agitation speed (rpm), and volume of solvent (mL), significantly enhanced the recovery of anti-diabetic compounds satisfying optimal – process requirements for the target values. The predicted optimal conditions were validated, showing a strong correlation between the observed and predicted values, thereby confirming the model's reliability. Furthermore, simulation studies were conducted to visualize the simulated – optimal target values as the misspecification parameter increases from zero to unity showing interactions among variables and to assess the process robustness under varying operational conditions. This work demonstrates the utility of RSM as a robust tool for optimizing and simulating complex extraction processes in natural product research. The optimized extraction protocol for *Cucumis Sativus* can serve as a foundation for the development of functional foods or phytopharmaceuticals targeting diabetes management.

*Corresponding author, e-mail: oeguasa@biu.edu.ng
DIO

1. INTRODUCTION

Diabetes mellitus is a chronic metabolic disorder characterized by elevated blood glucose levels due to impaired insulin secretion, action, or both. The global prevalence of diabetes has risen substantially over recent decades and is projected to affect over 640 million adults by 2040 (International Diabetes Federation (IDF), 2021). Despite the availability of synthetic ant-diabetic drugs, their prolonged use is often associated with adverse effects such as gastrointestinal discomfort, hypoglycemia, and drug resistance (Marrelli *et al*., 2016). Consequently, there has been a growing interest in identifying and developing natural, plant-based compounds with hypoglycemic potential, owing to their safety, accessibility, and affordability.

Cucumis Sativus, commonly known as cucumber, is a member of the Cucurbitaceae family and is widely consumed for its nutritional and medicinal properties. Traditionally, cucumber has been used in Ayurvedic and Unani medicine for managing various ailments including diabetes (Grover *et al*., 2002). Phytochemical studies have revealed that cucumber contains a diverse array of bioactive compounds such as flavonoids, tannins, phenolic acids, and saponins, which exhibit antioxidant, anti-inflammatory, and ant-diabetic activities (Sarkar *et al*., 2012; Dhanani *et al*., 2017). In particular, these compounds are known to inhibit key carbohydrate-hydrolyzing enzymes such as α-amylase and α-glucosidase, which play a central role in postprandial hyperglycemia management (Ali *et al*., 2006).

The efficiency of extracting these bioactive compounds from plant matrices is highly dependent on several process parameters, including extraction temperature, solvent composition, extraction time, and the solid-to-liquid ratio. Traditional extraction methods often fail to achieve high yields or require large quantities of solvent and energy. Therefore, optimization of these parameters is essential for maximizing extraction efficiency while ensuring compound stability and activity.

Response Surface Methodology (RSM) is a powerful statistical and mathematical tool used for modeling and optimizing complex processes involving multiple variables and their interactions. It reduces the number of experimental trials needed and provides a predictive model to assess the influence of each factor on the response (Myers *et al*., 2016). Central Composite Design (CCD), a commonly used design in RSM, allows for efficient estimation of quadratic response surfaces and identification of optimal conditions with minimal experimentation (Montgomery, 2017). Previous studies have successfully employed RSM in the optimization of bioactive compound extraction from various plants, confirming its applicability and reliability in natural product research (Rathore *et al.,* 2011; Norshazila *et al*., 2010).

This study aims to optimize and simulate the extraction conditions for anti-diabetic compounds from *Cucumis Sativus* using RSM. The specific objectives are: (i) to evaluate the effect of key extraction variables on the yield of enzyme-inhibitory compounds, (ii) to develop a predictive model using CCD, and (iii) to identify and validate the optimal extraction conditions through experimental confirmation and numerical simulation. The outcomes of this study are expected to provide a scientific

basis for the development of standardized extraction protocols for natural anti-diabetic agents from *Cucumis Sativus*.

Response Surface Methodology (RSM) is well-suited for optimizing a response variable *y* based on multiple explanatory variables ($x_{i1}, x_{i2}, \ldots, x_{ik}$), and can be represented using the following model:

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

ε_i represents the error term, which is assumed to be normally distributed with a mean of zero and a variance of σ^2. The function $f(x_{i1}, x_{i2}, \ldots, x_{ik})$ defines the response surface that characterizes the relationship between the response and the explanatory variables (Wan and Birch, 2011).

The true form of the response function *f* is generally unknown and must be approximated. In Response Surface Methodology (RSM), the objective is to establish an estimated functional relationship between the response variable *y* and the set of explanatory variables ($x_{i1}, x_{i2}, \ldots, x_{ik}$).

The conventional method for modeling the relationship between the *k*th explanatory variable and the *i*th response relies on the assumption that the underlying functional form can be adequately described using a parametric model. Such a model can be beneficial, assuming the user is able to correctly identify and specify a suitable functional form for the data.

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

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

where $\boldsymbol{y}$ is a vector of response, $\boldsymbol{X} = \boldsymbol{X}^{(OLS)}$ is the OLS model matrix, $\boldsymbol{\beta}$ is the unknown parameter vector and $\boldsymbol{\varepsilon}$ is the vector of error term assumed to be normally distributed with zero mean and constant variance property.

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

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

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

$$\widehat{\boldsymbol{y}}_i^{(OLS)} = \boldsymbol{x}_i'^{(OLS)}\widehat{\boldsymbol{\beta}}^{(OLS)} = \boldsymbol{x}_i'^{(OLS)}\left(\boldsymbol{X}'^{(OLS)}\boldsymbol{X}^{(OLS)}\right)^{-1}\boldsymbol{X}'^{(OLS)}\boldsymbol{y}, \qquad i = 1,2, \ldots, n \tag{4}$$

where $\boldsymbol{x}_i'^{(OLS)}$ is the i^{th} row of matrix $\boldsymbol{X}^{(OLS)}$, $n \times (k+1)$ vector.
$\boldsymbol{H}_i'^{(OLS)} = \boldsymbol{x}_i'^{(OLS)}\left(\boldsymbol{X}'^{(OLS)}\boldsymbol{X}^{(OLS)}\right)^{-1}\boldsymbol{X}'^{(OLS)}$ is the i^{th} row of the OLS "HAT" matrix of dimension $n \times n$, $\boldsymbol{H}^{(OLS)}$. The estimated response in the i^{th} location is given by:

$$\widehat{\boldsymbol{y}}^{(OLS)} = \boldsymbol{H}^{(OLS)}\boldsymbol{y}. \tag{5}$$

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

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

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

1.1.*THE QUADRATIC REGRESSION MODEL (QM)*

The Quadratic model is given as:

$$y_i = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C + \beta_4 D + \beta_{11} A^2 + \beta_{22} B^2 + \beta_{33} C^2 + \beta_{44} D^2 + \beta_{12} AB + \beta_{13} AC + \beta_{14} AD + \beta_{23} BC + \beta_{24} BD + \beta_{34} CD \quad (5)$$

where $A = x_1$, $B = x_2, C = x_3$, $D = x_4$ 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 (Jamal *et al*., 2011).

2. METHODOLOGY

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.

2.1. 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 $\alpha = \pm\sqrt[4]{2^k}$ from the center point and k_c, k^{th} central points (Bezerra *et al.*, 2008). The Circumscribed CCD can express geometrically as:

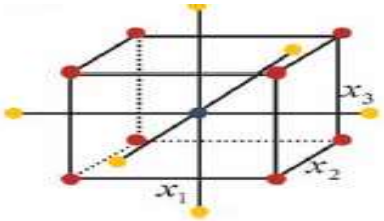

Figure 1: Circumscribed CCD (30 points, when k=4) with factorial design points (16 points), axial points (8 points) and with at least k^{th} central point (6 points).
Sources: Peasura (2015)

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:

$$CCCD = 2^k + 2k + k_c \quad (6)$$

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 $k_c = 6$ which from equation (6) sum up to 30 experimental runs.

A second-order linear regression model is given as:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \sum_{j=1}^{k} \beta_{jj} x_{ij}^2 + \sum_{j=1}^{k-1} \sum_{r=j+1}^{k} \beta_{jr} x_{ij} x_{ir} + \varepsilon_i, \; i = 1,2,\dots,n; \; r = j+1, j+2, \dots, k \quad (7)$$

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.

3. APPLICATION

We will evaluate the performance of the adaptive LLR_{AB} over the QM, focusing on the goodness-of-fit statistics and the optimal settings of the explanatory variables that maximize the response using the two RSM datasets.

3.1. SINGLE RESPONSE OPTIMIZATION PROBLEM

This paper focuses on optimizing a single response, aiming to identify the settings of the explanatory variables that will either maximize or minimize the fitted response (Eguasa *et al*., 2022). As such, the optimization criterion is based on the constrained minimization of the estimated Squared Distance from Target (SDT), defined as:

$$\text{Minimize} \quad \widehat{SDT} = (\hat{y}(\boldsymbol{x}) - T)^2 \quad (8)$$
$$s.t\, \boldsymbol{x} \epsilon\, \varphi\,,$$

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

3.2. APPLICATION: SINGLE RESPONSE EXTRACTION YIELD FOR RESPONSE SURFACE PROCESS DATA

The study outlined by Jamal *et al*. (2011) aimed to establish a relationship between maximum productivity of $\beta - glucosidase$ inhibition with process variables such as temperature (x_1), incubation time (x_2), agitation speed (x_3) and volume of solvent (x_4) with the goal of maximizing the inhibitory activity. Table 1 is the experimental range and level of independent process variables given below:

*Corresponding author, e-mail: oeguasa@biu.edu.ng
DIO

Table 1. Experimental design range and level of independent process variables (Jamal *et al*., (2011).

Independent Variables	Symbols	Levels of Independent Process Variables				
		-2	-1	0	+1	+2
Temperature (℃)	x_1	20	25	30	35	40
Incubation Time (h)	x_2	15	20	25	30	35
Agitation Speed (rpm)	x_3	50	75	100	125	150
Volume of Solvent (ml)	x_4	5	10	15	20	25

Table 2: CCD with actual values of factors and $\beta - glucosidase$ inhibition activity as a response (Jamal *et al*., (2011).

Run	Independent Variables				Actual Response y	Predicted Response $\hat{y}$	Predicted Response $QM\ \hat{y}$	Residual $: r = y - \hat{y}$
	x_1 (℃)	x_2 (h)	x_3 (rpm)	x_4 (ml)				
1	25	30	75	10	25.57	36.64	34.73	-11.06
2	35	30	75	10	45.57	30.26	29.03	15.31
3	25	20	125	10	89.34	108.02	107.88	-18.68
4	35	20	125	10	96.39	86.97	87.18	9.42
5	30	35	100	15	14.43	20.83	18.55	-6.40
6	30	15	100	15	58.36	50.73	51.35	7.63
7	30	25	100	15	95.74	79.32	78.95	16.42
8	35	20	75	20	18.95	3.66	4.43	15.29
9	30	25	100	15	82.13	79.32	78.95	2.81
10	30	25	100	15	78.98	79.32	78.95	-0.34
11	35	30	75	20	29.84	30.97	30.28	-1.14
12	20	25	100	15	91.80	72.30	71.15	19.50
13	25	30	125	10	54.10	50.81	49.23	3.29
14	30	25	100	5	49.84	60.25	59.45	-10.41
15	30	25	150	15	92.13	85.03	84.95	7.10
16	30	25	100	25	47.87	36.23	36.45	11.64
17	25	30	125	20	29.84	44.33	42.98	-14.49
18	25	20	75	20	24.92	54.00	54.13	-29.09
19	30	25	50	15	30.71	36.58	35.95	-5.87
20	30	25	100	15	76.62	79.32	78.95	-2.70
21	35	20	125	20	29.18	37.93	38.93	-8.75
22	35	30	125	20	19.85	18.63	18.28	1.23

23	30	25	100	15	68.52	79.32	78.95	-10.80
24	40	25	100	15	2.30	20.57	20.75	-18.27
25	25	30	75	20	70.82	61.66	59.98	9.16
26	25	20	125	20	86.56	83.29	83.63	3.27
27	30	25	100	15	73.93	79.32	78.95	-5.39
28	25	20	75	10	64.59	47.23	46.88	17.36
29	35	30	125	10	58.69	49.42	48.53	9.27
30	35	20	75	10	15.88	21.20	21.18	-5.32

Table 3: CCD with coded variables and $\beta - glucosidase$ inhibition activity as a response

Run	Independent Variables				Enzymes Inhibition (%)		
	x_1 (℃)	x_2 (h)	x_3 (rpm)	x_4 (ml)	Actual Response y	Predicted Response $\hat{y}$	Residual : $r = y - \hat{y}$
1	-1	-1	-1	-1	25.57	36.64	-11.06
2	1	-1	-1	-1	45.57	30.26	15.31
3	-1	1	-1	-1	89.34	108.02	-18.68
4	1	1	-1	-1	96.39	86.97	9.42
5	-1	-1	1	-1	14.43	20.83	-6.40
6	1	-1	1	-1	58.36	50.73	7.63
7	-1	1	1	-1	95.74	79.32	16.42
8	1	1	1	-1	18.95	3.66	15.29
9	-1	-1	-1	1	82.13	79.32	2.81
10	1	-1	-1	1	78.98	79.32	-0.34
11	-1	1	-1	1	29.84	30.97	-1.14
12	1	1	-1	1	91.80	72.30	19.50
13	-1	-1	1	1	54.10	50.81	3.29
14	1	-1	1	1	49.84	60.25	-10.41
15	-1	1	1	1	92.13	85.03	7.10
16	1	1	1	1	47.87	36.23	11.64
17	-2	0	0	0	29.84	44.33	-14.49
18	2	0	0	0	24.92	54.00	-29.09
19	0	-2	0	0	30.71	36.58	-5.87
20	0	2	0	0	76.62	79.32	-2.70
21	0	0	-2	0	29.18	37.93	-8.75
22	0	0	2	0	19.85	18.63	1.23
23	0	0	0	0	68.52	79.32	-10.80
24	0	0	0	0	2.30	20.57	-18.27
25	0	0	0	-2	70.82	61.66	9.16
26	0	0	0	2	86.56	83.29	3.27
27	0	0	0	0	73.93	79.32	-5.39
28	0	0	0	0	64.59	47.23	17.36
29	0	0	0	0	58.69	49.42	9.27
30	0	0	0	0	15.88	21.20	-5.32

3.3. DATA TRANSFORMATION USING CENTRAL COMPOSITE DESIGN (CCD)

In nonparametric regression methods applied to RSM, the explanatory variables are typically coded to fall within the interval [0, 1]. The data obtained through a Central Composite Design (CCD) is transformed using a mathematical relationship, as outlined in Eguasa *et al.* (2022):

$$x_{new} = \frac{Min(x_{old}) - x_0}{(Min(x_{old}) - Max(x_{old}))} \quad (9)$$

the transformed value is denoted as x_{new}, while x_o represents the target value to be transformed from the vector of original coded values, referred to as x_{old}. The terms Min (x_{old}) and $Max(x_{old})$ indicate the minimum and maximum values within the x_{old}vector, respectively (Eguasa *et al.*, 2022).

Table 4: Transformed CCD to RSM coded variables in the interval of zero and one inclusive.

Run	Independent Variables				Enzymes Inhibition (%)	
	x_1 (℃)	x_2 (h)	x_3 (rpm)	x_4 (ml)	Actual Response y	Predicted Response $\hat{y}$
1	0.2500	0.2500	0.2500	0.2500	25.57	36.64
2	0.7500	0.2500	0.2500	0.2500	45.57	30.26
3	0.2500	0.7500	0.2500	0.2500	89.34	108.02
4	0.7500	0.7500	0.2500	0.2500	96.39	86.97
5	0.2500	0.2500	0.7500	0.2500	14.43	20.83
6	0.7500	0.2500	0.7500	0.2500	58.36	50.73
7	0.2500	0.7500	0.7500	0.2500	95.74	79.32
8	0.7500	0.7500	0.7500	0.2500	18.95	3.66
9	0.2500	0.2500	0.2500	0.7500	82.13	79.32
10	0.7500	0.2500	0.2500	0.7500	78.98	79.32
11	0.2500	0.7500	0.2500	0.7500	29.84	30.97
12	0.7500	0.7500	0.2500	0.7500	91.80	72.30
13	0.2500	0.2500	0.7500	0.7500	54.10	50.81
14	0.7500	0.2500	0.7500	0.7500	49.84	60.25
15	0.2500	0.7500	0.7500	0.7500	92.13	85.03
16	0.7500	0.7500	0.7500	0.7500	47.87	36.23
17	0.0000	0.5000	0.5000	0.5000	29.84	44.33
18	1.0000	0.5000	0.5000	0.5000	24.92	54.00
19	0.5000	0.0000	0.5000	0.5000	30.71	36.58
20	0.5000	1.0000	0.5000	0.5000	76.62	79.32
21	0.5000	0.5000	0.0000	0.5000	29.18	37.93
22	0.5000	0.5000	1.0000	0.5000	19.85	18.63
23	0.5000	0.5000	0.5000	0.5000	68.52	79.32
24	0.5000	0.5000	0.5000	0.5000	2.30	20.57
25	0.5000	0.5000	0.5000	0.0000	70.82	61.66
26	0.5000	0.5000	0.5000	1.0000	86.56	83.29
27	0.5000	0.5000	0.5000	0.5000	73.93	79.32

*Corresponding author, e-mail: oeguasa@biu.edu.ng
DIO

28	0.5000	0.5000	0.5000	0.5000	64.59	47.23
29	0.5000	0.5000	0.5000	0.5000	58.69	49.42
30	0.5000	0.5000	0.5000	0.5000	15.88	21.20

4.0 DISCUSSION OF RESULTS

4.1. DESIGN OF EXPERIMENT AND STATISTICAL ANALYSIS

The impact of four process variables—temperature, incubation time, agitation speed, and extraction solvent volume was assessed to identify the conditions that maximize α-glucosidase inhibition. Table 2 presents the diagnostic case statistics comparing the actual experimental responses with the predicted values generated by the Design Expert software across all 30 experimental runs. The data were modeled using the quadratic polynomial Equation (10).

$$y = -739.45+14.88x_1+20.96x_2+4.94x_3+17.15x_4-0.33x_{12}-0.44x_{22} -0.0074x_{32}-0.31x_{42}+0.20x_1x_2+0.01x_1x_3-0.24x_1x_4-0.093x_2x_3+0.18x_2x_4 -0.063x_3x_4 \quad (10)$$

Table 5. The quadratic polynomial Equation for four factors

SOURCE	COEFFICIENTS
Intercept	**−739.45**
x_1	+14.88
x_2	+20.96
x_3	+4.94
x_4	+17.15
x_1^2	−0.33
x_2^2	−0.44
x_3^2	−0.0074
x_4^2	−0.31
$x_1 \cdot x_2$	+0.20
$x_1 \cdot x_3$	+0.01
$x_1 \cdot x_4$	−0.24
$x_2 \cdot x_3$	−0.093
$x_2 \cdot x_4$	+0.18
$x_3 \cdot x_4$	−0.063

Table 6. Comparison of the goodness-of-fit statistics for QM

Response	Model	R^2(%)
y	$Existing\ QM$	82.18
	QM	82.18

Table 7: Comparison of optimization results (process requirement) for the Extraction Process Data

Approach	x_1 (℃)	x_2 (h)	x_3 (rpm)	x_4 (ml)	$\hat{y}$ %
$Existing\ QM\ 1$	25.66	22.30	125	15.6	98.64
$Existing\ QM\ 2$	28.04	20.46	116	13.75	97.36
$Existing\ QM\ 3$	28.10	22.09	119	14.85	94.65
QM	27.20	15.48	150	6.30	136.55

From Table 7, provides the extraction yield for $QM's$ the four settings of the explanatory variables give the desired process satisfaction. The QM model provided optimal desirability over other existing Quadratic models.

4.2. SIMULATION STUDY

In this section, we compare the performances of the respective regression models, *QM* using data from the optimal responses for *QM* with interactive effects of temperature and incubation time, within the misspecification parameter γ simulated in the range 0.00, 0.25, 0.50, 0.75, 1.0; where the x_{1i},x_{2i} and x_{ij} are the explanatory variables.

4. 2.1. SIMULATION STUDY 1

The optimal response of QM, as reported in the literature, was used to simulate the polynomial model and assess the stability of the regression model under the misspecification parameter γ, which was required to lie within the interval $0 \leq \gamma \leq 1$.

The optimal response is given by:

$$Yield\ (\%) = 81.97 - 95.07 * \text{x1} + 46.36 * \text{x2} + 48.52 * \text{x1}.* \text{x2} + 25.11 * \text{x1}.\text{\^{}}2 - 96.29 * \text{x2}.\text{\^{}}2 \quad (11)$$

Table 8. Interaction effect between temperature and incubation time

SOURCE	COEFFICIENTS
Intercept	**+81.97**
x_1	−95.09
x_2	+46.36
x_1^2	+25.11
x_2^2	−96.29
$x_1 \cdot x_2$	+48.52

Model 1 `Z= 81.97-95.07*x1+46.36*x2+48.52*x1.*x2+25.11*x1.^2-96.29*x2.^2`

`+MS*(3*sin(3*pi*x1)-2*cos(2*pi*x2)+2*sin(4*pi*x1.*x2))`

$MS = \gamma = 0.0$

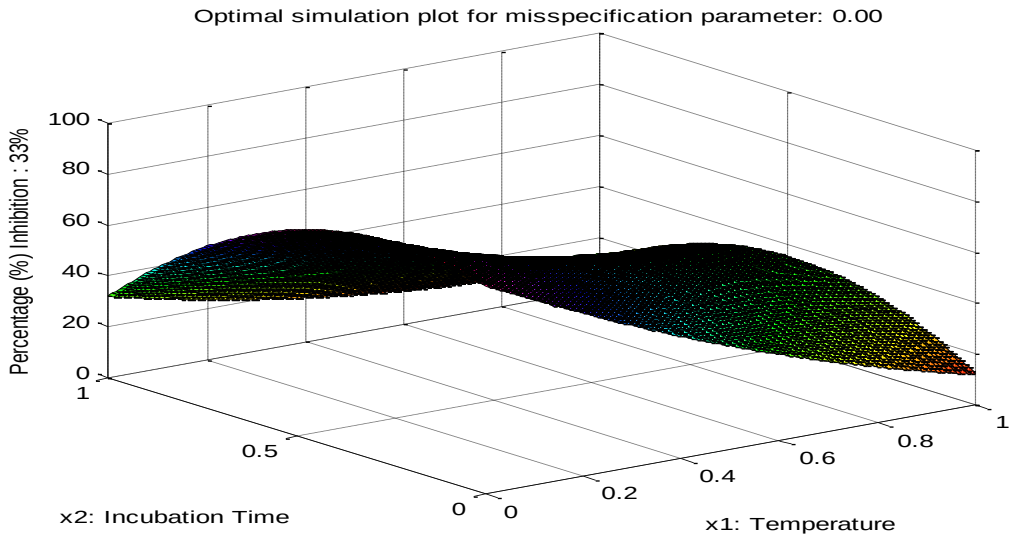

Figure 2: Simulated optimal response (33%) for QM Percentage $\beta - glucosidase$ inhibitory activity by *Cucumis Sativus* at $\gamma = 0.00$

Model 2 `Z= 81.97-95.07*x1+46.36*x2+48.52*x1.*x2+25.11*x1.^2-96.29*x2.^2 +MS*(3*sin(3*pi*x1)-2*cos(2*pi*x2)+2*sin(4*pi*x1.*x2))`
$MS = \gamma = 0.25$

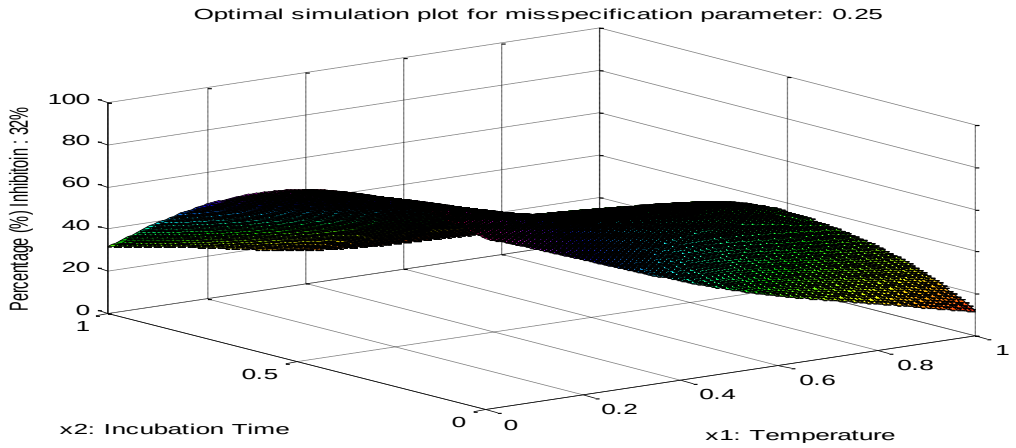

Figure 3: Simulated optimal response (32%) for QM Percentage $\beta - glucosidase$ inhibitory activity by *Cucumis Sativus* at $\gamma = 0.25$

Model 3 `Z=81.97-95.07*x1+46.36*x2+48.52*x1.*x2+25.11*x1.^2-96.29*x2.^2+MS*(3*sin(3*pi*x1)-2*cos(2*pi*x2)+2*sin(4*pi*x1.*x2))`
$MS = \gamma = 0.50$

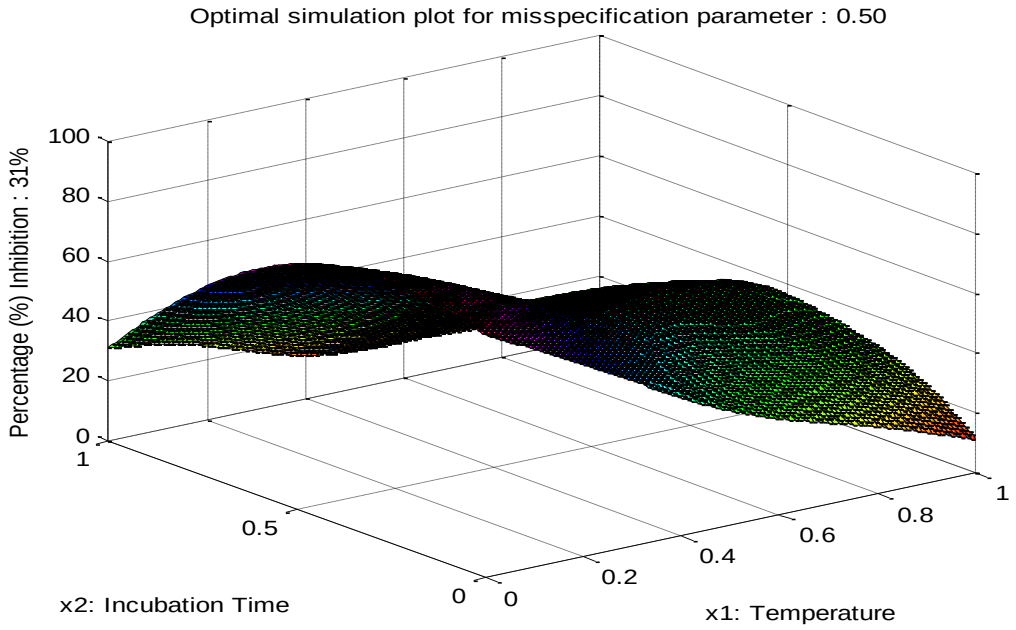

Figure 4: Simulated optimal response (31%) for QM Percentage $\beta - glucosidase$ inhibitory activity by *Cucumis Sativus* at $\gamma = 0.50$

Model 4 `Z=81.97-95.07*x1+46.36*x2+48.52*x1.*x2+25.11*x1.^2-96.29*x2.^2+MS*(3*sin(3*pi*x1)-2*cos(2*pi*x2)+2*sin(4*pi*x1.*x2))`

$MS = \gamma = 0.75$

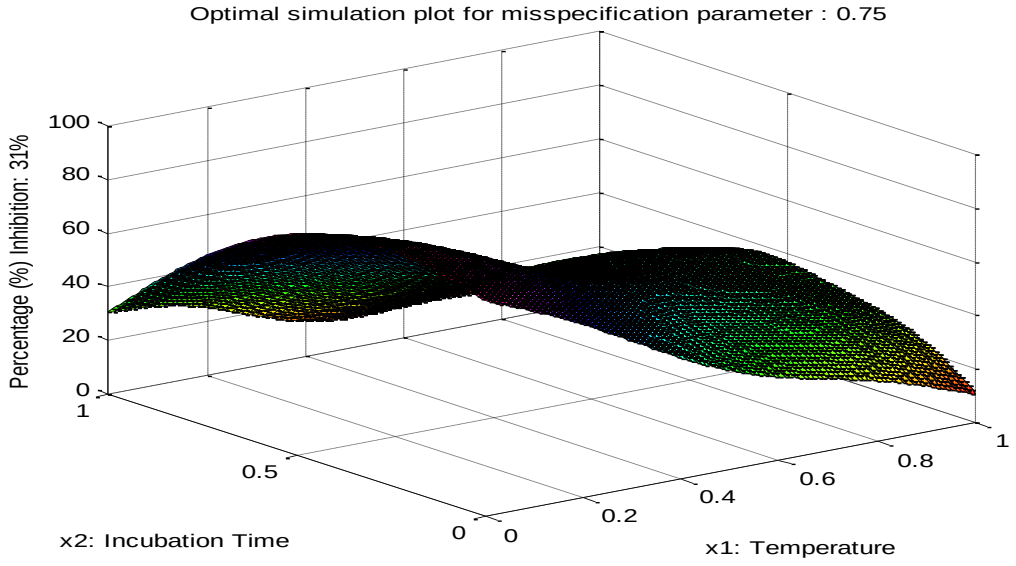

Figure 5: Simulated optimal response (31%) for QM Percentage $\beta - glucosidase$ inhibitory activity by *Cucumis Sativus* at $\gamma = 0.75$

Model 5 `Z=81.97-95.07*x1+46.36*x2+48.52*x1.*x2+25.11*x1.^2-96.29*x2.^2+MS*(3*sin(3*pi*x1)-2*cos(2*pi*x2)+2*sin(4*pi*x1.*x2))`

$MS = \gamma = 1.0$

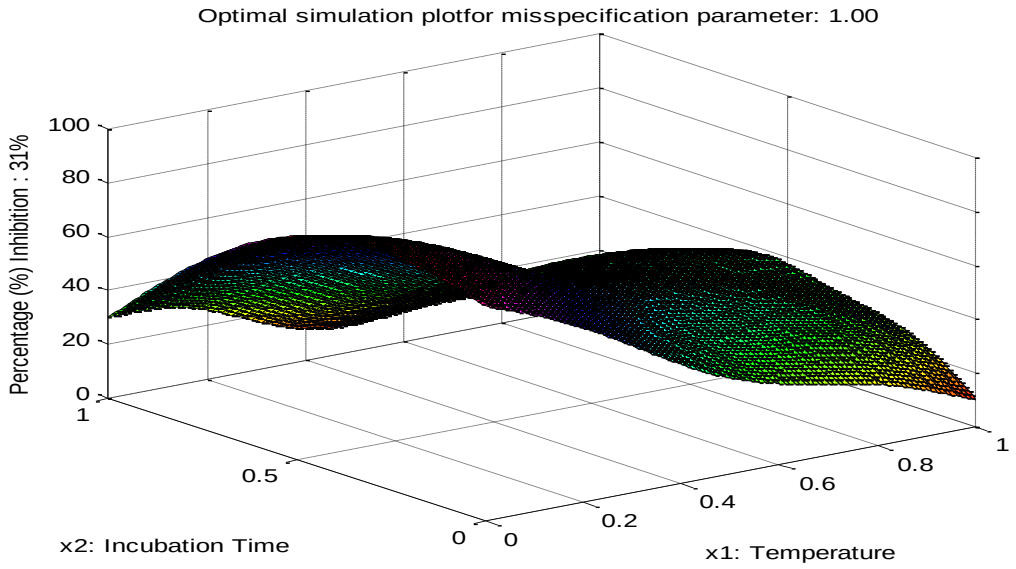

Figure 6: Simulated optimal response (31%) for QM Percentage $\beta - glucosidase$ inhibitory activity by *Cucumis Sativus* at $\gamma = 1.00$

4.2.2. SUMMARY OF SIMULATION STUDY 1

The estimated optimal percentage of $\beta - glucosidase$ inhibitory activity from *Cucumis Sativus*, derived using a genetic algorithm, was used to simulate two explanatory variables exhibiting positive interactive effects. This approach aimed to assess the stability of the regression model under the misspecification parameter γ, constrained within the range $0 \leq \gamma \leq 1$. Notably, as γ varies, the $\beta - glucosidase$ inhibitory activity remains stable, reflecting consistent positive interactive effects between temperature and incubation time.

Table 9. Optimal $\beta - glucosidase$ inhibition activity

Simulation Model	γ	Optimal $\beta - glucosidase$ inhibition activity (%)
Model 1	0.00	33
Model 2	0.25	32
Model 3	0.50	31
Model 4	0.75	31
Model 5	1.00	31

4.3.1. SIMULATION STUDY 2

The estimated optimal response derived from the genetic algorithm was employed to simulate the two explanatory variables, aiming to validate the stability of the regression model under the misspecification parameter (mixing parameter) γ, which was constrained to the interval $0 \leq \gamma \leq 1$.

Model 1 `Z=81.97-95.07*x1+46.36*x2+48.52*x1.*x2+25.11*x1.^2-96.29*x2.^2+MS*(2*sin(4*pi*x1)+2*cos(4*pi*x2)-2*sin(4*pi*x1*x2))`
$MS = \gamma = 0.0$

Therefore, the simulated plot gave optimal response of 37% when $\gamma = 0.0$ as given in Figure (7)

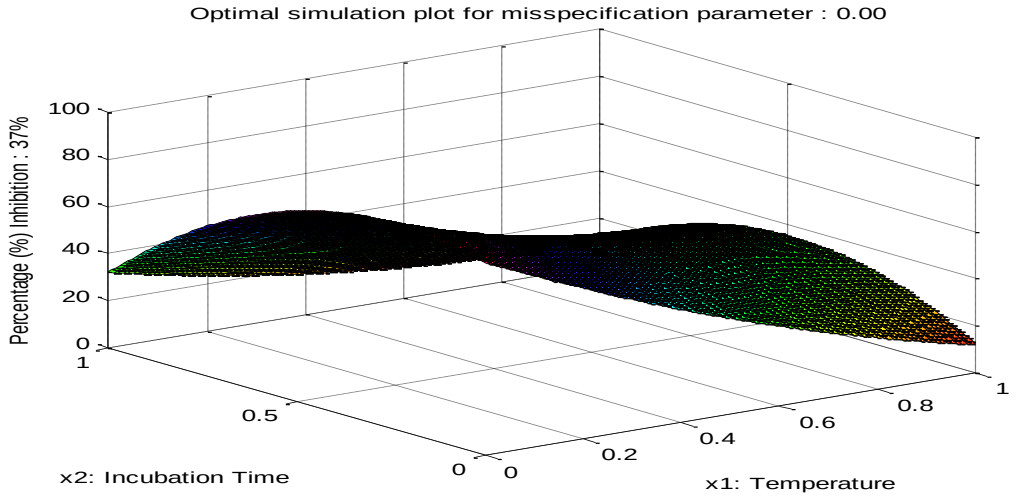

Figure 7: Simulated optimal response (37%) for QM Percentage $\beta - glucosidase$ inhibitory activity by *Cucumis Sativus* at $\gamma = 0.00$

Model 2 `Z=81.97-95.07*x1+46.36*x2+48.52*x1.*x2+25.11*x1.^2-96.29*x2.^2+MS*(2*sin(4*pi*x1)+2*cos(4*pi*x2)-2*sin(4*pi*x1*x2))`
$MS = \gamma = 0.25$.
Therefore, the simulated plot gave optimal response of 37% when $\gamma = 0.25$ as given in Figure (8)

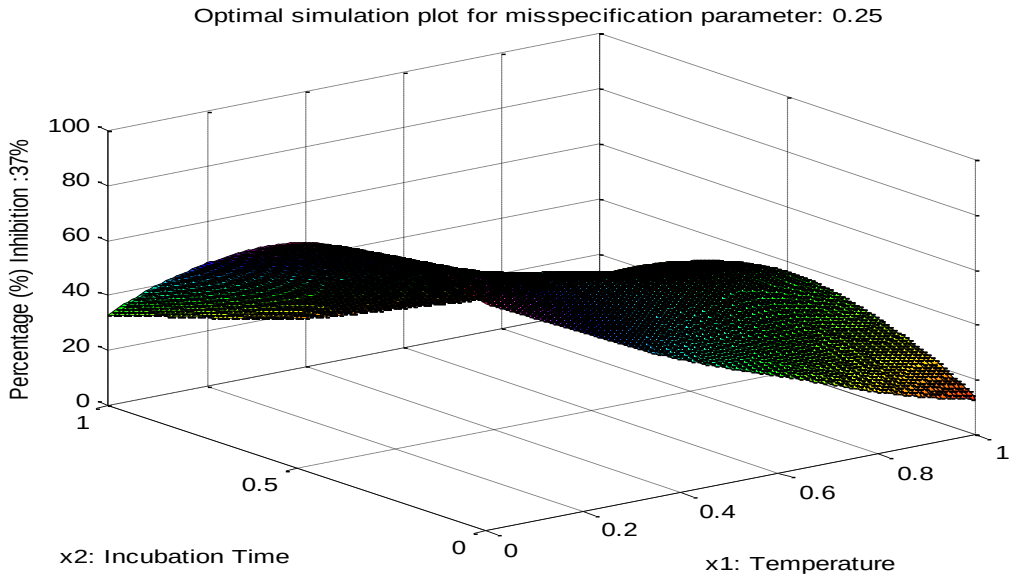

Figure 8: Simulated optimal response (37%) for QM Percentage $\beta - glucosidase$ inhibitory activity by *Cucumis Sativus* at $\gamma = 0.25$

Model 3 `Z=81.97-95.07*x1+46.36*x2+48.52*x1.*x2+25.11*x1.^2-96.29*x2.^2+MS*(2*sin(4*pi*x1)+2*cos(4*pi*x2)-2*sin(4*pi*x1*x2))`

$MS = \gamma = 0.50$

Therefore, the simulated plot gave optimal response of % when $\gamma = 0.50$ as given in Figure (9)

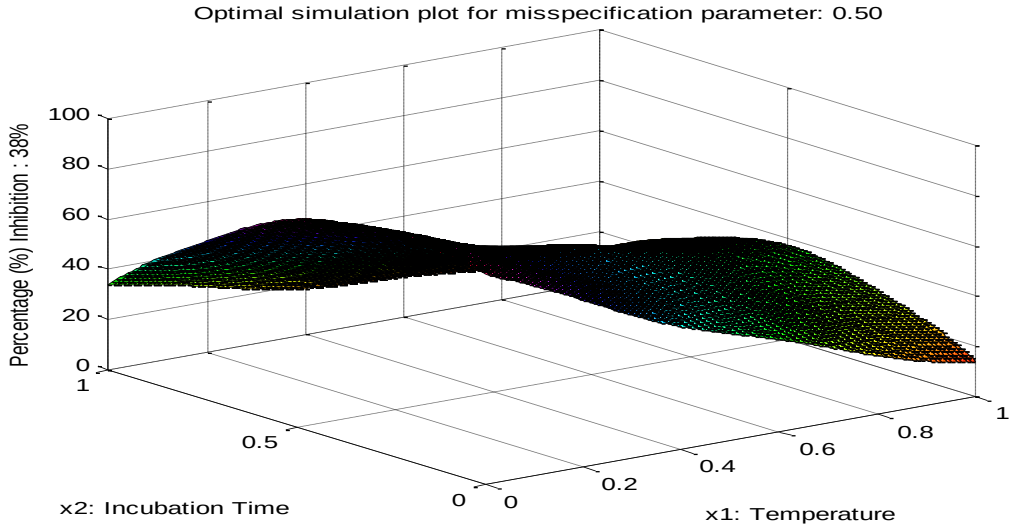

Figure 9: Simulated optimal response (38%) for QM Percentage $\beta - glucosidase$ inhibitory activity by *Cucumis Sativus* at $\gamma = 0.50$

Model 4 `Z=81.97-95.07*x1+46.36*x2+48.52*x1.*x2+25.11*x1.^2-96.29*x2.^2+MS*(2*sin(4*pi*x1)+2*cos(4*pi*x2)-2*sin(4*pi*x1*x2))`

$MS = \gamma = 0.75$

Therefore, the simulated plot gave optimal response of % when $\gamma = 0.75$ as given in Figure (10)

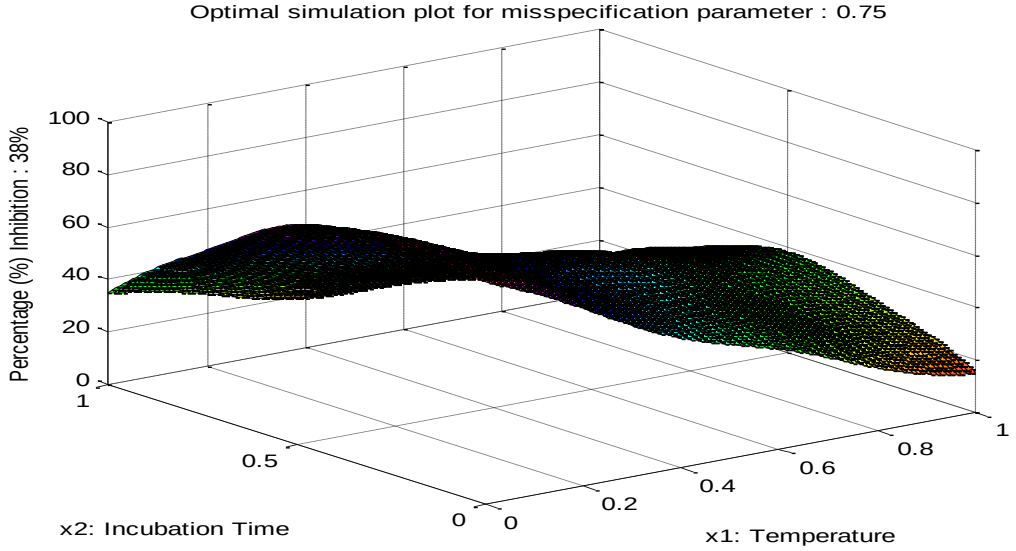

Figure 10: Simulated optimal response (38%) for QM Percentage $\beta - glucosidase$ inhibitory activity by *Cucumis Sativus* at $\gamma = 0.75$

Model 5 `Z=81.97-95.07*x1+46.36*x2+48.52*x1.*x2+25.11*x1.^2-96.29*x2.^2+MS*(2*sin(4*pi*x1)+2*cos(4*pi*x2)-2*sin(4*pi*x1*x2))`
$MS = \gamma = 1.00$

Therefore, the simulated plot gave optimal response of 38% when $\gamma = 1.00$ as given in Figure (11)

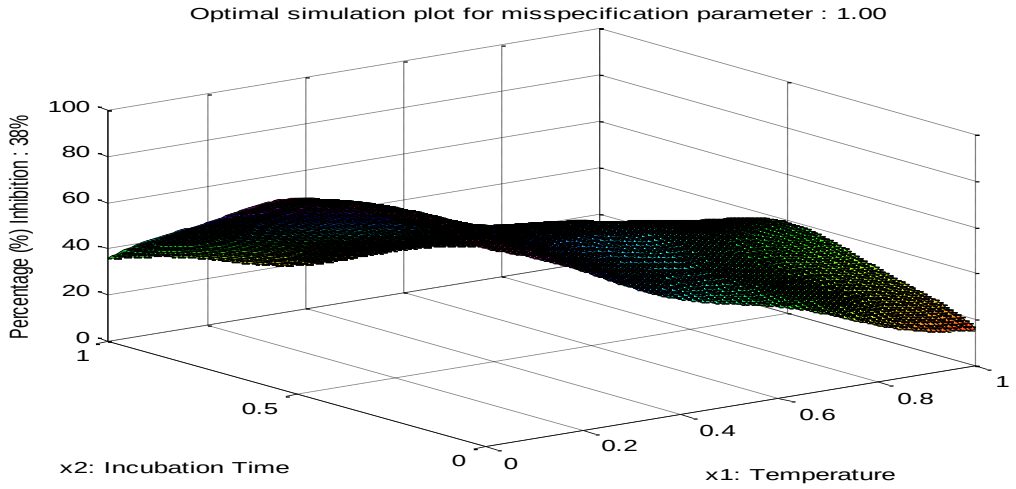

Figure 11: Simulated optimal response (38%) for QM Percentage $\beta - glucosidase$ inhibitory

activity by *Cucumis Sativus* at $\gamma = 1.00$

4.3.2. SUMMARY OF SIMULATION STUDY 2

The optimal estimated percentage of $\beta - glucosidase$ inhibitory activity from *Cucumis Sativus*, obtained through a genetic algorithm, was utilized to simulate two explanatory variables with positive interactive effects. This method aimed to evaluate the stability of the regression model under the misspecification parameter γ, restricted to the interval $0 \leq \gamma \leq 1$ Remarkably, the β-glucosidase inhibitory activity demonstrated stability across variations in γ, indicating consistent positive interactions between temperature and incubation time.

Table 10. Optimal $\beta - glucosidase$ inhibition activity

Simulation Model	γ	Optimal $\beta - glucosidase$ inhibition activity (%)
Model 1	0.00	37
Model 2	0.25	37
Model 3	0.50	38
Model 4	0.75	38
Model 5	1.00	38

5. CONCLUSIONS

The extraction of bioactive compounds with anti-diabetic potential from *Cucumis Sativus* (commonly known as cucumber) was optimized using Response Surface Methodology (RSM). This study focused on maximizing β-glucosidase inhibitory activity, a key indicator of anti-diabetic efficacy, by systematically evaluating the effects of critical process parameters, including extraction temperature, incubation time, and solvent concentration.

A central composite design (CCD) was employed to generate a predictive regression model, allowing for the analysis of linear, quadratic, and interactive effects of the independent variables on the response. The model demonstrated strong statistical significance with a high $R^2 = 82.18\%$, indicating good fit and reliability. The interactive effects between temperature and incubation time were found to be significantly positive, enhancing the inhibitory activity.

To further refine the optimization and ensure robustness of the model under uncertain conditions, a genetic algorithm (GA) was integrated with the RSM framework. This hybrid approach identified the optimal extraction conditions ($Temp = 27.20$℃, $Incubation\ Time = 15.48h, Agitation\ speed = 150rpm, Solvent = 6.30ml$ that yielded the maximum β-glucosidase inhibition of $\hat{y} = 136.55$. Simulations were conducted under a range of values for the misspecification parameter γ ($0 \leq \gamma \leq 1$), testing model sensitivity and stability. Results confirmed that the inhibitory activity remained stable across variations in γ, highlighting the resilience of the extraction process and the consistency of interactive effects.

Overall, this study demonstrates that combining RSM with genetic algorithm-based simulation provides an effective strategy for optimizing and validating the

extraction of anti-diabetic compounds from *Cucumis Sativus*. The findings offer a valuable foundation for the development of natural therapeutic agents targeting diabetes mellitus.

ACKNOWLEDGMENT

I am deeply grateful to my PhD supervisor, Prof. J. I. Mbegbu, for his guidance and unwavering support.

CONFLICT OF INTEREST

The authors stated that they have no conflicts of interest to disclose.

References

Ali, H., Houghton, P. J., and Soumyanath, A. (2006). α-Amylase inhibitory activity of some Malaysian plants used to treat diabetes; with particular reference to Phyllanthus amarus. Journal of Ethnopharmacology, 107(3), 449– 455.

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*.

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

Dhanani, T., Shah, S., Gajbhiye, N. A., and Kumar, S. (2017). Effect of extraction methods on yield, phytochemical constituents and antioxidant activity of Withania somnifera. Arabian Journal of Chemistry, 10, S1193– S1199.

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). 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,50(6), 1283 - 1309.*

Grover, J. K., Yadav, S., and Vats, V. (2002). Medicinal plants of India with antidiabetic potential. Journal of Ethnopharmacology, 81(1), 81–100.

IDF (International Diabetes Federation). (2021). IDF Diabetes Atlas, 10th ed.

Jamal, P., Barket, A.A. and Amid, A. (2011). Response surface optimization of the process conditions for anti-diabetic compounds from *Cucumis Sativus*. African Journal of Biotechnology, 10(81), 18788-18794

Marrelli, M., Qualtieri, A. and Conforti, F. (2016). Hypoglycemic agents from natural sources: a review on phytochemicals targeting key enzymes involved in type 2 diabetes. Current Medicinal Chemistry, 23(10), 1073–1095.

Montgomery, D. C. (2017). Design and Analysis of Experiments (9th ed.). Wiley.

Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M. (2016).

Response Surface Methodology: Process and Product Optimization Using Designed Experiments (4th ed.). Wiley.

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

Norshazila, S., Syed Zahir, I., Mustapha Suleiman, K., Aisyah, M. R., and Kamaruzaman, S. (2010). Antioxidant levels and activities of selected seeds of Malaysian tropical fruits. Malaysian Journal of Nutrition, 16(1), 149–159.

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

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*

Rathore, S., Prasad, S., and Sharma, R. R. (2011). Effect of ethylene and modified atmosphere packaging on shelf life and quality of cucumber (*Cucumis Sativus L.*) fruit. Journal of Food Science and Technology, 48(1), 42–48.

Rivers, D. L. (2009). A Graphical Analysis of Simultaneously Choosing the Bandwidth and Mixing Parameter for Semiparametric Regression Techniques, M. Sc. Thesis and Dissertations, Virginia Commonwealth University.

Sarkar, D., Sharma, A., Talukdar, P., and Chakraborty, M. (2012). Free radical scavenging activity and phytochemical analysis in the fruit extract of *Cucumis Sativus L*. International Journal of Pharmacy and Pharmaceutical Sciences, 4(1), 487–489.

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.*