

Article

Not peer-reviewed version

Methane Concentration Prediction in Anaerobic Codigestion Using Multiple Linear Regression with Integrated Microbial and Operational Data

[Iván Ostos](#)*, Iván Ruiz, [Diego Cruz](#)*, [Luz Marina Flórez-Pardo](#)

Posted Date: 28 August 2025

doi: 10.20944/preprints202508.2049.v1

Keywords: codigestion; metagenomics; biogas; MLR



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Methane Concentration Prediction in Anaerobic Codigestion Using Multiple Linear Regression with Integrated Microbial and Operational Data

Iván Ostos ^{1,*}, Iván Ruiz ¹, Diego Cruz ² and Luz Marina Flórez-Pardo ³

¹ Grupo de Investigación en Ingeniería Electrónica, Industrial, Ambiental, Metrología GIEIAM, Universidad Santiago de Cali, Cali, Colombia

² Grupo de Interacciones Microbianas GIM, Universidad de Salamanca, Salamanca, Spain

³ Grupo de Investigación en Modelado, Análisis y Simulación de Procesos Ambientales e Industriales PAI+, Universidad Autónoma de Occidente, Cali, Colombia

* Correspondence: iostos83@gmail.com

Abstract

Anaerobic codigestion of organic residues is a proven strategy for enhancing methane recovery. However, the complexity of microbial interactions and variability in operational conditions make it difficult to estimate methane concentration in real time, particularly in rural contexts. This study developed a multiple linear regression model to predict methane concentration using operational data and microbial community profiles derived from 16S rRNA gene sequencing. The system involved the codigestion of cassava by-product and pig manure in a two-phase anaerobic reactor. Predictor variables were selected through a hybrid approach combining statistical correlation with microbial functional relevance. The final model, trained on 70% of the dataset, demonstrated satisfactory generalization capability on the 30% test set, achieving a coefficient of determination (R^2) of 0.92 and a mean absolute error (MRE) of 6.50%. Requiring only a limited set of inputs and minimal computational resources, the model offers a practical and accessible solution for estimating methane levels in decentralized systems. The integration of microbial community data represents a meaningful innovation, improving prediction by capturing biological variation not reflected in operational parameters alone. This approach can support local decision-making and contribute to Sustainable Development Goal 7 by promoting reliable and affordable technologies for clean energy generation in rural and resource-constrained settings.

Keywords: codigestion; metagenomics; biogas; MLR

1. Introduction

The Valle del Cauca region is one of Colombia's most active agro-industrial areas, combining high agricultural productivity with unique ecological richness. The territory is sustained by ecosystems that range from coastal plains to montane forests, which support both biological diversity and productive capacity. It ranks as the third-largest producer and consumer of pork in the country, with a reported output of 88105 tons in 2023, equivalent to 15.6% of national production, and an average pig population exceeding 396000 animals [1,2]. This sector generates large volumes of pig manure (PM) that require appropriate handling to prevent environmental and public health risks. Another productive activity with growing regional relevance is cassava cultivation, which covered approximately 564 hectares in 2020, yielding a total of 9888 tons of fresh roots [1]. During starch extraction, each kilogram of cassava generates about 0.2 kilograms of starch, 0.65 kilograms of fibrous residue (cassava dregs (CD)), and between five and seven liters of wastewater [3]. Based on these ratios, the estimated annual generation of by-products in the region reaches nearly 6427 tons, most of which are not currently valorized [4–6].

To address the increasing accumulation of organic residues from pig farming and cassava processing, anaerobic digestion (AD) has been promoted in rural areas of Valle del Cauca as a strategy for energy recovery and waste management. In these settings, one phase tubular biodigesters are commonly employed due to their affordable construction, ease of installation and minimal infrastructure requirements, making them particularly attractive to smallholder producers [7,8]. AD is a biologically mediated process in which organic matter is sequentially transformed into methane through four main stages, each driven by specific microbial groups. In the hydrolysis phase, hydrolytic bacteria degrade complex macromolecules such as carbohydrates, proteins and lipids into soluble monomers. These compounds are then metabolized by acidogenic bacteria during acidogenesis, producing volatile fatty acids (VFA), alcohols and gases like hydrogen and carbon dioxide. In the subsequent acetogenesis stage, acetogenic microorganisms convert these intermediates into acetate, which, along with hydrogen and carbon dioxide, is used by methanogenic archaea in the final phase to generate methane [9].

For the process to remain stable and efficient, environmental conditions such as pH and temperature must be kept within optimal ranges, typically between 6.5 and 7.5 for pH and 30 to 38 °C under mesophilic conditions [10–13]. In addition, maintaining a C:N ratio between 20:1 and 30:1 is considered ideal for AD, as it ensures sufficient nitrogen for microbial growth without leading to ammonia inhibition or carbon limitation [14–17]. However, most rural systems lack monitoring tools and operate through empirical practices, without clear understanding of internal conditions or microbial dynamics [7]. This limitation frequently leads to process imbalance, reduced performance and early system failure.

To overcome the performance limitations of conventional digesters, several strategies have been developed to improve substrate biodegradability and enhance biogas production. Among them, mechanical pre-treatments, codigestion, and multiphase configurations have proven to be particularly effective in increasing system efficiency [18–22]. Mechanical pre-treatments have proven effective in enhancing the hydrolysis of lignocellulosic substrates by reducing particle size and fiber crystallinity, thus increasing surface area and enzymatic accessibility [23,24]. Depending on specific conditions, methane production improvements of 16% to 99% have been reported with mechanical treatments [25]. These results highlight the potential of simple mechanical treatments to enhance biodegradability and biogas productivity, especially during the hydrolysis and acidogenesis phases, which are often rate-limited in solid waste digestion.

Codigestion has emerged as a robust strategy to address the nutrient imbalances and low biodegradability often associated with single-substrate digestion. By combining complementary feedstocks, this approach improves the carbon to nitrogen (C:N) ratio, dilutes inhibitors, and stimulates microbial activity, allowing for higher energy yields [26]. For instance, it has been reported that co-digesting PM with cassava pulp at inclusion levels of up to 60% of the incoming volatile solids (VS) can increase the specific methane yield by around 41% compared with PM alone [27]. Similarly, results have also shown that mixtures containing 66% PM, 16% cassava pulp, and 16% bagasse achieve higher methane yields than those with high bagasse content alone, which led to pH imbalances and process failure [28]. Likewise, experimental trials combining sewage sludge with food waste reported an increase in methane yield from 159 to 799 mL CH₄/gVS, along with a reduction in hydraulic retention time (HRT) to less than four days [29]. These improvements are attributed to enhanced microbial synergy and substrate availability, which accelerate volatile solids degradation.

Multiphase AD systems have been developed to address the limitations of single-stage configurations by creating distinct operational environments for each metabolic phase. In two-phase systems, the acidogenic and methanogenic stages are physically separated, which enables more efficient substrate conversion, greater resilience to organic shocks, and better pH control [30,31]. This structural decoupling has led to increases in methane yields, improved volatile solids removal, and significant reductions in HRT without compromising performance [31]. Compared to single-stage systems, which often suffer from suboptimal compromises between the needs of different microbial

groups, two-phase configurations facilitate the coexistence of specialized communities under more stable conditions [31,32]. Although three-phase systems further refine process compartmentalization by isolating hydrolysis, acidogenesis, and methanogenesis, they often entail higher operational complexity, energy consumption, and maintenance requirements [33]. These drawbacks have limited their scalability, particularly in low-resource contexts. Consequently, two-phase systems strike a practical balance between performance enhancement and technical feasibility, making them a more accessible alternative for decentralized applications.

Among the strategies developed to improve AD performance, the integration of real-time monitoring systems has become increasingly relevant for enhancing process oversight and operational efficiency [34,35]. Basic and key variables such as pH, temperature, and methane concentration can be considered to infer the internal state of the reactor and anticipate potential imbalances. The use of cost-effective IoT platforms such as ESP32 microcontrollers coupled with sensor has proven suitable for real-time tracking, achieving deviations below 2% for CH₄ and 1.7% for pH when compared to laboratory-grade methods [34–36]. Systems incorporating the MQ-4 sensor (200–10,000 ppm CH₄) and platforms like ThingSpeak facilitate continuous data acquisition, cloud visualization, and automatic alerts, offering a practical solution to reduce manual intervention and increase system reliability [35,37,38].

In parallel, greater attention should be given to the microbial community (MC) involved in AD, as they are rarely considered in routine operation despite being responsible for driving the entire process [9]. Recent studies have highlighted that variations in microbial structure are strongly influenced by substrate type, operational parameters such as temperature and organic loading rates (OLR), and reactor configuration [39,40]. However, most operational strategies still rely exclusively on physicochemical parameters, overlooking microbial signals that often precede system imbalances [41]. Sequencing platforms like Illumina MiSeq and NextSeq have revealed both dominant and low-abundance taxa with key metabolic roles, including members of *Euryarchaeota* involved in methanogenesis and syntrophic bacteria mediating VFA conversion [40]. Understanding microbial shifts under stress conditions has provided valuable insights into process behavior and system dynamics, reinforcing the need to integrate microbial data into process understanding, particularly to elucidate how shifts in community structure and function impact methane levels [42,43].

Despite their central role in AD, MLR models have traditionally been developed using operational variables that capture external system conditions, parameters that are directly measurable or predefined during setup, while MC have often been treated as secondary inputs or excluded altogether. For example, recent studies have used MLR to predict specific methane production from dry AD of the organic fraction of municipal solid waste in pilot-scale plug-flow reactors. Six significant, mostly operational predictors were prioritized (VS, OLR, HRT, C/N ratio, lignin content, and VFA) via Pearson correlation and PCA. Simple regression showed low performance ($R^2=0.3$), while the full MLR reached $R^2=0.91$. A reduced model with four uncorrelated variables (VS, OLR, C/N ratio, lignin content) maintained strong accuracy ($R^2=0.87$) with fewer inputs. [44]. Similarly, MLR has been applied to predict VFA concentrations in AD of primary and secondary sludge using operational and physicochemical inputs. The model achieved R^2 values above 0.85 in several scenarios, offering high interpretability and low computational demand. Although less accurate than leading ensemble methods, MLR remains suitable for applications that require clear interpretation of variable influence [45].

Unlike models based solely on operational parameters, an MLR model using the relative abundances of archaeal and bacterial OTUs predicted methane production rates in 149 anaerobic digesters, explaining 66% of the variance with a standard error of 0.12 LCH₄/L-day. Through the integration of PCA, ANOVA, and cubic smoothing, it captured operational effects such as ammonia toxicity and achieved R^2 values of 0.93 for biogas generation and 0.95 for COD removal across laboratory, pilot, and industrial scales, underscoring its worth in translating complex MC data into scalable process predictions. [46].

Building on emerging evidence supporting the integration of microbial data into statistical modeling, this study evaluates the potential of MLR to predict methane concentrations in a low-cost, two-phase anaerobic digester treating PM and CD at laboratory scale. The work aligns with Sustainable Development Goal 7 by promoting accessible tools for energy recovery from organic waste.

This article is structured into four main sections. The Introduction outlines the context of AD in the Valle del Cauca region, highlighting environmental and operational challenges from agro-industrial organic waste, reviewing strategies to improve biogas systems, and emphasizing the need to integrate microbial data into predictive models. The Materials and methods detail the system setup, monitoring, sequencing, and the MLR approach used for variable selection and model construction. The Results and discussion section presents the modeling outcomes, identifies relevant predictors, and interprets their contribution to system behavior. The Conclusion summarizes the key findings and future perspectives for incorporating microbiota into data-driven frameworks for sustainable energy transitions.

2. Materials and Methods

This study aims to develop a predictive model for methane concentration based on a set of measurable variables, including VFAs, microbial populations, and operational parameters. This section first describes the dataset and the preprocessing steps undertaken. Subsequently, it details the initial linear modeling approach, followed by a feature selection process based on variable weighting to derive a simplified, yet robust, model. Finally, it presents the development of an adaptive predictive model using a moving window technique combined with a regularization method to prevent overfitting.

2.1. Substrate Selection

The substrates used in this study were fresh PM and CD. The inoculum, obtained from the same source as the manure, was included to ensure microbial compatibility with the feedstock. Both were collected at a small-scale pig farm located in the municipality of Florida, Valle del Cauca, where approximately 20 pigs are kept under semi-intensive conditions. Animal pens are washed twice daily, and the resulting wastewater, rich in organic matter, drains into a static open-air tank that served as the inoculum source. Fresh manure was manually collected after excretion using sanitized tools. CD were obtained from a medium-sized cassava starch-processing facility located in the rural area of Mandiba, Santander de Quilichao, Cauca. Processing nearly eight tons of cassava per day, the plant generates over two tons of lignocellulosic residue each week. This material was delivered in dry, milled form.

All samples were stored at 4 °C until physicochemical characterization, which included proximate analysis by gravimetric methods and determination of the carbon-to-nitrogen (C:N) ratio via high-temperature combustion. These procedures followed the Standard Methods for the Examination of Water and Wastewater (APHA, AWWA, WEF), ensuring analytical consistency as summarized in Table 1 [47].

Table 1. Substrate characerization.

Substrate	C(%)	N(%)	C:N	%Hum.	%TS	%VS	%VS/%TS	%FS
I	32.9	3.2	10.3	98.0%	2.2%	1.3%	60.8%	0.9%
CD	43.1	0.9	45.9	11.0%	89.0%	85.4%	96.0%	3.6%
PM	12.9	1.9	7.0	72.0%	28.0%	21.0%	75.0%	7.0%

I: Inoculum; CD: Cassava dregs; PM: Pig manure; C: Carbon content; N: nitrogen content.

2.2. Experimental Setup

The experimental setup consisted of a two-phase laboratory-scale anaerobic digester designed to operate without integrated control systems Figure 1. The system was constructed using 110 mm sanitary-grade PVC tubing due to its low cost, durability, and ease of assembly. Phase 1 (D1F1) (3 L) was expected to perform hydrolysis and acidogenesis, while phase 2 (D1F2) (4 L) supposedly supported acetogenesis and methanogenesis. Each chamber was operated at 80% of its total volume, 2.4 L in phase 1 and 3.2 L in phase 2, leaving the remaining headspace for biogas accumulation. To enable real-time monitoring, a low-cost IoT module was incorporated into the digester, integrating an Arduino UNO microcontroller with sensors for pH, temperature, and methane concentration. Data were transmitted through a mobile network to the ThingSpeak platform for remote visualization [38]. This setup allowed continuous monitoring without the need for sophisticated instrumentation.

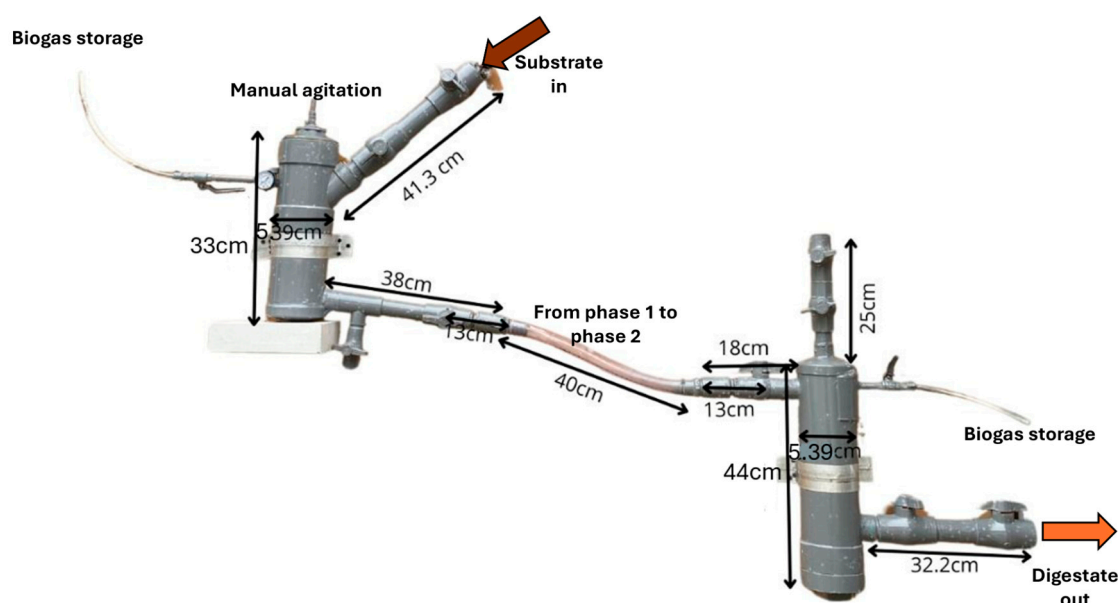


Figure 1. Two-phase digester made of PVC.

2.3. Operational Parameter

To establish an active MC, both phases were fed inoculum for five days, until its working volume. The inoculum had a C:N ratio of 10.3 and 2.2% TS. During start-up, the OLR, estimated with a five-day HRT, was 8.37 gVS/L-day. Thereafter, feeding used a 73:27 blend of PM and CD. The daily feed was 35 g fresh PM and 13 g CD, plus 166 g water to achieve 10% TS (214 g/day total). The theoretical C:N ratio was 21.55. With the defined working volumes, HRTs were 12 days for D1F1 and 15 days for D1F2. Corresponding OLRs were 7.7 and 5.7 gVS/L-day. VS inputs were 18.46 g/day (D1F1) and 18.45 g/day (D1F2). Daily manual feeding with graduated containers and isolation valves ensured accurate dosing and anaerobiosis.

The IoT-instrumented digester (D1) enabled incremental, data-driven feed adjustments in both phases (D1F1, D1F2) using real-time pH, temperature, and methane concentration. These signals guided when to lower the OLR and TS and when to apply temporary pH control, moving the reactors toward consistent operating conditions. Five feed formulations were implemented (Table 2). In D1F1, pH was briefly corrected with lime and then NaOH to keep it within 6.5–7.5; by mixture 5, recirculated digestate from D1F2 maintained pH without further chemicals. Mixture 4 used inoculum from an anaerobic digester at a university in Colombia treating food waste. Across mixtures, TS was reduced from 10% to 8–9%, OLR decreased from 12.4 gVS/L-day (inoculum step) to 5–6 gVS/L-day, and the C:N ratio increased in the final mixture due to recirculation while the contributions of PM and CD were reduced.

Table 2. Adjusted feeding regime.

D1F1															
Mix	OLR (gVS/L·day)	Mix load (g)	I (g)	PM (g)	CD (g)	H ₂ O added (g)	Daily load (g)	%TS	C:N	%I	%PM	%CD	HRT (day)	Period (day)	pH treatment
I	12.40	329	329			152	481	10%	10.3	100%	0%	0%	5	0-.5	
1	7.67	48		35	13	166	214	10%	21.6	0%	73%	27%	11	6-27	Lime
2	7.66	54		43	11	164	218	10%	18.3	0%	80%	20%	11	28-49	
3	6.60	48		39	9	162	210	9%	17.5	0%	81%	19%	11	50-89	NaOH
4	5.90	46	14	25	7	165	211	8%	15.7	30%	54%	15%	11	90-118	NaOH
5	5.87	71	35	26	10	136	207	8%	20.7	50%	36%	14%	11	119-161	
D1F2															
I	12.40	439	439			202	641	10%	10.3	100%	0%	0%	5	0-.5	
1	5.75	48		35	13	166	214	10%	21.6	0%	73%	27%	15	6-38	
2	5.74	54		43	11	164	218	10%	18.3	0%	80%	20%	15	39-60	
3	4.95	48		39	9	162	210	9%	17.5	0%	81%	19%	15	61-100	
4	4.41	46	14	25	7	165	211	8%	15.7	30%	54%	15%	15	101-127	
5	4.40	71	35	26	10	136	207	8%	20.7	50%	36%	14%	15	128-161	

I = inoculum; PM = pig manure; CD = cassava dreg; OLR = organic loading rate; TS = total solids; HRT = hydraulic retention time.

2.4. Steady-State

Identifying steady-state periods was essential to build a reliable dataset, define representative operating conditions, and guide downstream variable prioritization and modeling. pH, temperature, and methane concentration were monitored continuously for 161 days (24/7). The IoT system logged three readings per minute for each variable and was routinely cross-checked against bench measurements to validate operational reliability.

Data volume was substantial, D1F1 recorded 694110 samples per variable and D1F2 573215. Processing followed six steps: 1) splitting timestamp into date and time; 2) validity filtering (e.g., pH 3–12; 10–45 °C; CH₄ within instrument bounds) with out-of-range values set to blank; 3) multivariate imputation by chained equations (MICE) to preserve temporal continuity [48]; 4) resampling to hourly means (2893 rows in D1F1; 2389 in D1F2) and 5) to daily means (152 and 147, respectively), retaining trends while reducing computational load as shown in Table 3.

Table 3. Data processing.

D1F1					
Steps	pH	T(°C)	CH ₄	Total	%
All data	694110	694110	694110	2082330	100%
Day-hour	694110	694110	694110	2082330	100%
Filters	635953	635953	635953	1907859	92%
MICE	694110	694110	694110	2082330	100%
Data per hour	2893	2893	2893	8679	0.42%
Data per day	152	152	152	456	0.02%
D1F2					
All data	573215	573215	573215	1719645	100%
Day-hour	573215	573215	573215	1719645	100%
Filters	518897	518897	518897	1556691	91%
MICE	573215	573215	573215	1719645	100%
Data per hour	2389	2389	2389	7167	0.42%
Data per day	147	147	147	441	0.03%

Stable windows were then identified via rolling windows using relative standard deviation thresholds (<15%) around moving means for pH, temperature, and methane concentration, with a minimum continuous duration and compliance with predefined operating limits [49]. D1 showed extended steady windows, typically with pH 6.5–7.5, facilitated by high-frequency data and the ability to adjust operating conditions in real time.

2.5. VFA Quantification

Samples were collected every three days in 5 ml Eppendorf tubes and stored at -20°C until analysis. The final selection of samples for analysis was made considering the periods of system stabilization under IoT monitoring and budgetary constraints, prioritizing those most representative of the overall process behavior. Sampling was carried out during the active operation of the digester.

The quantification of VFAs was performed by gas chromatography, following the procedure described in section 5560D of the Standard Methods for the Examination of Water and Wastewater (APHA) [50], in the laboratory of the Department of Chemical Engineering and Analytical Chemistry at the University of Barcelona. Prior to chromatographic analysis, the samples were centrifuged and filtered through 0.45 µm nylon membranes to remove suspended solids. Each analysis vial contained 1 ml of sample, diluted or not depending on the estimated concentration level, along with 0.1 ml of 15% orthophosphoric acid containing a known concentration of 2-ethylbutyric acid (~500 mg/L) as an internal standard. This compound allowed verification of injection consistency and facilitated calibration of the equipment through the ratio of analyte to standard peak areas.

Analyses were carried out on a Shimadzu GC-2010 Plus gas chromatograph with a flame ionization detector, using a DB-FFAP capillary column (Agilent Technologies, 30 m × 0.25 mm × 0.25 µm). The oven temperature program started at 60°C with a two-minute hold, followed by an increase of 20°C/min up to 240°C, maintained for an additional two minutes. The total analysis time was 13 minutes.

The injector (SPL-1) operated at 220°C in split mode, with a split ratio of 50:1. Helium was used as the carrier gas at a pressure of 42.6 kPa, with a total flow of 233.4 ml/min, a column flow of 8.86 ml/min, and a linear velocity of 60 cm/s. The purge flow was set at 3 ml/min, and the makeup gas flow (nitrogen) at the detector was 10 ml/min. The injection volume was 2 ml, using helium, air, hydrogen, and nitrogen as auxiliary gases.

For equipment calibration, a commercial VFA standard (Supelco CRM46975) containing defined concentrations of acetic, propionic, isobutyric, butyric, isovaleric, valeric, isocaproic, caproic, hexanoic, and heptanoic acids was used. Serial dilutions were prepared in 1:1, 1:2, 1:4, 1:8, 1:16, and 1:32 ratios, to which orthophosphoric acid and the internal standard were also added. For alcohol analysis (ethanol, propanol, and butanol), defined-concentration standard solutions were prepared, applying the same dilutions and analytical conditions.

This procedure allowed precise and reproducible determination of VFAs in the samples, essential for evaluating the performance of the AD system and its relationship with operating conditions and microbiota.

2.6. Metagenomic Analysis

Samples for metagenomic analysis were collected directly from operational biodigester using 50 ml Falcon tubes. Sampling was performed every three days throughout the process, following the same prioritization criteria used for the quantification of VFAs, focusing on periods of greatest microbiological representativeness and considering the availability of resources. Once collected, samples were immediately frozen at -20°C and stored until further processing.

To analyse the MC, Falcon tubes were sent to Omega Bioservices (U.S.A) for DNA extraction using the kit E.Z.N.A.® Universal Pathogen Kit, library preparation and for sequencing the V3–V4 hypervariable region of the 16S rRNA gene using the primers 341F (CCTACGGGNGGCWGCAG) and 806R (GACTACHVGGGTATCTAATCC) which was conducted on an Illumina Miseq sequencing platform (Paired-end sequencing 300 bp). Illumina reads were then analysed using BaseSpace app (version 1.1.3) [51]. Thus, raw sequence data were demultiplexed and then quality filtered, denoised, merged, and chimera removed using the DADA2 [52] to generate amplicon sequence variants (ASVs). Taxonomic assignment was conducted using the SILVA database (version 138.2) [53].

To structure the analysis of microbial interactions, a subset of phyla of interest was defined from the general metagenomic dataset, considering the sequencing reads obtained for each taxonomic group. The selection was based on two main criteria. First, the sustained presence of each phylum throughout the monitoring period was evaluated, excluding those with very low or intermittent representation, as their variability would hinder the detection of consistent associations in the relational analysis. Second, functional relevance reported in previous studies on anaerobic digestion was reviewed, prioritizing phyla whose involvement in fermentative, acetogenic, or methanogenic pathways has been extensively documented in similar systems [9,54].

Once the representative periods were defined, the results from VFA quantification and metagenomic analysis were integrated, extending the characterization to the biochemical and microbiological components of the system. In several cases, the observed patterns were consistent with those reported in the specialized literature, which supported the robustness of the approach. The dataset included operational, biochemical, and microbiological variables [55–59].

Since the biochemical and microbiological measurements were less frequent than the operational records, imputation techniques were applied within the selected periods to expand the dataset

without distorting the relationships among variables. Methods such as KNN imputation, iterative imputation, and MICE were employed [56,60].

The analysis focused on the period between days 97 and 154, which, although not representing a fully stabilized phase, shows a trend toward stabilization and coincides with the selected VFA and microbiological samples. This ensured consistency between the experimental data and the operational conditions.

2.7. Preprocessing and Unified Database

Once the representative periods were defined, the results from VFA quantification and metagenomic analysis were incorporated to extend the characterization of the system to its biochemical and microbiological dimensions. The patterns obtained aligned with those reported in specialized literature, reinforcing the validity of the approach [55–59]. The unified dataset combined operational, biochemical, and microbiological variables.

Because biochemical and microbiological measurements were less frequent than operational records, imputation method MICE was applied to harmonize the dataset without altering the underlying relationships among variables [56,60].

The analysis focused on the period between days 97 and 154, which, while not fully stabilized, displayed a clear trend toward steady performance and coincided with the VFA and microbiological samples selected. This ensured coherence between experimental observations and operational conditions. The resulting dataset comprised daily averages over 58 days, which were further refined through linear interpolation to increase temporal resolution. This process expanded the series to 1000 points, enabling the application of moving window analyses, as illustrated in Figure 2.

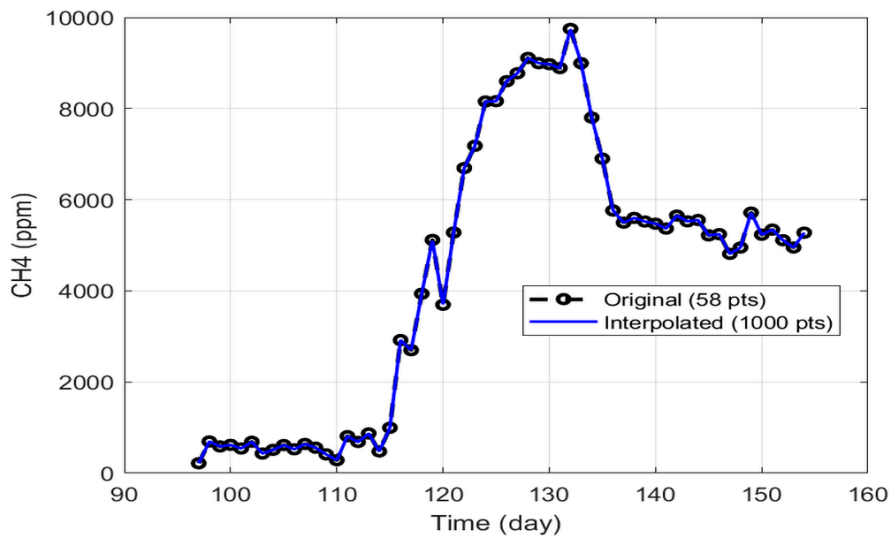


Figure 2. Original (58 points) vs. interpolated (1000 points) time-series data for CH₄ concentration.

The interpolation was validated for all variables, yielding R² values close to 1 and MRE values around 0.1%, confirming a high-fidelity representation of the original data.

2.8. Linear Modeling

To simplify the proposed equations and procedure, the suffixes associated with each fatty acid (Table 4) microorganism (Table 5), and operating condition (Table 6) are shown below.

Table 4. List of associated suffixes for fatty acids (a).

Acetic	1	Caproic	5
Propionic	2	Heptanoic	6
Butyric	3	Ethanol	7
Valeric	4	Propanol	8

Table 5. List of associated suffixes for microorganisms (m).

Firmicutes	9	Tenericutes	19
Bacteroidetes	10	Armatimonadetes	20
Actinobacteria	11	Cyanobacteria Chloroplast	21
Proteobacteria	12	Acidobacteria	22
Planctomycetes	13	Lentisphaerae	23
Synergistetes	14	BRC1	24
Spirochaetes	15	Candidatus Saccharibacteria	25
Euryarchaeota	16	Parcubacteria	26
Verrucomicrobia	17	Chloroflexi	27
Cloacimonetes	18		

Table 6. List of associated suffixes for operating conditions (p).

phi	28	pho	30
Ti	29	To	31

The equation (1) that linearly approximates CH₄ concentration as a function of the microorganisms, fatty acids, and operating conditions was proposed in the following linear form, based on the suffixes from Tables 4–6.

$$(CH_4)_{approx} = C_1(a_1) + C_2(a_2) + \dots + C_{25}(m_{25}) + C_{26}(m_{26}) + \dots + C_{30}(p_{30}) + C_{31}(p_{31})$$

(1)

In its matrix form (matrix A), equation (1) can be expressed as follows (equation (2)):

$$\begin{bmatrix} (CH_4)_1 \\ (CH_4)_2 \\ \vdots \\ (CH_4)_n \end{bmatrix} = \begin{bmatrix} (a_1)_1 & (a_2)_1 & \dots & (m_{26})_1 & \dots & (p_{31})_1 \\ (a_1)_2 & (a_2)_2 & \dots & (m_{26})_2 & \dots & (p_{31})_2 \\ \vdots & \vdots & & \vdots & & \vdots \\ (a_1)_n & (a_2)_n & \dots & (m_{26})_n & \dots & (p_{31})_n \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_{31} \end{bmatrix}$$

(2)

where the constants C_i are the approximation coefficients. This matrix form from equation (2) can be written more compactly as shown in equation (3):

$$\underline{CH_4} = \underline{A} \underline{x}$$

(3)

The matrix A contains data collected from fatty acids, microorganisms, and operating conditions, the vector CH₄ represents the collected methane production data, while the vector x contains the approximation coefficients that must be determined to formulate the model. The vector x, can be solved by rearranging equation (3) as follows:

$$\underline{x} = [\underline{A}^T \underline{A}]^{-1} \underline{A}^T \underline{CH_4}$$

(4)

where A^T is the transpose of matrix A.

2.8.1. Assessing Variable Importance

To determine the relative importance of each variable in the approximation, and subsequently define a smaller, more practical subset (as working with all 31 variables can be impractical and costly in terms of laboratory testing), a variable weighting method was used. Therefore, equation (5) appears as a modification of the equation (3) considering the minimal error ϵ .

$$\underline{CH_4} = \underline{A} \underline{x} + \epsilon$$

(5)

To quantify how much each variable "contributes" to the CH_4 production within the approximation, it is necessary to measure the relevance of each variable in the linear model. Since each variable may be measured on a different scale (e.g., microorganism abundance vs. fatty acid concentration in mg/L), directly comparing the raw coefficients C_i in vector \underline{x} can be misleading. Therefore, it is necessary to standardize the input data. In the same way, to compare the relative importance of each variable, the coefficients C_i that form the vector \underline{x} were standardized (as z-scores). The standardized coefficient C_i^* for each variable C_i was calculated as:

$$C_i^* = C_i \frac{\sigma_{C_i}}{\sigma_{CH_4}} \quad (6)$$

where σ_{C_i} and σ_{CH_4} are the standard deviations of the approximation coefficient C_i and the response variable CH_4 , respectively.

2.8.2. Proposing a Predictive Model

To capture the evolutionary nature of the anaerobic digestion process, a dynamic predictive model was developed based on a moving windows approach. The model operates iteratively. At each time step t , a linear regression model is trained using a window containing the last k observations (in this case, $k = 10$ was chosen). This model is then used to make a one-step-ahead prediction of CH_4 (denoted as $\widehat{CH_4}$), as a function of the weighted variables previously described.

However, the use of small data windows can lead to overfitting. To address this problem and improve the model's generalization capability, Ridge Regression was used instead of ordinary least squares. This regression introduces a penalty term into the least squares cost function. For each time window, the objective is to find the coefficient vector \underline{x} that minimizes the following function:

$$\min_{\underline{x}} \left(\left\| \underline{CH_4} - \underline{A} \underline{x} \right\|_2^2 + \lambda \left\| \underline{x} \right\|_2^2 \right) \quad (7)$$

where $\left\| \underline{CH_4} - \underline{A} \underline{x} \right\|_2^2$ is the sum of squared errors (the data fit term at time t), $\lambda \left\| \underline{x} \right\|_2^2$ is the regularization term applied at time t , λ is the regularization hyperparameter that controls the balance between the data fit and the model simplicity, and \underline{x} contains the coefficients C_i . The hyperparameter λ was selected to improve the model predictive performance ($\lambda = 2$). Thus, equation (4) is rewritten to obtain the predictive parameters ($\hat{\underline{x}}$) by solving the following equation:

$$(\hat{\underline{x}})_{t+1} = \left([\underline{A}^T \underline{A} + \lambda \underline{I}]^{-1} \underline{A}^T \underline{CH_4} \right)_t \quad (8)$$

where \underline{I} is the identity matrix. The goal of equation (8) is to find the value of the coefficients in vector \underline{x} at time $t + 1$ using the data available at time t . Using equation (8), it is possible to find the $\widehat{CH_4}$ values for a subsequent window, given a defined window size of $k = 10$. In this way, equation (3) becomes a prediction equation as follows:

$$(\widehat{CH_4})_{t+1} = (\underline{A})_{t+1} (\hat{\underline{x}})_{t+1} \quad (9)$$

2.8.3. Model Performance Evaluation

The precision of the predictive model was quantified using three standard statistical metrics. These metrics evaluate the divergence between the real observed values of CH_4 and the values predicted by the model, $\widehat{CH_4}$. On one hand, the Coefficient of Determination (R^2) indicates the proportion of the variance in methane production that is predictable from the independent variables. A value close to 1 indicates an almost perfect fit. Equation (10) shows how it was calculated for this case.

$$R^2 = 1 - \frac{\sum_{i=1}^n \left[(CH_4)_i - (\widehat{CH_4})_i \right]^2}{\sum_{i=1}^n \left[((CH_4)_i - \overline{CH_4}) \right]^2} \quad (10)$$

Next, the Root Mean Square Error (RMSE) represents the standard deviation of the prediction residuals. It is a measure of the average error of the model in the same units as the response variable (ppm of CH_4), which facilitates its interpretation and is expressed in equation (11).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [(CH_4)_i - (\widehat{CH_4})_i]^2} \quad (11)$$

Finally, the Mean Relative Error (MRE) measures the average error in relative or percentage terms with respect to the real value is defined in equation (12). The absolute value was used to prevent positive and negative errors from canceling each other out.

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{(CH_4)_i - (\widehat{CH_4})_i}{(CH_4)_i} \right| \quad (12)$$

In equations (10), (11), and (12), $(CH_4)_i$ is the real value of the i -th observation, $(\widehat{CH_4})_i$ is the value predicted by the model for the i -th observation, $\overline{CH_4}$ is the mean value of all real values, and n is the total number of observations used for the evaluation.

3. Results and Discussion

3.1. Digester Performance

During the implementation of the two-phase biodigester, one of the main technical challenges was controlling gas leaks and internal pressure, which required multiple structural adjustments and caused delays in the early stages of operation. The installed manometers failed, likely due to H_2S -induced corrosion, while the valves progressively stiffened with use, in some cases requiring replacement. In addition, certain PVC weld joints developed fractures, compromising system integrity. The manual stirring mechanism did not improve process performance but was associated with gas leaks, leading to its deactivation and reinforcement of seals. After several corrective interventions, continuous and functional operation was achieved for the duration of the experiment.

3.1.1. IoT Monitoring Advantages

Although low-cost systems might be unsuited for long-term use or high-precision data collection, these biodigesters provide a practical alternative for experimental applications at laboratory scale when resources are limited. The estimated cost of assembling a two-phase biodigester without IoT monitoring was 300 USD, while the addition of a digital monitoring system increased the total to 420 USD per unit. Although some sensors required replacement during operation, the IoT system performed reliably, providing consistent readings comparable to manual instruments. Its implementation enabled real-time, continuous data acquisition, which was essential for detecting operational variations, making timely adjustments, and improving process understanding.

Figure 3 (D1F1) and Figure 4 (D1F2) show the 24-hour profiles of pH, temperature, and CH_4 concentration during different operational stages. The days were randomly selected from both phases to provide representative snapshots of system behaviour under varying conditions. In all cases a consistent inverse relationship was observed, where pH increased during early morning hours as ambient temperature declined and then decreased progressively as temperature rose throughout the day. In D1F1 this trend was stronger and more reproducible, with correlation coefficients between -0.84 and -0.94, while in D1F2 the association was weaker (between -0.49 and -0.66) and accompanied by larger fluctuations in methane concentrations. These results highlight the direct effect of ambient thermal oscillations on microbial activity, especially on pH dynamics.

This behaviour may be linked to phases of microbial adaptation or to the accumulation of internal self-regulation mechanisms. As temperature dropped during the night and metabolic activity slowed, nitrogenous compounds likely continued decomposing and releasing ammonia (NH_3). This ammonia could react with dissolved CO_2 , which is more soluble at low temperatures, to form ammonium bicarbonate (NH_4HCO_3). The resulting increase in alkalinity buffered pH variations, preventing excessive acidification and contributing to system resilience [17,61].

Such dynamics are rarely captured in conventional laboratory-scale digesters where data are typically restricted to discrete measurements. In this case the use of IoT-based continuous monitoring provided hourly resolution and made it possible to identify fine-scale responses such as the rise in pH at lower nighttime temperatures that would otherwise remain unnoticed. This approach delivers a more realistic picture of system performance under environmental conditions and emphasizes the value of continuous monitoring strategies for interpreting anaerobic digestion behaviour beyond the limits of punctual sampling.

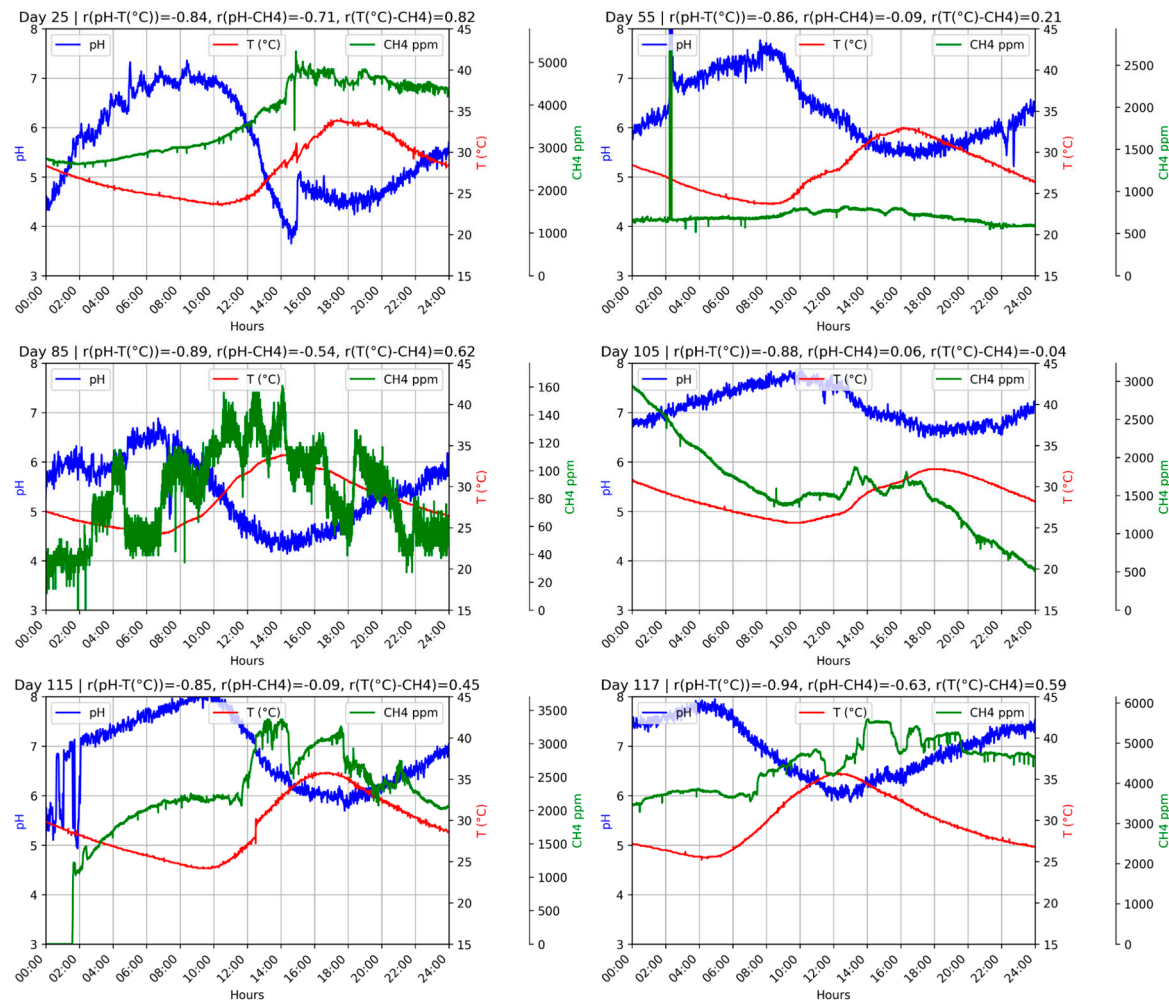


Figure 3. D1F1 operation during 24 hours on random days.

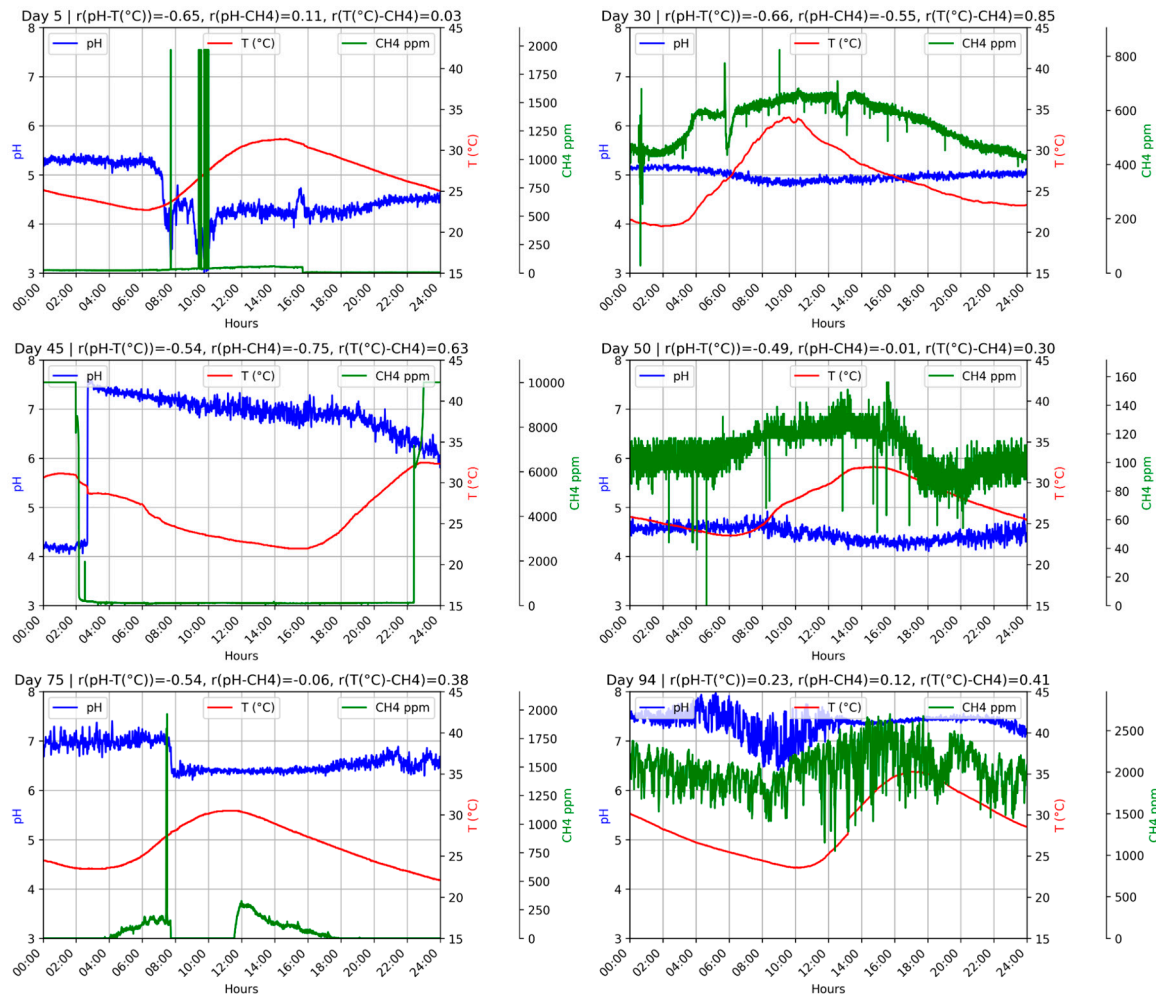


Figure 4. D1F2 operation during 24 hours on random days.

3.1.2. Stabilization of Anaerobic Codigestion

Achieving a steady-state is a critical milestone in AD, as it reflects the convergence of operational and microbial conditions that support sustained methanogenic activity [62,63]. In system D1, specific time segments were identified where pH, temperature, and methane concentration aligned within the functional ranges expected for AD. Figure 5 integrates these three variables across the full operational period, providing a comprehensive view of the transitions from unstable to stabilized phases. This visualization not only highlights the progression of the system under different operating conditions but also illustrates how corrective measures and phase-specific dynamics gradually steered the reactor toward a functional equilibrium.

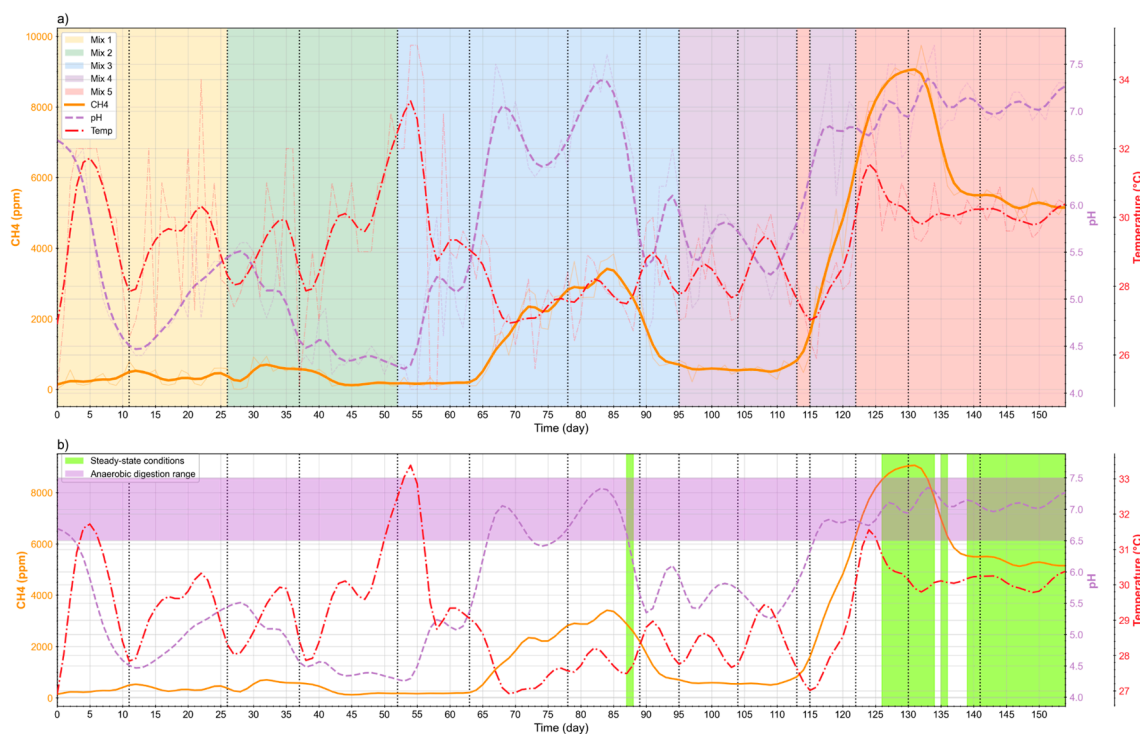


Figure 5. a) Mixture feeding over time b) Steady-state identification.

In Figure 5a, a shift becomes evident after day 120, when pH consistently remained above 6.5 while methane rose steadily, surpassing 8000 ppm by day 132. These conditions coincided with stable temperatures between 30 and 32 °C, an optimal mesophilic range that favours methanogenic activity [49,55]. The segmentation into D1F1 and D1F2 reveals the influence of each phase under a combined HRT of 27 days. During early mixtures (Mix1-Mix2), high OLR and the absence of adapted inoculum produced irregular methane signals dominated by acidogenesis. From Mix3 onwards, corrective measures such as alkalization promoted higher CH₄ concentrations, although fluctuations beyond $\pm 15\%$ prevented these periods from being classified as steady. Toward Mix4 and Mix5, adjustments including higher inoculum input and recirculation from D1F2 likely increased microbial density and functional diversity, progressively creating conditions more favourable to methanogenesis [64,65].

Figure 5b highlights two segments where the system approached steady-state behaviour. The first, between days 126 and 136, was characterized by pH values between 6.5 and 7.5, methane above 8000 ppm, and stable temperatures (30-32 °C), all within functional ranges and with fluctuations below $\pm 15\%$. A temporary decline in methane around day 137 disrupted this stage, but from day 139 onwards the system recovered, initiating a second steady segment that persisted until the end of the experiment.

Overall, the convergence of pH, CH₄, and temperature, demonstrates that the steady-state achieved in D1 was not the result of a single correction but the cumulative effect of progressive adjustments. This sequence of changes allowed the system to transition from acidogenic predominance to a consolidated methanogenic phase, representing a functional stabilization consistent with the goals of two-phase anaerobic digestion [65].

3.2. Volatile Fatty Acids (VFAs) and Metagenomic Analysis

Once stabilization was established from IoT-monitored variables, VFAs and microbiota were analyzed during the transition toward optimal operation (days 97-154, Mix 4 and Mix 5). Thirteen samples were taken, and missing data were inputted through MICE to ensure continuity.

Figure 6 shows that cumulative VFA concentrations dropped sharply after day 100, from above 14000 mg/L to 5800 mg/L, before oscillating between 5000 and 7000 mg/L. This decline reflects the mitigation of acidogenic pressure and the progressive adjustment of the microbial community,

setting conditions increasingly suitable for methanogenesis and linking metabolite dynamics with microbial responses in the path toward functional balance (Figure 8) [66,67].

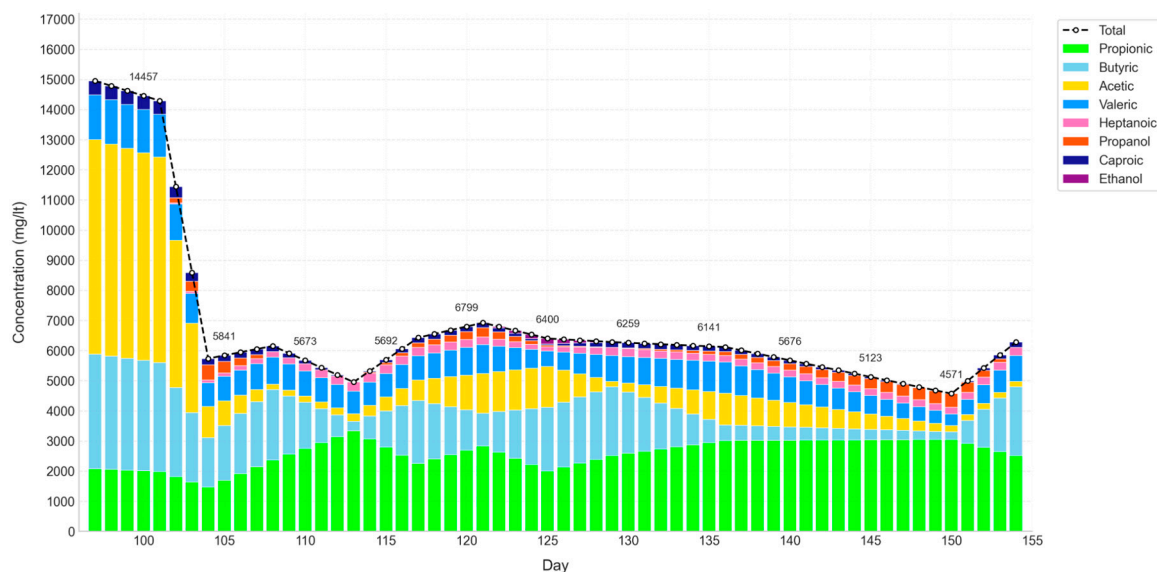


Figure 6. VFA and alcohol concentration in D1.

The individual analysis of VFAs confirmed that the steep decline after day 100 was largely driven by the reduction of acetic and butyric acids, both tied to early fermentative pathways [66]. Between days 103 and 118, however, propionic acid and medium-chain carboxylates (C5-C8), including caproic, heptanoic, and valeric, increased notably, reaching averages of 140 mg/L, 218 mg/L, and 820 mg/L, respectively [68]. These less common metabolites are typically linked to secondary fermentation processes or to transitional phases of temporary accumulation [69–71]. Their persistence, together with measurable levels of propanol (134 mg/L) and the absence of ethanol, suggests a fermentative stage dominated by chain-elongation routes, potentially hindered by propanol's inhibitory effect on methanogenic consortia [69,72,73]. After day 119, these acids gradually declined (e.g., caproic down to 109 mg/L, valeric to 758 mg/L), while ethanol reappeared (20 mg/L) and propanol rose to 191 mg/L. This pattern may indicate that, despite higher alcohol concentrations, microbes capable of degrading medium-chain acids regained activity, backing a functional shift toward methanogenesis [74,75].

Regarding microbial analysis, a total of 1815465 high-quality reads, with an average of 201718 ± 103945 reads per sample. Rarefaction analysis based on Shannon index showed that sequencing depth was adequate to capture most of the bacterial diversity across samples as shown in Figure 7.

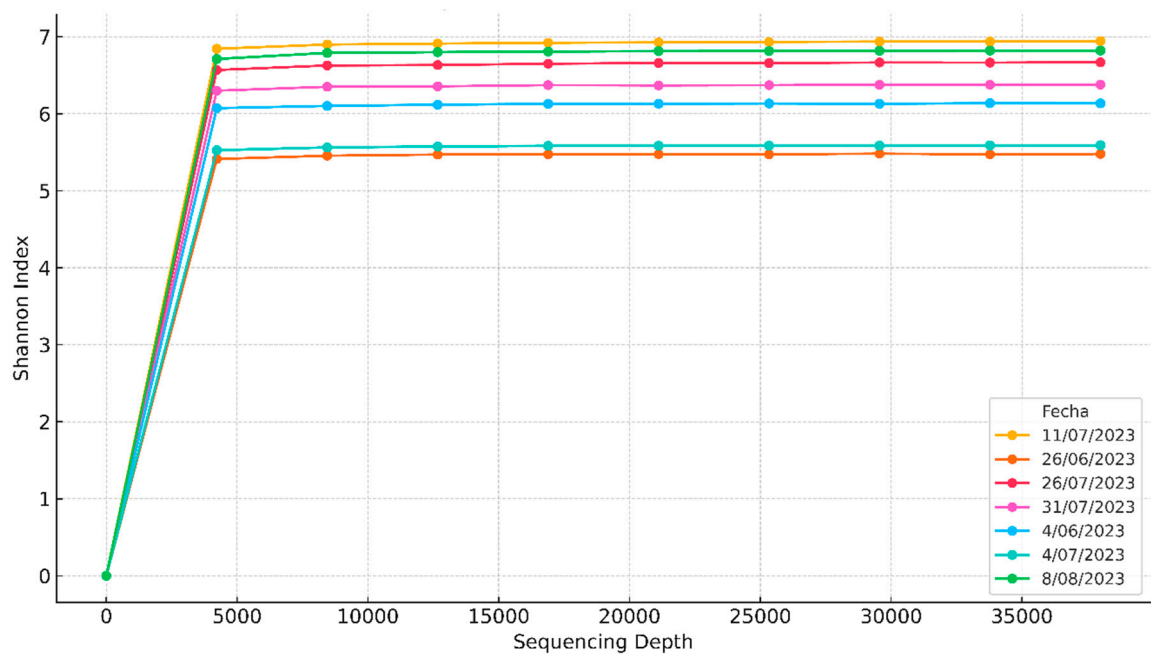


Figure 7. Alpha rarefaction curves of the alpha diversity (Shannon) of the 16S rRNA gene.

The microbial dynamics based on the phylum level aggregate counts (Supplementary Table S1) reflected in Figure 8 provide insights into the links between dominant phyla and the VFA and methane concentration profiles presented in Figures 6 and 5. During days 97 and 154, *Firmicutes* remained the prevailing group, averaging 49235 reads (64.2% of the total), underscoring its central role in the early stages of the process, particularly in hydrolysis and acidogenesis [9]. This activity likely promoted the production of fermentative precursors, consistent with the elevated provide insights into the links between dominant acids recorded at the beginning of this interval [66].

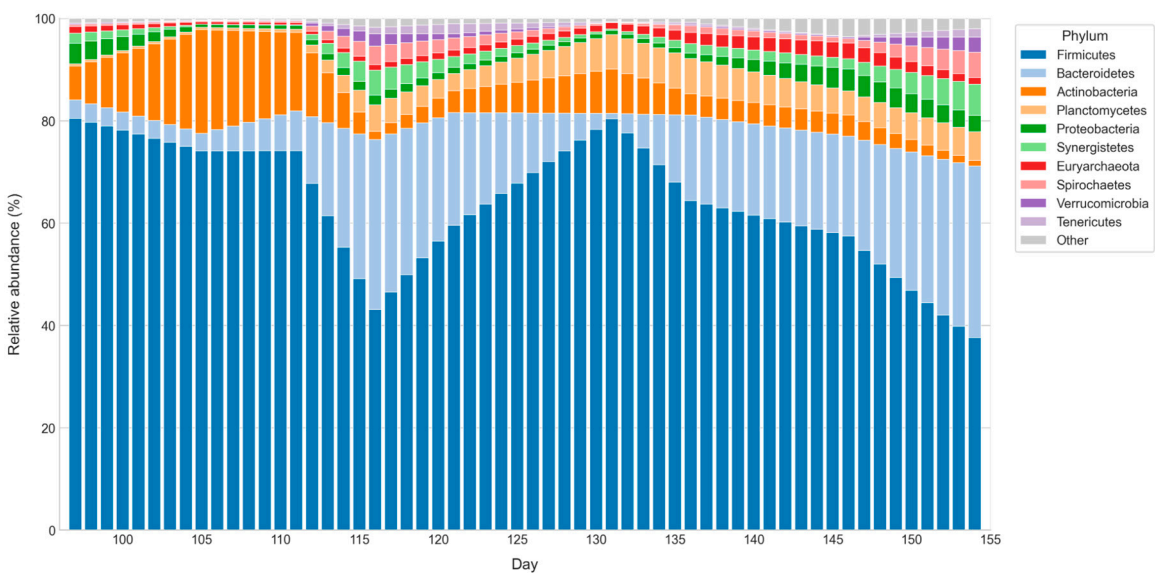


Figure 8. Temporal dynamics of microbial phyla in D1.

Along with *Firmicutes*, *Bacteroidetes* (15%) and *Actinobacteria* (7.5%) contributed to medium-chain fatty acids such as valeric and caproic during the accumulation phase (days 103-118) [69,73]. This functional diversity points to a bacterial consortium engaged in degrading complex polymers and extending fermentative pathways, buffering intermediates before methanogenic activity resumed

[76,77]. Toward the end, *Firmicutes* declined while *Euryarchaeota* increased to 1.5%, coinciding with reduced VFAs and steadier methane, suggesting activation of acetoclastic and hydrogenotrophic routes [9,78].

Minor groups might have played complementary roles, with *Planctomycetes* (3.9%) likely coupling sulfide oxidation to methanogenesis, *Proteobacteria* (1.9%) contributing to propionate and acetate turnover, and *Synergistetes* (1.9%) participating in syntrophic H₂ transfer [9,79].

Unexpectedly, *Verrucomicrobia* appeared in the community profile, a phylum typically restricted to volcanic habitats dominated by acidophilic methanotrophs [80]. These bacteria can oxidize methane as their main substrate and, to a lesser extent, hydrogen, carbon dioxide, ammonium and hydrogen sulfide, functioning as natural biofilters in extreme ecosystems. Their presence in a mesophilic anaerobic digester is unusual and may reflect residual inoculum or localized microredox niches rather than an active role in methanogenesis [80,81]. Alongside other low-abundance phyla such as *Lentisphaerae*, *Candidatus saccharibacteria*, and *Parcubacteria*, their detection expands the taxonomic spectrum and raises questions about potential ecological roles still unexplored in anaerobic bioenergy systems [82]. Many of these groups remain unresolved at the species level, even after advanced genomic assembly, forming part of the so called microbial dark matter. This hidden fraction highlights one of the major challenges in deciphering the functional complexity of anaerobic microbiomes [83,84].

3.3. Multiple Linear Regression (MLR)

Modelling phase was based on the Supplementary Table S2. The values in Table 7 show the coefficients C_i obtained by finding the vector \underline{x} after applying equation (4).

Table 7. Approximation coefficients.

C_1	-0.25	C_{16}	84.58
C_2	0.72	C_{17}	152.94
C_3	4.86	C_{18}	-154.01
C_4	18.20	C_{19}	-112.07
C_5	-61.33	C_{20}	165.65
C_6	-22.60	C_{21}	60.94
C_7	-12.08	C_{22}	15.47
C_8	14.33	C_{23}	-302.56
C_9	7.39	C_{24}	159.10
C_{10}	-18.99	C_{25}	-450.57
C_{11}	-21.21	C_{26}	81.69
C_{12}	-161.69	C_{27}	116.88
C_{13}	-58.79	C_{28}	-1562.95
C_{14}	67.72	C_{29}	6.72
C_{15}	123.17	C_{30}	68.50
		C_{31}	242.54

This method of finding the approximation coefficients is known by some authors as inverse modeling and can be considered a linear regression. When the vector \underline{x} was obtained, equation (1) was applied to generate the approximation curve. Figure 9 shows the resulting approximation.

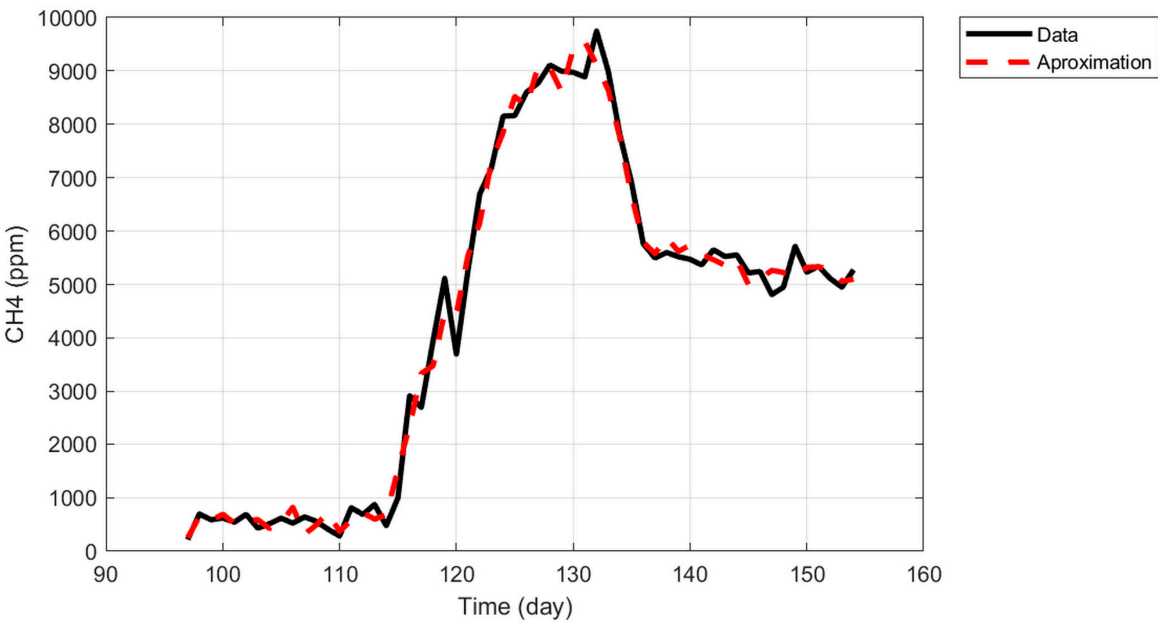


Figure 9. Approximation result.

3.3.1. Data Prioritization

While Figure 9 shows the overall approximation result, it is important to determine the contribution of each fatty acid, microorganism, or operating condition to the CH_4 production. As previously mentioned, a direct comparison of the C_i coefficients can be misleading, so it is important to perform a variable weighting process using equation (6). The values obtained from this process are shown in Table 8 and plotted in Figure 10.

Table 8. Weighting parameters C_i^* .

C_{10}^*	66.48	C_5^*	2.21
C_{12}^*	50.51	C_{27}^*	1.79
C_{17}^*	38.48	C_3^*	1.53
C_{13}^*	36.87	C_4^*	1.43
C_{15}^*	36.06	C_{26}^*	0.88
C_{11}^*	34.35	C_8^*	0.67
C_{20}^*	30.92	C_6^*	0.57
C_{14}^*	28.37	C_{22}^*	0.37
C_9^*	15.61	C_{21}^*	0.34
C_{16}^*	14.37	C_1^*	0.17
C_{19}^*	12.25	C_{28}^*	0.15
C_{18}^*	7.87	C_2^*	0.13
C_{23}^*	7.63	C_{31}^*	0.11
C_{25}^*	4.04	C_7^*	0.11
C_{24}^*	2.30	C_{30}^*	0.01
		C_{29}^*	0.01

To define how many variables are needed to achieve a good fit without a significant loss of precision, two limits were established, as can be seen in Figure 11 : an R^2 Coefficient greater than 0.9 and an MRE lower than 15%. Based on Table 8 and Figure 10, the 12 variables with the highest values, or greatest impact on CH_4 production, were selected. There are C_{10}^* , C_{12}^* , C_{17}^* , C_{13}^* , C_{15}^* , C_{11}^* , C_{20}^* , C_{14}^* , C_9^* , C_{16}^* , C_{19}^* and C_{18}^* . According to Table 5, these correspond to *Bacteroidetes*, *Proteobacteria*,

Verrucomicrobia, *Planctomycetes*, *Spirochaetes*, *Actinobacteria*, *Armatimonadetes*, *Synergistetes*, *Firmicutes*, *Euryarchaeota*, *Tenericutes*, and *Cloacimonetes*, respectively.

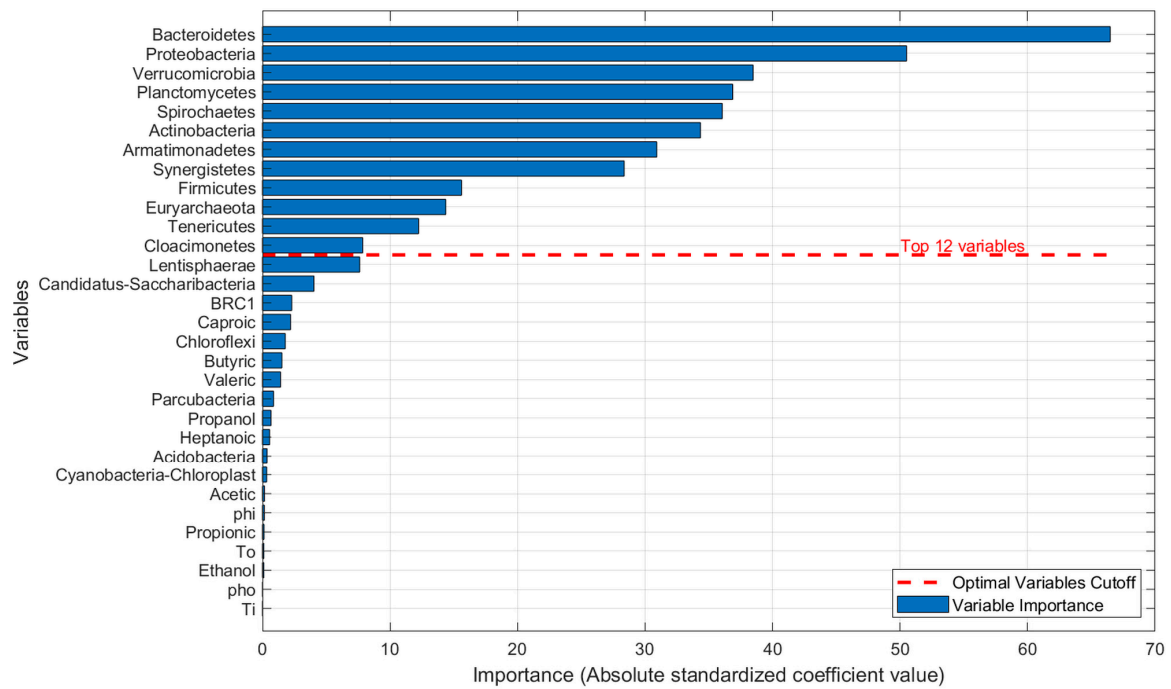


Figure 10. Relative variable importance with defined cut-off.

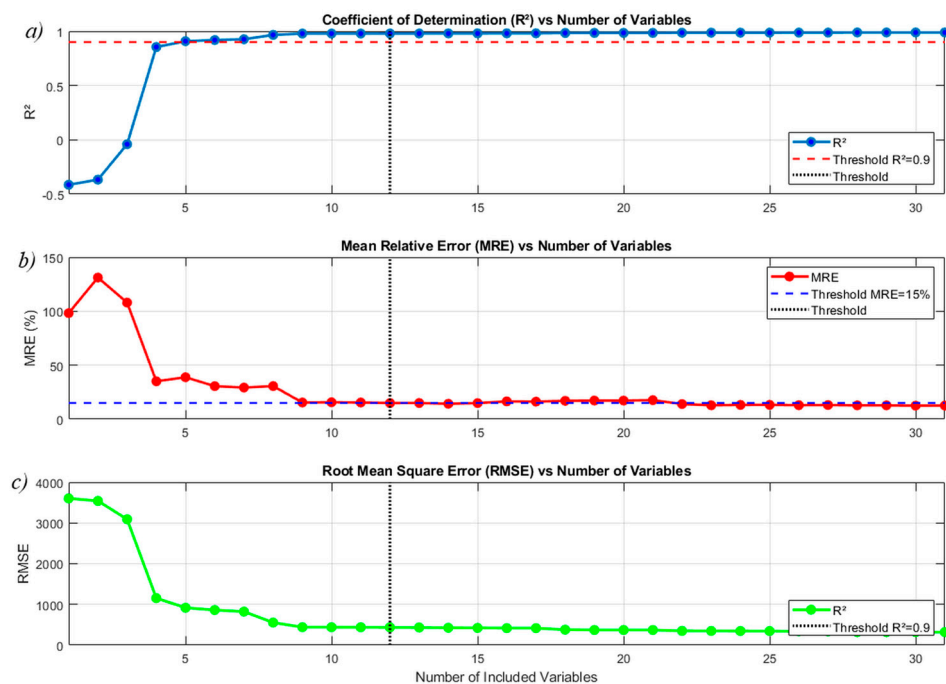


Figure 11. a) R^2 , b) MRE and c) RMSE vs Number of included variables in the approximation.

Thus, equations (1) and (2) can be rewritten in terms of the 12 variables that were found to be most important according to the weighting performed. This leads to equation (13):

$$(CH_4)_{WeightedApprox} = C_{10}(m_{10}) + C_{12}(m_{12}) + C_{17}(m_{17}) + C_{13}(m_{13}) + C_{15}(m_{15}) + C_{11}(m_{11}) + C_{20}(m_{20}) + C_{14}(m_{14}) + C_9(m_9) + C_{16}(m_{16}) + C_{19}(m_{19}) + C_{18}(m_{18}) \quad (13)$$

In its matrix form, equation (13) can be expressed as follows:

$$\begin{bmatrix} (CH_4)_1 \\ (CH_4)_2 \\ \vdots \\ (CH_4)_n \end{bmatrix} = \begin{bmatrix} (m_{10})_1 & (m_{12})_1 & \dots & (m_{18})_1 \\ (m_{10})_2 & (m_{12})_2 & \dots & (m_{18})_2 \\ \vdots & \vdots & \ddots & \vdots \\ (m_{10})_n & (m_{12})_n & \dots & (m_{18})_n \end{bmatrix} \begin{bmatrix} C_{10} \\ C_{12} \\ \vdots \\ C_{18} \end{bmatrix} \tag{14}$$

Table 9 shows the new values found for the approximation coefficients, calculated using equation (13) in conjunction with equation (4). These results in the weighted approximation shown in Figure 12.

Table 9. New weighted approximation coefficients.

C_{10}	-4.64	C_{20}	63.94
C_{12}	-42.40	C_{14}	16.20
C_{17}	-94.76	C_9	1.70
C_{13}	-7.66	C_{16}	-7.10
C_{15}	64.88	C_{19}	-22.13
C_{11}	-4.02	C_{18}	251.39

Table 10 shows a comparison of the metrics used to compare the fits discussed previously. While the general behavior is replicated by both curves, the R^2 coefficient for the approximation curve using equation (1) is 0.989, whereas with the weighted approximation from equation (13), the R^2 value is 0.979.

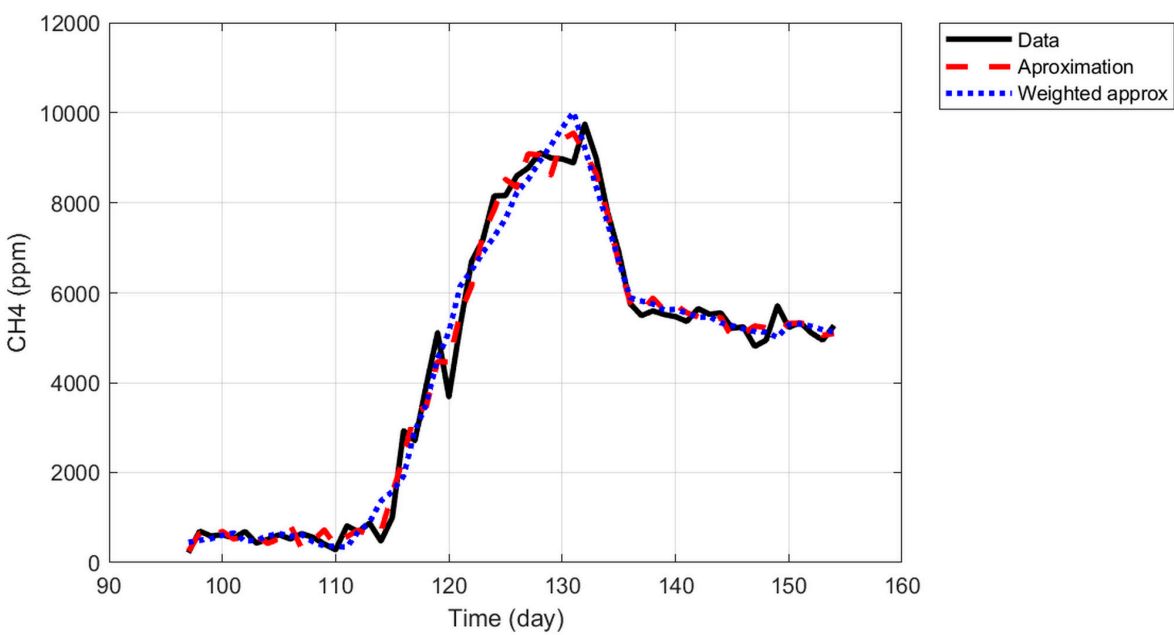


Figure 12. Weighted approximation.

Meanwhile, the mean relative error (MRE) for the approximation with equation (1) is around 12.59%, while for the weighted approximation with equation (13), it is 14.94%. Finally, the errors evaluated by the RMSE are below 450 ppm, which, considering the scale of Figure 5, are within an acceptable range. This approach successfully developed a simplified, dynamic model for predicting methane (CH_4) production in an anaerobic digestion process. The key achievement was the ability to reduce a complex system of 31 variables to a robust predictive model based on only the 12 most influential factors, without a significant loss of precision.

Table 10. Precision metrics for approximations and real data.

	R ²	MRE [%]	RMSE [ppm]
All variables	0.989	12.59	319.94
Weighted aproximation	0.979	14.94	435.82

The primary finding of this work is the overwhelming importance of microbial populations as indicators of CH_4 concentration compared to VFAs and operational parameters. The variable weighting analysis revealed that the 12 most significant variables were exclusively microorganisms, with groups like Bacteroidetes and Proteobacteria showing the highest importance scores. This suggests that, within the context of this study, the state of the microbial community is a more direct and powerful predictor of methanogenic activity than the concentration of intermediate substrates (VFAs) or the operational conditions measured. While VFAs are essential for methanogenesis, their concentrations can be transient. In contrast, the abundance of specific microbial groups likely represents the metabolic potential of the system, making them more robust indicators for modeling purposes.

3.3.2. Predictive Model

The simplification of the model from 31 variables to 12 demonstrates the practical value of the feature selection process. The weighted approximation model, using only the selected microorganisms, achieved an R^2 of 0.979, a negligible decrease from the 0.989 R^2 of the full model. By focusing only on the most critical microbial indicators, laboratory testing and data analysis efforts can be substantially reduced while still maintaining a high degree of accuracy. The slight increase in MRE and RMSE is an acceptable trade-off for the considerable reduction in model complexity.

In this context, using equation (13), it was possible to develop a predictive model to predict the components of the vector \hat{x} and, subsequently, the behavior of the $\widehat{CH_4}$ production by applying equations (8) and (9), respectively. This was done by using 70% of the dataset to train the model and the remaining 30% to test its performance. Thus, Figure 13 shows how, by applying equation (8), it is possible to obtain a prediction for the behavior of each of the coefficients for the weighted variables \hat{x} (originally listed in Table 9).

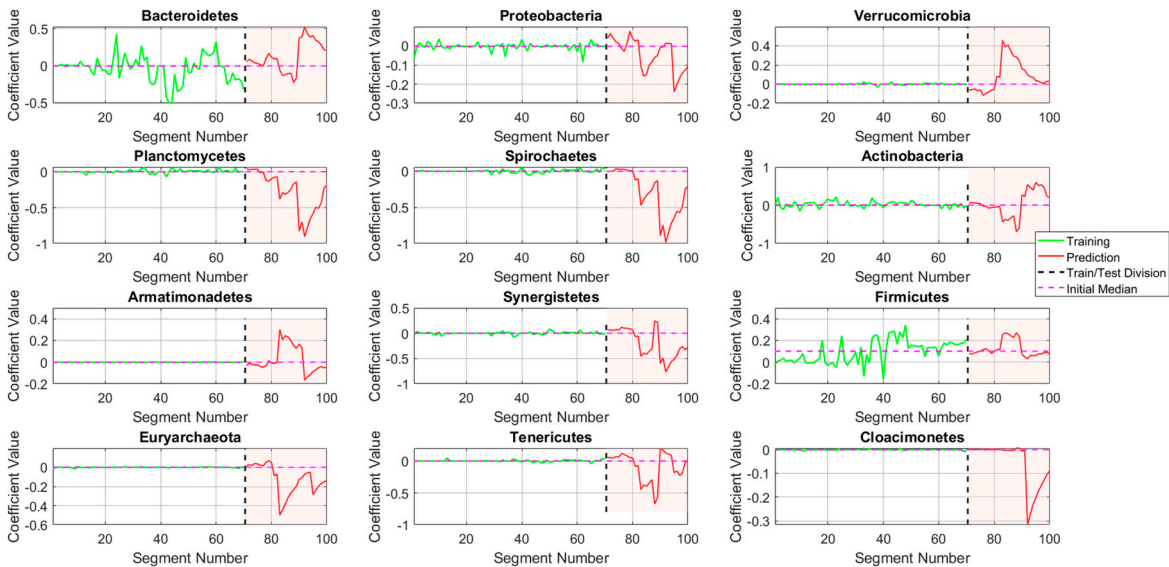


Figure 13. Behavior of the coefficients associated with the weighted variables.

By applying equation (9), the prediction for $\widehat{CH_4}$ was obtained, as shown in Figure 14. This was derived from the coefficients \hat{x} obtained via equation (8).

The development of the dynamic predictive model using a moving window approach combined with Ridge Regression proved to be highly effective. This strategy was designed to capture the evolutionary nature of the biological process and to prevent overfitting that can occur with small data windows. The performance of the final predictive model on the test data was excellent, with an R^2 of 0.920 and an MRE of 6.50%. This confirms that the model not only fits the training data well but also generalizes effectively to make accurate short-term predictions on unseen data. The use of Ridge Regression ($\lambda = 2$) was crucial in stabilizing the coefficients and ensuring the model's robustness.

Despite the promising results, this study has several limitations that should be acknowledged. First, the model was developed using a dataset from a single anaerobic digestion process spanning 58 days. Its performance and the relative importance of the selected variables may not be directly transferable to other digesters with different feedstock or operating conditions. Second, the initial dataset was expanded through linear interpolation to facilitate the moving window analysis. While this is a valid mathematical procedure, it does not generate new experimental information and could potentially mask high-frequency dynamics not captured by the daily sampling rate.

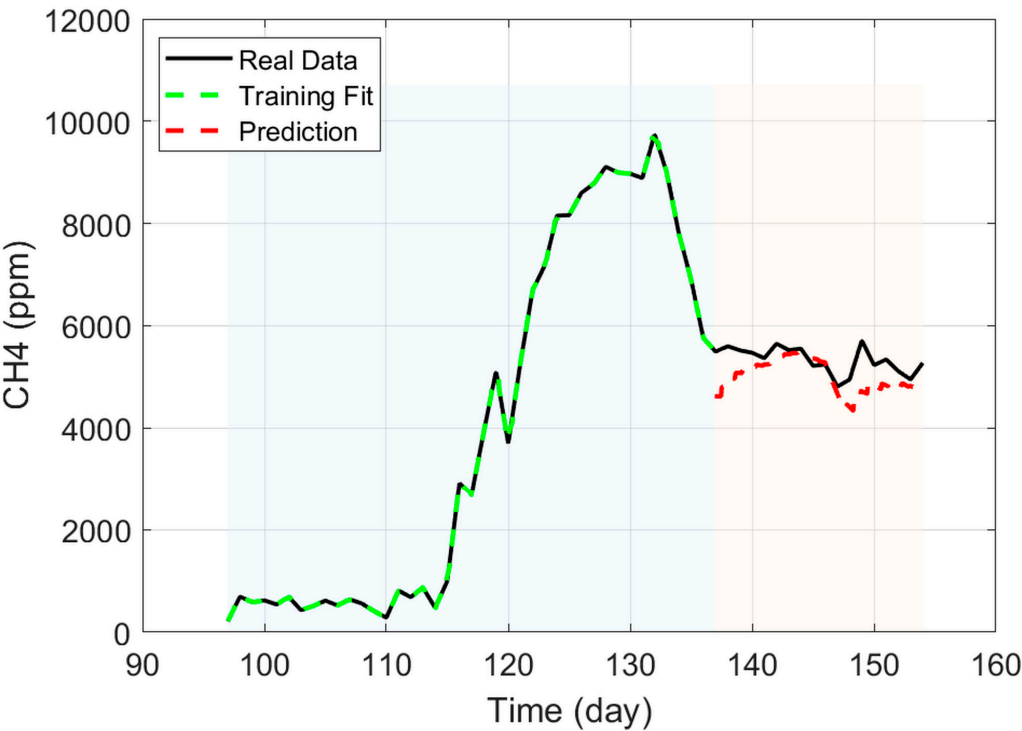


Figure 14. Methane prediction and training data.

The analysis of the data from the training and prediction curves is recorded in Table 11. The R^2 fit values show a strong correlation (greater than 0.9), and the MRE errors are considerably low (less than 7%), with a comparatively low RMSE for the prediction.

Table 11. Evaluation of the training model and predictive model.

	R^2	MRE [%]	RMSE [ppm]
Training Fit	0.999	0.35	20.77
Prediction	0.920	6.50	139.84

Future research should focus on addressing these limitations. Validating the simplified model on long-term datasets from diverse anaerobic digesters is a critical next step to assess its generalizability. Further investigation into the specific metabolic roles of the top 12 identified microorganisms could provide deeper biological insights to complement the model statistical findings. Finally, exploring non-linear modeling techniques, such as gradient boosting or recurrent

neural networks, could potentially capture more complex relationships within the data and further improve predictive accuracy. The ultimate goal would be to integrate such a validated, simplified model with online monitoring sensors for the key microbial populations, paving the way for advanced real-time control and optimization of anaerobic digesters.

4. Conclusions

This study successfully developed a multiple linear regression model to predict methane concentration in anaerobic codigestion using integrated microbial and operational data. The model demonstrated high predictive accuracy ($R^2 = 0.92$, MRE = 6.50%) while requiring only 12 key predictors, substantially reducing complexity compared to the initial 31 variable set. Among the relevant findings, the identification of *Verrucomicrobia* as a significant predictor was particularly noteworthy, as this phylum is typically associated with extreme environments rather than mesophilic digesters, suggesting previously unrecognized ecological adaptations. The overwhelming dominance of microbial indicators over conventional process parameters highlights the critical importance of community dynamics in driving methanogenic performance. Furthermore, the moving window approach with Ridge regularization effectively captured the system's biological evolution while maintaining robustness against overfitting. This modeling approach demonstrates significant potential for practical implementation in rural and resource-limited settings, offering a viable method for methane prediction without sophisticated computational requirements.

Future work should focus on validating this model across diverse reactor configurations and feedstock types to assess its generalizability. Additionally, developing cost-effective molecular monitoring tools for the identified key microbial groups could enable real-time implementation of this predictive approach in practical applications.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Table S1: Phylum level aggregate counts with no imputed data; Table S2: Cleaned database for initial modelling.

Author Contributions: Conceptualization, I.O. and I.R.; methodology, I.O., I.R. and D.C.; software, I.O., I.R. and D.C.; validation, I.O., I.R. and L.M.F.; formal analysis, I.O., I.R. and D.C.; investigation, I.O., and I.R.; resources, I.O., I.R.; data curation, I.O., I.R. and D.C.; writing—original draft preparation, I.O., I.R.; writing—review and editing, I.O., I.R. and L.M.F.; visualization, I.O., I.R. and D.C.; supervision, L.M.F; project administration, I.O, L.M.F; funding acquisition, I.O., I.R. and L.M.F. All authors have read and agreed to the published version of the manuscript.

Funding: The author(s) declare that financial support was received for the research, authorship and publication of this article. This research has been funded by Dirección General de Investigaciones of Universidad Santiago de Cali under call No. 01-2024, and the Universidad Autónoma de Occidente.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author(s).

Acknowledgments: The authors thank the Universidad Autónoma de Occidente and the Universidad Santiago de Cali for their invaluable support. This research has been partly funded by Dirección General de Investigaciones of Universidad Santiago de Cali under call No. 01-2024.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sidartha, Z.; Mendoza, J.C.; Gonzalez, L.S.; Kaiser, F.L.; Gebauer, A. Guía de Biogás Para El Sector Porcícola En Colombia. *J Chem Inf Model* **2020**, *53*, 1689–1699.

2. Porkcolombia Crecimiento Real, Estable Continuo, Distintivo de La Colombiana. *Revista porkcolombia* **2024**.
3. Jiang, H.; Qin, Y.; Gadow, S.I.; Li, Y.Y. The Performance and Kinetic Characterization of the Three Metabolic Reactions in the Thermophilic Hydrogen and Acidic Fermentation of Cassava Residue. *Int J Hydrogen Energy* **2017**, *42*, 2868–2877, doi:10.1016/j.ijhydene.2017.01.061.
4. Oghenejoboh, K.M.; Orugba, H.O.; Oghenejoboh, U.M.; Agarry, S.E. Value Added Cassava Waste Management and Environmental Sustainability in Nigeria: A Review. *Environmental Challenges* **2021**, *4*.
5. Padi, R.K.; Chimphango, A. Assessing the Potential of Integrating Cassava Residues-Based Bioenergy into National Energy Mix Using Long-Range Energy Alternatives Planning Systems Approach. *Renewable and Sustainable Energy Reviews* **2021**, *145*, 111071, doi:10.1016/j.rser.2021.111071.
6. Sánchez, A.S.; Silva, Y.L.; Kalid, R.A.; Cohim, E.; Torres, E.A. Waste Bio-Refineries for the Cassava Starch Industry: New Trends and Review of Alternatives. *Renewable and Sustainable Energy Reviews* **2017**, *73*, 1265–1275.
7. Castro, L.; Escalante, H.; Jaimes-Estévez, J.; Díaz, L.J.; Vecino, K.; Rojas, G.; Mantilla, L. Low Cost Digester Monitoring under Realistic Conditions: Rural Use of Biogas and Digestate Quality. *Bioresour Technol* **2017**, *239*, 311–317, doi:10.1016/j.biortech.2017.05.035.
8. World, D.; Kinyua, M.N.; Rowse, L.E.; Ergas, S.J. Review of Small-Scale Tubular Anaerobic Digesters Treating Livestock Waste in The; **2015**;
9. Ostos, I.; Flórez-Pardo, L.M.; Camargo, C. A Metagenomic Approach to Demystify the Anaerobic Digestion Black Box and Achieve Higher Biogas Yield: A Review. *Front Microbiol* **2024**, *15*.
10. Palau, E.; Carmen, V. Digestión Anaerobia de Residuos de Biomasa Para La Producción de Biogás; **2016**;
11. Uddin, M.M.; Wright, M.M. Anaerobic Digestion Fundamentals, Challenges, and Technological Advances. *Physical Sciences Reviews* **2023**, *8*, 2819–2837, doi:10.1515/psr-2021-0068.
12. Akindolire, M.A.; Rama, H.; Roopnarain, A. Psychrophilic Anaerobic Digestion: A Critical Evaluation of Microorganisms and Enzymes to Drive the Process. *Renewable and Sustainable Energy Reviews* **2022**, *161*.
13. Liu, C.M.; Wachemo, A.C.; Tong, H.; Shi, S.H.; Zhang, L.; Yuan, H.R.; Li, X.J. Biogas Production and Microbial Community Properties during Anaerobic Digestion of Corn Stover at Different Temperatures. *Bioresour Technol* **2018**, *261*, 93–103, doi:10.1016/j.biortech.2017.12.076.
14. Rajlakshmi; Jadhav, D.A.; Dutta, S.; Sherpa, K.C.; Jayaswal, K.; Saravanabhupathy, S.; Mohanty, K.T.; Banerjee, R.; Kumar, J.; Rajak, R.C. Co-Digestion Processes of Waste: Status and Perspective. In *Bio-Based Materials and Waste for Energy Generation and Resource Management*; Elsevier, **2023**; pp. 207–241.
15. TG, I.; Haq, I.; Kalamdhad, A.S. Factors Affecting Anaerobic Digestion for Biogas Production: A Review. In *Advanced Organic Waste Management*; Elsevier, **2022**; pp. 223–233.
16. Khanal, S.K.; Tirta Nindhia, T.G.; Nitayavardhana, S. Biogas From Wastes. In *Sustainable Resource Recovery and Zero Waste Approaches*; Elsevier, **2019**; pp. 165–174.
17. Velásquez, M.E.; Rincón, J.M. Estimación Del Potencial de Conversión a Biogás de La Biomasa En Colombia y Su Aprovechamiento; Colombia, **2018**;
18. AP, Y.; Farghali, M.; Mohamed, I.M.A.; Iwasaki, M.; Tangtaweewipat, S.; Ihara, I.; Sakai, R.; Umetsu, K. Potential of Biogas Production from the Anaerobic Digestion of Sargassum Fulvellum Macroalgae: Influences of Mechanical, Chemical, and Biological Pretreatments. *Biochem Eng J* **2021**, *175*, doi:10.1016/j.bej.2021.108140.
19. Gulsen Akbay, H.E.; Dizge, N.; Kumbur, H. Enhancing Biogas Production of Anaerobic Co-Digestion of Industrial Waste and Municipal Sewage Sludge with Mechanical, Chemical, Thermal, and Hybrid Pretreatment. *Bioresour Technol* **2021**, *340*, 125688, doi:10.1016/j.biortech.2021.125688.
20. Wu, L.J.; Kobayashi, T.; Li, Y.Y.; Xu, K.Q. Comparison of Single-Stage and Temperature-Phased Two-Stage Anaerobic Digestion of Oily Food Waste. *Energy Convers Manag* **2015**, *106*, 1174–1182, doi:10.1016/j.enconman.2015.10.059.
21. Maspolim, Y.; Zhou, Y.; Guo, C.; Xiao, K.; Ng, W.J. Comparison of Single-Stage and Two-Phase Anaerobic Sludge Digestion Systems - Performance and Microbial Community Dynamics. *Chemosphere* **2015**, *140*, 54–62, doi:10.1016/j.chemosphere.2014.07.028.

22. Velásquez Piñas, J.A.; Venturini, O.J.; Silva Lora, E.E.; Calle Roalcaba, O.D. Technical Assessment of Mono-Digestion and Co-Digestion Systems for the Production of Biogas from Anaerobic Digestion in Brazil. *Renew Energy* **2018**, *117*, 447–458, doi:10.1016/j.renene.2017.10.085.
23. Orlando, M.Q.; Borja, V.M. Pretreatment of Animal Manure Biomass to Improve Biogas Production: A Review. *Energies (Basel)* **2020**, *13*, 3573, doi:10.3390/EN13143573.
24. Jain, S.; Jain, S.; Wolf, I.T.; Lee, J.; Tong, Y.W. A Comprehensive Review on Operating Parameters and Different Pretreatment Methodologies for Anaerobic Digestion of Municipal Solid Waste. *Renewable and Sustainable Energy Reviews* **2015**, *52*, 142–154, doi:10.1016/j.rser.2015.07.091.
25. Mohammad Rahmani, A.; Gahlot, P.; Moustakas, K.; Kazmi, A.A.; Shekhar Prasad Ojha, C.; Tyagi, V.K. Pretreatment Methods to Enhance Solubilization and Anaerobic Biodegradability of Lignocellulosic Biomass (Wheat Straw): Progress and Challenges. *Fuel* **2022**, *319*, doi:10.1016/j.fuel.2022.123726.
26. Enokida, C.H.; Tapparo, D.C.; Antes, F.G.; Radis Steinmetz, R.L.; Magrini, F.E.; Sophiatti, I.V.M.; Paesi, S.; Kunz, A. Anaerobic Codigestion of Livestock Manure and Agro-Industrial Waste in a CSTR Reactor: Operational Aspects, Digestate Characteristics, and Microbial Community Dynamics. *Renew Energy* **2025**, *238*, doi:10.1016/j.renene.2024.121865.
27. Panichnumsin, P.; Nopharatana, A.; Ahring, B.; Chaiprasert, P. Production of Methane by Co-Digestion of Cassava Pulp with Various Concentrations of Pig Manure. *Biomass Bioenergy* **2010**, *34*, 1117–1124, doi:10.1016/j.biombioe.2010.02.018.
28. Martins, R.M. AVALIAÇÃO DA CO-DIGESTÃO ANAERÓBIA COMO ALTERNATIVA PARA VALORIZAÇÃO DE RESÍDUOS DO PROCESSAMENTO DE MANDIOCA E ESTERCO DE GADO LEITEIRO, Universidade Federal de Ouro Preto: Ouro Preto, 2022.
29. Xie, S.; Wickham, R.; Nghiem, L.D. Synergistic Effect from Anaerobic Co-Digestion of Sewage Sludge and Organic Wastes. *Int Biodeterior Biodegradation* **2017**, *116*, 191–197, doi:10.1016/j.ibiod.2016.10.037.
30. Tijani, H.; Yuzir, A.; Abdullah, N. Producing Desulfurized Biogas Using Two-Stage Domesticated Shear-Loop Anaerobic Contact Stabilization System. *Waste Management* **2018**, *78*, 770–780, doi:10.1016/j.wasman.2018.06.045.
31. Srisowmeya, G.; Chakravarthy, M.; Nandhini Devi, G. Critical Considerations in Two-Stage Anaerobic Digestion of Food Waste – A Review. *Renewable and Sustainable Energy Reviews* **2020**, *119*, 109587, doi:10.1016/j.rser.2019.109587.
32. Rincón, B.; Borja, R.; González, J.M.; Portillo, M.C.; Sáiz-Jiménez, C. Influence of Organic Loading Rate and Hydraulic Retention Time on the Performance, Stability and Microbial Communities of One-Stage Anaerobic Digestion of Two-Phase Olive Mill Solid Residue. *Biochem Eng J* **2008**, *40*, 253–261, doi:10.1016/j.bej.2007.12.019.
33. Jiraprasertwong, A.; Maitriwong, K.; Chavadej, S. Production of Biogas from Cassava Wastewater Using a Three-Stage Upflow Anaerobic Sludge Blanket (UASB) Reactor. *Renew Energy* **2019**, *130*, 191–205, doi:10.1016/j.renene.2018.06.034.
34. Yang, S.; Liu, Y.; Wu, N.; Zhang, Y.; Svoronos, S.; Pullammanappallil, P. Low-Cost, Arduino-Based, Portable Device for Measurement of Methane Composition in Biogas. *Renew Energy* **2019**, *138*, 224–229, doi:10.1016/j.renene.2019.01.083.
35. Kalamaras, S.D.; Tsitsimpikou, M.A.; Tzenos, C.A.; Lithourgidis, A.A.; Pitsikoglou, D.S.; Kotsopoulos, T.A. A Low-Cost IoT System Based on the ESP32 Microcontroller for Efficient Monitoring of a Pilot Anaerobic Biogas Reactor. *Applied Sciences* **2025**, *15*, doi:10.3390/app15010034.
36. Mabrouki, J.; Azrour, M.; Fattah, G.; Dhiba, D.; Hajjaji, S. El Intelligent Monitoring System for Biogas Detection Based on the Internet of Things: Mohammedia, Morocco City Landfill Case. *Big Data Mining and Analytics* **2021**, *4*, 10–17, doi:10.26599/BDMA.2020.9020017.
37. Gupta, A. Making Biogas SMART Using Internet of Things (LOT). *2020 4th International Conference on Electronics, Materials Engineering and Nano-Technology, IEMENTech 2020* **2020**, 2020–2023, doi:10.1109/IEMENTech51367.2020.9270067.
38. TheMathWorks ThingSpeak for Students and Educators Available online: <https://thingspeak.mathworks.com/pages/education> (accessed on 2 March 2025).

39. Treu, L.; Kougias, P.G.; Campanaro, S.; Bassani, I.; Angelidaki, I. Deeper Insight into the Structure of the Anaerobic Digestion Microbial Community; The Biogas Microbiome Database Is Expanded with 157 New Genomes. *Bioresour Technol* **2016**, *216*, 260–266, doi:10.1016/j.biortech.2016.05.081.
40. Tsapekos, P.; Kougias, P.G.; Treu, L.; Campanaro, S.; Angelidaki, I. Process Performance and Comparative Metagenomic Analysis during Co-Digestion of Manure and Lignocellulosic Biomass for Biogas Production. *Appl Energy* **2017**, *185*, 126–135, doi:10.1016/j.apenergy.2016.10.081.
41. Li, L.; Peng, X.; Wang, X.; Wu, D. Anaerobic Digestion of Food Waste: A Review Focusing on Process Stability. *Bioresour Technol* **2018**, *248*, 20–28, doi:10.1016/j.biortech.2017.07.012.
42. Fontana, A.; Campanaro, S.; Treu, L.; Kougias, P.G.; Cappa, F.; Morelli, L.; Angelidaki, I. Performance and Genome-Centric Metagenomics of Thermophilic Single and Two-Stage Anaerobic Digesters Treating Cheese Wastes. *Water Res* **2018**, *134*, 181–191, doi:10.1016/j.watres.2018.02.001.
43. Jünemann, S.; Kleinbölting, N.; Jaenicke, S.; Henke, C.; Hassa, J.; Nelkner, J.; Stolze, Y.; Albaum, S.P.; Schlüter, A.; Goesmann, A.; et al. Bioinformatics for NGS-Based Metagenomics and the Application to Biogas Research. *J Biotechnol* **2017**, *261*, 10–23, doi:10.1016/j.jbiotec.2017.08.012.
44. Rossi, E.; Pecorini, I.; Iannelli, R. Multilinear Regression Model for Biogas Production Prediction from Dry Anaerobic Digestion of OFMSW. *Sustainability* **2022**, *14*, doi:10.3390/su14084393.
45. Abubakar, U.A.; Lemar, G.S.; Bello, A.A.D.; Ishaq, A.; Dandajeh, A.A.; Jagun, Z.T.; Houmsi, M.R. Evaluation of Traditional and Machine Learning Approaches for Modeling Volatile Fatty Acid Concentrations in Anaerobic Digestion of Sludge: Potential and Challenges. *Environmental Science and Pollution Research* **2024**, doi:10.1007/s11356-024-33281-2.
46. Venkiteshwaran, K.; Milferstedt, K.; Hamelin, J.; Fujimoto, M.; Johnson, M.; Zitomer, D.H. Correlating Methane Production to Microbiota in Anaerobic Digesters Fed Synthetic Wastewater. *Water Res* **2017**, *110*, 161–169, doi:10.1016/j.watres.2016.12.010.
47. APHA; AWWA; WEF Standard Methods Online - Standard Methods for the Examination of Water and Wastewater; 2025;
48. Wilson, S. The MICE Algorithm Available online: <https://cran.r-project.org/web/packages/miceRanger/vignettes/miceAlgorithm.html> (accessed on 4 May 2025).
49. Gaspari, M.; Ghiotto, G.; Centurion, V.B.; Kotsopoulos, T.; Santinello, D.; Campanaro, S.; Treu, L.; Kougias, P.G. Decoding Microbial Responses to Ammonia Shock Loads in Biogas Reactors through Metagenomics and Metatranscriptomics. *Environ Sci Technol* **2024**, *58*, 591–602, doi:10.1021/acs.est.3c07840.
50. APHA; AWWA; WEF 5560 ORGANIC AND VOLATILE ACIDS Standard Methods For the Examination of Water and Wastewater, 24th; Braun-Howland E, editors, Ed.; 14th ed.; APHA Press: Washington DC, 2015;
51. Omega Bio-tek, I. E.Z.N.A.® Universal Pathogen DNA Kit; Norcross, 2018;
52. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High-Resolution Sample Inference from Illumina Amplicon Data. *Nat Methods* **2016**, *13*, 581–583, doi:10.1038/nmeth.3869.
53. Pruesse, E.; Quast, C.; Knittel, K.; Fuchs, B.M.; Ludwig, W.; Peplies, J.; Glockner, F.O. SILVA: A Comprehensive Online Resource for Quality Checked and Aligned Ribosomal RNA Sequence Data Compatible with ARB. *Nucleic Acids Res* **2007**, *35*, 7188–7196, doi:10.1093/nar/gkm864.
54. Zhang, L.; Loh, K.C.; Lim, J.W.; Zhang, J. Bioinformatics Analysis of Metagenomics Data of Biogas-Producing Microbial Communities in Anaerobic Digesters: A Review. *Renewable and Sustainable Energy Reviews* **2019**, *100*, 110–126, doi:https://doi.proxyuao.elogim.com/10.1016/j.rser.2018.10.021.
55. Navarro-Díaz, M.; Aparicio-Trejo, V.; Valdez-Vazquez, I.; Carrillo-Reyes, J.; Avitia, M.; Escalante, A.E. Levels of Microbial Diversity Affect the Stability and Function of Dark Fermentation Bioreactors. *Frontiers in Industrial Microbiology* **2024**, *2*, doi:10.3389/finmi.2024.1386726.
56. Xu, R.Z.; Cao, J.S.; Wu, Y.; Wang, S.N.; Luo, J.Y.; Chen, X.; Fang, F. An Integrated Approach Based on Virtual Data Augmentation and Deep Neural Networks Modeling for VFA Production Prediction in Anaerobic Fermentation Process. *Water Res* **2020**, *184*, doi:10.1016/j.watres.2020.116103.
57. Long, F.; Wang, L.; Cai, W.; Lesnik, K.; Liu, H. Predicting the Performance of Anaerobic Digestion Using Machine Learning Algorithms and Genomic Data. *Water Res* **2021**, *199*, doi:10.1016/j.watres.2021.117182.

58. Santinello, D.; Zampieri, G.; Agostini, S.; Müller, B.; Favaro, L.; Treu, L.; Campanaro, S. Process Stability in Anaerobic Digestion: Unveiling Microbial Signatures of Full-Scale Reactor Performance. *Chemical Engineering Journal* **2024**, *497*, doi:10.1016/j.cej.2024.154962.
59. Lu, D.; Li, M.; Nie, E.; Guo, R.; Fu, S. Microbial Volatile Organic Compounds Produced during the Anaerobic Digestion Process Can Serve as Potential Indicators of Microbial Community Stability. *Water Res* **2025**, *277*, doi:10.1016/j.watres.2025.123286.
60. Mumuni, A.; Mumuni, F. Data Augmentation: A Comprehensive Survey of Modern Approaches. *Array* **2022**, *16*.
61. Schnürer, A.; Jarvis, A. *Microbiology of the Biogas Process*; Sweden, 2018; ISBN 9789157695468.
62. Sun, Y.; Dai, H. liang; Moayed, H.; Nguyen Le, B.; Muhammad Adnan, R. Predicting Steady-State Biogas Production from Waste Using Advanced Machine Learning-Metaheuristic Approaches. *Fuel* **2024**, *355*, doi:10.1016/j.fuel.2023.129493.
63. de Jonge, N.; Moset, V.; Møller, H.B.; Nielsen, J.L. Microbial Population Dynamics in Continuous Anaerobic Digester Systems during Start up, Stable Conditions and Recovery after Starvation. *Bioresour Technol* **2017**, *232*, 313–320, doi:10.1016/j.biortech.2017.02.036.
64. Chen, H.; Zhang, W.; Wu, J.; Chen, X.; Liu, R.; Han, Y.; Xiao, B.; Yu, Z.; Peng, Y. Improving Two-Stage Thermophilic-Mesophilic Anaerobic Co-Digestion of Swine Manure and Rice Straw by Digestate Recirculation. *Chemosphere* **2021**, *274*, doi:10.1016/j.chemosphere.2021.129787.
65. Wu, C.; Huang, Q.; Yu, M.; Ren, Y.; Wang, Q.; Sakai, K. Effects of Digestate Recirculation on a Two-Stage Anaerobic Digestion System, Particularly Focusing on Metabolite Correlation Analysis. *Bioresour Technol* **2018**, *251*, 40–48, doi:10.1016/j.biortech.2017.12.020.
66. Ma, G.; Chen, Y.; Ndegwa, P. Association between Methane Yield and Microbiota Abundance in the Anaerobic Digestion Process: A Meta-Regression. *Renewable and Sustainable Energy Reviews* **2021**, *135*.
67. Raja Ram, N.; Nikhil, G.N. A Critical Review on Sustainable Biogas Production with Focus on Microbial-Substrate Interactions: Bottlenecks and Breakthroughs. *Bioresour Technol Rep* **2022**, *19*.
68. Kim, H.; Jeon, B.S.; Sang, B.-I. An Efficient New Process for the Selective Production of Odd-Chain Carboxylic Acids by Simple Carbon Elongation Using *Megasphaera Hexanoica*. *Sci Rep* **2019**, *9*, 11999, doi:10.1038/s41598-019-48591-6.
69. Sun, J.; Zhang, L.; Loh, K.C. Review and Perspectives of Enhanced Volatile Fatty Acids Production from Acidogenic Fermentation of Lignocellulosic Biomass Wastes. *Bioresour Bioprocess* **2021**, *8*, doi:10.1186/s40643-021-00420-3.
70. Harirchi, S.; Wainaina, S.; Sar, T.; Nojoumi, S.A.; Parchami, M.; Parchami, M.; Varjani, S.; Khanal, S.K.; Wong, J.; Awasthi, M.K.; et al. Microbiological Insights into Anaerobic Digestion for Biogas, Hydrogen or Volatile Fatty Acids (VFAs): A Review. *Bioengineered* **2022**, *13*, 6521–6557.
71. Franke-Whittle, I.H.; Walter, A.; Ebner, C.; Insam, H. Investigation into the Effect of High Concentrations of Volatile Fatty Acids in Anaerobic Digestion on Methanogenic Communities. *Waste Management* **2014**, *34*, 2080–2089, doi:10.1016/j.wasman.2014.07.020.
72. Lonkar, S.; Fu, Z.; Holtzapfel, M. Optimum Alcohol Concentration for Chain Elongation in Mixed-culture Fermentation of Cellulosic Substrate. *Biotechnol Bioeng* **2016**, *113*, 2597–2604, doi:10.1002/bit.26024.
73. Duber, A.; Zagrodnik, R.; Gutowska, N.; Brodowski, F.; Dąbrowski, T.; Dąbrowski, S.; Łężyk, M.; Oleskowicz-Popiel, P. Single- vs. Two-Stage Fermentation of an Organic Fraction of Municipal Solid Waste for an Enhanced Medium Chain Carboxylic Acids Production – The Impact of Different PH and Temperature. *Bioresour Technol* **2025**, *415*, doi:10.1016/j.biortech.2024.131697.
74. Zakaria, B.S.; Guo, H.; Kim, Y.; Dhar, B.R. Molecular Biology and Modeling Analysis Reveal Functional Roles of Propionate to Acetate Ratios on Microbial Syntrophy and Competition in Electro-Assisted Anaerobic Digestion. *Water Res* **2022**, *216*, 118335, doi:10.1016/j.watres.2022.118335.
75. Jabłoński, S.J.; Łukaszewicz, M. Mathematical Modelling of Methanogenic Reactor Start-up: Importance of Volatile Fatty Acids Degrading Population. *Bioresour Technol* **2014**, *174*, 74–80, doi:10.1016/j.biortech.2014.09.151.

76. Basile, A.; Campanaro, S.; Kovalovszki, A.; Zampieri, G.; Rossi, A.; Angelidaki, I.; Valle, G.; Treu, L. Revealing Metabolic Mechanisms of Interaction in the Anaerobic Digestion Microbiome by Flux Balance Analysis. *Metab Eng* **2020**, *62*, 138–149, doi:10.1016/j.ymben.2020.08.013.
77. Iglesias-Iglesias, R.; Campanaro, S.; Treu, L.; Kennes, C.; Veiga, M.C. Valorization of Sewage Sludge for Volatile Fatty Acids Production and Role of Microbiome on Acidogenic Fermentation. *Bioresour Technol* **2019**, *291*, 121817, doi:10.1016/j.biortech.2019.121817.
78. Yun, Y.M.; Sung, S.; Kang, S.; Kim, M.S.; Kim, D.H. Enrichment of Hydrogenotrophic Methanogens by Means of Gas Recycle and Its Application in Biogas Upgrading. *Energy* **2017**, *135*, 294–302, doi:10.1016/j.energy.2017.06.133.
79. Niya, B.; Yaakoubi, K.; Beraich, F.Z.; Arouch, M.; Meftah Kadmiri, I. Current Status and Future Developments of Assessing Microbiome Composition and Dynamics in Anaerobic Digestion Systems Using Metagenomic Approaches. *Heliyon* **2024**, *10*.
80. Schmitz, R.A.; Peeters, S.H.; Versantvoort, W.; Picone, N.; Pol, A.; Jetten, M.S.M.; Op Den Camp, H.J.M. Verrucomicrobial Methanotrophs: Ecophysiology of Metabolically Versatile Acidophiles. *FEMS Microbiol Rev* **2021**, *45*.
81. Dunfield, P.F.; Yuryev, A.; Senin, P.; Smirnova, A. V.; Stott, M.B.; Hou, S.; Ly, B.; Saw, J.H.; Zhou, Z.; Ren, Y.; et al. Methane Oxidation by an Extremely Acidophilic Bacterium of the Phylum Verrucomicrobia. *Nature* **2007**, *450*, 879–882, doi:10.1038/nature06411.
82. Basile, A.; Zampieri, G.; Kovalovszki, A.; Karkaria, B.; Treu, L.; Patil, K.R.; Campanaro, S. Modelling of Microbial Interactions in Anaerobic Digestion: From Black to Glass Box. *Curr Opin Microbiol* **2023**, *75*.
83. Zhang, L.; Loh, K.C.; Lim, J.W.; Zhang, J. Uncovering Microbiome Adaptations in a Full-Scale Biogas Plant: Insights from MAG-Centric Metagenomics and Metaproteomics. *Microorganisms* **2023**, *11*, doi:10.3390/microorganisms11102412.
84. Zhang, X.; Wang, Y.; Jiao, P.; Zhang, M.; Deng, Y.; Jiang, C.; Liu, X.-W.; Lou, L.; Li, Y.; Zhang, X.-X.; et al. Microbiome-Functionality in Anaerobic Digesters: A Critical Review. *Water Res* **2023**, *249*, 120891, doi:10.1016/j.watres.2023.120891.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.