

Optimization of Anaerobic Digestion through In-situ Hydrogen Biomethanation and Machine
Learning Prediction of Methane Production

Pouria Paydarnik

Thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the
Master of Applied Science in Civil Engineering

Department of Civil Engineering
Academic Supervisor: Prof. Hossein Bonakdari
Faculty of Engineering
University of Ottawa

© Pouria Paydarnik, Ottawa, Canada, 2026

Abstract

Biogas produced through anaerobic digestion (AD) represents an important renewable energy carrier, yet the relatively low methane (CH_4) content of raw biogas limits its value as fuel and grid-injectable gas. In-situ hydrogen (H_2) biomethanation (HBM) the direct injection of H_2 into an operating anaerobic digester to promote hydrogenotrophic conversion of carbon dioxide (CO_2) into additional CH_4 offers a promising route for biological biogas upgrading. However, the low aqueous solubility of H_2 means that gas-liquid mass transfer is frequently the rate-limiting step, and existing Anaerobic Digestion Model No. 1 (ADM1) implementations do not adequately represent the dynamic interplay between H_2 supply, transfer intensity, and process stability.

This thesis presents a hybrid mechanistic, data-driven framework to optimize in-situ HBM and predict CH_4 production using machine learning. The standard ADM1 model, implemented within the Benchmark Simulation Model No. 2 (BSM2) structure using the open-source PyADM1 platform, was extended in three ways: (i) an external H_2 injection term was incorporated into the gas-phase H_2 balance; (ii) gas-specific volumetric mass transfer coefficients ($k_L a$) for H_2 , CO_2 , and CH_4 were introduced using diffusivity-based scaling relationships; and (iii) an inorganic carbon limitation factor was added to the hydrogenotrophic methanogenesis rate to prevent unrealistic H_2 conversion under carbon-depleted conditions. The modified model was verified against published biomethanation benchmark data, achieving deviations below 8% for CH_4 production and below 1% for pH across the stable operating range.

Long-term dynamic simulations were conducted over 280-day periods across a matrix of $k_L a$ scenarios (60 to 1000 d^{-1}) and H_2 injection rates (0 to 2500 $\text{m}^3 \cdot \text{d}^{-1}$), yielding a dataset of approximately 280 daily observations per scenario. Results showed that CH_4 production increased consistently with H_2 injection under all $k_L a$ conditions, but the efficiency of H_2 utilization depended strongly on mass transfer intensity. Under low $k_L a$ conditions (60 d^{-1}), H_2 conversion efficiency reached only 57-61%, with substantial H_2 slip and declining CH_4 content at higher injection rates. Under high $k_L a$ conditions (1000 d^{-1}), H_2 conversion efficiency exceeded 96% across the full injection range, with CH_4 content in the outlet gas reaching up to 84%. An injection-efficiency evaluation framework was developed to identify the most favorable operating points under each mass-transfer scenario for subsequent machine-learning (ML) dataset preparation.

Four ML algorithms Support Vector Regression (SVR), Random Forest (RF), Extreme Learning Machine (ELM), and Evolutionary Polynomial Regression (EPR) were trained using ten process variables to predict CH₄ and total biogas flow rates. SVR achieved the highest testing accuracy, with R² values of 0.996 for CH₄ and 0.9948 for biogas, followed by ELM with 0.9899 and 0.9834, respectively. RF also demonstrated strong predictive performance, with testing R² values of 0.9879 for CH₄ and 0.9807 for biogas. EPR produced lower but still strong predictive performance, with testing R² values of 0.9561 and 0.9276, while offering the added advantage of explicit, interpretable mathematical expressions.

The findings demonstrate that in-situ H₂ biomethanation can substantially improve CH₄ yield and biogas quality when gas-liquid mass transfer is sufficiently strong, and that the combined ADM1 ML framework developed in this work provides both process insight and practical prediction tools for anaerobic digestion optimization. The EPR-derived explicit formula further provides a transparent and computationally inexpensive alternative for CH₄ estimation within the investigated operating domain.

List of Abbreviations

Abbreviation	Full term
AD	Anaerobic Digestion
ADM1	Anaerobic Digestion Model No. 1
ANN	Artificial Neural Network
ASM	Activated Sludge Model
BMP	Biomethane Potential
BSM2	Benchmark Simulation Model No. 2
PyADM1	Python implementation of ADM1
CH ₄	Methane
CHP	Combined Heat and Power
N ₂	Nitrogen
O ₂	Oxygen
H ₂ O	Water
CO	Carbon Monoxide
CO ₂	Carbon Dioxide
COD	Chemical Oxygen Demand
CSTR	Continuous Stirred Tank Reactor
RBF	Radial Basis Function
STP	Standard Temperature and Pressure
DAE	Differential Algebraic Equation
ELM	Extreme Learning Machine
EPR	Evolutionary Polynomial Regression
GA	Genetic Algorithm
GHG	Greenhouse Gas
H ₂	Hydrogen
H ₂ S	Hydrogen Sulfide
HBM	Hydrogen Biomethanation
HRT	Hydraulic Retention Time
IWA	International Water Association
$k_L a$	Volumetric Gas-Liquid Mass Transfer Coefficient
LCFA	Long-Chain Fatty Acid
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error

NH ₃	Ammonia
NRMSE	Normalized Root Mean Square Error
ODE	Ordinary Differential Equation
OLR	Organic Loading Rate
pH	Potential of Hydrogen
RF	Random Forest
RMSE	Root Mean Square Error
RNG	Renewable Natural Gas
R ²	Coefficient of Determination
SCADA	Supervisory Control and Data Acquisition
S _{IC}	Soluble Inorganic Carbon
S _{IN}	Soluble Inorganic Nitrogen
SLFN	Single-Hidden-Layer Feedforward Neural Network
SVM	Support Vector Machine
SVR	Support Vector Regression
TS	Total Solids
VFA	Volatile Fatty Acid
VS	Volatile Solids
WRRF	Water Resource Recovery Facility
WWTP	Wastewater Treatment Plant
XGBoost	Extreme Gradient Boosting

Acknowledgements

I would like to express my deepest gratitude to my parents, whose unwavering support and encouragement have been the foundation of everything I have achieved. Their belief in me through every challenge and every milestone has been a constant source of strength, and I could not have completed this journey without them. This work reflects as much their dedication and sacrifice as my own effort.

I am sincerely grateful to my supervisor, Prof. Hossein Bonakdari, for his invaluable guidance, expertise, and patience throughout the course of this research. His thoughtful feedback, constructive criticism, and commitment to academic excellence consistently pushed me to think more rigorously and write more precisely. It has been a privilege to conduct this research under his supervision, and I deeply appreciate the time and care he invested in my development as a researcher.

I would also like to extend my sincere thanks to Mohammad Salimi Delshad, for his generous support and collaboration during this work. His willingness to share his knowledge, engage in meaningful discussions, and help whenever it was needed made a genuine difference throughout this process. I am grateful for his collegiality and for the time he devoted to helping me navigate the challenges of this research.

Table of Contents

Abstract	ii
List of Abbreviations	iv
Acknowledgements	vi
Table of contents	vii
List of Figures	xii
List of Tables	xiv
1. Introduction.....	1
1.1. Background and Motivation	1
1.2. Problem Statement	3
1.3. Thesis Organization	4
2. Literature Review.....	6
2.1. Fundamentals of Anaerobic Digestion.....	6
2.1.1. Overview of Anaerobic Digestion	6
2.1.2. Products and Energy Potential	6
2.1.3. Biochemical Stages of Anaerobic Digestion Process	7
2.2. Gas-Liquid Mass Transfer in Anaerobic Systems	9
2.2.1. Gas-Liquid Mass Transfer	9
2.2.2. Factors Influencing kLa	10
2.3. Operating Conditions and Process Stability	10
2.3.1. Temperature Regimes	11
2.3.2. pH and Inhibition Mechanisms	12

2.3.3. Anaerobic Reactor Configurations	12
2.4. Biomethanation Technologies	13
2.4.1. Biomethanation	14
2.4.2. Hydrogenotrophic Pathway	14
2.4.3. Carbon Monoxide Injection Strategies	14
2.4.4. In-situ and Ex-situ Biomethanation	15
2.4.5. Mass Transfer Limitations in Biomethanation	16
2.5. Mathematical Modelling of Anaerobic Digestion	16
2.5.1. The ADM1 Framework.....	17
2.5.2. Benchmark Simulation Model No. 2 (BSM2)	17
2.5.3. Coupling ADM1 and BSM2	17
2.5.4. Limitations of ADM1 for Biomethanation	18
2.5.5. Extensions of ADM1 Toward Biomethanation	18
2.6. Data-Driven Modelling and Machine Learning in Anaerobic Digestion	19
2.6.1. Data-Driven Modelling Approaches and Algorithms.....	19
2.6.2. Hybrid Machine Learning and Mechanistic Simulation.....	20
2.6.3. Research Gaps and Thesis Objectives	21
3. Methodology	22
3.1. Research framework	22
3.2. ADM1/BSM2 Simulation Framework.....	22
3.3. Modified ADM1 model for in-situ HBM	23
3.3.1. H ₂ injection model	24
3.3.2. Scenario-based gas-liquid mass transfer model.....	24
3.4. Reactor configuration, input files, and operating conditions.....	25
3.4.1. Hydraulic and loading indicators	29

3.4.2. Inorganic carbon limitation of hydrogen uptake.....	29
3.5. Scenario design	30
3.5.1. Hydrogen injection scenarios.....	30
3.5.2. Gas-liquid mass-transfer scenarios	31
3.5.3. Scenario combination matrix	32
3.6. Injection-efficiency and biomethanation performance indicators	32
3.6.1. H ₂ utilization indicators	32
3.6.2. Gas-upgrading indicator.....	33
3.6.3. Role of the indicators in scenario evaluation.....	33
3.7. Model verification.....	33
3.8. Dataset collection and preparation.....	36
3.9. Feature selection and data preprocessing.....	37
3.10. Machine-learning model development	39
3.10.1. Random Forest	40
Hyperparameters of the Random Forest model	41
3.10.2. Support Vector Regression	42
Hyperparameters of the SVR model	43
3.10.3. Extreme Learning Machine.....	44
Hyperparameters of the ELM model	46
3.10.4. Evolutionary Polynomial Regression (EPR)	46
Evolutionary Polynomial Regression EPR Hyperparameters	47
3.10.5. Data partitioning and robustness assessment.....	48
3.10.6. Feature importance and sensitivity analysis	49
3.11. Model Evaluation Metrics.....	50
4. Results and discussion	52
4.1. Overview of ADM1 simulation results.....	52

4.2.	Reference performance of the baseline configuration	52
4.3.	Effect of hydrogen injection under different gas-liquid mass transfer scenarios	53
4.3.1.	CH ₄ production.....	53
4.3.2.	Outlet H ₂ flow under different injection and kLa scenarios	55
4.3.3.	Average pH under different H ₂ injection and kLa scenarios	57
4.3.4.	Maximum pH under different injection and kLa scenarios.....	58
4.3.5.	Average SIC under different injection and kLa scenarios.....	59
4.4.	Soluble Inorganic Nitrogen (SIN)	60
4.5.	Injection-efficiency evaluation	61
4.5.1.	Injection-efficiency evaluation for low kLa scenario.....	62
4.5.2.	Injection-efficiency evaluation under $kLa = 140$	62
4.5.3.	Injection-efficiency evaluation under $kLa = 240$	63
4.5.4.	Injection-efficiency evaluation under $kLa = 360$	64
4.5.5.	Injection-efficiency evaluation under $kLa = 650$	65
4.5.6.	Injection-efficiency evaluation under high mass-transfer scenarios.....	66
4.5.7.	Overall conclusion of the injection-efficiency evaluation	67
4.6.	Selection of efficient scenarios for the machine-learning dataset	68
4.7.	Machine-learning prediction of CH ₄ flow rate	69
4.7.1.	Support Vector Regression (SVR).....	69
4.7.2.	Random Forest (RF)	70
4.7.3.	Extreme Learning Machine (ELM)	72
4.7.4.	Evolutionary Polynomial Regression (EPR)	73
4.7.5.	Comparative discussion	76
4.7.6.	Feature Importance and Sensitivity Analysis	80
5.	Conclusion and Future Works	81

References	85
Appendix 1	93

List of Figures

Figure 1. AD main stages adapted from (Aslanzadeh, 2014; Dizaji, 2023)	8
Figure 2. Growth rate of methanogen bacteria in different thermal conditions (Lettinga et al., 2001)	12
Figure 3. Schematic comparison of in-situ and ex-situ biological biogas upgrading adapted from (Antukh et al., 2022)	16
Figure 4. Scenario-generation framework for dynamic simulations and machine-learning dataset preparation.	32
Figure 5 Comparison of CH ₄ production	35
Figure 6 comparison of pH	35
Figure 7. Representative decision tree structure within the Random Forest model	41
Figure 8. sample of the ELM structure for CH ₄ prediction from selected AD variables adapted from (Hammad et al., 2020).....	45
Figure 9. CH ₄ production under different scenarios	54
Figure 10. CH ₄ increment in each scenario	55
Figure 11. Average H ₂ output in different scenarios	56
Figure 12. Average pH for each condition.....	58
Figure 13. Maximum pH reaching 8.1 in highest <i>kLa</i> and injection scenario	59
Figure 14. Drop in S _{IC} in all scenarios due to H ₂ injection.....	60
Figure 15. Schematic behavior of S _{IN} under each condition.....	61
Figure 16. Actual versus predicted CH ₄ production using the SVR model for training and testing datasets. The red dashed line shows the 1:1 line, and the blue dashed lines represent the ±10% error bounds.	70
Figure 17. Actual versus predicted biogas production using the SVR model for training and testing datasets.	70
Figure 18. Actual versus predicted CH ₄ production using the RF model for testing and training datasets. The red dashed line shows the 1:1 line, and the blue dashed lines represent the ±10% error bounds.	71
Figure 19. RF prediction performance for train and test data set of biogas production	72

Figure 20 Actual versus predicted CH ₄ production using the ELM model for the testing and training datasets.	72
Figure 21 Actual versus predicted biogas production, using the ELM method.....	73
Figure 22 Actual versus predicted CH ₄ production using the EPR model for training and testing datasets. The red dashed line shows the 1:1 line, and the blue dashed lines represent the $\pm 10\%$ error bounds.	74
Figure 23 Actual versus predicted biogas production using the EPR model for testing and training datasets. The red dashed line shows the 1:1 line, and the blue dashed lines represent the $\pm 10\%$ error bounds.....	75
Figure 24 Comparative performance of SVR, RF, EPR, and ELM models for CH ₄ prediction on the testing dataset.....	77
Figure 25 Comparative performance of SVR, RF, EPR, and ELM models for biogas prediction on the testing dataset.....	77

List of Tables

Table 1. Composition of biogas produced during AD (Gomez, 2013).....	7
Table 2. Diffusivity coefficients and Henry’s law constants of key gases involved in biomethanation processes, adapted from (Pauss et al., 1990).....	25
Table 3. Initial ADM1 state variables and descriptions, used to initialize the digester model.....	26
Table 4. Average non-zero values of key variables in the dynamic influent file	28
Table 5. Gas-specific kLa values were derived from the H_2 reference value using diffusivity- based scaling factors.	31
Table 6. Input variables for ML algorithms.....	37
Table 7. Random Forest Selected Hyperparameters.....	42
Table 8. SVR Selected Hyperparameters.....	43
Table 9. ELM Selected Hyperparameters	46
Table 10. EPR Selected Hyperparameters	48
Table 11. Injection-efficiency evaluation under low kLa scenario	62
Table 12. Injection-efficiency evaluation under $kLa = 140$ scenario	63
Table 13. Injection-efficiency evaluation under $kLa = 240$ scenario	64
Table 14. Injection-efficiency evaluation under $kLa = 360$ scenario	65
Table 15. Injection-efficiency evaluation under $kLa = 650$ scenario	66
Table 16. Injection-efficiency evaluation under high kLa scenario	67
Table 17. Selected scenarios for ML development.....	68

1. Introduction

1.1. Background and Motivation

Biogas has emerged as an important renewable energy carrier within the broader transition toward low-carbon and circular energy systems. It is produced through anaerobic decomposition of organic biomass and consists mainly of methane (CH_4) and carbon dioxide (CO_2). Because CH_4 is an energy-bearing component, biogas can serve as a valuable substitute for fossil-based fuels. After basic cleaning, biogas can be converted into electricity and heat in combined heat and power units, used directly for thermal applications, or further upgraded to biomethane for uses similar to conventional natural gas. In this sense, biogas is not only a renewable fuel, but also a flexible energy carrier that can contribute to electricity generation, heating, transportation, and broader energy-system decarbonization (Gomez, 2013).

Biogas's importance extends beyond energy production. Since biogas is generated from organic residues, its production supports waste valorization by converting materials that would otherwise require disposal into useful energy and nutrient-rich products (Gomez, 2013). AD can process a wide variety of substrates, including agricultural residues, animal manure, municipal organics, sewage sludge, and food-processing by-products, which makes it a highly versatile technology within integrated waste-management systems. In addition to generating renewable energy, AD produces digestate, which can be reused as a fertilizer or soil improver and can therefore contribute to nutrient recycling. This dual role of biogas systems in both renewable energy generation and organic waste management makes them especially attractive in the context of circular economic strategies (Martín-Hernández et al., 2025; S. Zhang et al., 2015).

AD is the key biological process that enables biogas production. It involves the degradation of organic matter in the absence of oxygen through a sequence of biochemical stages that ultimately generate CH_4 -rich gas. As described in the literature, AD is widely applied to transform organic matter into biogas while stabilizing the residual material. Its practical importance lies in its ability to treat problematic organic waste while simultaneously recovering energy. For manure management in particular, AD offers several environmental advantages over conventional

practices, including reduced odor, reduced emissions of volatile compounds, improved sanitation due to pathogen reduction, and improved nutrient availability in the resulting digestate. Accordingly, AD is increasingly recognized as a core technology for sustainable biomass valorization and waste-to-energy conversion (Da Costa Gomez, 2013; S. Zhang et al., 2015).

Biogas and AD are particularly relevant in the Canadian context. Canada generates large quantities of organic waste, and a substantial share remains underutilized despite its energy-recovery potential. One review reported that Canada generates 35.5 million tonnes of waste containing 20% to 40% organic fraction municipal solid waste, while current waste-management infrastructure captures only a portion of this resource for diversion from landfill (Norouzi & Dutta, 2022). The same study emphasizes that AD is increasingly viewed as a suitable pathway for converting organic waste into renewable energy and fertilizer, especially as policies supporting renewable natural gas and landfill diversion continue to develop. In Ontario, organic waste is also a major opportunity for circular resource recovery, while landfill constraints and the environmental burden of waste disposal further strengthen the case for AD-based solutions (Ncube et al., 2021; Norouzi & Dutta, 2022).

Canada also presents sector-specific opportunities for biogas deployment. In British Columbia, livestock farms generate large amounts of manure, creating both environmental pressures and a strong feedstock base for AD. A study of a biogas-centered dairy farm-greenhouse system in British Columbia showed that AD could replace some natural gas consumption with renewable biogas, while digestate could be reused as bedding material, growing media, and fertilizer. The same study found that such an integrated system could reduce non-renewable energy use and several environmental impacts by more than 40% relative to conventional practice. This shows that in Canada, biogas extends beyond energy production to include agricultural integration, waste management, and environmental impact reduction (S. Zhang et al., 2015).

Additionally, AD already plays an important role in Canadian wastewater resource recovery facilities. Recent work on Canadian WRRFs notes that biogas is produced from sewage sludge digestion and is commonly used for thermal energy and combined heat and power, with a smaller number of facilities upgrading biogas to renewable natural gas. That study estimated that WRRFs account for a meaningful share of annual Canadian biogas production, highlighting the growing

strategic importance of AD within national waste-management and decarbonization efforts (Martín-Hernández et al., 2025).

A further important development is biogas upgrading, which converts raw biogas into biomethane by removing CO₂ and other unwanted components. Compared with direct use of raw biogas at the production site, upgrading offers several advantages. It enables local and temporal decoupling of production and utilization, provides storage capability, and allows biomethane to be used in multiple pathways including grid injection, vehicle fuel, electricity and heat production, and industrial applications. These characteristics make biomethane a more flexible and potentially more efficient energy carrier than untreated biogas, particularly when energy demand is located far from the digestion site or when year-round local heat demand is insufficient. As a result, biogas upgrading is increasingly viewed as an important step for maximizing the value of AD systems and supporting renewable gas integration into existing infrastructure (Beil & Beyrich, 2013).

The environmental performance of biogas systems depends on energy recovery, process efficiency, and emission control. CH₄ leakage across the biogas production and utilization chain can strongly affect greenhouse-gas performance because CH₄ has a much higher global warming potential than CO₂ (Martín-Hernández et al., 2025). This creates an additional need to optimize digester design, gas handling, upgrading, and operational management. Recent research has also pointed out that capturing biogenic CO₂ downstream of biogas utilization or within upgrading systems could help offset CH₄-related emissions and contribute to more carbon-neutral biogas production. These considerations help to explain the importance of improving the technical, environmental, and operational performance of AD (Gomez, 2013).

1.2. Problem Statement

Although AD is a well-established technology for organic waste stabilization and renewable energy recovery, the quality and utilization potential of raw biogas remain limited by its relatively high CO₂ content. Because only CH₄ contributes directly to biogas's fuel value, increasing the CH₄ fraction is necessary for high-value applications such as grid injection, vehicle fuel, or substitution for conventional natural gas. Conventional upgrading technologies can remove CO₂ from biogas, but they often require additional separation units, energy input, and capital investment. As a result,

biological biogas upgrading through biomethanation has gained increasing attention as a promising alternative.

In-situ (Hydrogen Biomethanation) HBM offers a potentially efficient route for improving biogas quality because H_2 can be injected directly into the anaerobic digester, where hydrogenotrophic methanogens convert CO_2 and H_2 into additional CH_4 . This approach can increase CH_4 production while simultaneously reducing the CO_2 content of the gas phase. However, the practical implementation of in-situ HBM is challenging. The low solubility of H_2 in the liquid phase makes gas-liquid mass transfer a major limiting factor. If the injected H_2 does not transfer efficiently into the liquid phase, a significant portion may leave the reactor through the gas outlet without being biologically converted. This reduces efficiency and may create operational and safety concerns.

Beyond mass-transfer limitations, excessive H_2 injections can affect the digester's chemical stability. The conversion of CO_2 into CH_4 consumes dissolved inorganic carbon and bicarbonate species, which are important components of the buffering system. As a result, intensive H_2 uptake may increase pH and reduce inorganic carbon availability, potentially limiting hydrogenotrophic methanogenesis and disturbing process stability. Therefore, the performance of in-situ HBM depends not only on the amount of H_2 injected, but also on the interaction between H_2 loading, gas-liquid transfer intensity, pH response, and inorganic carbon availability.

These challenges create a need for modelling tools that can evaluate different operating conditions before practical implementation. Mechanistic models such as ADM1 provide a strong basis for representing AD, but the standard ADM1 framework was not originally developed to analyze H_2 injection, gas-liquid transfer limitations, and inorganic carbon limitation during in-situ biomethanation. At the same time, machine-learning models can support CH_4 prediction and process optimization, but their performance depends strongly on the availability and quality of suitable datasets. Therefore, a combined mechanistic and data-driven approach is needed to evaluate biomethanation scenarios and develop predictive tools for CH_4 production.

1.3. Thesis Organization

This thesis is presented in a monograph format and is organized into five chapters as follows:

Chapter 1: Introduction presents the background and motivation for this research, identifies the limitations of current AD modelling frameworks with respect to in-situ HBM, and defines the research gaps.

Chapter 2: Literature Review provides a comprehensive review of AD fundamentals, gas-liquid mass transfer theory, biomethanation technologies, and the ADM1 and BSM2 modelling frameworks. It also reviews existing machine learning applications in AD and consolidates the research gaps that this thesis addresses.

Chapter 3: Methodology describes the full research framework, including the mechanistic extensions made to the standard ADM1 model to incorporate external hydrogen (H_2) injections, diffusivity-based gas-specific mass transfer coefficients, and an inorganic carbon limitation factor. It also details the scenario design matrix, the injection-efficiency evaluation framework, and the development and validation of the four machine learning algorithms used for methane prediction.

Chapter 4: Results and Discussion present the outcomes of the 280-day dynamic simulations across the full $k_L a$ and H_2 injection scenario matrix, followed by the injection-efficiency evaluation and scenario screening. The machine learning results for all four algorithms SVR, RF, ELM, and EPR are then reported and compared using standardized performance metrics.

Chapter 5: Conclusion summarizes the principal findings of this study, discusses their practical implications for the design and operation of in-situ biomethanation systems, acknowledges the limitations of the present work, and outlines directions for future research.

2. Literature Review

2.1. Fundamentals of Anaerobic Digestion

2.1.1. Overview of Anaerobic Digestion

AD is a biological process that involves the degradation and stabilization of organic matter by a complex microbial community in the absence of oxygen (Lovato et al., 2017). This process naturally occurs in anoxic environments such as lake sediments, wetlands, and the digestive systems of ruminant animals (Murphy & Thamsiriroj, 2013). AD is widely implemented for the treatment of municipal sludge, agricultural residues, and organic wastes due to its ability to simultaneously achieve waste stabilization and renewable energy recovery (Chen et al., 2016). The biological conversion results in the production of biogas and a nutrient-rich residue known as digestate (Sadrimajd et al., 2021). AD is widely considered a sustainable and energy-efficient technology for waste management and environmental protection (Balagurusamy & Chandel, 2021). It reduces the environmental impact of organic waste by limiting greenhouse gas (GHG) emissions that would otherwise occur during uncontrolled decomposition in landfills. By capturing CH_4 , a powerful greenhouse gas, AD allows it to be used as a renewable energy source (Balagurusamy & Chandel, 2021). In addition, the process supports the concept of a circular bioeconomy by recovering renewable energy and returning valuable nutrients to the production cycle (Pullen, 2015).

2.1.2. Products and Energy Potential

The biogas produced during AD consists primarily of CH_4 (50-75%) and CO_2 (25-50%) (Gomez, 2013). It also contains minor amounts of other compounds, including nitrogen (N_2), H_2 , water vapor (H_2O), oxygen (O_2), and trace gases such as hydrogen sulfide (H_2S) and ammonia (NH_3) (Angelidaki et al., 2018; Dizaji, 2023). As shown in Table 1, CH_4 is the central element of biogas and is the key product as a fuel due to its high energy density (Angelidaki et al., 2018; Sanjay Shah, 2016). The energy content of biogas is proportional to its CH_4 content; a higher CH_4 concentration among the components of biogas results in greater energy potential for utilization. (Angelidaki et al., 2018). Biogas can be utilized for heat and electricity generation in combined

heat and power units, or it can be upgraded for injection into natural gas distribution grids or used as vehicle fuel (Gomez, 2013; Lovato et al., 2017)

Table 1. Composition of biogas produced during AD (Gomez, 2013).

Component	Chemical formula	Typical volumetric fraction
Methane	CH ₄	50-75 %
Carbon dioxide	CO ₂	25-50 %
Other gases	H ₂ O, H ₂ S, NH ₃ , H ₂ , N ₂ , O ₂	traces

Digestate contains essential minerals such as nitrogen, potassium, and phosphorus, making it a valuable organic fertilizer and soil conditioner (Pullen, 2015). The AD process improves the availability of these nutrients, enabling plants to absorb them more efficiently compared with raw manure (Balagurusamy & Chandel, 2021).

2.1.3. Biochemical Stages of Anaerobic Digestion Process

The conversion of complex organic matter into biogas is a multi-step process involving parallel and sequential biochemical reactions (Wang et al., 2020). This transformation requires the coordinated activity of diverse microbial groups, each performing specific tasks within the degradation network (García Diéguez, 2009; Murphy & Thamsiriroj, 2013). The process is traditionally divided into four interrelated stages: hydrolysis, acidogenesis, acetogenesis, and methanogenesis, which is displayed in figure 1 (J. Rajesh Banu & R. Yukesh Kannah, 2019).

Hydrolysis: In the first stage, hydrolytic bacteria release extracellular enzymes that decompose water-insoluble polymers, including carbohydrates, proteins, and lipids, into soluble monomers such as sugars, amino acids, and fatty acids. Hydrolysis is commonly regarded as the rate-limiting step in the AD process (J. Rajesh Banu & R. Yukesh Kannah, 2019).

Acidogenesis: during acidogenesis, which is considered the fastest phase of the process, fermentative bacteria transform the soluble monomers into volatile fatty acids (VFAs), alcohols,

H₂, and CO₂ (Ijäs, 2022). An excessive accumulation of VFAs may cause reactor acidification and can ultimately lead to process instability or failure (Nguyen, 2014).

Acetogenesis: During the third stage, acetogenic bacteria convert the intermediate products formed during acidogenesis, including longer-chain fatty acids and alcohols, into acetic acid, H₂, and CO₂ (Adewale Aromolaran, 2021).

Methanogenesis: The final and most sensitive stage is performed by methanogenic archaea, which convert acetic acid and H₂ into CH₄ and CO₂. Approximately 70% of the CH₄ produced during this phase originates from the conversion of acetate through the acetoclastic pathway (Sanjay Shah, 2016).

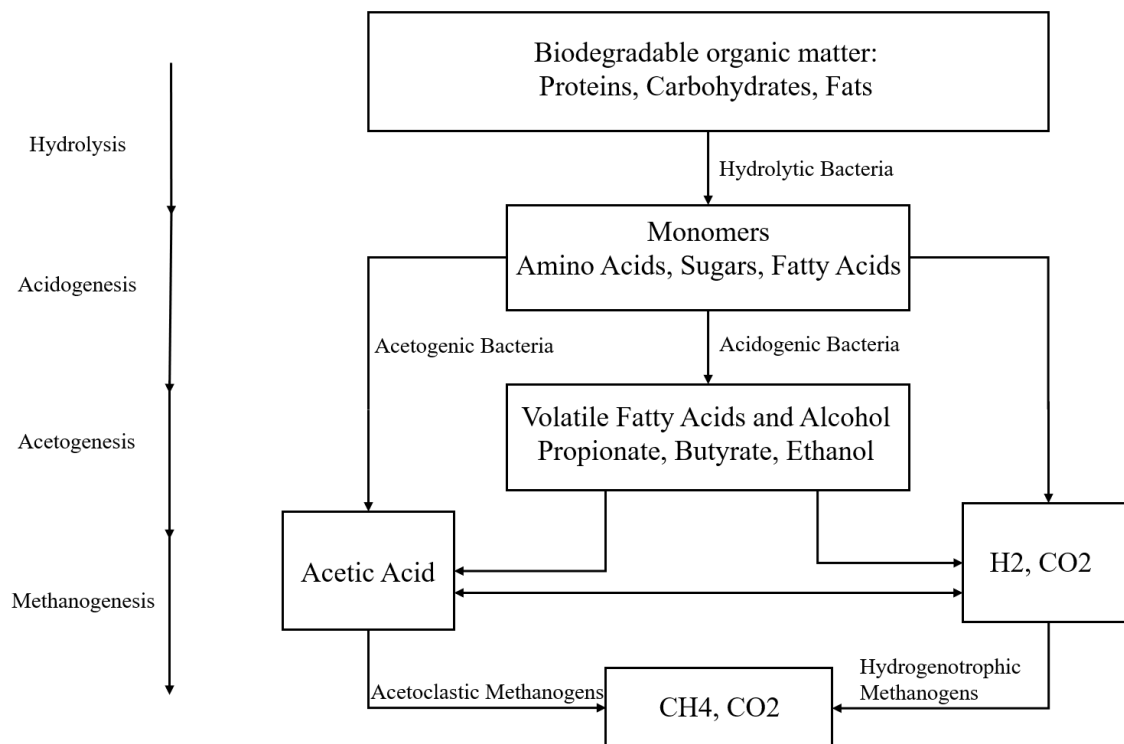


Figure 1. AD main stages adapted from (Aslanzadeh, 2014; Dizaji, 2023)

2.2. Gas-Liquid Mass Transfer in Anaerobic Systems

2.2.1. Gas-Liquid Mass Transfer

Gas-liquid mass transfer is a physicochemical phenomenon that involves the exchange of gaseous components between the gas and liquid phases within an anaerobic digester (Nguyen, 2014). In AD systems, this process includes both the transfer of generated gases such as CH₄, CO₂ and H₂ from the liquid phase to the headspace, and the transfer of injected gases, into the liquid phase for biological conversion (D. A. Poggio, 2015; Sanjay Shah, 2016).

$$r = k_L a (C^* - C) \quad (2.1)$$

Here the symbol r is the mass transfer rate, $k_L a$ is the volumetric mass transfer coefficient, C^* represents the equilibrium concentration, and C is the actual dissolved concentration of the gas in the liquid phase.

Based on Henry's law:

$$C^* = H \times p \quad (2.2)$$

Where H is Henry's constant and P is partial pressure of the gas

The effectiveness of this transfer is important because the limited solubility of certain gases, particularly H₂ and carbon monoxide (CO), can become a rate-limiting constraint for biological reactions (Acosta-Pavas et al., 2023). In addition, mass transfer rates affect key performance parameters such as biogas composition, effluent pH, and the possible inhibition of acetogenesis caused by elevated concentrations of dissolved H₂ (D. A. Poggio, 2015).

Henry's law forms the theoretical foundation for describing the equilibrium relationship between the partial pressure of a gas in the headspace and its dissolved concentration in the liquid phase (D. A. Poggio, 2015).

The volumetric mass transfer coefficient, $k_L a$, is commonly used to quantify the rate of gas-liquid transfer (Acosta-Pavas et al., 2022). This parameter represents the product of the liquid-phase mass transfer coefficient (k_L) and the specific interfacial area (a) available for mass transfer (Pauss et al., 1990).

2.2.2. Factors Influencing k_La

The volumetric mass transfer coefficient is strongly influenced by reactor hydrodynamics and the physical properties of the medium (D. A. Poggio, 2015). One of the most effective approaches to increase k_La is reducing the size of gas bubbles introduced into the reactor. Smaller bubbles provide a significantly larger gas-liquid interfacial area for the same volume of injected gas, which enhances the contact surface between the phases and improves mass transfer efficiency (Hornung, 2014). For this reason, porous spargers are commonly used in anaerobic reactors to generate fine bubbles, thereby improving gas dispersion and promoting more effective gas dissolution within the liquid phase (Acosta-Pavas et al., 2023; Dizaji, 2023).

Mixing intensity is another important factor that directly affects mass transfer rates (Singh et al., 2020). Increased agitation enhances turbulence within the reactor, which promotes the breakup of gas bubbles and improves their distribution throughout the liquid phase. This enhanced dispersion increases the gas-liquid contact area and reduces concentration gradients near the interface, ultimately leading to higher k_La values (Dizaji, 2023). In continuous stirred tank reactors (CSTRs), the volumetric mass transfer coefficient is closely related to the gassed power input per unit volume. The level of power input is determined by factors such as the impeller design, rotational speed, and reactor geometry. Higher power input generally increases turbulence and gas dispersion, which can significantly improve the rate of gas-liquid exchange (Acosta-Pavas et al., 2023). Gas recirculation represents another operational strategy used to enhance mass transfer. By recirculating the gas phase back into the reactor, the residence time of gases within the liquid phase is increased. This extended contact time allows additional opportunities for gas dissolution and biological conversion (Lecker et al., 2017). Gas recirculation can also contribute to improved mixing and turbulence within the reactor, further supporting higher mass transfer rates (Balagurusamy & Chandel, 2021).

2.3. Operating Conditions and Process Stability

AD performance is strongly shaped by operating conditions because the process relies on interconnected microbial and physicochemical phenomena (Acosta-Pavas et al., 2023). These conditions influence not only CH₄ production efficiency, but also the stability and operational

reliability of the digestion system (Yenigün & Demirel, 2013). For this reason, reviewing the principal operating conditions is necessary for understanding AD behavior under different process environments (Neuner et al., 2024). The following subsections therefore examine temperature regimes, pH and inhibition mechanisms, and anaerobic reactor configurations.

2.3.1. Temperature Regimes

AD is a temperature-dependent process in which the growth and metabolic activity of microbial communities are strongly affected by the thermal conditions within the reactor. Microorganisms involved in AD are generally classified into three temperature ranges: psychrophilic, mesophilic, and thermophilic (J. Rajesh Banu & R. Yukesh Kannah, 2019).

Psychrophilic digestion takes place at temperatures below approximately 20 °C to 25 °C. In this temperature range, microbial growth and metabolic activity occur at significantly lower rates. Nevertheless, psychrophilic treatment can be advantageous for treating moderate to low temperature wastewaters, as it may improve the overall energy balance of the treatment plant (Lettinga et al., 2001; Murphy & Thamsiriroj, 2013). Mesophilic digestion, typically operating between 25 °C and 40 °C, represents the most widely applied regime because of its relatively high process stability and reduced heating requirements (Neuner et al., 2024). Thermophilic digestion occurs within a temperature range of approximately 45 °C to 70 °C and is associated with faster reaction rates, increased biogas production, and more effective pathogen reduction compared with mesophilic systems. However, thermophilic operation is more sensitive to temperature variations and disturbances, requires greater energy input for heating, and exhibits higher susceptibility to ammonia inhibition (J. Rajesh Banu & R. Yukesh Kannah, 2019; Neuner et al., 2024).

Temperature significantly influences digestion kinetics and microbial metabolic pathways. As it is shown in figure 2, metabolic and biochemical reaction rates increase with temperature (Neuner et al., 2024).

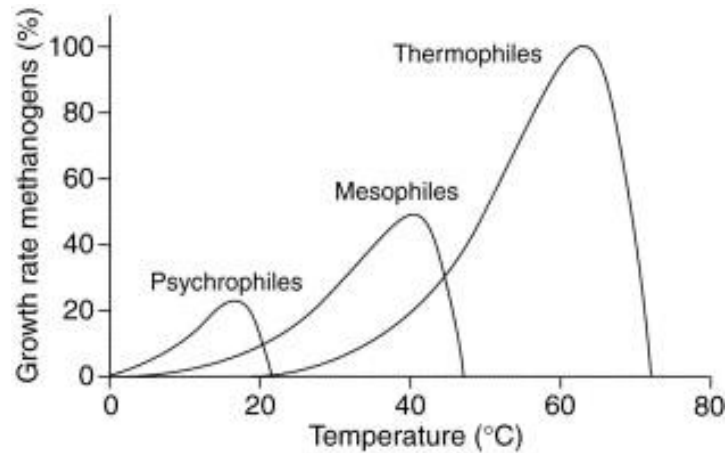


Figure 2. Growth rate of methanogen bacteria in different thermal conditions (Lettinga et al., 2001)

2.3.2. pH and Inhibition Mechanisms

The stability of the AD process is controlled by complex biochemical interactions and several inhibition mechanisms. Among these, pH is a key parameter because methanogenic archaea typically function optimally within a narrow, near-neutral pH range of approximately 6.5 to 8.5 (Maria Crolla & Kennedy, 2017).

Ammonia inhibition is another important limitation, particularly when digesting substrates rich in proteins. Total ammonia nitrogen exists in the forms of ammonium ions and free ammonia, with free ammonia being a strong inhibitor that can penetrate microbial cell membranes (Dizaji, 2023). The concentration of free ammonia increases with both higher pH and elevated temperature, which makes thermophilic digestion systems more vulnerable to ammonia inhibition (Yenigün & Demirel, 2013; J. Rajesh Banu, Schnürer. A & Westerholm. M, 2019).

In addition, the availability of inorganic carbon can affect hydrogenotrophic methanogenesis, since dissolved CO₂ is required for this metabolic pathway (Bensmann et al., 2014).

2.3.3. Anaerobic Reactor Configurations

The design and configuration of anaerobic reactors depend on the characteristics of the feedstock and the desired operational performance of the system (Murphy & Thamsiriroj, 2013). The performance of these configurations is commonly assessed using key operational parameters, particularly the hydraulic retention time (HRT) and the organic loading rate (OLR) (Ling et al., 2024; Mo et al., 2023). HRT refers to the average time that the liquid substrate remains inside the

digester (Aslanzadeh, 2014). A sufficient HRT is essential to allow sufficient time for the microbial community to grow and efficiently degrade organic matter. When the retention time becomes shorter than the microbial doubling time, active biomass may be washed out of the reactor, which can ultimately result in process failure (Hornung, 2014; Maria Crolla & Kennedy, 2017).

In addition to that, the organic loading rate (OLR) represents the quantity of organic material introduced into the reactor per unit volume per day (Almeida Streitwieser, 2017). OLR is an important parameter for evaluating biological performance and determining the maximum sustainable loading capacity of a digester (Aslanzadeh, 2014). Although higher OLR values enable smaller reactor volumes and greater CH₄ productivity, excessive loading can lead to the accumulation of volatile fatty acids and subsequent acidification of the reactor (Balagurusamy & Chandel, 2021).

One of the most widely used configurations is the continuously stirred tank reactor (CSTR). CSTR systems provide uniform heat distribution and allow biogas bubbles to rise through the reactor (Lecker et al., 2017). The geometry and structural design of the reactor influence both capital costs and the effectiveness of contact between microorganisms and the substrate (Lecker et al., 2017). Mixing represents an essential operational element that ensures uniform substrate distribution, supports microbial activity, and prevents the accumulation or settling of solids (Neuner et al., 2024). However, the intensity of mixing must be carefully controlled. While mixing enhances mass transfer, excessive agitation can increase operational costs and may disrupt the sensitive syntrophic microbial associations present within granules or biofilms (Balagurusamy & Chandel, 2021).

2.4. Biomethanation Technologies

Biomethanation is increasingly recognized as a promising approach for enhancing CH₄ production and improving biogas quality within AD and renewable gas systems (Angelidaki et al., 2018; Luo & Angelidaki, 2013). Its relevance has grown further because it can be implemented in different technological forms and integrated with existing biogas infrastructure (Jiang et al., 2022). However, its practical performance depends on several technological factors, including process configuration, gas addition strategy, reactor design, and transfer constraints (Sastraatmaja et al., 2022).

2.4.1. Biomethanation

Biomethanation is a technology that employs methanogenic archaea to generate renewable CH₄ from CO₂ and H₂ (Jensen et al., 2021). Technology functions as a biogas upgrading method by biologically converting the CO₂ produced during AD into additional CH₄ (Angelidaki et al., 2018). Through this conversion, higher CH₄ yields can be achieved for a given amount of organic waste, while the overall costs associated with upgrading biogas to natural gas grid standards are reduced (Jensen et al., 2021). In addition, biomethanation contributes to a circular bioeconomy by capturing and reutilizing internally generated CO₂ (Angelidaki et al., 2018).

2.4.2. Hydrogenotrophic Pathway

H₂ injection is the most widely applied method for biomethanation, where H₂ is introduced into the anaerobic digester (Fu et al., 2021). When suitable operating conditions are maintained, H₂ injections can increase the CH₄ concentration in biogas to levels exceeding 90% (Ding et al., 2015; Sanjay Shah, 2016).



Several operational challenges are associated with H₂ injection. One major concern is the potential increase of reactor pH above 8.5 (Angelidaki et al., 2018). This occurs because the biological consumption of dissolved CO₂ and bicarbonate, which constitute the primary buffering components of the system, reduces the buffering capacity. In addition, elevated H₂ partial pressure can make the oxidation of intermediate compounds such as propionate and butyrate thermodynamically unfavorable (Lecker et al., 2017). Consequently, volatile fatty acids (VFAs) may accumulate, which can destabilize the system and potentially result in process failure if the microbial community is not gradually adapted through strategies such as controlled loading or bioaugmentation (Fu et al., 2021).

2.4.3. Carbon Monoxide Injection Strategies

CO injection is typically implemented by syngas, which is a mixture of CO, H₂, and CO₂ produced from the thermochemical gasification of biomass (Sanjay Shah, 2016). The biological conversion of CO into CH₄ involves several metabolic pathways. CO may first be converted indirectly into

H₂ through the water-gas shift reaction, a process known as carboxydophilic hydrogenogenesis, or it can be transformed into acetate through CO-acetogenesis or CO-homoacetogenesis. The acetate produced through these pathways can be utilized by acetoclastic methanogens for CH₄ formation (Acosta-Pavas et al., 2023; Sanjay Shah, 2016).



A major limitation of CO injections is the toxicity to various methanogenic and acetogenic microorganisms, which can inhibit microbial activity even at relatively low concentrations (Sanjay Shah, 2016). Moreover, CO has very low solubility in water, which restricts its availability to microorganisms and causes gas-mass transfer to become the primary limiting factor in the process (Shah et al., 2017). To address this limitation, increasing reactor mixing intensity is commonly applied to enhance CO mass transfer and improve overall conversion efficiency (Dizaji, 2023).

2.4.4. In-situ and Ex-situ Biomethanation

In-situ biomethanation involves injecting H₂ directly into an operating anaerobic digester, where it reacts with the CO₂ produced internally during digestion (Angelidaki et al., 2018). One of the main advantages of this approach is the relatively low capital cost, as it makes use of existing reactors and infrastructure (Acosta-Pavas et al., 2023). However, the in-situ process can be negatively affected by high gas concentrations that may inhibit the early stages of AD, and the CH₄ production rate is typically lower compared with ex-situ systems (Lecker et al., 2017).

Ex-situ biomethanation, in contrast, is carried out in a separate reactor that contains enriched hydrogenotrophic microbial cultures (Acosta-Pavas et al., 2023). This configuration protects the stability of the primary AD process and provides greater operational flexibility, as external CO₂ sources can also be utilized (Angelidaki et al., 2018). Although ex-situ systems can handle larger volumes of substrate gases and achieve shorter gas retention times, they require higher capital investment due to the need for additional reactor units (Ijäs, 2022). Figure 3 explains difference between and in-situ and an ex-situ biomethanation in terms of steps.

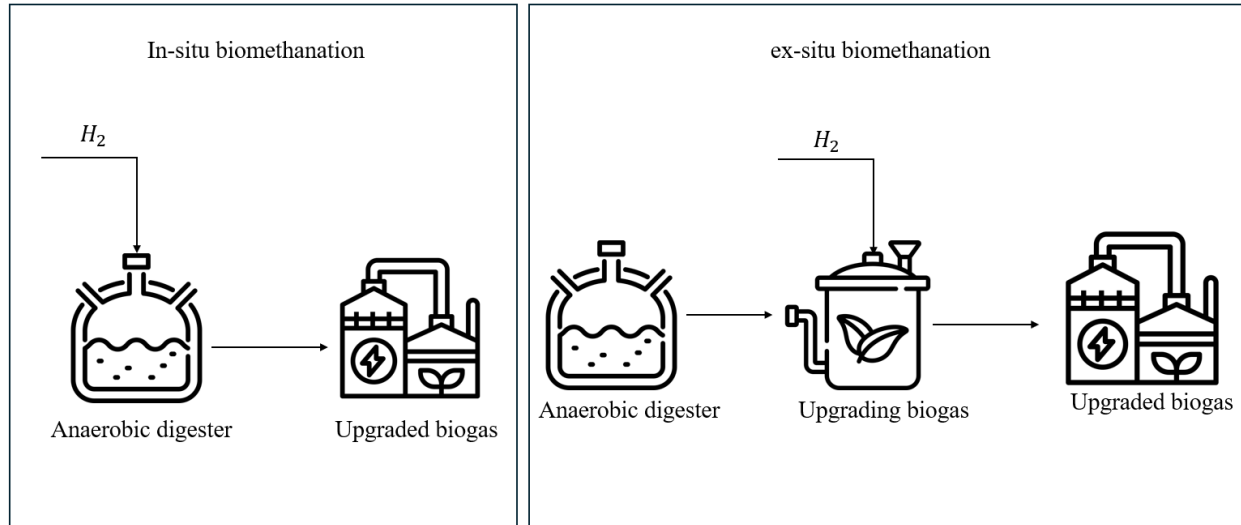


Figure 3. Schematic comparison of in-situ and ex-situ biological biogas upgrading adapted from (Antukh et al., 2022)

2.4.5. Mass Transfer Limitations in Biomethanation

Gas-liquid mass transfer is frequently identified as the rate-limiting bottleneck for the entire biomethanation process (Jensen et al., 2021). The relatively low solubility of H_2 in water represents a major limitation for its availability to microorganisms during biomethanation. Under standard temperature and pressure conditions, H_2 dissolves in water at much lower concentrations than CO_2 . In many cases, the biological consumption capacity of microorganisms exceeds the rate at which H_2 can dissolve into the liquid phase (Angelidaki et al., 2018). Consequently, CH_4 production is governed more by the efficiency of the gas-liquid mass transfer system than by the metabolic capability of the microbial community (Fu et al., 2021).

2.5. Mathematical Modelling of Anaerobic Digestion

Mathematical modelling plays an important role in the analysis of AD because it provides a systematic way to describe the complex interactions governing process performance. Such models are widely used to support simulation, interpretation, optimization, and control of digestion systems under different operating conditions. Over time, structured modelling frameworks have become central to AD research and have enabled more consistent plant-scale and benchmark-based studies.

2.5.1. The ADM1 Framework

The Anaerobic Digestion Model No. 1 (ADM1) was developed by a task group of the International Water Association (IWA) to provide a generalized and standardized framework for modelling the AD process (Batstone et al., 2002). The model is structured to integrate various biochemical and physicochemical processes involved in the conversion of organic matter into CH_4 (Sanjay Shah, 2016). In addition to these biological reactions, the framework incorporates physicochemical equations that describe non-biological processes, including ion association and dissociation as well as gas-liquid mass transfer. To represent these interconnected processes, the model includes a complex system of dynamic state variables governed by differential and algebraic equations (Batstone et al., 2002).

2.5.2. Benchmark Simulation Model No. 2 (BSM2)

The Benchmark Simulation Model No. 2 (BSM2) is a standardized plant-wide simulation framework developed to assess and compare control strategies for wastewater treatment plants (WWTPs) (Alex et al., 2018). Earlier benchmark models focused primarily on localized activated sludge processes, whereas BSM2 represents the entire treatment plant configuration, including wastewater pre-treatment and the full sludge treatment line (Rosen & Jeppsson, 2006). The model layout consists of multiple activated sludge reactors, primary and secondary clarifiers, thickeners, an anaerobic digester, and a dewatering unit. This integrated configuration enables the evaluation of supervisory control strategies that account for the interactions among individual treatment units, thereby reducing the risk of sub-optimization within the system (Rosen & Jeppsson, 2006).

2.5.3. Coupling ADM1 and BSM2

The integration of ADM1 with BSM2 is achieved by incorporating the AD process as the main sludge treatment component within the plant-wide simulation framework (Alex et al., 2018). To facilitate the exchange of information between the activated sludge model (ASM) and the anaerobic digestion model (ADM1), dedicated interface models are employed. The ASM/ADM interface converts the state variables from the activated sludge process into a format compatible with ADM1, whereas the ADM/ASM interface translates the digester outputs so they can be reintroduced into the wastewater treatment process (Alex et al., 2018). These interfaces preserve elemental, mass, and charge balances by fractionating incoming chemical oxygen demand (COD)

into components such as carbohydrates, proteins, and lipids (Rosen & Jeppsson, 2006). Because the resulting system of equations is highly numerically stiff, the ADM1 implementation within BSM2 typically employs a Differential Algebraic Equation (DAE) formulation, in which fast-reacting variables such as pH and dissolved H_2 are solved algebraically. This approach significantly improves computational efficiency (Alex et al., 2018; Rosen & Jeppsson, 2006).

2.5.4. Limitations of ADM1 for Biomethanation

Despite being a highly advanced modelling framework, ADM1 has certain limitations in its application to biomethanation systems (Mo et al., 2023). A major limitation of the original ADM1 framework is the assumption of a fixed volumetric mass transfer coefficient ($k_L a$), which is commonly assigned a constant value. This simplification is not suitable for biomethanation processes because it does not account for the dynamic influence of gas injection and variations in reactor hydrodynamics on mass transfer rates (Acosta-Pavas et al., 2023). In addition, the original ADM1 structure does not explicitly include gas injection strategies involving syngas components such as H_2 and CO. These externally supplied gases are not represented as substrates within the standard biochemical reaction matrix, which requires researchers to introduce additional diffusive connections or structural modifications to simulate their biological uptake (Acosta-Pavas et al., 2023).

The calibration of ADM1 is also challenging due to the large number of model parameters. This high degree of parameterization makes it difficult to uniquely estimate parameter values using typical experimental data (Ling et al., 2024).

2.5.5. Extensions of ADM1 Toward Biomethanation

To address the limitations of the original framework, several extensions have been proposed by the research community to adapt ADM1 for advanced applications such as biomethanation. These modifications are typically grouped into three main categories: model structure, kinetics, and stoichiometry (Mo et al., 2023). Structural modifications generally include the introduction of additional state variables and biochemical reactions to represent intermediate compounds and microbial pathways that were not considered in the original generalized model but become important during the biological conversion of externally supplied gases (Mo et al., 2023).

Kinetic and stoichiometric modifications are also commonly implemented to more accurately describe microbial growth dynamics, substrate utilization yields, and specific inhibition mechanisms that arise under the operating conditions associated with biogas upgrading (Bertacchi et al., 2021).

These developments indicate that although the conventional ADM1 framework provides a strong modelling basis, it should be applied as a flexible and modular platform that can be adapted to properly represent the complex behavior of advanced anaerobic systems (Mo et al., 2023).

2.6. Data-Driven Modelling and Machine Learning in Anaerobic Digestion

Machine learning (ML) is a branch of computer science concerned with pattern recognition and computational learning theory. Rather than relying on predefined programming instructions, ML algorithms construct predictive models by learning hidden patterns from input data. These models are increasingly being applied in AD research because the biological and physicochemical mechanisms governing the process are highly complex, non-linear, and difficult to represent accurately using conventional modelling approaches (Ling et al., 2024). Traditional mechanistic models often encounter calibration challenges due to the large number of parameters involved and the need for detailed substrate characterization (De Clercq et al., 2020). In contrast, data-driven ML models offer a more flexible alternative, as they can identify latent relationships between input variables and outputs without requiring explicit prior knowledge of the underlying system mechanisms (Long et al., 2021; L. Wang et al., 2020). In addition, ML techniques can enhance real-time monitoring and control by enabling early detection of disturbances and supporting the maintenance of process stability (Ling et al., 2024).

2.6.1. Data-Driven Modelling Approaches and Algorithms

A range of data-driven methods has been employed in AD modelling to estimate biogas production and overall system performance (Ling et al., 2024). Artificial neural networks are among the most widely used approaches because they can represent relationships within complex systems and approximate non-linear mappings without the need for fundamental equations (Asadi et al., 2020).

Extreme Learning Machine (ELM) is a feed-forward neural network with a single hidden layer that is recognized for its rapid training capability and strong generalization performance, particularly when applied to large datasets (Ling et al., 2024). Tree-based ensemble algorithms, including random forest and gradient boosting methods such as Extreme Gradient Boosting (XGBoost), are also commonly used because of their capacity to capture complex non-linear interactions among variables (De Clercq et al., 2020). Random forest operates by constructing a large number of independent decision trees and aggregating their predictions to improve accuracy while reducing model correlation (L. Wang et al., 2020). Support vector machines (SVM) and support vector regression (SVR) project input data into higher-dimensional feature spaces to achieve linear separability, making them suitable for generating reliable solutions even when the available dataset is relatively limited (Dong & Chen, 2019).

2.6.2. Hybrid Machine Learning and Mechanistic Simulation

The integration of ML with mechanistic models such as the ADM1 represents a combined strategy for overcoming the limitations associated with individual modelling frameworks (Sadrimajd et al., 2021). Mechanistic models provide a structured theoretical foundation for understanding AD, but they often suffer from high computational stiffness and require extensive substrate characterization (De Clercq et al., 2020). In contrast, although ML offers considerable predictive power and flexibility, it is frequently criticized for its black-box nature, which provides limited transparency regarding the underlying biochemical mechanisms (Ling et al., 2024). Hybrid modelling helps address these limitations by using data-driven algorithms to predict parameters or to act as software sensors for variables that are difficult to measure directly (Jimenez et al., 2015; Ling et al., 2024).

This hybrid strategy is particularly relevant for the development of digital twins and advanced control systems in industrial biogas plants (Sadrimajd et al., 2021). By connecting ML models with existing supervisory control and data acquisition systems, plant operators can identify abnormalities at an early stage and evaluate operational scenarios with reduced risk (Ling et al., 2024). Python-based implementations such as PyADM1 further support this integration by linking mechanistic models with commonly used ML libraries, thereby enabling more flexible and accessible bioprocess simulations (Sadrimajd et al., 2021). In addition, the combination of different algorithms, including hybrid genetic algorithm-artificial neural network models, has been

shown to achieve higher accuracy than conventional modelling methods used in isolation (Ling et al., 2024).

2.6.3. Research Gaps and Thesis Objectives

Despite the considerable body of research on AD modelling, several important gaps remain that limit the practical effectiveness of existing tools, particularly in the context of biological biogas upgrading (Mo et al., 2023). Although extensions for biomethanation have been proposed, they commonly assume a constant mass transfer coefficient with limited operational condition, which does not adequately represent the rate-limiting nature of H_2 transfer in upgraded systems. Second, many ML studies in anaerobic digestion are based on relatively limited experimental datasets, while fewer studies have examined machine-learning prediction using long-duration dynamic process data that capture a broad range of operating states and process conditions. (De Clercq et al., 2020; D. Poggio et al., 2023). Furthermore, modelling approaches for biological biogas upgrading remain limited, particularly in representing unresolved mechanistic behavior and gas-liquid mass transfer constraints during H_2 assisted biomethanation (Mo et al., 2023; D. Poggio et al., 2023).

Moreover, due to complexity and the highly nonlinear nature of this process, ML approaches are increasingly explored as alternative or complementary modelling methods since data-driven ML models do not require prior knowledge of system kinetics and can handle complex systems with intricate relationships between input and output variables. Compared with conventional mechanistic models, ML algorithms can provide greater flexibility and faster responses to variations in process conditions (Acosta-Pavas et al., 2023; Ling et al., 2024).

Accordingly, this research seeks to address these gaps by developing a hybrid mechanistic data-driven framework that uses long-duration dynamic simulation data to investigate H_2 -assisted biomethanation, evaluate gas-liquid mass-transfer effects, and develop ML models for predicting CH_4 and biogas production across the investigated operating conditions.

3. Methodology

3.1. Research framework

This study adopted a hybrid mechanistic data-driven methodology to investigate the optimization of AD through in-situ HBM and the prediction of CH₄ production using ML. The overall framework consisted of two integrated components. First, a dynamic mechanistic model based on the ADM1 within the Benchmark Simulation Model No. 2 (BSM2) framework was implemented in Python and further extended to represent in-situ H₂ injection and associated gas-liquid transfer effects. Second, the simulation outputs generated under multiple operating scenarios were used to construct a dataset for machine-learning-based prediction of CH₄ production. This hybrid strategy was selected because ADM1 provides a physically and biologically interpretable description of AD, while ML offers strong capability for nonlinear prediction once sufficient process data are available.

The methodological workflow of the study comprised the following stages: (i) selection of the ADM1-BSM2 framework as the mechanistic basis; (ii) modification of the original model to account for H₂ injection, scenario-based gas-liquid mass transfer, and inorganic carbon limitation of hydrogenotrophic methanation; (iii) execution of long-term dynamic simulations under different H₂ injection and mass-transfer scenarios; (iv) verification of the modified model using implementation literature and biomethanation studies; (v) extraction and evaluation of simulation outputs; and (vi) development and evaluation of ML models for predicting CH₄ production, expressed as CH₄ flow rate. This workflow was designed to address an important gap identified in recent ADM1 and biomethanation literature, namely the need for extended AD models that better represent H₂ injection and gas-transfer limitations. This integrated framework was designed to support both mechanistic interpretation of the biomethanation process and the development of predictive tools for process performance assessment.

3.2. ADM1/BSM2 Simulation Framework

The simulation framework used in this study was based on ADM1 implemented according to the BSM2 structure and executed in Python using PyADM1 as the computational starting point. ADM1 was selected because it provides the state and process structure needed to represent AD,

while the BSM2 implementation is specifically designed for dynamic influent operation, 15-minute stepping, and long-duration simulations under changing input conditions (Rosen & Jeppsson, 2006).

A BSM2-oriented implementation was adopted because the AD model is numerically stiff, especially when dynamic inputs are used. For this reason, the present work followed the common DAE-style formulation in which the main process states are integrated over each time step and the fastest equilibrium-related states are then solved separately. In practice, this approach is consistent with both the BSM2 ADM1 implementation and PyADM1, where the model advances stepwise and then uses a dedicated routine to calculate the remaining algebraic states (Sadrimajd et al., 2021).

PyADM1 was selected as the modelling platform because it is an open Python implementation of ADM1 that is directly aligned with the BSM2 benchmark structure and supports both ODE and DAE formulations (Liu et al., 2025; Sadrimajd et al., 2021). This made it suitable for the present thesis, where the base model needed to be modified for H_2 injection, scenario-based gas transfer, and additional biomethanation related effects. Its use is also justified by recent work that employed PyADM1 in a BSM2 layout to generate AD high resolution data intervals for CH_4 flow prediction, showing that PyADM1 can serve as a process simulator and as a practical data-generation platform for machine-learning studies.

3.3. Modified ADM1 model for in-situ HBM

The base ADM1 framework was extended in this study to represent in-situ HBM under scenario-based operating conditions. The modifications were introduced to capture the main phenomena associated with H_2 assisted CH_4 enhancement, namely external H_2 addition, gas-liquid transfer limitations, and carbon availability effects on hydrogenotrophic methanation. This extension strategy is consistent with recent ADM1 biomethanation studies, which report that conventional ADM1 must be adapted when external gas injection is included.

3.3.1. H₂ injection model

Pure H₂ was introduced into the digester as an external injection term. In the code, H₂ injection scenarios were first defined volumetrically in m³.d⁻¹, then converted using the ideal gas relation to *kmol H₂.d⁻¹*, and finally expressed in ADM1-consistent COD units, which represents the oxygen-equivalent electron-donating capacity of the substrate and is used in ADM1 as a common basis for organic and gaseous components, before being added to the H₂ balance. This implementation reflects the principle of in-situ biomethanation, where externally supplied H₂ reacts with the CO₂ already present in biogas through hydrogenotrophic methanogenesis rather than in a separate methanation reactor. This modelling choice was made because in-situ biomethanation studies commonly describe CH₄ enhancement as the result of direct H₂ addition to the AD system, where the available microbial community produces extra CH₄ and improves the quality of biogas by the chemical reaction of H₂ and CO₂.

3.3.2. Scenario-based gas-liquid mass transfer model

Gas-liquid mass transfer was modified in the present model because the default BSM2 ADM1 implementation commonly adopts a simplified treatment in which a constant volumetric mass-transfer coefficient is used for the main dissolved gases. In ADM1 implementations based on (Rosen & Jeppsson, 2006), the default value of $k_L a$ is often taken as approximately 200 d⁻¹, which is generally acceptable for well-mixed digesters and medium-to-high transfer conditions, but it does not explicitly distinguish between H₂, CH₄, and CO₂ transfer behavior. Moreover, previous ADM1 implementation work has noted that $k_L a$ depends on mixing, temperature, liquid properties, and, to a lesser extent, on the specific gas considered.

In the context of in-situ HBM, this simplified assumption was considered insufficient because H₂ transfer plays a critical role in determining H₂ availability in the liquid phase, while CH₄ and CO₂ also participate in dynamic gas exchange affecting biogas composition and carbon conversion. For this reason, the present study did not retain a single identical $k_L a$ value for all gases. Instead, $k_L a$ for H₂ was assigned directly according to selected assumed cases, and the corresponding $k_L a$ values for CO₂ and CH₄ were estimated from the H₂ value using scaling factors derived from the gas-transfer relationships discussed by (Pauss et al., 1990)., which shows that gas-specific transfer behavior in anaerobic systems is linked to diffusivity and that transfer limitations are especially important for H₂ and CH₄ (Pauss et al., 1990).

Table 2 summarizes the diffusivity coefficients and Henry's law constants adopted for the key gases involved in the present biomethanation model. Based on these diffusivity values, the transfer coefficients of CO₂ and CH₄ were calculated relative to the selected H₂ transfer coefficient, as expressed in Eqs. (3.1)-(3.3).

Table 2. Diffusivity coefficients and Henry's law constants of key gases involved in biomethanation processes, adapted from (Pauss et al., 1990)

Gas	Diffusivity (D) (cm ² .s ⁻¹ × 10 ⁻⁵)	Henry's Constant (K _H) (mol·L ⁻¹ ·Pa ⁻¹)
H ₂	4.65	7.40 × 10 ⁻⁹
CH ₄	1.57	1.12 × 10 ⁻⁸
CO ₂	1.98	2.70 × 10 ⁻⁷

$$\frac{k_L a_i}{k_L a_{H_2}} = \sqrt{\frac{D_i}{D_{H_2}}} \quad (3.1)$$

$$\frac{k_L a_{CO_2}}{k_L a_{H_2}} = \sqrt{\frac{D_{CO_2}}{D_{H_2}}} \quad (3.2)$$

$$\frac{k_L a_{CH_4}}{k_L a_{H_2}} = \sqrt{\frac{D_{CH_4}}{D_{H_2}}} \quad (3.3)$$

3.4. Reactor configuration, input files, and operating conditions

The simulations were carried out for a continuously stirred anaerobic digester represented within the ADM1/BSM2 framework. The reactor configuration followed the anaerobic digester layout,

with a liquid volume of 3400 m³ and a headspace volume of 300 m³, giving a total reactor volume of 3700 m³. The model was operated under mesophilic conditions at 35 °C (308.15 K), which is consistent with the base ADM1/BSM2 parameterization and with the implementation adopted in the Python code.

Two external input variables were used to define the simulation states. The first set of data contained the initial reactor conditions, including soluble, particulate, ionic, and gas-phase state variables at the beginning of the simulation. The second data set contained the time-varying influent data, which were read sequentially during the dynamic simulation. At each simulation step, the influent file supplied the current feed composition and flow conditions, and these values were used to update the reactor input states before numerical integration. This stepwise use of dynamic influent is consistent with the BSM2 simulation philosophy, where plant units are not assumed to remain at steady state and are instead driven continuously by changing inputs.

The initial-state file was used to define the starting condition of the anaerobic digester at the beginning of the simulation. It contained 38 ADM1 state variables representing the initial soluble, particulate, ionic, and gas-phase conditions in the reactor. These values were loaded once at the start of the simulation and used to construct the initial model state vector, which then served as the starting point for dynamic integration. Table 3 shows the initial parameters for the reactor.

Table 3. Initial ADM1 state variables and descriptions, used to initialize the digester model

ADM1 State Variable	Description	Initial Value
S_{su}	Monosaccharides	0.01239
S_{aa}	Amino acids	0.00554
S_{fa}	Long-chain fatty acids	0.10741
S_{va}	Valerate	0.01233
S_{bu}	Butyrate	0.014
S_{pro}	Propionate	0.01758
S_{ac}	Acetate	0.08932
S_{H2}	Dissolved hydrogen	2.51E-07
S_{CH4}	Dissolved methane	0.05549
S_{IC}	Inorganic carbon	0.09515
S_{IN}	Inorganic nitrogen	0.09447
S_I	Soluble inerts	0.13087
X_{xc}	Composite material	0.10792
X_{ch}	Carbohydrates	0.02052
X_{pr}	Proteins	0.08422

X_{li}	Lipids	0.04363
X_{su}	Sugar-degrading biomass	0.31222
X_{aa}	Amino-acid-degrading biomass	0.93167
X_{fa}	LCFA-degrading biomass	0.33839
X_{c4}	Valerate and butyrate degrading biomass	0.33577
X_{pro}	Propionate-degrading biomass	0.10112
X_{ac}	Acetate-degrading biomass	0.67724
X_{H2}	Hydrogenotrophic biomass	0.28484
X_I	Particulate inerts	17.2162
S_{cation}	Cations	1.08E-47
S_{anion}	Anions	0.00521
S_{H-ion}	Hydrogen ions	5.46E-08
S_{va-ion}	Ionized valerate	0.01228
S_{bu-ion}	Ionized butyrate	0.01395
$S_{pro-ion}$	Ionized propionate	0.01751
S_{ac-ion}	Ionized acetate	0.08904
$S_{hco3-ion}$	Bicarbonate ions	0.08568
S_{CO2}	Dissolved carbon dioxide	0.00947
S_{nh3}	Free ammonia	0.00188
$S_{nh4-ion}$	Ammonium ions	0.09258
S_{gas-H2}	Hydrogen in gas phase	1.10E-05
$S_{gas-CH4}$	Methane in gas phase	1.6535
$S_{gas-CO2}$	Carbon dioxide in gas phase	0.01354

The influent description followed the ADM1 state representation. Accordingly, the input file contained soluble substrates and intermediates such as sugars, amino acids, long-chain fatty acids, volatile fatty acids, dissolved H_2 , CH_4 , inorganic carbon, inorganic nitrogen, and soluble inert, together with particulate states representing composites, carbohydrates, proteins, lipids, microbial groups, and particulate inert. Additional ionic variables, including cations and anions, were also included to preserve charge balance. This structure is consistent with the ADM1/BSM2 input definition, in which the influent to the digester is expressed directly in ADM1 terminology.

The dynamic influent data were applied sequentially during the simulation. At each time step, the influent state variables and the influent flow rate (Q) were updated from the influent data before solving the model equations. Therefore, the reactor was not simulated under steady feed conditions, but under time-varying influent composition and flow, which formed the baseline digestion behavior for all biomethanation scenarios. This is consistent with the dynamic influent philosophy adopted in BSM2-based simulations.

The influent composition was expressed directly in ADM1 state variables, including soluble substrates and intermediates such as S_{su} , S_{pro} , dissolved H_2 , dissolved CH_4 , inorganic carbon, and inorganic nitrogen. This characteristic was shown in table 4, the influent file also included particulate states such as, microbial biomass groups, and particulates, together with cation and anion concentrations. This structure allowed the feedstock behavior to be represented in a form fully compatible with the ADM1 state-space description.

Table 4. Average non-zero values of key variables in the dynamic influent file

Influent Variable	Description	Average of Non-Zero Values
S_{su}	Monosaccharides	0.01003
S_{aa}	Amino acids	0.03743
S_{IC}	Inorganic carbon	0.00805
S_{IN}	Inorganic nitrogen	0.00203
S_I	Soluble inerts	0.02851
X_{ch}	Particulate carbohydrates	3.86787
X_{pr}	Particulate proteins	15.92325
X_{li}	Particulate lipids	8.32658
X_I	Particulate inerts	17.50299
S_{anion}	Anions	0.00526
Q	Average Influent flow rate	183.50555
T	temperature	35

To summarize the dynamic influent used in the simulations, the average values of the non-zero influent variables were calculated from the full influent dataset. Variables that remained zero throughout the dataset were excluded from this summary, and the temperature was maintained at 35 degrees during the simulation.

3.4.1. Hydraulic and loading indicators

The HRT and OLR were derived from the dynamic influent data rather than prescribed as fixed operating conditions. the HRT varied with time according to:

$$HRT(t) = \frac{V_{liq}}{Q(t)} \quad (3.4)$$

$$V_{liq} = 3400 \text{ m}^3$$

$$\text{Average } Q = 183.51 \frac{\text{m}^3}{\text{d}}$$

$$HRT(avg) = \frac{3400}{183.51} \rightarrow HRT(avg) \approx 18.53 \text{ d}$$

Similarly, the OLR was calculated dynamically from the influent flow rate and the total influent COD concentration:

$$OLR(t) = \frac{Q(t) \times COD_{in}(t)}{V_{liq}} \quad (3.5)$$

Here, Q is the influent feed to the reactor, COD_{in} represents the sum of the COD-based influent ADM1 variables, and V_{liq} refers to the liquid-phase volume in the reactor.

$$OLR(avg) = \frac{183.51 \times 45.54}{3400} = 2.45 \frac{\text{kg COD}}{\text{m}^3 \cdot \text{d}}$$

These values are reported as average influent characteristics for summary purposes only. In the actual simulation model and corresponding output files, the full dynamic influent dataset was used, and the influent variables were updated at each time step to preserve the time-varying feeding behavior of the reactor.

3.4.2. Inorganic carbon limitation of hydrogen uptake

To improve the representation of in-situ HBM, an additional inorganic carbon limitation term was introduced into the hydrogenotrophic methanogenesis rate. This modification was adopted from the model extension proposed by (Bensmann et al, 2014), who noted that the standard ADM1 structure does not explicitly limit H_2 conversion when dissolved inorganic carbon becomes depleted, even though hydrogenotrophic CH_4 formation requires CO_2 as a reactant according to chemical reaction between H_2 and CO_2 .

$$I_{IC,lim} = \frac{1}{1 + 10^{-4} / S_{IC}} \quad (3.6)$$

where S_{IC} is the dissolved inorganic carbon concentration. This factor approaches 1 when inorganic carbon is sufficiently available, meaning that H_2 uptake is not restricted. However, when S_{IC} decreases to very low values, the factor becomes smaller and reduces the H_2 uptake rate. In this way, the model accounts for the fact that hydrogenotrophic methanation cannot proceed indefinitely in the absence of sufficient carbon sources. (Bensmann et al., 2014) selected the threshold value 10^{-4} so that normal AD operation would remain unaffected, while carbon limitation would become active only under strong H_2 driven depletion of inorganic carbon. H_2 conversion decreases when inorganic carbon becomes depleted.

3.5. Scenario design

The simulation was designed to evaluate the effect of H_2 addition and gas-liquid mass transfer intensity on the performance of the modified ADM1 biomethanation model. In this study, these two factors were selected as the explicit scenario variables because H_2 availability and transfer limitation are consistently identified as the main constraints in in-situ biomethanation. (Bensmann et al., 2014) showed that direct HBM within a biogas plant is bound by both a biological limit and a transfer limit, which makes H_2 feed rate and transfer capacity appropriate variables for scenario-based analysis.

3.5.1. Hydrogen injection scenarios

H_2 injection was represented as a constant gas-phase input to the reactor. A range of H_2 injection scenarios was considered to examine the response of CH_4 production and gas composition to increasing H_2 availability. The H_2 injection scenarios used in this work were labeled according to the imposed daily volumetric flow rate, from zero injection as baseline scenario to higher volumes such as $2500 \text{ m}^3.\text{d}^{-1}$. These scenarios were not intended to reproduce a single plant operating policy, but to provide a structured range for sensitivity analysis within the modified ADM1 framework. The purpose was to determine how CH_4 production changes when H_2 supply is progressively increased.

3.5.2. Gas-liquid mass-transfer scenarios

The H₂-specific volumetric mass-transfer coefficient was investigated using six sensitivity levels of 60, 140, 240, 360, 650, and 1000 d⁻¹. The conventional BSM2/ADM1 value of approximately 200 d⁻¹ was used as a reference rather than as a calibrated value for the present reactor. The selected scenarios were intentionally distributed both below and substantially above this reference value in order to examine the transition from strongly transfer-limited conditions to conditions in which H₂ transfer becomes progressively less restrictive. Accordingly, 60 and 140 d⁻¹ represent low transfer capacities, 240 and 360 d⁻¹ represent intermediate to enhanced transfer capacities, and 650 and 1000 d⁻¹ represent high and very high transfer capacities. The purpose of this range was to determine how increasing gas-liquid transfer capacity affects H₂ utilization, CH₄ production, gas quality, and process stability.

The selected $k_L a_{H_2}$ cases were defined as sensitivity scenarios and included values such as 60, 140, 240, 360, 650 and 1000 d⁻¹. These values were not assigned to specific measured operating conditions; rather, they were assumed to investigate how the model responds to different transfer capacities. The corresponding transfer coefficients for CO₂ and CH₄ were then calculated from the H₂ value using diffusivity-based relationships derived from (Pauss et al., 1990).

Table 5. Gas-specific $k_L a$ values were derived from the H₂ reference value using diffusivity-based scaling factors.

Label	$k_L a_{H_2}$	$k_L a_{CO_2}$	$k_L a_{CH_4}$
scenario 1	60	39	34.8
scenario 2	140	91	81.2
scenario 3	240	156	139.2
scenario 4	360	234	208.8
Scenario 5	650	422.5	377
scenario 6	1000	650	580

3.5.3. Scenario combination matrix

The overall simulation set was generated by combining the H₂ injection scenarios with the $k_L a$ scenarios. In this way, each simulation case represented one pair of conditions, while all other baseline digester characteristics, influent dynamics, and operating temperature were kept unchanged. This matrix-based design enabled the effect of H₂ supply and transfer intensity to be evaluated systematically within the same modelling environment. In fig 4 the overall flow of the process and combination matrix is introduced shortly.

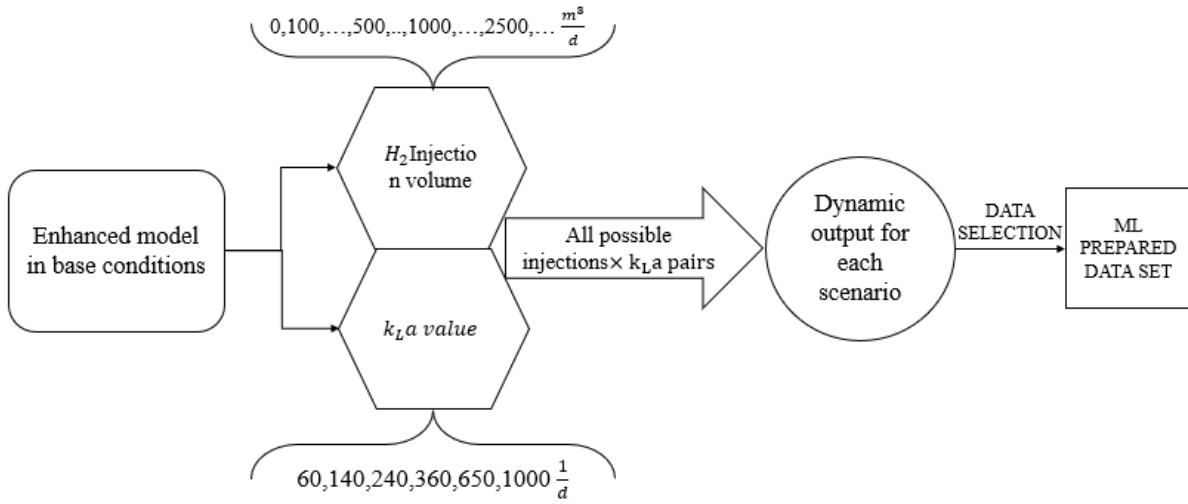


Figure 4. Scenario-generation framework for dynamic simulations and machine-learning dataset preparation.

3.6. Injection-efficiency and biomethanation performance indicators

To evaluate the performance of the in-situ HBM scenarios, a set of scenario level performance indicators was calculated from the simulation outputs. These indicators were derived from the average gas-phase and liquid-phase variables obtained for each simulation file after excluding the initial row, which represented the starting zero condition in the exported results.

3.6.1. H₂ utilization indicators

The amount of H₂ consumed was calculated as the difference between the imposed H₂ injection rate and the average H₂ flow rate remaining in the outlet gas.

$$H_{2,consumed} = H_{2,inj} - H_{2,out} \quad (3.7)$$

H₂ utilization in relative form, the H₂ conversion efficiency was calculated based on:

$$X_{H_2}(\%) = \frac{H_{2,inj} - H_{2,out}}{H_{2,inj}} \times 100 \quad (3.8)$$

H₂ slip represents the proportion of the injected H₂ that is not consumed and leaves the reactor in the outlet gas. A higher H₂ slip indicates less efficient H₂ utilization.

$$H_{2,slip}(\%) = \frac{H_{2,out}}{H_{2,inj}} \times 100 \quad (3.9)$$

3.6.2. Gas-upgrading indicator

To describe the CH₄ enrichment of the final gas mixture, a gas-quality indicator termed biomethanation extent was calculated from the average CH₄ and CO₂ gas flow rates.

$$CH_4 \text{ content } (\%) = \frac{q_{CH_4}}{(q_{CH_4} + q_{CO_2} + q_{H_2})} \times 100 \quad (3.10)$$

3.6.3. Role of the indicators in scenario evaluation

These indicators were used to evaluate the performance of each biomethanation scenario in a consistent and comparable way. For every combination of H₂ injection rate and mass transfer condition, the calculated metrics showed how effectively the reactor utilized the injected H₂.

The purpose of this evaluation was to identify the scenarios that performed efficiently and to exclude those that were not suitable for further analysis, such as scenarios with low H₂ conversion efficiency, high H₂ slip, and limited CH₄ improvement. These cases were removed from the next steps of the study, since they did not represent effective or stable biomethanation conditions. As a result, only the efficient scenarios were retained for the subsequent analysis, dataset preparation, and machine-learning stage.

3.7. Model verification

The modelling framework used in this study was based on the open-source PyADM1 code, which is a Python implementation of ADM1 derived from the BSM2 structure. The original PyADM1 code was previously benchmarked against the BSM2 ADM1 reference implementation, and the reported absolute errors for all state variables were below 10⁻⁵. Therefore, the core ADM1 structure adopted in this work was considered already verified, and the verification effort in this

thesis focused on the correctness of the modifications introduced for H₂ injection and gas transfer analysis.

The modifications introduced in this work included the implementation of a gas-liquid mass transfer formulation through gas-specific $k_L a$ terms and the addition of an inorganic carbon limitation term in the hydrogenotrophic methanogenesis rate. The latter follows the extension proposed by (Bensmann et al., 2014), in which the H₂ uptake rate is additionally limited by dissolved inorganic carbon availability. The explicit treatment of gas transfer is also physically justified, since (Pauss et al., 1990) showed that mass transfer limitations are relevant for poorly soluble gases such as CO₂ and CH₄ in anaerobic systems.

To verify the modified implementation, the code was operated under literature-based benchmark conditions reconstructed from the study of (Bensmann et al., 2014). For this purpose, the substrate characterization of maize silage was extracted from the selected literature sources and translated into the ADM1-compatible input format used by the model (based on table S₁ in appendix). The numerical values adopted for the feed characterization are presented separately. The benchmark case was reconstructed using the reactor and operating conditions applied in the present simulations, namely a liquid reactor volume of $V_{liq} = 1000 \text{ m}^3$, a gas volume of $V_{gas} = 300 \text{ m}^3$, and temperature of $T = 308.15 \text{ K}$. The dilution rate was set to $D = 0.01 \text{ d}^{-1}$, corresponding to a feed flow rate of $Q=10 \text{ m}^3 \cdot \text{d}^{-1}$.

The benchmark comparison was carried out at H₂ injection levels of 0, 0.5, 1.0, 1.5, 2.0, 2.5, and $3.0 \text{ m}^3_{STP} \text{ m}^{-3}_{liq} \text{ d}^{-1}$. Based on the selected comparison periods, the simulated methane production values were 1.046, 1.160, 1.275, 1.391, 1.510, 1.631, and $1.754 \text{ m}^3_{STP} \text{ m}^{-3}_{liq} \text{ d}^{-1}$, whereas the corresponding reference values were 1.13, 1.25, 1.37, 1.49, 1.61, 1.73, and $1.85 \text{ m}^3_{STP} \text{ m}^{-3}_{liq} \text{ d}^{-1}$. The associated deviations were 7.43%, 7.20%, 6.93%, 6.64%, 6.21%, 5.72%, and 5.19%, respectively. These results show that methane production followed the same increasing trend as the reference model and remained within an acceptable deviation range, consistently below 10% over the stable operating region.

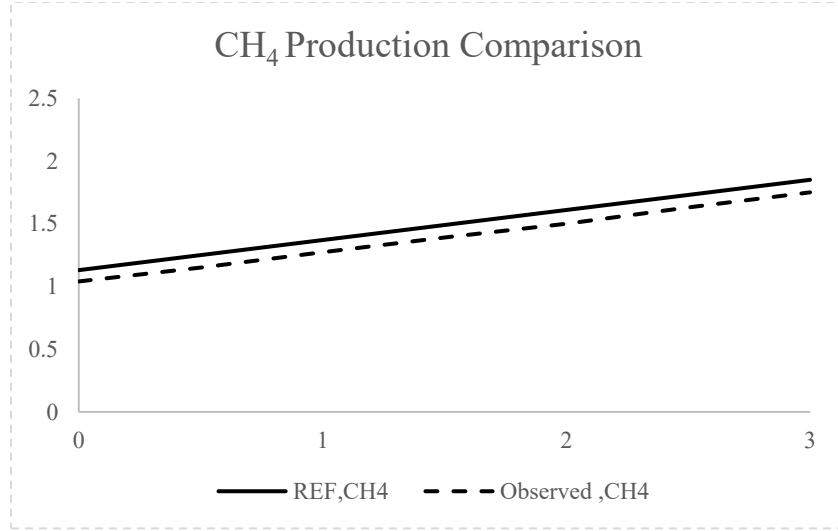


Figure 5 Comparison of CH4 production

A similarly strong agreement was observed for pH. The simulated pH values were 7.43, 7.49, 7.54, 7.5985, 7.64, 7.77, and 7.81, while the corresponding reference values were 7.42, 7.4, 7.50, 7.56, 7.62, 7.71, and 7.81. This confirms that the pH response was also reproduced well across the same injection range, with the model capturing the overall increasing trend and remaining very close to the published values. The highest pH deviation was observed at the $2.5 \text{ m}^3_{STP} \text{ m}^{-3}_{liq} \text{ d}^{-1}$ scenario, where the difference was 0.0615, corresponding to only 0.8%.

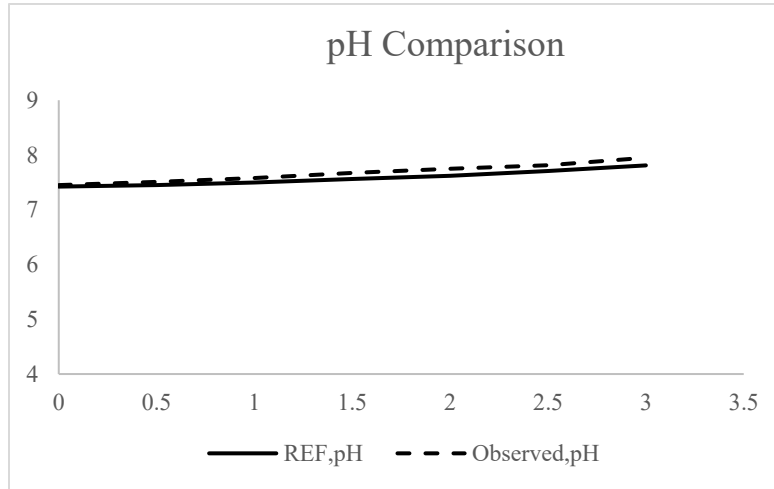


Figure 6 comparison of pH

In addition, other variables such as S_{IC} , S_{IN} , X_{H_2} , and S_{H_2} followed the same qualitative pattern as the reference study: S_{IC} and S_{IN} decreased with increasing H_2 injection, X_{H_2} increased, and

S_{H_2} remained low within the stable operating region and increased only when the system approached its limit.

At the same time, the benchmark comparison showed that this agreement was maintained only up to a certain H_2 injection level. In the reference study, the biological limit under ideal mass transfer occurs at approximately $4 m_{STP}^3 m_{liq}^{-3} d^{-1}$, beyond which methane production ceases, and the microbial community is washed out. In the present implementation, however, acceptable steady agreement could be maintained up to $3.0 m_{STP}^3 m_{liq}^{-3} d^{-1}$ ($3340 m^3 \cdot d^{-1}$). At $3.5 m_{STP}^3 m_{liq}^{-3} d^{-1}$ ($3897 m^3 \cdot d^{-1}$), the model no longer showed an acceptable steady operating window, indicating that collapse occurred earlier than in the reference case. Therefore, while the modified implementation captures the expected trends and operating behavior well in the feasible region, its upper stable injection threshold is more restrictive than that reported by (Bensmann et al., 2014).

3.8. Dataset collection and preparation

The dataset used for machine-learning model development was generated from the outputs of the modified ADM1/BSM2-PyADM1 simulation framework described in the previous sections. For each selected simulation case, the model produced a dynamic output file containing the time evolution of reactor states, gas-phase variables, pH, gas-liquid mass-transfer coefficients, and gas production variables. Among these outputs, the CH_4 flow rate (q_{CH_4}) was selected as the principal target variable for the machine-learning stage because it directly represents CH_4 production performance and is one of the most relevant operational indicators in AD studies. The use of ADM1-generated datasets for CH_4 prediction model development has also been reported in recent work, where dynamic simulation outputs were used as a data source to overcome the limited size of experimental datasets (Liu et al., 2025).

In the present study, each simulation was carried out over a period of 280 days, resulting in 26,881 records per simulation file. The high-resolution simulation outputs were retained in full during the modelling stage to preserve the transient process behavior associated with dynamic influent conditions, H_2 injection, and gas-transfer assumptions (Sadrimajd et al., 2021).

However, for the machine-learning stage, the raw 15-minute outputs were considered too detailed relative to the number of simulation cases and the final prediction objective. Therefore, the output time series was aggregated to daily resolution before model training. This reduced each simulation

file from 26,881 rows to about 280 daily observations, which made the dataset more compact and more suitable for regression-based learning while still preserving the long-term process behavior across the simulation.

The final dataset was assembled by combining the daily outputs from all simulation scenarios into a single machine-learning table. Each row of the final dataset therefore represented one daily observation under a specific pair of biomethanation conditions, defined by H₂ injection scenario and gas-liquid mass-transfer scenario. The final dataset included the selected predictive input features, the relevant scenario descriptors, and the target output (q_{CH_4}).

3.9. Feature selection and data preprocessing

The feature selection process was carried out to reduce the input space while still preserving the most important information for CH₄ prediction. Although including many variables can sometimes improve model accuracy, it can also make the model more complex and less interpretable. Therefore, the final feature set was selected by considering both the predictive value of the variables and their relevance to the AD process.

The initial set of available parameters included a broader range of process and state variables. However, not all of them were equally suitable for the final model. Some were highly model-specific, some were difficult to interpret clearly, and others were likely to provide overlapping information. For this reason, the final subset was designed to give a more balanced representation of the process while keeping the model simpler and easier to justify.

Based on this, the final input set was reduced to ten variables which are reported in table 6.

Table 6. Input variables for ML algorithms

Variable	Definition
pH	Reactor pH
Q	Influent flow rate to the digester
Hydrogen Injection	Hydrogen gas injection rate
$k_L a$	Gas-liquid mass transfer coefficient
S _{IC}	Soluble inorganic carbon concentration
S _{IN}	Soluble inorganic nitrogen concentration

X_{ch}	Particulate carbohydrate concentration
X_{pr}	Particulate protein concentration
X_{Li}	Particulate lipid concentration
S_{pro}	Soluble propionate concentration

These variables were selected because together they describe several important dimensions of AD. pH was included because it reflects the acidity or alkalinity of the digester environment and directly affects microbial activity. $k_L a$ was retained as an indicator of gas-liquid mass transfer conditions, which can influence the transfer of dissolved gases within the system. Q represents the feed flow rate and describes the loading condition of the reactor. H_2 injection was included because it represents an important operating input that can influence the biochemical pathway of CH_4 generation.

The variables S_{IC} and S_{IN} were selected to represent the inorganic chemical condition of the system. S_{IC} is the concentration of soluble inorganic carbon and is related to the carbonate system and buffering behavior of the digester. S_{IN} is the concentration of soluble inorganic nitrogen and reflects the nitrogen status of the system, which is important for microbial activity and process stability. Previous ADM1-based CH_4 prediction work also showed that variables such as S_I , S_{IC} , and S_{IN} could improve model performance when included in ML models (Liu et al., 2025).

The variables X_{ch} , X_{pr} , and X_{Li} were retained to describe the main biodegradable fractions of the substrate. X_{ch} represents particulate carbohydrates, X_{pr} represents particulate proteins, and X_{Li} represents particulate lipids. These three fractions were included because substrate composition plays an important role in AD kinetics, hydrolysis behavior, and CH_4 production potential. Keeping these variables allows the model to account for differences in the composition of the organic feedstock.

Finally, S_{pro} was retained as a representative intermediate process variable. S_{pro} is the concentration of propionate, which is commonly regarded as an indicator of process imbalance or stress in AD. Since propionate accumulation is often associated with instability in downstream conversion steps, its inclusion provides useful information about the internal biochemical condition of the reactor.

The selected inputs are interconnected through several biochemical, physicochemical, and hydraulic mechanisms represented in ADM1. A central interaction in the present biomethanation system occurs between H_2 injection and $k_L a$. H_2 injection determines the amount of H_2 supplied, whereas $k_L a$ affects its transfer from the gas to the liquid phase; consequently, their combined effect governs the availability of H_2 for hydrogenotrophic methanogenesis. Under insufficient mass-transfer conditions, increasing H_2 injection may result in H_2 accumulation or gas-phase slip, whereas greater transfer capacity can support higher H_2 utilization and CH_4 formation (Acosta-Pavas et al., 2023).

H_2 conversion also establishes relationships with S_{IC} and pH. Hydrogenotrophic methanogenesis consumes CO_2 , and increasing H_2 utilization can therefore reduce the available inorganic-carbon pool. Because S_{IC} participates in the carbonate acid-base equilibrium represented in ADM1, changes in inorganic carbon can alter buffering conditions and contribute to changes in pH (Rosen and Jeppsson, 2006). A related coupling exists between S_{IN} and pH through the NH_4^+/NH_3 equilibrium, while X_{pr} is linked to S_{IN} through protein hydrolysis, amino-acid degradation, and microbial nitrogen assimilation.

Relationships are also expected among X_{ch} , X_{pr} , X_{Li} , and S_{pro} . Although ADM1 represents the three particulate substrate fractions through distinct degradation pathways, they originate from common influent organic matter and may therefore co-vary with changes in feed composition or loading. Hydrolysis products of carbohydrates and proteins subsequently enter acidogenesis and can contribute to propionate formation. S_{pro} is additionally coupled to the H_2 state of the system because elevated dissolved H_2 can inhibit propionate oxidation and favour its accumulation.

Finally, Q can influence several state variables simultaneously through changes in hydraulic retention time, loading, dilution, and advective transport. Its relationships with individual state variables may therefore be complex rather than consistently positive or negative.

3.10. Machine-learning model development

In this study, four machine-learning approaches were selected for CH_4 production prediction: Random Forest (RF), SVR, ELM, and Evolutionary Polynomial Regression (EPR). ML offers a powerful data-driven alternative because these algorithms can identify hidden patterns and model intricate non-linear relationships between operational inputs and outputs without requiring explicit

knowledge of the underlying biochemical mechanisms (L. Wang et al., 2020; Yildirim & Ozkaya, 2023).

3.10.1. Random Forest

Random Forest (RF) is a supervised ensemble learning method that combines many decision trees to improve predictive accuracy and generalization. In (Breiman, 2001) original formulation, a random forest is a collection of tree predictors generated using independent random vectors, where the randomness is introduced through data resampling and random feature selection during tree growth. For regression problems, the final prediction is obtained by averaging the outputs of all trees in the ensemble (Biau & Scornet, 2016; Breiman, 2001).

$$\hat{q}_{CH_4}(x) = \frac{1}{M} \sum_{j=1}^M \hat{q}_{CH_4}^j(x) \quad (3.11)$$

The ML input vector $x_i = [pH_i, inj_i, Q_i, \dots]$ represents the features for sample i , where M is the total number of trees in the model and j denotes the index of an individual tree.

The RF training procedure consists of two main randomization steps. First, for each tree, a bootstrap sample is drawn from the original training dataset. Second, at each node, the split is searched only among a randomly selected subset of predictor variables rather than across all available features. The tree is then grown recursively until a predefined stopping condition is reached, such as a minimum number of samples in the terminal node. This combination of bootstrap aggregation and random feature selection reduces inter-tree correlation and improves model robustness (Breiman, 2001; Salman et al., 2024) A sample from one of trees used in the RF model was extracted and shown in figure 5.

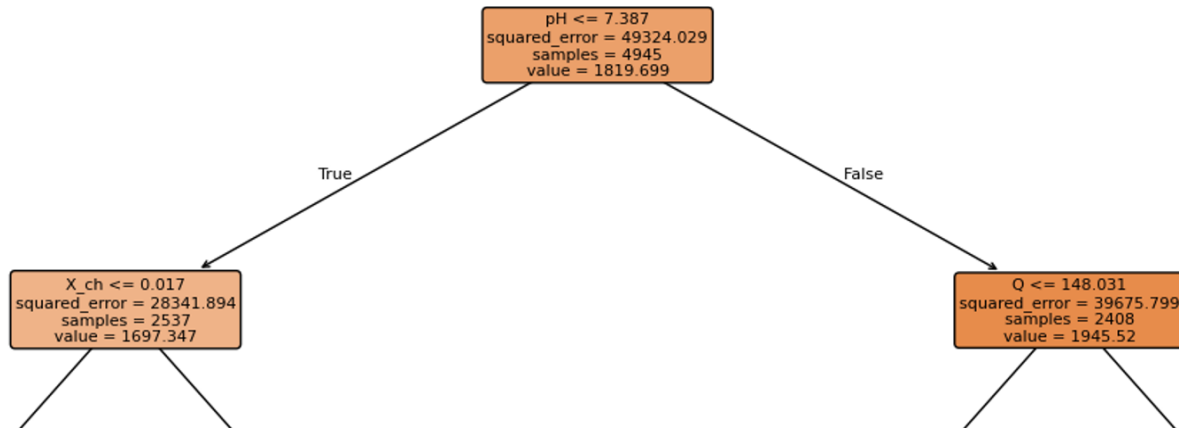


Figure 7. Representative decision tree structure within the Random Forest model

Hyperparameters of the Random Forest model

In ML, hyperparameters are settings defined before model training begins. Unlike model parameters, which are learned automatically from the data, hyperparameters control how the model is constructed and how the learning process is performed.

For the Random Forest (RF) model, the main hyperparameters are related to the number of trees, the structure of the trees, and the level of randomness introduced during training. The main hyperparameters include the number of trees, which determines how many decision trees are combined in the forest; increasing it usually improves prediction stability but also increases computation time.

The number of variables tested at each split max features function controls how many input variables, such as pH, $k_L a$, H₂ injection, and Q, are randomly considered when searching for the best split; this affects the balance between tree diversity and tree strength.

The minimum number of samples required to split a node and the minimum number of samples allowed in a leaf node control how easily the tree continues growing and therefore affect model complexity and sensitivity to local data patterns.

The maximum tree depth limits how deep each tree can grow, with deeper trees allowing more complex relationships to be captured but also increasing the risk of overfitting.

Moreover, the bootstrap option determines whether each tree is trained on a resampled version of the training dataset, which increases randomness among trees and improves generalization. These

hyperparameters control the size, complexity, and diversity of the forest, and directly affect prediction accuracy, and performance.

After testing different Random Forest hyperparameter combinations and comparing their prediction performance, the final model was implemented using the table 7 optimal settings:

Table 7. Random Forest Selected Hyperparameters

Hyper parameter	Final value
Number of estimators	300
Maximum features	4
Minimum samples required for split	3
Minimum samples required in leaf node	2
Maximum depth	12

RF was applied in regression mode to predict CH₄ related output from the simulation-derived dataset. The method was selected because of its robustness, its ability to model nonlinear relationships, and its suitability for high-dimensional process datasets with interacting variables.

3.10.2. Support Vector Regression

SVR is a supervised machine-learning method used for predicting continuous outputs. It extends the SVM framework from classification to regression by constructing a regression function that balances prediction accuracy and model complexity. The main advantage of SVR is that it can handle nonlinear relationships effectively through kernel functions, while maintaining good generalization performance, especially in high-dimensional problems.

For nonlinear data, SVR applies a kernel transformation that maps the original input variables into a higher-dimensional feature space, where linear regression can then be performed. In kernel form, the regression function is: (Zhang & O'Donnell, 2019)

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) K(x_i, x) + b \quad (3.12)$$

$K(x_i, x)$ represents the kernel similarity between the i -th training observation and a new input x , α_i and α_i^* are the coefficients determined during training, b is the bias or intercept term that shifts the regression function upward or downward, i denotes the index of each training sample, and l represents the total number of training samples used in the model.

Before model fitting, both input and output data were standardized. Input scaling was applied within the regression pipeline, while output scaling was performed as well. This preprocessing step was helpful because SVR is sensitive to variable magnitude, and scaling improves numerical stability and model performance.

Hyperparameters of the SVR model

For the SVR model, the main hyperparameters are related to the regression margin, the penalty assigned to prediction errors, and the kernel function used to describe nonlinear relationships. The main hyperparameters include the regularization parameter C , which controls the trade-off between minimizing training error and maintaining a smooth regression function; increasing C usually forces the model to fit the training data more closely, while smaller values improve generalization by allowing a smoother function.

The epsilon parameter (ε) defines the width of the ε -insensitive tube around the regression function, meaning that prediction errors smaller than this margin are ignored, while larger deviations are penalized; therefore, this parameter controls the tolerance of the model to small prediction errors. The kernel type determines how the relationship between the input variables and the target output is represented. In this study, the radial basis function (RBF) kernel was used because it is suitable for capturing nonlinear relationships. The tuned hyper parameters are mentioned in table 8.

Table 8. SVR Selected Hyperparameters

Hyper parameter	Final value
kernel	rbf
C	30
epsilon	0.01

3.10.3. Extreme Learning Machine

ELM is a learning algorithm developed for single-hidden-layer feedforward neural networks (SLFNs) (J. Wang et al., 2022). In ELM, the weights connecting the input layer to the hidden layer and the hidden-layer biases are assigned randomly and then kept fixed, while the output weights are determined analytically rather than by iterative gradient-based optimization. This structure makes ELM computationally efficient and avoids the repeated weight-updating procedure used in conventional backpropagation-based neural networks (Ding et al., 2015).

For a training set with N samples, the ELM model can be written as (Ding et al., 2015):

$$\sum_{i=1}^{\tilde{N}} \beta_i g(a_i \times x_j + b_i) = t_j, \quad (3.13)$$
$$j = 1, 2, \dots, N$$

where x_j is the input vector for sample j , a_i is the input weight vector connected to hidden neuron i , b_i is the bias of hidden neuron i , $g()$ is the hidden-layer activation function, β_i is the output weight associated with hidden neuron i , and t_j is the target output. In the present study, the target output is CH₄ related output, while the input vector contains the selected process variables used for machine-learning prediction.

The ELM formulation can be expressed in matrix form as (Ding et al., 2015):

$$H\beta = T \quad (3.14)$$

where H is the hidden-layer output matrix, β is the output-weight matrix, and T is the target matrix. After the hidden-layer output matrix is formed, the output weights are calculated analytically using the Moore–Penrose generalized inverse, which converts network training into a least-squares solution rather than an iterative optimization problem. Figure 6 indicates a sample of ELM structure for CH₄ predictions and input variables.

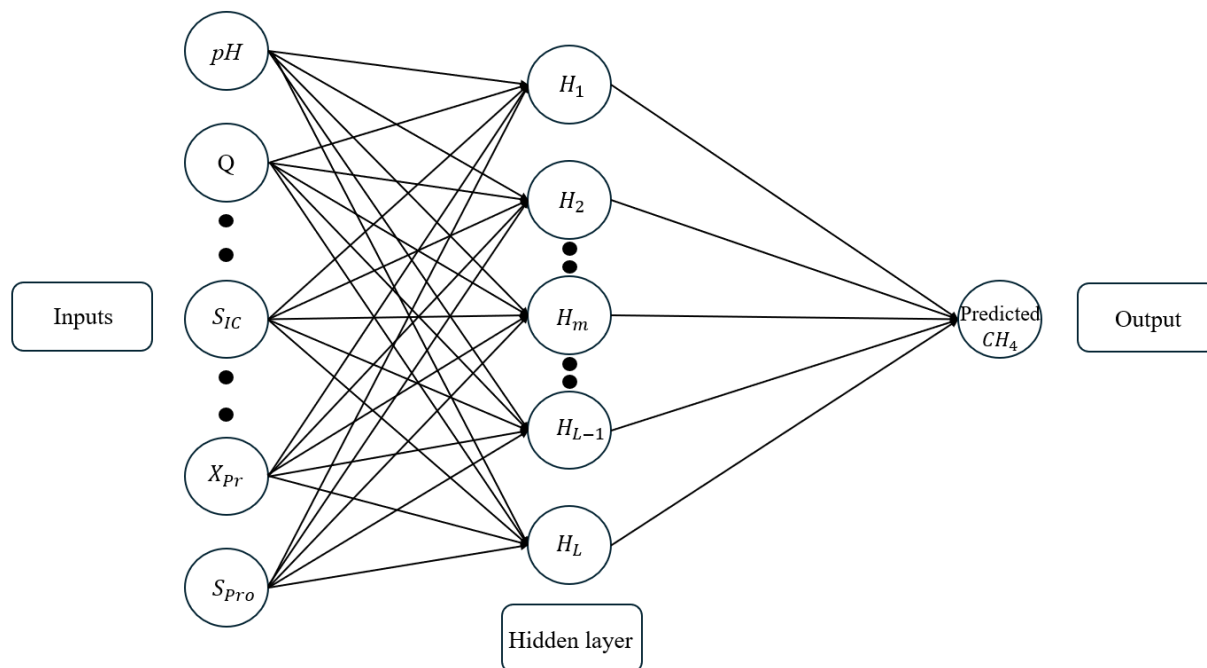


Figure 8. sample of the ELM structure for CH₄ prediction from selected AD variables adapted from (Hammad et al., 2020)

In the present study, ELM was implemented as a single-hidden-layer feedforward regression model using the selected AD and biomethanation variables as inputs and CH₄ production as the target output. Before model training, the input variables were standardized using parameters calculated only from the training dataset. This preprocessing step was particularly important for ELM because the weights connecting the input layer to the hidden layer are randomly generated and remain fixed during model training. Since the selected input variables have substantially different numerical ranges and units, using the unscaled data can cause variables with larger magnitudes to exert a disproportionate influence on the hidden-layer activation values.

Standardization placed the predictors on comparable numerical scales and improved the numerical representation of their combined effects in the hidden layer. This was particularly relevant with the sigmoid activation function, because large weighted inputs can move the activation function toward its saturated regions and reduce its sensitivity to changes in the predictors. After the hidden-layer output matrix was generated, the output weights were calculated analytically using the Moore-Penrose generalized inverse. The substantial improvement observed after preprocessing showed that appropriate input scaling was an important requirement for effective application of ELM to the present dataset.

Hyperparameters of the ELM model

The number of hidden neurons is an important ELM hyperparameter because it determines the representation capacity of the hidden layer. A network with too few hidden neurons may not adequately represent the nonlinear relationships between the process variables and CH₄ production, whereas unnecessarily large networks increase computational complexity without necessarily improving prediction accuracy. Therefore, after standardizing the input variables, a sensitivity analysis was conducted by evaluating different numbers of hidden neurons while maintaining the remaining ELM settings unchanged. Model performance was compared using R^2 and RMSE, and the final number of hidden neurons was selected based on the combination of predictive accuracy and performance stability. The analysis resulted in the selection of 250 hidden neurons for the final ELM model. The sigmoid function was retained as the hidden-layer activation function. The finalized selected hyper parameters are shown in table 9.

Table 9. ELM Selected Hyperparameters

Hyper parameter	Final value
hidden neurons	250
iteration number	10000
activation function	Sigmoid
orthogonal	true

3.10.4. Evolutionary Polynomial Regression (EPR)

Evolutionary Polynomial Regression is a hybrid regression technique that integrates the strengths of conventional numerical regression with evolutionary computing through genetic programming (Giustolisi & Savic, 2006). The methodology operates in two distinct stages: first, it employs a Genetic Algorithm (GA) to perform a global search for the most suitable symbolic polynomial structure that represents the system. In the second stage, it utilizes numerical least squares optimization to calculate the constant parameters and coefficients for the identified structure (Giustolisi & Savic, 2006; Shahin, 2016). EPR is particularly advantageous because it produces transparent mathematical formulas that allow for physical interpretation and scientific knowledge discovery, rather than operating purely as an opaque black box (Giustolisi & Savic, 2006).

The formulation of EPR is based on representing the output as a sum of polynomial-type terms composed of the candidate input variables raised to selected exponents.

$$Y = a_0 + \sum_{j=1}^m a_j \left(X_1^{ES(j,1)} X_2^{ES(j,2)} \dots X_k^{ES(j,k)} \right) \quad (3.15)$$

where Y is the predicted output, a_0 is the intercept term, a_j are numerical coefficients, $X_1, \dots, X_k$ are the input variables, and $ES(j, k)$ shows the exponent assigned to each variable in each model term.

A useful feature of EPR is that it searches within a family of candidate polynomial expressions rather than assuming a fixed model structure in advance (Giustolisi & Savic, 2006). This enables the method to capture nonlinear relationships while still producing a compact analytical formula. At the same time, EPR usually requires fewer fitted constants than many neural-network-based approaches, which helps reduce model complexity and makes the final expression more suitable for engineering interpretation (Marasco et al., 2021).

Another important aspect of EPR is the balance between model accuracy and model simplicity. Since the method can generate many possible equations, model selection is typically guided not only by goodness of fit but also by the complexity of the resulting formula. In this way, EPR aims to identify equations that are sufficiently accurate while remaining parsimonious and interpretable. This characteristic is particularly valuable in scientific and engineering studies, where the final model is expected to provide both prediction capability and insight into the role of the governing variables (Giustolisi & Savic, 2006; Marasco et al., 2021).

Evolutionary Polynomial Regression EPR Hyperparameters

The EPR model uses several hyperparameters to control the structure of the final equation and the way the genetic algorithm searches for the best model. The number of terms defines how many polynomial components can appear in the final equation. The candidate exponent vector determines the possible powers assigned to each input variable in the exponent matrix. The degree of pressure controls the balance between model accuracy and equation complexity.

In the genetic algorithm, the maximum number of iterations defines how many optimization cycles are performed, while the population size determines how many candidate equations are evaluated

in each cycle. The selection method determines how candidate equations are selected for reproduction; in this model, random selection was used. Crossover parameters control how candidate equations are combined, while the mutation parameters control how randomly the exponent matrix is modified during the search process. These hyperparameters influence the model accuracy, stability, complexity, and ability to explore different equation structures.

Table 10 illustrates the value selected for each hyper parameter in EPR model for prediction of biogas and CH₄.

Table 10. EPR Selected Hyperparameters

Hyper parameter	Final value
Number of terms	5
Candidate exponent vector	-3 to 3
Degree of pressure	0.35
Maximum iterations	200
Population size	300
Selection method	Random
Cross over percentage	0.8
Mutation rate	0.2

3.10.5. Data partitioning and robustness assessment

Although the simulation outputs were generated sequentially, the objective of the ML stage was not to forecast CH₄ at a future time from preceding observations, but to establish a state-based relationship between the selected process variables and CH₄ production over the investigated operating domain. For this reason, the dataset was treated as a collection of process-state observations generated under different H₂-injection and scenarios. A random 70/30 train-test partition was therefore applied separately within each scenario, ensuring that both the training and testing datasets contained observations from every investigated operating condition rather than restricting one dataset to a particular portion of the process trajectory. Similar non-chronological partitioning has been used in AD machine-learning studies focused on predicting biogas production from process variables. For example, Yildirim and Ozkaya (2023) randomly separated industrial-scale AD data into training and testing sets and subsequently used cross-validation for

model development, while Han et al. (2024) used a 70/30 training-test division for methane and biogas prediction from continuously collected anaerobic-fermentation data.

A limitation of a single random partition is that model performance can depend on the particular observations assigned to the training and testing subsets. Therefore, the 70/30 partitioning procedure was repeated 30 times using different random seeds, while preserving the same scenario-wise sampling structure for all four ML models. This follows the principle of repeated hold-out validation, in which multiple independent random partitions are used to reduce dependence on an unusually favorable or unfavorable split and to assess the stability of model performance. For CH₄ prediction, the repeated analysis produced mean testing R² values of 0.9968 ± 0.0003 for SVR, 0.9896 ± 0.0006 for ELM, 0.9887 ± 0.0006 for RF, and 0.9601 ± 0.0034 for EPR. The small variability across the 30 repetitions indicates that the comparative performance of the models was not dependent on a single random train-test realization.

3.10.6. Feature importance and sensitivity analysis

To improve the interpretability of the ML models, feature-importance and sensitivity analyses were conducted for RF, SVR, ELM, and EPR using the same testing dataset. Permutation Feature Importance (PFI) was used to quantify the contribution of each input variable to prediction accuracy. Each feature was randomly shuffled while the remaining inputs were unchanged, and the resulting decrease in R² was calculated as:

$$I_{PFI,j} = R_{base\ line}^2 - R_{permuted,j}^2 \quad (3.16)$$

The permutation was repeated 20 times for each feature, with larger reductions in R² indicating greater predictive importance. Because the absolute PFI magnitudes differed among models, the importance values were normalized within each model and averaged to obtain a consensus ranking across the four algorithms.

A one-at-a-time sensitivity analysis was also performed. Each input was varied across its training-data percentiles P_{10} , P_{25} , P_{50} , P_{75} , and P_{90} while the remaining variables were retained at their original values. The resulting change in mean predicted CH₄ was used to quantify model sensitivity. The response range was normalized within each model and combined to obtain a consensus sensitivity ranking.

These results were interpreted as model-response sensitivity rather than direct physical causality, since AD process variables can be interdependent and may not vary independently under actual operating conditions.

3.11. Model Evaluation Metrics

Model evaluation metrics were used to assess how well the ML models reproduced the target outputs and how reliably they captured the relationship between the input and output variables. Since model performance cannot be judged accurately from visual comparison alone, quantitative indicators were required to compare prediction quality in a consistent and objective manner.

In this study, five metrics were considered for evaluation including: mean squared error (MSE), mean absolute error (MAE), root mean square error (RMSE), normalized root mean square error (NRMSE), and R^2 .

Mean Squared Error (MSE)

MSE was used to measure the average squared difference between predicted and actual values. Because the errors are squared, this metric gives more weight to large deviations and is therefore useful for identifying models that produce significant prediction errors.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.17)$$

y_i = observed value

$\hat{y}_i$ = predicted value

n = number of observations

Mean Absolute Error (MAE)

MAE was used to determine the average absolute difference between predicted and actual values. Unlike MSE, it does not emphasize large errors more strongly than small ones. For this reason, MAE provides a direct and easily interpretable measure of the typical prediction error.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.18)$$

Root Mean Square Error (RMSE)

RMSE was used as another measure of prediction error, derived from the squared differences between predicted and actual values. It is commonly applied because it expresses error in the same unit as the target variable while still remaining sensitive to large deviations.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.19)$$

Normalized Root Mean Square Error (NRMSE)

NRMSE was used to express RMSE in normalized form relative to the scale of the observed data. This makes it easier to compare prediction performance across different variables or datasets, especially when their ranges or units are different.

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \quad (3.20)$$

Coefficient of Determination (R^2)

The coefficient of determination was used to evaluate how well the model explained the variation in the target variable. A higher R^2 indicates that the model predictions follow the overall trend of the actual data more closely and that a greater proportion of the variance is captured by the model.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (3.21)$$

Where $\bar{y}_i$ is the mean of observed values

4. Results and discussion

This chapter presents and discusses the findings obtained from the modified ADM1 simulations and the subsequent machine-learning analysis. The results are organized in two parts. The first part examines the performance of the AD system under baseline conditions and across the full matrix of H_2 injection rates and gas-liquid mass-transfer scenarios, with specific attention to CH_4 production, H_2 utilization, and liquid-phase stability indicators. The second part evaluates the predictive performance of four machine-learning algorithms trained on the simulation-derived dataset and compares their suitability for CH_4 and biogas flow rate prediction.

4.1. Overview of ADM1 simulation results

The ADM1 simulation results demonstrated that H_2 injection consistently increased CH_4 production across all investigated scenarios, but the efficiency and quality of that increase were strongly dependent on the gas-liquid mass transfer condition. In general, higher $k_L a$ values allowed the reactor to dissolve and biologically convert a greater fraction of the injected H_2 , yielding stronger CH_4 gains, lower H_2 slip, and improved outlet gas composition. In contrast, low $k_L a$ conditions limited the availability of dissolved H_2 to hydrogenotrophic methanogens, such that increasing the injection rate produced diminishing biomethanation returns and progressive deterioration in gas quality. Alongside these gas-phase responses, the simulations also captured systematic changes in reactor liquid chemistry: the biological consumption of dissolved CO_2 during hydrogenotrophic methanogenesis progressively reduced the buffering capacity of the system and elevated pH, with the magnitude of both effects increasing under higher transfer and injection conditions. These observations are discussed in detail in the following sections.

4.2. Reference performance of the baseline configuration

The baseline configuration was first evaluated under the no H_2 condition to establish a reference state for all subsequent comparisons. Under this condition, the reactor showed stable performance across all gas-liquid mass transfer scenarios, as expected, since H_2 transfer does not actively affect the system when no external H_2 is supplied. The average CH_4 flow rate (q_{CH_4}) remained within a narrow range, with an overall mean value of 1613.99, while the average H_2 flow rate in the outlet gas was negligible, with a mean value of 0.045.

The liquid-phase chemistry of the baseline condition also indicated stable operation. The average pH across all no-injection cases was 7.26, while the minimum and maximum pH values varied only slightly from 7.13-7.14 and 7.36-7.37, respectively. The narrow spread of these values across six different k_La conditions demonstrate that the reactor's carbonate buffering system remained fully intact in the absence of CO₂ consumption by biomethanation. These baselines are having a mean values of CH₄ flow rate of approximately 1614 m³.d⁻¹ and a stable pH near 7.26, serve as the reference points for evaluating the incremental effects of H₂ injection in the following sections.

4.3. Effect of hydrogen injection under different gas-liquid mass transfer scenarios

The effect of H₂ injection on reactor performance was evaluated across six response variables: CH₄ production, outlet H₂ flow rate, average pH, maximum pH, average soluble inorganic nitrogen (S_{IN}) and average soluble inorganic carbon (S_{IC}) concentration. Each variable captures a distinct aspect of the biomethanation response from the primary production output to the gas-phase utilization efficiency and the liquid-phase chemical stability of the system. The results for each variable are presented separately in the following subsections, with each analysis covering the full matrix of k_La conditions (60 to 1000 d⁻¹) and H₂ injection rates (0 to 2500 m³.d⁻¹). Taken together, these six indicators provide a comprehensive characterization of how the modified ADM1 model represented the coupled effects of H₂ supply and gas-liquid mass transfer on both the productive and stability dimensions of in-situ biomethanation.

4.3.1. CH₄ production

CH₄ production increased with H₂ injection under all investigated k_La scenarios, as shown in Figure 7. However, the magnitude of this increase was strongly dependent on the gas-liquid mass transfer condition and differed substantially between scenarios.

At the no-injection condition, CH₄ increment was nearly identical across all cases, confirming that transfer intensity had no effect on CH₄ formation in the absence of external H₂. Once injection was introduced, the CH₄ response curves diverged progressively, and the CH₄ response increasingly depended on the gas-liquid mass-transfer condition.

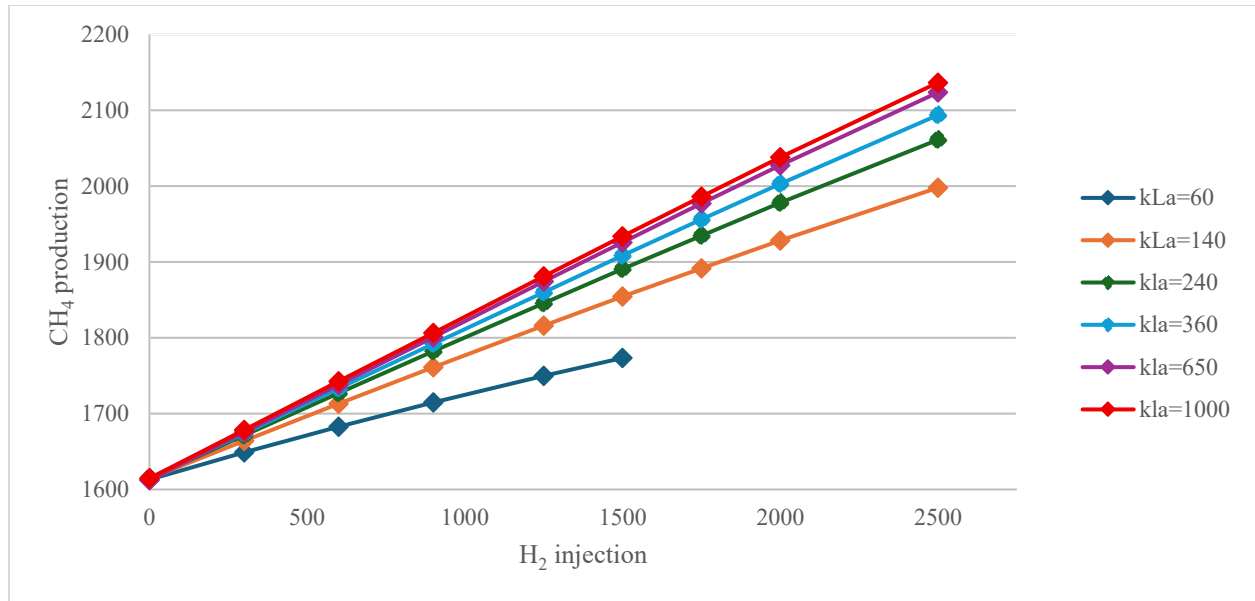


Figure 9. CH₄ production under different scenarios

As it is shown in figure 7 under the lowest transfer condition ($k_L a = 60 \text{ d}^{-1}$), CH₄ production increased from $1613 \text{ m}^3 \cdot \text{d}^{-1}$ at the baseline to $1773 \text{ m}^3 \cdot \text{d}^{-1}$ at the maximum injection rate of $1500 \text{ m}^3 \cdot \text{d}^{-1}$ a gain of approximately 10%. Under the highest transfer condition, the same baseline rose to $2137 \text{ m}^3 \cdot \text{d}^{-1}$ at $2500 \text{ m}^3 \cdot \text{d}^{-1}$ injection, representing an increase exceeding 32%. This difference illustrates the central role of mass transfer: higher $k_L a$ values enabled the reactor to dissolve and convert a greater proportion of the injected H₂ into additional CH₄, while low $k_L a$ conditions limited H₂ availability in the liquid phase and therefore constrained the biomethanation response.

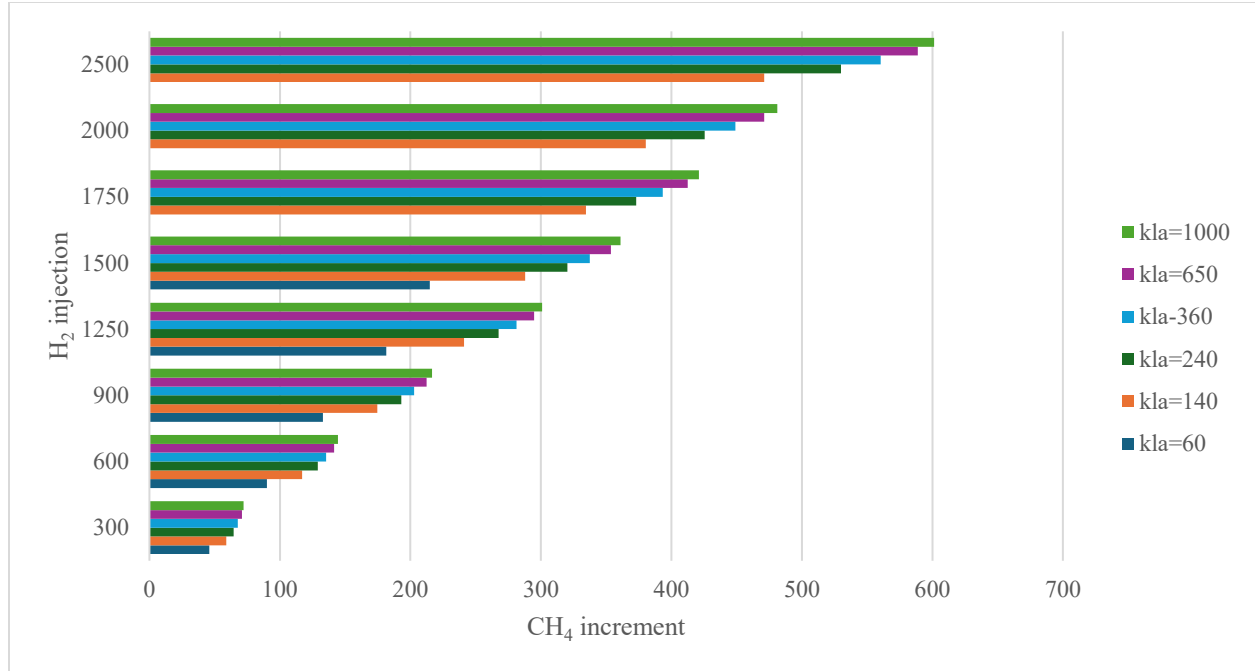


Figure 10. CH₄ increment in each scenario

Another important trend was shown in figure 8 it is observed that the difference between scenarios became more visible at higher injection levels. At low injection, the CH₄ values under different k_La conditions remained relatively close to each other. However, as H₂ injection increased, the separation between the curves widened, meaning that the benefit of improving k_La became increasingly important under stronger H₂ loading. This shows that mass-transfer limitations were not equally important across the full operating range; they became much more restrictive as more H₂ was supplied to the system.

The results also suggest that the effect of increasing k_La was strongest when moving from low to intermediate transfer conditions, while the improvement became less pronounced at the highest k_La values. Thus, CH₄ production continued to increase with k_La , but the marginal benefit gradually decreased at the upper end of the transfer range. This indicates that, once gas-liquid transfer became sufficiently strong, further increases in k_La still improved CH₄ production, but with a smaller incremental advantage than in the lower k_La region.

4.3.2. Outlet H₂ flow under different injection and k_La scenarios

The outlet H₂ results showed a clear increase with H₂ injection under all investigated k_La scenarios, but the rate of this increase depended strongly on the mass-transfer condition this

difference in rates is shown in figure 9. At the no-injection case, outlet H_2 was negligible in all scenarios, confirming that the baseline reactor operation was not associated with any meaningful H_2 loss. Once H_2 injection was introduced, however, the curves separated clearly and revealed a strong dependence on $k_L a$.

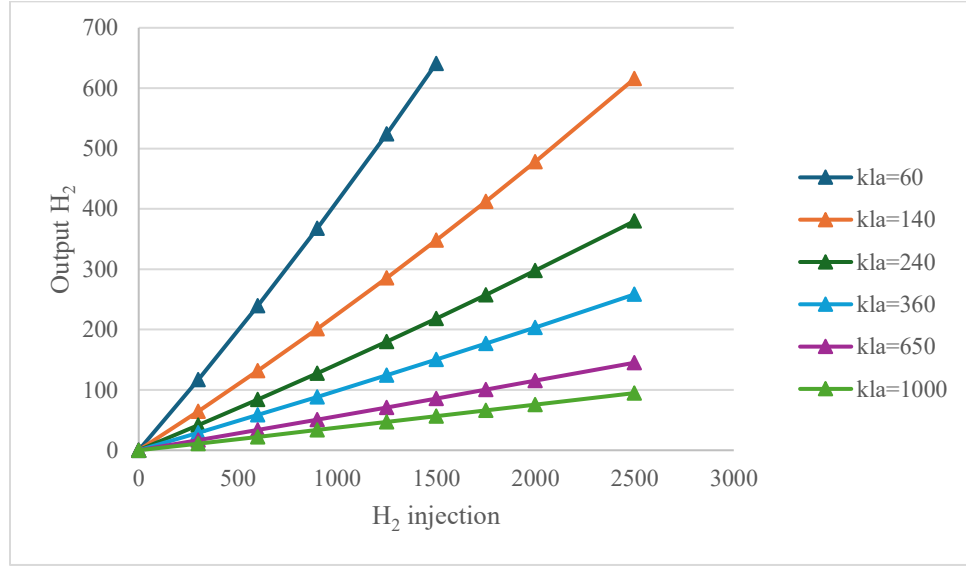


Figure 11. Average H_2 output in different scenarios

The low $k_L a$ scenarios, particularly $k_L a=60 \text{ d}^{-1}$ and $k_L a=140 \text{ d}^{-1}$, exhibited the steepest increase in outlet H_2 with rising injection rate. At the maximum injection level under $k_L a=60 \text{ d}^{-1}$, H_2 slip exceeded 40% of the injected volume, indicating that a large fraction of the supplied hydrogen was leaving the reactor unconverted. This behavior is a direct consequence of H_2 low aqueous solubility: when the gas-phase supply rate exceeds the dissolution capacity, dissolved H_2 concentrations in the liquid phase remain insufficient for complete hydrogenotrophic conversion, and the surplus accumulates in the headspace before exiting with the biogas stream.

Under the high $k_L a$ scenarios ($k_L a=650 \text{ d}^{-1}$ and $k_L a=1000 \text{ d}^{-1}$), outlet H_2 remained low across the full injection range. Even at the maximum injection level of $2500 \text{ m}^3 \cdot \text{d}^{-1}$, H_2 slip was below 4% under $k_L a=1000 \text{ d}^{-1}$, confirming that enhanced dissolution rates kept dissolved H_2 within the range that hydrogenotrophic methanogens could consume continuously

Another visible result is that the difference between scenarios became wider as injections increased. At low injection, the outlet H_2 values were still relatively close, but at higher injection rates the curves diverged strongly. This indicates that mass-transfer limitations became

increasingly important as more H₂ was added. In other words, when the H₂ loading was low, the reactor could still accommodate the supplied gas reasonably well across multiple scenarios, but as injections increased, the advantage of higher $k_L a$ became much more pronounced.

The figure also suggests that the major improvement occurred when moving from low and intermediate $k_L a$ values toward the high-transfer range. The reduction in outlet H₂ between $k_L a = 650$ and $k_L a = 1000$ was still visible, but it was smaller than the reduction observed when moving upward from the low- $k_L a$ cases.

The outlet H₂ indicator was subsequently used as one of the primary criteria in the injection-efficiency evaluation presented in Section 4.5, given its direct relevance to both process efficiency and gas quality for grid injection purposes.

4.3.3. Average pH under different H₂ injection and $k_L a$ scenarios

The average pH results showed a gradual increase with H₂ injection under all $k_L a$ scenarios as shown in Figure 10. In the no-injection condition, all cases started from nearly the same pH level, indicating that variation in $k_L a$ alone had little influence on reactor acidity or alkalinity before external H₂ was introduced. Once injection began, however, pH increased progressively, showing that the addition and conversion of H₂ affected the carbonate equilibrium and buffering behavior of the system. Figure 10 shows average pH value under each scenario.

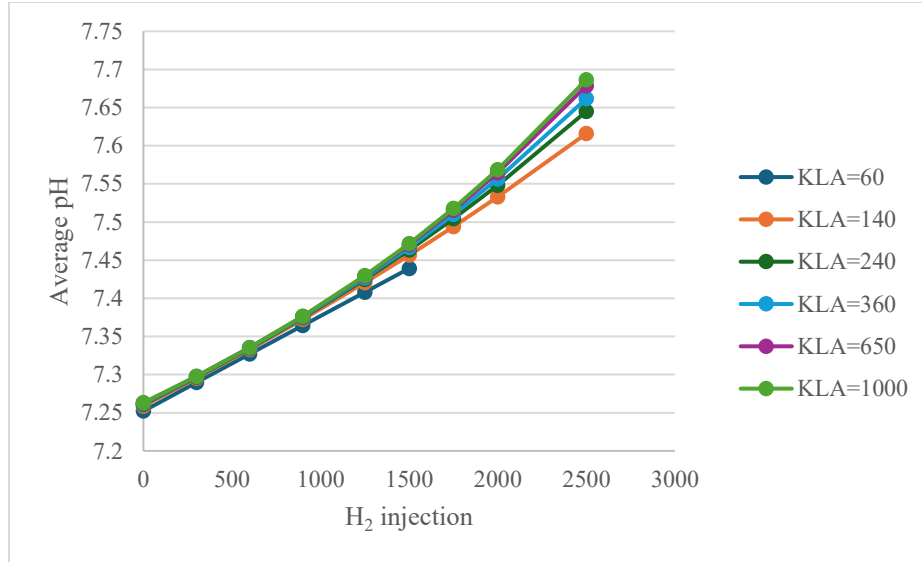


Figure 12. Average pH for each condition

A clear trend in the figure is that higher $k_L a$ generally produced slightly higher average pH values at the same injection level. This means that stronger gas-liquid transfer intensified the liquid-phase chemical response to H₂ addition. Under low- $k_L a$ conditions, the pH increase remained more limited, whereas under high- $k_L a$ conditions the pH rise became more pronounced. This behavior is consistent with the stronger CH₄ response and lower outlet H₂ losses observed at high $k_L a$, since more effective H₂ utilization also implies stronger consumption of dissolved inorganic carbon species that contribute to the reactor buffering system.

Average pH provides a direct indication of the overall chemical response and process stability of the reactor. A gradual increase in pH reflects the progressive consumption of CO₂ and bicarbonate during biomethanation, but if the increase becomes too strong, it may indicate movement toward an unfavorable operating region.

4.3.4. Maximum pH under different injection and $k_L a$ scenarios

The maximum pH results showed a clear upward trend with increasing H₂ injection in all investigated $k_L a$ scenarios which was predictable based on average pH increment as shown in figure 11. Compared with average pH, the maximum pH responded more strongly, which makes it a more sensitive indicator of the chemical stress imposed on the reactor. This means that even when the overall pH change appeared gradually, peak pH values could increase more sharply and reveal a stronger shift in reactor conditions.

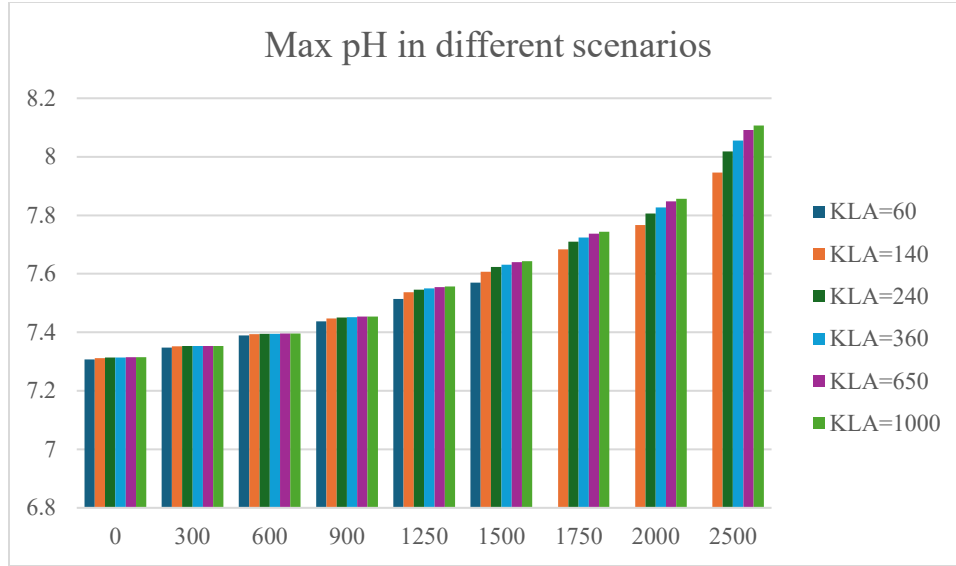


Figure 13. Maximum pH reaching 8.1 in highest k_La and injection scenario

The effect of k_La was also more pronounced in this parameter. At low injection levels, the differences between scenarios remained relatively small, but as H_2 injection increased, the higher- k_La scenarios showed a much faster rise in maximum pH than the lower- k_La cases.

In $k_La = 650$ and $k_La = 1000$, the highest maximum pH values were reached and approached the upper stability range much more rapidly than the other scenarios. In practical terms, this shows that the conditions that were most favorable for CH_4 production and H_2 utilization were also the ones most likely to drive the reactor toward excessive alkalinization at high injection levels.

Maximum pH trend acts as an important parameter in the scenario-evaluation step. Whereas average pH describes the overall chemical state of the reactor, maximum pH provides a more cautious measure of potential instability. A scenario may still seem acceptable when only the average pH is considered, even though the peak pH values are already moving toward an unfavorable region. In this way, maximum pH was especially useful for distinguishing scenarios in which better biomethanation performance was achieved at the expense of an increasing risk of chemical imbalance.

4.3.5. Average S_{IC} under different injection and k_La scenarios

The average S_{IC} results declined progressively with increasing H_2 injection under all investigated k_La scenarios. At the no-injection condition, the S_{IC} values were very similar in all cases,

indicating that changing $k_L a$ alone had little influence on inorganic carbon concentration before external H_2 was introduced. After H_2 injection, S_{IC} declined gradually, showing that biomethanation increasingly consumed the available inorganic carbon pool.

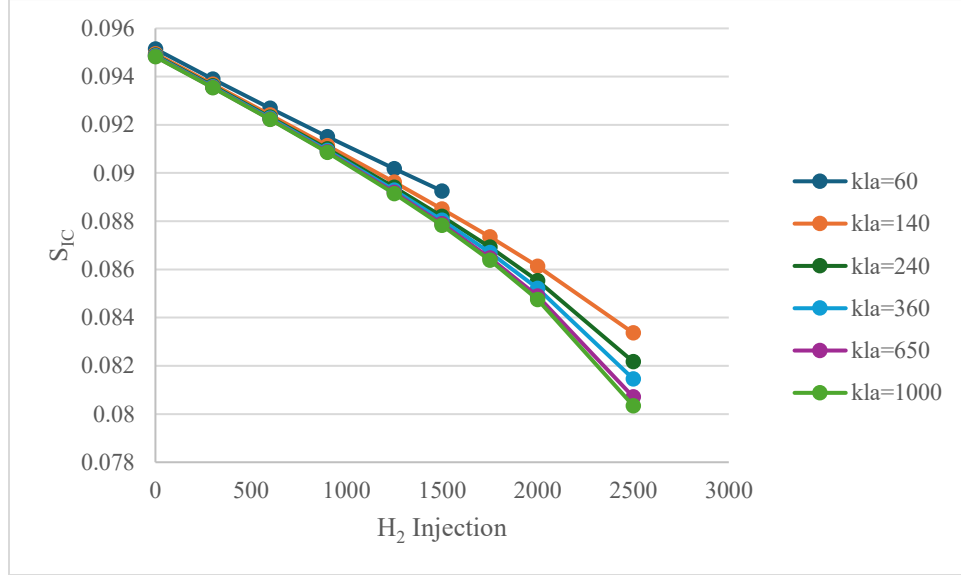


Figure 14. Drop in S_{IC} in all scenarios due to H_2 injection

As H_2 injection increased, S_{IC} decreased in all scenarios, with the extent of depletion being more pronounced under higher $k_L a$ conditions. This pattern directly reflects the stoichiometry of the hydrogenotrophic reaction: each mole of CH_4 produced via this pathway consumes one mole of CO_2 , reducing both the dissolved inorganic carbon concentration and the associated buffering capacity. The scenarios that produced the strongest CH_4 gains and highest pH values also showed the S_{IC} depletion in figure 12, confirming the mechanistic coupling between these three response variables.

4.4. Soluble Inorganic Nitrogen (S_{IN})

The variation of S_{IN} across the investigated scenarios was gradual and relatively limited compared with the more pronounced changes observed for pH and S_{IC} . As shown in Figure 13, S_{IN} decreased slightly with increasing H_2 injection under all $k_L a$ conditions. At the no-injection condition, the average S_{IN} values were nearly identical across all scenarios, remaining close to 0.0943. With increasing H_2 injection, a progressive decline was observed, reaching approximately 0.0936 at $k_L a = 60$ and about 0.0921 at $k_L a = 1000$ at the highest evaluated injection levels.

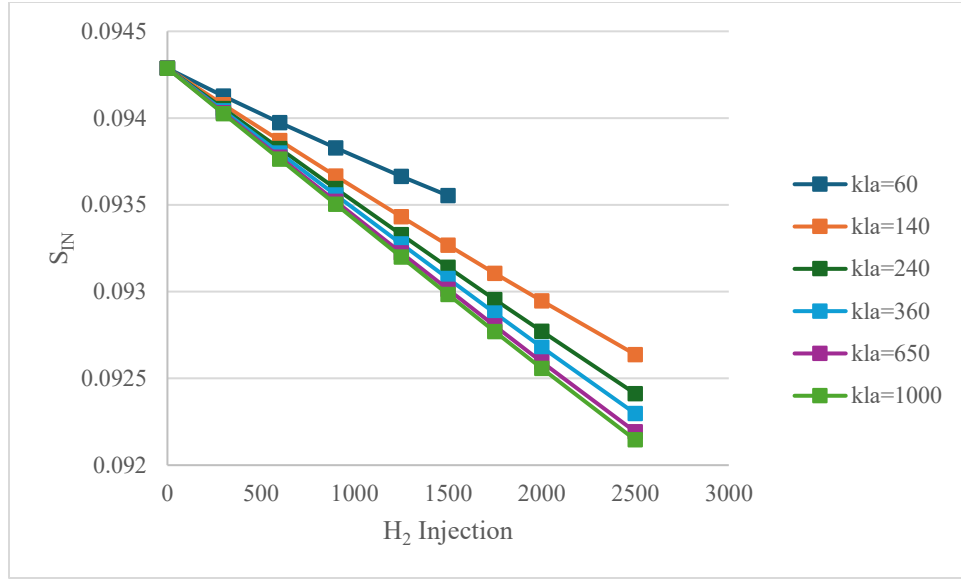


Figure 15. Schematic behavior of S_{IN} under each condition

A secondary trend was observed whereby higher k_La conditions produced marginally lower S_{IN} values for a given injection level, consistent with the generally greater biological activity under more favorable transfer conditions. However, the total variation in S_{IN} across the entire scenario matrix was less than 2%, which is small in comparison to the changes observed for pH and S_{IC} . These results indicate that nitrogen variation was not a limiting factor under the operating conditions investigated and that the dominant chemical effects of in-situ H_2 injection were mediated through the carbon cycle. The standard ADM1 nitrogen balance was therefore retained without modification, which is consistent with these findings.

4.5. Injection-efficiency evaluation

After presenting the ADM1 simulation results, a further evaluation step was carried out to identify which scenarios represented the most efficient and practically meaningful biomethanation conditions. A higher H_2 injection rate consistently produced more CH_4 in absolute terms, but not always with proportionate efficiency.

The evaluation was based on the scenario-level indicators defined in Section 3.6, as H_2 consumed $m^3 \cdot d^{-1}$, H_2 conversion efficiency (%), H_2 slip (%), and CH_4 content in the outlet gas (%).

These indicators were assessed jointly for each scenario, since no single metric is sufficient to characterize biomethanation performance. A scenario that maximizes absolute CH_4 production

while producing high H₂ slip and declining gas quality is less operationally justified than one that achieves a moderate CH₄ gain with strong H₂ utilization and stable gas composition. The evaluation was conducted separately for each $k_L a$ condition, as the efficiency of a given injection rate depends fundamentally on the corresponding transfer intensity.

4.5.1. Injection-efficiency evaluation for low $k_L a$ scenario

Under the lowest mass-transfer condition, the efficiency evaluation showed a clear trade-off between H₂ use efficiency and CH₄ production benefit. As injections increased from 300 to 1500 m³.d⁻¹ CH₄ production rose from 1648.89 to 1773.38 m³.d⁻¹, while H₂ conversion efficiency declined from 61.05% to 57.27% and H₂ slip increased from 38.95% to 42.73%. CH₄ content in the outlet gas fell from 64.14% to 57.70% over the same range. This pattern confirms that under $k_L a = 60$, the system remained strongly transfer-limited, and additional H₂ could not be utilized proportionally. Table 11 summarizes efficiency conditions for each injection scenario.

Table 11. Injection-efficiency evaluation under low $k_L a$ scenario

Injection	Average q_{CH_4}	Average q_{H_2}	H ₂ consumed	$X_{H_2}(\%)$	H ₂ slip (%)	CH ₄ content (%)
300	1648.89	116.86	183.14	61.05	38.95	64.14
600	1682.87	239.42	360.58	60.1	39.9	62.53
900	1714.95	367.69	532.31	59.15	40.85	60.67
1250	1749.97	524.45	725.55	58.04	41.96	59.06
1500	1773.38	641.01	858.99	57.27	42.73	57.7

The 300 m³.d⁻¹ injection level offered the highest H₂ conversion efficiency and the lowest slip within this scenario, though at the cost of a modest CH₄ gain. The 600 m³.d⁻¹ level provided a greater CH₄ benefit, but it was not selected since it decreased CH₄ content in output gas due to unconsumed H₂ in the system, besides the average output H₂ from the reactor increased more than two times. Based on these observations lowest injection scenario is a better option.

4.5.2. Injection-efficiency evaluation under $k_L a = 140$

Table 12 shows that under $k_L a = 140$, the efficiency evaluation had a more favorable response to H₂ injection than the low-transfer case, indicating that the increase in mass-transfer capacity improved the usefulness of H₂ addition over a broader range. As the injection level increased, CH₄

production continued to improve, while H_2 conversion efficiency declined gradually and H_2 slip increased at a relatively slow rate. At the same time, the CH_4 content of the outlet gas improved through the lower and intermediate injection range before reaching a near-plateau. This indicates that, under this condition, the process retained a wider operating window in which increasing H_2 injection remained justifiable.

Table 12. Injection-efficiency evaluation under $k_L a = 140$ scenario

Injection	Average q_{CH_4}	Average q_{H_2}	H_2 consumed	$X_{H_2}(\%)$	H_2 slip (%)	CH_4 content (%)
300	1664.27	64.74	235.26	78.42	21.58	65.99
600	1713.27	131.81	468.19	78.03	21.97	66.25
900	1761.22	201.32	698.68	77.63	22.37	66.42
1250	1816.04	285.61	964.39	77.15	22.85	66.48
1500	1854.28	347.96	1152.04	76.8	23.2	66.45
1750	1891.68	412.14	1337.86	76.45	23.55	66.35
2000	1928.19	478.18	1521.82	76.09	23.91	66.19
2500	1997.84	615.83	1884.17	75.37	24.63	65.68

Based on the evaluation table, the most relevant comparison is between the 600, 900 injection levels. The 600-case remained attractive from a conservative efficiency perspective, since it still preserved relatively strong H_2 -use performance while providing a clear CH_4 benefit. However, the 900 case represented a stronger overall compromise, because it combined a higher CH_4 improvement with the highest CH_4 content in the outlet gas among the low injection cases, while the associated penalty in H_2 -use efficiency remained limited.

4.5.3. Injection-efficiency evaluation under $k_L a = 240$

$k_L a = 240$ efficiency evaluation results are shown table 13, where the evaluations showed an improvement relative to $k_L a = 140$ in the same injection levels, indicating that the higher mass-transfer capacity allowed the reactor to use injected H_2 more effectively over a wider operating range that kept H_2 conversion efficiency above 84.8% across all injection levels, and H_2 slip stayed below 15.2%. Across the reported injection levels, CH_4 production increased steadily, while H_2 conversion efficiency decreased only gradually and H_2 slip increased at a relatively slow rate. In

addition, the CH₄ content of the outlet gas improved through the lower and intermediate injection range and then remained nearly stable before showing a slight decline at higher injections. This pattern suggests that the process under $k_L a = 240$ retained a broad and technically favorable operating window, but that the most efficient operating point should be identified by balancing CH₄ benefit against H₂-use performance rather than by selecting the highest CH₄ output alone.

Table 13. Injection-efficiency evaluation under $k_L a = 240$ scenario

Injection	Average q_{CH_4}	Average q_{H_2}	H ₂ consumed	$X_{H_2}(\%)$	H ₂ slip (%)	CH ₄ content (%)
300	1671.19	41.5	258.5	86.17	13.83	66.99
600	1726.28	84.05	515.95	85.99	14.01	67.44
900	1782.2	127.7	772.3	85.81	14.19	67.79
1250	1844.73	180.05	1069.95	85.6	14.4	68.02
1500	1890.5	218.4	1281.6	85.44	14.56	68.04
1750	1934.74	257.58	1492.42	85.28	14.72	68.01
2000	1977.33	297.55	1702.45	85.12	14.88	67.93
2500	2060.96	379.83	2120.17	84.81	15.19	67.65

At 1250 m³.d⁻¹ injection level was selected as the most balanced operating point in $k_L a = 240$ d⁻¹, corresponding to the near-peak CH₄ content while maintaining high H₂ conversion efficiency. 1500 m³.d⁻¹ level was an acceptable alternative when greater emphasis is placed on absolute CH₄ production, though the efficiency advantage over 1250 m³.d⁻¹ was marginal. Injection scenarios up to 1750 m³.d⁻¹ were retained for the machine-learning dataset, as all cases within this range maintained a satisfactory efficiency balance.

4.5.4. Injection-efficiency evaluation under $k_L a = 360$

For $k_L a = 360$, the evaluation showed that the reactor had entered a much more favorable operating regime, where increasing H₂ injection continued to improve CH₄ performance without causing a strong efficiency penalty. Across the reported injection levels, CH₄ production increased steadily, H₂ conversion efficiency remained close to 90%, and H₂ slip increased only slightly. CH₄ content improved continuously from 67.30% at 300 m³.d⁻¹ to 77.70% at 2500 m³.d⁻¹, a gain of more than 10 percentage points achieved without a substantial efficiency penalty. This pattern

indicates that the trade-off between CH₄ benefit and H₂ utilization efficiency had weakened considerably compared to the lower $k_L a$ scenarios. Table 14 summarizes efficiency conditions for this case.

Table 14. Injection-efficiency evaluation under $k_L a = 360$ scenario

Injection	Average q_{CH_4}	Average q_{H_2}	H ₂ consumed	$X_{H_2}(\%)$	H ₂ slip (%)	CH ₄ content (%)
300	1674.06	28.99	271.01	90.34	9.66	67.3
600	1733.12	58.48	541.52	90.25	9.75	68.92
900	1791.87	88.52	811.48	90.16	9.84	70.49
1250	1859.96	124.29	1125.71	90.06	9.94	72.24
1500	1908.19	150.32	1349.68	89.98	10.02	73.43
1750	1955.95	176.76	1573.24	89.9	10.1	74.58
2000	2003.06	203.6	1796.4	89.82	10.18	75.67
2500	2093.58	258.38	2241.62	89.66	10.34	77.7

Both 1500 and 1750 m³.d⁻¹ injection levels represented strong operating points under this condition. The 1500 m³.d⁻¹ case offered marginally better H₂ conversion efficiency, while the 1750 m³.d⁻¹ case provided a higher CH₄ content in the outlet gas, with a very small efficiency penalty between the two. The full injection ranges up to 2500 m³.d⁻¹ was retained for the machine-learning dataset, given that the reactor maintained favorable efficiency indicators throughout.

4.5.5. Injection-efficiency evaluation under $k_L a = 650$

As it is shown in table 15 under $k_L a = 650$, H₂ conversion efficiency exceeded 94% at all injection levels tested, declining by only 0.21 percentage points across the full range from 300 to 2500 m³.d⁻¹. H₂ slip remained below 5.8%, and CH₄ content in the outlet gas improved continuously from 67.76% to 82.17%. These results confirm that the reactor had moved into a high-efficiency operating regime in which increasing H₂ injection rates could be sustained without the efficiency penalties or gas-quality deterioration that characterized the lower- $k_L a$ scenarios.

Table 15. Injection-efficiency evaluation under $k_L a = 650$ scenario

Injection	Average q_{CH_4}	Average q_{H_2}	H_2 consumed	$X_{H_2}(\%)$	H_2 slip (%)	CH_4 content (%)
300	1675.35	16.77	283.23	94.41	5.59	67.76
600	1738.01	33.67	566.33	94.39	5.61	69.86
900	1801.02	50.78	849.22	94.36	5.64	71.93
1250	1874.2	70.99	1179.01	94.32	5.68	74.29
1500	1925.74	85.58	1414.42	94.29	5.71	75.94
1750	1976.98	101.99	1648.01	94.27	5.73	77.56
2000	2027.16	115.14	1884.86	94.24	5.76	79.14
2500	2123.35	145.08	2354.92	94.2	5.8	82.17

2000 and 2500 $m^3 \cdot d^{-1}$ injection levels were both identified as technically favorable operating points under this condition. 2000 $m^3 \cdot d^{-1}$ case maintained a marginally higher H_2 conversion efficiency, while 2500 $m^3 \cdot d^{-1}$ injection provided a stronger CH_4 -production benefit and the highest outlet gas quality. Given the negligible efficiency difference between these two levels, both were retained in the machine-learning dataset along with the full injection range.

4.5.6. Injection-efficiency evaluation under high mass-transfer scenarios

Under the $k_L a = 1000$, the efficiency evaluation showed the best overall performance with conversion efficiency exceeding 96% at every injection level. Across the reported injection levels, CH_4 production increased steadily, while H_2 conversion efficiency remained almost constant and H_2 slip increased only marginally. At the same time, CH_4 content in the outlet gas improved from 67.97% at 300 $m^3 \cdot d^{-1}$ to 84.24% at 2500 $m^3 \cdot d^{-1}$, the highest value recorded across the entire simulation matrix, which is illustrated in table 16.

Table 16. Injection-efficiency evaluation under high $k_L a$ scenario

Injection	Average q_{CH_4}	Average q_{H_2}	H_2 consumed	$X_{H_2}(\%)$	H_2 slip (%)	CH_4 content (%)
300	1678.43	11.14	288.86	96.29	3.71	67.97
600	1742.51	22.3	577.7	96.28	3.72	70.3
900	1806.43	33.54	866.46	96.27	3.73	72.61
1250	1880.92	46.77	1203.23	96.26	3.74	75.26
1500	1933.74	56.28	1443.72	96.25	3.75	77.13
1750	1986.17	65.85	1684.15	96.24	3.76	78.97
2000	2037.77	75.46	1924.54	96.23	3.77	80.78
2500	2137.08	94.74	2405.26	96.21	3.79	84.24

The 2500 $m^3.d^{-1}$ injection level was the most productive operating point under this condition, delivering the highest CH_4 output and outlet gas quality with only a negligible efficiency penalty relative to lower injection rates.

4.5.7. Overall conclusion of the injection-efficiency evaluation

The injection-efficiency evaluation showed that the effect of increasing H_2 injection depended strongly on the gas-liquid mass-transfer condition, and that the most favorable injection level shifted upward as $k_L a$ increased.

Under low- $k_L a$ conditions ($k_L a = 60 d^{-1}$), increasing H_2 injection beyond moderate levels intensified H_2 slip, reduced H_2 conversion efficiency, and degraded CH_4 content in the outlet gas. Under these conditions, additional H_2 supply could not be dissolved and converted proportionately, such that higher injection rates delivered more absolute CH_4 but at a progressively lower efficiency and quality. As $k_L a$ increased to intermediate values (140-240 d^{-1}), the efficiency penalty associated with higher injection rates diminished, and a broader range of operating conditions remained technically justifiable. Under high $k_L a$ conditions (360-1000 d^{-1}), the trade-off between

CH₄ enhancement and H₂ utilization efficiency weakened substantially. H₂ conversion efficiency remained high and H₂ slip remained low across the full injection range, allowing larger injection rates to be applied without significant efficiency loss.

This systematic shift in the preferred injection range with $k_L a$ is explained directly by the role of gas-liquid mass transfer in in-situ biomethanation. The low aqueous solubility of H₂ means that, under weak transfer conditions, increasing the gas-phase H₂ supply enlarges the mismatch between injection and biological uptake, leading to accumulation in the headspace and losses through the outlet stream. Under strong transfer conditions, dissolved H₂ concentrations in the liquid phase are maintained within the range that hydrogenotrophic methanogens can consume effectively, and larger injections translate into proportionate biomethanation gains. These results confirm that scenario selection for the machine-learning stage should be based on the combined assessment of injection rate and transfer intensity, rather than on either variable in isolation.

4.6. Selection of efficient scenarios for the machine-learning dataset

Based on the injection-efficiency evaluation described in Section 4.5, a scenario-screening step was carried out prior to machine-learning dataset preparation. The objective was to retain only those operating conditions that represented a meaningful and balanced biomethanation outcome, and to exclude scenarios in which increasing injection produced disproportionate H₂ losses or declining gas quality without sufficient CH₄ benefit. The retained scenarios for each $k_L a$ condition are summarized in Table 17.

Table 17. Selected scenarios for ML development

$k_L a$	Injection scenarios retained for ML
60	0, 300
140	0, 300, 600, 900
240	0, 300, 600, 900, 1250, 1500, 1750
360	0, 300, 600, 900, 1250, 1500, 1750, 2000, 2500
650	0, 300, 600, 900, 1250, 1500, 1750, 2000, 2500
1000	0, 300, 600, 900, 1250, 1500, 1750, 2000, 2500

The reason for excluding some scenarios from the machine-learning stage was not that they were numerically invalid, but that they were less suitable as representative efficient operating

conditions. Under low mass-transfer conditions, especially $k_L a = 60$, higher H_2 injections led to progressively higher H_2 slip, lower H_2 conversion efficiency, and a reduction in CH_4 content in the outlet gas. In this case, additional H_2 did not translate into proportionate biomethanation benefit, so the higher-injection scenarios were not considered appropriate for training the machine-learning models. A similar but less severe pattern was observed under $k_L a = 140$, where the lower and intermediate injections provided more balanced performance, while the higher injections became less attractive from an efficiency-oriented point of view.

As $k_L a$ increased to 240, the practical injection window expanded, allowing a broader range of scenarios to remain efficient enough for inclusion. At $k_L a = 360$, 650, and 1000, the reactor maintained high H_2 conversion efficiency and low H_2 slip across the full evaluated range, and higher H_2 injections continued to improve CH_4 production and gas quality without causing the same efficiency seen at lower $k_L a$. For this reason, the full injection range was retained for these higher-transfer scenarios.

4.7. Machine-learning prediction of CH_4 flow rate

Following the scenario selection described above, the filtered dataset was used to train and evaluate the four ML models for the prediction of CH_4 flow rate and total biogas production. Model performance was assessed using R^2 , MAE, RMSE, and NRMSE, applied to both training and testing subsets. The results for each algorithm are presented in the following subsections, followed by a comparative discussion in Section 4.7.5.

4.7.1. Support Vector Regression (SVR)

The SVR model achieved the highest predictive accuracy among all four algorithms evaluated in this study. For CH_4 flow rate prediction, the model yielded R^2 values of 0.997 and 0.996 for the training and testing datasets, respectively. The negligible difference between training and testing performance indicates that the model generalized well to unseen data and did not overfit to training-specific patterns. In the predicted-versus-actual scatter plots (Figures 16), data points remained tightly clustered around the 1:1 line across nearly the full output range, with only a slight broadening of scatter in a limited intermediate region. All predictions fell within the $\pm 10\%$ error band in both datasets, confirming consistent and accurate predictions across the full range of simulated operating conditions.

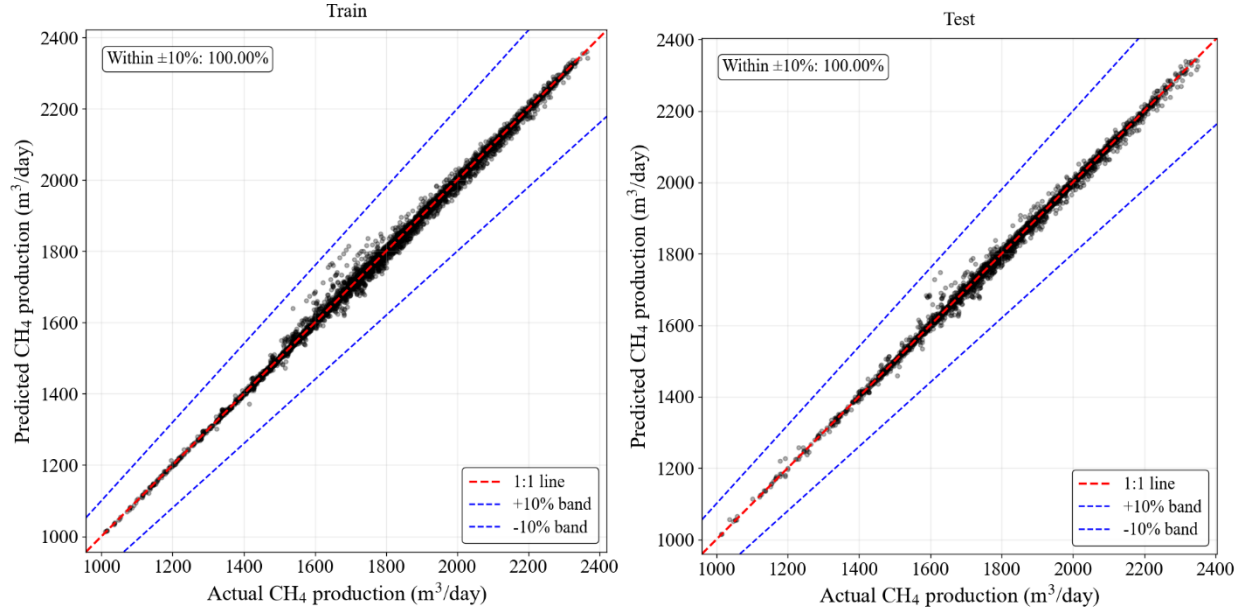


Figure 16. Actual versus predicted CH₄ production using the SVR model for training and testing datasets. The red dashed line shows the 1:1 line, and the blue dashed lines represent the $\pm 10\%$ error bounds.

SVR also performed very strongly in predicting total biogas flow rate. The scatter plots for both the training and testing datasets followed the 1:1 line very closely, and the scatter remained narrow across nearly the full biogas range. The R^2 values of 0.9955 for training and 0.9948 for testing.

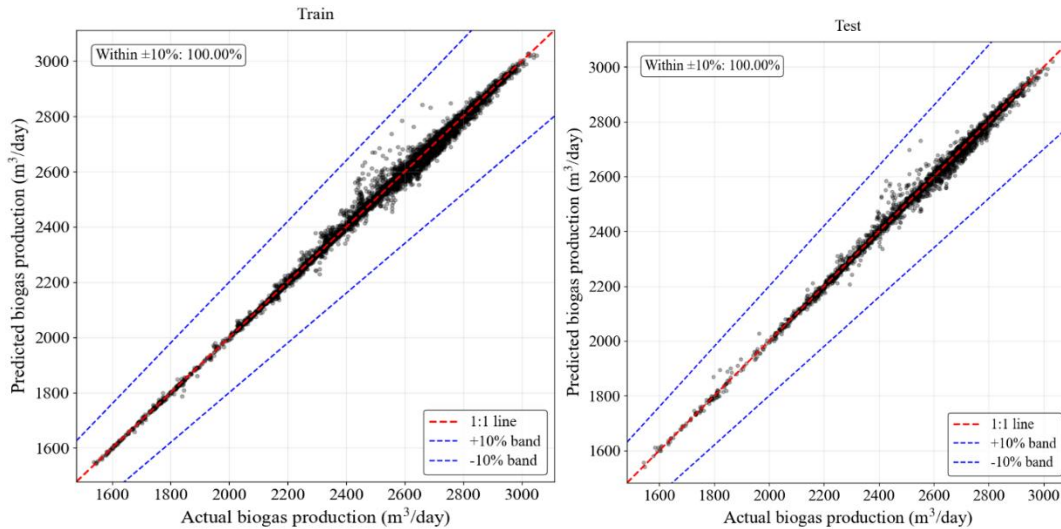


Figure 17. Actual versus predicted biogas production using the SVR model for training and testing datasets.

4.7.2. Random Forest (RF)

Random forest has shown a strong performance at CH₄ prediction as well however; this model was not as precise and accurate as SVR. The model achieved R^2 values of 0.9951 for training and

0.9879 for testing, which shows it explained most of the variation in CH₄ flow rate and had a high level of predictive strength on test data. The predicted-versus-actual plots support this statement, since most points remained concentrated near the 1:1 line over the full output range as shown in figure 18. However, compared with SVR, the RF scatter was slightly broader, especially in the testing dataset and in the higher CH₄ region. This suggests that RF captured the overall nonlinear pattern of the process well but reproduced the highest output conditions with slightly less precision.

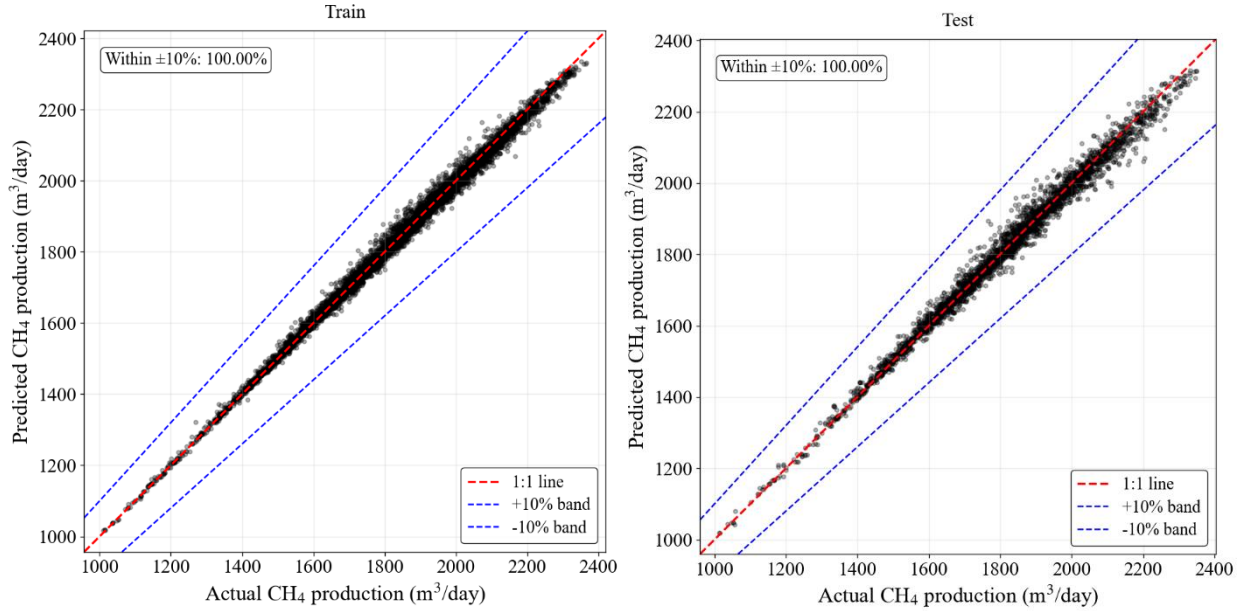


Figure 18. Actual versus predicted CH₄ production using the RF model for testing and training datasets. The red dashed line shows the 1:1 line, and the blue dashed lines represent the $\pm 10\%$ error bounds.

A similar pattern was observed for biogas prediction in figure 19. The RF model achieved R^2 values of 0.9899 for training and 0.9807 for testing, which still indicates strong predictive performance, strong overall performance, and a slightly wider training-to-testing gap than SVR. Moreover, the spread in the testing plot was somewhat wider than in the SVR case, particularly in the higher biogas range. This indicates that RF preserved the main production trend effectively, but its predictions became less compact as output increased.

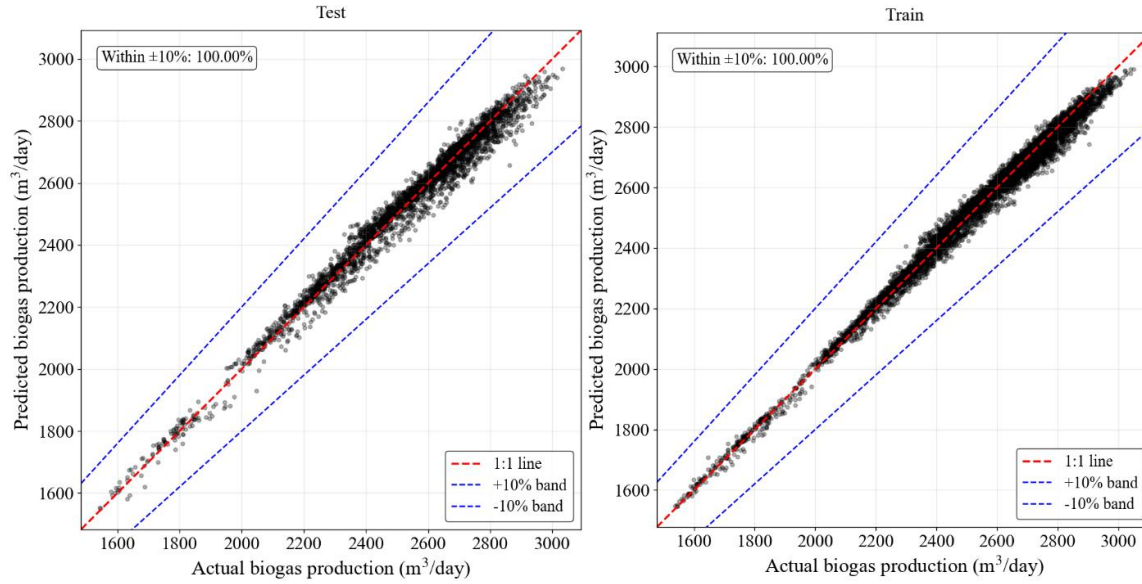


Figure 19. RF prediction performance for train and test data set of biogas production

4.7.3. Extreme Learning Machine (ELM)

The ELM model showed strong predictive performance for CH₄ production, achieving R² values of 0.99 and 0.989 for the training and testing datasets, respectively. As shown in Figure 20, the predicted values were closely distributed around the 1:1 line for both datasets, indicating good agreement between the simulated and predicted CH₄ flow rates. The small difference between the training and testing values also indicates consistent model performance on unseen observations.

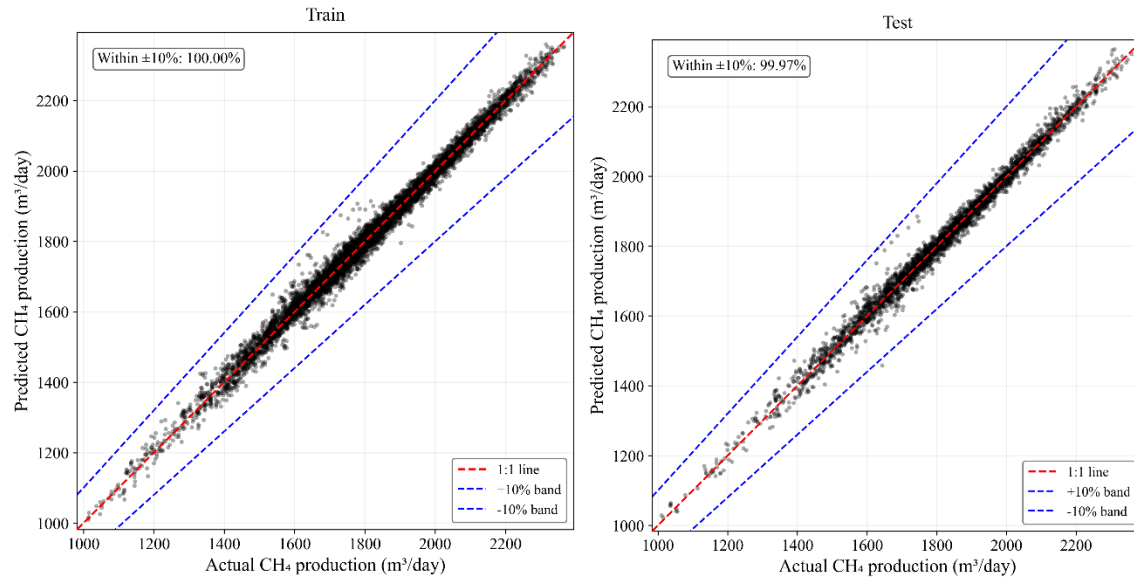


Figure 20 Actual versus predicted CH₄ production using the ELM model for the testing and training datasets.

A similarly strong result was obtained for biogas prediction. The ELM model achieved training and testing R^2 values of 0.985 and 0.983, respectively. Figure 21 shows that the predicted biogas values generally followed the 1:1 relationship across the investigated output range. Although the prediction accuracy was slightly lower than that obtained for CH_4 , ELM maintained a high level of agreement between the actual and predicted biogas flow rates.

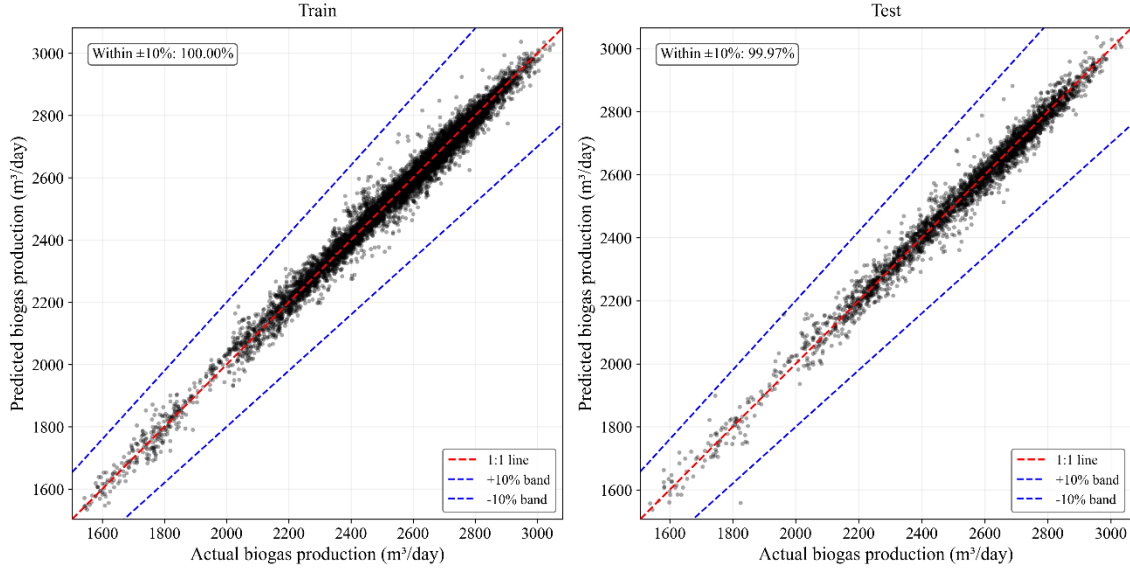


Figure 21 Actual versus predicted biogas production, using the ELM method.

4.7.4. Evolutionary Polynomial Regression (EPR)

The EPR model showed good predictive performance for CH_4 production, achieving R^2 values of 0.954 and 0.956 for the training and testing datasets, respectively. As shown in Figure 22, most predicted values were distributed close to the 1:1 reference line across the investigated CH_4 production range, although the scatter was visibly wider than that observed for the higher-performing ML models. The very small difference between the training and testing values indicates that the model maintained a similar level of accuracy when applied to unseen observations. This consistency suggests that the identified EPR expression captured the main nonlinear relationships between the selected process variables and CH_4 production without a noticeable loss of generalization between the training and testing datasets.

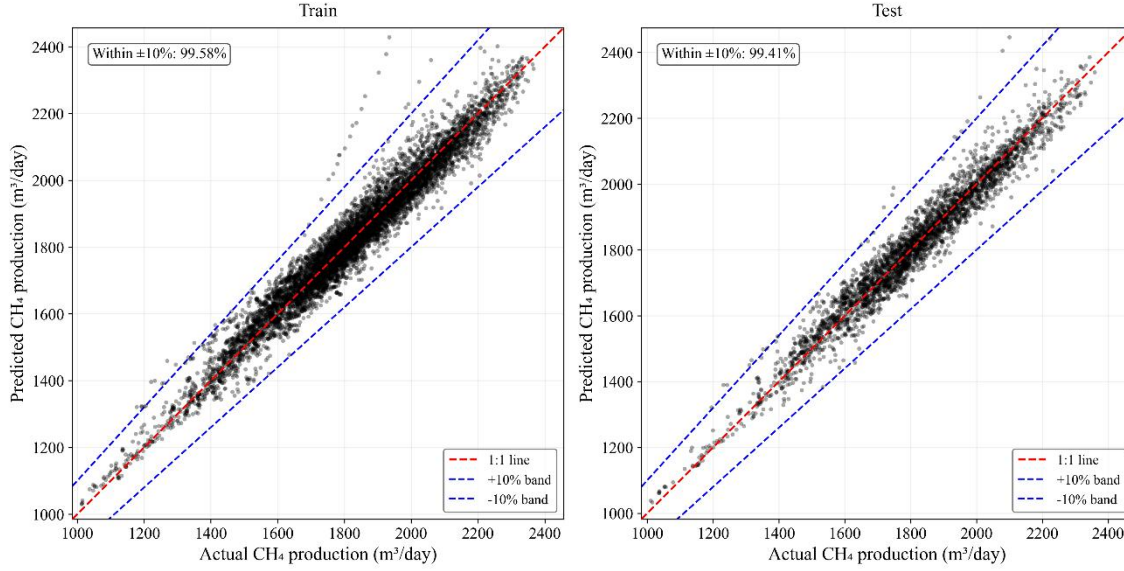


Figure 22 Actual versus predicted CH₄ production using the EPR model for training and testing datasets. The red dashed line shows the 1:1 line, and the blue dashed lines represent the $\pm 10\%$ error bounds.

In addition to prediction accuracy, EPR provides an explicit mathematical expression relating the input variables to the target output. The final EPR model for CH₄ production consisted of five nonlinear terms. The corresponding exponent matrix is presented below.

The final exponent matrix for CH₄ production is:

$$\begin{bmatrix} 0 & 0 & 0 & 0.5 & 1 & 0 & 0 & 1 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0.5 \\ -0.5 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0.5 & -0.5 & 1.5 & 0.5 & 0 & -1 & 0.5 & 0.5 & 0 \end{bmatrix}$$

Each row of the exponent matrix represents one term of the EPR equation, while each column corresponds to one normalized input variable. A zero exponent indicates that the corresponding variable is not included in that term.. The final EPR formula for CH₄ production was constructed as:

$$\begin{aligned} CH_{4,norm} &= -0.069 - 0.55(Inj_{norm}^{0.5})(S_{IC,norm})(X_{pr,norm})(X_{Li,norm}^{0.5}) + 0.618(S_{IN,norm})(S_{pro,norm}^{0.5}) \\ &+ 0.345(pH_{norm}^{-0.5})(inj_{norm})(X_{Li,norm}^{0.5}) + 0.45(S_{pro,norm}) \\ &+ 0.135(k_L a_{norm}^{0.5})(Q_{norm}^{-0.5})(Inj_{norm}^{1.5})(S_{IC,norm}^{0.5})(X_{ch,norm}^{-1})(X_{pr,norm}^{0.5})(X_{Li,norm}^{0.5}) \end{aligned}$$

The explicit EPR formula provides an important practical advantage because it translates the prediction model into a direct mathematical relationship between the selected input variables and CH₄ production. Unlike black-box models, the EPR expression can be evaluated directly once the required process variables are available, without requiring repeated model training. In addition, the explicit formulation improves model transparency by showing how the prediction is constructed from combined nonlinear effects among the input variables. Therefore, although EPR showed lower predictive accuracy than SVR, ELM, and RF in the present study, it provides an interpretable and computationally simple approach for rapid state-based estimation of CH₄ production within the investigated operating domain.

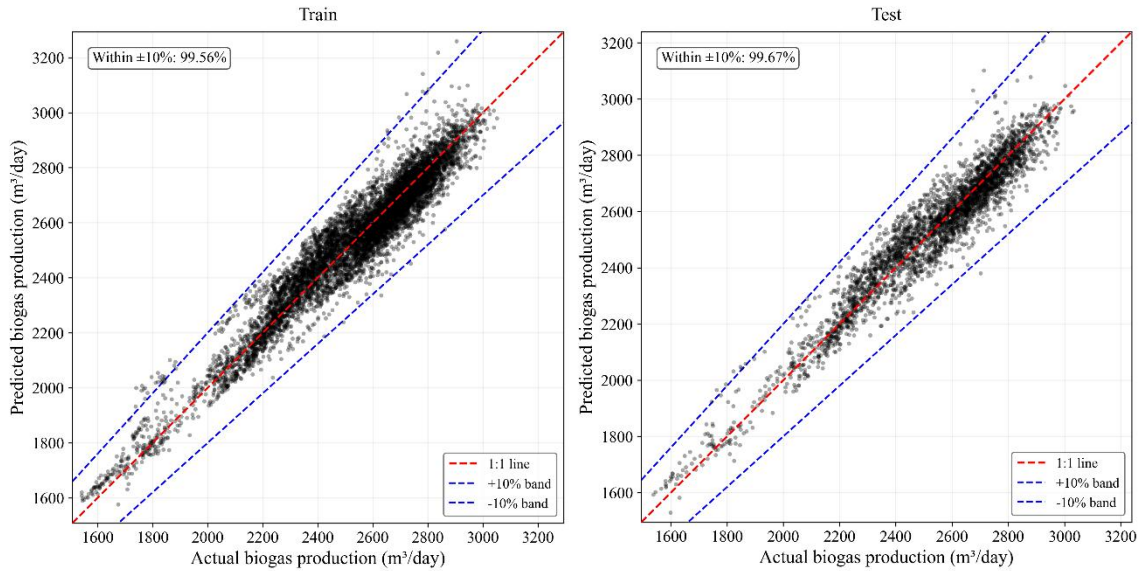


Figure 23 Actual versus predicted biogas production using the EPR model for testing and training datasets. The red dashed line shows the 1:1 line, and the blue dashed lines represent the $\pm 10\%$ error bounds.

The EPR model also showed good predictive performance for biogas production, achieving R^2 values of 0.9276 for both the training and testing datasets. As shown in Figure 23, the predicted values generally followed the 1:1 reference line across the investigated biogas production range, although a wider distribution around the reference line was observed compared with CH₄ prediction. The nearly identical training and testing values indicate that the model maintained consistent predictive performance when applied to unseen observations

The provided matrix for biogas prediction is:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ -0.5 & 0 & 0 & 0 & 1 & 0 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & -0.5 & 0 & 0.5 & 0 & 0 & 2.5 & 0.5 & 0 \\ 0 & -0.5 & 0 & 1.5 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 1 \end{bmatrix}$$

Based on this structure, the final normalized EPR equation for biogas prediction was obtained as:

$$\begin{aligned} Biogas_{norm} = & -0.052 + 1.163(S_{IC,norm})(S_{pro,norm}) \\ & + 0.134 (pH_{norm}^{-0.5})(S_{IC,norm})(X_{ch,norm}^{0.5}) \\ & - 0.523 (pH_{norm}^{0.5})(Q_{norm}^{-0.5})(S_{IC,norm}^{0.5})(X_{pr,norm}^{2.5})(X_{Li,norm}^{0.5}) \\ & + 0.251 (k_L a_{norm}^{-0.5})(Inj_{norm}^{1.5})(S_{IC,norm}^{0.5}) \\ & + 0.403 (Q_{norm}^{-0.5})(Inj_{norm}^{0.5})(X_{pt,norm}^{0.5})(S_{pro,norm}) \end{aligned}$$

4.7.5. Comparative discussion

The comparative performance of the four ML models for CH₄ and biogas prediction is presented in Figures 24 and 25, respectively. Considering the testing results, the overall performance followed the same order for both prediction targets: SVR, ELM, RF, and EPR. For CH₄ prediction, SVR achieved the highest testing R² of 0.996, followed by ELM at 0.9899, RF at 0.9879, and EPR at 0.9561. The error metrics showed the same general pattern. SVR produced the lowest MAE and RMSE, at 7.4 and 12.2 m³·d⁻¹, respectively. ELM followed with an MAE of 16.51 m³·d⁻¹ and RMSE of 22.22 m³·d⁻¹, while RF produced slightly higher errors of 17.5 and 24.2 m³·d⁻¹. EPR showed the largest errors among the four models, with an MAE of 34.94 m³·d⁻¹ and RMSE of 46.43 m³·d⁻¹. The corresponding NRMSE values remained low for all models, but again showed the advantage of SVR, followed closely by ELM and RF.

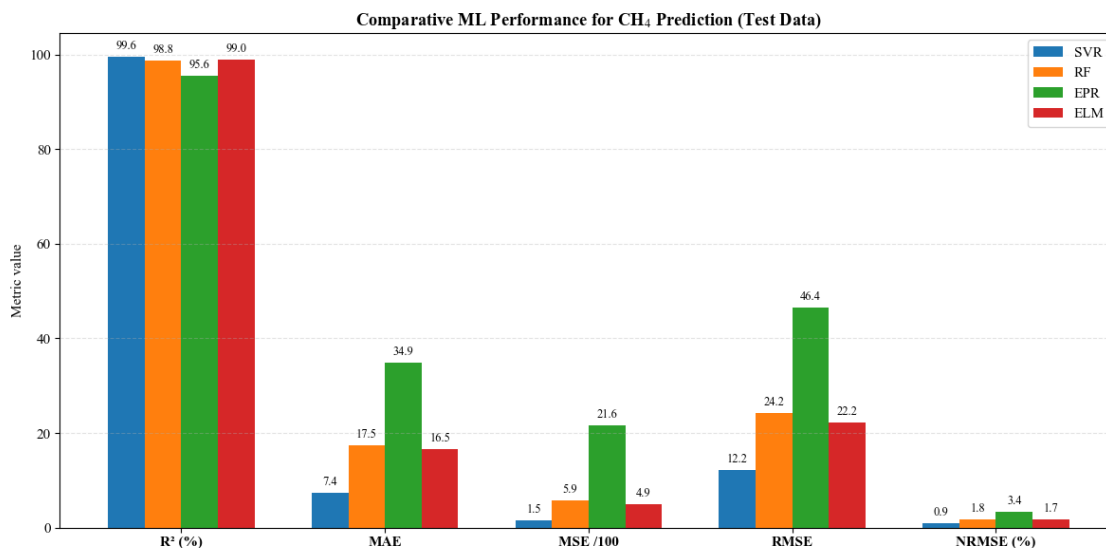


Figure 24 Comparative performance of SVR, RF, EPR, and ELM models for CH₄ prediction on the testing dataset

A similar performance hierarchy was observed for total biogas prediction. SVR again achieved the highest testing R², reaching 0.9948, followed by ELM at 0.9834, RF at 0.9807, and EPR at 0.9276. The error metrics provide an important additional distinction between the models. SVR achieved the lowest MAE and RMSE of 11.0 and 18.3 m³·d⁻¹, respectively. ELM produced an MAE of 24.15 m³·d⁻¹ and an RMSE of 32.70 m³·d⁻¹, slightly outperforming RF, for which the corresponding values were 26.6 and 35.4 m³·d⁻¹. EPR showed a larger prediction error, with an MAE of 52.87 m³·d⁻¹ and RMSE of 68.31 m³·d⁻¹. Therefore, the ranking obtained from R² was also supported by the magnitude-based error metrics rather than being dependent on one indicator.

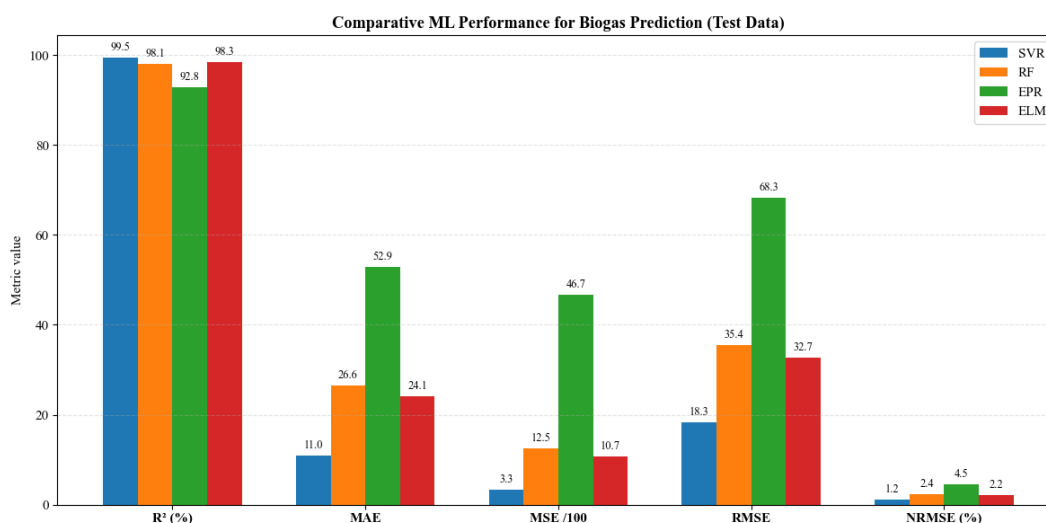


Figure 25 Comparative performance of SVR, RF, EPR, and ELM models for biogas prediction on the testing dataset

The superior performance of SVR can be explained by the mathematical structure of the model and the characteristics of the present dataset. The AD and biomethanation variables have continuous and strongly nonlinear relationships with CH₄ production, particularly through interactions among $k_L a$, H₂ injection, pH, S_{IC}, S_{IN}, and the substrate-related variables. The RBF kernel used by SVR is well suited to representing such smooth nonlinear relationships because it constructs the regression function from similarities between observations in the transformed feature space. At the same time, the regularization parameter C and the ε -insensitive loss restrict excessive sensitivity to individual observations and provide a balance between fitting accuracy and model complexity. This combination appears particularly suitable for the simulation-derived dataset, where the process response varies continuously across operating conditions. Consequently, SVR was able to reproduce both the overall process trend and local variations with the lowest MAE and RMSE among the four algorithms.

ELM achieved the second-best overall performance and remained very close to RF, particularly for CH₄ prediction. Its strong performance can be related to the nonlinear transformation produced by the hidden layer together with the preprocessing adopted in this study. In ELM, the input weights and hidden-layer biases are generated randomly and remain fixed, while the output weights are obtained analytically using a least-squares solution. This makes the representation generated by the hidden layer particularly sensitive to the numerical scale of the predictors. Standardization therefore played an important role by placing variables with different units and magnitudes on comparable scales before the sigmoid transformation. This reduced scale-related dominance among the predictors and limited excessive saturation of the sigmoid activation function. In addition, the selected configuration of 250 hidden neurons provided sufficient nonlinear representation capacity for the process relationships contained in the dataset. The remaining small performance gap relative to SVR can partly be related to the fact that ELM relies on a randomly generated hidden representation, whereas the SVR regression surface is directly determined through optimization of the support vectors and kernel-based objective function.

RF also provided high predictive accuracy, but its performance was slightly below ELM and SVR. The strength of RF lies in its ability to capture nonlinear relationships and interactions without requiring a predefined mathematical form. Individual trees recursively partition the input domain, while bootstrap sampling, random feature selection, and ensemble averaging improve robustness

and reduce the variance associated with a single decision tree. These characteristics explain the strong RF results obtained in this study. However, tree-based regression represents a continuous response through a collection of discrete partitions and approximately piecewise-constant predictions. The present ADM1-derived data show relatively smooth changes in CH₄ and biogas production with variations in the process states and operating conditions. As a result, the smooth regression surfaces generated by SVR and the nonlinear hidden-layer transformation of ELM were able to reproduce these continuous variations slightly more precisely. Ensemble averaging in RF can also smooth predictions near the extreme regions of the response range, which is consistent with the somewhat larger scatter previously observed at higher production levels.

EPR produced lower predictive accuracy than the other three algorithms, but this difference should be interpreted in relation to its fundamentally different modelling objective. Unlike SVR, ELM, and RF, EPR searches for a compact symbolic equation composed of a limited number of polynomial-type terms and candidate exponents. In the present study, the final equations were intentionally restricted to five terms. This restriction reduces the flexibility available to reproduce all local nonlinearities and higher-order interactions present in the simulation dataset, which explains the larger MAE and RMSE values. However, the close agreement between EPR training and testing indicates that the resulting expressions were stable and did not obtain their accuracy through excessive fitting of the training data. More importantly, EPR provides an explicit mathematical relationship between the process variables and the predicted output. Therefore, its lower predictive accuracy represents a trade-off between model flexibility and interpretability, rather than simply poor performance. For applications in which transparency, direct calculation, and understanding of the mathematical contribution of the predictors are important, this remains a significant advantage of EPR.

All four algorithms showed slightly lower predictive performance for biogas than for CH₄, although the magnitude of the reduction differed among models. This can be related to the nature of the two prediction targets. CH₄ production represents one specific product of the AD and biomethanation pathways, whereas total biogas flow reflects the combined behavior of several gas-phase components. In the present system, changes in H₂ injection and $k_L a$ simultaneously influence CH₄ formation, CO₂ consumption, and unconverted H₂ leaving the reactor. Consequently, total biogas flow contains the combined response of several interacting gas-phase

processes and is therefore somewhat more difficult to represent using the same ten predictors. Despite this additional complexity, SVR, ELM, and RF retained testing values above 0.98, showing that all three models captured the dominant relationships governing both outputs.

4.7.6. Feature Importance and Sensitivity Analysis

Feature-importance and sensitivity analyses were performed to identify the variables that contributed most strongly to CH₄ prediction across RF, SVR, ELM, and EPR. The consensus PFI analysis identified $k_L a$ as the most influential variable, followed by pH, S_{IC}, S_{IN}, and S_{pro}. A similar pattern was obtained from the sensitivity analysis, where $k_L a$ again ranked first, followed by pH, S_{IN}, S_{IC}, and S_{pro}. The agreement between the two analyses indicates that the dominant predictive variables were relatively consistent despite the different mathematical structures of the four ML models.

The most consistent result was obtained for $k_L a$. It ranked first in both PFI and sensitivity analyses, and all four models showed an increase in predicted CH₄ as increased. This behavior agrees directly with the mechanistic ADM1 simulations. The agreement between the mechanistic simulations and four independent ML models therefore provides strong evidence that mass-transfer intensity is a dominant variable controlling CH₄ production under the investigated biomethanation conditions.

pH was also consistently identified as an important predictor. This result is reasonable because pH changed systematically as biomethanation intensified. Greater hydrogenotrophic activity consumes dissolved CO₂ and bicarbonate-related inorganic carbon, thereby reducing part of the carbonate buffering capacity and causing pH to increase.

S_{IC} was another important variable, but its relationship with CH₄ differed from that of pH and $k_L a$. $k_L a$ conditions associated with higher CH₄ production were generally accompanied by lower S_{IC} concentration. The importance of S_{IC} in the ML models therefore reflects its role as an indicator of the extent to which inorganic carbon has been consumed during biomethanation.

S_{IN} also appeared among the more influential predictors, although its physical response was considerably weaker than that of S_{IC}.

The substrate-related variables X_{ch} , X_{pr} , and X_{Li} showed lower overall importance than $k_L a$, pH, S_{IC} , and S_{IN} , although they still contributed useful predictive information. These variables represent biodegradable substrate fractions and therefore influence CH_4 production indirectly through hydrolysis and subsequent conversion pathways. Their lower individual importance does not indicate that substrate composition is unimportant; rather, part of their information is shared with other process-state variables and with each other.

The relatively low importance assigned to H_2 injection requires interpretation. The mechanistic simulations clearly demonstrated that increasing H_2 injection increases CH_4 production. However, the effectiveness of a given H_2 injection rate depends strongly on the corresponding $k_L a$ condition. In addition, variables such as pH, S_{IC} , and S_{pro} already reflect part of the reactor response generated by H_2 addition. Consequently, once these process-state variables are provided to the ML model, H_2 injection may contribute relatively little additional predictive information by itself. Therefore, its low PFI and sensitivity ranking should not be interpreted as evidence of low physical importance.

5. Conclusion and Future Works

This thesis sets out to address a gap that persists in the AD modelling literature: the standard ADM1 framework, despite its broad adoption, requires additional mechanistic components to represent in-situ HBM, and existing hybrid approaches for linking simulation-derived process insight to real-time CH_4 estimation remain limited. Both gaps were addressed through a sequential framework in which a mechanistically extended ADM1, implemented within the BSM2 structure using PyADM1, generated a systematic simulation dataset that subsequently trained four machine-learning algorithms. The model extensions comprised an external H_2 injection term, diffusivity-scaled gas-specific mass-transfer coefficients for H_2 , CO_2 , and CH_4 , and an inorganic carbon limitation factor on hydrogenotrophic methanogenesis. Three modifications that, taken together, allowed the simulation to capture the dynamic coupling between hydrogen supply, gas-liquid dissolution, and biological conversion that standard ADM1 cannot represent. Verification against published benchmark data confirmed deviations below 8% for CH_4 production and below 1% for pH, establishing that the extended model reproduced known biomethanation behavior with sufficient fidelity to serve as a reliable basis for scenario generation.

The most consequential finding from the simulation study is that the relationship between H_2 injection rate and process benefit is not uniformly proportional, because the magnitude and efficiency of the response are strongly governed by the gas-liquid mass-transfer coefficient. Across 280-day dynamic simulations spanning $k_L a$ values from 60 to 1000 d^{-1} and injection rates from 0 to 2500 $m^3.d^{-1}$, CH_4 production increased with injection under every condition tested, yet H_2 conversion efficiency ranged from 57-61% at the lowest transfer intensity to above 96% at the highest. This divergence has a direct physical explanation: hydrogen's low solubility means that, under weak transfer conditions, additional H_2 supply outpaces the rate at which H_2 dissolves into the liquid phase where hydrogenotrophic methanogens can access it. The result is intensified H_2 slip and a declining CH_4 fraction in the outlet gas the reactor delivers more CH_4 in absolute terms but does so less efficiently and at lower gas quality. Under strong mass-transfer conditions, the physical dissolution limitation becomes less dominant, while other process constraints, including inorganic carbon availability, pH, and biological conversion capacity, become increasingly relevant. The pH and soluble inorganic carbon dynamics observed across the scenario matrix support this interpretation: CO_2 depletion and pH elevation toward 8.1 under the highest $k_L a$ and H_2 injection conditions indicated that inorganic carbon availability became an important constraint when mass transfer was no longer limiting, confirming the relevance of the S_{IC} limitation factor introduced in the modified model., confirming the necessity of the S_{IC} limitation factor introduced in the modified model rather than treating it as a precautionary assumption. The injection-efficiency evaluation framework, constructed from complementary performance indicators, translated these mechanistic trends into a structured operational envelope for each mass-transfer regime and provided the screening basis for selecting representative scenarios before machine-learning dataset preparation.

The ML comparison produced a clear performance hierarchy of SVR, ELM, RF, and EPR for both prediction targets. SVR achieved the strongest overall performance, with testing R^2 values of 0.996 for CH_4 and 0.9948 for biogas, together with the lowest prediction errors among the evaluated algorithms. This performance can be attributed to the suitability of the RBF kernel for representing the smooth and nonlinear relationships present in the ADM1-derived process data. ELM achieved the second-highest accuracy, with testing R^2 values of 0.9899 for CH_4 and 0.9834 for biogas. Its strong performance was supported by input standardization, which placed predictors with different numerical ranges on comparable scales before the sigmoid transformation, together with the

selected hidden-layer configuration of 250 neurons. RF also demonstrated strong predictive capability, achieving testing R^2 values of 0.9879 and 0.9807 for CH₄ and biogas, respectively. Its slightly lower accuracy relative to SVR and ELM can be associated with the piecewise prediction structure of tree-based regression when approximating the relatively smooth process responses generated by the mechanistic model. EPR achieved testing R^2 values of 0.9561 for CH₄ and 0.9276 for biogas. Although its predictive accuracy was lower than that of the other three models, EPR provided explicit mathematical expressions and therefore represented a useful trade-off between predictive accuracy and model interpretability. The small differences between training and testing performance, together with the repeated-random partitioning analysis, indicated that the comparative performance of the models was stable for held-out observations within the investigated operating domain.

On one hand, the findings reframe the design priority for in-situ biomethanation systems: improving gas-liquid mass transfer should precede any increase in H₂ supply capacity, because additional H₂ delivered under weak transfer conditions worsens the efficiency balance rather than improving it. From a predictive perspective, SVR provided the strongest overall accuracy, while the explicit EPR equations offer a complementary option when model transparency and computational simplicity are prioritized. The predictive models developed in this study represent the operating domain covered by the modified ADM1 simulations, including the investigated mesophilic conditions, dynamic influent characteristics, H₂ injection rates, and gas-liquid mass-transfer scenarios. Their application to substantially different reactor configurations or operating domains would therefore require process data representative of those conditions.

In addition, time-series prediction models could be used to forecast future reactor performance. Future work should further improve the representation of gas-liquid mass transfer in the biomethanation model. In this thesis, the $k_L a$ was treated as a scenario-based variable to represent different transfer intensities. However, in digesters, mass transfer is affected by numerous factors including operating conditions. Therefore, future studies could link the ADM1/BSM2-based biomethanation model with experimentally measured $k_L a$ or computational fluid dynamics.

Another important future direction is the development of dynamic H₂ injection strategies. In the present study, H₂ was injected at fixed rates for scenario comparison. In real operation, however, the optimum H₂ injection rate may change over time depending on influent composition, CH₄

production, outlet H_2 concentration, pH, S_{IC} availability, and renewable electricity supply. Future studies could therefore develop control strategies in which H_2 injection is adjusted dynamically based on process indicators. This could help reduce H_2 slip, prevent excessive pH increase, protect inorganic carbon availability, and improve the overall efficiency of in-situ HBM. The modelling framework should also be tested under broader operating conditions and feedstock types. The present thesis was based on a specific dynamic influent structure and mesophilic operating condition. Future work could apply the same framework to different substrates, including sewage sludge, food waste, manure, agricultural residues, and co-digestion mixtures, and could also examine the influence of different HRT, OLR, and temperature regimes, including thermophilic operation. This would help determine whether the model structure and selected input variables remain reliable across different AD systems.

Future studies should also include energy-balance and techno-economic analysis. Although this thesis identified technically efficient H_2 injection and $k_L a$ scenarios, practical implementation depends on the energy and cost required for H_2 production, gas compression, sparging, mixing, and process control. The additional CH_4 gained through biomethanation should therefore be compared with the energy input and economic cost of supplying and transferring H_2 . In addition, the source of H_2 should be considered explicitly, since the environmental benefit of in-situ biomethanation depends strongly on whether H_2 is produced from renewable electricity or from fossil-based pathways. Future work could connect the AD model with renewable H_2 production pathways such as alkaline electrolysis or PEM electrolysis powered by surplus wind or solar electricity.

Lastly, the ML part of the work could also be expanded. In this thesis, the ML models were mainly used to predict CH_4 and biogas flow rates from simulation-generated data. Future studies could develop multi-output models capable of predicting several important process variables at the same time, such as CH_4 flow rate, biogas flow rate, H_2 slip, pH, S_{IC} , and CH_4 content. As a separate future objective, time-series forecasting models could also be developed to predict reactor performance several hours or days ahead using historical process trajectories. This would represent a different prediction task from the state-based regression framework adopted in the present study.

References

- Acosta-Pavas, J. C., Morchain, J., Dumas, C., Ngu, V., Arnaud, C., & Aceves-Lara, C. A. (2022). Towards Anaerobic Digestion (ADM No.1) Model's Extensions and Reductions with In-situ Gas Injection for Biomethane Production. *IFAC-PapersOnLine*, 55(20), 635–640. <https://doi.org/10.1016/j.ifacol.2022.09.167>
- Acosta-Pavas, J. C., Robles-Rodríguez, C. E., Morchain, J., Dumas, C., Cockx, A., & Aceves-Lara, C. A. (2023). Dynamic modeling of biological methanation for different reactor configurations: An extension of the anaerobic digestion model No. 1. *Fuel*, 344. <https://doi.org/10.1016/j.fuel.2023.128106>
- Aromolaran, A. (2021). *Enhancement of biogas production from organic wastes through leachate blending and co-digestion* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).
- Alex, J., Benedetti, L., Copp, J., Gernaey, K. V., Jeppsson, U., Nopens, I., Pons, M.-N., Rosen, C., Steyer, J.-P., & Vanrolleghem, P. (2018). *Benchmark Simulation Model no. 2 (BSM2)*.
- Almeida Streitwieser, D. (2017). Comparison of the anaerobic digestion at the mesophilic and thermophilic temperature regime of organic wastes from the agribusiness. *Bioresource Technology*, 241, 985–992. <https://doi.org/10.1016/j.biortech.2017.06.006>
- Angelidaki, I., Treu, L., Tsapekos, P., Luo, G., Campanaro, S., Wenzel, H., & Kougias, P. G. (2018). Biogas upgrading and utilization: Current status and perspectives. *Biotechnology Advances*, 36(2), 452–466. <https://doi.org/10.1016/j.biotechadv.2018.01.011>
- Antukh, T., Lee, I., Joo, S., & Kim, H. (2022). Hydrogenotrophs-based biological biogas upgrading technologies. *Frontiers in Bioengineering and Biotechnology*, 10, 833482. <https://doi.org/10.3389/fbioe.2022.833482>
- Asadi, M., Guo, H., & McPhedran, K. (2020). Biogas production estimation using data-driven approaches for cold region municipal wastewater anaerobic digestion. *Journal of Environmental Management*, 253, 109708. <https://doi.org/10.1016/j.jenvman.2019.109708>

- Aslanzadeh, S. (2014). *Pretreatment of cellulosic waste and high rate biogas production* (Doctoral dissertation, University of Borås, School of Engineering).
<http://hdl.handle.net/2320/12853>
- Balagurusamy, N., & Chandel, A. K. (2021). Biogas production: From anaerobic digestion to a sustainable bioenergy industry. In *Biogas Production: From Anaerobic Digestion to a Sustainable Bioenergy Industry*. Springer International Publishing.
<https://doi.org/10.1007/978-3-030-58827-4>
- Batstone, D. J., Keller, J., Angelidaki, I., Kalyuzhnyi, S. V., Pavlostathis, S. G., Rozzi, A., Sanders, W. T. M., Siegrist, H., & Vavilin, V. A. (2002). *The IWA Anaerobic Digestion Model No. 1 (ADM1)*. IWA Publishing.
- Beil, M., & Beyrich, W. (2013). Biogas upgrading to biomethane. In *The Biogas Handbook: Science, Production and Applications* (pp. 342–377). Elsevier Inc.
<https://doi.org/10.1533/9780857097415.3.342>
- Bensmann, A., Hanke-Rauschenbach, R., Heyer, R., Kohrs, F., Benndorf, D., Reichl, U., & Sundmacher, K. (2014). Biological methanation of hydrogen within biogas plants: A model-based feasibility study. *Applied Energy*, 134, 413–425.
<https://doi.org/10.1016/j.apenergy.2014.08.047>
- Bertacchi, S., Ruusunen, M., Sorsa, A., Sirviö, A., & Branduardi, P. (2021). Mathematical analysis and update of adm1 model for biomethane production by anaerobic digestion. *Fermentation*, 7(4). <https://doi.org/10.3390/fermentation7040237>
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227.
<https://doi.org/10.1007/s11749-016-0481-7>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Chen, X., Chen, Z., Wang, X., Huo, C., Hu, Z., Xiao, B., & Hu, M. (2016). Application of ADM1 for modeling of biogas production from anaerobic digestion of *Hydrilla verticillata*. *Bioresource Technology*, 211, 101–107. <https://doi.org/10.1016/j.biortech.2016.03.002>

- Da Costa Gomez, C. (2013). Biogas as an energy option: an overview. In A. Wellinger, J. Murphy, & D. Baxter (Eds.), *The Biogas Handbook: Science, Production and Applications* (pp. 1–16). Woodhead Publishing. <https://doi.org/10.1533/9780857097415.1>
- De Clercq, D., Wen, Z., Fei, F., Caicedo, L., Yuan, K., & Shang, R. (2020). Interpretable machine learning for predicting biomethane production in industrial-scale anaerobic co-digestion. *Science of the Total Environment*, 712. <https://doi.org/10.1016/j.scitotenv.2019.134574>
- Ding, S., Zhao, H., Zhang, Y., Xu, X., & Nie, R. (2015). Extreme learning machine: algorithm, theory and applications. *Artificial Intelligence Review*, 44(1), 103–115. <https://doi.org/10.1007/s10462-013-9405-z>
- Parichehreh Dizaji, P. (2023). *Biogas production through bio-methanation of syngas* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).
- Dong, C., & Chen, J. (2019). Optimization of process parameters for anaerobic fermentation of corn stalk based on least squares support vector machine. *Bioresource Technology*, 271, 174–181. <https://doi.org/10.1016/j.biortech.2018.09.085>
- Fu, S., Angelidaki, I., & Zhang, Y. (2021). In situ Biogas Upgrading by CO₂-to-CH₄ Bioconversion. In *Trends in Biotechnology* (Vol. 39, Number 4, pp. 336–347). Elsevier Ltd. <https://doi.org/10.1016/j.tibtech.2020.08.006>
- García Diéguez, C. (2009). *Modelling and control of anaerobic wastewater treatment processes Memoria presentada por*.
- Ge, Y., Tao, J., Wang, Z., Mu, L., Guo, W., Cheng, Z., Yan, B., Shi, Y., Su, H., & Chen, G. (2024). A hybrid approach of anaerobic digestion model no. 1 and machine learning to model and optimize continuous anaerobic digestion processes. *Biomass and Bioenergy*, 184. <https://doi.org/10.1016/j.biombioe.2024.107176>
- Giustolisi, O., & Savic, D. A. (2006). A symbolic data-driven technique based on evolutionary polynomial regression. *Journal of Hydroinformatics*, 8(3), 207–222.
- Gomez, C. D. C. (2013). Biogas as an energy option: an overview. *The Biogas Handbook*, 1–16.

- Hammad, M. A., Jereb, B., Rosi, B., & Dragan, D. (2020). Methods and Models for Electric Load Forecasting: A Comprehensive Review. *Logistics & Sustainable Transport*, 11(1), 51–76. <https://doi.org/10.2478/jlst-2020-0004>
- Hornung, Andreas. (2014). *Transformation of Biomass : Theory to Practice*. Wiley.
- Ijäs, M. (2022). *BIOMETHANATION POTENTIAL IN BIOGAS AND BIOETHANOL PLANTS IN FINLAND*.
- Jensen, M. B., Jensen, B., Ottosen, L. D. M., & Kofoed, M. V. W. (2021). Integrating H₂ injection and reactor mixing for low-cost H₂ gas-liquid mass transfer in full-scale in situ biomethanation. *Biochemical Engineering Journal*, 166. <https://doi.org/10.1016/j.bej.2020.107869>
- Jiang, H., Hao, W., Li, Y., & Zhou, H. (2022). Biological methanation of H₂ and CO₂ in a continuous stirred tank reactor. *Journal of Cleaner Production*, 370. <https://doi.org/10.1016/j.jclepro.2022.133518>
- Jimenez, J., Latrille, E., Harmand, J., Robles, A., Ferrer, J., Gaida, D., Wolf, C., Mairet, F., Bernard, O., Alcaraz-Gonzalez, V., Mendez-Acosta, H., Zitomer, D., Totzke, D., Spanjers, H., Jacobi, F., Guwy, A., Dinsdale, R., Premier, G., Mazhegrane, S., ... Steyer, J. P. (2015). Instrumentation and control of anaerobic digestion processes: a review and some research challenges. In *Reviews in Environmental Science and Biotechnology* (Vol. 14, Number 4, pp. 615–648). Springer Netherlands. <https://doi.org/10.1007/s11157-015-9382-6>
- Lecker, B., Illi, L., Lemmer, A., & Oechsner, H. (2017). Biological hydrogen methanation – A review. In *Bioresource Technology* (Vol. 245, pp. 1220–1228). Elsevier Ltd. <https://doi.org/10.1016/j.biortech.2017.08.176>
- Lettinga, G., Rebac, S., & Zeeman, G. (2001). Challenge of psychrophilic anaerobic wastewater treatment. In *TRENDS in Biotechnology* (Vol. 19, Number 9). <http://tibtech.trends.com0167>
- Ling, J. Y. X., Chan, Y. J., Chen, J. W., Chong, D. J. S., Tan, A. L. L., Arumugasamy, S. K., & Lau, P. L. (2024). Machine learning methods for the modelling and optimisation of biogas production from anaerobic digestion: a review. In *Environmental Science and Pollution*

- Research* (Vol. 31, Number 13, pp. 19085–19104). Springer.
<https://doi.org/10.1007/s11356-024-32435-6>
- Liu, L., Tian, Y., Zhao, J., Xia, Z., Wang, N., & Wang, D. (2025). Machine learning for predicting methane production and optimizing parameter in anaerobic digestion process. *Fuel*, 396. <https://doi.org/10.1016/j.fuel.2025.135206>
- Long, F., Wang, L., Cai, W., Lesnik, K., & Liu, H. (2021). Predicting the performance of anaerobic digestion using machine learning algorithms and genomic data. *Water Research*, 199. <https://doi.org/10.1016/j.watres.2021.117182>
- Lovato, G., Alvarado-Morales, M., Kovalovszki, A., Peprah, M., Kougias, P. G., Rodrigues, J. A. D., & Angelidaki, I. (2017). In-situ biogas upgrading process: Modeling and simulations aspects. *Bioresource Technology*, 245, 332–341.
<https://doi.org/10.1016/j.biortech.2017.08.181>
- Luo, G., & Angelidaki, I. (2013). Hollow fiber membrane based H₂ diffusion for efficient in situ biogas upgrading in an anaerobic reactor. *Applied Microbiology and Biotechnology*, 97(8), 3739–3744. <https://doi.org/10.1007/s00253-013-4811-3>
- Marasco, S., Fiore, A., Greco, R., Cimellaro, G. P., & Marano, G. C. (2021). Evolutionary Polynomial Regression Algorithm Enhanced with a Robust Formulation: Application to Shear Strength Prediction of RC Beams without Stirrups. *Journal of Computing in Civil Engineering*, 35(6). [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000985](https://doi.org/10.1061/(asce)cp.1943-5487.0000985)
- Crolla, A. M. (2017). *Anaerobic Digestion of Dairy Manure with Food and Industry Wastes–Enhanced Biogas Production and Digestate Quality* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).
- Martín-Hernández, E., Chen, K., Santoro, D., & Omelon, S. (2025). Decarbonizing anaerobic digestion in Canada’s wastewater resource recovery facilities: an opportunity to attain carbon-neutral biogas production and its potential to offset waste management sector carbon emissions. *RSC Sustainability*. <https://doi.org/10.1039/d5su00688k>
- Meegoda, J. N., Li, B., Patel, K., & Wang, L. B. (2018). A review of the processes, parameters, and optimization of anaerobic digestion. In *International Journal of Environmental*

Research and Public Health (Vol. 15, Number 10). MDPI AG.

<https://doi.org/10.3390/ijerph15102224>

Mo, R., Guo, W., Batstone, D., Makinia, J., & Li, Y. (2023). Modifications to the anaerobic digestion model no. 1 (ADM1) for enhanced understanding and application of the anaerobic treatment processes – A comprehensive review. In *Water Research* (Vol. 244). Elsevier Ltd. <https://doi.org/10.1016/j.watres.2023.120504>

Murphy, J. D., & Thamsiriroj, T. (2013). Fundamental science and engineering of the anaerobic digestion process for biogas production. *The Biogas Handbook: Science, Production and Applications*, 103(3), 104–130. <https://doi.org/10.1533/9780857097415.1.104>

Ncube, A., Cocker, J., Ellis, D., & Fiorentino, G. (2021). Biogas from source separated organic waste within a circular and life cycle perspective. A case study in Ontario, Canada. *Environmental and Sustainability Indicators*, 11. <https://doi.org/10.1016/j.indic.2021.100134>

Neuner, T., Meister, M., Pillei, M., Senfter, T., Draxl-Weiskopf, S., Ebner, C., Winkler, J., & Rauch, W. (2024). Impact of Design and Mixing Strategies on Biogas Production in Anaerobic Digesters. *Water (Switzerland)*, 16(15). <https://doi.org/10.3390/w16152205>

Nguyen, H. H. (2014). *Modelling of food waste digestion using ADM1 integrated with Aspen Plus*.

Norouzi, O., & Dutta, A. (2022). The Current Status and Future Potential of Biogas Production from Canada's Organic Fraction Municipal Solid Waste. In *Energies* (Vol. 15, Number 2). MDPI. <https://doi.org/10.3390/en15020475>

Pauss, A., Andre, G., Perrier, M., & Guiot, S. R. (1990). Liquid-to-Gas Mass Transfer in Anaerobic Processes: Inevitable Transfer Limitations of Methane and Hydrogen in the Biomethanation Process. In *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* (Vol. 56, Number 6). <https://journals.asm.org/journal/aem>

Poggio, D. A. (2015). *Modification and experimental calibration of ADM1 for modelling the anaerobic digestion of solid wastes in demand driven applications*.

- Poggio, D., Sastraatmaja, A., Walker, M., Michailos, S., Nimmo, W., & Pourkashanian, M. (2023). Experimental Evaluation of Continuous In-Situ Biomethanation of CO₂ in Anaerobic Digesters Fed on Sewage Sludge and Food Waste and the Influence of Hydrogen Gas–Liquid Mass Transfer. *Processes*, 11(2). <https://doi.org/10.3390/pr11020604>
- Pullen, T. (2015). *Anaerobic Digestion-Making Biogas-Making Energy*.
- Rosen, C., & Jeppsson, U. (2006a). *Aspects on ADM1 implementation within the BSM2 framework*.
- Sadrimajd, P., Mannion, P., Howley, E., & Lens, P. N. L. (2021a). *PyADM1: a Python implementation of Anaerobic Digestion Model No. 1*. <https://doi.org/10.1101/2021.03.03.433746>
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. In *Babylonian Journal of Machine Learning* (Vol. 2024, pp. 69–79). Mesopotamian Academic Press. <https://doi.org/10.58496/BJML/2024/007>
- Sanjay Shah. (2016). *Methane from Syngas by Anaerobic Digestion*. <http://www.usn.no>
- Sastraatmaja, A., Pourkashanian, M., Ingham, D., Poggio, D., & Walker, M. (2022). *Experimental Process Development and Techno-economic Assessment of In-Situ Biomethanation of Carbon Dioxide in Biowaste Anaerobic Digesters*.
- Shah, S., Bergland, W. H., & Bakke, R. (2017). Methane from Syngas by Anaerobic Digestion. *Proceedings of the 58th Conference on Simulation and Modelling (SIMS 58) Reykjavik, Iceland, September 25th – 27th, 2017*, 138, 114–120. <https://doi.org/10.3384/ecp17138114>
- Shahin, M. A. (2016). State-of-the-art review of some artificial intelligence applications in pile foundations. *Geoscience Frontiers*, 7(1), 33–44. <https://doi.org/10.1016/j.gsf.2014.10.002>
- Singh, B., Szamosi, Z., & Siménfalvi, Z. (2020). Impact of mixing intensity and duration on biogas production in an anaerobic digester: a review. In *Critical Reviews in Biotechnology* (Vol. 40, Number 4, pp. 508–521). Taylor and Francis Ltd. <https://doi.org/10.1080/07388551.2020.1731413>

- Wang, J., Lu, S., Wang, S. H., & Zhang, Y. D. (2022). A review on extreme learning machine. *Multimedia Tools and Applications*, 81(29), 41611–41660. <https://doi.org/10.1007/s11042-021-11007-7>
- Wang, L., Long, F., Liao, W., & Liu, H. (2020). Prediction of anaerobic digestion performance and identification of critical operational parameters using machine learning algorithms. *Bioresource Technology*, 298. <https://doi.org/10.1016/j.biortech.2019.122495>
- Yenigün, O., & Demirel, B. (2013). Ammonia inhibition in anaerobic digestion: A review. In *Process Biochemistry* (Vol. 48, Numbers 5–6, pp. 901–911). <https://doi.org/10.1016/j.procbio.2013.04.012>
- Yildirim, O., & Ozkaya, B. (2023). Prediction of biogas production of industrial scale anaerobic digestion plant by machine learning algorithms. *Chemosphere*, 335. <https://doi.org/10.1016/j.chemosphere.2023.138976>
- Zhang, F., & O'Donnell, L. J. (2019). Support vector regression. In *Machine Learning: Methods and Applications to Brain Disorders* (pp. 123–140). Elsevier. <https://doi.org/10.1016/B978-0-12-815739-8.00007-9>
- Zhang, S., Bi, X. T., & Clift, R. (2015). Life cycle analysis of a biogas-centred integrated dairy farm-greenhouse system in British Columbia. *Process Safety and Environmental Protection*, 93, 18–30. <https://doi.org/10.1016/j.psep.2014.02.017>

Appendix 1

Table S1. variables and parameters based on (Bensmann et al., 2014)

Stoichiometric parameter	Value	Kinetic parameter	Value
$f_{sl,xc}$	0.0253	k_{dis}	$0.5 d^{-1}$
$f_{xl,xc}$	0.0709	$k_{hyd,ch}$	$10 d^{-1}$
$f_{ch,xc}$	0.7852	$k_{hyd,pr}$	$10 d^{-1}$
$f_{pr,xc}$	0.0932	$k_{hyd,li}$	$10 d^{-1}$
$f_{li,xc}$	0.0253	$k_{dec,i}$	$0.02 d^{-1}$
N_{xc}	0.0007	$k_{S,nh3,all}$	$10^{-4} kmol.m^{-3}$
$f_{fa,li}$	0.95	$pH_{UL,5-10}$	5.5
$f_{H_2,su}$	0.19	$pH_{LL,5-10}$	4
$f_{bu,su}$	0.13	$k_{m,su}$	$30 COD COD^{-1}d^{-1}$
$f_{pro,su}$	0.27	$K_{S,su}$	$0.5 kgCOD.m^3$
$f_{ac,su}$	0.41	$k_{m,aa}$	$50 COD COD^{-1}d^{-1}$
$f_{H_2,aa}$	0.06	$K_{S,aa}$	$0.3 kgCOD m^3$
N_{aa}	0.007	$k_{m,fa}$	$6 COD COD^{-1}d^{-1}$
$f_{va,aa}$	0.23	$K_{S,fa}$	$0.4 kgCOD m^3$
$f_{bu,aa}$	0.26	$K_{I,H_2,fa}$	$5 \times 10^{-6} kgCOD m^3$
$f_{pro,aa}$	0.05	$k_{m,c4}$	$20 COD COD^{-1}d^{-1}$
$f_{ac,aa}$	0.4	$K_{S,c4}$	$0.3 kgCOD m^3$
Y_{su}	0.1	$K_{I,H_2,c4}$	$1 \times 10^{-5} kgCOD m^3$
Y_{aa}	0.08	$k_{m,pro}$	$4 COD COD^{-1}d^{-1}$
Y_{fa}	0.06	k_{dis}	$0.1 kgCOD m^3$
Y_{c4+}	0.06	$K_{I,H_2,pro}$	$3.5 \times 10^{-6} kgCOD m^3$
Y_{pro}	0.04	$k_{m,ac}$	$4.1 COD COD^{-1}d^{-1}$
Y_{ac}	0.05	$K_{S,ac}$	$0.15 kgCOD m^3$
Y_{H_2}	0.06	$pH_{UL,11}$	7
		$pH_{LL,11}$	6
		$K_{I,nh3,ac}$	$0.0018 mol m^{-3}$
		k_{m,H_2}	$5 COD COD^{-1}d^{-1}$
		K_{S,H_2}	$7 \times 10^{-6} kgCOD m^3$
		$pH_{UL,12}$	6
		$pH_{LL,12}$	5