# Modeling Anaerobic Decomposition: JMP Application with Biomass Data

# A. M. Abubakar<sup>1,2,\*</sup>, N. Elboughdiri<sup>3, 4</sup>, A. Chibani<sup>5</sup>, E. C. Nneka<sup>6</sup>, M. U. Yunus<sup>2</sup> and D. Ghernaout<sup>3, 7</sup>

 <sup>1</sup>Department of Chemical Engineering, Modibbo Adama University, Girei LGA, Adamawa State, Nigeria
 <sup>2</sup>Department of Chemical Engineering, University of Maiduguri, Borno State, Nigeria <sup>3</sup>Chemical Engineering Department, University of Ha'il, Ha'il, Saudi Arabia
 <sup>4</sup>Chemical Engineering Process Department, University of Gabes, Gabes, Tunisia
 <sup>5</sup>Research Center in Industrial Technologies CRTI, Cheraga, Algiers, Algeria
 <sup>6</sup>Department of Chemical Engineering, Faculty of Engineering, Chukwuemeka Odumegwu Ojukwu University, Igbariam, Anambra State, Nigeria
 <sup>7</sup>Chemical Engineering Department, University of Blida, Blida, Algeria \*Corresponding author: abdulhalim@mau.edu.ng

> Received 23/11/2023; accepted 18/04/2024 https://doi.org/10.4152/pea.2025430604

#### Abstract

Modern predictive modeling techniques, such as regression, NN and decision trees can be used to build better and more useful models. JMP 17.2.0 was used in this study to develop a fitting model for microbial growth observed data from chicken manure and banana peels labelled as Sample A, and a single chicken manure substrate, identified as Sample B. Statistical metrics, including COD (R<sup>2</sup>), RASE, MAD, negative log-likelihood and SSE were used to determine best predictions for Ct (X) from biomass of 22 and 24 samples (A and B) on SC (S) and SGR of microorganisms ( $\mu$ ). Along with estimated Monod parameters, TanH function SAS codes for 3 declared hidden layers, also demonstrated by surface plots, portrayed Sample B predicted model as the best one, even though the 2 samples datasets R<sup>2</sup> values for training (A: 0.9887916 and B: 1.0000) and validation (A: 0.9787637 and B: 0.9999999) pointed to a good fit. According to findings, optimal conditions for datasets were: A- biomass = 899868717 mg/L and SC =  $4.62 \times 10^9 \text{ mg/L}$ , correspondent to high  $\mu$  (0.010201 h<sup>-1</sup>); and B- biomass = 15351147 mg/L and SC = 9.2322 x 10<sup>9</sup> mg/L, consistent with  $\mu$  of 0.007316 h<sup>-1</sup>. RMSE, which is the standard method of choice for evaluating the accuracy of predictive models, including those based on NN, should be activated in future studies. This research is both timely and relevant in the pursuit of sustainable waste management and renewable energy generation.

Keywords: ANN; JMP; Monod; SAS code; SGR; SC; TanH function.

#### Introduction•

Decomposition of organic wastes in an oxygen-free environment requires anaerobic microorganisms, which feed on the nutrients it contains, thereby converting it to biofuel. The development of several microbial growth kinetics model has helped explaining this process. Inherent to most models, such as Andrew, Contois, Monod, Moser and Verhulst, are independent factors like biomass and SC, in the bioreactor under consideration, where the microorganism's SGR is the output variable. Since exponential microbial growth plays a key role in organic waste conversion, in most environmental engineers' agenda it is crucial to optimize their development, especially during biodegradable waste anaerobic decomposition. Kinetic models have been employed in recent years, of which parameters were used to predict biogas optimal production and the growth rate of isolated microbes that help digesting feedstock. Others scholars have employed RSM to generate DOE, proposing different combinations of input variables for single or sets of responses or outputs. Currently, NN predictive modeling has evolved as a new optimization tool in bioprocessing, for biogas yield prediction, with random combination of factors, from palm oil mill effluents, cassava wastewater, cow dung and food waste, etc., using MATLAB [2, 3, 9, 12].

Thorough research shows that prior studies using typical microbial growth kinetic parameters for biomass and SC are non-existent, especially a practical application of JMP software [7] linked to the anaerobic process. Applying NN predictive modeling technique to banana peels and chicken manure anaerobic decomposition can provide valuable insights onto processing behavior and enhance its management, leading to more efficient and reliable biogas and biofertilizer production. This study aimed to: use two different digesters separately charged with Sample A (banana peels + chicken manure) and Sample B (chicken manure) for microbial growth study; mathematically and experimentally determine and generate a dataset for the respective depleting Ct of SC and biomass in the exponential growth phase in their respective chambers, to calculate their microbes SGR; use datasets distinctly recorded in JMP software, specifying an orthogonal CCD, before programming it to generate a neural predictive model; analyze SGR statistical outputs of Sample A and B, for training and validation fitted plots; employ OriginPro 2018 software for a user-defined regression analysis, specifying the Monod equation, to estimate kinetic parameters for each dataset; and compare fits, surface plots, kinetic parameters and the predictive model for the two datasets, to determine the optimal combination of input variables. Similar to this work, [8] have carried out a combined demonstration of kinetics (using modified Gompertz model) and ANN, for biogas production via anaerobic digestion observations. [5] have also examined the behavior of biogas yield curve from lignocellulosic material using ANN. The present study suggests a different and more enhanced approach of merging the idea behind regression, CCD for RSM and neural modeling, to achieve the stated goals.

<sup>•</sup> The abbreviations list is in pages 392-393.

## Methodology and materials

# Materials sourcing

Banana peels and chicken manure were collected from Kasuwan Shanu Market and UNIMAID Faculty of Agriculture Poultry Farm, respectively, both in Maiduguri, Borno State, Nigeria. Drinkable water was utilized for anaerobic digestion and cell density determination. Before digestion, organic biomass was processed into Sample A and Sample B. Sample A is a mixture of chicken manure (4 kg), banana peels (0.5 kg) and water (7.5 kg) in a digester. Sample B is a mixture of chicken manure (7.5 kg) and water in equal proportions by weight, decomposed in a different digester. The two digesters were operated in batch mode, as described by [17].

#### **Biomass and SC datasets**

Viable cells in samples A and B digesters were determined using hypothetical unitscolony-forming units (CFU/mL) converted to SC in mg/L units. In microbiology and cell growth experiments, biomass means the viable cells counted in colonies, which stand as Ct from biomass. This determination was carried out by manual NA preparation. From the experiment onset, it was assumed that chicken manure contains microorganisms which are paramount to the anaerobic process [10]. Using a biomass-to-SC ratio of Y = 400, SC was determined using Eq. 1 [1] for samples A and B, respectively.

$$S = S_0 - \frac{X - X_0}{Y} \tag{1}$$

Depletion of nutrients or feedstock by biomass during anaerobic fermentation, or the feedstock amount left at a particular time, are typified as S. In Eq. 1,  $X_0$  and  $S_0$  represent initial biomass and SC, which are 899868717.4 and 4620000000 mg/L, (Sample A), and 15351147.09 and 9232210402.48 mg/L, respectively (Sample B). By convention,  $\mu$  and S in Monod plots are often extracted from the growth phase of the microorganism acting on the samples. Hence, datasets for S and  $\mu$  are inverted, initiating at S = 0 and  $\mu$  = 0, before plotting Monod curve (Figs. 1 and 2 (b)).



Figure 1: SGR against X and S (sample A).



Figure 2: SGR against X and S (sample B).

# RSM

JMP® Trial 17.2.0 (701896), Serial Number: T-VHFY3P0J09 was installed on Microsoft Windows 10 Pro (10.0.10240.0). The software was developed by JMP Statistical Discovery LLC, by Neil Hodgson (neilh@scintilla.org). Under DOE menu, and Classical drop-down menu, Response Surface Design was selected. For RSM analysis of Sample A and B, the response ( $\mu$ ), called SGR (h<sup>-1</sup>) was to be maximized. The specification of upper and lower limits of X and S factors is shown in Figs. 3 and 4.

File E	Edit Tables	Rows	Cols	DOE	Analyze	Graph	Tools	View	Window	Help		
1 💌 R	lesponse S	urface	Desig	m								
⊿ Re	esponses											
A	dd Response	- Rem	ove	Numb	er of Respo	onses						
	Response	Name			Goal		Lower	.imit	Upper	Limit	Importance	Un
m	mu			Maximize	aximize		0.0				/hr	
⊿ Fa	actors											
	Add 1	Contin	uous									
Re	emove Selecte	:d										
	Name		Role			Valu	Jes				Unit	ts
1	X Continuous 0			899868717.4 n		mg/l	8					
	S Continuous 0				4620	0000000	ma/l					

Figure 3: Sample A analysis -factor boundary values and response goal specification.

ile Edit Tables	Rows C	ols DOE	Analyze	Graph	Tools	View	Window	Help		
Response S	urface D	esign								
A Responses										
Add Response		e Numb	er of Resp	onses						
Response	Name		Goal		Lower	Limit	Upper	Limit	Importance	Uni
mu			Maximize		0		0.007	32		/hr
⊿ Factors										
Add 1	Continuo	us								
Remove Selecte	d									
Name	R	ole		Valu	ies				Uni	its
<b>⊿</b> X	X Continuous		us 0		D D		15351147.09		mg/	L
	C	ontinuous		0			9232	210402.4	8 ma/	L

Figure 4: Sample A analysis-factor boundary values and response goal specification.

Then, CCD-Orthogonal Design Type, over 16 runs and 8 center points, was chosen, and a randomized output option run order was selected. A table of the runs that was generated by JMP was then assessed for implementation. If the estimates of X and S given in the table are not realistic, those empirically obtained (Figs. 1 and 2) are entered after JMP proposed outcomes are deleted.

#### Incorporating a NN design

Using sets of empirical X, S and  $\mu$ , which were used to visualize microbial growth via Figs. 1 and 2, NN was generated for  $\mu$  outcomes. To do that, NN was selected under 'Predictive Modeling' drop-down arrow, under "Analyze" menu in JMP application. NN is a modern predictive modeling technique that predicts the response variable, using a flexible function of input variables (e.g., X and S). The respective factors and the response were then chosen, and a Random Seed of 0 was defined, to generate a reproducible sequence of random numbers, starting from 0. In Model Launch window, a Holdback Validation Method used to assess the performance of a trained model was selected. A default Holdback Proportion of 0.333, referring to a fraction of the dataset reserved for validation/testing purposes, or a determinant of how the dataset is split into training, validation and test sets, were allowed, as conducted by [9]. Afterwards, 3 was entered as the number of hidden nodes, and the model was launched using 'Go' button.

Expected results after the model run were: samples A and B dataset statistical parameter estimates; NN; prediction profiler model interpretation and sensitivity analysis representations; 3D surface profiles [16]; Actual by Prediction plots; a table containing formulas for the predicted response and hidden layers' nodes; and SAS code that can be used to score a new dataset. Using prediction profiler, optimal settings for predictor variables that led to desired predicted outcomes were found, in order to optimize the process. Based on 40 experimental runs, JMP simulated several DOE for microbial growth process involving X, S and  $\mu$  average estimates.

#### Determination of microbial growth optimization parameters

Parameters in basic microbial growth models were determined for observed datasets and NN model predicted values, by performing regression based on Monod equation. Predicted parameters were then assessed based on  $\mu$  optimal prediction.

# **Results and discussion**

# Training/validation and statistical measure based on predicted $\mu$

A holdback validation method is a crucial technique for assessing and fine-tuning NN models, which helps to prevent overfitting, and ensure they generalize well to new and unseen data. A holdback of 0.333, as specified, implies that 66.7% of the data were allocated for training, while 33.3% was set aside for validation or testing. Commonly, the specific choice of the holdback proportion depends on the nature of the problem, the dataset size, and the goals of the machine learning experiment. In the regression analysis context,  $R^2$  is a metrics often used for training and validation

datasets [3, 13]. Higher  $R^2$  of 0.9888 and 1.0000 for training, and lower  $R^2$  of 0.9788 and 0.9999 for validation, obtained for Samples A and B statistical metric (Table 1), respectively, indicated overfitting.

	-		• •			
Маалина	San	iple A	Sample B			
Measure	Training value	Validation value	Training value	Validation value		
R <sup>2</sup>	0.9887916	0.9787637	1	0.9999999		
RASE	0.0002279	0.0004126	1.8517 x 10 <sup>-7</sup>	4.8578 x 10 <sup>-7</sup>		
Mean abs dev	0.0001933	0.0002962	1.5785 x 10 <sup>-7</sup>	4.2239 x 10 <sup>-7</sup>		
–Log likelihood	-97.5472	-50.99224	-225.3291	-104.9486		
SSE	7.2715 x 10 <sup>7</sup>	1.3621 x 10 <sup>-6</sup>	5.486 x 10 <sup>-7</sup>	1.888 x 10 <sup>-12</sup>		
Sum freq	14	8	16	8		

**Table 1:** Sample A and B statistical model fitting predictions.

Overfitting occurs when the model is able to perfectly fit training data, but fails to generalize them to unseen data. In such cases, the model may capture noise and idiosyncrasies in training data, which results in high R<sup>2</sup> value for training. When applied to validation dataset, the model performs poorly, resulting in lower R<sup>2</sup> value. Viz-a-viz, higher R<sup>2</sup> for validation and lower R<sup>2</sup> for training are often a sign of underfitting. However, if R<sup>2</sup> values for both training and validation are reasonably high and close to each other, as shown in Table 1, this indicates that the model is able to capture underlying patterns in the data without overfitting or underfitting [6]. SSE can be explained in terms of training and validation. Training SSE close to zero implies that the model is perfectly fitting training data, while lower validation SSE indicates better generalization. Clearly, results in Table 1 establish an ideal scenario, where validation SSE is reasonably close to the training one, for the respective samples, it shows that the model learnt underlying patterns in the data without overfitting. Similarly, lower training RASE of 0.0002279 (Sample A) and 1.8517 x 10<sup>-7</sup> (Sample B), shows that the model predicted training data with smaller errors.

MAD or MAE behavior, which quantifies how far predictions are from true values on average, is consistent whether one is looking at training or validation dataset. If  $R^2$  is high for training dataset, MAD will typically be lower on it (e.g., 0.0001933 for Sample A). Alternatively, low  $R^2$  for validation dataset will result in higher MAD on validation dataset (viz. 0.0002962). As for equal  $R^2$  (dataset B), MAD's estimates can be used to select models. In order to minimize prediction errors in practical applications where accuracy is crucial, lower MAD (i.e., 1.5785 x 10<sup>-7</sup> for training) is suitable. Sum of frequencies show how many data points are present in each dataset, which is crucial in assessing reliability and generalizability of NN predictive model. Larger datasets (as in Sample B) are often beneficial for training more robust and accurate models. The difference itself in "Sum Freq" between training and validation datasets does not inherently indicate an issue. "Loglikelihood" is a term used to quantify how well a statistical model's predictions match observed data, in which lower values (i.e., -97.5472 and -225.3291 for Samples A and B training, respectively) indicate a better fit.

#### Hidden layers' structure

In a NN, hidden nodes, as earlier specified in the methodology, are computational units that are part of hidden layers, which stand between input (X and S blue box) and output layers ( $\mu$  square box), as shown in Fig. 5.



Figure 5: NN configuration for predicting  $\mu$  from sample A and B datasets.

Hidden nodes are responsible for processing input data and learning complex patterns and representations from them [2]. Hidden layers with 3 hidden nodes mean that this layer contains 3 computational units. The units in the green field circle containing 'S' symbol, as shown in Fig. 5, are called TanH [14].

Due to selected alike nodes, Fig. 5 is similar for samples A and B data. Changing the number of hidden nodes from 3 to 5 and 10, for Sample A dataset, will affect the extent of fit and R<sup>2</sup> values [4, 11]. At hidden nodes of 3, 4 and 5, R<sup>2</sup> values were: for training, 0.9887916, 0.9678015 and 0.9926753; and for validation, 0.9787637, 0.9786579 and 0.9786468. Initially, increasing hidden nodes number might lead to an improvement in  $R^2$  for training dataset, since a more complex model is able to learn intricate patterns on data. Further increase in nodes (to 10) will lead to overfitting, as the model starts to fit noise or random variations in training data, which results in artificially high  $R^2$ . Likewise, increasing hidden nodes may, at first, improve  $R^2$  for validation dataset, because the model can capture more nuances in data. Nevertheless, if the model starts to overfit training data, R<sup>2</sup> for validation dataset is likely to decrease. Generally, increasing the hidden nodes number can affect the dependent variable  $(\mu)$  prediction. NN with more hidden nodes is computationally more complex, and it may require additional computational power and memory. What is called "black box" may occur [2], a situation in which it has signally became challenging to interpret how specific features or variables impact  $\mu$ , due to more hidden nodes.

# Prediction profile under maximum desirability

In JMP software, prediction profiler is a tool used for exploring and visualizing X and S effects on predictions made by a statistical model. It is particularly useful for understanding how changes in input variables (predictors) impact predicted outcome ( $\mu$ ) from the model. As portrayed in Figs. 6 and 7, desirability refers to a measure used to assess the overall quality of a set of predicted outcomes for a given set of input conditions or factors.



Figure 6: Predicted SGR from sample A under X and S values' maximal desirability.

In Fig. 6, a high desirability score of 0.994319 indicates that a combination of X = 899868717 mg/L and  $S = 4.62 \text{ x } 10^9 \text{ mg/L}$  led to optimal or desirable values of  $\mu = 0.010201 \text{ h}^{-1}$ , which met the desired goal. This is because, when working with prediction profilers and desirability, the aim is to find factor settings or input conditions that maximize overall desirability score. Hence, in the same context, X = 15351147 mg/L and  $S = 9.2322 \text{ x } 10^9 \text{ mg/L}$ , which, at high

desirability of 0.981835, is the desired combination for  $\mu_{max}$  of 0.007316 h<sup>-1</sup> (Fig. 7).



Figure 7: Predicted SGR from sample b under maximal desirability of X and S values.

### Simulation experiment

'Simulation' feature, under prediction profiler in JMP software, was used to perform Monte Carlo simulations, for predicting X and S values, and statistical properties: mean and SD of  $\mu$  for 40 experimental runs (Tables 2 and 3).

Run	X	S (T)	µ mean	ц SD
	(mg/L)	(mg/L)	<u>(h)</u>	1.015.17
1	830648046.83	3968461538.50	0.008959737	1.21E-17
2	819111268.40	2310000000.00	0.006560012	1.04E-17
3	888331938.97	36130/6923.10	0.008/31199	1.21E-17
4	/84500933.12	402/69230/./0	0.00881895	1.21E-17
5	461471137.13	3316923076.90	0.00617417	1.65E-17
6	634522813.55	3020769230.80	0.006670966	2.60E-18
7	530691807.70	3790769230.80	0.007191932	1.04E-17
8	484544693.98	3139230769.20	0.00605546	8.67E-18
9	669133148.84	2783846153.80	0.006513329	1.73E-18
10	449934358.70	2606153846.20	0.005140131	0
11	611449256.69	4383076923.10	0.008428288	1.21E-17
12	715280262.55	4501538461.50	0.009129081	2.43E-17
13	599912478.27	4560769230.80	0.00860883	2.26E-17
14	646059591.98	3198461538.50	0.006976319	1.73E-17
15	703743484.12	3435384615.40	0.007595131	1.91E-17
16	576838921.41	308000000.00	0.006458603	1.73E-18
17	807574489.97	4146153846.20	0.009096647	1.21E-17
18	519155029.27	3909230769.20	0.007291284	1.73E-17
19	876795160.54	2487692307.70	0.007076688	5.20E-18
20	796037711.55	462000000.00	0.009698155	1.56E-17
21	772964154.69	2428461538.50	0.006513781	1.65E-17
22	692206705.69	2369230769.20	0.006043987	1.56E-17
23	565302142.98	2902307692.30	0.006153266	8.67E-19
24	507618250.84	3257692307.70	0.006340537	1.04E-17
25	588375699.84	4323846153.80	0.008225424	0
26	865258382.12	4205384615.40	0.009456864	0
27	542228586.13	2961538461.50	0.006115486	8.67E-18
28	853721603.69	2665384615.40	0.007225682	6.07E-18
29	473007915.56	3731538461.50	0.006800266	3.47E-18
30	899868717.40	3553846153.80	0.008700551	1.21E-17
31	657596370.41	4086923076.90	0.008263504	1.21E-17
32	680669927.26	4264615384.60	0.008626408	1.21E-17
33	622986035.12	3672307692.30	0.007514357	1.73E-17
34	761427376.26	4442307692.30	0.009280284	1.21E-17
35	553765364.55	2843076923.10	0.006011707	1.56E-17
36	749890597.83	3376153846.20	0.007739604	1.13E-17
37	496081472.41	3850000000.00	0.00708624	0
38	842184825.26	2546923076.90	0.007003912	1.73E-17
39	738353819.41	2724615384.60	0.006768006	1.04E-17
40	726817040.98	3494615384.60	0.007792039	1.91E-17

Table 2: JMP predicted experimental runs for sample A dataset.

Higher number of runs provides more precise estimates of  $\mu$ 's mean and SD. The relationship between  $\mu$  and limiting S given in the tables, over 40 runs, can be described using Monod equation (Eq. 2), for a typical controlled microbial environment to which Samples A and B were subjected to.

$$\mu = \frac{\mu_{max} S}{K_s + S} \tag{2}$$

where  $\mu_{max}$  is maximum SGR, i.e., rate at which microorganisms can grow when S is not limiting, and K<sub>s</sub> is half-saturation constant or S, at which  $\mu$  is half of  $\mu_{max}$ . Since there are no estimates of  $\mu_{max}$  and K<sub>s</sub>, to aid selecting an optimal design, a run with higher  $\mu$  in Tables 2 and 3 may hint on X and S best combinations. Thus, Sample A dataset's X = 899868717.4 mg/L and S = 4.62 x 10<sup>9</sup> mg/L (Table 2), at  $\mu_{max} = 0.009698155$  h<sup>-1</sup> below maximum desirability, as shown in Fig. 6. On the other hand, Sample B dataset's X = 15351147.09 mg/L and S = 9232210402.5 mg/L (Table 3), at  $\mu_{max}$  of 0.00728 h<sup>-1</sup> below maximum desirability, as shown in Fig. 7, since estimates were not practically experimented and compared with the ones given by JMP.

Table 3: JMP predicted experimental runs for sample B dataset.

Dun	Χ	S	µ mean	SD
Kull	(mg/L)	(mg/L)	· (h)	μδυ
1	14367099.20	7811870340.60	0.007152	5.67255E-16
2	13776670.47	5562998575.90	0.006648	1.76074E-16
3	12005384.26	8166955356.00	0.00713	6.65267E-16
4	8856431.01	8640402043.30	0.007108	5.73326E-16
5	8659621.44	7456785325.10	0.006924	1.04951E-16
6	13383051 31	6864976965 90	0.006965	2.6975E-16
7	11808574 69	5918083591 30	0.006676	4 25875E-16
8	9053240 59	5089551888 50	0.006274	5.96745E-16
0	12080422 15	6500801050 50	0.006274	6 24501E 16
9 10	12909432.13	0303031330.30	0.000873	0.243011-10
10	8200002.28	00//12330/.00	0.007122	9.200//E-1/
11	/8/2383.12	85220403/1.50	0.00/069	1.0850/E-15
12	1002//1/.22	3320273232.20 8005487058 80	0.000433	1.40399E-13 1.1162E-15
13	14170289.62	5799721919 50	0.007193	1.0365E-15
15	15154337 51	7930232012.40	0.007191	1.0505L-15 1.97759E-16
16	10234098.06	7575146996 90	0.006993	1.81105E-15
17	8462811.86	7220061981.40	0.006873	2.09902E-16
18	9250050.17	6154806935.00	0.006645	7.61544E-16
19	11611765.11	5681360247.70	0.006596	7.31186E-16
20	15351147.09	8758763715.20	0.00728	1.5786E-16
21	13973480.04	4852828544.90	0.006397	1.7486E-15
22	12595813.00	6273168606.80	0.006803	6.03684E-16
23	13579860.89	5444636904.00	0.006601	8.92515E-16
24	14957527.93	8285317027.90	0.007225	9.6971E-16
25	13186241.73	6983338637.80	0.006982	1.81279E-15
26	11414955.53	/101/00309.60	0.006945	1.00614E-16
27	12/92622.58	8048593684.20	0.00/13/	2.09902E-16
28	14563908.78	6628253622.30	0.006959	6.21031E-16
29	12202193.84	/693508668./0	0.00/0/1	1.39339E-13
30	/0/33/3.33	84030/8099./0	0.00/049	1.4/452E-1/
22	9043009.33	4010103201.20	0.006094	1.44303E-13 1.20021E-15
32	14760718 36	6301530278.60	0.000803	1.29951E-15 4.42355E-17
34	11021336 37	6036445263 20	0.000914	1 31319E-15
35	12399003 42	9232210402.50	0.007249	1 13798E-15
36	9840478.90	4971190216.70	0.00626	2.8276E-16
37	10430907.64	4734466873.10	0.006184	1.5786E-15
38	11218145.95	9113848730.70	0.007213	5.67255E-16
39	10037288.48	7338423653.30	0.006945	1.3262E-15
40	8069192.70	5207913560.40	0.006279	1.12497E-15

#### Surface profile

Just as previously demonstrated by [11, 15], predictors' (S and X) effect on the response can be simulated with NN model, and analyzed using 3D surface plots. Fig. 8 shows that  $\mu$  increased gradually with X, before attaining  $\mu_{max}$  of 0.015 h<sup>-1</sup>, and then declined, after nutrients became low. Likewise,  $\mu$  increased as S decreased and X rose, as shown in Fig. 9.



**Figure 8:** Surface plot of X, S and  $\mu$  showing patterns of changes in dataset A.



Figure 9: Surface plot of X, S and  $\mu$  showing patterns of changes in dataset B.

#### Fitted plots and model equation command code

In the specified fit, plot for training set with actual and predicted values on Y- and X-axes, respectively, and validation set are shown in Figs. 10 and 11. They are generated by JMP based on their number of data points or Sum Freq. or counts given earlier in Table 1.





Figure 10: (a)- Actual vs. predicted correlation based on dataset A.



Figure 11: Actual vs. predicted correlation based on dataset B.

Under training linear fitting plots in both figures, random data splitting process instigates more data points, to end up in the training dataset by chance. As previously described, MAD and R<sup>2</sup> are complementary metrics. High and low R<sup>2</sup> often correspond to lower and higher MAD, indicating better and worse model fits, respectively (Fig. 10a and b). Fig. 11, for chicken manure model fitting of  $\mu$ , shows that R<sup>2</sup> values for training and validation datasets are equally high (Table 1). This suggests that the model performed well in terms of explaining the variance in the target variable, and generalized effectively to unseen data [9]. This is actually a positive sign. However, if precise and accurate predictions are the analysis' priority, the training model with less errors (SSE = 5.486 x 10<sup>-13</sup>) should be chosen based on its lower MAD. A model with higher MAD (validation) should be selected if one wishes more sensitivity to potential outliers or extreme cases. Another reason for better fit in both training graphs is that their negative log-likelihoods are lower, while higher values in validation plots suggest poorer fit [7].

Figs. 12 and 13 depict Ntan(H) model equations for 3 hidden layers. Notably, [13] utilizes 2 hidden layers and both neural and fuzzy logic to optimize biogas yield from datasets of two reactor setups TanH function's advantage is that it is nonlinear, allowing to map more intricate functions and stack layers, and to obtain high precision [14].

A. M. Abubakar / Portugaliae Electrochimica Acta 43 (2025) 377-394



Figure 12: Model equation of  $\mu$  for sample A dataset.



Figure 13: Model equation of  $\mu$  for sample B dataset.

#### Experimental versus predicted $\mu$ datasets

In samples A and B (Figs. 14 and 15), increased  $\mu$  means microbial growth faster rate, which rapidly depletes S, as it further consumes the substrate. In that case,  $\mu$  in Monod equation represents microbial growth in relation to limiting S. Predicted values (Figs. 14 and 15), using model equations in Figs. 12 and 13, showed almost 100% fit demonstrated by the earlier shown plots. When regression was performed using OriginPro 2018 software,  $\mu_{max}$  and K<sub>s</sub> values obtained for experimental and predicted samples were almost equal (Table 4).

Although model equation for predicted  $\mu$  of A dataset indicated that it was reliable, resulting in values of  $\mu_{max} = 0.00997 \text{ h}^{-1}$  and  $K_s = 9.12975 \text{ x } 10^7 \text{ mg/L}$  (0.01 h<sup>-1</sup> and 9.498595 x 10<sup>7</sup> mg/L), the same parameters estimated for experimental results gave the best combination, coupled with their corresponding X and S based on maximal desirability. Model equation for Sample B accurately mimicked Eq. 2, since it gave 100% equal experimental and predicted  $\mu_{max}$  and K<sub>s</sub> (0.00762 h<sup>-1</sup> and 3.838 x 10<sup>7</sup> mg/L, respectively) values and suitable chicken manure data (Fig. 13). These parameters, along with X and S maximal desirability (Fig. 7), indicate that Sample B study was the optimal combination.

Run	X (mg/L)	S (mg/L)	mu (/hr)	H1_1	H1_2	H1_3	Predicted mu (/hr)
1	0	0	0	0.723263	-0.01976	-0.98171	-0.000884752
2	5076637617.00	38055555	0.00286	0.172238	-0.67866	0.535796	0.003520116
3	5024148704.00	48378093	0.003374	0.179851	-0.67418	0.512069	0.003798366
4	4958968838.00	61500609	0.00393	0.18928	-0.66854	0.481433	0.004132281
5	4878510894.00	78182597	0.004514	0.20088	-0.66148	0.441848	0.0045266
6	4779921767.00	99389560	0.005113	0.21503	-0.65267	0.390761	0.004982759
7	4660198965.00	12600000.00	0.005708	0.232108	-0.64173	0.325183	0.005496354
8	4516392491.00	16100000.00	0.006284	0.252486	-0.62825	0.241507	0.006057686
9	4345907884.00	20400000.00	0.006825	0.276382	-0.61177	0.137228	0.006643514
10	4146909860.00	26000000.00	0.007321	0.303946	-0.59187	0.010466	0.007228778
11	3918796046.00	33000000.00	0.007765	0.335022	-0.56817	-0.13624	0.007779432
12	3662669056.00	41900000.00	0.008154	0.369231	-0.54044	-0.29588	0.008268546
13	3381694060.00	53300000.00	0.008488	0.405894	-0.50869	-0.45637	0.008677732
14	3081209665.00	67800000.00	0.008771	0.444051	-0.47328	-0.60337	0.008995165
15	2768486086.00	86200000.00	0.009007	0.482595	-0.43492	-0.72572	0.009213515
16	2452105901.00	110000000.00	0.009202	0.520483	-0.39477	-0.81926	0.009337625
17	2141058774.00	139000000.00	0.009361	0.556559	-0.35411	-0.88483	0.009357154
18	1843742952.00	177000000.00	0.009491	0.590442	-0.31464	-0.92913	0.009339258
19	1567101512.00	225000000.00	0.009595	0.621659	-0.27777	-0.95757	0.009328058
20	1316070831.00	286000000.00	0.009679	0.650263	-0.24478	-0.97539	0.00940782
21	1093410185.00	364000000.00	0.009745	0.676613	-0.21671	-0.98634	0.009675515
22	899868717.40	462000000.00	0.009799	0.701075	-0.19415	-0.99282	0.010201279

Figure 14: Sample A's experimental dataset and predicted SGR.

Constant "H" in Figs. 14 and 15, as herein mentioned, likely represents a parameter in model equations used for predicting  $\mu$  of microorganisms based on X and S input variables. Positive "H" constant may suggest that certain conditions or factors it represents have a stimulatory effect on microbial growth, and indicate a direct relationship with  $\mu$ , meaning that they increased together. Conversely, negative "H" constant may imply inhibitory factors or conditions that negatively impact microbial growth, which indicates an inverse relationship with  $\mu$ . In this case, as "H" value decreased,  $\mu$  increased.

Parameters, such as  $\mu_{max}$  and K<sub>s</sub> allow scientists to predict how fast microbial growth is under specific nutrient Ct. It is quite important in this study, since  $\mu$  specifically impacts biogas generation, when Sample A or B are used, with their associated production rates.

Optimization of S in their respective bioreactors to maximize microbial growth is aided with the knowledge of these parameters. In the Monod equation, an increase in  $\mu_{max}$  would lead to a higher  $\mu$  that can be achieved under optimal conditions. An increase in  $\mu_{max}$  implied higher microbial growth potential, which could lead to faster substrate consumption and, consequently, a decrease in S, over time [1].

Run	X (mg/L)	S (mg/L)	mu (/hr)	H1_1	H1_2	H1_3	Predicted mu (/hr)
1	0	0	0	0.103758	0.232189	0.685976	2.56E-12
2	268892417.89	165180857.40	0.002293	-0.41633	0.99596	0.197786	0.0022931
3	253581226.33	198328228.73	0.002596	-0.39555	0.994338	0.247839	0.002596054
4	237353737.26	238127389.16	0.002918	-0.37368	0.991896	0.301097	0.002917681
5	220417927.77	285913174.50	0.003253	-0.3512	0.988203	0.356825	0.003253491
6	203024574.73	343288286.32	0.003598	-0.32878	0.982628	0.414156	0.003598238
7	185453576.79	412177045.47	0.003946	-0.30721	0.974274	0.472164	0.003946272
8	167996450.94	494889932.40	0.004292	-0.28741	0.961923	0.529943	0.004291895
9	150937246.68	594201079.08	0.00463	-0.27034	0.944024	0.586672	0.004629645
10	134534480.78	713441311.43	0.004955	-0.25706	0.918724	0.641654	0.004954519
11	119006462.74	856609862.84	0.005262	-0.24859	0.884003	0.694311	0.005262176
12	104521632.16	1028508505.70	0.005549	-0.24599	0.837861	0.744153	0.005549164
13	91194509.84	1234902599.40	0.005813	-0.25029	0.778548	0.790723	0.005813124
14	79086872.12	1482714456.40	0.006053	-0.2625	0.704735	0.833538	0.006052873
15	68213033.03	1780255512.00	0.006269	-0.28362	0.615577	0.872065	0.006268269
16	58547770.00	2137505083.50	0.00646	-0.31457	0.510651	0.905741	0.006459901
17	50035433.20	2566445069.90	0.006629	-0.35612	0.389857	0.934058	0.006628786
18	42599029.44	3081461816.30	0.006776	-0.40873	0.253415	0.956694	0.006776241
19	36148436.55	3699828621.50	0.006904	-0.47222	0.102127	0.973661	0.006903925
20	30587274.76	4442285072.70	0.007014	-0.54541	-0.06197	0.9854	0.007013863
21	25818268.04	5333732636.30	0.007109	-0.62568	-0.23466	0.992764	0.007108252
22	21747145.08	6404069835.60	0.007189	-0.70868	-0.4086	0.996865	0.007189113
23	18285257.16	7689195026.40	0.007258	-0.7886	-0.57337	0.998845	0.007257954
24	15351147.09	9232210402.50	0.007316	-0.85913	-0.7172	0.99965	0.007315504

Figure 15: Sample B's experimental dataset and predicted SGR.

	Sample A		Sample B			
	Experimental	Predicted	Experimental	Predicted		
Number of points	22	22	24	24		
Degrees of freedom	20	20	22	22		
Reduced chi-square	4.98155E-12	1.01704E-07	1.06662E-19	4.54345E-14		
Residual sum of squares	9.46310E-11	2.03408E-06	2.34655E-18	9.99560E-13		
$R^2$ (COD)	1	0.98769	1	1		
Adj. R <sup>2</sup>	1	0.98707	1	1		
Fit status	Succeeded (100)	Succeeded (100)	Succeeded (100)	Succeeded (101)		
$\mu_{max}$ (h <sup>-1</sup> )	0.01	0.00997	0.00762	0.00762		
$K_s(mg/L)$	9.49859E + 07	9.12975E + 07	3.838E + 08	3.83747E + 08		

 Table 4: Statistics and estimated Monod parameters.

# Conclusion

 $R^2$  values near or equal to 1, obtained on training and validation sets, show that TanH neural models for Samples A and B datasets correctly predicted  $\mu$ . Under 99.43%

desirability score,  $\mu_{max}$  of 0.01 h<sup>-1</sup> and K<sub>s</sub> of 9.498595 x 10<sup>7</sup> mg/L, the combination of X and S (899868717 mg/L and 4.62 x 10<sup>9</sup> mg/L, respectively), for optimal µ of 0.010201 h<sup>-1</sup>, was valid for Sample A's experimental observations. Estimated Monod parameters for  $\mu_{max}$  and K<sub>s</sub> (0.00762 h<sup>-1</sup> and 3.838 x 10<sup>7</sup> mg/L, respectively), based on 98.18% desirability for X and S (15351147 and 9 x 10<sup>9</sup> mg/L), made optimal microbial growth condition that resulted in maximum  $\mu$  of 0.007316 h<sup>-1</sup>, which was valid for Sample B empirical and predicted outputs. Overall, Sample B's anaerobic chamber produced higher biogas yields that those from Sample A's bioreactor. This was due to exact R<sup>2</sup> and kinetic parameter values obtained for experimental and forecasted  $\mu$ . In future researches, neural predictive modeling of two datasets, employing linear and Gaussian representative models, can be run using JMP, of which outcome may be compared with estimates in this study and with the created TanH equation. In addition, hidden nodes variation may improve R<sup>2</sup> estimation and predicted values, especially for Sample B estimates, which still falls below observed variables and obtained parameters. Applying advanced statistical tools like JMP and NN allows for data-driven decision making, which can result in optimized process parameters and improved waste-to-resource conversion.

#### **Declaration of interests**

Authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work herein reported.

## **Authors' contributions**

A. M. Abubakar: study conceptualization; integrated new ideas with existing literature; analyzed results using JMP and OriginPro; studied and discussed results. N. Elboughdiri: carried out literature survey; reviewed and edited the manuscript; provided project administration and support. A. Chibani: chose the research problem; gave outlines for the manuscript preparation. E. C. Nneka and D. Ghernaout: reviewed and edited the manuscript; guided the paper writing. M. U. Yunus: completed experimental part; helped in raw material sourcing.

# Abbreviations

ANN: artificial neural network CCD: central composite design COD: coefficient of determination (R<sup>2</sup>) DoE: design of experiment MAD: mean absolute deviation MAE: mean absolute error NA: nutrient agar NN: neural network RASE: root average squared error RMSE: root mean square error RSM: response surface methodology SAS: statistical analysis system SC: substrate concentration SD: standard deviation SGR: specific growth rate SSE: sum of squared errors TanH: hyperbolic tangent

# References

- 1. Abubakar AM, Soltanifar Z, Kida MM. Sensitivity analysis of kinetic growth model data: Monod equation. Schol: J Nat Med Educ. 2022;1:1-13. https://univerpubl.com/index.php/scholastic/article/view/17
- Chen W-Y, Chan YJ, Lim JW et al. Artificial neural network (ANN) modelling for biogas production in pre-commercialized integrated anaerobic-aerobic bioreactors (IAAB). Water. 2022;14(1410):1-36. https://doi.org/10.3390/w14091410
- Cruiz IA, Nascimento VRS, Felisardo RJA et al. Evaluation of artificial neural network models for predictive monitoring of biogas production from cassava wastewater: Comparison of training algorithms. Biom Bioener. 2023;175(106869):1-24. https://doi.org/10.1016/j.biombioe.2023.106869
- 4. Fakharudin AS, Sulaiman N, Salihon J et al. Implementing artificial neural networks and genetic algorithms to solve modeling and optimisation of biogas production. Proceed 4th Inter Conf Comp Inform Saraw, Mal. Univ Ut Mal. 2013;088:121-126. http://www.uum.edu.my
- Ghatak MD, Ghatak A. Artificial neural network model to predict behavior of biogas production curve from mixed lignocellulosic co-substrates. Fuel. 2018;232:78-189. https://doi.org/10.1016/j.fuel.2018.05.051
- Jaroenpoj S, Yu QJ, Ness J. Development of artificial neural network models for biogas production from co-digestion of leachate and pineapple peel. Glob Environ Eng. 2014;1(2):42-47. https://doi.org/10.15377/2410-3624.2014.01.02.2
- Klimberg R, Mccullough BD. 2012. Logistic regression. In R. Klimberg & B. D. McCullough (Eds.). Fundamentals of Predictive Analytics with JMP (2nd ed., 104-133). SAS Institute Inc. https://support.sas.com/content/dam/SAS/support/en/books/fundamentalsof-predictive-analytics-with-jmp/68711\_excerpt.pdf
- 8. Mougari NE, Largeau JF, Himrane N et al. Application of artificial neural network and kinetic modeling for the prediction of biogas and methane production in anaerobic digestion of several organic wastes. Int J Green Ener. 2021;18(15):1584-1596. https://doi.org/10.1080/15435075.2021.1914630
- 9. Obileke K, Tangwe S, Makaka G et al. Comparison of prediction of biogas yield in a batch mode underground fixed dome digester with cow dung. Biom Conv Biorefin. 2023:1-16. https://doi.org/10.1007/s13399-023-04593-z

- 10. Olufunmi AO. Microbiological potentials of co-digestion of chicken droppings and banana peels as substrates for biogas production. J Chem Pharm Res. 2014;6(4):1088-1092. www.jocpr.com
- 11. Palaniswamy D, Ramesh G, Sivasankaran S et al. Optimising biogas from food waste using a neural network model. Proceed Instit Civ Eng Munic Eng. 2017;170(ME4):221-229. https://doi.org/10.1680/jmuen.16.00008
- Qdais HA, Hani KB, Shatnawi N. Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. Res Cons Recyc. 2010;54(6):359-363. https://doi.org/10.1016/j.resconrec.2009.08.012
- Rego ASC, Leite SAF, Leite BS et al. Artificial neural network modelling for biogas production in biodigesters. Chem Eng Transac. 2019;74:25-30. https://doi.org/10.3303/CET1974005
- Rodriguez-Granrose D, Jones A, Loftus H et al. Design of experiment (DOE) applied to artificial neural network architecture enables rapid bioprocess improvement. Bioproc Biosyst Eng. 2021;44(1):1-7. https://doi.org/10.1007/s00449-021-02529-3
- 15. Sakiewicz P, Piotrowski K, Ober J et al. Innovative artificial neural network approach for integrated biogas wastewater treatment system modelling: Effect of plant operating parameters on process intensification. Renew Sustain Energ Rev. 2020;124(109784):1-14. https://doi.org/10.1016/j.rser.2020.109784
- Sudirman, Rodríguez-Nieto CA, Dhidden LZB et al. Ways of thinking 3D geometry: Exploratory case study in junior high school students. Polyhed Int J Math Educ. 2023;1(1):15-34. https://nakiscience.com/index.php/pijme
- Um-e-Habiba, Khan MS, Raza W et al. A study on the reaction kinetics of anaerobic microbes using batch anaerobic sludge technique for beverage industrial wastewater. Separations. 2021;8(43):1-16. https://doi.org/10.3390/ separations8040043