27th INTERNATIONAL CONGRESS OF MECHANICAL ENGINEERING
COBEM 2023

ABCM
Brazilian Society of Mechanical Sciences and Engineering
27th ABCM International Congress of Mechanical Engineering
December 04-08, 2023, Florianópolis-SC, Brazil

COB-2023-1040
DEVELOPMENT OF ARTIFICIAL NEURAL NETWORKS (ANN) MODELS TO PREDICT THE PRODUCTION OF CUMULATIVE BIOGAS FROM FOOD WASTE (FW), FRUITS AND VEGETABLES WASTE (FVW) AND THEIR CODIGESTION (CD)

Michel Angelo Oliveira Werneck de Carvalho
Florian Pradelle
Departamento de Engenharia Mecânica (DEM), Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rua Marquês de São Vicente, 225, Gávea, Rio de Janeiro, RJ, CEP:22451-900, Brazil
michelwerneckc@hotmail.com; pradelle@puc-rio.br

Brunno Ferreira dos Santos
Departamento de Engenharia Química e de Materiais (DEQM), Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rua Marquês de São Vicente, 225, Gávea, Rio de Janeiro, RJ, CEP:22451-900, Brazil
bsantos@puc-rio.br

Abstract. *Neural Network (ANN) models for estimating the cumulative volume of biogas produced by a diverse range of biodigester configurations. A comprehensive database was created, incorporating information from 47 literature references, resulting in a total of 2098 conditions. The database included eight variables: biomass type, reactor type, volatile solids (VS), hydraulic retention time (HRT), organic load rate (OLR), temperature, pH, and reactor volume. Data wrangling analysis was conducted to prepare the database, including the removal of outliers and missing data using histograms. The ANNs, developed in Matlab software, were evaluated using various topologies, with the number of neurons in the hidden layer ranging from 8 to 11 and different activation functions, including the output layer. The models' performance was assessed using the coefficient of determination (R^2) and the sum of squared errors (SSE). Additionally, response surface assessments were conducted to evaluate the models' applicability across a range of operational conditions. The final database consisted of 2,098 conditions, and the investigations demonstrated the feasibility of developing a predictive model for cumulative biogas production with acceptable performance indexes. The response surfaces identified regions with enhanced production performance, characterized by a combination of process variables consistent with previous literature.*

Keywords: *Biofuel production; Biomass; Database analysis, Response surface assessment.*

1. INTRODUCTION

The global challenge of combating climate change is driving the increasing adoption of renewable energy sources to produce biofuels. Economic factors associated with fossil fuels also contribute to the production of biofuels, as they face growing obstacles such as price volatility and a medium to long-term trend of decreasing supply (BNDES, 2021). Food waste (FW) has great potential for conversion into biogas. Unfortunately, worldwide, approximately a quarter and a third of the food produced annually for human consumption is lost or wasted, respectively. This amounts to approximately 1.3 billion tons of food, encompassing 30% of cereals, 40 to 50% of roots, fruits and vegetables (FVW), and oil seeds, 20% of meat and dairy products, and 35% of fish (FAO, 2023). Brazil's biofuel production has been growing due to the high availability biomass from various sources, including forest residues, agricultural and livestock waste, urban waste, and industrial waste (ECBGROUP, 2023).

Biomass, as an energy carrier, can undergo various conversion processes to be transformed into fuels (Abou Rjeily et al., 2021). Among the potential pathways, anaerobic digestion stands out as an environmentally friendly technology for converting solid or liquid organic wastes into biogas, which can be further utilized as a valuable source of energy in the form of electricity and/or heat This approach is widely reported in literature such as Bandgar et al. (2022), Carlini et al. (2021), Yang et al. (2022), Shi et al. (2023) and others. Biogas consists of mixture of methane (CH_4), carbon dioxide (CO_2), hydrogen sulfide (H_2S), moisture, traces of dust particles, siloxanes, aromatic and halogenated compounds (Mao et al., 2015; Zhang et al., 2017; Isha et al., 2020). To increase the biogas production, the literature showed a significant potential of co-digestion (CD) of FW and FVW. Pal et al. (2022) studied the conversion potential under co-digestion mode of rice straw and cow dung in ratio 1/2. The authors mentioned that the maximum value reached has been recorded in 35 days with average methane composition of 58.3%. Jayanth et al. (2020) investigated the influence of volatile organic loading rate (OLR) on a semi-commercial biomethanation plant comprising three anaerobic gas lift reactors. The plant was operated for 47 weeks, utilizing mono and co-digestion of the organic fraction of municipal solid waste and landfill leachate. The results indicated the potential for producing substantial quantities of biogas.

The biogas production is known for its relatively long duration, resulting in high time costs for engineering optimization and scientific research in this field (Weinrich et al., 2021; Zhao et al., 2021). Consequently, the use of theoretical or numerical simulation to model the anaerobic digestion process serves as an effective approach to explore optimal regulation strategies (Cruz et al., 2022). There are numerous approaches to developing process models, such as white-box and black-box models. White-Box models are constructed using kinetic and thermodynamic parameters of the reactions involved in the transformation process, based on various phenomena (Mandapati et al., 2021). On the other hand, Black-Box models, for instance Artificial Neural Networks (ANN), Fuzzy logic and Fuzzy Neural Networks, are developed using Artificial Intelligence techniques. These models primarily emphasize the statistical characteristics of data, but often omit the specific details of the reactions involved in the mass-to-energy transformations during the modeling process (Tang et al., 2022).

The ANN is an attractive model that involves the processing of interconnected elements called neurons, which are activated by specific functions. The process of adjusting the Artificial Neural Network's parameters is the subject of many studies in different areas of knowledge, that is, the definition of a topology. The ANN exhibits the capability to address nonlinear problems and achieve faster prediction by learning from examples. It possesses the potential to establish the relationship between input and output variables of a process using a limited amount of data. Many works report the use of Artificial Neural Network models as predictors of biogas production from waste, as seen in Cruz et al. (2023), Almomani (2020), Vondra et al. (2019) etc. Li et al. (2022) studied the development of ANN models to predict biogas production in an industrial-scale biogas plant treating food waste. The performance's model was not satisfactory with a coefficient of determination R^2 of 0.48 and a root mean square error (RMSE) of 0.88. Neto et al. (2021) constructed ANN models using a comprehensive database comprising 205 experimental conditions of the anaerobic digestion process. The database was compiled by combining experimental data from their own study and 53 previously published works in the literature, aiming to establish correlations between cumulative biogas production (PCB) at the end of the biodigestion. Prior to model development, thorough evaluations were conducted on the database (through the analysis of histograms, the final database had 108 cases), revealing promising performance in determining the optimal region for biogas production. Olatunji et al. (2023) utilized a genetic algorithm to optimize an adaptive neuro-fuzzy inference system (ANFIS) to predict cumulative biogas production. The benchmark dataset used for this analysis was constructed by Neto et al. (2021). The authors conducted a comprehensive comparison among different intelligent models and identified an effective tool for scaling up anaerobic digestion units and for conducting techno-economic studies aimed at enhancing energy utilization efficiency.

Most of the works reported in the literature for predicting biogas production from FW using Artificial Neural Network models portray the use of databases of only one experimental biodigester or type of biomass developed by the researchers, with very few models that can predict multiple scenarios (Neto et al., 2021). In other words, studies covering different conditions of experimental operations of biogas production, such as the construction of a benchmark datasets. There is a natural limitation on the physical ability of one or more groups of scientists to produce data representative of vast experimental process conditions. However, scientific literature through published works can provide valuable information for a good database. It is important to highlight, simultaneously studying the interactive effects of multiple experimental scenarios yields valuable insights for predicting and optimizing key characteristics in the experimental process of any given scenario. Conducting a comprehensive review of relevant literature is a crucial step in gathering information for numerous regulatory tasks (Pare and Kitsiou, 2017). Thus, exploratory data techniques, such as data wrangling, can be used to identify the presence of outliers, missing data, and ensure that statistical assumptions are met by gaining insights from the data (Shreffler and Huecker, 2023). Several other factors, including data quality, integration of multiple sources, reproducibility of processes, and managing data provenance, need to be taken into consideration. Histograms serve as a prime example of these techniques. They provide an empirical distribution of the population within a dataset, displaying the observed values of a random variable based on a finite sample (Bityukov et al., 2016). Moreover, histograms help identify outliers, which can be a result of incorrect data entry or data points that significantly deviate from the rest of the sample (Weissman, 1997).

To enhance the understanding of biogas production, this study aimed to construct an extensive database, following the approach of Neto et al. (2021), by incorporating a larger amount of information from previously published literature, including the transient behavior of the process. Through data analysis using histograms, Artificial Neural Network models were developed, utilizing key variables of the anaerobic digestion process as inputs. These variables included biomass type, reactor type, volatile solids content (VS), pH, organic loading rate (OLR), hydraulic retention time (HRT), temperature, and reactor volume. Various topologies were explored to evaluate the effects of parameters and determine the optimal conditions that maximize cumulative biogas production, visualized through response surface charts.

2. METHODOLOGY

This work was based on the work by Neto et al. (2021). Modifications were necessary in order to adapt the topology to the availability of data and characteristics of the proposed problem.

2.1 Database construction and data analysis

A comprehensive database was compiled through a rigorous process involving selection of article and data analysis, accompanied by meticulous review to ensure accuracy and reliability of the information utilized. Standardization of data and conversion of measurement units into a common scale were also undertaken to facilitate integrated analysis. These efforts were crucial to ensure the validity and robustness of the study. An extensive literature search was conducted to gather scientific articles containing data on key variables in the anaerobic digestion process for biogas production. Consequently, information from 47 previously published scientific papers in the period 2003-2020 was gathered, composing 2,098 samples.

Data analysis involved constructing histograms using Excel software to identify, address potential outliers and preserve the representativeness of the data from the original sample without generating empty classes. The database was then divided into three subsets: training, testing, and validation. The division was done randomly while maintaining the proportions of 60%, 20%, and 20%, respectively. The "*dividerand*" function of the MATLAB R2020b software was used to perform this step.

2.2 ANN modelling

The modelling of biogas production was carried out using various topologies of artificial neural networks. These Artificial Neural Networks consisted of three or more layers, with interconnected artificial neurons. The first layer, known as the input layer, represented the variable values of the process, while the output layer represented the response of the ANN. Different numbers of neurons, specifically 8, 9, 10, and 11, were investigated. For each independent topology, the training process was conducted using MATLAB R2020b with the *Deep Learning Toolbox* for optimizing weights and bias values based on backpropagation algorithm. For the hidden layer, two activation functions were examined: hyperbolic tangent (*tansig*) and logistic sigmoid (*logsig*). The output layer utilized a linear transfer function (*purelin*), as represented by equations (1), (2), and (3) respectively.

$$f(x) = \frac{2}{(1+e^{-2x})} - 1 \quad (1)$$

$$f(x) = \frac{1}{(1+e^{-x})} \quad (2)$$

$$f(x) = x \quad (3)$$

Three different training algorithms were employed: *traingdx, trainlm* and *trainbr*. The *traingdx* algorithm utilizes a gradient descent approach, where the negative gradient is utilized to minimize the error. On the other hand, the *trainlm* algorithm employs the Levenberg-Marquardt algorithm along with a Quasi-Newton method. In this case, the Hessian matrix is approximated from the Jacobian matrix, which is computed using the backpropagation technique. The inclusion of a µ parameter in this technique allows for weighing the influence of the Hessian matrix over the update of the Artificial Neural Network parameters in relation to the Jacobian. Additionally, the *trainbr* algorithm, also utilizing a Quasi-Newton method, aims to minimize the error. In this case, Bayesian regularization is employed, where the weight values are included in the algorithm in addition to the error. Subsequently, the Levenberg-Marquardt algorithm is applied, which typically yields improved generalization performance.

Evaluation of the Artificial Neural Network model's quality was carried out by sum of squared error (SSE), which makes a comparison between the experimental and calculated values as shown in Equation (4). The model´s predictive accuracy can be estimated by the coefficient of determination (R^2), as shown in equation (5).

$$SSE = \sum_{i=1}^{n}\left(Y_{observed} - Y_{predicted}\right)^2 \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_{observed} - Y_{predicted}\right)^2}{\sum_{i=1}^{n}(Y_{observed} - Y_{mean})^2} \quad (5)$$

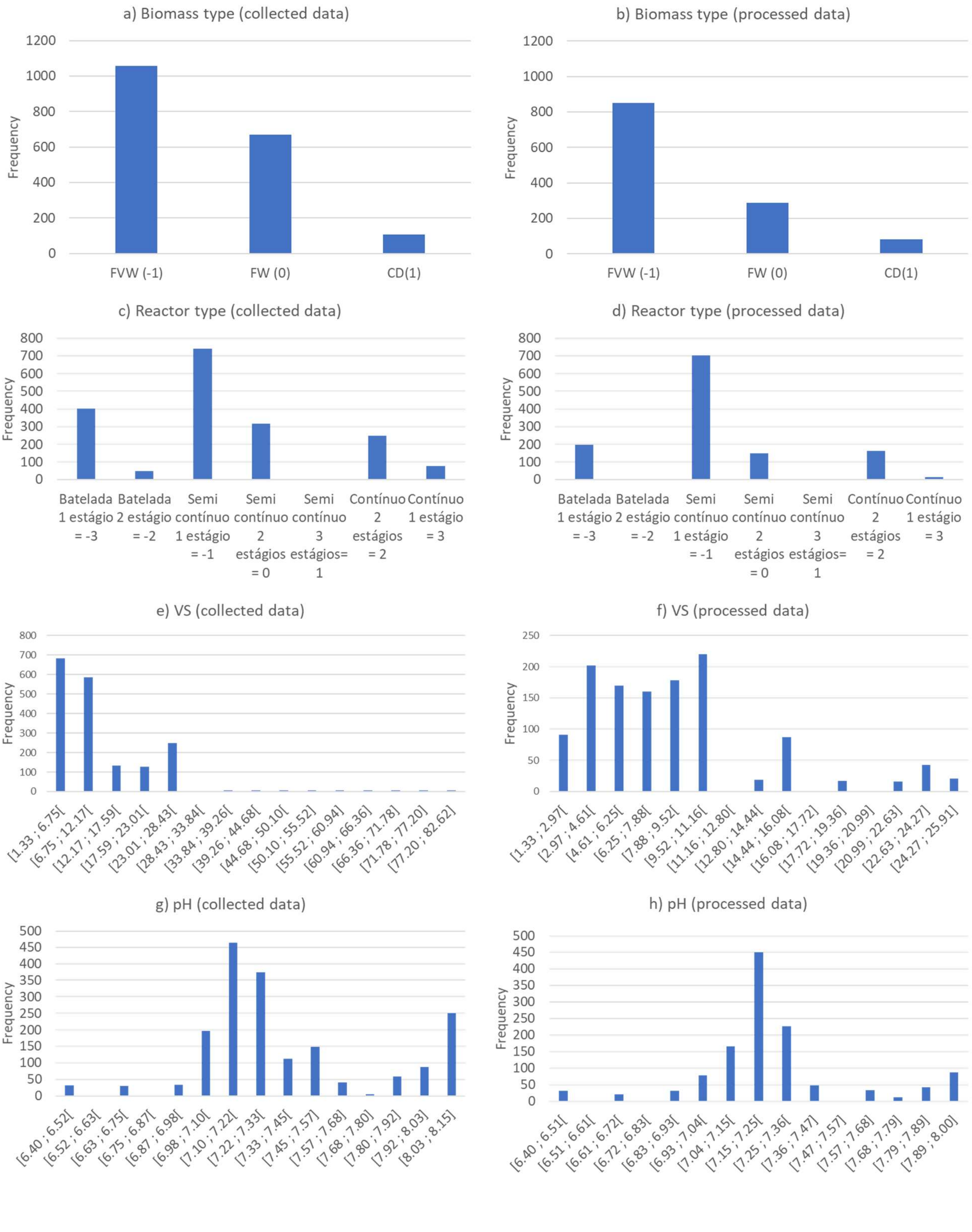
2.3 Sensitively analysis

The sensitivity analysis consisted of evaluating the effects on the response variable in the ranges of the variables. Thus, response surfaces were built with the help of MATLAB R2020b to investigate the so-called optimal regions. This process allows a qualitative validation of the model and a quantitative assessment of the best values of each parameter

3. RESULTS AND DISCUSSION

3.1 Database analysis

From the data collected in the literature to build the database, it was possible to develop a systematic analysis to predict the cumulative production of biogas (PCB). Initially, missing data without statistical representation were excluded from these databases. The process enabled reformatting, validation, standardization, enrichment, and integration of data from diverse sources, offering a wide scope for comprehensive analysis and facilitating iterative discovery of patterns within the datasets. To accomplish this, histograms were generated to examine the data distribution of each variable and response variable. Histograms for the cumulative biogas production database are shown in Figures.



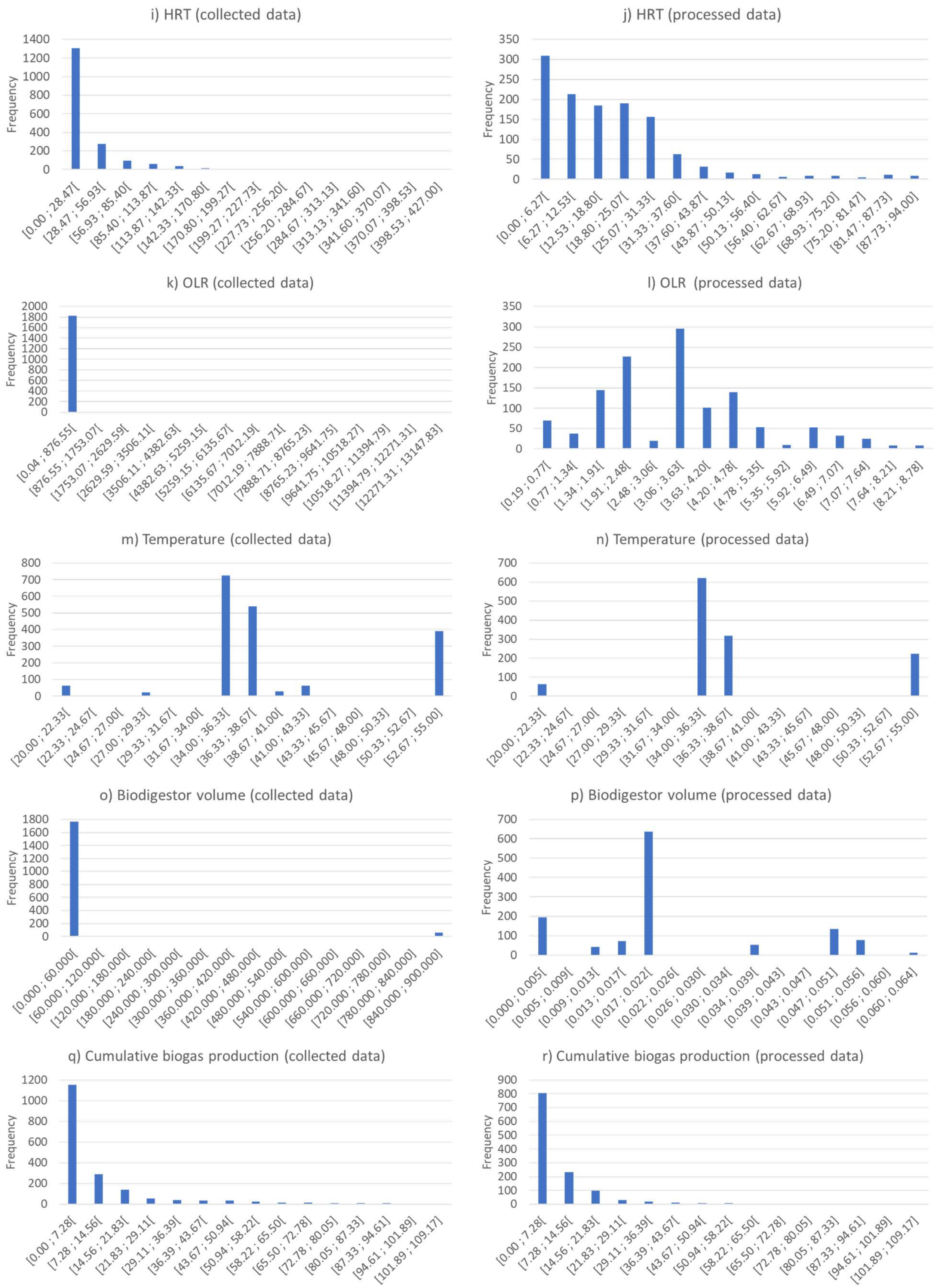


Figure 1: Histograms for the cumulative biogas production databases: comparison of “collected data” and “processed data” distributions for biomass type, reactor type, volatile solids content (VS), pH, organic loading rate (OLR), hydraulic retention time (HRT), temperature, reactor volume and cumulative biogas production

It was possible to understand which data regions are more populated. In particular, regions with biodigestion of FVW, in a semi-continuous rector with 1 stage, volatile solid contents above 28.43%, pH below 6.98, hydraulic retention time above 56.93 days, temperature below 34°C and biodigester volume above 60 m^3 were poorly sampled. As for the organic loading rate variable, the data were mainly concentrated in the first class of the histogram. At the end of the analysis, the database was reduced to 1,832 vectors per variables (database named "collected data" with 87% of the original data from the literature). Additional refinement of the database, restricted to volume lower than 0.1 m^3, reduced organic loading rate (very high values were introduced when calculating um pseudo-OLR for batch reactor must be removed) and HRT (up to 94 days) led to the database named "processed database with 1,224 vectors (58% of the original dataset). Thus, the filtering process is not impacting equally all properties. Looking to the distribution of the smaller dataset, it can be observed that the co-digestion condition is poorly represented, as batch reactor with 2 stages, semi-continuous reactor with 3 stages or continuous reactor with 1 stage. Regarding the volatile solid content, the experimental data are showing a quasi-homogeneous distribution in the range 2.97 to 11.16%, while the pH is almost following a normal distribution centered in the range 7.15-7.25. HRT values are concentrated in a duration up to 1 month, as reported in the literature. The range of organic loading rate with higher population is 3.06-3.63 gVS/(l.d), but values as high as 8.78 gVS/(l.d) can be found. The temperature distribution is mainly bimodal with most of the data in the mesophilic temperature (around 35ºC) and others close to 55ºC. Reactors with volume of approximately 20L are the most represented, confirming that the data are collected at laboratory or pilot scales. The refinement of the dataset did not significantly affect the histogram of PCB, since the lowest values are still the most abundant and the classes with higher values were still represented.

3.2 Artificial Neural Network modelling

As an illustration of the next step, the treatment of the dataset with the collected data is commented (the same procedure is applied to the refined dataset. Out of the 1,832 samples available in the biogas database, 1100 samples (60%) were allocated to the training database, 366 samples (20%) were assigned to the testing database, and another 366 samples (20%) were designated for the validation database. Then, the weights and biases for each topology were determined using the tools provided by the MATLAB software. The results of these calculations are presented in Table 1.

Table 1: Performance metrics for each investigated topology

Identification of RNA	Training algorithm	Hidden layer activation function	Total number of neurons	SSE	R^2 Training	R^2 Test
net1	Traingdx	tansig	8	8.110728	0.923292	0.867873
net2			9	9.429183	0.91111	0.859587
net3			10	11.83361	0.888425	0.832878
net4			11	9.719988	0.908076	0.855406
net5		logsig	8	19.50261	0.815647	0.732019
net6			9	22.39626	0.792302	0.717801
net7			10	17.11557	0.838305	0.776688
net8			11	14.23207	0.865949	0.804186
net9	Trainbr	tansig	8	1.500297	0.985803	0.980214
net10			9	1.195906	0.988683	0.985174
net11			10	1.151157	0.989107	0.985033
net12			11	1.14029	0.98921	0.986271
net13		logsig	8	1.212851	0.988523	0.986454
net14			9	1.161367	0.98901	0.986273
net15			10	1.16002	0.989023	0.987011
net16			11	1.114876	0.98945	0.985843
net17	**Trainlm**	tansig	8	1.224031	0.988417	0.985327
net18			9	1.132132	0.989287	0.985296
net19			10	1.509916	0.985712	0.978761
net20			11	1.015256	0.990393	0.985881
net21		**logsig**	8	1.400528	0.986747	0.982676
net22			9	1.024400	0.990306	0.985219
net23			**10**	**0.816672**	**0.996130**	**0.994330**
net24			11	0.904292	0.991443	0.987241

Based on the evaluation of various topologies, multiple models with commendable performance can be obtained. The selected model, highlighted in Table 1, possesses the following topology: training algorithm "trainlm," activation function "logsig," and 10 neurons and the observed performance metrics include an SSE of 0.8166, training R^2 of 0.9923, and test R^2 of 0.9887, when the “collected data” are used. Similarly, for the “processed data”, the best topology is training algorithm "trainbr," activation function "tansig," and 11 neurons and the SSE, training R^2 and test R^2 are equal to of 0.1808, 0.9950 and 0.9924, respectively. The metrics showed that the obtained ANN is slightly more robust, in particular regarding the R^2 of the validation dataset and the value of SSE. The models demonstrated a strong capability in accurately predicting the cumulative volume of biogas produced, achieving an R^2 value of 0.9955 when evaluated against a validation database that was not utilized during the parameter adjustments of the ANN model. The coefficient of determination (R^2) is widely employed in the literature to assess the performance of Artificial Neural Networks in predicting biogas production. For instance, in a study conducted by Tian et al. (2021) using a dataset encompassing various substrates for biogas production, R^2 values exceeding 0.9 were obtained in certain cases.

3.3 Sensitively analysis

Regarding the sensitivity analysis, a comparison of the models obtained with the two dataset is performed. First of all, the surface dealing with the variation of PCB with the reactor volume and pH is plotted for each biomass. Such surface must point out higher cumulative production of biogas for pH close to 7.2 (Figure 3).

a) ANN model trained with the “collected data” b) ANN model trained with the “processed data”

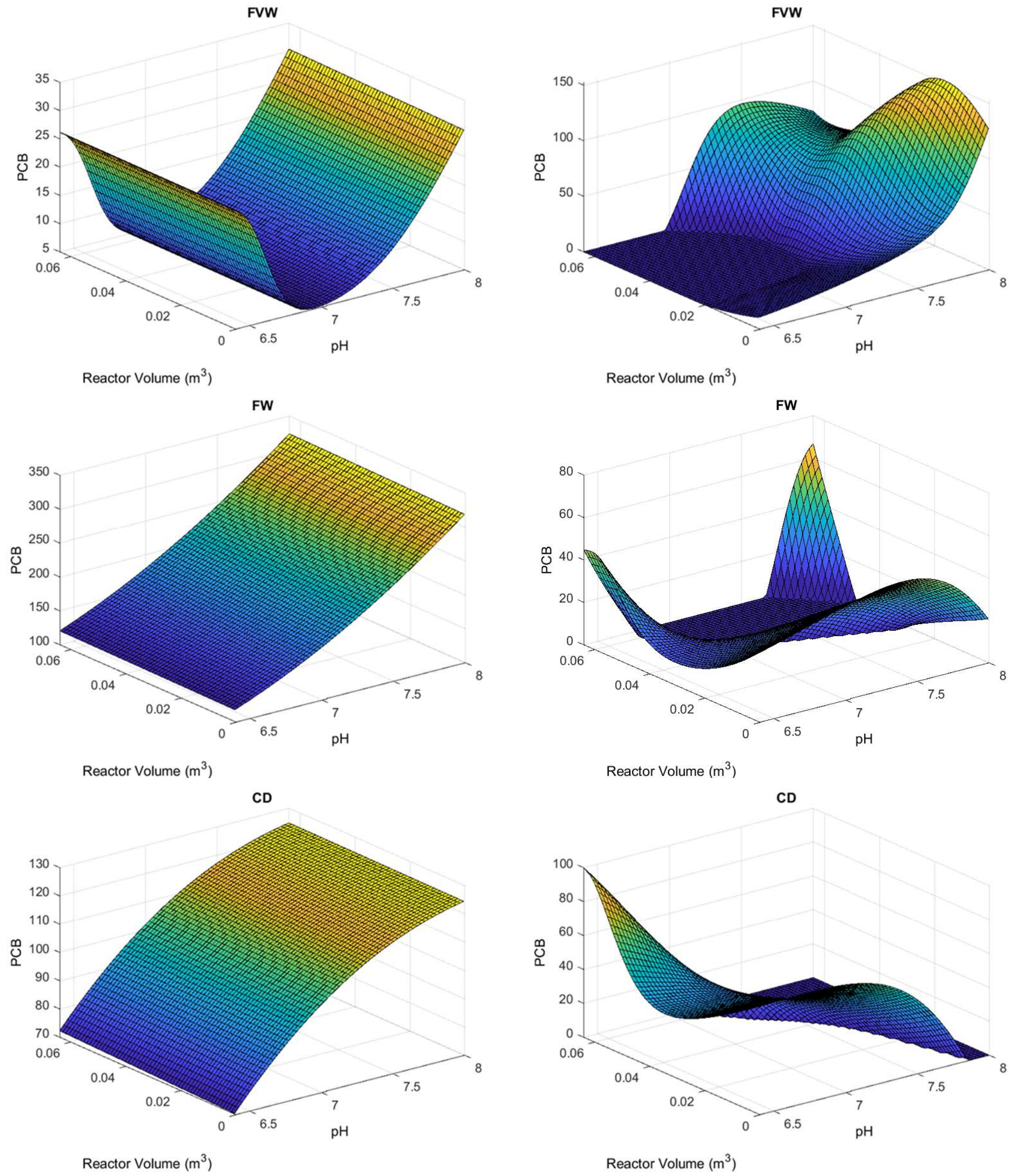


Figure 2: Variation of the cumulative biogas production (PCB) as a function of the pH and the reactor volume

As only two factors can be changed simultaneously, the other factors must adopt a constant value. It is assumed the median values of the processed data for each variable: VS equal to 8.70%; pH value of 7.20; OLR equal to 3.17 g VS/(l.d)); HRT of 15.23 days; temperature equal to 35.00°C and a reactor volume of 0.018 m³. The reactor type is a semi-continuous reactor with 1 stage.

The Figure 2 showed that the ANN model trained with the "collected data" is not sensitive to volume variation. For Food Wastes and Codigestion, the values of PCB can reach values higher than the maximum values in the dataset and the model was not able to identify optimal condition close to a pH equal to 7.2. For FVW, the PCB values are within the expected range. Nonetheless, the surface showed a minimum PCB for pH close to 7 which is not compatible to the physics of the problem. For this reason, this model is not reliable for further investigation. On the other hand, the Artificial Neural Network model trained with the "processed data" showed a coherent behavior for FW and CD for volume up to 0.020 m³ (i.e., the most populated range of the dataset). For higher values, the represented behavior is not consistent with the expected behavior of a biodigester, but the dataset showed some values in the region. For FVW, an additional comment must be done since the calculated values extrapolated the maximum value of PCB in the data dataset for the lowest values of volume. Thus, at this time, the ANN model trained with the "processed data" since to be the most reliable in the region with volume inferior to 20 L for FV and CD.

The Figure 4 is showing the variation of the cumulative biogas production (PCB) in L/gVS as a function of the temperature and the reactor volume

a) ANN model trained with the "collected data" b) ANN model trained with the "processed data"

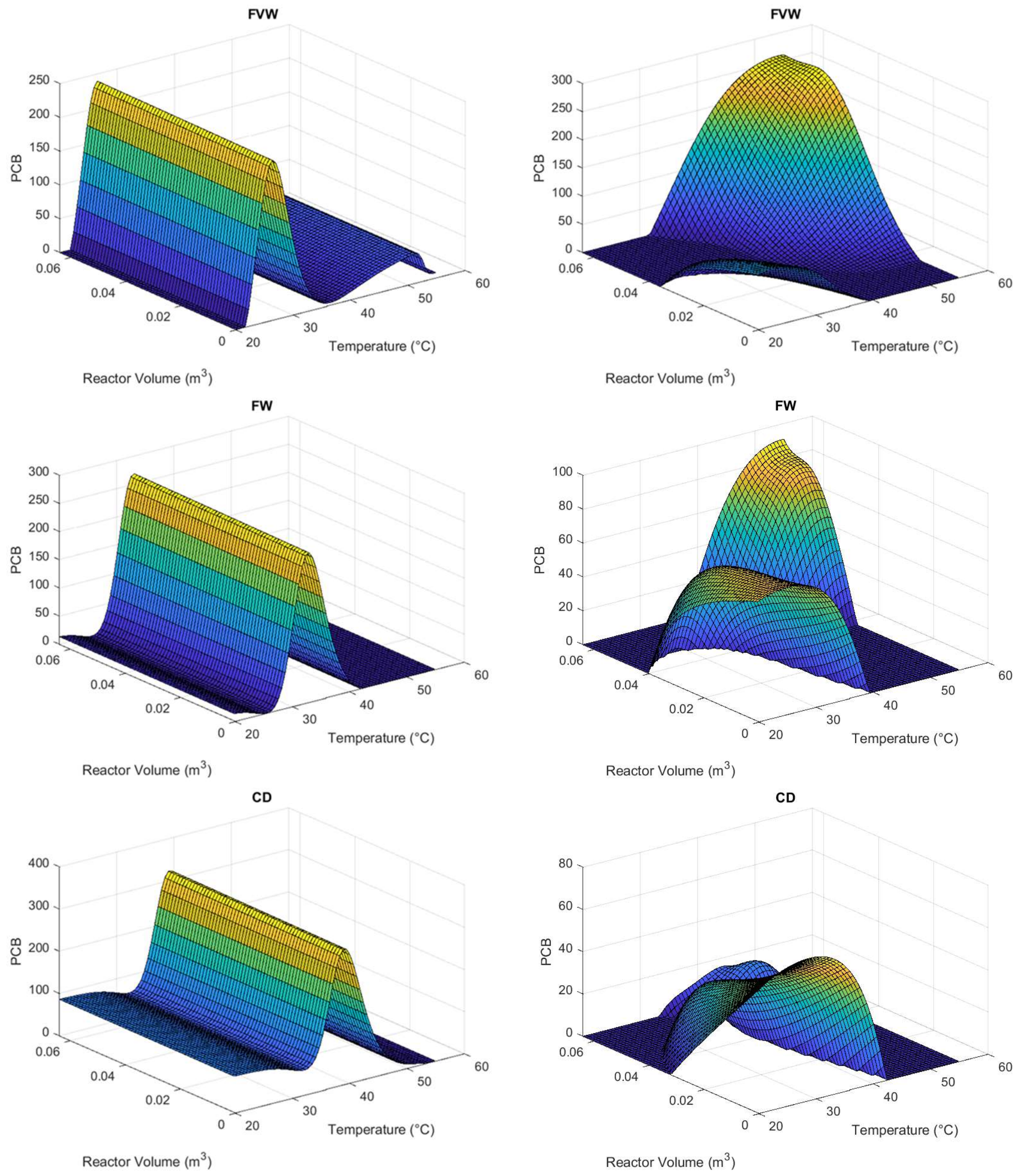


Figure 3: Variation of the cumulative biogas production (PCB) as a function of the temperature and the reactor volume

The Figure 3 is strengthening the restriction regarding the reliability of the models since Artificial Neural Network model trained with the "collected data" is not sensitive to volume variation (as seen before) and the optimum temperature is not consistent neither to mesophilic nor thermophilic values for FVW and CD. Regarding the ANN model trained with the "processed data", the surfaces for FW and CD respectively, showed consistent values of PCB and identified optimal conditions at mesophilic temperature for low volume and thermophilic temperature at volume higher than 0.04 m^3. For FVW, the calculated values extrapolated the maximum value of PCB in the data dataset for the highest values of volume and temperature.

4. CONCLUSION

In the present study, a database consisted of 2,098 conditions identified from experimental data reported in 47 articles was built. After processing the data to remove the outliers, obtain a more homogeneous distribution of the values within the investigated experimental domain preserving the database diversity and improve the quality of the dataset for Artificial Neural Network training, the investigations demonstrated the feasibility of developing a predictive model for cumulative biogas production with acceptable performance indexes. The response surfaces identified regions with enhanced production performance, characterized by a combination of process variables consistent with previous literature for the ANN model trained with the "processed data" while prediction obtained with the model trained with the "collected data" are not sensitive to reactor volume or even predict improper values (higher order of magnitude). Nonetheless, the qualitative and quantitative analysis of such surface showed that the model seems reliable for volume up to 20 l for Food Wastes and Codigestion. In the case of Fruit and Vegetable Wastes, the Artificial Neural Network prediction in the investigated experimental domain lead to very high values of PCB, inconsistent with the phenomenon.

Future works must study more precisely the reliability of the model, in particular identifying subregion of the experimental domain with higher quantity of data in the training data set and compare the calculated results to a new experimental dataset to guarantee that the results are not dependent to the selected sample (the current validation is mainly based on qualitative assessment of the behaviour of the model and order of magnitude of PCB, based on the state-of-the-art available in the literature). The procedure described in this paper can be applied to other parameters such as cumulative biomethane production or percentage of biomethane in the mixture. Additionally, it can be interesting to add new variables to the model (for instance, nutrients, C:N ratio, pressure, inoculum) and use other machine learning approach.

5. ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001 and the grant awarded to Michel Angelo Oliveira Werneck de Carvalho. The authors would like to thank the CNPq/MCTIC for the financial support to the Department of Mechanical Engineering (DEM) at the Pontifical Catholic University of Rio de Janeiro (PUC-Rio) and the FAPERJ for the. Jovem Cientísta do Nosso Estado (JCNE) grants awarded to Florian Pradelle (process Number E26/201.402/2021).

6. REFERENCES

Abou Rjeily, M., Gennequin, C., Pron, H., Abi-Aad, E., Randrianalisoa, J.H., 2021, Pyrolysis-catalytic upgrading of bio-oil and pyrolysis-catalytic steam reforming of biogas: a review Environ. Chem. Lett., 19, pp. 2825-2872

Almomani, F., 2020, Prediction of biogas production from chemically treated co-digested agricultural waste using artificial neural network, Fuel, 280.

Bandgar, P.S., Jain, S., Panwar, N.L., 2022, A comprehensive review on optimization of anaerobic digestion technologies for lignocellulosic biomass available in India, Biomass and Bioenergy, v. 161.

Bityukov, S.I., Maksimushkina, A.V., Smirnova, V.V., 2016, Comparison of histograms in physical research, Nuclear Energy and Technology, Volume 2, Issue 2.

BNDES, 2021, Biogas: Recent evolution and potential of a new energy border renewable for brazil, Final Report (in Portuguese), Complexo Agroalimentar e de Biocombustíveis da Área de Indústria e Serviços e Comércio Exterior do BNDES., Rio de Janeiro, https://web.bndes.gov.br/bib/jspui/bitstream/1408/20801/1/PR_Biogas_215276_P_BD.pdf. Accessed 23 June 2023.

Carlini, M., Monarca, D., Castellucci, S., Mennuni, A., Casini, L., Selli, S., 2021, Beer spent grains biomass for biogas production: Characterization and anaerobic digestion-oriented pre-treatments, Energy Reports, v. 7, p. 921-929.

Cruz, I.A., Chuenchart, W., Long, F., Surendra, K.C., Andrade, L.R.S., Bilal, M., Liu, H., Figueiredo, R.T., Khanal, S.K., Ferreira, L.F.R., 2022, Application of machine learning in anaerobic digestion: perspectives and challenges Bioresour. Technol., 345.

Cruz, I.A, Nascimento, V.R.S., Felisardo, R.J.A., Santos, A.M.G., Jesus, A.A., Vasconcelos, B.R., Kumar, V., Cavalcanti, E.B., Souza, R.L., Ferreira, L.F.R., 2023, Evaluation of artificial neural network models for predictive monitoring of biogas production from cassava wastewater: A training algorithms approach, Biomass and Bioenergy, v. 175.

ECBGROUP, 2023, Biofuels gain prominence in the Brazilian energy matrix, Final Report (in Portuguese), Biocombustíveis ganham destaque na matriz energética brasileira, Rio Grande do Sul, https://www.ecbgroup.com.br/pt/noticia/biocombustiveis-ganham-destaque-na-matriz-energetica-brasileira-diz-aie. Accessed 18 June 2023.

FAO, 2023, Losses and food waste in Latin America and the Caribbean, https://www.fao.org/americas/noticias/ver/en/c/239392/. Accessed 18 June 2023.

Isha, A., Kumar, S., Jha, B., Subbarao, P.M.V., Chandra, R., Vijay, V.K., 2020, Development of stabilization methods using a pilot scale anaerobic digester for seasonal variations in kitchen wastes for improved methane production with zero breakdowns. Clean Eng Technol, v. 1.

Jayanth T.A.S, Mamindlapelli, N.K., Begum, S., Arelli, V., Juntupally, S., Ahuja, S., Dugyala, S.K., Anupoju, G.R., 2020, Anaerobic mono and co-digestion of organic fraction of municipal solid waste and landfill leachate at industrial scale: Impact of volatile organic loading rate on reaction kinetics, biogas yield and microbial diversity, Science of The Total Environment, v. 748.

Li, C., He, P., Peng, W., Lü, F., Du, R., Zhang, H., 2022, Exploring available input variables for machine learning models to predict biogas production in industrial-scale biogas plants treating food waste, Journal of Cleaner Production, v. 380.

Mandapati, R.N., Ghodke, P.K., 2021, Kinetics of pyrolysis of cotton stalk using model-fitting and model-free methods Fuel, 303.

Mao C., Feng Y., Wang X., Ren G., 2015, Review on research achievements of biogas from anaerobic digestion. Renewable and Sustainable Energy Reviews; 45: 540–55.

Neto, J. G.; Ozorio, L. V.; Abreu. C. C.; Santos, B. F.; Pradelle, F. Modelling of biogas production from food, fruits and vegetables wastes using artificial neural network (ANN). Fuel journal. 2020.

Olatunji, O.O., Adedeji, P.A., Madushele, N., 2023, Evolutionary optimization of biogas production from food, fruit, and vegetable (FFV) waste. Biomass Conv. Bioref.

Pal, B.D., Tiwari, A.K., Mohammad, A., Prasad, N., Srivastava, N., Srivastava, K.R., Singh, R., Yoon, T., Syed, A., Bahkali, A.H., Gupta, V.K., 2022, Enhanced biogas production potential analysis of rice straw: Biomass characterization, kinetics and anaerobic co-digestion investigations, Bioresource Technology, v. 358.

Pare, G., Kitsiou, S., 2017, Methods for literature reviews Handbook Of eHealth Evaluation: an Evidence-Based Approach, Victoria (BC): University of Victoria

Shi, M., Duan, H., Feng, L., Xiao, M., He, Q., Yan, S., 2023, Sustainable ammonia recovery from anaerobic digestion effluent through pretreating the feed by biomass ash, Separation and Purification Technology, v. 307.

Shreffler J., Huecker, M.R., 2023, Exploratory Data Analysis: Frequencies, Descriptive Statistics, Histograms, and Boxplots, [Updated 2023 Feb 13]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing.

Tang, J., Xia, H., Aljerf, L., Wang, D., Ukaogo, P.O., 2022, Prediction of dioxin emission from municipal solid waste incineration based on expansion, interpolation, and selection for small samples J. Environ. Chem. Eng., 10.

Vondra, M., Touš, M., Teng, S.Y., 2019, Digestate evaporation treatment in biogas plants: A techno-economic assessment by Monte Carlo, neural networks and decision trees, Journal of Cleaner Production, v. 238.

Weinrich, S., Mauky, E., Schmidt, T., Krebs, C., Liebetrau, J., Nelles, M., 2021, Systematic simplification of the Anaerobic Digestion Model No. 1 (ADM1)-Laboratory experiments and model application Bioresour. Technol., 333.

Weissman C., 1997, Analyzing intensive care unit length of stay data: problems and possible solutions. Crit Care Med. Sep;25(9):1594-600

Yang, H., Dou, X., Pan, F., Wu, Q., Li, C., Zhou, B., Hao, L., 2022, Optimal planning of local biomass-based integrated energy system considering anaerobic co-digestion, Applied Energy, v. 316.

Zhang J., Loh, K.C., Li, W., Lim, J.W., Dai Y., Tong, Y.W., 2017, Three-Stage Anaerobic Digester for Food Waste. Applied Energy, v. 194.

Zhao, W.X., Yang, H.Z., He, S.F., Zhao, Q.L., Wei, L.L., 2021, A review of biochar in anaerobic digestion to improve biogas production: Performances, mechanisms and economic assessments Bioresour. Technol., 341.

7. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.