Vol. 24, No. 3(2025) IA25570
Revista Mexicana de Ingeniería Química
Ingeniería Ambiental

Influence of chemical indicators on biogas production in a bioreactor from water hyacinth using canonical correlation analysis

Influencia de los indicadores químicos en la producción de biogás en un biorreactor a partir de lirio acuático mediante análisis de correlación canónico

D. B. Benitez-Suarez [1], H. Bautista-Zaragoza[1], J. P. Molina-Aguilar[2], J. Apolinar-Cortés[1], M. C. Chávez-Parga[1‡]
[1]Graduate Studies Division, Master of Science in Environmental Engineering, Chemical Engineering Faculty, Universidad Michoacana de San Nicolás de Hidalgo, Building V-1, Morelia, Michoacán, México.
[2]Civil Engineering Faculty, Universidad Michoacana de San Nicolás de Hidalgo, Building -C, Morelia Michoacán, México.
Sent date: April 14, 2025; Accepted: August 6, 2025

Abstract
The global energy crisis is currently a matter of great concern. With the rapid growth of the population, especially in emerging economies, energy supply often struggles to keep pace with demand. Biogas production from water hyacinth (*Eichhornia crassipes*) in an upflow anaerobic sludge blanket reactor offers a promising renewable energy alternative with multiple benefits. The anaerobic digestion of organic matter generates biogas, which holds significant potential for electricity and heat generation and can also be upgraded into a usable fuel. This study applied canonical correlation analysis to evaluate the relationship between sets of independent and dependent variables involved in the biogas production process. A 20 L anaerobic reactor upflow fed with water hyacinth pretreated with calcium oxide. Physicochemical variables were measured for the substrate, inoculum, effluent, and the biogas produced. The analysis yielded a canonical correlation coefficient of 0.8467 between the two variable sets, indicating a relatively strong relationship. Moreover, biogas production was estimated based on the input variables using the canonical variables derived from the analysis. These results demonstrate that canonical correlation analysis is a valuable tool for monitoring and optimizing the biogas production process, as it helps identify critical variables and their effects on reactor performance.
Keywords: anaerobic digestion, biogas, canonical correlation, canonical variables.

Resumen
La problemática energética mundial es un tema de gran relevancia en la actualidad. Con el rápido crecimiento de la población, especialmente en las economías emergentes, la oferta a menudo no puede mantener el ritmo de la demanda energética. La producción de biogás, a partir de lirio acuático (*Eichhornia crassipes*) en un reactor anaerobio de flujo ascendente es una alternativa para producir energía renovable que ofrece múltiples beneficios. La digestión anaerobia de la materia orgánica produce biogás, el cual tiene un gran potencial para su uso en la generación de energía eléctrica y calorífica, así como también puede ser transformado en combustible. El presente trabajo aplicó el análisis de correlación canónica para evaluar la relación entre los conjuntos de las variables independientes y dependientes del proceso de producción de biogás. Se utilizó un reactor anaerobio de flujo ascendente de 20 L alimentado con lirio acuático pretratado con óxido de calcio donde se midieron las variables fisicoquímicas del sustrato, el inóculo, el efluente y el biogás producido. Los resultados del análisis muestran un coeficiente de correlación canónica de 0.8467 entre ambos conjuntos, lo que indica una relación relativamente alta entre las variables, además se estimó la producción de biogás en función de las variables de entrada usando las variables canónicas obtenidas por el análisis. Estos resultados muestran que el análisis de correlación canónica es una herramienta útil para monitorear y optimizar el proceso de biogás, debido a que permite identificar las variables críticas y su efecto sobre el rendimiento del biorreactor.
Palabras clave: digestión anaerobia, biogás, correlación canónica, variables canónicas.

‡*Corresponding author. E-mail*: cparga@umich.mx;
https://doi.org/10.24275/rmiq/IA25570
ISSN:1665-2738, *issn-e*: 2395-8472

1 Introduction

Energy is essential for human well-being and daily life. Many countries, especially those in the developing world, face energy crises due to their heavy reliance on fossil fuels (Mohammed *et al.,* 2024). Approximately 80% of global energy is derived from fossil sources, such as oil, coal, and natural gas, which are the primary emitters of greenhouse gases, including carbon dioxide (CO_2) and methane (CH_4). These gases significantly contribute to global warming and environmental degradation (El-Araby, 2024; Gupta *et al.,* 2023; Hou *et al.,* 2023). In 2023, global natural gas consumption reached nearly four trillion cubic meters (Statista, 2024). Considering this demand, there is an urgent need to explore renewable energy alternatives, such as biogas, that could potentially replace natural gas in the future (Cisneros *et al.,* 2021).

Biogas is produced through the anaerobic degradation of organic matter and typically contains 55–70% CH_4, approximately 35–40% CO_2, and nitrogen (N_2) (Jameel *et al.,* 2024). It also includes trace amounts of other gases such as hydrogen sulfide (H_2S), hydrogen (H_2), ammonia (NH_3), oxygen (O_2), and carbon monoxide (CO) (Khan *et al.,* 2017). Biogas can be utilized for electricity generation, heating, and as a biofuel (Martínez-Gutiérrez, 2018).

Biowaste generated from commercial, industrial, and household activities can be broadly classified into two main fractions. The first is the *organic fraction of municipal solid waste* (OFMSW), primarily consisting of food scraps and kitchen waste, and typically representing the largest share of the total waste stream. The second is the *lignocellulosic or vegetal fraction*, which comprises plant residues from public and private green areas. Proper separation and management of both fractions, particularly the OFMSW, is essential for enhancing waste valorization and minimizing landfill impacts, as evidenced in recent urban waste management studies (Rolewicz-Kalińska *et al*., 2020; Alves *et al*., 2023). Effective management of these waste streams is critical due to their environmental impact, including greenhouse gas emissions and contamination issues (Triviño-Pineda *et al.,* 2024). The anaerobic digestion of organic waste enables resource recovery by producing biogas and digestate, thereby closing the loops of energy and nutrient cycles (González *et al.,* 2024). This process is framed within the circular economy model, in which waste is reintegrated as a resource in productive systems, promoting more efficient and sustainable use of natural resources (Stylianou *et al.,* 2023).

In recent years, the use of lignocellulose-rich organic waste for anaerobic digestion aimed at producing biotechnologically valuable byproducts such as methane (CH_4) has become an increasingly complex challenge (Amiri & Karimi, 2018). Currently, the use of water hyacinth (*Eichhornia crassipes*) as a substrate to produce these byproducts has emerged as an efficient and cost-effective alternative for biogas generation. This invasive aquatic weed can double its biomass within 6 to 28 days, forming dense mats on the water's surface. These mats obstruct sunlight penetration, leading to eutrophication and the eventual degradation of aquatic ecosystems (Pottipati *et al.,* 2021). Biogas production is carried out through anaerobic digestion, a process composed of four sequential stages: hydrolysis, acidogenesis, acetogenesis, and methanogenesis. During this process, the substrate is degraded by a consortium of facultative and strict anaerobic microorganisms operating under controlled conditions, such as those provided by upflow anaerobic sludge blanket (UASB) reactors (Chiemchaisri & Visvanathan, 2018; Themelis & Uloa, 2007).

Optimizing process parameters, such as inoculum concentration and incubation temperature, significantly contributes to enhancing biogas production (Armah *et al.,* 2018). According to Kumar *et al.* (2019), mesophilic conditions (40°C) are more favorable for biogas generation compared to lower temperatures, such as 30 °C. Biogas production is closely linked to the diversity and dynamics of the anaerobic digestion process, which are strongly influenced by operational factors, including temperature, pH, hydraulic retention time, carbon-to-nitrogen (C/N) ratio, organic loading rate, substrate composition, and nutrient bioavailability (Nakasima-López *et al.,* 2017). Therefore, microbial community diversity and activity are critical variables that can be affected by environmental and biological factors (Rehman *et al.,* 2019). However, monitoring and analysis of these variables in anaerobic reactors remain limited.

In this context, multivariate analytical tools such as canonical correlation analysis (CCA) have gained importance, as they enable the simultaneous evaluation of multiple variables, facilitating the identification of those that most significantly influence the process (Badii & Castillo, 2017). CCA stands out as a powerful multivariate technique for assessing and optimizing the performance of anaerobic bioreactors, enabling the identification and interpretation of relationships between multiple input and output variables and offering a comprehensive view of process dynamics.

This study applied CCA to improve the control and management of the anaerobic digestion (AD) process. By examining both input and output variables, CCA enabled the identification of relationships between independent variables and the output variable (biogas), thereby clarifying their impact on biogas production. The findings highlight the potential of CCA to identify critical variables and enhance overall system performance (Rehman *et al.,* 2019; Zhuang *et al.,* 2020;

Hotelling, 1936; Molina *et al.,* 2019).

Therefore, this study contributes to current knowledge by applying CCA to identify the most relevant physicochemical predictors of biogas production in a mono-digestion system using lime-pretreated water hyacinth. Unlike traditional studies based on bivariate correlations or univariate models, the multivariate approach employed here enables a more comprehensive assessment of operational interactions under real-world conditions. This type of analysis has proven to be more robust and representative in anaerobic systems, as it simultaneously captures the complex dynamics between operational parameters and biological responses, as reported by Otto *et al.* (2024) in a comparative study of 80 full-scale digesters. The findings of this work support the development of predictive frameworks for process monitoring and control, particularly in systems that utilize invasive aquatic biomass under mesophilic conditions.

2 Materials and methods

The methodological framework employed in this study integrates the design and operation of a pilot-scale UASB bioreactor fed with water hyacinth pretreated with calcium hydroxide, along with standardized analytical procedures for measuring pH, electrical conductivity, and alkalinity across various system streams. Physicochemical characterization was complemented by a multivariate statistical approach based on CCA, which enabled the identification of optimal statistical projections between the sets of independent and dependent variables. The implementation of this methodological strategy ensures the statistical robustness and practical relevance of the findings under mesophilic conditions, which are representative of real-world operations.

2.1 Bioreactor characteristics

The system from which data were obtained for the canonical correlation analysis (CCA) was a pilot-scale upflow anaerobic sludge blanket (UASB) reactor (**Figure 1**), constructed from polymethylmethacrylate (acrylic) with a total volume of 33 L. A valve is located at the bottom of the reactor for purging sludge and effluent, connected to a diffuser that ensures even distribution of the feed within the reactor. At the top, the effluent outlet is equipped with a gas trap to prevent the loss of generated biogas. Additionally, a biogas outlet is located at the top and is connected to an external high-density polyethylene (HDPE) system with a 20 L capacity for biogas storage and quantification. Biogas volume is measured by liquid displacement.

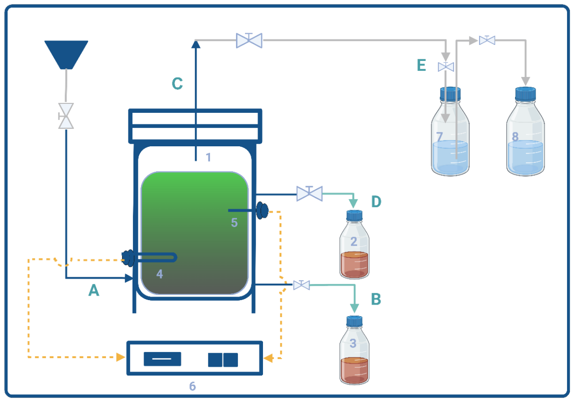

Figure 1. Schematic diagram of the pilot-scale UASB reactor. (**1**) UASB reactor; (**2**) Effluent outlet; (**3**) Sludge purge outlet; (**4**) Electric heating element; (**5**) Temperature sensor; (**6**) Temperature controller; (**7**) Biogas storage unit; (**8**) Displaced liquid collector; (**A**) Influent (feed); (**B**) Sludge purge; (**C**) Biogas outlet; (**D**) Effluent (discharge); (**E**) Liquid displacement.

The bioreactor is equipped with a temperature control system designed for automatic regulation, utilizing an I-Power Electronics STC-1000 controller. Two 60 W heating elements are installed inside the reactor to maintain a constant operating temperature of 35°C ± 1 °C.

2.2 Substrate characteristics

The water hyacinth used as a substrate in the bioreactor was collected from the "Grande" River, which runs through the city of Morelia, Michoacán, at coordinates 19.685611, -101.242114. After collection, the biomass underwent a pretreatment process that involved grinding the hyacinth in a disc mill (Estrella) to obtain particles with diameters ranging from 0.2 to 0.8 cm. Subsequently, a CaO solution heated to 60 °C was added to solubilize the organic matter. The mixture was then filtered to separate larger lignocellulosic fibers.

The preparation and feeding procedure of the substrate is illustrated in **Figure 2**. After harvesting, the water hyacinth was mechanically processed using a disc mill to reduce particle size and then filtered to eliminate coarse lignocellulosic residues. The resulting homogeneous slurry was transferred to the bioreactor using a peristaltic pump and combined with the inoculum before anaerobic digestion. The system incorporated temperature regulation at 35 °C and included dedicated outlets for biogas collection, effluent discharge, and sludge purging.

The filtered substrate was used to feed the UASB reactor, which had a total volume of 5 L. The volume of biogas produced was determined by measuring the volume of displaced liquid (Walker *et al.,* 2009).

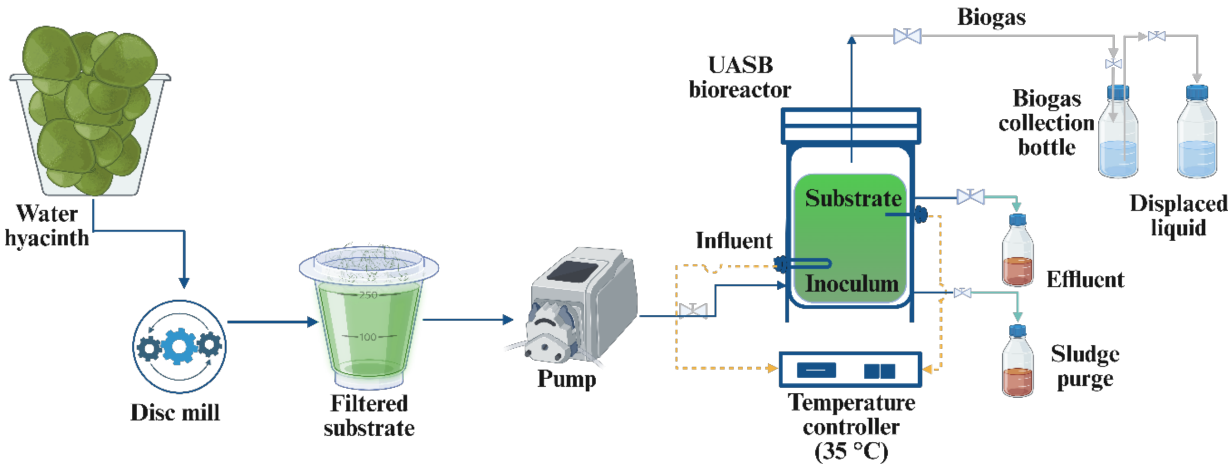

Figure 2. Schematic representation of the feeding and operation process of the UASB bioreactor. Water hyacinth is initially processed in a disc mill to reduce particle size, followed by filtration to remove coarse lignocellulosic fibers and obtain a homogeneous substrate. This filtered substrate is then pumped into the reactor and mixed with the inoculum. The bioreactor operates under mesophilic conditions (35 °C), maintained by an internal heating system and monitored via a temperature sensor. During anaerobic digestion, the generated biogas is collected and quantified by liquid displacement in an external gas collection and measurement system. The system also features separate outlets for sludge purge and effluent discharge, enabling operational monitoring and maintenance.

2.3 *Analytical parameter determination*

pH and conductivity measurements were conducted at the influent, effluent, and inoculum valves every 72 hours using a multiparameter instrument (Thermo Scientific Orion Star A325 portable pH/conductivity meter), by Mexican standards NMX-AA-008-SCFI-2016 and NMX-AA-093-SCFI-2000, respectively.

Alkalinity was determined in the influent, effluent, and inoculum of the reactor, following the NMX-AA-036-SCFI-2001 standard, with a frequency of every 72 hours, to calculate the total alkalinity as $CaCO_3$ in mg/L.

2.4 *Canonical correlation analysis*

Before conducting the canonical correlation analysis, the dataset, consisting of 22 paired observations collected over a 258-day monitoring period, was evaluated to ensure compliance with statistical assumptions. All variables were standardized using z-score transformation to eliminate the effect of differing measurement scales. Multicollinearity was assessed by analyzing the Pearson correlation matrix of the independent variable set. No correlation coefficients exceeded $|r| = 0.9$, indicating no significant redundancy among predictors (Hair *et al*., 2009). Additionally, potential outliers were identified graphically using boxplots for all ten variables included in the model. Although some mild outliers were observed, particularly in inoculum alkalinity and effluent pH, no extreme values were detected. Consequently, all data points were retained in the canonical correlation analysis, as their presence did not compromise the robustness or validity of the model.

Despite the moderate sample size (22 paired cases), the model meets established statistical criteria for multivariate analysis. Recent findings by Helmer *et al*. (2024) suggest that canonical correlation can yield stable results in moderately sized datasets, particularly when strong relationships exist between variable sets. Furthermore, the widely accepted "one-in-ten" rule recommends a minimum of 10 observations per variable, a condition fulfilled in this study by including 10 variables and surpassing the 1:1 ratio. Therefore, the statistical model complies with the methodological requirements for its application.

All statistical analyses, including canonical correlation analysis, were performed using the Statistica® software.

The use of linear multivariate statistical analysis allows for the identification of optimal correlations between two sets of variables. According to Hair *et al*. (2009), this type of correlation simultaneously links p dependent variables with q independent variables by linearly combining both sets to establish two weight vectors that maximize the corresponding correlation coefficient.

The variables $x_1, x_2, \ldots, x_p$ are grouped to form an independent set (X), which generates the composite variable (U). Similarly, the variables $y_1, y_2, \ldots, y_q$ form the dependent set (Y), resulting in the composite variable(W). Both U and W are referred to as canonical variables.

These canonical variables are obtained by multiplying the transposed weight vectors $a = \left(a_1, a_2, \ldots, a_p\right)$ y $b = \left(b_1, b_2, \ldots, b_q\right)$ by the variable sets X and Y, respectively. The result defines their linear combinations with the maximum possible variance, based on the orthogonality between the canonical variables.

$$U = a^T X = \left(a_1, a_2, \ldots, a_p\right)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = a_1 x_1 + a_2 x_2 + \ldots + a_p x_p \tag{1}$$

$$W = b^T Y = \left(b_1, b_2, \ldots, b_q\right)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{pmatrix} = b_1 y_1 + b_2 y_2 + \ldots + b_q y_q \tag{2}$$

Based on both results, the sample variance-covariance matrix (S), of order *p x q,* is established, noting that $C_{YX} = C_{XY}^T$

$$S = \begin{pmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{pmatrix} \tag{3}$$

Thus, the correlation coefficient between the canonical variables is referred to as the canonical correlation coefficient, and is defined as follows:

$$corr(U, W) = \frac{cov(U, W)}{\sqrt{var(U)}\sqrt{var(W)}} = \frac{a^T C_{XY} b}{\sqrt{a^T C_{XX} a}\sqrt{b^T C_{YY} b}} \tag{4}$$

In the above equation, the denominator imposes two normalization constraints, both of which determine the weight vectors that favor maximization (Khalil *et al.*, 2011).

$$var(U) = a^T C_{XX} a = 1 \tag{5}$$

$$var(W) = b^T C_{YY} b = 1 \tag{6}$$

Additionally, both constraints standardize the canonical variations, thereby establishing the solution to the canonical correlation problem as the maximization of the numerator.

$$corr(U, W) = a^T C_{XY} b \tag{7}$$

To achieve this, a two-parameter Lagrangian function is used, which enables the calculation of multivariate gradients by considering the solution weight vectors.

$$L(a, b) = a^T C_{XY} b - \tau_1 \left(a^T C_{XX} a - 1\right) - \tau_2 \left(b^T C_{YY} b - 1\right) \tag{8}$$

Each gradient requires the partial derivative of Equation (8) to be defined with respect to each vector, respectively.

$$\frac{\partial L(a, b)}{\partial a} = C_{XY} b - 2\tau_1 C_{XX} a = 0$$

$$C_{XY} b = 2\tau_1 C_{XX} a \tag{9}$$

$$\frac{\partial L(a, b)}{\partial b} = C_{XY}^T a - 2\tau_2 C_{YY} b = 0 \tag{10}$$

$$C_{XY}^T a = 2\tau_2 C_{YY} b \tag{11}$$

Both gradients yield a system of equations which, when solved, define the weight vectors a and b that maximize the numerator in Equation (7).

This solution is based on Equations (5) and (6), which are respectively multiplied by the transposed vectors from Equations (9) and (11) for simplification.

$$(C_{XY} b = 2\tau_1 C_{XX} a) a^T$$

$$a^T C_{XY} b = 2\tau_1 \tag{12}$$

$$\left(C_{XY}^T a = 2\tau_2 C_{YY} b\right) b^T$$

$$b^T C_{XY}^T a = 2\tau_2 \tag{13}$$

The algebraic simplification of the second term in Equations (12) and (13) defines the scalar of the solution, ensuring that the normalization constraints are satisfied

$$\left(a^T C_{XY} b\right)^T = b^T C_{XY}^T a = 1 \tag{14}$$

The equations are equivalent, defined as $2\tau_1 = 2\tau_2$, which represent the eigenvalue (λ). Therefore, Equations (9) and (11) can be rewritten in their simplified form

$$C_{XY} b = \lambda C_{XX} a \tag{15}$$

$$C_{XY}^T a = \lambda C_{YY} b \tag{16}$$

By solving for vector, a from Equation (15) and substituting it into Equation (16), the system is reduced as follows:

$$C_{XY}^T \left[\frac{1}{\lambda} C_{XX}^{-1} C_{XY} b\right] = \lambda C_{YY} b$$

$$C_{YY}^{-1} C_{XY}^T C_{XX}^{-1} C_{XY} b - \lambda^2 b = 0 \tag{17}$$

Analogously, we obtain:

$$C_{XX}^{-1} C_{XY} C_{YY}^{-1} C_{XY}^T a - \lambda^2 a = 0 \tag{18}$$

Finally, solving Equations (17) and (18) yields the vectors a and *b,* which represent the eigenvectors that maximize the correlation between the canonical variables for the given variable sets in the problem.

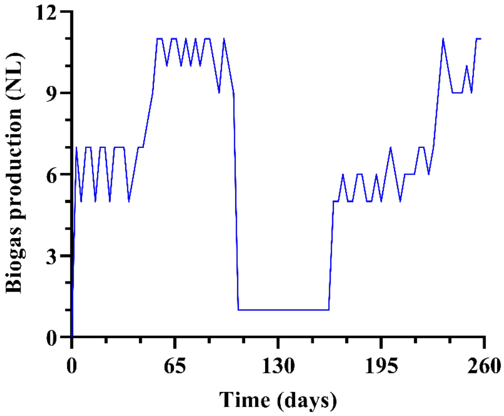

Figure 3. Normalized biogas production (NL) over a 258-day mono-digestion of calcium hydroxide pretreated water hyacinth. Biogas was measured every 72 h and standardized to 25 °C and 1 atm. Temporal fluctuations highlight periods of low production and recovery, corresponding to shifts in physicochemical parameters.

3 Results and discussion

The canonical correlation analysis enabled the identification of significant relationships between the input and output variables of the bioreactor over a 258-day monitoring period, with measurements taken every 72 hours (**Figure 3**).

The evaluated variables included pH, electrical conductivity, and alkalinity of the substrate and inoculum as input factors, while pH, electrical conductivity, alkalinity, and biogas production were considered as output variables (**Table 1** and **Table 2**). These variables were selected based on their operational relevance and their documented influence on anaerobic digestion performance (Ali *et al*., 2021; Chen *et al*., 2024). They serve as key indicators of buffering capacity, ionic strength, and microbial activity in anaerobic systems. They are easily measurable through routine monitoring, making them suitable for real-time decision-making and process control.

Table 1. Input variables used in the canonical correlation analysis.

Periods Days	Initial conditions					
	Substrate pH (S_{pH})	Inoculum pH (I_{pH})	Substrate electrical conductivity ($S_{E.C}$) μS/cm	Inoculum electrical conductivity ($I_{E.C}$) μS/cm	Substrate alkalinity (S_A) g $CaCO_3$/L	Inoculum alkalinity (I_A) g $CaCO_3$/L
0	10.60	8.01	4.96	3.52	4.30	11.00
12	11.21	8.08	3.13	4.97	5.50	36.00
24	9.47	7.62	8.37	6.66	1.80	12.00
36	9.13	7.63	8.56	8.33	2.00	15.00
48	9.85	7.36	4.32	7.68	2.00	26.00
60	9.16	7.62	6.47	6.22	2.00	13.00
72	9.24	7.53	7.18	7.86	2.50	17.00
84	9.32	7.54	9.32	7.54	2.00	15.00
96	9.81	7.52	5.01	8.21	2.70	18.00
108	9.09	7.35	8.32	8.09	2.70	17.00
120	9.04	7.50	8.12	12.24	2.90	10.00
132	8.19	7.19	9.33	10.65	2.50	28.00
144	7.88	7.70	12.93	9.66	3.40	20.00
156	7.74	7.69	11.94	10.18	3.10	20.00
168	7.37	7.73	4.76	9.00	1.60	19.00
180	7.73	7.47	12.59	10.66	3.10	26.00
192	7.73	7.56	14.09	10.90	3.60	16.00
204	7.62	7.71	16.60	13.46	3.20	15.00
216	7.85	7.60	10.83	10.88	4.60	16.00
228	8.02	7.62	10.97	11.25	3.70	10.00
240	7.61	7.59	12.05	11.64	2.80	13.00
252	7.73	7.43	9.99	10.57	4.10	15.00
258	7.91	7.46	13.79	11.59	3.80	18.00

Table 2. Response variables used in the canonical correlation analysis.

Periods Days	Response variable			
	Biogas produced NL	Effluent pH (E_{pH})	Effluent electrical conductivity ($E_{E.C}$) μS/cm	Effluent alkalinity (E_A) g $CaCO_3$/L
0	7.33	7.20	5.11	2.10
12	7.33	8.09	4.96	4.50
24	5.50	7.59	8.62	2.50
36	5.50	7.61	9.67	2.40
48	8.25	7.30	8.88	2.20
60	10.08	7.43	7.11	1.80
72	10.99	7.34	9.46	2.80
84	10.99	7.50	7.12	2.90
96	10.99	7.54	10.01	3.70
108	0.92	7.76	9.76	2.80
120	0.92	7.48	11.42	3.60
132	0.92	8.02	10.40	3.90
144	0.92	7.80	16.36	4.20
156	0.92	7.83	11.94	3.80
168	5.50	7.82	11.82	3.30
180	6.41	7.78	12.58	3.30
192	6.41	7.66	12.53	4.70
204	6.41	7.67	14.35	4.70
216	6.41	7.69	12.45	5.00
228	7.33	7.72	12.90	5.70
240	9.16	7.73	14.19	5.90
252	9.16	7.89	12.74	5.60
258	10.99	7.59	13.61	5.80

Their inclusion enabled a robust physicochemical characterization of the system, which is essential for developing multivariate statistical models, such as canonical correlation analysis, that require well-defined relationships between sets of independent and dependent variables.

Figure 4. Illustrates the relationship between the canonical variables **U** and **W**, obtained through canonical correlation analysis, which identifies the optimal linear combinations of input and output variables that maximize their mutual correlation. The linear fit between these variables resulted in a slope of **0.9202** and a correlation coefficient of **R^2 = 0.8467**, indicating that approximately **84.67%** of the variability in the dependent variable **W** can be explained by the independent variable **U**. The canonical functions were defined by the Equations (19) and (20):

$$U = a_1 S_{pH} + a_2 I_{pH} + a_3 S_{C.E} + a_4 I_{C.E} + a_5 S_A + a_6 I_A \quad (19)$$

$$W = b_1 E_{pH} - b_2 E_{C.E} - b_3 E_A \quad (20)$$

Each coefficient represents the statistical weight assigned to the original variables within the canonical combination. These functions serve as new axes in a transformed space, analogous to a Cartesian coordinate system, allowing for a graphical representation of the observations projected onto these axes.

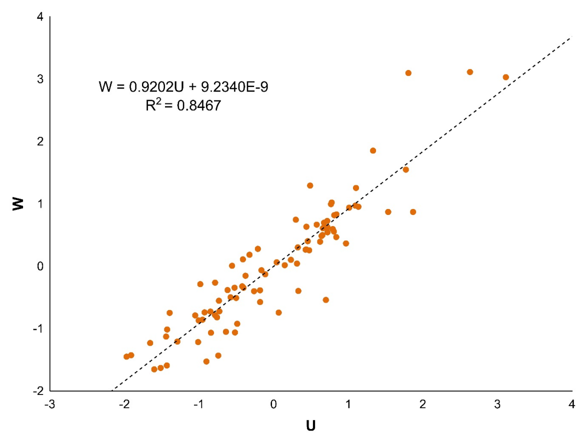

Figure 4. Canonical correlation analysis between the recorded values of the physicochemical variables in the bioreactor and the corresponding canonical variable values for estimating biogas production.

This transformation facilitates the visual interpretation of the multivariate relationship between the sets of predictor and response variables, providing a powerful tool for analyzing underlying structures in complex systems.

The eigenvalues obtained for the canonical functions were λ_1 = 0.8467, λ_2 = 0.2996, λ_3 = 0.0973, and λ_4 = 0.0697, with λ_1 representing the highest explained variance. These results highlight the importance of the first canonical function, which accounts for the most significant portion of the relationship between the two variable sets. The coefficients of the first weight vectors were a = (-0.2211, 0.0801, 0.3313, 0.5170, 0.1056, -0.0717) and b = (-0.0327, 0.0064, 0.8338, 0.2094), indicating that substrate electrical conductivity and alkalinity are key factors influencing biogas production (Cao *et al.,* 2019). This finding is consistent with previous studies that have identified electrical conductivity as a potential indicator of microbial activity in anaerobic digestion systems (Callegari *et al.,* 2025). Additionally, Alvarado-Reyna *et al*. (2024) demonstrated that electrical conductivity displays greater sensitivity and earlier deviation compared to conventional indicators such as pH or methane content, reinforcing its value as a predictive variable within multivariate models such as canonical correlation analysis.

This finding suggests that the electrochemical characteristics of the environment, represented by electrical conductivity, may modulate nutrient availability and microbial performance in anaerobic systems. Similar associations have been reported by Callegari *et al*. (2025) and Hasani *et al*. (2025), who linked shifts in conductivity and alkalinity to microbial community adaptation and process efficiency under varying loading conditions. This reinforces the role of these variables not only as statistical predictors but also as functional indicators of biological responses.

These relationships show that increases in both electrical conductivity and alkalinity may reflect not only changes in buffering capacity and ion exchange but also shifts in microbial metabolic networks and community structure, which in turn impact biogas yield and system stability. For instance, empirical dynamic modeling of anaerobic digesters converting sucrose has revealed that temporal variations in microbial interaction networks (particularly among fermentative, syntrophic, and methanogenic functional groups) closely match performance dynamics, such as hydrogen concentrations and methane production, throughout operation (Chang *et al*., 2025; Goux *et al*., 2015). Highlighting similar mechanisms in lignocellulosic digestion, Al Hasani *et al*. (2025) reported that modifications in the electrochemical matrix enhance direct interspecies electron transfer (DIET), promoting microbial resilience and improving methane output (Negi *et al*., 2025). These findings validate that the observed statistical associations with conductivity and alkalinity are supported by functional microbial responses, reinforcing their usefulness for both process monitoring and mechanistic understanding.

Based on the canonical correlation equation derived from the analysis, a biogas production volume of 3.2579 L was estimated under the initial experimental conditions. However, by modifying the substrate's electrical conductivity and alkalinity values in Equation (21), the calculations suggest a potential biogas yield of up to 18.48 L. What suggests that proper regulation of these variables could optimize system performance, aligning with previous studies that highlight the importance of alkalinity in maintaining the stability of the anaerobic digestion process (Soto *et al.,* 2021).

$$Biogas = \frac{0.9202\ (a_1\ S_{pH} + a_2 I_{pH} + a_3 S_{C.E} + a_4 I_{C.E} + a_5 S_A + a_6 I_A) + 0.0000000092349\ (-b_1 E_{pH} - b_2 E_{C.E} - b_3 E_A)}{b_4} \tag{21}$$

Although the correlation values obtained suggest a strong association between the input variables and biogas production, it is essential to note that this relationship does not imply causation. The 15.33% variability in biogas production not explained by the model may be attributed to other unaccounted factors, such as substrate composition, temperature, the presence of inhibitors, or specific microbial activity within the bioreactor. This unexplained variance may also reflect the heterogeneous composition of the water hyacinth biomass or adaptive fluctuations in the inoculum´s microbial consortia, as suggested by Karouach *et al*. (2023). Recent studies have demonstrated that substrate composition, particularly the content of biodegradable organic matter, has a significant impact on biogas production efficiency (Chew *et al.,* 2021).

Despite the high correlation observed, it is important to recognize that biological systems exhibit inherent variability that is not always fully captured by mathematical models. Factors such as microbial dynamics, substrate composition, and operational conditions of the bioreactor remain sources of uncertainty in predicting biogas yield. Therefore, it is recommended that complementary studies be conducted with greater control over internal system variables to improve model accuracy and further optimize the anaerobic digestion process.

In the present study, electrical conductivity and alkalinity were identified as key predictors in the canonical model. These variables were also identified by Hernández-Eugenio *et al*. (2025) as being strongly associated with shifts in microbial community structure, particularly under varying organic loading rates and varying buffering capacities. While their study focused on microbial adaptation in co-digestion systems using metagenomic tools, our approach applied a statistical framework (canonical correlation analysis) to a mono-digestion system using lime-pretreated water hyacinth. This convergence reinforces the idea that physicochemical indicators such as those measured in this study may also reflect microbial activity and resilience. Therefore, the observed statistical relationships gain biological relevance when interpreted in the context of known microbial responses to environmental conditions.

Another essential aspect to consider is the applicability of the canonical equation in closed biological systems. Although the predictions suggest an increase in biogas production with specific adjustments to the input variables, in practice, the ability to modify these variables is limited by the system's intrinsic conditions. Microbial dynamics within a bioreactor can be influenced by subtle changes in pH or ion concentration, which may result in adverse effects on the system's equilibrium (Martínez *et al.,* 2016).

Finally, it is recommended to conduct controlled experiments that evaluate not only the physical and chemical variables but also those related to the metabolism of the microorganisms present in the bioreactor. The integration of more complex mathematical models, incorporating factors such as organic loading rate and microbial kinetics, could provide a more accurate representation of the system and enhance the predictive capacity of the proposed canonical model. In this regard, future studies could focus on the inclusion of additional variables and the experimental validation of the predictions presented in this work.

4 Conclusions

4.1 General conclusions

This study demonstrated the influence of input and output physicochemical variables on biogas production in an Upflow anaerobic sludge blanket (UASB) reactor, using canonical correlation analysis as a multivariate statistical tool. The results indicated that substrate electrical conductivity and alkalinity are key factors in system efficiency, with a predictive model that accounted for 84.67% of the variability in biogas production. These findings demonstrate the potential of canonical correlation analysis as a valuable tool for analyzing and optimizing biotechnological processes.

From a practical perspective, the ability to adjust variables such as electrical conductivity and alkalinity represents a viable strategy for enhancing system performance. However, implementing these adjustments under real operating conditions requires controlled experimental trials to evaluate their impact on microbial dynamics, system stability, and the quality of the biogas produced.

Despite the model's high explanatory power, it is essential to acknowledge that biological systems exhibit inherent variability, which may limit the predictive capacity of purely statistical approaches. To strengthen the model's validity, factors such as substrate composition, specific microbial activity, temperature, and the presence of inhibitors should be considered in future studies.

4.2 Future research directions

It is recommended that this statistical approach be integrated with complementary tools such as metagenomic analysis, kinetic modeling, and real-time monitoring of critical parameters to develop more precise and robust control strategies. Such integration would facilitate the advancement toward more efficient, stable, and scalable anaerobic digestion systems for applications in organic waste treatment and renewable energy generation.

Additionally, future research should integrate the analysis of microbial consortium dynamics or functional metabolic monitoring to validate the statistical associations observed. What would strengthen the causal interpretation of physicochemical predictors, as proposed by Helmer *et al*. (2024).

Acknowledgements

We express our sincere gratitude to the SECIHTI for awarding the master's scholarship (No. 002528) and to the Graduate Studies Division of the Faculty of Chemical Engineering at the Universidad Michoacana de San Nicolás de Hidalgo (UMSNH) for providing the space and resources necessary for the completion of this work.

Nomenclature

$x_1, x_2, \ldots, x_p$	Independent variables
$y_1, y_2, \ldots, y_q$	Dependent variables
X	Vector grouping all independent variables
Y	Vector grouping all dependent variables
U	Canonical variate associated with the independent variable set X
W	Canonical variate associated with the dependent variable set Y
$a = (a_1, a_2, \ldots, a_p)^T$	Weight vector associated with the independent variable set
$b = (b_1, b_2, \ldots, b_q)^T$	Weight vector associated with the dependent variable set
λ	Eigenvalue representing the squared canonical correlation
τ_1, τ_2	Lagrange multipliers used in the optimization process

C_{XX}	Covariance matrix of the independent variables
C_{YY}	Covariance matrix of the dependent variables
C_{XY}	Cross-covariance matrix between independent and dependent variables
$C_{YX} = C_{XY}{}^T$	Transposed cross-covariance matrix
S	Combined covariance matrix (block matrix of order $p \times q$)
$var(U), var(W)$	Variance of canonical variables U and W
$cov(U, W)$	Covariance between the canonical variables
$corr(U, W)$	Canonical correlation coefficient
$L(a, b)$	Lagrangian function used to solve the canonical correlation maximization problem
$\partial L/\partial a, \partial L/\partial b$	Partial derivatives of the Lagrangian concerning a and b

References

Ali, H.A., Faraj, J.J., Hussien, F.M. (2021). Effect of pH on biogas production during anaerobic digestion. *Journal of University of Shanghai for Science and Technology*, *23*(8), 224–231. https://doi.org/10.51201/jusst/21/08369

Alvarado-Reyna, P., Albalate-Ramírez, A., García-Balandrán, E., Escamilla-Alvarado, C., Galván-Arzola, U., Miramontes-Martínez, L., & Rivas-García, P. (2024). Evaluation of the reaction capacity of early warning indicators to failures in biogas production systems. *Revista Mexicana de Ingeniería Química*, *23*(3), 1-30. https://doi.org/10.24275/rmiq/bio24313

Alves, D., Villar, I., Mato, S. (2023). Community composting strategies for biowaste treatment: methodology, bulking agent and compost quality. *Environmental science and pollution research*, *31*(7), 9873–9885. https://doi.org/10.1007/s11356-023-25564-x

Amiri, H., & Karimi, K. (2018). Pretreatment and hydrolysis of lignocellulosic wastes for butanol production: Challenges and perspectives. *Bioresource Technology*, *270*, 702-721. https://doi.org/10.1016/j.biortech.2018.08.117

Armah, E. K., Boamah, B. B., & Boakye, G. O. (2018). Impact of Water Hyacinth (Eicchornia crassipes) as a Feedstock for Biogas Production. *Chemical And Biomolecular Engineering*, *2*(4), 184. https://doi.org/10.11648/j.cbe.20170204.13

Badii, M. H., & Castillo, J. (2017). Análisis de correlación canónica (ACC) e investigación científica. *Revista Innovaciones de Negocios*, *4*(8). https://doi.org/10.29105/rinn4.8-9

Callegari, A., Tucci, M., Aulenta, F., Viggi, C. C., & Capodaglio, A. G. (2025). Anaerobic sludge digestion enhancement with bioelectrochemical and electrically conductive materials augmentation: A state of the art review. *Chemosphere*, *372*, 144101. https://doi.org/10.1016/j.chemosphere.2025.144101

Cao, Z., Jung, D., Olszewski, M. P., Arauzo, P. J., & Kruse, A. (2019). Hydrothermal carbonization of biogas digestate: Effect of digestate origin and process conditions. *Waste Management*, *100*, 138-150. https://doi.org/10.1016/j.wasman.2019.09.009

Chang, C., Chang, C., Lu, H., Hsieh, C., & Wu, J. (2025). Bioenergetically constrained dynamical microbial interactions govern the performance and stability of methane-producing bioreactors. *Npj Biofilms And Microbiomes*, *11*(1). https://doi.org/10.1038/s41522-025-00668-z

Chen, B., Azman, S., Crauwels, S., Dewil, R., & Appels, L. (2024). Mild alkaline conditions affect digester performance and community dynamics during long-term exposure. *Bioresource Technology*, *406*, 131009. https://doi.org/10.1016/j.biortech.2024.131009

Chew, K. R., Leong, H. Y., Khoo, K. S., Vo, D. N., Anjum, H., Chang, C., & Show, P. L. (2021). Effects of anaerobic digestion of food waste on biogas production and environmental impacts: a review. *Environmental Chemistry Letters*, *19*(4), 2921-2939. https://doi.org/10.1007/s10311-021-01220-z

Chiemchaisri, C., & Visvanathan, C. (2008). Greenhouse Gas Emission Potential of the Municipal Solid Waste Disposal Sites in Thailand. *Journal Of The Air & Waste Management Association*, ***58***(5), 629-635. https://doi.org/10.3155/1047-3289.58.5.629

El-Araby, R. (2024). Biofuel production: exploring renewable energy solutions for a greener future. *Biotechnology For Biofuels And Bioproducts*, *17*(1). https://doi.org/10.1186/s13068-024-02571-9

González, M. M. R., De León, E. M. D., Luque, L., Pitty, N., Arias, J., & Chen, A. (2024). Codigestión Anaeróbica de Lodos y Residuos Orgánicos Municipales en Clima Tropical. *I+D Tecnológico*, *20*(1), 82-93. https://doi.org/10.33412/idt.v20.1.3902

Goux, X., Calusinska, M., Lemaigre, S., Marynowska, M., Klocke, M., Udelhoven, T., Benizri, E., & Delfosse, P. (2015). Microbial community dynamics in replicate anaerobic digesters exposed sequentially to increasing organic loading rate, acidosis, and process recovery. *Biotechnology For Biofuels*, *8*(1). https://doi.org/10.1186/s13068-015-0309-9

Gupta, P., Kurien, C., & Mittal, M. (2022). Biogas (a promising bioenergy source): A critical review on the potential of biogas as a sustainable energy source for gaseous fuelled spark ignition engines. *International Journal Of Hydrogen Energy*, *48*(21), 7747-7769. https://doi.org/10.1016/j.ijhydene.2022.11.195

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C. (2009). *Multivariate data analysis* (7th ed.). Prentice Hall, Upper Saddle River, NJ.

Hasani, Z.A., Nayak, J.K., Balushi, N.J.A., Al-Mamun, A., Samal, K. (2025). Prospect of conductive materials in the anaerobic digester matrix for methane production: electron transfer and microbial communication. *Water*, *17*(9), 1321. https://doi.org/10.3390/w17091321

Helmer, M., Warrington, S., Mohammadi-Nejad, A., Ji, J.L., Howell, A., Rosand, B., Anticevic, A., Sotiropoulos, S.N., Murray, J.D. (2024). On the stability of canonical correlation analysis and partial least squares with application to brain-behavior associations. *Communications biology*, *7*(1). https://doi.org/10.1038/s42003-024-05869-4

Hernández-Eugenio, G., Espinosa-Solares, T., López-Ortiz, C., Meneses-Reyes, J.C., Ochoa-Bernal, T.G. (2025). Changes in microbial diversity and methane yield caused by overloading in systems of chicken litter, microalgae oil-free and glycerol in co-digestion. *Revista mexicana de ingeniería química*, *24*(2), IA25488. https://doi.org/10.24275/rmiq/IA25488

Hotelling, H. (1936). Relations between two sets of variants. *Biometrika*, *28*(3/4), 321–377.

Jameel, M. K., Mustafa, M. A., Ahmed, H. S., Mohammed, A. J., Ghazy, H., Shakir, M. N., Lawas, A. M., Mohammed, S. K., Idan, A. H., Mahmoud, Z. H., Sayadi, H., & Kianfar, E. (2024). Biogas: Production, properties, applications, economic and challenges: A review. *Results In Chemistry*, ***7***, 101549. https://doi.org/10.1016/j.rechem.2024.101549

Karouach, F., Bakrim, W.B., Ezzariai, A., Mnaouer, I., Ibourki, M., Kibret, M., Sobeh, M., Hafidi, M., Kouisni, L. (2023). Valorization of water hyacinth to biomethane and biofertilizer through anaerobic digestion technology. *Fuel*, *358*, 130008. https://doi.org/10.1016/j.fuel.2023.130008

Khalil, B., Ouarda, T., & St-Hilaire, A. (2011). Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *Journal Of Hydrology*, *405*(3-4), 277-287. https://doi.org/10.1016/j.jhydrol.2011.05.024

Khan, I. U., Othman, M. H. D., Hashim, H., Matsuura, T., Ismail, A., Rezaei-DashtArzhandi, M., & Azelee, I. W. (2017). Biogas as a renewable energy fuel – A review of biogas upgrading, utilisation and storage. *Energy Conversion And Management*, *150*, 277-294. https://doi.org/10.1016/j.enconman.2017.08.035

Kumar, V., Singh, J., Nadeem, M., Kumar, P., & Pathak, V. V. (2018). Experimental and Kinetics Studies for Biogas Production Using Water Hyacinth (*Eichhornia crassipes* [Mart.] Solms) and Sugar Mill Effluent. *Waste And Biomass Valorization*, *11*(1), 109-119. https://doi.org/10.1007/s12649-018-0412-9

Martínez, E. J., Gil, M. V., Fernandez, C., Rosas, J. G., & Gómez, X. (2016). Anaerobic Codigestion of Sludge: Addition of Butcher's Fat Waste as a Cosubstrate for Increasing Biogas Production. *PLoS ONE*, *11*(4), e0153139. https://doi.org/10.1371/journal.pone.0153139

Mohammed, M., Belkair, A., Hamad, T., Jirhiman, I., Hassan, R., & Ahmeedah, A. (2022). Improving biogas production from animal manure by batch anaerobic digestion. *Algerian Journal of Engineering and Technology*, 6, 79–84. https://doi.org/10.5281/zenodo.6561086

Molina-Aguilar, J. P., Gutiérrez-López, A., & Cruz-Paz, I. M. (2019). Correlación canónica entre volúmenes de almacenamiento en presas e intensidades de precipitación durante huracanes. *Tecnología y Ciencias del Agua*, *10*(6), 25-56. https://doi.org/10.24850/j-tyca-2019-06-02

Nakasima-López, M., Taboada-González, P., Aguilar-Virgen, Q., & Velázquez-Limón, N. (2017).

Adaptación de Inóculos Durante el Arranque de la Digestión Anaerobia con Residuos Sólidos Orgánicos. *Información Tecnológica*, *28*(1), 199-208. https://doi.org/10.4067/s0718-07642017000100020

Negi, S., Chai, J., Tjhin, A. C. T., & Pan, S. (2025). Electro-anaerobic digestion as carbon–neutral solutions. *Chemical And Biological Technologies In Agriculture*, *12*(1). https://doi.org/10.1186/s40538-025-00776-0

Otto, P., Puchol-Royo, R., Ortega-Legarreta, A., Tanner, K., Tideman, J., De Vries, S., Pascual, J., Porcar, M., Latorre-Pérez, A., & Abendroth, C. (2024). Multivariate comparison of taxonomic, chemical and operational data from 80 different full-scale anaerobic digester-related systems. *Biotechnology For Biofuels And Bioproducts*, *17*(1). https://doi.org/10.1186/s13068-024-02525-1

Pottipati, S., Yadav, K. D., & Kalamdhad, A. S. (2021). The Potential of Biogas Production from Water Hyacinth by Using a Floating Drum Biogas Reactor. *Springer eBooks*, pp. 215-223. https://doi.org/10.1007/978-3-030-70463-6_20

Rolewicz-Kalińska, A., Lelicińska-Serafin, K., Manczarski, P. (2020). The circular economy and organic fraction of municipal solid waste recycling strategies. *Energies*, *13*(17), 4366. https://doi.org/10.3390/en13174366

Soto, M., Ruiz, I. (2021). Co-digestión anaerobia para la obtención de biogás a partir de residuos forestales. Tesis de doctorado, *Universidad de A Coruña*. https://ruc.udc.es/dspace/handle/2183/29172

Statista. (2024). Global natural gas consumption 1998–2023. *Statista*. Disponible en: https://www.statista.com/statistics/282717/global-natural-gas-consumption/

Stylianou, M., Laifi, T., Bennici, S., Dutournie, P., Limousy, L., Agapiou, A., Papamichael, I., Khiari, B., Jeguirim, M., & Zorpas, A. A. (2023). Tomato waste biochar in the framework of circular economy. *The Science Of The Total Environment*, *871*, 161959. https://doi.org/10.1016/j.scitotenv.2023.161959

Themelis N.J. y Ulloa P.A. (2007). Methane generation in landfills. Renew. *Energy*, *32* (7), 1243-1257. https://doi.org/10.1016/j.renene.2006.04.020

Triviño-Pineda, J., Sanchez-Rodriguez, A., & Peláez, N. P. (2024). Biogas production from organic solid waste through anaerobic digestion: A meta-analysis. *Case Studies In Chemical And Environmental Engineering*, *9*, 100618. https://doi.org/10.1016/j.cscee.2024.100618

Walker, M., Zhang, Y., Heaven, S., & Banks, C. (2009). Potential errors in the quantitative evaluation of biogas production in anaerobic digestion processes. *Bioresource Technology, 100*(24), 6339-6346. https://doi.org/10.1016/j.biortech.2009.07.018

Zhuang, X., Yang, Z., & Cordes, D. (2020). A technical review of canonical correlation analysis for neuroscience applications. *Human Brain Mapping, 41*(13), 3807-3833. https://doi.org/10.1002/hbm.25090