Modeling and dynamic optimization of biomethanation : dynamic and multiphysics study of biological couplings and gas-liquid transfers to optimize the operation of industrial reactors

Juan Camilo Acosta Pavas

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut National des Sciences Appliquées de Toulouse (INSA Toulouse)

Présentée et soutenue par

Juan Camilo ACOSTA PAVAS

Le 11 septembre 2023

Modélisation de réacteur de Biométhanation : étude multiphysique des couplages biologiques et des transferts gaz-liquide pour l'optimisation du design et de la conduite de réacteurs industriels

---

Ecole doctorale : **MEGEP-Mécanique, Energétique, Génie civil, Procédés**
Spécialité : **Génie des Procédés et de l'Environnement**

Unité de recherche : **TBI - Toulouse Biotechnology Institute, Bio & Chemical Engineering**

Thèse dirigée par
Jérôme MORCHAIN et césar Arturo ACEVES LARA

Jury
Arnaud COCKX, Président du jury
Céline CASENAVE, Rapportrice
Jean-Philippe STEYER, Rapporteur
Carlos Eduardo Robles Rodriguez, Examinateur
Oscar Andrés Prado Rubio, Examinateur
David GRIOL BARRES, Examinateur

Abstract

This thesis aims at studying biological methanation to find the optimal conditions to produce high purity biomethane as a value-added product. The objective is addressed from a modeling point of view, based on the use of model-based control strategies and data-driven soft sensors. A bibliography synthesis was carried out to set the theoretical framework that includes dynamic models, control strategies, and monitoring tools applied to biological methanation. An extension of the Anaerobic Digestion Model No.1 (ADM1_ME) was proposed to describe the dynamics of the biological methanation process with the use of syngas (H_2, CO_2, and CO) as substrate. The variation of the volumetric mass transfer coefficient is considered as a function of two types of reactors, a bubble column reactor (BCR) and a Continuous Stirred Tank Reactor ($CSTR$). The ADM1_ME was accurately calibrated and validated in different operating conditions using experimental data from the literature. A Multi-Objective Dynamic Optimization (MODO) strategy was proposed to optimize the biological methanation performance. The MODO strategy was designed to consider three different objective functions to maximize: (i) yield (Y_{CH4}) and productivity (P_{CH4}) of methane, (ii) Y_{CH} and P_{CH4} simultaneously complemented by a switch to maximize acetate yields (Y_{ac}) and productivities (P_{ac}), and (iii) economic optimality in terms of ($Gain$) and ($Profit\ margin$). The results demonstrated the feasibility of the MODO strategy and its robustness to switch between products of interest and the key role of the manipulated variables (*i.e.*, inlet liquid and gas flow rates) in the biological methanation process. Furthermore, data-driven soft sensors were applied to detect deviations from the optimal operation points when disturbances occurred in the manipulated variables. Specifically, Support Vector Machine (SVM) showed promising results and a potential application by using 2D visualizations constructed by pair of features.

Keywords: Biological Methanation, Dynamic Modeling, Muti-Objective Optimization, (Economic) Model Predictive Control, Soft Sensors, Machine Learning, Fault Detection

Résumé

Cette thèse vise à étudier la méthanation biologique afin de trouver les conditions optimales pour produire du biométhane de haute pureté en tant que produit à valeur ajoutée. L'objectif est abordé du point de vue de la modélisation, en se basant sur l'utilisation de stratégies de commande basées sur des modèles et de capteurs souples pilotés par des données. Une synthèse bibliographique a été réalisée pour établir le cadre théorique comprenant les modèles dynamiques, les stratégies de commande et les outils de surveillance utilisés pour la méthanation biologique. Une extension du modèle de digestion anaérobie N°1 (ADM1_ME en anglais) a été proposée pour décrire la dynamique du processus de méthanation biologique avec l'utilisation de gaz de synthèse (H_2, CO_2 et CO) comme substrat. La variation du coefficient de transfert de matière volumétrique est considérée en fonction de deux types de réacteurs, un réacteur à colonne à bulles et un réacteur à réservoir agité continu. L'ADM1_ME a été calibré avec précision et validé dans différentes conditions de fonctionnement en utilisant des données expérimentales tirées de la littérature. Une stratégie d'optimisation dynamique multi-objectifs (MODO en anglais) a été proposée pour optimiser les performances de la méthanation biologique. La stratégie MODO a été conçue pour prendre en compte trois fonctions objectives différentes afin de maximiser : (i) le rendement (Y_{CH4}) et la productivité (P_{CH4}) du méthane, (ii) maximiser Y_{CH} et P_{CH4} simultanément, complété par un commutateur pour maximiser les rendements (Y_{ac}) et les productivités (P_{ac}) de l'acétate, et (iii) l'optimalité économique en termes de $(Gain)$ et $(Profit\ margin)$. Les résultats ont démontré la faisabilité de la stratégie MODO et sa robustesse pour passer d'un produit à l'autre, ainsi que le rôle clé des variables manipulées (c'est-à-dire les débits d'entrée du liquide et du gaz) sur le processus de méthanation biologique. En outre, des capteurs souples pilotés par les données ont été appliqués pour détecter les écarts par rapport aux points de fonctionnement optimaux lorsque des perturbations se produisent dans les variables manipulées. En particulier, la machine à vecteur de support (SVM en anglais) a montré des résultats prometteurs et une application potentielle en utilisant des visualisations en 2D construites par paire de prédicteurs.

Mots clefs : Méthanation biologique, Modèle dynamique, Optimisation Muti-Objective, (Economique) commande prédictive basée modèle, Apprentissage automatique, Capteurs souples, Détection des défauts

"All models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind"

George Edward Pelham Box

Acknowledgment

I want to thank all the people who contributed to the development of this thesis.

Initially, I would like to thank all the members of the jury. I am honored that its members have agreed to judge this thesis.

I want to thank César Aceves and Jérôme Morchain for their valuable knowledge in the direction of this thesis.

I want to thank Cesar Aceves for his confidence and friendship. Thank you for all the support, patience, and advice on academic and personal aspects.

Special thanks to Carlos Robles for all his advice and constructive comments on this thesis.

Thank you to all the people who have collaborated on the different works throughout this thesis: David Corrales, Arnaud Cockx, Claire Dumas, Camilo Suarez, David Griol, and Zoraida Callejas.

I want to thank all the institutions that contributed to the support of this thesis. Ministerio de Ciencias, Tecnología e Innovación (MINCIENCIAS) through the Scholarship Program No. 860. Agencia de Educación Postsecundaria de Medellín (SAPIENCIA) through the program Enlaza Mundos 2020-2. This work has benefited from a State grant managed by the National Research Agency under the "Investissements d'Avenir" programme with the reference ANR-18-EURE-0021.

I want to thank my family for their support in the distance. Thank you for every word of encouragement.

Finally, I want to thank Susana for all the love, patience and always being there to support me.

Publications and Communications

International Journals

Acosta-Pavas, J. C., Robles-Rodríguez, Carlos. 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, 128106. https://doi.org/10.1016/j.fuel.2023.128106

Acosta-Pavas, J. C., Robles-Rodríguez, C. E., Méndez Suarez, C. A., Morchain, J., Dumas, C., Cockx, A., & Aceves-Lara, C. A. (2022). Dynamic Multi-Objective Optimization Applied to Biomethanation Process. *Chemical Engineering Transactions*, 96, 319–324. https://doi.org/10.3303/CET2296054

Blaiech K., **Acosta-Pavas J.C.**, Ben Gaida L., Gannoun H., Liebgott P.P., Aceves-Lara C.A. (2023). Multi-Objective Optimization of Thermophilic Biohydrogen Production. *Chemical Engineering Transactions* (Accepted).

Acosta-Pavas, J. C., Robles-Rodríguez, C. E., Morchain, J., Corrales, C., Dumas, C., Cockx, A., & Aceves-Lara, C. A. (2023). Economic Multi-Objective Dynamic Optimization (EMODO) as a Decision-Making tool in Biomethanation Process. *Chemical Engineering Transactions* (Accepted).

Book Chapter

Acosta-Pavas, J. C., Griol, D., Callejas, Z., Corrales, D. C., Robles-Rodríguez, C. E., Morchain, J., & Aceves-Lara, C. A. (2023). Fault Detection in Biological Methanation Process Using Machine Learning: A Comparative Study of Different Algorithms. In P. García Bringas, et al. (eds.), *18th International Conference on Soft Computing Models in Industrial and Environmental Applications. Lecture Notes in Networks and Systems* (Vol. 749, pp. 132-142). Springer, Cham. https://doi.org/10.1007/978-3-031-42529-5_13

Acosta-Pavas, J. C., Robles-Rodríguez, C. E., Morchain, J., Corrales, D. C., Dumas, C., Cockx, A., & Aceves-Lara, C. A. (2023). Switching Multi-Objective Dynamic Optimization (MODO) for the Production of Value-Added Products. In A. C. Kokossis, M. C. Georgiadis, E. Pistikopoulos (eds.), *33rd European Symposium on*

Computer-Aided Process Engineering (Vol. 52, pp. 583-588). Elsevier. https://doi.org/10.1016/B978-0-443-15274-0.50092-5

Acosta-Pavas, J.C., Robles-Rodríguez, C.E., Dumas, C., Cockx, A., Morchain, J., Aceves-Lara, C.A. (2022). Use of Support Vector Machine to Fault Detection in Biomethanation Process. In S. Omatu, R. Mehmood, P. Sitek, S. Cicerone, S. Rodríguez (eds), *Distributed Computing and Artificial Intelligence, 19th International Conference*. *Lecture Notes in Networks and Systems* (Vol. 583, pp. 176–186). Springer, Cham. https://doi.org/10.1007/978-3-031-20859-1_18

International Conferences with Online Proceedings

Acosta-pavas, J. C., Morchain, J., Dumas, C., Ngu, V., Cockx, A., Aceves-Lara C.A. (2022). Towards Anaerobic Digestion (ADM No. 1) Model's Extensions and Reductions with In-situ Gas Injection for Biomethane Production. In A. Kugi, A. Körner, W. Kemmetmüller, A. Deutschmann-Olek, F. Breitenecker, I. Troch (eds.), *10th Vienna International Conference on Mathematical Modelling* (Vol. 55(20), pp. 635-6409). IFAC-PapersOnLine. https://doi.org/10.1016/j.ifacol.2022.09.167

Index of Content

List of Figures

List of Tables

Nomenclature and Abbreviations

Nomenclature

P_{atm}: Atmospheric pressure
$\boldsymbol{g_i}$: Associated response to the input vectors
W_0: Bias term in an SVM model
X_k: Concentration of biomass k in the liquid phase
$S_{liq,j}$: Concentration of the component j in the liquid phase
$S_{gas,i}$: Concentration of component i in the gas phase
$V(\xi_{\mathbf{u}}(\theta_{\mathbf{u}}))$: Conditional variance for the subvector $\theta_{\mathbf{u}}$ for $\mathbf{u} \subset \{1, \dots n\}$
b_0, b_1, and b_2: Constant parameters in $k_L a_{,i}$ equations
u: Control variables/manipulated variables
$\gamma_{COD,i}$: Conversion factor between the $moles$ and $gCOD$ of a component
φ: Cost function that evaluates the final estate of the system
Tc: Critical temperature
Pc: Critical pressure
Vc: Critical molar volume
$\mathcal{D}$: Dataset
$K_{k,dec}$: Decay biomass constant of biomass k
$\mu_{k,dec}$: Decay rate of biomass k
D_i: Diffusivity of component i
δ: Distance metric in machine learning algorithms
EE_i^j: Elementary effects for each parameter i over the j^{th} parameter disturbance
E: Entropy of the dataset $\mathcal{D}$
$\hat{q}_{gas,i}$: Experimental measure of the outlet gas flow rate of component i
$\hat{p}_{gas,i}$: Experimental measure of the percent composition of component i in the gas phase
SI_i: First-order sensitivity index
SI_{is}: First-order sensitivity indices with a scalar characteristic
$Gain^*$: Gain obtained from the MOO
V_{gas}: Gas volume
$Power$: Gassed power input
$Gini$: Gini index
μ_k: Growth rate of biomass k
ΔH_c°: Heat combustion
H_i: Henry's law equilibrium constant*
$F = \Phi(X)$: High-dimensional feature space
λ_i and ψ_i: Inequality and equality restrictions of the objective function
G: Information gain
$KI_{H2,k}$: Inhibition constants over butyrate, propionate, and acetate due to H_2
$KI_{CO,k}$: Inhibition constants over acetate and H_2 due to CO
$S_{liq,j}^{in}$: Inlet concentration of the component j in the liquid phase
$S_{gas,i}^{in}$: Inlet concentration of component i in the gas phase
X_k^{in}: Inlet concentration of biomass k in the liquid phase
q_{gas}^{in}: Inlet gas flow rate
q_{liq}^{in}: Inlet liquid flow rate
COD_{bio}^{in}: Inlet organic loading rate as a biomass
COD_{liq}^{in}: Inlet organic loading rate in liquid phase
COD_{gas}^{in}: Inlet organic loading rate in gas phase
Φ: Input-output mapping in SVM algorithms
$\boldsymbol{X} = (\boldsymbol{x_i}, \boldsymbol{g_i})$: Input space to train machine learning algorithms

L: Instantaneous cost function that measures the performance at each instant time t
$\boldsymbol{x_i}$: Input vector with N instances and p features (training data)
K: Kernel function in SVM models
V_{liq}: Liquid volume
ℓ: Loss function to measure the quality of the prediction of the machine learning model $s(x)$
u^L and u^U: Lower and upper limits of the model inputs
$s(x)$: Machine learning model
b: Margin hyperplane in SVM
C_i: Measurement error covariance matrix
N_i: Mass transfer rate of component i
d_{max}: Maximum Euclidean length from the origin on the normalized coordinates
$\mu_{m,k}$: Maximum specific growth rate of biomass k
$\hat{Y}_{mean}$: Mean experimental model data
μ_i^j : Mean of the absolute value of the distribution of the elementary effects.
SI_n^m: Measures the sensitivity of the state Y_m concerning the parameter θ_n
$Gain$: CH_4 and acetate gain
$profit\ margin$: CH_4 and acetate profit margin
Y: Model outputs of the dynamic model dY/dt defined as $Y = (Y_1, Y_2, \dots, Y_m)$
$\partial Y/\partial\theta$: Model outputs partial derivatives to the parameters
ξ: Model system
Cp: Molar heat capacity
$\theta_{i,nom}$: Nominal value of the parameter θ_i
μ_i^*: Normalized mean of the distribution of the elementary effects
n_{OP1} and n_{OP2}: Number of observations in OP1 and OP2, respectively
θ: n-dimensional parameter vector defined as $\theta = (\theta_1, \theta_2, \dots, \theta_n)$
J: Objective function
J^*: Objective function determined in the MOO
$q_{gas,i}$: Outlet gas flow rate of component i
COD_{liq}^{out}: Outlet organic loading rate in liquid phase
COD_{gas}^{out}: Outlet organic loading rate in gas phase
COD_{bio}^{out}: Outlet organic loading rate as a biomass
$P_{gas,i}$: Partial pressure of component i
$W_{u,1}$ and $W_{u,2}$: Parameters that weight the importance of the control effort
K_P: Parameter related to the friction in the gas outlet
$p_{gas,i}$: Percent composition of component i in the gas phase
H_p and H_c: Prediction and control horizons
P: Probability function
t_i: Process time
P_{ac}: Productivity of acetate
P_{CH4}: Productivity of CH_4
P_{ac}^*: Productivity of acetate obtained from the MOO
P_{CH4}^*: Productivity of CH_4 obtained from the MOO
$Profit\ margin^*$: Profit margin obtained from the MOO
A: Reactor cross-sectional area
r: Reference trajectory
s^2: Residual mean square
Ks_k: Saturation constant of biomass k
SI_{ij}: Second order-sensitivity index
$SI_n^m(t_t)$: SI_i sensitivity of the state Y_m concerning the parameter θ_n at the time t_t
ζ_i: Slack variables in an SVM model measured deviation from the ideal for data i
σ_i^j: Standard deviation of the distribution of the elementary effects
$\Delta_f H^\circ$: Standard molar enthalpy of formation

$\Delta_f G^\circ$: Standard molar Gibbs energy of formation
So: Standard molar entropy
Δ_i: Step between two consecutive input space points of the trajectory
$f_{j,k}$: Stoichiometric coefficients
$t_{N-p}^{\alpha/2}$: Student-t distribution for a confidence level specified as $100(1-\alpha)$ %
U_G: Superficial gas velocity
W_i: Square matrix with weighting coefficients (relative weight in cost functions)
T: Temperature of the process
q_{gas}: Total outlet gas flow rate
P_{gas}: Total pressure
SI_i^T: Total Sobol' índices
$V(Y)$: Total variance
R: Universal gas constant
q: Unknown sample to be classified in machine learning models
SI: Variance-based sensitivity index
$\widehat{Y}$: Vector of experimental measurements
e_i: Vector of zeros but with a one in its i^{th} component
$k_L a_{,i}$: Volumetric mass transfer coefficient of component i
W_{OP} and W_{OP2}: Weights to trade-off the estimation of OP1 and OP2
w_f: Weight for each feature in machine learning algorithms
W: Weight vector in an SVM model
Y_k: Yield of biomass k
Y_{CH4}: Yield of CH_4
Y_{ac}: Yield of acetate
Y_{CH4}^*: Yield of CH_4 obtained from the MOO
Y_{ac}^*: Yield of acetate obtained from the MOO

Abbreviations

$RMSE_{est}$: Adapted root mean square error
ADM1: Anaerobic Digestion Model No. 1
ADM1_ME: Anaerobic Digestion Model Extension
AD: Anaerobic Digestion
BCR: Bubble Column Reactor
BPNN: Back Propagation Neural Networks
BS: Benchmark Simulation Model
Cov: Covariance matrix
CH_4: Methane
COD: Chemical Oxygen Demand
CO: Carbon Monoxide
CO_2: Carbon Dioxide
CI: Confidence intervals
$CSTR$: Continuous Stirred Tank Reactor
DT: Decision Trees
EMPC: Economic Model Predictive Control
FIM: Fisher Information Matrix
GLR: Gas Loading Rate
GSA: Global sensitivity analysis
$GSIM$: Global Sensitivity Information Matrix
H_2: Hydrogen
HRT: Hydraulic Retention Time
k-NN: k-Nearest Neighbor

LCFA: Long-Chain Fatty Acids
MOO: Multi-Objective Optimization
MODO: Multi-Objective Dynamic Optimization
MPC: Model Predictive Control
NB: Naïve Bayes
OAT: One-Factor-at-a-Time
OP1: Operational Condition One
OP2: Operational Condition Two
OLR: Organic Loading Rate
POS: Pareto optimal set
POP: Pareto optimal point
PID: Proportional Integral Derivative Control
QDA: Quadratic Discriminant Analysis
RF: Random Forests
R^2: Coefficient of Determination
SA: Sensitivity Analysis
SSE: Sum Squared Error
SVM: Support Vector Machine
VFA: Volatile Fatty Acids
VS: Volatile Solids

General Introduction

The depletion of non-renewable fossil-derived fuels associated with the increased energy demand and the environmental problems related to fossil energy (Brémond *et al.*, 2021; Li *et al.*, 2018), encourage the "green" transition towards the use of renewable energies (Dar *et al.*, 2021; Hupfauf *et al.*, 2020). The European Union (EU) presented a long-term strategy that aims climate neutrality by 2050, where the use of renewable energies is expected to increase by at least 32% by 2030 (EC-European Commission, 2018).

Biogas produced through Anaerobic Digestion (AD) is considered one of the most promising renewable energy sources (Calise *et al.*, 2021; Li *et al.*, 2018). However, this biogas presents a low caloric value with 50-75% CH_4 and 25-50% CO_2 (Hupfauf *et al.*, 2020; Mulat *et al.*, 2017).

Biological methanation has recently gained attention (Angelidaki *et al.*, 2018; Bensmann *et al.*, 2014; Dumas *et al.*, 2020; Voelklein *et al.*, 2019) because it is a promising technology to upgrade biogas by adding syngas (Rafrafi *et al.*, 2020). The aim to perform biological methanation is to use the CO_2 contained in biogas as a carbon source that combined with H_2 can produce CH_4 and thus increase the CH_4 content between 95 and 99% (Iglesias *et al.*, 2021; Rusmanis *et al.*, 2019).

In this regard, it can cope with gas impurities (biogas may contain H_2S), CH_4 purity is increased (higher energetic power for a purer CH_4), and the CO_2 from biogas can be valorized (avoiding CO_2 emissions). The biomethane essential advantages are: once purified, it is used as a natural gas grid, energy storage, vehicle fuel (Luo and Angelidaki, 2012), and its use for generating electricity (Hupfauf *et al.*, 2020).

This work focuses on studying biological methanation from a process simulation perspective. In this context, this thesis aims **to develop a model for biological methanation (biomethanation) that can be used to optimize process operation, especially for** producing value-added products such as methane and acetate. This objective is accomplished thanks to **Multi-Objective Model Predictive Control** schemes and the use of **data-driven Soft Sensors**, which are based on the use of a **dynamic model** for biological methanation.

This thesis addresses the following three specific objectives:

- To propose a model for biological methanation capable of simulating the effect of the syngas addition at different operational conditions using $CSTR$ and BCR
- To maximize simultaneously multiple variables of interest such as yields, productivities, or economic aspects
- To develop data-driven soft sensors for fault detection based on the optimal operation of the biological methanation process

Figure 0.0.1 displays the link between the three specific objectives:

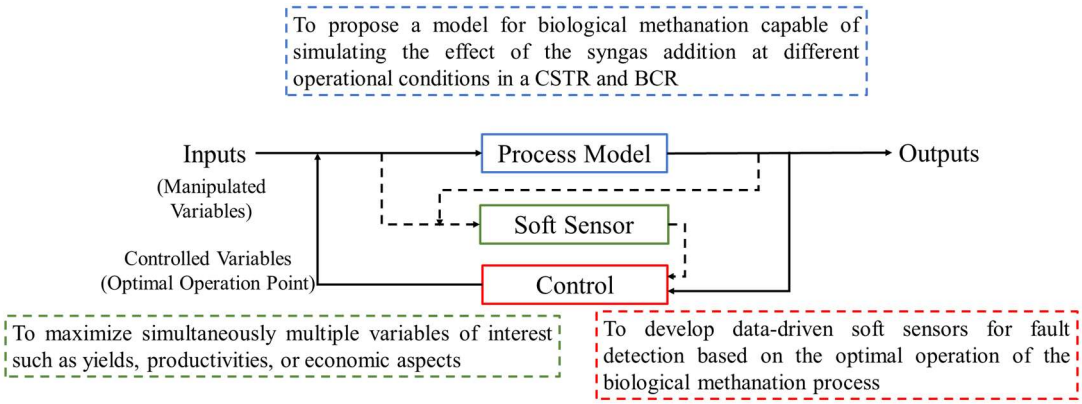


Figure 0.0.1. Thesis development scheme.

The contributions of this thesis include the formulation of a dynamic model for biological methanation capable to describe accurately the dynamics of biological methanation at different conditions. The advantage of the model over other models proposed in the literature was the generalization of the operational conditions. This model considers the volumetric mass transfer coefficient for two different reactor configurations: a bubble column reactor (BCR), transforming glucose, and a continuous stirred tank reactor ($CSTR$), using primary sludge and activated ticked-disintegrated waste. The model also considers the biological transformation of CO into acetate and H_2 by carboxydrotrophic acetogens and carboxydrotrophic hydrogenogens and the uptake of CO_2.

The biological methanation process was optimized using control schemes such as Multi-Objective Model Predictive Control, which simultaneously maximizes several variables of interest using Pareto Optimal solutions.

Another contribution of this thesis is the use of machine learning soft sensors oriented for fault detection in the biomethanation process, which has helped to detect disturbances in the manipulated variables.

This Ph.D. thesis is organized as follows. The first chapter presents the bibliography study and sets the theoretical framework for the biological methanation process. This first chapter is reported in four sections. The first and second sections provide insights into the anaerobic digestion and biological methanation processes and the different models used in the literature to represent those processes. The third section reviews some of the model-based control approaches and optimization tools that will be used in this work. The fourth section describes data-driven soft sensors, especially machine learning soft sensors, whose application is monitoring and process fault detection.

The second chapter is divided into seven sections presenting the main results. The first section is an introduction that summarizes and links the six following sections, which are presented as an article type. The second section details the formulation of the biological methanation model. The third, fourth, and fifth sections show different case studies showing model predictive control (MPC) and multi-objective optimization applications for biomethanation. The third section focuses on the maximization of yield (Y_{CH4}) and productivity (P_{CH4}) of methane. The fourth section describes the simultaneous maximization of (Y_{CH4}, P_{CH4}) which is complemented by a switch for the maximization of yields (Y_{ac}) and productivities (P_{ac}) of acetate. The fifth section accomplishes the Economic Multi-Objective Dynamic Optimization (EMODO) for the maximization of economic variables, ($Gain$) and ($Profit\ margin$). The Sixth and Seventh sections exhibit the application of machine learning soft sensors for fault detection in the biological methanation process. The Sixth section presents the training of several machine learning algorithms to check their ability to detect deviations from the optimal operation when there are disturbances in the liquid and gas flow rates. Section seventh shows the use of Support Vector Machines (SVM) in fault detection with an emphasis on training pairs of features to build 2D visualization diagrams.

Finally, chapter three draws the conclusions and sets the perspectives of this Ph.D. thesis based on the obtained results, making particular emphasis on a digital twin perspective.

0.1 References

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

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

Brémond, U., Bertrandias, A., Steyer, J.-P., Bernet, N., Carrere, H., 2021. A vision of European biogas sector development towards 2030: Trends and challenges. Journal of Cleaner Production 287, 125065. https://doi.org/10.1016/j.jclepro.2020.125065

Calise, F., Cappiello, F.L., Cimmino, L., d'Accadia, M.D., Vicidomini, M., 2021. A Review of the State of the Art of Biomethane Production: Recent Advancements and Integration of Renewable Energies. Energies 14, 4895. https://doi.org/10.3390/en14164895

Dar, R.A., Parmar, M., Dar, E.A., Sani, R.K., Phutela, U.G., 2021. Biomethanation of agricultural residues: Potential, limitations and possible solutions. Renewable and Sustainable Energy Reviews 135, 110217. https://doi.org/10.1016/j.rser.2020.110217

Dumas, C., Ottosen, L. D. M., Escudié, R., & Jensen, P. 2020. Editorial: Biological Methanation or (Bio/Syn)-Gas Upgrading. *Frontiers in Energy Research*, *8*, 30. https://doi.org/10.3389/fenrg.2020.00030

EC-European Commission, 2018. A Clean Planet for all–A European long-term strategic vision for a prosperous, modern, competitive and climate neutral economy depth analysis in support of the commission; Communication COM (2018) (No. 773).

Hupfauf, S., Winkler, A., Wagner, A.O., Podmirseg, S.M., Insam, H., 2020. Biomethanation at 45 °C offers high process efficiency and supports hygienisation. Bioresource Technology 300. https://doi.org/10.1016/j.biortech.2019.122671

Iglesias, R., Muñoz, R., Polanco, M., Díaz, I., Susmozas, A., Moreno, A.D., Guirado, M., Carreras, N., Ballesteros, M., 2021. Biogas from Anaerobic Digestion as an Energy Vector: Current Upgrading Development. Energies 14, 2742. https://doi.org/10.3390/en14102742

Li, W., Huusom, J.K., Zhou, Z., Nie, Y., Xu, Y., Zhang, X., 2018. Multi-objective optimization of methane production system from biomass through anaerobic digestion. Chinese Journal of Chemical Engineering 26, 2084–2092. https://doi.org/10.1016/j.cjche.2018.01.001

Luo, G., Angelidaki, I., 2012. Integrated biogas upgrading and hydrogen utilization in an anaerobic reactor containing enriched hydrogenotrophic methanogenic culture. Biotechnology and Bioengineering 109, 2729–2736. https://doi.org/10.1002/bit.24557

Mulat, D.G., Mosbæk, F., Ward, A.J., Polag, D., Greule, M., Keppler, F., Nielsen, J.L., Feilberg, A., 2017. Exogenous addition of H 2 for an in situ biogas upgrading through biological reduction of carbon dioxide into methane. Waste Management 68, 146–156. https://doi.org/10.1016/j.wasman.2017.05.054

Rafrafi, Y., Laguillaumie, L., Dumas, C., 2020. Biological Methanation of H2 and CO2 with Mixed Cultures: Current Advances, Hurdles and Challenges. Waste and Biomass Valorization. https://doi.org/10.1007/s12649-020-01283-z

Rusmanis, D., O'Shea, R., Wall, D.M., Murphy, J.D., 2019. Biological hydrogen methanation systems – an overview of design and efficiency. Bioengineered 10, 604–634. https://doi.org/10.1080/21655979.2019.1684607

Voelklein, M. A., Rusmanis, D., & Murphy, J. D. 2019. Biological methanation: Strategies for in-situ and ex-situ upgrading in anaerobic digestion. *Applied Energy*, *235*, 1061–1071. https://doi.org/10.1016/j.apenergy.2018.11.006

Chapter 1 Bibliography and Theoretical Framework

Detailed Content Chapter 1

Section 1 Anaerobic Digestion & Biological Methanation

Summary of Section 1

This section describes the anaerobic digestion process, its stages, and the factors that could affect the process. Then, challenges related to the optimization of biological methanation as a biogas upgrading technology are presented. It is described the reactions involved in the process, its modes of operation (in-situ and ex-situ), possible factors that limit the proper development of the process, and some of the reactors commonly used.

Anaerobic Digestion (AD) for biogas production has been used since the ancient ages. There is evidence of its implementation in ancient China (Bond & Templeton, 2011) and the use of biogas to heat bath water in Assyria and Persia in the 10th century B.C. (Meegoda *et al.*, 2018). Significant developments in the use of AD date back to the 19th century. In 1808 Humpy Davy demonstrated that it was possible to produce methane from the AD of cattle manure (Lusk, 1998). In 1859, India built the first AD plant to treat sewage. Later, in 1895, England built an anaerobic digestion plant prototype to recover biogas for light street lamps (Lusk, 1998; Meegoda *et al*., 2018). In the same century, China (1921) and Germany (1920) initiated the construction of large-scale digesters (He, 2010). During the late 19th and early 20th century, AD became a more developed technology. However, it was not until the middle of the 20th century that many countries became aware of the need to manage their waste, given the scarcity of landfills and the pollution effects on human health associated with fossil fuels (Klinkner, 2014).

The attractiveness of biogas comes from its high content of methane (~60%), which has interesting properties (see Table 1.1.1) and an extensive list of possible uses: in natural gas grids, vehicle fuels, to generate electricity and heat, and chemical feedstock (Dar *et al.*, 2021; Hupfauf *et al.*, 2020; Rafrafi *et al.*, 2020).

Table 1.1.1. Physical properties of methane (Chemical Rubber Company, 2005; Perry & Green, 1999).

Property	Value
Molecular weight	16.04 g/mol
Melting point	90.65K
Boiling point (Tb)	111.65K
Heat Value	55 MJ/mol
Molar heat capacity (Cp)	35.7 $J/mol\ K$
Standard molar enthalpy of formation ($\Delta_f H^\circ$)	-74.6 kJ/mol
Standard molar Gibbs energy of formation ($\Delta_f G^\circ$)	-50.5 kJ/mol
Standard molar entropy (So)	186.3 $J/mol\ K$
Critical temperature (Tc)	190.56 K
Critical pressure (Pc)	$4.599\times10^{-6}\ Pa$
Critical molar volume (Vc)	0.099 m^3/mol
Heat combustion (ΔH_c°)*	890.8 *kJ/mol*

*https://www.engineeringtoolbox.com/

In the last 20 years, research on AD and biogas has increased considerably, especially in fields such as environmental, energy chemical engineering, and agricultural sciences. Figure 1.1.1

shows the number of annual articles reported by Scopus using the keywords "anaerobic digestion" and "biogas", indicating a continuous effort of the scientific community to work in a renewable and sustainable process.

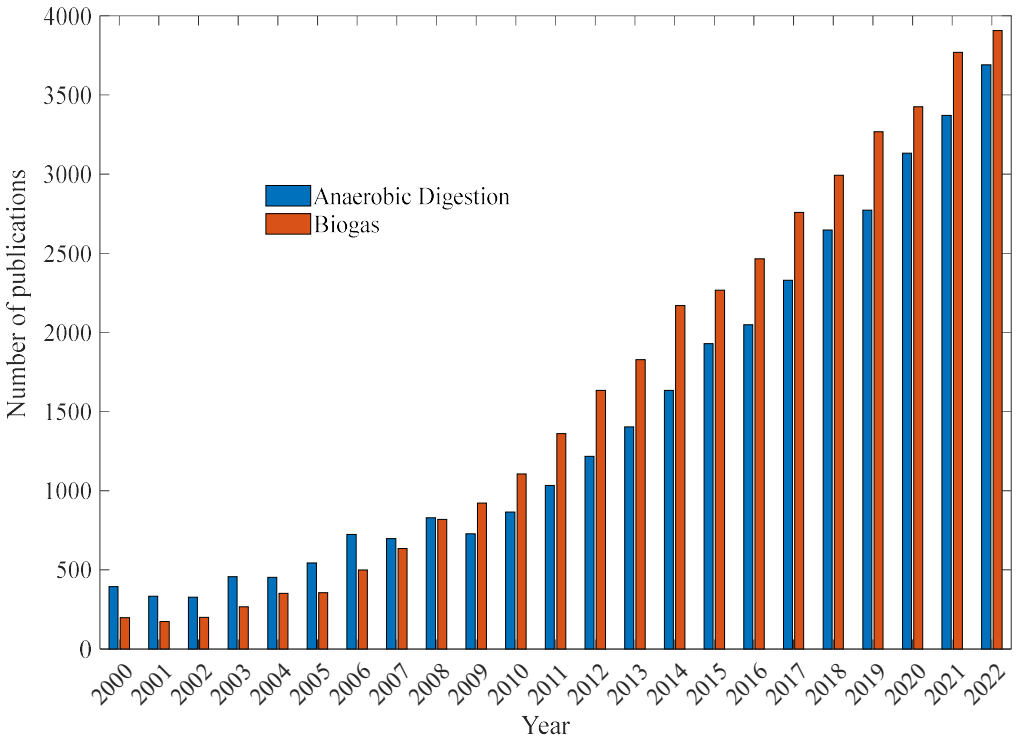


Figure 1.1.1. Scientific research articles dedicated to anaerobic digestion and biogas production. Data obtained from https://www.scopus.com/, June 2023.

1.1 Anaerobic Digestion

AD is a complex biological process in which organic matter is anaerobically degraded synergistically into a mixture of methane (CH_4), carbon dioxide (CO_2), and other gases in a minor way by microbial consortia of fermenting bacteria, anaerobic oxidizing bacteria, and methanogenic archaea (Angelidaki *et al.*, 2011).

AD involves biological transformations, physicochemical processes, and mass transfer between phases (Merkel & Krauth, 1999). **Physicochemical** processes are represented by components such as anions and cations or ionized forms of the compounds generated and consumed through AD, which is why multiple studies have focused on the analysis and variation of pH within this

type of system (Bashir & Aggarwal, 2017; Begum *et al.*, 2017; Chandra *et al.*, 2012; Czatzkowska *et al.*, 2020; Krishania *et al.*, 2013). **Biological degradation** process is the fundament of CH_4 production through AD in the liquid phase. Concerning **mass transfer**, compounds such as CH_4, H_2, and CO_2 are produced in the liquid phase and then released into the gas phase through a concentration driving force or gradient between the two phases.

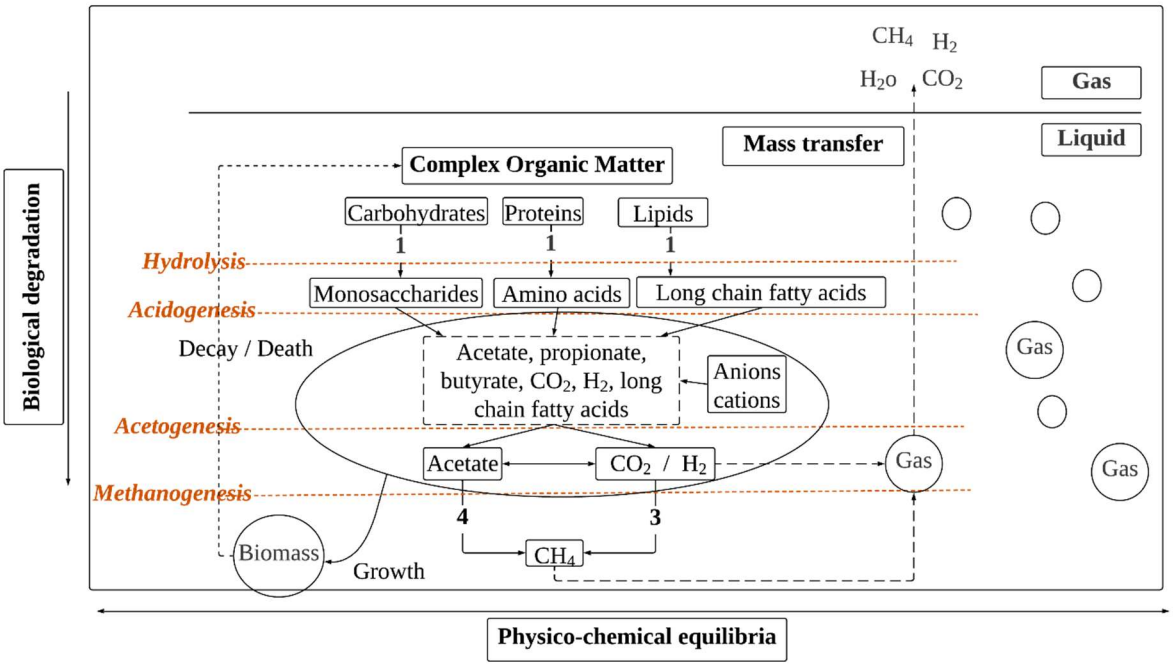


Figure 1.1.2. Overall anaerobic digestion process. Adapted from (Schön, 2009).

Figure 1.1.2 presents the interaction of the three processes, the biological component through AD, and its stages: hydrolysis, acidogenesis, acetogenesis, and methanogenesis. The interaction of components such as acetate, propionate, butyrate, and bicarbonate with their anions and cations obeys the physicochemical principles. Finally, mass transfer between the liquid and gas phases is graphically represented as the formation of bubbles in the liquid phase. In practice, both nucleation and mass transfer through an interphase are involved.

1.1.1 Anaerobic Digestion Process Stages

AD is divided into four stages associated with the degradation and conversion of organic biomass: hydrolysis, acidogenesis, acetogenesis, and methanogenesis (Chandra *et al.*, 2012; Ferry, 2011; Henze *et al.*, 2019; Mao *et al.*, 2015; Roopnarain & Adeleke, 2017; Saha *et al.*,

2020). This biogas produced in the AD contains between 50 - 75% of CH_4, 25 – 50 % of CO_2, and 2–7% of water vapor (Iglesias *et al.*, 2021; Laguillaumie *et al.*, 2022; Zupančič *et al.*, 2022).

Figure 1.1.3 describes the AD and each component associated with each step in the process.

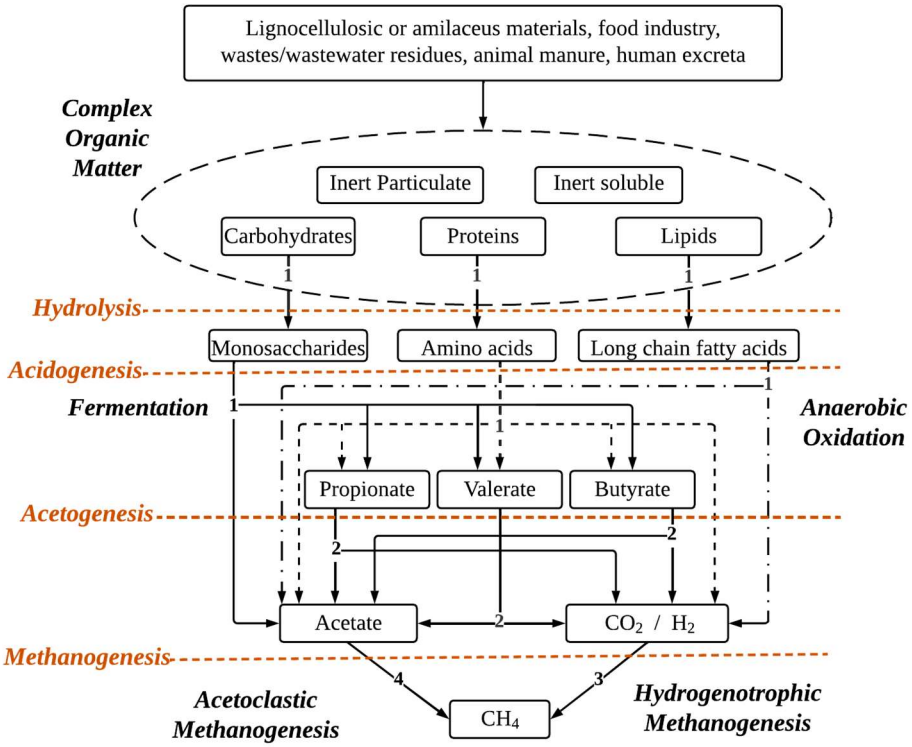


Figure 1.1.3. Anaerobic digestion process. The numbers indicate the microorganisms groups: 1. Hydrolytic and fermentative bacteria, 2. Acetogenic bacteria, 3. Hydrogenotrophic methanogens, 4. Acetoclastic methanogens. Adapted from (Batstone *et al.*, 2002; Henze *et al.*, 2019).

1.1.1.1 Hydrolysis

The first step of the AD process is hydrolysis, in which the organic matter, carbohydrates, proteins, and fats (complex undissolved compounds or polymeric organic compounds) are depolymerized into monomers or oligomers, sugars, glycerol, amino acids, long-chain fatty acids (LCFA) (less complex dissolved compounds).

This process generally takes place on the surface of acidogenic bacteria as it involves exo-enzymes excreted by hydrolytic bacteria, such as *Clostridia, Bacteroides*, *Fusobacterium*, *Butyrivibrio*, *Micrococci, Streptococcus* and *Selenomonas* (Chandra *et al.*, 2012; Czatzkowska *et al.*, 2020). In most cases, this stage is the rate-limiting step of the overall digestion process because the accessibility of enzymes reacting site is limited by the complex structure of substrate particles (Henze *et al.*, 2019).

1.1.1.2 Acidogenesis

The second step is acidogenesis; throughout this stage, the dissolved monomers or oligomers, amino acids, LCFA, and the components produced in the hydrolysis step undergo a degradation reaction. These components are assimilated into the acidogenic bacteria through the cell membrane and later fermented or anaerobically oxidized (Henze *et al.*, 2019) to produce volatile fatty acids (VFA), such as propionate, butyrate, valerate, and in minor form amounts of lactic, formic, and carbonic acid, alcohols, carbon dioxide, hydrogen, ammonia, as well as new cell material. This step is carried out by the action of bacteria of the genera *Bacillus* sp., *Pseudomonas* sp., *Clostridium* sp., and *Bifidobacterium* sp. (Czatzkowska *et al.*, 2020; Dar *et al.*, 2021).

The monosaccharides and the amino acids are the most abundant substrates for fermentation. Monosaccharides enter either the Emben-Meyerhof-Parnas (EMP) or the Entner Doudorof (ED) pathway, and later they are fermented via the acetyl-CoA pathway. At the same time, amino acids utilize the Stickland reaction, where these substrates are degraded into acetate in a coupled oxidation/reduction reaction (Angelidaki *et al.*, 2011). The acidogenesis stage is the most rapid in anaerobic conversion due to the high free energy change of the acidifying reactions. Furthermore, acidogenic bacteria are able to metabolize the substrates in a pH between 4 to 5.

1.1.1.3 Acetogenesis

In the third step, acetogenesis, the VFA produced in acidogenesis are reduced and transformed into acetate, hydrogen, and carbon dioxide, as well as in new cellular material, by the action of

bacteria of the genera *Clostridium*, *Syntrophomonas* sp., *Syntrophobacter* sp. (Chandra *et al.*, 2012; Czatzkowska *et al.*, 2020).

An important aspect to consider at this stage is the inhibitory effect of H_2. The interaction between H_2-producing acetogenic bacteria and H_2-consuming methanogenic bacteria regulates the H_2 levels. By themselves, the reactions involved in the acetogenesis (see Table 1.1.2) are thermodynamically unfavorable, presenting a $\Delta G^{\circ\prime}>0$ (Henze *et al.*, 2019). Once H_2 levels are regulated through these syntrophic associations, the partial pressure ranges between 10^{-4}-10^{-6} atm, making the reactions thermodynamically favorable with a $\Delta G^{\circ\prime}< 0$ (Henze *et al.*, 2019; Luo *et al.*, 2012).

1.1.1.4 Methanogenesis

In the last step, methanogenesis, the acetate, bicarbonate, and hydrogen are transformed into methane and carbon dioxide, as well as in new cellular material in two types of reactions, hydrogenotrophic methanogenesis and acetoclastic methanogenesis by the strictly anaerobic methanogens of the order Euryarcheota: *Methanobacteriales*, *Methanococcales*, *Methanomicrobiales*, *Methanosarcinales,* and *Methanocellales*.

In hydrogenotrophic methanogenesis, CO_2 is reduced into CH_4 using H_2 as a reduction agent (Ashraf *et al.*, 2020) by the action of hydrogenotrophic methanogens such as *Methanobacterium*, *Methanospirillum*, *Methanothermobacter*, and *Methanosarcina*. In acetoclastic methanogenesis, acetate is decarboxylated and converted into CH_4 by the action of acetoclastic methanogens (*e.g.*, *Methanosaeta*, *Methanococcoides, and Methanosarcina*) (Bharathiraja *et al.*, 2016; Czatzkowska *et al.*, 2020; Dar *et al.*, 2021; Dev *et al.*, 2019; Henze *et al.*, 2019; Saha *et al.*, 2020), Figure 1.1.4 shows the metabolic pathway, from acetate and CO_2 to CH_4.

Acetogenesis and methanogenesis usually run in parallel, as the symbiosis of two groups of microorganisms. Among the total CH_4 produced, about 60%-70% originates from the decarboxylation of acetate (by acetoclastic methanogens), while the remaining CH_4 is produced from CO_2 reduction and conversion of H_2 (by hydrogenotrophic methanogens) (Bharathiraja *et al.*, 2016; Malinowsky *et al.*, 2021; Pan *et al.*, 2021).

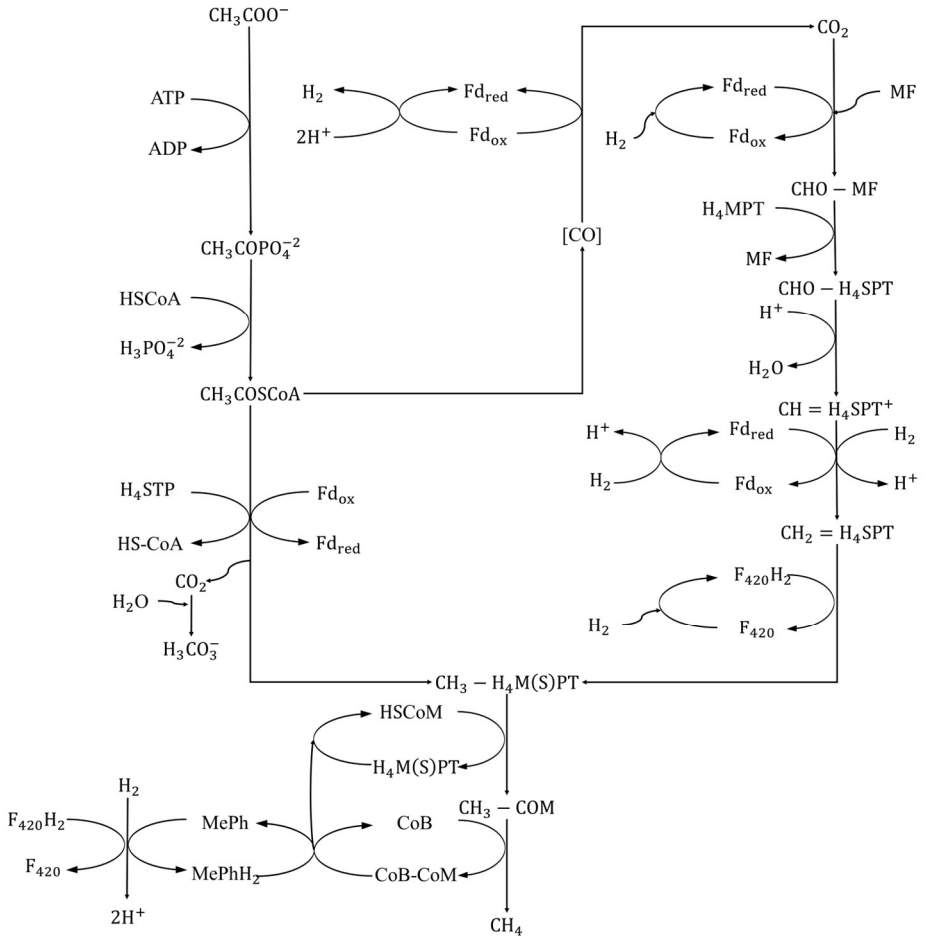


Figure 1.1.4. Methanogenesis pathways in anaerobic digestion by *Methanosarcina barkeri*. Adapted from (Ferry, 2011; Lyu & Whitman, 2019; Welander & Metcalf, 2005).

1.1.1.5 Homoacetogenesis & Sintrophic Acetate Oxidation

Two other types of microorganisms participate in the AD process: homoacetogenic bacteria and syntrophic acetate bacteria, which are activated when the concentration of H_2 is high. The homoacetogenesis implies the conversion of H_2 and CO_2 into acetate by bacteria such as *Moorella thermoacetica* (*Clostridium thermoaceticum*), *Acetobacterium woodii*, and *Clostridium ljungdahlii* (Ashraf *et al.*, 2020; Ferry, 2011; Liu *et al.*, 2016; Pan *et al.*, 2021; Westerholm *et al.*, 2016). Syntrophic acetate oxidation implies the conversion of acetate into H_2 and CO_2 by bacteria such as *Clostridium ultunense*, *Thermacetogenium phaeum*, and *Syntrophaceticus schinkii* (Pan *et al.*, 2021).

As previously mentioned, acetogenesis refers to acetate production by heterotrophic microorganisms through butyrate, propionate, valerate oxidation, *etc*. Conversely, homoacetogenesis is acetate production by autotrophic acetogenic microorganisms through CO_2 reduction with H_2 (Pan *et al.*, 2021; Saady, 2013) (Figure 1.1.5).

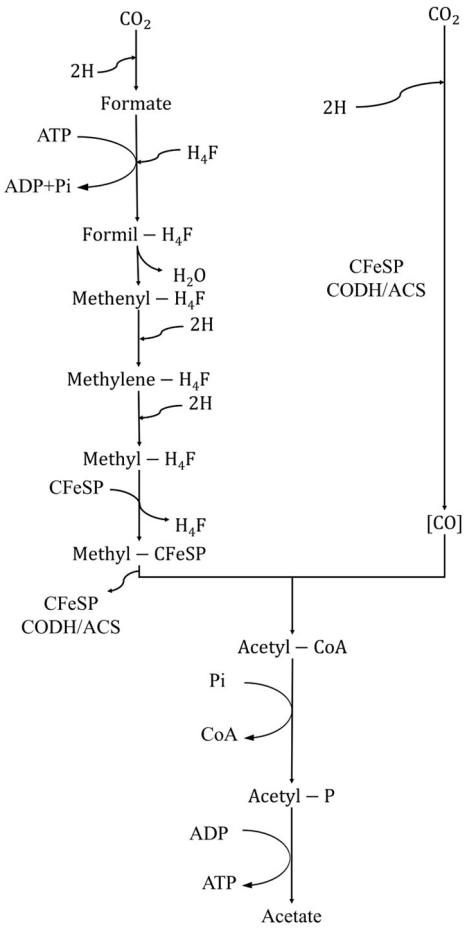


Figure 1.1.5. Acetyl-CoA Pathway. The reductive pathways comprise two branches (methyl and carbonyl) through which the methyl and carboxyl groups of acetate are synthesized, respectively. Adapted from (Saady, 2013; Westerholm *et al.*, 2016).

Acetogens grow slowly in a syntrophic relationship with methanogens which allows to keed H_2 partial pressures low $<10^{-6}\ atm$ (Henze *et al.*, 2019). Homoacetogens can grow faster than acetogens (with organic substrates) in the presence of H_2 and CO_2, which means high partial pressures $> 10^{-3}\ atm$ (Ashraf *et al.*, 2020; Liu *et al.*, 2016). It implies that homoacetogens are not inhibited by high H_2 concentrations (Saady, 2013), and they can change their metabolism

under conditions of stress or depletion of organic compounds (Ashraf *et al.*, 2020; Liu *et al.*, 2016).

Table 1.1.2. Main anaerobic bioreactions during the whole anaerobic digestion. Adapted from (Angelidaki *et al.*, 2011; Ashraf *et al.*, 2020; Liu *et al.*, 2016; Pan *et al.*, 2021; Sun *et al.*, 2021).

Reactions	$\Delta G°$ (kJ/mol)
Hydrolysis reactions:	
$(C_6H_{10}O_5)_n + nH_2O \rightarrow nC_6H_{12}O_6$	-
Acidogenesis reactions:	
Acetate:	
$C_6H_{12}O_6 + 2H_2O \rightarrow 2CH_3COOH + 4H_2 + 2CO_2$	-206
Butyrate:	
$C_6H_{12}O_6 \rightarrow 2CH_3CH_2CH_2COOH + 2H_2 + 2CO_2$	-254
Propionate:	
$C_6H_{12}O_6 + 2H_2 \rightarrow 2CH_3CH_2COOH + 2H_2O$	-279.4
Valerate:	-143.3
$CH_3CH_2COO^- + 2CO_2 + 6H_2 \rightarrow CH_3(CH_2)_3COO^- + 4H_2O$	-96.7
$3CH_3COO^- + 3H_2 + 2H^+ \rightarrow CH_3(CH_2)_3COO^- + 4H_2O$	-48.0
$CH_3(CH_2)_2COO^- + CH_3COO^- + 2H_2 + H^+ + CH_3(CH_2)_3COO^- + 2H_2O$	
Acetogenesis reactions:	
Propionate:	
$CH_3CH_2COOH + 2H_2O \rightarrow CH_3COOH + 3H_2 + CO_2$	+76.2
Butyrate:	
$CH_3CH_2CH_2COOH + 2H_2O \rightarrow 2CH_3COOH + 2H_2$	+48.4
Methanogenesis reactions:	
Hydrogen:	
$4H_2 + CO_2 \rightarrow CH_4 + 2H_2O$	-130.7
Acetate:	
$CH_3COOH \rightarrow CH_4 + CO_2$	-31.0
Syntrophic Acetate Oxidation (SAO) reactions:	
Acetate:	
$CH_3COOH + 2H_2O \rightarrow 2CO_2 + 4H_2$	+94.9
Homoacetogenesis reactions:	
Autotrophic:	
$2CO_2 + 4H_2 \rightarrow CH_3COOH + 2H_2O$	-94.9
SAO coupled with hydrogenotrophic methanogenesis:	
$CH_3COO^- + 2H_2O \rightarrow CH_4 + HCO_3^-$	-31.0

Syntrophic acetate oxidation is the process in that methyl groups of acetate are converted to CO_2 with the generation of H_2 (Pan *et al.*, 2021) and competing with the acetoclastic methanogens by the action of bacterias such as *Syntrophaceticus schinkii Clostridium ultunense*, *Thermacetogenium phaeum*, and *Tepidanaerobacter acetatoxydans*, (Ashraf *et al.*,

2020; Ferry, 2011; Liu *et al.*, 2016; Pan *et al.*, 2021; Westerholm *et al.*, 2016). The syntrophic acetate oxidation process is unfavorable ($\Delta G^{\circ\prime}$=+94.9 kJ/mol). However, at low concentrations of H_2 (low H_2 partial pressures), they couple with hydrogenotrophic bacteria, which allows the overall reaction to be exergonic ($\Delta G^{\circ\prime}$=-36.0 kJ/mol) and produce CH_4 with the equal stoichiometric of acetoclastic methanogens (Pan *et al.*, 2021).

Table 1.1.2 presents the reactions and their change in free energy associated. Acidogenesis presents a $\Delta G^{\circ\prime}$<0, indicating that it is an exergonic process. Acetogenesis presents a $\Delta G^{\circ\prime}$>0, indicating that it is an endergonic process. Methanogenesis shows a $\Delta G^{\circ\prime}$<0 (exergonic process), and syntrophic acetate oxidation and homoacetogenesis present a $\Delta G^{\circ\prime}$=-94.9, and $\Delta G^{\circ\prime}$=+94.9, respectively. However, the coupled reaction between syntrophic acetate oxidation with hydrogenotrophic methanogenesis presents a $\Delta G^{\circ\prime}$=-31.0, indicating that reactions can occur.

1.1.2 Factors Affecting the Anaerobic Digestion Process

Several factors influence AD performance (Chew *et al.*, 2021). Proper control of these factors is critical to maximizing CH_4 production and ensure the stability of the process.

1.1.2.1 Substrate composition

Several substrates have been used in AD. Li *et al.* (2018) presented a review of a large variety of substrates used in AD. The goal was to explore the characteristics of these substrates (high organic matter concentration, salt, oil, and protein contents; low Carbon/Nitrogen ratio) and their effect on AD efficiency. Nasir *et al.* (2012) reviewed the potential of AD for biogas production from livestock manure treatment and compared operating and performance data for various AD configurations. They checked livestock manure such as cattle, swine, and poultry manure. They concluded that the AD of livestock wastes could be an alternative disposal option with CH_4 yields between 0.01-0.5 L/gVS. Sendilvadivelu *et al.* (2022) reviewed the composition of municipal solid waste and input feedstock characteristics that affect the quality of products, such as digestate and CH_4 in AD. Values between 131-693 $LCH_4/kgVS$ were reported from municipal solid substrates such as food waste, kitchen waste, vegetable waste. Overall, a wide variety of studies refer to AD from multiple substrates.

1.1.2.2 Temperature

Temperature is one of the most critical factors that affect the AD process. Smaller fluctuations in temperature affect the biological activity of the microorganisms (Laiq Ur Rehman *et al.*, 2019). The AD process can be operated at three temperature ranges, psychrophilic AD (4–20°C, although 15°C is usually used as optimal), mesophilic AD (20–45°C, optimal at 37°C), and thermophilic AD (45–70°C, optimal at 55°C) (Ossa-Arias & González-Martínez, 2021). The mesophilic and thermophilic temperatures are the most common conditions in this process (Raposo *et al.*, 2012; Hupfauf *et al.*, 2018; P. Wang *et al.*, 2018). The correct choice depends on the AD objectives, *e.g.*, the operation of the AD process at thermophilic conditions implies high biogas yields and deactivation of pathogens. However, temperature values between 40 and 50°C inhibit the activity of methanogens (Laiq Ur Rehman *et al.*, 2019). On the other hand, the operation at mesophilic conditions can maintain high organic loading rates but has lower conversion rates (Laiq Ur Rehman *et al.*, 2019; Sendilvadivelu *et al.*, 2022; Van *et al.*, 2019).

Hupfauf *et al.* (2018) studied the AD process at five different temperatures, 10, 20, 37, 45, and 55°C, using cattle slurry and maize straw as a substrate, with an OLR of 2.04 $gVS/L/d$. The authors demonstrated that 45°C is a good alternative with similar results in CH_4 content (22.4%) concerning the thermophilic and mesophilic conditions (20.2 and 19.9%). Conversely, Mortezaei *et al.* (2023) investigated the effects on AD with changes in the solids retention times operating at two temperatures, 35°C and 55°C, where the latest improved the biogas production between 34-42% concerning the mesophilic operation. Nie *et al.* (2021) presented an interesting review of the effect of temperature in the different stages of AD. This work summarizes different research works that studied the influence of temperature on AD and allowed us to conclude some relevant aspects:

- Substrate hydrolysis rate increased with the temperature due to the increased activity of the extracellular enzymes, *i.e.*, thermophilic hydrolysis increases the activity of the extracellular enzymes
- The production of VFA can be increased from psychrophilic and mesophilic temperatures. However, there is still no agreement concerning the increase in VFA when AD is developed at mesophilic and thermophilic conditions; During

methanogenesis, most hydrogenotrophic and acetoclastic methanogens may shift with the bioreaction temperature. For instance, acetoclastic methanogens predominate at 35°C, while hydrogenotrophic methanogens predominate at 45°C; Syntrophic acetate oxidation is favored at low acetate concentrations, while acetoclastic methanogenesis is favored at high acetate concentrations

1.1.2.3 pH

The pH in AD affects the activity of the microorganisms (Laiq Ur Rehman *et al.*, 2019). The CH_4 formation in AD processes ranges between 6.5-8.5 and becomes unstable when the pH drops below 6.0 or increases above 8.5 (Weiland, 2010). The optimal pH values for each stage are 6-8 for hydrolysis, 5.5-6.5 for acidogenesis, 6.0-6.2 for acetogenesis, and 7.0-8.0 for methanogenesis (Raposo *et al.*, 2012; Van *et al.*, 2019).

An interesting study was developed by Lindner *et al.* (2015), which performed a two-phase biogas plant composed of (i) a 124 L continuous acidification reactor operating at 60°C to develop the first steps of the AD and (ii) two 62 L anaerobic filters to perform the methanogenesis step at 37°C using as substrate maize silage. The aim was to evaluate the effect of different pH values 5.5, 6.0, 7.0, and 7.5 over the AD process. In the continuous acidification reactor, values closed to 12.0, 14.0, 7.0, and 2.0 g/kg were obtained for the VFA: caproic, valerate, butyrate, propionate, and acetate (values estimated from a graphic). At pH 5.5 and 6.0, the representative VFA was acetate with values of 6.5 and 10.0 g/kg, respectively, while propionate was the representative VFA at pH 7.0 and 7.5 with values of 4.4 and 0.8 g/kg, respectively. The CH_4 content values of 0.4, 35.1, 48.0[2], and 50.42% were obtained for pH of 5.5, 6.0, 7.0, and 7.5, respectively. In the anaerobic filters, a CH_4 of 64.3% was obtained with a pH of 5.5, while the other pH reached a CH_4 content higher than 71%. It is concluded that acidogenesis and methanogenesis work better in the previously mentioned ranges. However, when the entire system is analyzed in terms of CH_4, biogas yields are higher with increasing pH. For pH 5.5, values of 194.19 and 483.57 NL/kg for organic dry matter were obtained for CH_4 and biogas yield, respectively. On the other hand, for a pH of 7.5, values of 336.71 and 641.00 NL/kg of organic dry matter were obtained for CH_4 and biogas yield, respectively.

1.1.2.4 Hydraulic retention time and organic loading rate

Hydraulic Retention Time (HRT) is the average time interval that a liquid or dissolved component remains in the reactor (Dong *et al*., 2022; Eggen and Vogelsang, 2015; Lindmark *et al*., 2014). It is calculated as reactor volume over the input flow. A longer HRT contributes to the high reduction of VFA, resulting in improved AD efficiency (Malinowsky *et al.*, 2021; Zamri *et al.*, 2021). Tena *et al.* (2021) investigated the impact of HRT on CH_4 production in the two-stage thermophilic and mesophilic AD process. Two Continuous Stirred Tank Reactors ($CSTR$) were used in series to perform the acidogenesis and methanogenesis steps. Eight HRT, 20, 16, 10, 6,5,4,3, and 2 days were imposed while the feed flow was made of a mixture of sewage sludge and wine vinasse (0.5:0.5). According to their results, it can be concluded that an AD operating at an HRT of 4 days is an excellent option to reduce the time of the process and increase the CH_4 yield to 159.4 $mLCH_4/gCOD$ (COD: Chemical Oxygen Demand. Amount of oxygen needed to degrade the organic matter into CO_2 and H_2O). Decreasing the HRT below 4 days resulted in the accumulation of VFA. Sillero *et al.* (2023) investigated the influence on methanogenesis of the HRT during co-digestion of a substrate mixture of sewage sludge, wine vinasse, and poultry manure (0.495:0.495:0.01). The authors developed the AD process using two 5 L $CSTR$ connected in series. The first reactor was used to develop the acidogenic stage at 55°C and pH 5.5. The second reactor was used to perform the methanogenesis step at 35°C and pH 7.5 with seven HRT, 15, 12, 10, 8, 5, 4, and 3 days. The results showed that 12 days is the best HRT for the methanogenesis with a maximum CH_4 yield was 391 mL/gVS.

The organic loading rate (OLR) is the amount of organic matter added to the AD system per unit reactor volume per day (Grangeiro *et al.*, 2019). High values in the OLR cause acidification and inhibit the activity of different microorganisms, leading to reduced biogas production. Zhou *et al.* (2022) proposed AD for food waste in a horizontal flow reactor operated in a semi-continuous condition at mesophilic temperatures with OLR ranging from 1.00 to 13.80 $kgVS/m^3/d$. Values between 0.173 - 0.516 $L/g/d$ and 0.25 -5.69 $L/L/d$ were obtained for CH_4 yields and volumetric CH_4 production, respectively. The authors demonstrated that high OLR improved the AD process. Ünyay *et al.* (2022) developed an AD process for raw switchgrass in a sequential batch reactor (daily feed) and semi-$CSTR$ at three different OLR

0.75, 1.0, and 1.5 $gVS/L/d$. They found the optimum OLR in the semi-continuous configuration at 1.0 $gVS/L/d$, where 35% of the switchgrass theoretical CH_4 yield and 38% energy recovery were attained. Higher OLR (1.5 $gVS/L/d$) caused low CH_4 content due to the VFA accumulation.

When there is a continuous or semi-continuous flow rate added to the system, the HRT and OLR are related by the following Equation (Labatut & Pronto, 2018):

$$OLR = \frac{S_0}{HRT} = S_0 \frac{q_{liq}^{in}}{V} \quad (1.1.1)$$

where OLR is expressed as VS or COD basis ($g/L/d$); HRT is expressed in days (d). S_0 refers to the influent substrate concentration, VS or COD basis (g/L); q_{liq}^{in} and V are the inlet liquid flow rate (L/d); and volume reactor (L), respectively.

1.1.2.5 Mixing

The AD process could be carried out with continuous mixing, intermittent mixing, or not be mixed at all (Lindmark *et al*., 2014). Zhang *et al.* (2019) investigated the effect of different mixing strategies on AD of food waste. Three reactors operating at 35°C with an HRT of 5 days were used. Reactor 1 (R1) was operated with semi-continuous mixing of 2 min/h at 80 rpm. The Reactor 2 (R2) was operated with continuous mixing of 80 rpm for the duration of the experiment. Reactor 3 (R3) was used as an unmixed control. The results showed that the semi-continuously mixed R1 achieved AD efficiencies of 74.4%, which is higher than the continuously mixed R2 (66.9%) and unmixed R4 (14.9%).

Ma *et al.* (2019) evaluated the effect of the mixing velocity using a two-phase AD system of sewage sludge. The first phase consisted of the hydrolysis and acidogenesis process (HAP), where it was used a 600 $mL\ CSTR$ operating at 37°C with an OLR of 7.96 $gVS/L/d$. The second phase was referred to as the methanogenesis process (MP). It was performed in a 600 $mL\ CSTR$ operating at 37°C with an OLR of 0.796 $gVS/L/d$. Eight $CSTR$ were used and divided into four groups according to the mixing power in the HAP: group 1 (HAP1 at 30 rpm and MP1 at 120 rpm), group 2 (HAP2 at 30 rpm and MP2 at 120 rpm), group 3 (HAP3 at 30

rpm and MP3 at 120 rpm), group 4 (HAP4 at 30 rpm and MP4 at 120 rpm). Results showed that the concentration of soluble COD and total VFA produced in group 3 (2134 ± 58 mg/L and 1311 ± 22 mg/L) and group 4 (2030 ± 39 mg/L and 1281 ± 21 mg/L) was significantly higher than those in group 1 (1346 ± 32 $mg/\ L$ and 730 ± 43 mg/L) and group 2 (1693 ± 62 mg/L and 1144 ± 32 $mg/\ L$). On the other hand, Yang and Deng (2020) proposed a mixing method for AD ("air mixing"), treating animal wastewater using air as a momentum source of agitation. Four 4.0 L batch reactors operating at 35°C were used. Reactor 1 (R1) and reactor 2 (R2) were mixed with air and biogas. Reactor 3 (R3) was mixed using an axial flow impeller, and Reactor 4 (R4) was used as control (unmixing). It was concluded that the addition of air (R1) improved CH_4 production by 6.4, 11.9, and 19.6% compared to the addition of biogas (R2), mechanical agitation (R2), and control (R4), respectively. Additionally, the CH_4 yield in R1 improved by 6.5, 11.7, and 19.90 %, concerning R2, R3, and R4, concluding that adding air improved the AD. The authors also analyzed the effect of mixing in the mass transfer process. The degree of mixing (homogeneity) was calculated through the coefficient of variation (cv) of the total solids concentrations in the system. The authors defined that cv-value less than 0.02 indicated homogeneous mixing. For R1 and R2, stable values close to 0.016 were obtained after 1.5 min of agitation. For R3, a value of 0.025 was obtained after 10 min of agitation, while a value close to 0.17 were achieved in R4. They interpreted this result as a measure of the mass transfer effect, where R1 and R2 presented the best improvement due to the substrate's and sludge's physical movement when the bubbles were ascending.

1.2 Biological Methanation

Biological methanation, or biomethanation, is a promising technology to upgrade biogas by adding syngas (Rafrafi *et al.*, 2020). The aim is to upgrade residual components such as CO_2 and increase the CH_4 content (95 – 99 % CH_4) towards the end of the process (Iglesias *et al.*, 2021; Rusmanis *et al.*, 2019). One of the essential advantages of upgrading biogas is its use as a natural gas grid, energy storage, and vehicle fuel (Luo & Angelidaki, 2012). Biological methanation occurs in the last stage of AD. The occurrence of biological methanation requires the addition of H_2. Several sources of H_2 can be used as either pure components or mixed with others. An interesting mixture containing H_2 is syngas. The syngas loading (commonly a

combination of H_2:CO:CO_2) can improve the process and convert the H_2 and CO_2 into CH_4 (Grimalt-Alemany *et al.*, 2018; Rusmanis *et al.*, 2019). The hydrogenotrophic methanogens with CO_2 consumption transform the H_2. Although this route is well-known, carbon monoxide (CO) consumption remains unclear (Sun *et al.*, 2021).

The conversion of CO can be divided into direct reactions and indirect reactions. Directly, the CO is transformed to CH_4 by the action of microorganisms such as *Methanobacterium thermoautotrophicum, etc.* (Guiot *et al.*, 2011). Indirectly, the CO is converted to acetate by some species from genera, *e.g.*, *Clostridium, Acetobacterium,* and *Sporomusa*, which can produce acetate and alcohols (Karekar *et al.*, 2022; Paniagua *et al.*, 2022).

Then, this reaction is flowed by acetoclastic methanogenesis to obtain CH_4. The CO is also transformed into H_2 by carboxydotrophic hydrogenogenesis (water gas shift). Some species of the gender *Rhodospirillum, Thermincola,* convert CO to H_2 and CO_2 (Y. Li *et al.*, 2020; Paniagua *et al.*, 2022). Then, this reaction is followed by hydrogenotrophic methanogenesis. The homoacetogens and syntrophic acetate oxidizers have a role in biological methanation. The homoacetogens transform H_2 and CO_2 into acetate, and the syntrophic acetate oxidizing bacteria oxidized the acetate to produce H_2 and CO_2 (Grimalt-Alemany *et al.*, 2020; Pan *et al.*, 2021). Figure 1.1.6 presents the reaction pathways in biological methanation.

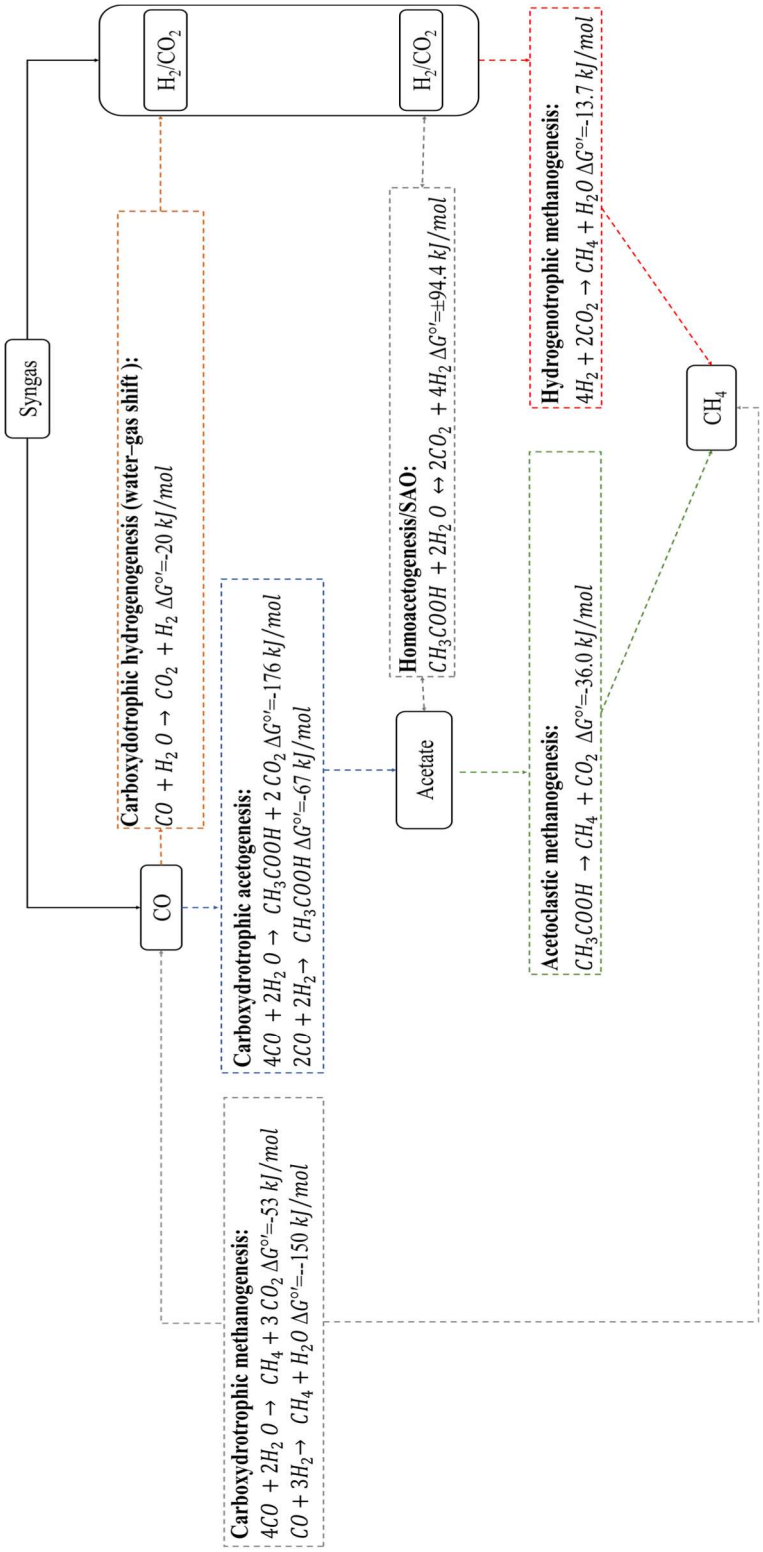


Figure 1.1.6. Reaction pathway in biological methanation. Adapted from (Paniagua *et al.*, 2022).

1.2.1 Biological Methanation strategies

There are two strategies to develop biological methanation (Figure 1.1.7): *in-situ* biological methanation (directly in the AD reactor), where the syngas is added during the AD process, and *ex-situ* biological methanation (in a separate unit), where an external reactor is used to enhance the process using specialized methanogenic microorganisms (Jensen *et al.*, 2018). In both cases, concentrations above 90% are reported, where the conversion of biogas, specifically H_2, is limited by the gas-liquid mass transfer rate (Luo *et al.*, 2012; Rusmanis *et al.*, 2019).

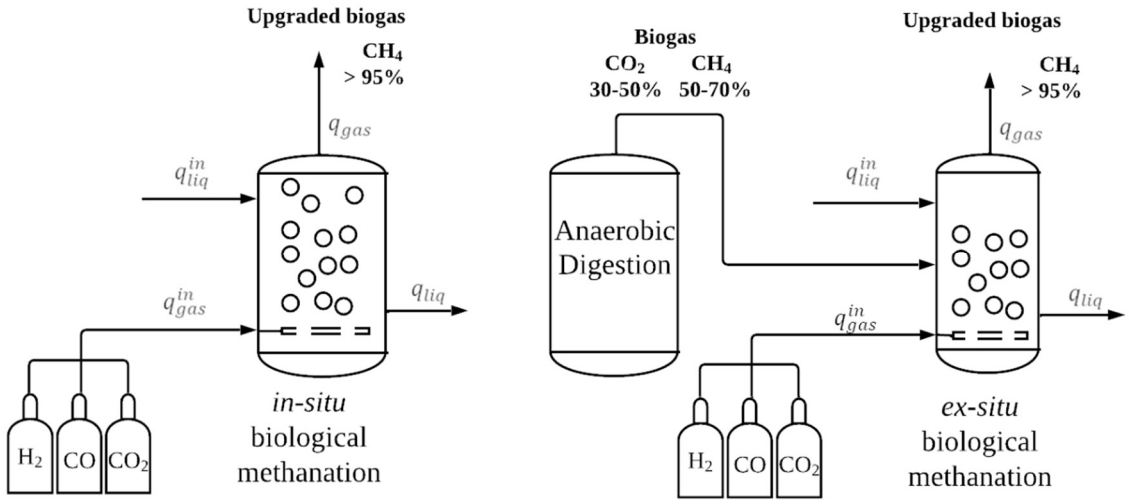


Figure 1.1.7. Scheme of *in-situ* and *ex-situ* biological methanation process. Adapted from (Voelklein *et al.*, 2019).

The advantage of *in-situ* biological methanation is the reduction of infrastructure costs, using only one reactor where syngas is directly added, allowing AD and biological methanation to occur simultaneously. The main drawback of *in-situ* biological methanation is that AD can be affected by the high concentration of gas. The H_2 added to the system can inhibit the hydrolysis and acetogenesis steps (Rafrafi *et al.*, 2020). On the other hand, *ex-situ* biomethanation takes place in a separate external reactor, which implies constructing an additional physical system. The *ex-situ* biomethanation is typically adapted to suit the hydrogenotrophic methanogens due to the facility to dissociate the process conditions, such as temperature and pressure of the hydrolysis and acidogenesis steps to the methanogenesis step (Voelklein *et al.*, 2019). One of the advantages of this technology is that it can process higher H_2 loading rates and obtain

excellent gas conversion rates compared with the *in-situ* technology (Wu *et al.*, 2021). Regardless of the type of configuration, microorganisms, and metabolic pathways are similar (Mulat *et al.*, 2017; Rafrafi *et al.*, 2020; Rusmanis *et al.*, 2019).

Although *ex-situ* technology is currently preferred at the industrial level due to its properties (Mulat *et al.*, 2017; Rafrafi *et al.*, 2020), studies have been developed using both technologies to understand better. Luo *et al.* (2012) investigated the feasibility of converting H_2 to CH_4 and upgrading biogas through two types of experiments *in-situ*. The first experiment corresponded to the study of the potential inhibition of H_2 on VFA degradation in batch operation. In contrast, the second experiment involved CH_4 production from H_2 continuously. In the batch operation, they tested three initial partial pressure, 0.25, 0.5, and 1.0 atm; and 2 agitation speeds, 100 and 300 rpm. The results showed that at 100 rpm, the H_2 consumption rate decreased 66, 30, and 16 $mL/L/h$ with the reduction of H_2 partial pressure 1.00, 0.50, and 0.25 atm, respectively. However, at 300 rpm, the H_2 consumption rate was almost constant at 270 $mL/L/h$, *i.e.*, independent of the H_2 partial pressure. Their results showed that lower mixing intensity is crucial to achieving H_2 utilization without inhibiting propionate and butyrate degradation. Additionally, two $CSTR$ were used to analyze the biogas production in the continuous operation. Both reactors were fed with cattle manure until they reached a steady state. Then, H_2 was added at a flow rate of 28.6 $mL/L/h$ to one of those reactors and the second one was used as a control; both reactors were operated for 1.5 months until they reached a new steady state. After the H_2 addition, 80% of the added H_2 was consumed. The H_2 consumption rate was around 22.8 $mL/L/h$, indicating that gas liquid mass transfer is the limitating factor. The biogas production increased concerning the control from 25.1±1.8 to 29.1±2 $mL/L/h$ with a CH_4 content increased from 62±2.5 to 65±3.3%, and CH_4 production increased from 15.5±1.1 to 18.9±0.9 $mL/L/h$).

Voelklein *et al.* (2016) conducted several alternatives to upgrade the biological methanation. One of those investigations started to understand the thermophilic degradation of grass silage at increasing the OLR (more related to AD). The authors used a $CSTR$ operating at 55°C with five different OLR, 7, 6, 5, 4, and 3 $gVS/L/d$. They found yields between 351 and 405 $L_{CH4}/kgVS$ with a CH_4 content in the biogas produced between 51.5 and 52.9%. Therefore,

another investigation was performed to study *in-situ* and *ex-situ* biological methanation strategies for biogas upgrading potential (Voelklein *et al.*, 2019). The experiment was developed using three reactors with a volume of 9.5 L: Batch *in-situ*, batch *ex-situ*, and continuous *ex-situ*. The *in-situ* and *ex-situ* upgrading was performed at 55°C and ambient pressure with an OLR of 4 $gVS/L/d$. The *in-situ* strategy achieved CH_4 yields between 382 and 640 $L/kgVS$ and a CH_4 formation rate between 0.33-2.52 $L/L/d$ with a CH_4 content between 32.1 and 60.3%. In the *ex-situ* strategy, the batch operation achieved a CH_4 formation rate between 1.7-3.7$L/kgVS/d$ with a CH_4 content between 92-96%, while the continuous operation achieved a CH_4 formation rate between 0.85-9.1$L/L/d$ with a CH_4 content between 15-85%. In both cases, a significant increase in CH_4 content was observed concerning the first experiment, especially in the *ex-situ* batch operation, which elucidates the efficacy of improving CH_4 content by adding syngas.

1.2.2 Process Limitations

In section 1.1.3. was presented the factors that affect the AD process. These factors have also been considered in biological methanation. However, considering the addition of gases such as H_2, CO, and CO_2 in the biological methanation process, it is necessary to address several of these factors from a different perspective.

1.2.2.1 pH

pH is one of the most critical variables to consider in biological methanation, as it ensures the correct stability of the different microbial consortia (Giwa *et al.*, 2019). However, the selection of pH depends on the type of microorganism desired to predominate in biological methanation. The accumulation of VFA in the system causes a reduction in pH, leading to inhibitory processes (Czatzkowska *et al.*, 2020). This was corroborated by Rafrafi *et al.*, (2020), who mentioned that regardless of operation mode, *in-situ* or *ex-situ*, an increase in pH can inhibit acetoclastic methanogenesis. Methanogenic bacteria prefer to work in neutral environments. However, some strains can work in both basic and acidic environments, so the pH range in which they work well is between 5.0 and 8.5 (Strobel *et al.*, 2020). The addition of H_2 will prefer to react with CO_2 instead of CO, resulting in rapid consumption of this H_2 and an increase in pH, which can inhibit the activity of CO-consuming microorganisms (Paniagua *et al.*, 2022).

Several studies have focused on the effect of pH in biological methanation. Li *et al.* (2020) studied the impact of different substrates over biological methanation and added different syngas ratios. The findings revealed inhibitory effects due to the pH increase associated with adding H_2, which was solved using phosphate buffer for pH control. Ashraf *et al.* (2020) investigated the pH effect as a control strategy to endure the process in stable conditions using a thermophilic trickling filter to degrade cow manure. Among the two control strategies, phosphate buffer addition and control CO_2/H_2 feed ratio, the latest was selected as the best method to control de pH below 8.5. The results highlighted that the pH is maintained at 8.5 in the trickling filter by varying CO_2/H_2 with ratios between 0.25 and 0.5.

Wang *et al.* (2013) developed *in-situ* biological methanation using coke oven gas (H_2/CO 92/8) and sewage sludge. The experiment consisted in upgrading the biogas using a 2 L $CSTR$ coupled with a hollow fiber membrane. The gas addition was developed in four stages. Stage I (1-30 days): there was no gas added; stage II (31-60 days): 1300±300 mL/d; stage III (61-72 days): 2600-2900 mL/d; stage IV (73-103 days): 2882±32 mL/d. The CH_4 content was around 64.4, 89.9, 97, and 98.8% in stages I, II, III, and IV. Nonetheless, the authors found some relevant aspects in the experiment. In stage II, H_2 and CO were not detected, which indicated they were utilized efficiently by the microorganisms. The VFA rounded 4.8 g/L (close to the value in stage I), showing that adding coke oven gas did not negatively affect the process. The pH in stages I and II was 7.0 and 7.5, respectively. In stage III, the increase in the gas flow rate occasioned an increase of pH from 7.5 to 9.0, which reduced CH_4 content to 64%, accumulation of VFA, and detection of H_2 and CO, signs of process inhibition. To solve this problem, in stage IV, the pH of the reactor was controlled close to 8.0. The concentration of VFA came back to 4.9 g/L, H_2 and CO were not detected, and the AD of sewage sludge was not affected. The results showed that the *in-situ* addition of coke oven gas to the AD of sewage sludge was successfully achieved. However, the need to control the pH as the coke oven gas is added exists.

1.2.2.2 Temperature

The temperature impacts the gas-liquid mass transfer and the microorganism's interactions (Paniagua *et al.*, 2022). Strobel *et al.* (2020) state that methanogens grow and live in mesophilic, thermophilic, hyperthermophilic, and even psychrophilic environments. The lowest and highest

temperatures reported are 15 and 98°C, respectively. Studies in biological methanation from CO indicate acetate as the primary precursor of methanogenesis at mesophilic conditions and H_2 as the primary precursor at thermophilic conditions (Grimalt-Alemany *et al.*, 2018). Asimakopoulos *et al.* (2020) compared the performance of two continuous trickle bed reactors operated at mesophilic and thermophilic conditions. In both reactors, a mixed microbial consortium was used. The results revealed that the operation at thermophilic conditions achieved higher CH_4 productivities (8.49 $mmol/L/h$) and higher conversion efficiencies (H_2/CO 89 and 73%, respectively) of H_2 and CO concerning the mesophilic conditions. Nevertheless, Y. Li *et al.* (2020) showed that there were minor differences in the CH_4 content between both operations when a continuous reactor under same conditions was used.

1.2.2.3 Type of culture

Biological methanation can be performed in pure or mixed cultures. Pure cultures imply operating the biological methanation process to conditions favoring its growth and process performance, *i.e.*, the optimal growth conditions for a selected microorganism with an adequate metabolism. However, a mixed consortium rarely shares the same optimal growth conditions. In recent years, the study of biological methanation has progressed from using pure cultures to understanding carboxydotrophic microorganisms to some studies with mixed cultures for use in the industrial sector (Grimalt-Alemany *et al.*, 2018). Both types of cultures have been investigated concerning biological methanation. Nevertheless, mixed cultures are more robust and do not require sterile conditions (Rachbauer *et al.*, 2017).

Most authors have used enriched mixed cultures to develop the biological process with different objectives. Rachbauer *et al.* (2017) evaluated the effect of process parameters such as acetic acid concentration on carbon conversion in a trickle bed reactor using an enrichment culture of hydrogenotrophic methanogens adapted from sewage sludge. Grimalt-Alemany *et al.* (2020) characterized the syngas conversion routes utilized by a mixed consortium enriched at mesophilic and thermophilic conditions. Figueras *et al.* (2021) used a mixed consortium to explore high-pressure effects on a continuous lab-scale pilot using a pressurized agitated column operating at thermophilic conditions. Laguillaumie *et al.* (2022) performed *ex-situ*

biological H_2 and CO methanation with a mixed culture in a bubble column reactor (BCR) operating at 55°C.

In those studies, overall findings were elucidated:

- The adaptation of hydrogenotrophic microorganisms is affected by the addition of components such as acetate, resulting in a reduced carbon conversion (Rachbauer *et al.*, 2017)
- Mixed cultures are frequently used in the biological methanation process. They can support operational changes due to the high microbial diversity and resilience to large storage periods at different temperatures (Laguillaumie *et al.*, 2022)
- Carboxydotrophic microorganisms are considered to be more sensitive to CO than methanogens (Figueras *et al.*, 2021)
- The substrate competition between different microorganisms is driven by kinetic competition and thermodynamic limitations. Additionally, the activity patterns differ between mesophilic and thermophilic enriched consortia, where the latest could be more suitable for industrial applications (Grimalt-Alemany *et al.*, 2020)

1.2.2.4 Gas-liquid mass transfer

Gas-liquid mass transfer is the main limiting factor in biological methanation (Ngu *et al.*, 2023; Paniagua *et al.*, 2022; Rusmanis *et al.*, 2019). Andreides *et al.* (2022) studied mass transfer in biological methanation using mechanical and pneumatic agitation. The authors performed an experiment using a $CSTR$ operated at 55°C with mechanical agitation (55 rpm). The syngas a mixture of H_2 and CO (0.55/0.45) was added to the system at different flow rates in 5 periods, 3.15, 7.35, 10.5, 15.75, 15.75 L/d, for period 1 (36-51 days), period 2 (51-81 days), period 3 (81-118 days), period 4 (118-130 days), and period 5 (130-150 days), respectively Nevertheless, the increase in the syngas flow rate from period 3 to 4 affected negatively the process, reducing the conversion efficiency of H_2 and CO from 60.7±4.3 and 58.1±2.7 to 54.4±3.0 and 54.1±3.6%, respectively. They used pneumatic agitation to correct the negative effect, in period 5 of the process., The internal gas mixing rate was set at 7 $L/L/d$ and the gas flow rate was maintained constant along of period 4. This permitted to achieved conversion efficiency of H_2 and CO of 84.3±4.0 and 73.9±6.6%, respectively. The modification to

pneumatic mixing instead of mechanical stirring mixing ensured the efficient gas-liquid mass transfer rate, with an increase in the hydrogen and carbon monoxide conversion efficiency close to 30%.

The pressure increase, through bubble size reduction, ensures a longer residence time and improves the mass transfer of components such as H_2 (Ullrich *et al.*, 2018). As mentioned in Section 1.2.1.3, Figueras *et al.* (2021) used a pressurized 10 $L\ CSTR$ operating at 55°C and 4 bars to explore high-pressure effects on a continuous lab-scale pilot. The reactor was made of three Rushton turbines operated at 1000 rpm. Syngas was added with a flow rate of 7.5 NL/h. It was composed by CO, H_2, and CO_2 with a ratio 0.4:0.4:0.2, respectively. The biological methanation of syngas was successfully developed with a conversion of CO and H_2 to 97 and 98%, respectively and values close to 6.8 $mmol/L/h$ was achieved in CH_4 productivity.

Recently, Ngu *et al.* (2023) explored the H_2 gas–liquid mass transfer of biological methanation. The experiment used a 22 L bubble column with an initial liquid height of 1200 mm and internal diameter of 150 mm operating at 55°C and atmospheric pressure. H_2 and CO_2 were fed with a ratio of 4:1 to the column. The unreacted gas was recirculated with a flow rate of 2 NL/min to increase the H_2 conversion. Two types of gas sparger were tested, a 4-point porous sparger (made of 4 small glass sintered diffusers) and a uniform porous plate (single porous sintered diffuser occupying the bottom cross-section of the column). The results showed that a uniform porous plate favored a more intense mass transfer, hence the biological methanation process. At an inlet gas flow rate of 0.14 NL/min, a CH_4 production rate of 90±2 and 77±3 $mL/L/h$, with a CH_4 content of 80±2 and 44±3% were obtained, using the porous plate and 4-points sparger, respectively. Additionally, experimental results were compared with simulations obtained with 1D and Computational Fluid Dynamics (CFD) hydrodynamic models. Lower inlet gas flowrate leads to higher CH_4 purity but with lower productivity, which was consistent with the experimental results. Finally, the 1D model was used to analyze and compare the effect of increasing the inlet gas flow rate and the inlet bubble diameter with the experimental results. It was found that productivity increased when the inlet gas flow rate increased. Nevertheless, it did not imply higher CH_4 purity. Moreover, a decrease in the inlet

bubble diameter leads to an increase in CH_4 purity. It means that smaller bubbles offer a higher interfacial area for mass transfer and higher mass transfer efficiency.

1.2.3 Process Configurations

As previously mentioned, the biological methanation process is commonly limited by the gas-liquid mass transfer, especially the H_2 mass transfer (Jensen *et al.*, 2018). The biological methanation process has been studied on several reactor types to overcome this limitation. The most common reactors are $CSTR$, trickling beds reactor, and BCR.

1.2.3.1 Stirred tank reactors

Stirred thank reactors are the typical reactors used to develop the biomethanation process. The agitation mechanism guarantees the homogeneous mixing between the gas phase, components in the liquid phase, microorganisms, and the correct temperature distribution over the system. The volumetric mass transfer coefficient is affected by several factors in this type of reactor, such as the geometry of the reactor, impeller configuration, agitation speed, and gas flow rate (Paniagua *et al.*, 2022). Mass transfer rate increases are often related to high agitation speeds and high syngas flow rates, causing the break up of large bubbles into smaller ones (Diender *et al.*, 2018; Jensen *et al.*, 2021; Jiang *et al.*, 2022).

Luo and Angelidaki (2012) proposed increasing the agitation speed for upgrading biogas. The experiments were developed in an *ex-situ* operation to enrich the hydrogenotrophic methanogens for two months. Two substrates were used: sewage sludge at 37°C and manure at 55°C with a syngas content of H_2:CO_2 (4:1). Thermophilic conditions showed more efficiency than the mesophilic condition, with a conversion rate of 320 $mLCH4/gVS/h$. Therefore, a $CSTR$ with a working volume of 600 mL was proposed to evaluate the effect of different operating conditions. The $CSTR$ was operated at 55°C with a volume of 600 mL in 5 stages, stage I (0-10 days), stage II (11-43 days), stage III (44-73 days), stage IV (74-96 days), and stage V (97-135 days), with syngas (H_2, CH_4, and CO_2: 0.6/0.25/0.15) additions of 3, 6, 12, 12 and 24 $L/Lr/d$, respectively. The agitation was kept constant at 500 rpm for the first 3 stages and then at 800 rpm for the last two stages to increase the mass transfer rate. The authors demonstrated the feasibility of *ex-situ* biological methanation and exploration in the gas-liquid

mass transfer limitations. They produced biogas (95% CH_4) with a decrease of 90% between 44-73 days due to the gas-liquid mass transfer limitations and CH_4 yields close to 0.23 LCH_4/LH_2 over all the process.

Jensen *et al.* (2018) studied the development of mass transfer technologies for *in-situ* biological methanation in a full-scale AD reactor (agitated at 19 rpm) by increasing the syngas flow rates. The authors developed a full-scale venturi-type H_2 injector to upgrade the biogas. Seven experiments were developed in a reactor with a volume of 1110 m^3 operating at 52°C. They used as substrate a mixture of manure, straw briquettes, grass, and maize silage that was added to the system at different OLR. The volume of H_2 added to the system ranged from 3.4 to 33.5 m^3, which were added using the venturi-type injector system at flow rates ranging from 20 to 65 m^3/h. Six of the seven experiments used a recirculation of headspace gas (~100–120 m^3/h). It was found that H_2 consumption rates ranged between 0.03 and 0.25 $L/m^3/min$ during the increased addition of H_2. A value lower than 0.01 $L/m^3/min$ was obtained in the system without recirculation, contrary to values between 0.02 to 0.07 $L/m^3/min$ obtained with the recirculation. These results indicate that recirculation improved overall H_2 consumption.

1.2.3.2 Trickling bed filters

Trickling bed filters comprise a column packed with inert materials of high specific surface area, on which biofilm is developed. The syngas is added through the reactor, and the liquid phase is trickled and recycled over the packing material (Paniagua *et al.*, 2022). This type of reactor is more efficient in terms of the gas-liquid mass transfer due to the low gas and liquid flow rates and higher contact surface area between the gas and liquid phase (Grimalt-Alemany *et al.*, 2018).

Asimakopoulos *et al.* (2020) used trickle bed reactors to explore the effect of mesophilic and thermophilic conditions. They used a reactor consisting of a trickle bed column of borosilicate glass with a total packed bed volume of 180 mL and a height/diameter ratio of 4.18. The packed bed consisted of polypropylene/polyethylene packing material. The liquid and gas phase were flowing as co-currents entering at the top and leaving at the bottom. The recirculation of the

liquid phase was performed at a constant flow rate of 200 mL/min. As mentioned in Section 1.2.1.2., high temperatures achieved higher CH_4 productivities (8.49 $mmol/L/h$) and higher conversion efficiencies (H_2/CO 89 and 73%, respectively) of H_2 and CO concerning the mesophilic conditions. Sieborg *et al.* (2020) used ex-*situ* trickle bed reactors for the biological CO_2 methanation with polyurethane foam as packing material and cattle manure. The reactor was operated under different gas retention times at mesophilic conditions. The reactor consisted of a trickle bed column of polyvinyl chloride pipe with a total length of 60 cm and an internal diameter of 2.72 cm with a total packed bed volume of 291 mL and a height/diameter ratio of 8.4. The liquid and gas phase were flowing co-current entering at the top and leaving at the bottom. A distributor plate was fitted at the top of the reactor to provide even distribution of the liquid phase. The trickle filters performance in terms of outlet gas composition, conversion efficiency, and the specific CH_4 production capacity was investigated at five different values of gas retention time (GRT), $4h$, $3h$, $2.25h$, $1.75h$, $1.32h$ for period 1 (0-16 days), period 2 (17-25 days), period 3 (26-39 days), period 4 (40-51 days), and period 5 (51-60 days), respectively. The best results were obtained with the shortest gas retention time ($1.32h$) under thermophilic conditions. The specific CH_4 productivity was 2.08±0.04 $Nm^3/m^3/d$ and CH_4 yield with respect to CO_2 and H_2 were 1.04 ±0.09 mol_{CH4}/mol_{CO2}, and 0.18±0.01 mol_{CH4}/mol_{H2}, respectively.

1.2.3.3 Bubble column and gas-lift bioreactors

Bubble column and gas-lift reactors have been explored in biological methanation processes due to their advantages, such as high gas-liquid interfacial, high volumetric mass transfer coefficient, non-mechanical mixing, and relatively low cost of operation (Grimalt-Alemany *et al.*, 2018). In the research performed by Laguillaumie *et al.* (2022), a pilot-scale BCR with a working volume of 20 L was operated for *ex-situ* biological methanation of H_2:CO_2 at 55°C with a mixed microbial culture. The aim was to investigate the reactivity of the biological methanation process in a dynamic operation mode, such as gas load variations and feed intermittence. The authors found that CH_4 production rates increased linearly with the loading rate, indicating the system's non-limiting gas-liquid mass transfer capacity.

Kougias *et al.* (2017) compared $CSTR$ and BCR for *ex-situ* biogas upgrading. The gas and liquid flow rates in both reactors were 3 $L/L/d$ and 80 mL/d, respectively. The liquid and gas HRT were 15 mL/d and 8 h, respectively. Two different gas recirculation rates were applied to evaluate the gas-liquid mass transfer process; 4 L/h during the days 0–18 (Period I) and 12 L/h during the days 19–51 (Period II). The author concluded that the most efficient biological methanation was performed in the BCR with a CH_4 content of 76% and 97–98% for periods I and II, respectively. Concerning the $CSTR$, a value close to 54% was achieved for both periods. The poor conversion of the gas substrates in the $CSTR$ was mainly due to the limited gas-liquid mass transfer rate, which the increase of the agitation speed can improve. The increase in the CH_4 content between the two periods highlighted the recirculation effect in the BCR.

Guiot *et al.* (2011) used a closed-loop gas lift reactor with a working liquid volume of 30L using granular sludge and supplied with CO. The authors evaluated the production of methane and other metabolites, at different gas dilutions, feeding, and recirculation rates. The reactor temperature and pH were controlled at 35±2°C and 7.1±0.2, respectively. The experiment was developed in 6 stages, where the CO loading rate ranged from 15.0 to 122 $mmol/gVS/d$, the gas recirculation ratio from 4:1 to 20:1, and the partial pressure of CO between 0.42 and 0.96 atm, regarding the stage one: 1.744 $mmol/gVS/d$, 20:1, and 0.62 atm for CO loading rate, gas recirculation ratio and CO partial pressure, respectively. The authors achieving a CO conversion efficiency of 75% with a CH_4 production of 2.92±0.09 $mmol/gVS/d$. However, in Stage six: 122 $mmol/gVS/d$, 4:1, and 0.96 atm for CO loading rate, gas recirculation ratio, and CO partial pressure, respectively. The CO conversion efficiency was reduced to 17%, concluding that high gas recirculation can be more effective in improving the CO gas-liquid mass transfer compared to higher CO partial pressures in a gas lift reactor.

1.3 Conclusions Anaerobic Digestion & Biological Methanation

The biological methanation process could be applied as a biogas upgrading technology after the methanogenesis step in the AD. Its application and development are challenging given the multiple factors to consider, which leads to questions such as what type of substrates? Which strategy could be considered, *in-situ* or *ex-situ*? Which type of reactor configuration? What are

the key phenomena affecting the process? All those questions need to be evaluated to obtain optimal performance of biological methanation.

Several authors studied biological methanation to upgrade its performance in terms of yields and productivities. Some of these researchers have focused on exploring the effect of various factors. The flow rate of gases such as H_2 and CO could lead to an accumulation of VFA and posterior inhibition of the biological methanation process (Ashraf *et al.*, 2020; C. Li *et al.*, 2020; W. Wang *et al.*, 2013). Concerning temperature, the operation at mesophilic and thermophilic conditions are widely used. Nevertheless, the latest could be more suitable for industrial applications due to the thermodynamic limitations at mesophilic conditions (Grimalt-Alemany *et al.*, 2020). Mixed cultures are more robust and do not require sterile conditions (Laguillaumie *et al.*, 2022; Rachbauer *et al.*, 2017). Increasing the agitation in $CSTR$, the pressure in a pressurized agitated column, and the recirculation ratio in the gas lift reactor improves the mass transfer in the biological methanation (Guiot *et al.*, 2011; Figueras *et al.*, 2021; Andreides *et al.*, 2022). The reactor configuration plays an essential role in biological methanation. For a $CSTR$, mass transfer rate increases are often related to the increase in the agitation speed and syngas flow rates, both contributing to the high gas-liquid interfacial area (Jensen *et al.*, 2018; Luo & Angelidaki, 2012; Ngu *et al.*, 2023; Paniagua *et al.*, 2022). The *ex-situ* BCR showed better results than *ex-situ* $CSTR$ reactors in terms of CH_4 content; the poor conversion of the gas substrates in the $CSTR$ is limited by the gas-liquid mass transfer rate (Kougias *et al.*, 2017).

Table 1.1.3 summarizes the characteristics of the biological methanation reported through the literature used in this thesis, which allowed us to highlight some questions that will be addressed in this thesis:

- Which are the best conditions to carry out biological methanation in different types of reactors such as $CSTR$ or BCR and what is the effect of using different kinds of substrates varying GLR and OLR?
- Which are the optimal operating conditions to improve the yields and productivities of biological methanation? Can we set them automatically?
- Can the variations in the composition of the syngas be used to upgrade the biological methanation and its effect on the mass transfer process?

- How does the on-line monitoring of products and substrates such as sugars, VFA, H_2, and CH_4 over time help to improve the biological methanation?
- Can the characterization of microorganisms provide information on the biological methanation process performance?

Table 1.1.3. Summary of performances and operating conditions of biological methantion reported in the literature.

Substrate	Reactor design and strategy *(ex-situ / in-situ)*	Temperature (°C)	pH	Pressure (atm)	Agitation (rpm)	Time	Syngas Content	CH_4			Reference
								Yield	Content (%)	Productivity	
Sewage sludge	600 mL $CSTR$ - *ex-situ* operation	37	NR	1.5	300	135d	H_2 / CH_4 /CO_2 (60/25/15)	NR	NR	NR	(Luo & Angelidaki, 2012)
Digested manure		55	7.8	1*	500-800			0.23 L/LH_2	89.9-95.4	1.5-5.3 $L/L/d$	
Cattle Manure	Serum bottles - *ex-situ* batch operation	55	8	0.25-1.0	100-300	25h	H_2/CO_2 (4/1)	NR	NR	3.7-15.5 $mL/L/h$ 61-67.8 $mL/L/h$	(Luo *et al.*, 2012)
	3.5 L $CSTR$		8.3	NR	65	60d		200 mL/gVS	65	18.9 $mL/L/h$	
Manure-based Straw Briquettes Grass silage Maize silage	1110 m^3 Anaerobic digester - *in-situ* batch operation	52	5.5-7.8	1	19	120d	H_2	0.003-0.31 mol/mol_{H2}	44.8-63.6	NR	(Jensen *et al.*, 2018)
Sewage sludge	2 L $CSTR$ with a hollow fiber membrane module - *in-situ* operation	37	8.0	2.6	200	103d	Coke oven gas H_2:CO (0.92:0.08)	NR	89.9-98.8	436.5– 625.5$mL/L/d$	(W. Wang *et al.*, 2013)

¶estimated from a figure; *assumed atmospheric pressure; $conditions for H_2:CO effect; ±Initial proportion not specified; TSS: Total suspended solids; NR: Not reported.

Table 1.1.3. Continuation summary of performances and operating conditions of biological methantion reported in the literature.

Substrate	Reactor design and strategy *(ex-situ / in-situ)*	Temperature (°C)	pH	Pressure (atm)	Agitation (rpm)	Time	Syngas Content	CH_4			Reference
								Yield	Content (%)	Productivity	
Granular sludge	110 mL bottles	37	7.2	2.6	100	230d	Blast furnace gas CO/ $CO_2/H_2/N_2$ (22/22/4/52)	5.32 $mmol/gVS$	NR	4.19 $mmol/gVS/d$	(Y. Wang *et al.*, 2018)
Maize leaf	120 mL bottles - *in-situ* batch operation	52	7-8	1*	100	24d	H_2	NR	87.9-89.4	267-272 $mL/L/d$	(Mulat *et al.*, 2017)
Cattle manure Sewage sludge Gaseous H_2:CO_2	118 mL Serum bottles - batch operation	55	7.8-8.8	1*	200	500h	CO: CO_2:H_2 (1:0:3)	NR	95	NR	(C. Li *et al.*, 2020)
Sludge	10 L $CSTR$	55	6.0	3.95 (4 bar)	1000	74d	CO: H_2:CO_2 (0.4:0.4:0.2)	NR	NR	6.8 $mmol/L/h$	(Figueras *et al.*, 2021)

¶estimated from a figure; *assumed atmospheric pressure; $conditions for H_2:CO effect; ±Initial proportion not specified; TSS: Total suspended solids; NR: Not reported.

Table 1.1.3. Continuation summary of performances and operating conditions of biological methantion reported in the literature.

Substrate	Reactor design and strategy *(ex-situ / in-situ)*	Temperature (°C)	pH	Pressure (atm)	Agitation (rpm)	Time	Syngas Content	CH_4			Reference
								Yield	Content (%)	Productivity	
Sludge mixture	180 mL Trickle bed reactors *polypropylene /polyethylene as packing material with a mass flow of 10 mL/min	37 60	7.0	1.3	NA	9d	H_2/CO_2 CO/N2 (45/25/20/10)	NR	NR	3.13 $mmol/L_{bed}/h$ 8.49 $mmol/L_{bed}/h$	(Asimakopoulos *et al.*, 2020)
Anaerobic sludge	1.7 L semi-continuous stirred tank reactor	37	7.2	1*	400	95d[¶]	H_2/CO (3/1, 4/1 and 5/1)	NR	66.37-73.35	NR	(Y. Li *et al.*, 2020)[$]
Sewage sludge	0.00578 m^3 Trickle-bed reactor *Polypropylene packing rings as packing material with a mass flow of 250 mL/min	38	7.0	1*	125	4d	H_2/CO_2 (0.8/0.2)	NR	NR	0.55 $mmol$	(Rachbauer *et al.*, 2017)

[¶]estimated from a figure; *assumed atmospheric pressure; [$]conditions for H_2:CO effect; [±]Initial proportion not specified; TSS: Total suspended solids; NR: Not reported.

Table 1.1.3. Continuation summary of performances and operating conditions of biological methantion reported in the literature.

Substrate	Reactor design and strategy *(ex-situ / in-situ)*	Temperature (°C)	pH	Pressure (atm)	Agitation (rpm)	Time	Syngas Content	CH_4			Reference
								Yield	Yield Content (%)	Productivity	
Anaerobic sludge	100mL flasks - *in-situ* operation	37	6.9	2	100	17d	CO:CO_2[±]	1.83$mmol/gVS/h$	NR	NR	(Grimalt-Alemany *et al.*, 2020)
		60	7.1					33.48 $mmol/gVS/h$	NR	NR	
Household wastes Duck manures, Bovine manures	20 L BCR - *ex situ* operación	55	6-8[¶]	1*	NR	405d	CO:H_2 (ratio 4.2)	1 $mol/gTSS/d$	NR	3.97 $NL/L/d$	(Laguillaumie *et al.*, 2022)
Thermophilic digestate	1.4 L STR - *ex-situ* operation	52	8.0	1*	300	51d	CH_4:CO_2:H_2 (23:15:62)	NR	54-76	NR	(Kougias *et al.*, 2017)
	1.2 L Bubble column				NA				76-98		

[¶]estimated from a figure; *assumed atmospheric pressure; [$]conditions for H_2:CO effect; [±]Initial proportion not specified; TSS: Total suspended solids; NR: Not reported.

Table 1.1.4. Continuation performance characteristics in biological methanation reported in the literature.

Substrate	Reactor design and strategy *(ex-situ / in-situ)*	Temperature (°C)	pH	Pressure (atm)	Agitation (rpm)	Time	Syngas Content	CH_4			Reference
								Yield	Yield Content	Productivity	
Cow manure	291×10^{-6} m^3 Trickled filter – *ex-situ* operation *polyurethane foam (surface area 560–580 m2 m-3) as packing material with a mass flow 50 mL/min	52	7.9	1	NA	215d	CO_2:H_2 0.273	NR	NR	8.54 $Nm^3/m^3_{bed}/d$	(Ashraf *et al.*, 2020)
Granular sludge	30L closed-loop gas lift reactor	35	6.9-7.8	2.5	100-400	100d	CO	4.77 $mmol/gVS/d$	NR	NR	(Guiot *et al.*, 2011)
Primary sludge and thickened-disintegrated waste (3:1)	10.5 L $CSTR$ – *in-situ* operation	55	7.6-8.6	1*	55	150d	H_2/CO (0.55/0.45)	0.36-0.63 L/gVS	34.8-65.1	0.31-0.57 $L/L/d$	(Andreides *et al.*, 2022)

¶estimated from a figure; *assumed atmospheric pressure; $conditions for H_2/CO effect; ±Initial proportion not specified; TSS: Total suspended solids; NR: Not reported.

1.4 References

Andreides, D., Pokorna, D., Zabranska, J., 2022. Assessing the syngas biomethanation in anaerobic sludge digestion under different syngas loading rates and homogenisation. Fuel 320, 123929. https://doi.org/10.1016/j.fuel.2022.123929

Angelidaki, I., Ellegaard, L., Ahring, B.K., 1999. A comprehensive model of anaerobic bioconversion of complex substrates to biogas. Biotechnol. Bioeng. 63, 363–372. https://doi.org/10.1002/(SICI)1097-0290(19990505)63:3<363::AID-BIT13>3.0.CO;2-Z

Angelidaki, I., Karakashev, D., Batstone, D.J., Plugge, C.M., Stams, A.J.M., 2011. Biomethanation and its potential, 1st ed, Methods in Enzymology. Elsevier Inc. https://doi.org/10.1016/B978-0-12-385112-3.00016-0

Ashraf, M.T., Sieborg, M.U., Yde, L., Rhee, C., Shin, S.G., Triolo, J.M., 2020. Biomethanation in a thermophilic biotrickling filter — pH control and lessons from long-term operation. Bioresource Technology Reports 11, 100525. https://doi.org/10.1016/j.biteb.2020.100525

Asimakopoulos, K., Łężyk, M., Grimalt-Alemany, A., Melas, A., Wen, Z., Gavala, H.N., Skiadas, I.V., 2020. Temperature effects on syngas biomethanation performed in a trickle bed reactor. Chemical Engineering Journal 393, 124739. https://doi.org/10.1016/j.cej.2020.124739

Bashir, K., Aggarwal, M., 2017. Physicochemical, thermal and functional properties of gamma irradiated chickpea starch. International Journal of Biological Macromolecules 97, 426–433. https://doi.org/10.1016/j.ijbiomac.2017.01.025

Batstone, D.J., Balthes, C., Barr, K., 2010. Model assisted startup of anaerobic digesters fed with thermally hydrolysed activated sludge. Water Science 6. https://doi.org/10.2166/wst.2010.487

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). Water Science and Technology 45, 65–73. https://doi.org/10.2166/wst.2002.0292

Begum, S., Ahuja, S., Anupoju, G.R., Kuruti, K., Juntupally, S., Gandu, B., Ahuja, D.K., 2017. Process intensification with inline pre and post processing mechanism for valorization of poultry litter through high rate biomethanation technology: A full scale experience. Renewable Energy 114, 428–436. https://doi.org/10.1016/j.renene.2017.07.049

Bharathiraja, B., Sudharsanaa, T., Bharghavi, A., Jayamuthunagai, J., Praveenkumar, R., 2016. Biohydrogen and Biogas – An overview on feedstocks and enhancement process. Fuel 185, 810–828. https://doi.org/10.1016/j.fuel.2016.08.030

Bond, T., Templeton, M.R., 2011. History and future of domestic biogas plants in the developing world. Energy for Sustainable Development 15, 347–354. https://doi.org/10.1016/j.esd.2011.09.003

Chandra, R., Takeuchi, H., Hasegawa, T., 2012. Methane production from lignocellulosic agricultural crop wastes: A review in context to second generation of biofuel production. Renewable and Sustainable Energy Reviews 16, 1462–1476. https://doi.org/10.1016/j.rser.2011.11.035

Chew, K.R., Leong, H.Y., Khoo, K.S., Vo, D.-V.N., Anjum, H., Chang, C.-K., Show, P.L., 2021. Effects of anaerobic digestion of food waste on biogas production and environmental impacts: a review. Environ Chem Lett 19, 2921–2939. https://doi.org/10.1007/s10311-021-01220-z

Czatzkowska, M., Harnisz, M., Korzeniewska, E., Koniuszewska, I., 2020. Inhibitors of the methane fermentation process with particular emphasis on the microbiological aspect: A review. Energy Science and Engineering 8, 1880–1897. https://doi.org/10.1002/ese3.609

Dar, R.A., Parmar, M., Dar, E.A., Sani, R.K., Phutela, U.G., 2021. Biomethanation of agricultural residues: Potential, limitations and possible solutions. Renewable and Sustainable Energy Reviews 135, 110217. https://doi.org/10.1016/j.rser.2020.110217

Dev, S., Saha, S., Kurade, M.B., Salama, E.S., El-Dalatony, M.M., Ha, G.S., Chang, S.W., Jeon, B.H., 2019. Perspective on anaerobic digestion for biomethanation in cold environments. Renewable and Sustainable Energy Reviews 103, 85–95. https://doi.org/10.1016/j.rser.2018.12.034

Diender, M., Uhl, P.S., Bitter, J.H., Stams, A.J.M., Sousa, D.Z., 2018. High Rate Biomethanation of Carbon Monoxide-Rich Gases via a Thermophilic Synthetic Coculture. ACS Sustainable Chem. Eng. 6, 2169–2176. https://doi.org/10.1021/acssuschemeng.7b03601

Dong, R., Qiao, W., Guo, J., Sun, H., 2022. Manure treatment and recycling technologies, in: Circular Economy and Sustainability. Elsevier, pp. 161–180. https://doi.org/10.1016/B978-0-12-821664-4.00009-1

Eggen, T., Vogelsang, C., 2015. Occurrence and Fate of Pharmaceuticals and Personal Care Products in Wastewater, in: Comprehensive Analytical Chemistry. Elsevier, pp. 245–294. https://doi.org/10.1016/B978-0-444-63299-9.00007-7

Ferry, J.G., 2011. Fundamentals of methanogenic pathways that are key to the biomethanation of complex biomass. Current Opinion in Biotechnology 22, 351–357. https://doi.org/10.1016/j.copbio.2011.04.011

Figueras, J., Benbelkacem, H., Dumas, C., Buffiere, P., 2021. "Biomethanation of syngas by enriched mixed anaerobic consortium in pressurized agitated column." Bioresource Technology 338, 125548. https://doi.org/10.1016/j.biortech.2021.125548

Giwa, A.S., Chang, F., Xu, H., Zhang, X., Huang, B., Li, Y., Wu, J., Wang, B., Vakili, M., Wang, K., 2019. Pyrolysis of difficult biodegradable fractions and the real syngas bio-methanation performance. Journal of Cleaner Production 233, 711–719. https://doi.org/10.1016/j.jclepro.2019.06.145

Grangeiro, L.C., de Almeida, S.G.C., Mello, B.S. de, Fuess, L.T., Sarti, A., Dussán, K.J., 2019. New trends in biogas production and utilization, in: Sustainable Bioenergy. Elsevier, pp. 199–223. https://doi.org/10.1016/B978-0-12-817654-2.00007-1

Grimalt-Alemany, A., Łężyk, M., Kennes-Veiga, D.M., Skiadas, I.V., Gavala, H.N., 2020. Enrichment of Mesophilic and Thermophilic Mixed Microbial Consortia for Syngas Biomethanation: The Role of Kinetic and Thermodynamic Competition. Waste Biomass Valor 11, 465–481. https://doi.org/10.1007/s12649-019-00595-z

Grimalt-Alemany, A., Skiadas, I.V., Gavala, H.N., 2018. Syngas biomethanation: state-of-the-art review and perspectives. Biofuels, Bioprod. Bioref. 12, 139–158. https://doi.org/10.1002/bbb.1826

Guiot, S.R., Cimpoia, R., Carayon, G., 2011. Potential of Wastewater-Treating Anaerobic Granules for Biomethanation of Synthesis Gas. Environ. Sci. Technol. 45, 2006–2012. https://doi.org/10.1021/es102728m

He, P.J., 2010. Anaerobic digestion: An intriguing long history in China. Waste Management 30, 549–550. https://doi.org/10.1016/j.wasman.2010.01.002

Henze, M., van Loosdrecht, M.C.M., Ekama, G.A., Brdjanovic, Damir., 2019. Biological Wastewater Treatment: Principles, Modeling and Design. IWA Publishing. https://doi.org/10.2166/9781780408613

Hupfauf, S., Plattner, P., Wagner, A.O., Kaufmann, R., Insam, H., Podmirseg, S.M., 2018. Temperature shapes the microbiota in anaerobic digestion and drives efficiency to a maximum at 45 °C. Bioresource Technology 269, 309–318. https://doi.org/10.1016/j.biortech.2018.08.106

Hupfauf, S., Winkler, A., Wagner, A.O., Podmirseg, S.M., Insam, H., 2020. Biomethanation at 45 °C offers high process efficiency and supports hygienisation. Bioresource Technology 300. https://doi.org/10.1016/j.biortech.2019.122671

Iglesias, R., Muñoz, R., Polanco, M., Díaz, I., Susmozas, A., Moreno, A.D., Guirado, M., Carreras, N., Ballesteros, M., 2021. Biogas from Anaerobic Digestion as an Energy Vector: Current Upgrading Development. Energies 14, 2742. https://doi.org/10.3390/en14102742

Jensen, M.B., Kofoed, M.V.W., Fischer, K., Voigt, N.V., Agneessens, L.M., Batstone, D.J., Ottosen, L.D.M., 2018. Venturi-type injection system as a potential H2 mass transfer technology for full-scale in situ biomethanation. Applied Energy 222, 840–846. https://doi.org/10.1016/j.apenergy.2018.04.034

Jensen, M.B., Ottosen, L.D.M., Kofoed, M.V.W., 2021. H2 gas-liquid mass transfer: A key element in biological Power-to-Gas methanation. Renewable and Sustainable Energy Reviews 147, 111209. https://doi.org/10.1016/j.rser.2021.111209

Jiang, B., Zhang, D., Hu, X., Söderlind, U., Paladino, G., Gamage, S., Hedenström, E., Zhang, W., Arrigoni, J., Lundgren, A., Tuvesson, M., Yu, C., 2022. Low-Grade Syngas Biomethanation in Continuous Reactors with Respect to Gas–Liquid Mass Transfer and Reactor Start-Up Strategy. Fermentation 9, 38. https://doi.org/10.3390/fermentation9010038

Karekar, S., Stefanini, R., Ahring, B., 2022. Homo-Acetogens: Their Metabolism and Competitive Relationship with Hydrogenotrophic Methanogens. Microorganisms 10, 397. https://doi.org/10.3390/microorganisms10020397

Klinkner, B.A., 2014. Anaerobic Digestion as a Renewable Energy Source and Waste Management Technology: What Must be Done for This Technology to Realize Success in the United States? University of Massachusetts Law Revie.

Kougias, P.G., Treu, L., Benavente, D.P., Boe, K., Campanaro, S., Angelidaki, I., 2017. Ex-situ biogas upgrading and enhancement in different reactor systems. Bioresource Technology 225, 429–437. https://doi.org/10.1016/j.biortech.2016.11.124

Krishania, M., Kumar, V., Vijay, V.K., Malik, A., 2013. Analysis of different techniques used for improvement of biomethanation process: A review. Fuel 106, 1–9. https://doi.org/10.1016/j.fuel.2012.12.007

Labatut, R.A., Pronto, J.L., 2018. Sustainable Waste-to-Energy Technologies: Anaerobic Digestion, in: Sustainable Food Waste-To-Energy Systems. Elsevier, pp. 47–67. https://doi.org/10.1016/B978-0-12-811157-4.00004-8

Laguillaumie, L., Rafrafi, Y., Moya-Leclair, E., Delagnes, D., Dubos, S., Spérandio, M., Paul, E., Dumas, C., 2022. Stability of ex situ biological methanation of H2/CO2 with a mixed microbial culture in a pilot scale bubble column reactor. Bioresource Technology 354, 127180. https://doi.org/10.1016/j.biortech.2022.127180

Laiq Ur Rehman, M., Iqbal, A., Chang, C., Li, W., Ju, M., 2019. Anaerobic digestion. Water Environment Research 91, 1253–1271. https://doi.org/10.1002/wer.1219

Li, C., Zhu, X., Angelidaki, I., 2020. Carbon monoxide conversion and syngas biomethanation mediated by different microbial consortia. Bioresource Technology 314, 123739. https://doi.org/10.1016/j.biortech.2020.123739

Li, L., Peng, X., Wang, X., Wu, D., 2018. Anaerobic digestion of food waste: A review focusing on process stability. Bioresource Technology 248, 20–28. https://doi.org/10.1016/j.biortech.2017.07.012

Li, Y., Wang, Z., He, Z., Luo, S., Su, D., Jiang, H., Zhou, H., Xu, Q., 2020. Effects of temperature, hydrogen/carbon monoxide ratio and trace element addition on methane production performance from syngas biomethanation. Bioresource Technology 295, 122296. https://doi.org/10.1016/j.biortech.2019.122296

Lide, D.R., 2005. CRC handbook of chemistry and physics: a ready-reference book of chemical and physical data, 86th ed. CRC Press, Florida, FL.

Lindmark, J., Thorin, E., Bel Fdhila, R., Dahlquist, E., 2014. Effects of mixing on the result of anaerobic digestion: Review. Renewable and Sustainable Energy Reviews 40, 1030–1047. https://doi.org/10.1016/j.rser.2014.07.182

Lindner, J., Zielonka, S., Oechsner, H., Lemmer, A., 2015. Effect of different pH-values on process parameters in two-phase anaerobic digestion of high-solid substrates. Environmental Technology 36, 198–207. https://doi.org/10.1080/09593330.2014.941944

Liu, R., Hao, X., Wei, J., 2016. Function of homoacetogenesis on the heterotrophic methane production with exogenous H2/CO2 involved. Chemical Engineering Journal 284, 1196–1203. https://doi.org/10.1016/j.cej.2015.09.081

Luo, G., Angelidaki, I., 2012. Integrated biogas upgrading and hydrogen utilization in an anaerobic reactor containing enriched hydrogenotrophic methanogenic culture. Biotechnology and Bioengineering 109, 2729–2736. https://doi.org/10.1002/bit.24557

Luo, G., Johansson, S., Boe, K., Xie, L., Zhou, Q., Angelidaki, I., 2012. Simultaneous hydrogen utilization and in situ biogas upgrading in an anaerobic reactor. Biotechnology and Bioengineering 109, 1088–1094. https://doi.org/10.1002/bit.24360

Lusk, P., 1998. Methane Recovery from Animal Manures The Current Opportunities Casebook (No. NREL/SR--580-25145, 1364). https://doi.org/10.2172/1364

Lyu, Z., Whitman, W.B., 2019. Transplanting the pathway engineering toolbox to methanogens. Current Opinion in Biotechnology 59, 46–54. https://doi.org/10.1016/j.copbio.2019.02.009

Ma, S., Ma, H., Hu, H., Ren, H., 2019. Effect of mixing intensity on hydrolysis and acidification of sewage sludge in two-stage anaerobic digestion: Characteristics of dissolved organic matter and the key microorganisms. Water Research 148, 359–367. https://doi.org/10.1016/j.watres.2018.10.058

Malinowsky, C., Nadaleti, W., Debiasi, L.R., Gonçalves Moreira, A.J., Bayard, R., Borges de Castilhos Junior, A., 2021. Start-up phase optimization of two-phase anaerobic digestion of food waste: Effects of organic loading rate and hydraulic retention time. Journal of Environmental Management 296, 113064. https://doi.org/10.1016/j.jenvman.2021.113064

Mao, C., Feng, Y., Wang, X., Ren, G., 2015. Review on research achievements of biogas from anaerobic digestion. Renewable and Sustainable Energy Reviews 45, 540–555. https://doi.org/10.1016/j.rser.2015.02.032

Meegoda, J., Li, B., Patel, K., Wang, L., 2018. A Review of the Processes, Parameters, and Optimization of Anaerobic Digestion. IJERPH 15, 2224. https://doi.org/10.3390/ijerph15102224

Merkel, W., Krauth, K., 1999. Mass transfer of carbon dioxide in anaerobic reactors under dynamic substrate loading conditions. Water Research 33, 2011–2020. https://doi.org/10.1016/S0043-1354(98)00434-5

Mortezaei, Y., Williams, M.R., Demirer, G.N., 2023. Effect of temperature and solids retention time on the removal of antibiotic resistance genes during anaerobic digestion of sludge. Bioresource Technology Reports 21, 101377. https://doi.org/10.1016/j.biteb.2023.101377

Mulat, D.G., Mosbæk, F., Ward, A.J., Polag, D., Greule, M., Keppler, F., Nielsen, J.L., Feilberg, A., 2017. Exogenous addition of H 2 for an in situ biogas upgrading through biological reduction of carbon dioxide into methane. Waste Management 68, 146–156. https://doi.org/10.1016/j.wasman.2017.05.054

Nasir, I.M., Mohd Ghazi, T.I., Omar, R., 2012. Anaerobic digestion technology in livestock manure treatment for biogas production: A review: Anaerobic digestion technology in livestock manure treatment for biogas production. Eng. Life Sci. 12, 258–269. https://doi.org/10.1002/elsc.201100150

Ngu, V., Fletcher, D.F., Kavanagh, J.M., Rafrafi, Y., Dumas, C., Morchain, J., Cockx, A., 2023. H2 mass transfer – A key factor for efficient biological methanation: Comparison between pilot-scale experimental data, 1D and CFD models. Chemical Engineering Science 268, 118382. https://doi.org/10.1016/j.ces.2022.118382

Nie, E., He, P., Zhang, H., Hao, L., Shao, L., Lü, F., 2021. How does temperature regulate anaerobic digestion? Renewable and Sustainable Energy Reviews 150, 111453. https://doi.org/10.1016/j.rser.2021.111453

Ossa-Arias, M. del M., González-Martínez, S., 2021. Methane Production from the Organic Fraction of Municipal Solid Waste Under Psychrophilic, Mesophilic, and Thermophilic Temperatures at Different Organic Loading Rates. Waste Biomass Valor 12, 4859–4871. https://doi.org/10.1007/s12649-021-01354-9

Pan, X., Zhao, L., Li, C., Angelidaki, I., Lv, N., Ning, J., Cai, G., Zhu, G., 2021. Deep insights into the network of acetate metabolism in anaerobic digestion: focusing on syntrophic acetate oxidation and homoacetogenesis. Water Research 190, 116774. https://doi.org/10.1016/j.watres.2020.116774

Paniagua, S., Lebrero, R., Muñoz, R., 2022. Syngas biomethanation: Current state and future perspectives. Bioresource Technology 358, 127436. https://doi.org/10.1016/j.biortech.2022.127436

Perry, R.H., Green, D.W. (Eds.), 1999. Perry's chemical engineers' handbook, 7th ed, Perry's chemical engineers' platinum edition. McGraw-Hill, New York, NY.

Rachbauer, L., Beyer, R., Bochmann, G., Fuchs, W., 2017. Characteristics of adapted hydrogenotrophic community during biomethanation. Science of The Total Environment 595, 912–919. https://doi.org/10.1016/j.scitotenv.2017.03.074

Rafrafi, Y., Laguillaumie, L., Dumas, C., 2020. Biological Methanation of H2 and CO2 with Mixed Cultures: Current Advances, Hurdles and Challenges. Waste and Biomass Valorization. https://doi.org/10.1007/s12649-020-01283-z

Raposo, F., De la Rubia, M.A., Fernández-Cegrí, V., Borja, R., 2012. Anaerobic digestion of solid organic substrates in batch mode: An overview relating to methane yields and experimental procedures. Renewable and Sustainable Energy Reviews 16, 861–877. https://doi.org/10.1016/j.rser.2011.09.008

Roopnarain, A., Adeleke, R., 2017. Current status, hurdles and future prospects of biogas digestion technology in Africa. Renewable and Sustainable Energy Reviews 67, 1162–1179. https://doi.org/10.1016/j.rser.2016.09.087

Rusmanis, D., O'Shea, R., Wall, D.M., Murphy, J.D., 2019. Biological hydrogen methanation systems – an overview of design and efficiency. Bioengineered 10, 604–634. https://doi.org/10.1080/21655979.2019.1684607

Saady, N.M.C., 2013. Homoacetogenesis during hydrogen production by mixed cultures dark fermentation: Unresolved challenge. International Journal of Hydrogen Energy 38, 13172–13191. https://doi.org/10.1016/j.ijhydene.2013.07.122

Saha, S., Basak, B., Hwang, J.H., Salama, E.S., Chatterjee, P.K., Jeon, B.H., 2020. Microbial Symbiosis: A Network towards Biomethanation. Trends in Microbiology 1–17. https://doi.org/10.1016/j.tim.2020.03.012

Schön, M., 2009. Numerical Modelling of Anaerobic Digestion Processes in Agricultural Biogas Plants Dissertation.

Sendilvadivelu, A., Dhandapani, B., Vijayasimhan, S., 2022. A Short Review on Feedstock Characteristics in Methane Production from Municipal Solid Waste. Architecture, Civil Engineering, Environment 15, 75–85. https://doi.org/10.2478/acee-2022-0032

Sieborg, M.U., Jønson, B.D., Ashraf, M.T., Yde, L., Triolo, J.M., 2020. Biomethanation in a thermophilic biotrickling filter using cattle manure as nutrient media. Bioresource Technology Reports 9, 100391. https://doi.org/10.1016/j.biteb.2020.100391

Sillero, L., Solera, R., Pérez, M., 2023. Thermophilic-mesophilic temperature phase anaerobic co-digestion of sewage sludge, wine vinasse and poultry manure: Effect of hydraulic retention time on mesophilic-methanogenic stage. Chemical Engineering Journal 451, 138478. https://doi.org/10.1016/j.cej.2022.138478

Siqueira, J.C. de, Assemany, P., Siniscalchi, L.A.B., 2022. Microbial dynamics and methanogenic potential of co-digestion of sugarcane vinasse and dairy secondary effluent in an upflow anaerobic sludge blanket reactor. Bioresource Technology 361, 127654. https://doi.org/10.1016/j.biortech.2022.127654

Strobel, G., Hagemann, B., Huppertz, T.M., Ganzer, L., 2020. Underground bio-methanation: Concept and potential. Renewable and Sustainable Energy Reviews 123, 109747. https://doi.org/10.1016/j.rser.2020.109747

Sun, H., Yang, Z., Zhao, Q., Kurbonova, M., Zhang, R., Liu, G., Wang, W., 2021. Modification and extension of anaerobic digestion model No.1 (ADM1) for syngas biomethanation simulation: From lab-scale to pilot-scale. Chemical Engineering Journal 403, 126177. https://doi.org/10.1016/j.cej.2020.126177

Tena, M., Perez, M., Solera, R., 2021. Effect of hydraulic retention time on the methanogenic step of a two-stage anaerobic digestion system from sewage sludge and wine vinasse: Microbial and kinetic evaluation. Fuel 296, 120674. https://doi.org/10.1016/j.fuel.2021.120674

Ullrich, T., Lindner, J., Bär, K., Mörs, F., Graf, F., Lemmer, A., 2018. Influence of operating pressure on the biological hydrogen methanation in trickle-bed reactors. Bioresource Technology 247, 7–13. https://doi.org/10.1016/j.biortech.2017.09.069

Ünyay, H., Yılmaz, F., Başar, İ.A., Altınay Perendeci, N., Çoban, I., Şahinkaya, E., 2022. Effects of organic loading rate on methane production from switchgrass in batch and semi-continuous stirred tank reactor system. Biomass and Bioenergy 156, 106306. https://doi.org/10.1016/j.biombioe.2021.106306

Van, D.P., Fujiwara, T., Leu Tho, B., Song Toan, P.P., Hoang Minh, G., 2019. A review of anaerobic digestion systems for biodegradable waste: Configurations, operating parameters, and current trends. Environmental Engineering Research 25, 1–17. https://doi.org/10.4491/eer.2018.334

Voelklein, M.A., Rusmanis, D., Murphy, J.D., 2019. Biological methanation: Strategies for in-situ and ex-situ upgrading in anaerobic digestion. Applied Energy 235, 1061–1071. https://doi.org/10.1016/j.apenergy.2018.11.006

Voelklein, M.A., Rusmanis, D., Murphy, J.D., 2016. Increased loading rates and specific methane yields facilitated by digesting grass silage at thermophilic rather than mesophilic temperatures. Bioresource Technology 216, 486–493. https://doi.org/10.1016/j.biortech.2016.05.109

Wang, P., Wang, H., Qiu, Y., Ren, L., Jiang, B., 2018. Microbial characteristics in anaerobic digestion process of food waste for methane production–A review. Bioresource Technology 248, 29–36. https://doi.org/10.1016/j.biortech.2017.06.152

Wang, W., Xie, L., Luo, G., Zhou, Q., Angelidaki, I., 2013. Performance and microbial community analysis of the anaerobic reactor with coke oven gas biomethanation and in situ biogas upgrading. Bioresource Technology 146, 234–239. https://doi.org/10.1016/j.biortech.2013.07.049

Wang, Y., Yin, C., Liu, Y., Tan, M., Shimizu, K., Lei, Z., Zhang, Z., Sumi, I., Yao, Y., Mogi, Y., 2018. Biomethanation of blast furnace gas using anaerobic granular sludge *via* addition of hydrogen. RSC Adv. 8, 26399–26406. https://doi.org/10.1039/C8RA04853C

Weiland, P., 2010. Biogas production: current state and perspectives. Appl Microbiol Biotechnol 85, 849–860. https://doi.org/10.1007/s00253-009-2246-7

Welander, P. V., Metcalf, W.W., 2005. Loss of the mtr operon in Methanosarcina blocks growth on methanol, but not methanogenesis, and reveals an unknown methanogenic pathway. Proceedings of the National Academy of Sciences of the United States of America 102, 10664–10669. https://doi.org/10.1073/pnas.0502623102

Westerholm, M., Moestedt, J., Schnürer, A., 2016. Biogas production through syntrophic acetate oxidation and deliberate operating strategies for improved digester performance. Applied Energy 179, 124–135. https://doi.org/10.1016/j.apenergy.2016.06.061

Wu, B., Lin, R., Kang, X., Deng, C., Dobson, A.D.W., Murphy, J.D., 2021. Improved robustness of ex-situ biological methanation for electro-fuel production through the addition of graphene. Renewable and Sustainable Energy Reviews 152, 111690. https://doi.org/10.1016/j.rser.2021.111690

Xu, F., Li, Yangyang, Ge, X., Yang, L., Li, Yebo, 2018. Anaerobic digestion of food waste – Challenges and opportunities. Bioresource Technology 247, 1047–1058. https://doi.org/10.1016/j.biortech.2017.09.020

Yang, H., Deng, L., 2020. Using air instead of biogas for mixing and its effect on anaerobic digestion of animal wastewater with high suspended solids. Bioresource Technology 318, 124047. https://doi.org/10.1016/j.biortech.2020.124047

Zamri, M.F.M.A., Hasmady, S., Akhiar, A., Ideris, F., Shamsuddin, A.H., Mofijur, M., Fattah, I.M.R., Mahlia, T.M.I., 2021. A comprehensive review on anaerobic digestion of organic fraction of municipal solid waste. Renewable and Sustainable Energy Reviews 137, 110637. https://doi.org/10.1016/j.rser.2020.110637

Zhang, J., Mao, L., Nithya, K., Loh, K.-C., Dai, Y., He, Y., Wah Tong, Y., 2019. Optimizing mixing strategy to improve the performance of an anaerobic digestion waste-to-energy system for energy recovery from food waste. Applied Energy 249, 28–36. https://doi.org/10.1016/j.apenergy.2019.04.142

Zhou, H., Jiang, J., Zhao, Q., Li, L., Wang, K., Wei, L., 2022. Effects of organic loading rates on high-solids anaerobic digestion of food waste in horizontal flow reactor: Methane production,

stability and mechanism. Chemosphere 293, 133650. https://doi.org/10.1016/j.chemosphere.2022.133650

Zupančič, M., Možic, V., Može, M., Cimerman, F., Golobič, I., 2022. Current Status and Review of Waste-to-Biogas Conversion for Selected European Countries and Worldwide. Sustainability 14, 1823. https://doi.org/10.3390/su14031823

Section 2 Model and Simulation

Summary of Section 2

This section describes the evolution of AD models, from models describing a few compounds and microorganisms, inhibition by VFA, free ammonium, or high H_2 pressures, to the establishment of more structured models that served as a tool to study the AD process in more detail. Some of these models have been used to model biological methanation, considering the addition of gases such as H_2 and CO to the system. The evolution of these models increased the number of parameters and their effect on the model outputs. Finally, sensitivity analysis methods are presented to study the importance of the parameters in the process.

2.1 Anaerobic Digestion Models

Over the last decades, several studies have been focused on modeling and simulation of the AD process. The complexity of these models has evolved according to the need to represent better the various phenomena occurring in the bioprocess.

Andrews and Graef (1971) developed a model to simulate the liquid, gas, and biological phases (Figure 1.2.1). Their model was capable of following the dynamic response of five variables, volatile acids concentration, alkalinity, pH, gas flow rate, and gas composition. This model included an inhibition function for volatile acids concentration in a specific growth rate for the CH_4 microorganisms, and it considered the non-ionized volatile acids and the growth-limiting substrate. The model also considered the interactions between the liquid, gas, and biological phases.

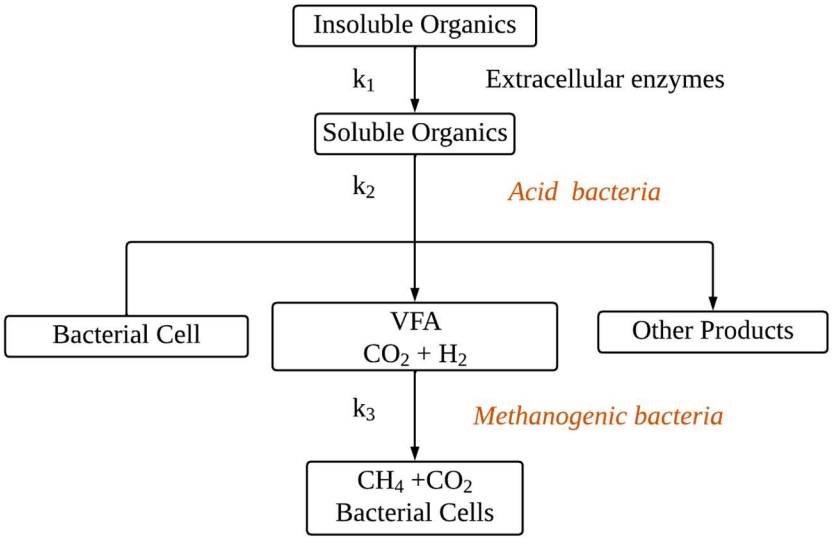


Figure 1.2.1. Anaerobic digestion model proposed by Andrews and Graef (1971).

The model simulated correctly batch and continuous conditions and these simulations permitted to infer some relevant aspects without intervening in the real process, *e.g.*, in batch operation. Thanks to the model, it was possible to regulate pH by controlling the carbon dioxide content of the recirculated gas used for mixing (Figure 1.2.2-A) in batch mode. On the other hand, it was possible to determinate the cause of some faults (organic or hydraulic overloading)

according to the type of response in the concentration of microorganisms in continuous operation (Figure 1.2.2-B).

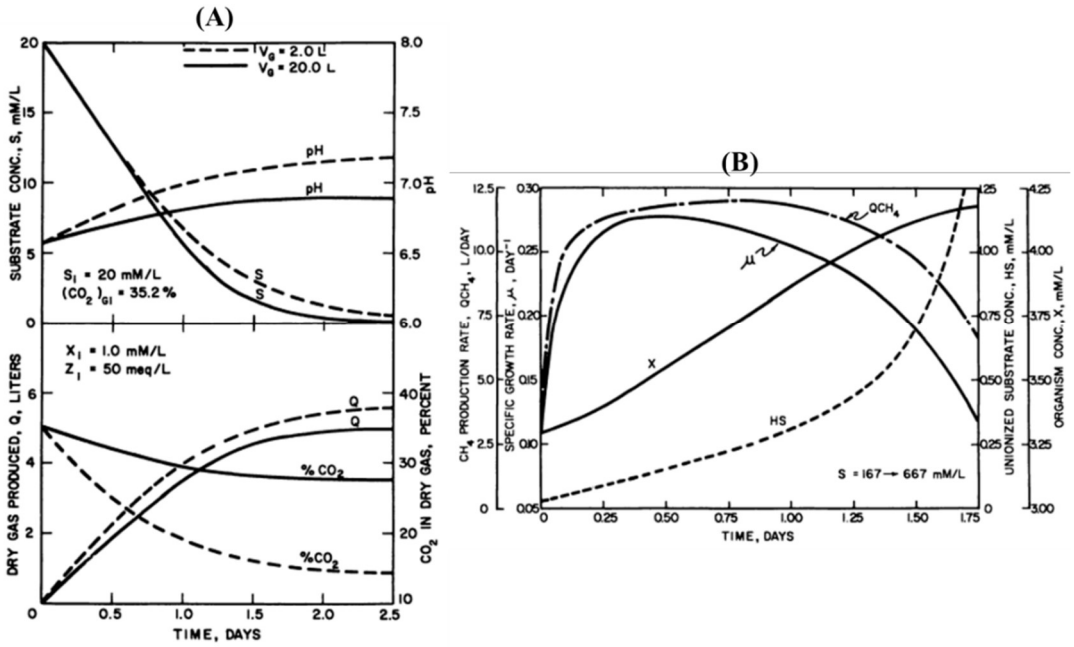


Figure 1.2.2. Simulation results using the model proposed by Andrews and Graef (1971). (A) batch operation, (B) continuous operation.

Hill and Barth (1977) proposed a model to describe the dynamic behaviors of components such as volatile matter, soluble organic, volatile acids, acids bacteria (facultative heterotrophs), CH_4 formers (obligate anaerobes), and CO_2 following the scheme in Figure 1.2.3. This model considered the inhibition by an organic acid or high ammonia concentrations. The model considered the inhibition by non-ionized acids and non-ionized ammonia on the growth kinetics of the CH_4 formers. The growth rates of kinetic acid bacteria considered only ionized acids inhibition. The pH was calculated using the solution of the electroneutrality equation solution coupled to the CO_2 mass balance.

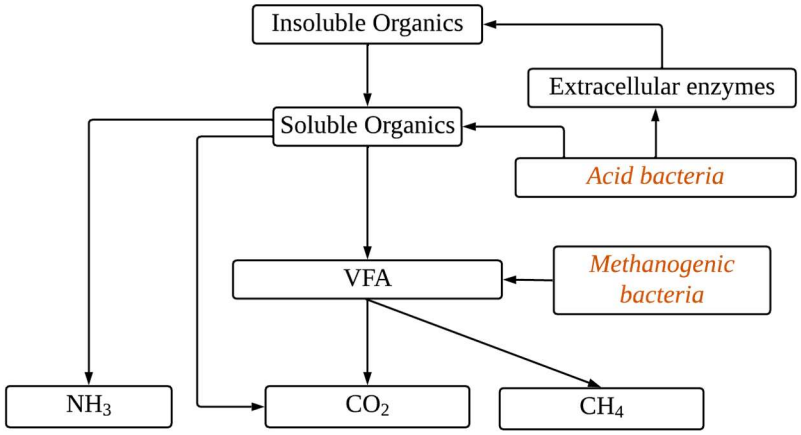


Figure 1.2.3. Anaerobic digestion model proposed by Hill and Barth (1977).

The authors compared the simulation results with experimental data on a lab scale (4.5 L) and a pilot scale (3785 L). The lab scale operated at 25°C using raw poultry waste with an OLR ranging between 400 and 3203 $gVS/m^3/d$. In the pilot-scale raw swine waste was used as a substrate. Figure 1.2.4. shows some results obtained.

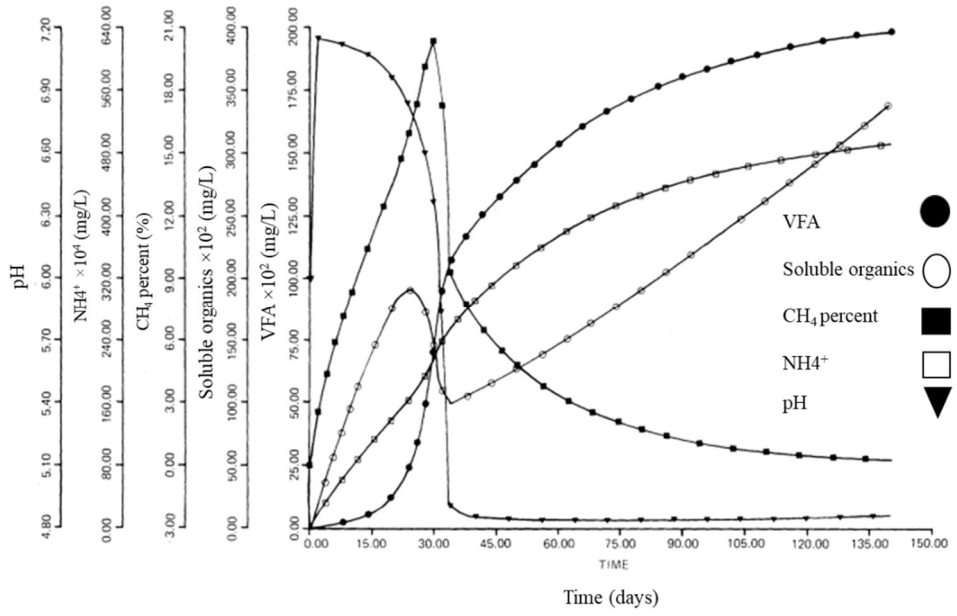


Figure 1.2.4. Simulation results using the model proposed by Hill and Barth (1977).

Mosey (1983) proposed a more sophisticated model, including the complex patterns of volatile acid production and the characterization of the H_2 utilizing microorganisms. The model is summarized in Figure 1.2.5.

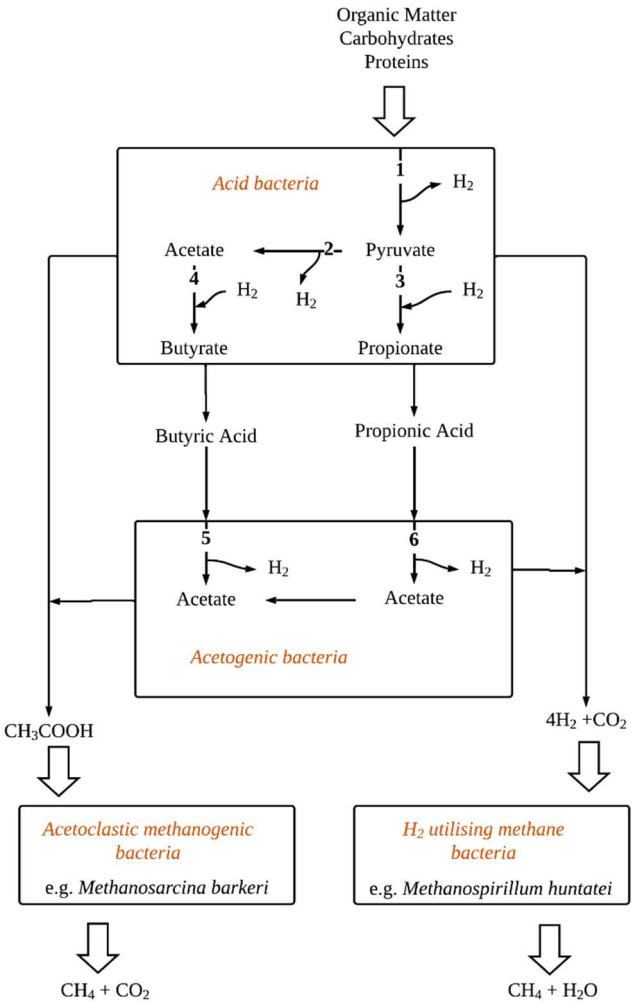


Figure 1.2.5. Anaerobic digestion model proposed by Mosey (1983).

The acid-forming bacteria transform glucose into acetic acid, propionic, and butyric acids, explaining that acetic acid is preferred. The acetogenic bacteria transform propionic and butyric acids into acetic acid. The acetoclastic bacteria convert acetic acid into a mixture of CO_2 and CH_4. The H_2-utilizing bacteria remove the H_2 from the system by generating CH_4. Some assumptions were considered for the development of this model: (i) The internal pH is maintained neutral and constant, (ii) The H_2 in the gas phase is diffused freely and rapidly

through the obligate anaerobic cells, which implies that the partial pressure inside the cells is the same as the system, (iii) The redox potential of the bacteria is the same as the potential of the growth. Model simulations were compared with steady-state data for the AD of 2000 mg/L of glucose for 20 days operating at 35°C and pH 7.0. Then, a dynamic simulation of a 1 L digester working at 35°C with an HRT of 10 days was performed using synthetic wastewater (2000 mg/L glucose for substrate and 150 mg/L ammonium hydroxide). At day 10, glucose concentration was increased until 12000 mg/mL leading the digester close to failure due to the accumulation of volatile acids in the digester (pH reduction from 6.5 to 5.5). The simulation results described some of the common problems in the AD process, such as the accumulation of VFA. The limitations of the model were the assumption of a neutral pH and the restriction on the use of microorganisms that high H_2 pressures can inhibit. Figure 1.2.6 presents some results obtained.

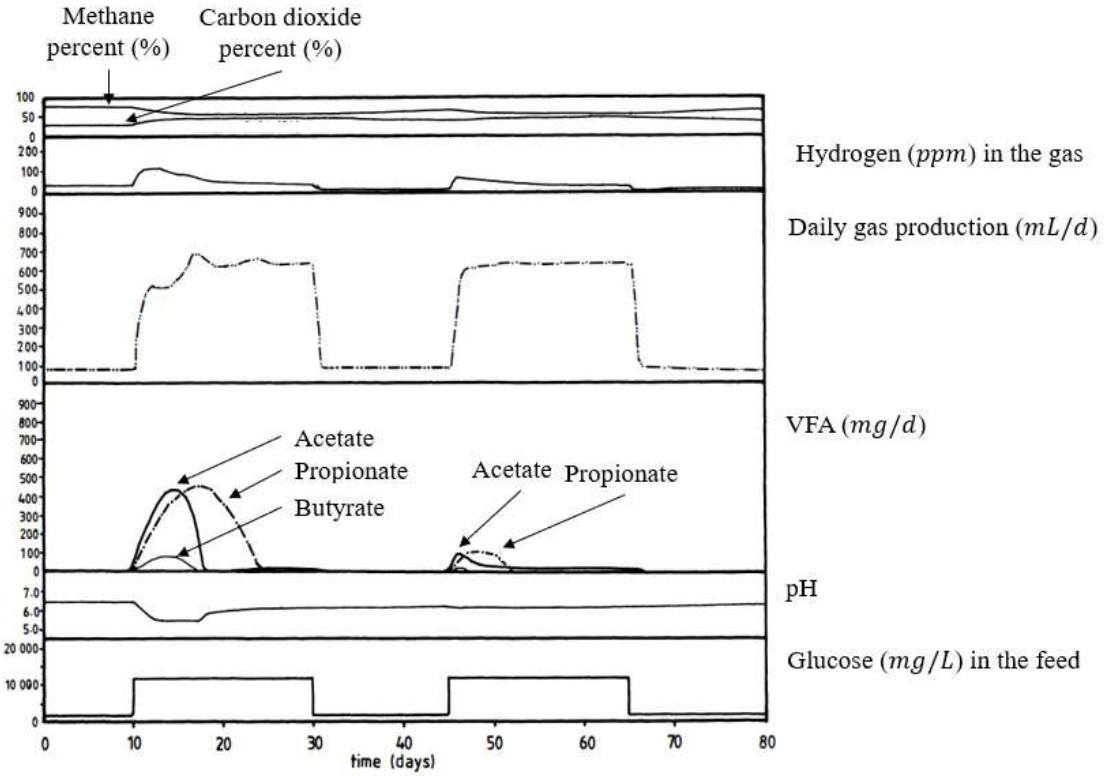


Figure 1.2.6. Simulation results using the model proposed by Mosey (1983).

Angelidaki *et al.* (1993) proposed a model that includes more species in the chemical equilibrium. In particular, the authors considered the role of ammonia, CO_2, and VFA on pH. They also implemented growth inhibition by these species for some bacteria. The model was developed according to Figure 1.2.7, where the inhibition phenomena inside the AD are represented as discontinuous lines. The following items summarize the idea of the developed model:

- The substrate composition and VFA produced to determine the pH of the process
- The pH and temperature affect the ionization degree of ammonia
- Free ammonia controls the rate of the methanogenic step
- Inhibition of the methanogenic step results in an acetate accumulation
- Acetate accumulation inhibits the acetogenic steps, generating propionate and butyrate accumulation.
- VFA accumulation inhibits the hydrolysis step
- VFA accumulation lowers pH, which in turn causes a reduction in the free ammonia inhibition

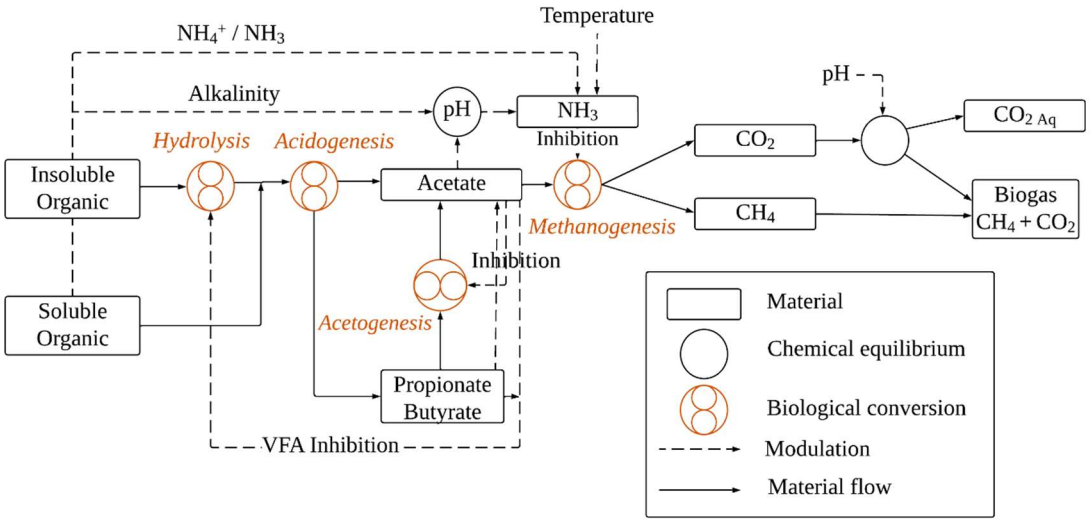


Figure 1.2.7. Anaerobic digestion model proposed by Angelidaki *et al.* (1993).

The model was validated with experimental data. The experiment was performed in a $CSTR$ with cattle manure as a substrate, operating at 55°C and atmospheric pressure. The HRT was

15 days. The ammonia levels at steady state were 2.5 g/L. The model successfully reproduced the experiments, with a minor CH_4 yield and content deviation. Once validated, the model was used to simulate temperature disturbances changing from 55°C to 50°C. The results showed an increase in AD performance at 50°C. This improved in the performance was based on the decrease and stabilization of acetate and propionate, a decrease in butyrate concentrations, and an increase in the CH_4 yield and content. Figure 1.2.8 presents some results obtained.

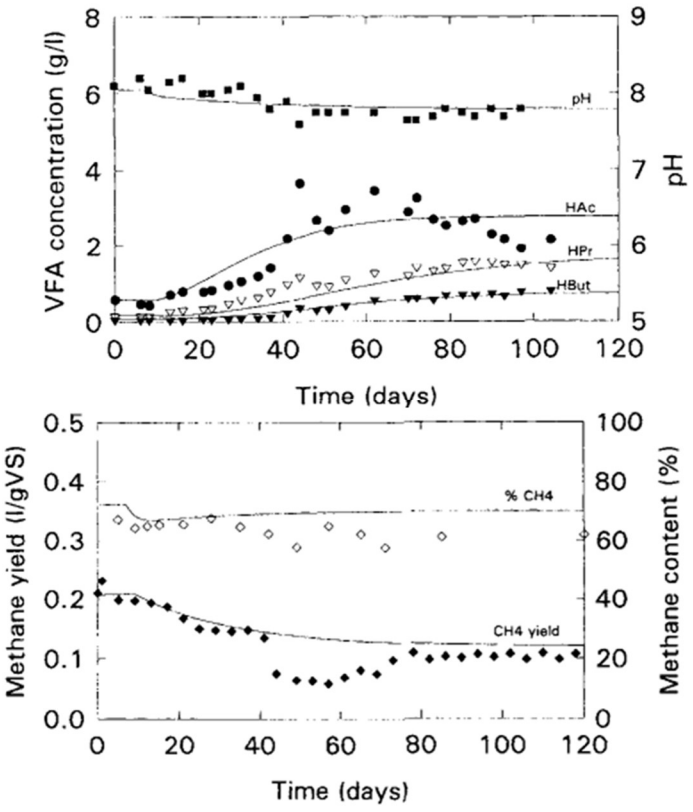


Figure 1.2.8. Simulation results using the model proposed by Angelidaki *et al.* (1993).

Moletta (1986) developed a model considering a two-step process, transforming glucose into acetate by acetogenic bacteria and acetate into CH_4 and CO_2 by methanogenic bacteria, separating growth and metabolite production expressions. Kleinstreuer and Poweigha (1982) used as a base the model proposed by Andrews and Graef (1971) to simulate the production of CH_4 from biomass for two cultures. Mata-Alvarez (1987) presented a model to simulate a two-phase system for the digestion of wastes with high solid content.

However, in 1997, the International Water Association (IWA) Anaerobic Digestion Model Task Group focused on proposing a generalized model for AD, the "Anaerobic Digestion Model No 1 (ADM1)". This model is probably the most used, extended, and modified model to explore the different applications of AD**.** ADM1 was based on experience acquired over the previous years in modeling and simulating the AD process (Esposito *et al.*, 2011). The ADM1 proposed by Batstone *et al.* (2002) was a structured model with multiple steps describing biochemical and physicochemical processes (Figure 1.2.9).

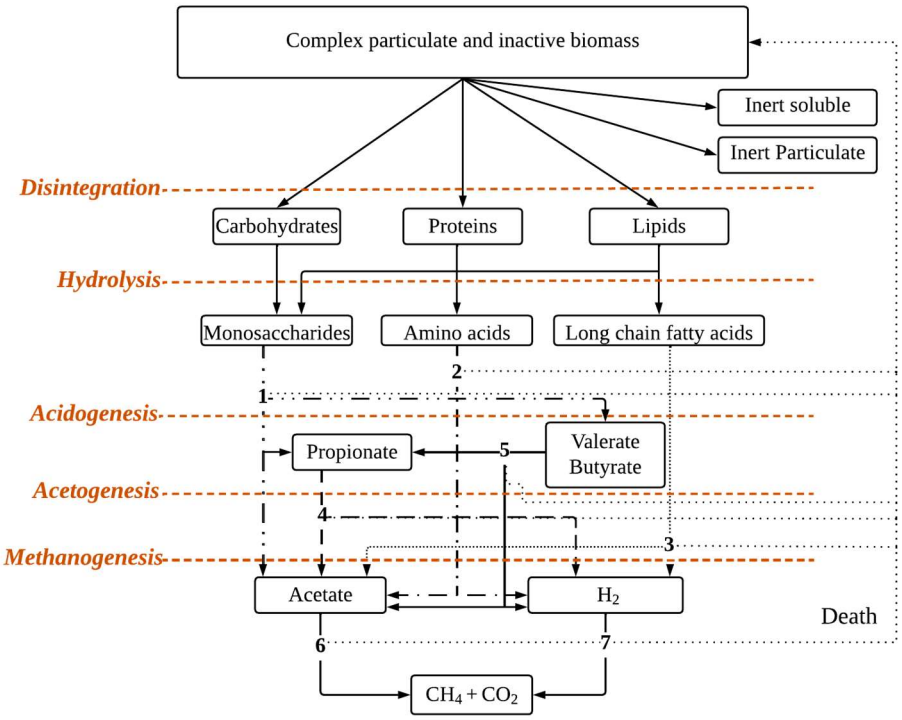


Figure 1.2.9. Anaerobic digestion model proposed by Batstone *et al.* (2002). (1) acidogenesis from sugars, (2) acidogenesis from amino acids, (3) acetogenesis from LCFA, (4) acetogenesis from propionate, (5) acetogenesis from butyrate and valerate, (6) acetoclastic methanogenesis, and (7) hydrogenotrophic methanogenesis.

The biochemical processes include:

- The degradation of organic matter into carbohydrates, proteins, and lipids;
- The hydrolysis of these substrates to monosaccharides (sugar), amino acids, and long-chain fatty acids (LCFA);

- The acidogenesis from sugars and amino acids to VFA and H_2;
- The acetogenesis of LCFA and VFA to acetate
- The separation of methanogenesis steps from acetate and H_2:CO_2

The physicochemical processes refer to: (i) the ionic association/dissociation of components such as propionate, butyrate, valerate, and bicarbonate, (ii) the gas-liquid mass transfer of components generated during digestion as H_2, CO_2, and CH_4.

The ADM1 reports concentrations of most of the components in Chemical Oxygen Demand (COD) per volume unit. COD is the amount of oxygen needed to degrade the organic matter into CO_2 and H_2O (see Annex 1.1). As CO_2 could not be expressed in COD, its concentration is presented as mol per volume.

Batstone *et al.* (2002) presented Equations (1.2.1)-(1.2.3) to summarize the model.

$$\frac{dS_{liq,i}}{dt} = \frac{q_{in}S_{in,i}}{V_{liq}} - \frac{S_{liq,i}q_{out}}{V_{liq}} + \sum_{j=1-19} \rho_j v_{i,j} \qquad (1.2.1)$$

$$\rho_{10,T} = k_L a_{CO2} \left(S_{liq,CO2} - K_{H,CO2} p_{CO2,gas}\right) \qquad (1.2.2)$$

$$\rho_{\mathrm{A/B\ CO2}} = -\rho_{\mathrm{A/B\ HCO_3^-}} = \mathrm{k_{A/B\ HCO3}} \left(\mathrm{S_{H.C.O_3^-}} \cdot \mathrm{S_{H^+}} - \mathrm{K_{a,CO2}} \cdot \mathrm{S_{CO2}}\right) \qquad (1.2.3)$$

Equation (1.2.1) refers to the mass balances of each component in the liquid phase. The term $\sum_{j=1-19} \rho_j v_{i,j}$ is the sum of the kinetic rates for process j multiplied by $v_{i,j}$. Equation (1.2.2) refers to transferring gas components to the gas headspace (the example presented was CO_2) where $k_L a_{CO}$ is the volumetric mass transfer coefficient, $K_{H,CO}$ is Henry's law equilibrium constant and $p_{CO\ ,gas}$ is the CO_2 gas-phase partial pressure. Equation (1.2.3) is the dynamic rate equation used for acid-base reactions.

The model solution of the differential equations can generate stiffer problems and introduce errors in the ADM1. Therefore, Rosen and Jeppsson (2006) proposed the use of Hill functions to model the process to avoid stiff problems.

The ADM1 has been widely accepted and validated by multiple authors. It has been used to simulate the dynamic behavior of a pilot-scale for anaerobic two-stage digestion of sewage sludge (Blumensaat and Keller, 2005), to simulate a full-scale anaerobic sludge digester (Ersahin, 2018; Ozgun, 2019), or even to simulate the AD process of source-sorted organic fractions of municipal solid wastes (Calise *et al.*, 2020). The ADM1 has also been used as a base of multiple models that intend to reduce their complexity (Hassam *et al.*, 2015; Arzate *et al.*, 2017; Weinrich and Nelles, 2021). Other authors have used extensions of the ADM1 to consider the co-digestion of mixed substrates such as sewage sludge/municipal solid waste (Esposito *et al.*, 2011). Additionally, it has been used to predict interactions between phosphorus, sulfur, and iron in plant-wide simulation (Flores-Alsina *et al.*, 2016) and to study the inhibition phenomena by ammonium concentrations (Bai *et al.*, 2017; Li *et al.*, 2019).

Other models used the ADM1 for plant layout, for instance, the Benchmark Simulation Model number 1 (BSM1) (Henze *et al.*, 1987) and number 2 (BSM2) (Alex *et al.*, 2018). The BSM1 plant was designed considering five compartments of activated sludge reactor, two anoxic tanks, and three aerobic tanks, combining nitrification with pre-denitrification in a configuration typical for achieving biological nitrogen removal in full-scale plants. The BSM2 included BSM1 for the biological treatment of wastewater sludge and the implementation of the ADM1 model for anaerobic digestion (Rosen and Jeppsson, 2006).

2.2 Biological Methanation Models

The ADM1 has been successfully accepted as a general dynamic model to represent AD. However, its application in biological methanation is limited, especially in the gas-liquid mass transfer aspects and biochemical processes such as the transformation of CO (Sun *et al.*, 2021). Only a few researchers have worked on generating advances in modeling biological methanation, whose works will be presented in the following paragraphs.

Grimalt-Alemany *et al.* (2020) proposed a model to simulate the biological methanation using a mixed microbial consortium for the first time. The biological methanation of syngas was carried out at mesophilic and thermophilic conditions intended to study possible control strategies through the modulation of key operating parameters. The model used the structure of

the ADM1 for the concentration of components in liquid and gas phases, the growth of different biomass groups, and other physicochemical processes such as acid dissociation and gas-liquid mass transfer. The novelty of this model consisted in considering possible syntrophic pathways to the hydrogenotrophic methanogens and homoacetogens without assuming a strictly kinetically driven competition between them. They considered a thermodynamic consistency with a potential factor.

Their experiments were developed in 100 mL flasks (batch reactors) operating at 37°C and 60°C with an average pH of 7.2 and agitation of 100 rpm. The syngas was added, ranging the CO partial pressure from 0.2 to 0.8 atm, and fixed the partial pressure of H_2 and CO_2 at 1.0 and 0.2 atm, respectively.

The thermodynamic potential factor (F_T), factor was considered in the modeling of the biomass growth $\mu = f(F_T)$ as follows,

$$F_T = \begin{cases} 1 - exp\left(-\dfrac{\Delta G_A - \Delta G_C}{\chi^{RT}}\right), \Delta G_A \geq \Delta G_C \\ 0, \Delta G_A \leq \Delta G_C \end{cases} \quad (1.2.4)$$

where ΔG_A refers to the negative Gibbs free energy change of each biochemical reaction, $\Delta G_C = Y_{ATP}\Delta G_p$ is the free energy conserved through each metabolic pathway calculated by multiplying the ATP yield with the Gibbs free energy of phosphorylation (ΔG_p); and χ^{RT} is a parameter to weigh the contribution of ΔG_A to the reaction and ΔG_C to the overall F_T. If $\Delta G_A \approx \Delta G_C$, then $F_T = 0$, which indicates that the thermodynamic drive for the reaction to proceed forward disappears, and the metabolism stops. The results correctly simulate the process, the specific growth of the various microorganisms, and CH_4 productivity over time, even when the partial pressure of the CO varied in the system.

Sun *et al.* (2021) modified and extended the ADM1 to consider syngas addition (CO + H_2) from lab and pilot scale experiments. The authors considered that CO could be uptaken in two steps. In the first step, the CO is transformed into H_2 and CO_2 by the carboxydotrophic hydrogenogens. In the second step, the CO is converted into acetate and CO_2 by

carboxydotrophic acetogens (Figure 1.1.6 in Section 1). Then, the methanogens metabolize H_2 and acetate to produce CH_4.

The authors proposed to use the volumetric mass transfer coefficient ($k_L a$) based on the two-film theory, expressed as Equation (1.2.5).

$$k_L a = \frac{1/\left(\frac{1}{k_y} + \frac{m_j}{k_x}\right)}{P} \qquad (1.2.5)$$

where $k_y = D_g/RT\delta_g\,(P/Pbm)$ and $k_x = D_l/\delta_l\,(Cm/Cbm)$ are the gas and liquid mass transfer coefficient, respectively, with $1/k_y$ and$1/k_x$ being the gas and liquid mass transfer resistance. P refers to the gas pressure and $m_j = E_j/P$ is the solubility coefficient. E_j is the Henry constant, R is the Universal gas constant, T represents the temperature of the system, δ_g and δ_l are the theoretical gas and liquid film thickness. D_g and D_l are the diffusion coefficients of the gas and liquid phase, respectively. The terms (Cm/Cbm) and (P/Pbm) are the drift and diffusion factors of the gas and liquid phases, with Cbm and Pbm as the logarithmic average of the concentrations and pressure on both sides of the stationary fluid and gas layer.

The authors validated their model experimentally. The experiment at lab scale was developed in a bubble column reactor (BCR) with a working volume of 37.5 L operating at 37°C for 207 days. The OLR was 0.5 $g/L/d$ of glucose (HRT= 20 days). The syngas addition (H_2/CO = 0.5/0.5) was performed at five stages ranging from 0.2 to 1.0 $L/L/d$. The results were used to calibrate the model, with R^2 of 0.97, 0.86, and 0.87 for CH_4, H_2, and CO outlet gas flow rates, respectively. The experiment in pilot scale was performed in a working volume of 6 m^3 operating at 35°C with an OLR of 1 $g/L/d$ of glucose (HRT= 20 days). The syngas flow rate was 1.10 m^3/d (H_2/CO/CH_4/CO_2: impurity 0.15/0.22/0.17/0.35/0.11). Model validations gave a R^2 between 0.83-0.89, 0.78-0.84, and 0.61-0.73 for CH_4, H_2, and CO, respectively. The inconvenient of this model is that it does not consider the modeling of CO_2, an important compound found in syngas mixtures which is also uptaken since the carbon from CO_2 is used for methane while the O_2 is combined with H_2 to produce water. Additionally, the modified ADM1 used a $k_L a$ based on the film theory, which generates a significant increase in the parameters of the model where some of them must be estimated experimentally.

Tsapekos *et al.* (2022) developed an unstructured kinetic model to study the influence of the variables such as partial pressure of H_2, CO_2, and pH in the hydrogenotrophic, homoacetogenesis, and acetoclastic pathways with an inoculum adapted and non-adapted to H_2/CO_2 as substrates. The main assumption of the model is that an increase in the overall pressure could alter conversion efficiency and favors homoacetogens over methanogens. The relevant aspect of the model lies in the thermodynamic aspect to study the competition between homoacetogens and methanogens for H_2 conversion. The proposed model considers hydrogenotrophic methanogens, acetogenic bacteria, and acetoclastic methanogens; the concentration of CO_2 and acetate in the liquid phase, and H_2, CH_4, and CO_2 in the gas phase. The model was used to simulate four cases varying the pressure and pH of the system. Case 1: 1.0 atm and pH of 8.39, Case 2: 1.0 atm and pH of 7.0, Case 3: 0.2 atm and pH of 8.39, and Case 4: 0.2 atm and pH of 7.00. These results were compared with the experimental data. The best coefficients of determination $R^2 > 0.94$ were obtained in case 2 for CO_2, H_2, and CH_4 in the gas phase and the acetate concentration in the liquid phase.

Santus *et al.* (2022) proposed an *ex-situ* biological methanation model based on a simplified and modified version of the ADM1, just considering hydrogenotrophic methanogenesis. The model is made of 8 differential equations, describing particular mass balances of H_2 and CH_4 in the liquid phase, inorganic carbon and nitrogen, H_2, CH_4, and CO_2 in the gas phase. It is completed by the definition of growth law for hydrogenotrophic methanogens. The equations for the components in the gas phase were modified to consider the addition of H_2 and CO_2 to the system. The model was used to simulate two data sets. The first was in an up-flow reactor for ex-situ biological methanation operating at 55°C, and the other performed in a 380 L $CSTR$ with a 460 L of headspace with H_2 and CO_2 as a substrate maintaining a relation 4:1. The model predicted the CH_4 outlet flow rate of two different configurations successfully.

2.3 Sensitivity Analysis and Confidence Intervals

The previous research presented models to describe AD and biological methanation. These models are constituted by several parameters, *e.g.*, stoichiometric, biochemical, and physiochemical, and each one brings a degree of uncertainty to the model. Consequently, it is necessary to consider a Sensitivity Analysis (SA), which intends to determine how the

uncertainty of the parameters influences the outputs (Saltelli *et al.*, 2004; Damblin *et al.*, 2013; Sepulveda *et al.*, 2013; Sohier *et al.*, 2014; Tosin *et al.*, 2020).

SA can be classified into two main types: local sensitivity analysis (LSA) and global sensitivity analysis (GSA) (Morio, 2011; Ochoa *et al.*, 2016a). LSA analyses minor disturbances of the model inputs or parameters near the nominal value one by one. This means one factor at a time (OAT). In contrast, the other parameters are fixed at the corresponding nominal value (Zi, 2011). However, this technique does not study all the parameter space over output variables and the interaction between these parameters (Saltelli *et al.*, 2017). GSA was developed to identify the contribution of each uncertainty input (or parameter) to the outputs (Feng *et al.*, 2019; Kucherenko *et al.*, 2015; Zhang *et al.*, 2015). The sensitivity addresses the exploration of the entire range of variation of the model parameters (Kiparissides *et al.*, 2008), using a probability density function associated with each input parameter and repeated simulations of the model (Iooss and Lemaître, 2015; Ochoa *et al.*, 2016a; Tosin *et al.*, 2020).

2.3.1 Global Sensitivity Analysis Methods

GSA includes three groups: regression methods, screening methods, and variance-based methods (Sepúlveda *et al.*, 2014; Iooss and Lemaître, 2015). One of the most commonly used GSA in bioprocess are the Morris and Sobol' Methods (Ashraf and Abu-Reesh, 2022; Kiparissides *et al.*, 2009; Ochoa *et al.*, 2016b, 2016a; Rapadamnaba *et al.*, 2021; Ruano *et al.*, 2011; Tosin *et al.*, 2020).

2.3.1.1 Sobol' Method

The Sobol method (Sobol′, 2001) is an interesting variance-based method in which the variance of the model output can be decomposed into partial variances that represent the contribution of the inputs over the overall uncertainty of the model output (Morio, 2011; Ochoa *et al.*, 2016a; Sepulveda *et al.*, 2013; Sobol′, 2001; Tosin *et al.*, 2020).

Consider a model defined by ξ as,

$$Y = \xi(\boldsymbol{\theta}) \qquad (1.2.6)$$

where $Y \in \mathcal{R}^m$ is the model output of interest, and $\boldsymbol{\theta} \in \mathcal{R}^n$ is an n-dimensional parameter vector defined as $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)$ and characterized by a probability density function (PDF).

The function $\xi(\boldsymbol{\theta})$ can be decomposed into summands of different dimensions, Equation (1.2.7).

$$Y = \xi_0 + \sum_{i=1}^{n} \xi_i(\theta_i) + \sum_{1 \le i < j \le n} \xi_{ij}(\theta_i, \theta_j) + \cdots + \xi_{1\dots n}(\theta_i, \dots, \theta_n) = \sum_{\mathbf{u} \subseteq \{1 \dots n\}} \xi_{\mathbf{u}}(\theta_{\mathbf{u}}) \qquad (1.2.7)$$

where:

$$\begin{cases} \xi_0 = \mathbb{E}[Y] = \int \xi(\boldsymbol{\theta}) f_{\theta_i}(\theta_i) d\boldsymbol{\theta} \\ \xi_i(\theta_i) = \mathbb{E}[Y|\theta_i] - \xi_0 = \int \xi(\boldsymbol{\theta}) \prod_{i=1}^{n} \left[f_{\theta_i}(\theta_i) d\theta_i \right] - \xi_0 \\ \xi_{ij}(\theta_i, \theta_j) = \mathbb{E}[Y|\theta_i, \theta_j] - \xi_0 - \xi_i(\theta_i) - \xi_j(\theta_j) = \int \xi(\boldsymbol{\theta}) \prod_{i=1}^{n} \left[f_{\theta_i, \theta_j}(\theta_i, \theta_j) d\theta_i d\theta_j \right] - \xi_0 - \xi_i(\theta_i) - \xi_j(\theta_j) \end{cases}$$

ξ_0 is the mean of the function, $\xi_i(\theta_i)$ and $\xi_{ij}(\theta_i, \theta_j)$ are the expectation terms of increasing order and the conditional expectations defined recursively, $f_{\theta_i}(\theta_i)$ and $f_{\theta_i, \theta_j}(\theta_i, \theta_j)$ are the marginal PDF of θ_i and the interaction θ_i, θ_j $(i = 1,2, \dots, n)$.. This decomposition is unique, provided that the inputs are independent and the individual terms are square-integrable (Efron and Stein 1981).

The so-called ANOVA decomposition could be obtained from Equation (1.2.7) as follows (Sobol′, 2001),

$$V(Y) = \sum_{i=1}^{n} V_i(\theta_i) + \sum_{1 \le i < j \le n} V_{ij}(\theta_i, \theta_j) + \cdots + V_{1\dots n}(\theta_i, \dots, \theta_n) = \sum_{\mathbf{u}} V(\xi_{\mathbf{u}}(\theta_{\mathbf{u}})) \quad \text{for } \mathbf{u} \subset \{1, \dots n\} \qquad (1.2.8)$$

where $V(\xi_{\mathbf{u}}(\theta_{\mathbf{u}}))$ express the conditional variance for the subvector $\theta_{\mathbf{u}}$, containing the variables whose indices are indicated by the subset $\mathbf{u}$.

The variance of the output can be decomposed into terms depending on the parameters and their interactions. In order to normalize the variances, it is possible to define a variance-based sensitivity index (SI) associated with the subset $\mathbf{u}$, which is the ratio between the contribution given by the interaction among the components of $\mathbf{u}$ for the model variance and the total variance, Equation (1.2.9).

$$SI_{\mathbf{u}} = \frac{V\left(\xi_{\mathbf{u}}(\theta_{\mathbf{u}})\right)}{V(Y)} \qquad (1.2.9)$$

Based on this, for $\mathbf{u} \subset \{1, \dots n\}$, and $\mathbf{u} \neq \mathbf{0}$,

$$\sum_{\mathbf{u}} SI_{\mathbf{u}} = \sum_{i=1}^{n} SI_i + \sum_{1 \le i < j \le n} SI_{ij} + \cdots + SI_{1 \dots n} = 1 \qquad (1.2.10)$$

The term SI_i is the first-order sensitivity index, which measures the fraction of the total output variance explained by the parameter θ_i alone as,

$$SI_i = \frac{V\left(\xi_i(\theta_i)\right)}{V(Y)} \qquad i = 1, \dots, n \qquad (1.2.11)$$

Similarly, SI_{ij} is the second order-sensitivity index that measures the amount of variance caused by the interaction between the parameters θ_i and θ_j as,

$$SI_{ij} = \frac{V\left(\xi_{ij}(\theta_{ij})\right)}{V(Y)} \quad 1 \le i < j \le n \qquad (1.2.12)$$

It is possible to construct the *SI* for all orders until the n^{th} order index $SI_{1 \dots n}$, which represents the contribution of the interactions between all the parameters in $\boldsymbol{\theta}$. The total Sobol' indices are used to measure the full contribution of the i^{th} random parameter θ_i for the total variance either by its single effect or by its interaction with others:

$$SIT_i = \sum_{\substack{\mathbf{u} \subset \{1,\dots,n\} \\ i \in \mathbf{u}}} SI_{\mathbf{u}} \qquad i = 1, \dots, n \tag{1.2.13}$$

Equation (1.2.14) indicates that the total sensitivity index does not only include the marginal contribution of θ_i to the variance of the output, but it also contains its cooperative contribution with all the other inputs.

2.3.1.2 Morris Method

Morris method (Morris, 1991) is the most frequently used screening method to perform SA by analyzing one-factor-at-a-time (OAT). This is generally used when the number of model parameters is higher and the computation of the model simulations is expensive (Feng *et al.*, 2019; Sepúlveda *et al.*, 2014). This method provides qualitative sensitivity measures, ranking the factors according to their importance. Nevertheless, it does not quantify the importance of one factor concerning another (Saltelli *et al.*, 2004). The Morris method applied to parameter sensitivity discretizes the space of each parameter and performs a given number of OAT designs. These designs and variation directions are randomly chosen from the parameter space. The repetition of these steps allows the estimation of elementary effects $\left(EE_i^j\right)$ for each parameter i, which represents the relative difference between the outputs and the j^{th} parameter disturbance (Feng *et al.*, 2019; Iooss and Lemaître, 2015; Morio, 2011; Morris, 1991; Saltelli *et al.*, 2007).

Consider a trajectory in the parameter space as,

$$\theta_i^{j+1} = \theta_i^j + e^j \Delta^j \qquad \mathrm{j} = 1, \dots, \mathrm{r} \tag{1.2.15}$$

where $\mathrm{j} = 1, \dots, \mathrm{r}$ corresponds to the number of repetititons and $\boldsymbol{\theta} \in \mathcal{R}^n$ is an n-dimensional parameter vector defined as $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)$.

The effect of parameter variation can be evaluated by estimating the difference between the model output with the actual parameter θ_i^j and the updated parameter $\theta_i^j + e^j\Delta^j$ over a given increment Δ^j. e^j is a vector of zeros but with a unit as its j^{th} component (canonical base). This variation is referred as elementary effects, which can be calculated as follows,

$$EE_i^j = \frac{\xi\left(\theta_i^j + e^j\Delta^j\right) - \xi\left(\theta_i^j\right)}{\Delta^j} \quad (1.2.16)$$

where θ_i^j is a sample of input $\boldsymbol{\theta}$ and $\xi\left(\theta_i^j\right)$ is the corresponding model output. Δ^j is a step between two consecutive input space points of the trajectory. The term $\left(\theta_i^j + e^j\Delta^j\right)$ represents a new sample by moving the i^{th} parameter input from θ_i^j to $\theta_i^j + \Delta^j$, with the respective model output $\xi\left(\theta_i^j + e^j\Delta^j\right)$.

The index j of EE_i^j expresses the ratio of the change of the output Y when the i^{th} parameter θ_i^j is given a particular change Δ^j. Then, EE_i^j can measure the effect of θ_i^j in a given scope of output Y. The sensitivity measures are expressed in terms of means μ_i^*, and standard deviations σ_i^j are defined as Equations (1.2.17) and (1.2.18).

$$\mu_i^* = \frac{\sum_{j=1}^{r}\left|EE_i^j\right|}{r} \quad (1.2.19)$$

$$\sigma_i^j = \sqrt{\frac{\sum_{j=1}^{r}\left(EE_i^j - \mu_i^*\right)^2}{\mathrm{r} - 1}} \quad (1.2.20)$$

where EE_i^j is the elementary effect of the i^{th} parameter obtained at the j^{th} repetition. The sensitivity measures μ_i^j and σ_i^j are the mean of the absolute value and standard deviation of the distribution of the elementary effects, respectively. μ_i^* measures the influence of the i^{th} parameter on the output. σ_i^j is a measure of non-linear and interaction effects of the i^{th} parameter. A high value of μ_i^* indicates that the parameter θ_i^j has a more important effect on the output. A high value of σ_i^j indicates that the elementary effect of θ_i^j varies significantly

from one to another, which shows that the value of EE_i^j is strongly influenced by the selected sample points.

2.3.2 Parameter Estimation and Confidence Regions

Parameter estimation is one of the most relevant aspects in the formulation and calibration of models. The calibration must guarantee that the accuracy obtained by the parametric estimation is maintained, even when there is a slight variation in the parameters.

Parameter estimation is frequently calculated by minimizing a quadratic cost function that compares the experimental data set with the results obtained by the model. Then, the confidence intervals for these estimated parameters are computed via scalar functions of the Fisher information matrix (FIM) (Rodríguez-Fernández *et al.*, 2005). However, the determination of the FIM depends on the parameters' values and the responses' behavior (calculated with the use of the partial derivatives $\partial Y/\partial\theta$), which makes difficult to perform the parameter estimation especially when there are non-linear interactions between parameters (Rodriguez-Fernandez *et al.*, 2007).

2.3.2.1 Fisher Information Matrix

Consider again the model previously defined in Section 1.3.1.1 represented by ξ, Equation (1.2.21).

$$Y = \xi(\boldsymbol{\theta}) \quad (1.2.18) \qquad (\ 1.2.21\ )$$

where $Y \in \mathcal{R}^m$ is the vector of model outputs, and $\boldsymbol{\theta} \in \mathcal{R}^n$ is an n-dimensional parameter vector defined as $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)$. The parameter estimation is often conducted to minimize a quadratic function as, Equation (1.2.22) (Dochain and Vanrolleghem, 2015).

$$J(\boldsymbol{\theta}) = \sum_{i=1}^{N} \left(Y(\theta) - \hat{Y}\right)^T \cdot W_i \cdot \left(Y(\theta) - \hat{Y}\right) \qquad (\ 1.2.22\ )$$

where $\hat{Y}$ and $Y(\theta)$ are the vectors of experimental measurements and model predictions at the time t_i (i=1 to N) respectively. W_i is a square matrix with weighting coefficients. The expected value of the objective function for a parameter set slightly different from the optimal one is defined as:

$$\mathbb{E}[J(\theta+\Delta\theta)] \cong \Delta\theta^T \left[\sum_{i=1}^{N}\left(\frac{\partial Y}{\partial\theta}(t_i)\right)^T W_i \left(\frac{\partial Y}{\partial\theta}(t_i)\right)\right]\Delta\theta + \sum_{i=1}^{N} t_r(C_i W_i) \qquad (1.2.23)$$

C_i represents the measurement error covariance matrix (W_i is typically chosen as C_i and the second term reduces to a scalar). $\partial Y/\partial\theta$ are the partial derivatives of each output concerning each parameter, *i.e.*, the output sensitivity functions, which quantify the dependence of the model predictions on the parameter values. The term between brackets in Equation (1.2.23) is the so-called Fisher Information Matrix (FIM):

$$FIM = \sum_{i=1}^{N}\left(\frac{\partial Y}{\partial\theta}(t_i)\right)^T W_i \left(\frac{\partial Y}{\partial\theta}(t_i)\right) \qquad (1.2.24)$$

FIM equation expresses the information content of the experimental data (Rodriguez-Fernandez *et al.*, 2013) by ensuring that the fit of a parameter set slightly different from the best parameter set is significantly worse. This matrix is the inverse of the parameter estimation error covariance matrix (Cov) of the best linear unbiased estimator, Equation (1.2.25).

$$Cov = FIM^{-1} = \left(\sum_{i=1}^{N}\left(\frac{\partial Y}{\partial\theta}(t_i)\right)^T W_i \left(\frac{\partial Y}{\partial\theta}(t_i)\right)\right)^{-1} \qquad (1.2.25)$$

The formulation of the FIM leads to the delimitation of confidence regions around the best parameter estimates for different confidence levels. Once the Cov is calculated, approximate standard errors (σ) for the parameters can be formulated by evaluating the residual mean square (s^2), Equations (1.2.26)-(1.2.27).

$$\sigma(\theta_i) = s\sqrt{Cov_{ii}} \qquad (1.2.26)$$

with

$$s^2 = \frac{J(\theta)}{N-p} \quad (1.2.27)$$

The confidence intervals (CI) for each estimated parameter can be calculated as, Equation (1.2.28).

$$CI = \theta \pm t_{N-p}^{\alpha/2}\, \sigma(\theta_i) \quad (1.2.28)$$

with a confidence level specified as $100(1-\alpha)$ % and t-values obtained from the Student-t distribution.

2.3.2.2 Global Sensitivity Information Matrix

Rodriguez-Fernandez *et al.* (2007) proposed a novel methodology for optimal experimental design based on Sobol' global sensitivity indices to increase the parameter's precision. The idea was to use the information proportioned by the GSA and construct the confidence intervals for the estimated parameters. Similar to the FIM, the authors proposed the use of a matrix called the Global Sensitivity Information Matrix ($GSIM$) based on the first-order SI_i.

The $GSIM$ is calculated as, Equation (1.2.29).

$$GSIM = \sum_{t=1}^{N} [Q^T(t_t) \cdot W_t^{-1} \cdot Q(t_t)] \quad (1.2.29)$$

where ${W_t}^{-1}$ is a weighting matrix usually chosen as the measurement error covariance matrix, and $Q(t_t)$ is defined as, Equation (1.2.30).

$$Q(t_t) = \begin{bmatrix} SI_1^1(t_t) & SI_2^1(t_t) & \dots & SI_n^1(t_t) \\ SI_1^2(t_t) & SI_2^2(t_t) & \dots & SI_n^2(t_t) \\ \vdots & \vdots & \ddots & \vdots \\ SI_1^m(t_t) & SI_2^m(t_t) & \dots & SI_n^m(t_t) \end{bmatrix} \quad (1.2.30)$$

In this case, $SI_n^m(t_t)$ measures the sensitivity of the state Y_m concerning the parameter θ_n at the time t_t. Then the variance of each parameter θ_i can be approximated by $\sigma^2(\theta_i) \approx GSIM_{ii}^{-1}$ and used to evaluate the confidence intervals as presented in Equations (1.2.26)-(1.2.27).

2.4 Sensitivity Analysis Applications in AD and Biological Methanation

There are at least fifteen investigations in the literature related to the LSA and twelve related to GSA applied over models that represent the AD (Barahmand and Samarakoon, 2022). Most of these researchers considered the ADM1 model complex due to the many parameters and variables involved. They have opted to work with simplified versions to study some relevant parameters in an easier manner.

Donoso-Bravo *et al.* (2013) used a GSA to determine the most sensitive parameters of a model that describes AD. The model was made of five mass balances over acidogenic and methanogenic microorganisms, acetate, CH_4, and inorganic carbon. The Sobol' GSA technique was implemented to analyze the effect of 12 parameters over four outputs, biogas flow, pH, glucose, and acetate. Despite being a simplified version of the AD process, the GSA found four key parameters that affect the outputs of the process, the maximum specific growth rates of both biomass and the stoichiometric coefficients for the substrates glucose and acetate. The authors complemented the GSA by estimating these parameters using a Classical least-squares estimator as a cost function and the covariance that can be obtained through the evaluation of the FIM to determine the confidence intervals. Schroyen *et al.* (2018) performed a GSA over a reduced version of the ADM1 and assessed the biogas production of 7 substrates with different lignin content. The SA was performed using Monte Carlo simulations and assuming a uniform distribution. This model described the process in four steps. The insoluble organic matter is hydrolyzed to volatile dissolved solids through first-order kinetics. The acidogenic bacteria transform the volatile dissolved solids to VFA, which are then transformed by methanogenic microorganisms into CH_4 following Monod kinetics. The authors found that the most sensitive parameters were those related with the hydrolysis step.

All these works provided advances in the identification of the most influential parameters of AD and the estimation of parameters. However, finding any information concerning the SA and parameter estimation has been scarce for biological methanation.

The model constructed by Grimalt-Alemany *et al.* (2020) for biomethanation was analyzed with LSA to evaluate the model outputs sensitivity with the estimated parameter values. They performed a parameter estimation, especially of the parameters related to the specific growth rate of carboxydotrophic acetogens and hydrogenotrophic methanogens. Santus *et al.* (2022) proposed an *ex-situ* biological methanation model based on a simplified and modified version of the ADM1, just considering hydrogenotrophic methanogenesis. The authors performed a SA to assess the most significant parameters via an individual parameter-based sensitivity analysis. The analyzed parameters were the volumetric mass transfer coefficient of oxygen and all the related with the hydrogenotrophic kinetics and stoichiometric. The analyzed outputs were the H_2 transfer rate, CH_4 production rate, and CH_4 concentration in the gas phase. As conclusion they found that CH_4 concentration in gas phase was highly influenced by the maximum specific uptake of hydrogenotrophic methanogens, the volumetric mass transfer coefficient of H_2, and the half-saturation constant on dissolved H_2.

2.5 Conclusions of Modeling and Simulation

Several models for AD have been developed over the last decades, and several approaches were consolidated during this time. The emergence of powerful computational machinery has allowed the possibility to explore AD by mathematically complex models. The modeling of AD started with models that represent the hydrolysis, acetogenesis, and methanogenesis, which was focused on the production of volatile acids as a whole and the subsequent production of CH_4 and CO_2 (Andrews and Graef, 1971). Then, more sophisticated models were developed. Those models permitted the characterization of primary volatile acids, such as propionate, valerate, butyrate, and acetate, as well as the differentiation in the CH_4 and CO_2 produced from acetate (acetoclastic methanogens) and H_2 (hydrogenotrophic methanogens) (Batstone *et al.*, 2002). In this regard, the ADM1 proposed by the IWA Anaerobic Digestion Model Task Group is probably the most used model to simulate and predict AD. This model has been extended to study ammonia inhibition (Bai *et al.*, 2017; Li *et al.*, 2019), the prediction of interactions

between phosphorus, sulfur, and iron in plant-wide simulation (Flores-Alsina *et al.*, 2016), and for model reduction to explore a specific task desire for the researchers (Hassam *et al.*, 2015; Arzate *et al.*, 2017; Weinrich and Nelles, 2021).

The application of ADM1 in biological methanation is quite recent. Therefore, a few investigations have intended to develop a model representing biological methanation. They used well-known theories, such as the two-film theory, to calculate the different volumetric mass transfer coefficients (Sun *et al.*, 2021) or to explore different reactor configurations for biological methanation (Santus *et al.*, 2022).

Another highlighted aspect is assessing the SA and parameter estimation considering confidence intervals. Some researchers explored the GSA over models representing the AD, which are frequently simplified versions of the ADM1 model (Donoso-Bravo *et al.*, 2013; Schroyen *et al.*, 2018). Nevertheless, in biological methanation, GSA has not been explored yet.

Unfortunately, these models present weaknesses concerning biological methanation, *e.g.*, restrictions for the generalization of the model given its formulation for particular conditions, the increased complexity of the model, or the incomplete representation of the most relevant variables and phenomena in the biological methanation.

Based on this, some questions can be drawn:

- Can a mathematical model of biological methanation accurately reproduce multiple operational conditions with emphasis on different liquid OLR, syngas addition, and varying GLR?
- How can the transformation of CO into acetate and H2 and their inhibitions be described in a model for biological methanation?

2.6 References

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). 2018.

Andrews, J.F., Graef, S.P., 1971. Dynamic Modeling and Simulation of the Anaerobic Digestion Process, in: Anaerobic Biological Treatment Processes, Advances in Chemistry. American Chemical Society, pp. 126–162. https://doi.org/10.1021/ba-1971-0105.ch008

Angelidaki, I., Ellegaard, L., Ahring, B.K., 1993. A mathematical model for dynamic simulation of anaerobic digestion of complex substrates: Focusing on ammonia inhibition. Biotechnol. Bioeng. 42, 159–166. https://doi.org/10.1002/bit.260420203

Arzate, J.A., Kirstein, M., Ertem, F.C., Kielhorn, E., Ramirez Malule, H., Neubauer, P., Cruz-Bournazou, M.N., Junne, S., 2017. Anaerobic Digestion Model (AM2) for the Description of Biogas Processes at Dynamic Feedstock Loading Rates. Chemie Ingenieur Technik 89, 686–695. https://doi.org/10.1002/cite.201600176

Ashraf, H.M., Abu-Reesh, I.M., 2022. Variance-based global sensitivity analysis of a multi-population, single-chamber microbial fuel cell operating in continuous flow mode at steady state. Biomass Conv. Bioref. https://doi.org/10.1007/s13399-022-03429-6

Bai, J., Liu, H., Yin, B., Ma, H., Chen, X., 2017. Modified ADM1 for modeling free ammonia inhibition in anaerobic acidogenic fermentation with high-solid sludge. Journal of Environmental Sciences 52, 58–65. https://doi.org/10.1016/j.jes.2016.03.004

Barahmand, Z., Samarakoon, G., 2022. Sensitivity Analysis and Anaerobic Digestion Modeling: A Scoping Review. Fermentation 8, 624. https://doi.org/10.3390/fermentation8110624

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). Water Science and Technology 45, 65–73. https://doi.org/10.2166/wst.2002.0292

Blumensaat, F., Keller, J., 2005. Modelling of two-stage anaerobic digestion using the IWA Anaerobic Digestion Model No. 1 (ADM1). Water Research 39, 171–183. https://doi.org/10.1016/j.watres.2004.07.024

Calise, F., Cappiello, F.L., Dentice d'Accadia, M., Infante, A., Vicidomini, M., 2020. Modeling of the Anaerobic Digestion of Organic Wastes: Integration of Heat Transfer and Biochemical Aspects. Energies 13, 2702. https://doi.org/10.3390/en13112702

Damblin, G., Couplet, M., Iooss, B., 2013. Numerical studies of space-filling designs: optimization of Latin Hypercube Samples and subprojection properties. Journal of Simulation 7, 276–289. https://doi.org/10.1057/jos.2013.16

Dochain, D., Vanrolleghem, P., 2015. Dynamical Modelling & Estimation in Wastewater Treatment Processes. Water Intelligence Online 4, 9781780403045–9781780403045. https://doi.org/10.2166/9781780403045

Donoso-Bravo, A., Mailier, J., Ruiz-Filippi, G., Vande Wouwer, A., 2013. Identification in an anaerobic batch system: global sensitivity analysis, multi-start strategy and optimization criterion selection. Bioprocess Biosyst Eng 36, 35–43. https://doi.org/10.1007/s00449-012-0758-5

Efron, B., Stein, C., 1981. The Jackknife Estimate of Variance. Ann. Statist. 9. https://doi.org/10.1214/aos/1176345462

Ersahin, M.E., 2018. Modeling the dynamic performance of full-scale anaerobic primary sludge digester using Anaerobic Digestion Model No. 1 (ADM1). Bioprocess Biosyst Eng 41, 1539–1545. https://doi.org/10.1007/s00449-018-1981-5

Esposito, G., Frunzo, L., Panico, A., Pirozzi, F., 2011. Model calibration and validation for OFMSW and sewage sludge co-digestion reactors. Waste Management 31, 2527–2535. https://doi.org/10.1016/j.wasman.2011.07.024

Feng, K., Lu, Z., Yang, C., 2019. Enhanced Morris method for global sensitivity analysis: good proxy of Sobol' index. Struct Multidisc Optim 59, 373–387. https://doi.org/10.1007/s00158-018-2071-7

Flores-Alsina, X., Solon, K., Kazadi Mbamba, C., Tait, S., Gernaey, K.V., Jeppsson, U., Batstone, D.J., 2016. Modelling phosphorus (P), sulfur (S) and iron (Fe) interactions for dynamic simulations of anaerobic digestion processes. Water Research 95, 370–382. https://doi.org/10.1016/j.watres.2016.03.012

Grimalt-Alemany, A., Asimakopoulos, K., Skiadas, I. V., Gavala, H.N., 2020. Modeling of syngas biomethanation and catabolic route control in mesophilic and thermophilic mixed microbial consortia. Applied Energy 262, 114502. https://doi.org/10.1016/j.apenergy.2020.114502

Hassam, S., Ficara, E., Leva, A., Harmand, J., 2015. A generic and systematic procedure to derive a simplified model from the anaerobic digestion model No. 1 (ADM1). Biochemical Engineering Journal 99, 193–203. https://doi.org/10.1016/j.bej.2015.03.007

Henze, M., Grady Jr, C.P.L., Gujer, W., Larais, G. v. R., Matsou, T., 1987. Activated Sludge Model No 1. IAWQ Scientific and Technical Report n°1, IAWQ, London, UK.

Hill, D.T., Barth, C.L., 1977. A Dynamic Model for Simulation of Animal Waste Digestion. Water Pollution Control Federation 49, 2129–2143.

Iooss, B., Lemaître, P., 2015. A Review on Global Sensitivity Analysis Methods. Operations Research/Computer Science Interfaces Series 59. https://doi.org/10.1007/978-1-4899-7547-8_5

Kiparissides, A., Kucherenko, S.S., Mantalaris, A., Pistikopoulos, E.N., 2009. Global Sensitivity Analysis Challenges in Biological Systems Modeling. Ind. Eng. Chem. Res. 48, 7168–7180. https://doi.org/10.1021/ie900139x

Kiparissides, A., Rodriguez-Fernandez, M., Kucherenko, S., Mantalaris, A., Pistikopoulos, E., 2008. Application of global sensitivity analysis to biological models, in: Computer Aided Chemical Engineering. Elsevier, pp. 689–694. https://doi.org/10.1016/S1570-7946(08)80120-4

Kleinstreuer, C., Poweigha, T., 1982. Dynamic simulator for anaerobic digestion processes. Biotechnol. Bioeng. 24, 1941–1951. https://doi.org/10.1002/bit.260240903

Kucherenko, S., Delpuech, B., Iooss, B., Tarantola, S., 2015. Application of the control variate technique to estimation of total sensitivity indices. Reliability Engineering & System Safety 134, 251–259. https://doi.org/10.1016/j.ress.2014.07.008

Li, X., Yang, Z., Liu, G., Ma, Z., Wang, W., 2019. Modified anaerobic digestion model No.1 (ADM1) for modeling anaerobic digestion process at different ammonium concentrations. Water Environment Research 91, 700–714. https://doi.org/10.1002/wer.1094

Mata-Alvarez, J., 1987. A dynamic simulation of a two-phase anaerobic digestion system for solid wastes. Biotechnol. Bioeng. 30, 844–851. https://doi.org/10.1002/bit.260300706

Moletta, R., 1986. Dynamic modelling of anaerobic digestion. Water Research 20, 427–434. https://doi.org/10.1016/0043-1354(86)90189-2

Morio, J., 2011. Global and local sensitivity analysis methods for a physical system. Eur. J. Phys. 32, 1577–1583. https://doi.org/10.1088/0143-0807/32/6/011

Morris, M.D., 1991. Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics 33, 161–174. https://doi.org/10.1080/00401706.1991.10484804

Mosey, F.E., 1983. Mathematical Modelling of the Anaerobic Digestion Process: Regulatory Mechanisms for the Formation of Short-Chain Volatile Acids from Glucose. Water Science and Technology 15, 209–232. https://doi.org/10.2166/wst.1983.0168

Ochoa, M.P., Estrada, V., Di Maggio, J., Hoch, P.M., 2016a. Dynamic global sensitivity analysis in bioreactor networks for bioethanol production. Bioresource Technology 200, 666–679. https://doi.org/10.1016/j.biortech.2015.10.069

Ochoa, M.P., Estrada, V., Hoch, P., 2016b. Wastewater stabilisation ponds system: global sensitivity analysis on network design. Chemical Engineering Transactions 50, 187–192. https://doi.org/10.3303/CET1650032

Ozgun, H., 2019. Anaerobic Digestion Model No. 1 (ADM1) for mathematical modeling of full-scale sludge digester performance in a municipal wastewater treatment plant. Biodegradation 30, 27–36. https://doi.org/10.1007/s10532-018-9859-4

Rapadamnaba, R., Ribatet, M., Mohammadi, B., 2021. Global sensitivity analysis for assessing the parameters importance and setting a stopping criterion in a biomedical inverse problem. Int J Numer Meth Biomed Engng 37. https://doi.org/10.1002/cnm.3458

Rodríguez-Fernández, M., Alonso, A.A., Banga, J.R., 2005. Robust parameter estimation in nonlinear dynamic process models, in: Computer Aided Chemical Engineering. Elsevier, pp. 37–42. https://doi.org/10.1016/S1570-7946(05)80128-2

Rodriguez-Fernandez, M., Kucherenko, S., Pantelides, C., Shah, N., 2007. Optimal experimental design based on global sensitivity analysis, in: Computer Aided Chemical Engineering. Elsevier, pp. 63–68. https://doi.org/10.1016/S1570-7946(07)80034-4

Rodriguez-Fernandez, M., Rehberg, M., Kremling, A., Banga, J.R., 2013. Simultaneous model discrimination and parameter estimation in dynamic models of cellular systems. BMC Syst Biol 7, 76. https://doi.org/10.1186/1752-0509-7-76

Rosen, C., Jeppsson, U., 2006. Aspects on ADM1 Implementation within the BSM2 Framework, Technical report. Department of Industrial Electrical Engineering and Automation (IEA), Lund University, Lund, Sweden.

Ruano, M.V., Ribes, J., Ferrer, J., Sin, G., 2011. Application of the Morris method for screening the influential parameters of fuzzy controllers applied to wastewater treatment plants. Water Science and Technology 63, 2199–2206. https://doi.org/10.2166/wst.2011.442

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., 2017. Why So Many Published Sensitivity Analyses Are False. A Systematic Review of Sensitivity Analysis Practices. https://doi.org/10.48550/ARXIV.1711.11359

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2007. Global Sensitivity Analysis. The Primer, 1st ed. Wiley. https://doi.org/10.1002/9780470725184

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, 1st ed. Wiley. https://doi.org/10.1002/0470870958

Santus, A., Corbellini, V., Trionfini, M., Malpei, F., Ferretti, G., 2022. Modelling and Parameter Identification of Ex-Situ Biological Biogas Upgrading. IFAC-PapersOnLine 55, 624–629. https://doi.org/10.1016/j.ifacol.2022.09.165

Schroyen, M., Vervaeren, H., Raes, K., Van Hulle, S.W.H., 2018. Modelling and simulation of anaerobic digestion of various lignocellulosic substrates in batch reactors: Influence of

lignin content and phenolic compounds II. Biochemical Engineering Journal 134, 80–87. https://doi.org/10.1016/j.bej.2018.03.017

Sepúlveda, F.D., Cisternas, L.A., Gálvez, E.D., 2014. The use of global sensitivity analysis for improving processes: Applications to mineral processing. Computers & Chemical Engineering 66, 221–232. https://doi.org/10.1016/j.compchemeng.2014.01.008

Sepulveda, F.D., Cisternas, L.A., Gálvez, E.D., 2013. Global sensitivity analysis of a mineral processing flowsheet, in: Computer Aided Chemical Engineering. Elsevier, pp. 913–918. https://doi.org/10.1016/B978-0-444-63234-0.50153-6

Sobol′, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation 55, 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6

Sohier, H., Farges, J.-L., Piet-Lahanier, H., 2014. Improvement of the Representativity of the Morris Method for Air-Launch-to-Orbit Separation. IFAC Proceedings Volumes 47, 7954–7959. https://doi.org/10.3182/20140824-6-ZA-1003.01968

Sun, H., Yang, Z., Zhao, Q., Kurbonova, M., Zhang, R., Liu, G., Wang, W., 2021. Modification and extension of anaerobic digestion model No.1 (ADM1) for syngas biomethanation simulation: From lab-scale to pilot-scale. Chemical Engineering Journal 403, 126177. https://doi.org/10.1016/j.cej.2020.126177

Tosin, M., Côrtes, A.M.A., Cunha, A., 2020. A Tutorial on Sobol' Global Sensitivity Analysis Applied to Biological Models, in: da Silva, F.A.B., Carels, N., Trindade dos Santos, M., Lopes, F.J.P. (Eds.), Networks in Systems Biology, Computational Biology. Springer International Publishing, Cham, pp. 93–118. https://doi.org/10.1007/978-3-030-51862-2_6

Tsapekos, P., Alvarado-Morales, M., Angelidaki, I., 2022. H2 competition between homoacetogenic bacteria and methanogenic archaea during biomethanation from a combined experimental-modelling approach. Journal of Environmental Chemical Engineering 10, 107281. https://doi.org/10.1016/j.jece.2022.107281

Weinrich, S., Nelles, M., 2021. Systematic simplification of the Anaerobic Digestion Model No. 1 (ADM1) – Model development and stoichiometric analysis. Bioresource Technology 333, 125124. https://doi.org/10.1016/j.biortech.2021.125124

Zhang, X., Trame, M., Lesko, L., Schmidt, S., 2015. Sobol Sensitivity Analysis: A Tool to Guide the Development and Evaluation of Systems Pharmacology Models. CPT Pharmacometrics Syst. Pharmacol. 4, 69–79. https://doi.org/10.1002/psp4.6

Zi, Z., 2011. Sensitivity analysis approaches applied to systems biology models. IET Systems Biology 5, 336–346. https://doi.org/10.1049/iet-syb.2011.0015

Section 3 Bioprocess Control and Optimization

Summary of Section 3

This section introduces model-based control approaches, such as MPC, which has worked adequately in bioprocess, and it has even been extended to optimize the economic process performance rather than tracking to a set point. One of the common needs of biological processes is the optimization of multiple variables simultaneously, and those variables are commonly conflicting. Optimal solution approaches such as Pareto optimal sets are presented as a strategy to find the trade-offs between conflicting variables, such as yields, productivities, or others related to the economic aspects of the process.

3.1 Model-Based Control

The industry continues to have difficulties developing products based on biological processes, particularly when trying to obtain these products from raw materials, which are wastes from other processes. The implementation of biological processes can be expensive due to different factors, for example, the cost of equipments and materials necessary to carry out the process. Consequently, different model-based optimization techniques have been used to analyze and improve them. However, this kind of optimization begins with experimentation (usually at a laboratory scale) that enables an understanding of biological processes. Then, through modeling and simulation, the strategy to optimize these biological processes is developed without the need to continue with experimental development (Mitsos *et al.*, 2018).

Biological processes are complex systems whose dynamics are governed by the non-linear behavior of the microorganisms involved (Van Impe, 1996). The challenge in its development is maintaining the best (frequently called "optimal") environmental conditions for the proper growth of the different microorganisms. Nevertheless, the characteristics of the bioprocesses are time-varying, making them sensitive to disturbances, resulting in deviations from the desired operating conditions.

The formulation of a model allows the development of monitoring, optimization, and control techniques (Rathore *et al.*, 2021). Bioprocess control focuses on maintaining an optimal condition for microorganisms to grow, multiply, and generate the desired product (Alford, 2006). However, the complexity and non-linearity of the bioprocesses can make them difficult to control (Doran, 2013).

The principles in biological processes are similar to chemical processes; the only difference is the nature of the catalyst (bacteria, fungi, or enzymes). Indeed, control approaches used in chemical processes can be applied to biological processes (Alford, 2006; Luo *et al.*, 2021).

The most common closed-loop controller is the Proportional Integral Derivative (PID). In this controller, the difference between the controlled variable and the set point (desired value) at time t, is used to calculate the control action (Marlin, 1995). This control action is determined

in proportion to the error, the integral of the error, and the derivative of the error concerning time (Doran, 2013). PID is a controller that works well for processes with a linear behavior or in a pseudo-linear region (Bastin and Dochain, 1990; Pind *et al.*, 2003; Alford, 2006; Rathore *et al.*, 2021). However, it does not consider the future behavior of the process (predictions) on the current control actions (Rossiter, 2017). The model-based control is an alternative since they use a model as a basis. The most commonly used model-based controls are adaptive control, optimal control, optimal adaptive control, and model predictive control.

3.1.1 Adaptive Control

The adaptive control used in biological processes is based on the fact that some kinetics are unknown. Parametric estimation is implemented and confers the property of adapting itself to variations in the kinetics. In other words, it can modify its behavior in response to changes in the dynamics of the process and the character of the disturbances (Åström and Wittenmark, 2008).

There are two approaches used commonly in adaptive control. One is the self-tuning regulator model adaptive control, commonly known as indirect adaptive control, which first recursively identifies the unknown model parameters and then uses these estimates to update the controller parameters through some fixed transformation. Another approach is the reference adaptive control or direct adaptive control, which updates the controller parameters directly from the measurements of the prediction error (Sastry and Bodson, 1989).

Adaptive control presents some disadvantages. The determination of an optimal strategy varied and rich in information is necessary to guarantee that the model parameters can be identified. In other words, it is necessary to know the maximum prior knowledge concerning the process in advance. The difficulty in tuning the controller parameters and sensitivity to system variability make the adaptive control susceptible to adapt inappropriately (Sastry and Bodson, 1989; Van Impe, 1996).

Adaptive control has been applied in AD processes. In Renard *et al.* (1988), an adaptive control was used to regulate the substrate in the output at reference values, despite fluctuations of the input concentration by acting in the HRT. The model considered a simplified AD process,

including the net accumulation of biomass, substrate, and CH_4 gas production rate. The control strategy was evaluated in a 60 L $CSTR$ mechanically agitated, operating at $35°C$. As substrate was used spent liquor from citric acid fermentation (50% volatile solids), which was recirculated with a flow rate of 10 L/h. Petre *et al.* (2013) used an adaptive control for AD in wastewater treatment. The authors proposed a dynamic model representing the substrates, glucose, acetate, CO_2, CH_4, and the acidogenic and acetoclastic methanogens involved in AD. The design of the controller was performed considering that: (i) biomass dynamic is not available, (ii) the specific reaction rates are complety unknown, (iii) the on-line measurements are the inlet and outlet glucose concentrations, and (iv) model output are glucose and acetate. In other words, the control problem of output pollution level using an appropriate control input. A comprehensive review of adaptive control strategies is presented by Pind *et al.* (2003) where different objectives were studied, *e.g.*, regulation at a particular reference point or optimization of process performance.

3.1.2 Optimal Control

The Optimal Control approach aims at optimizing a reference trajectory (Bryson and Ho, 1975; Sastry and Bodson, 1989).

Consider the following objective function, generated from the dynamic system ξ, Equation (1.3.1).

$$J = \varphi\left(\xi\big(Y(t_f), u(t_f), t_f\big)\right) + \int_{t_0}^{t_f} L\big(Y(t_f), u(t_f), t_f\big)\, dt \qquad (\ 1.3.1\ )$$

where φ is the cost function that evaluates the final state of the system (terminal cost), L is the instantaneous cost function that measures the performance at each instant time t. The second term represents the integral of the instantaneous cost function over time (the sum of the instantaneous cost functions over time).

Optimal control is formulated as the optimization problem of Equation (1.3.2), subject to the constraints in Equation (1.3.3).

$$\min_{u(t)} J \tag{1.3.2}$$

$$Subject\ to \begin{cases} \dfrac{dY}{dt} = \xi(Y, u, \theta, t) & t \in [0, t_f] \\ \lambda_i(Y,\ u, \theta, t) \leq 0 & i = 1,2, \dots, n_\lambda \\ \psi_i(Y,\ u, \theta, t) = 0 & i = 1,2, \dots,\ n_\psi \\ u^L \leq u \leq u^U \end{cases} \tag{1.3.3}$$

where Y corresponds to the states variables, $\lambda_i(Y)$ and $\psi_i(Y)$ are the n_λ inequality and n_ψ equality constraints, respectively. u^L, u^U correspond to the lower and upper bounds for the manipulated inputs u.

Optimization techniques such as the principle of Pontryagin or the Bellman method are used to solve this problem (Bertsekas, 2007; Zabczyk, 2008).

Both control strategies present some limitations. Optimal control could fail due to modeling uncertainties, while adaptive control requires complete knowledge of the kinetic functions (Bastin and Van Impe, 1995; Lewis *et al.*, 2012b; Vrabie *et al.*, 2013) and it does not guarantee the optimality of the results (Lewis *et al.*, 2012a; T. Nguyen, 2014; Van Impe and Bastin, 1995).

3.1.3 Optimal Adaptive Control

The optimal adaptive control arises from the need to integrate the best of optimal and adaptive controllers (Van Impe and Bastin, 1995).

Smets *et al.* (2004) proposed an optimal adaptive control approach for biological processes. Their methodology was derived in three steps. **In step 1,** the process model was assumed to be well-known. Then, an optimal control solution for a given optimization problem was computed. **In step 2**, a nearly optimal heuristic controller was constructed based on analyzing the optimal control from biological and mathematical perspectives. To do this, the process variables that characterize optimal control solutions were selected and the reference trajectory was constructed for the characteristic process variable as a function of time. **In the last step**, the heuristic model controller was incorporated inside a linearized controller, and the adaptive estimation of the states and parameters was performed on-line.

This approach was used to design a substrate feeding rate controller of a fed-batch reactor. Three implementations were tested to optimize a penicillin G fed-batch fermentation process, depending on which variables are available for on-line measurements, (i) substrate and biomass concentration in the reactor, (ii) only substrate concentration, and (iii) carbon dioxide evolution rate (Van Impe and Bastin, 1995).

3.1.4 Model Predictive Control

Model Predictive Control (MPC) is the type of controller where the control actions are based on the optimization of a criterion (Camacho and Bordons, 2007). This criterion is associated with the future behavior of the system, predicted by a dynamic model (Rossiter, 2017).

The MPC implementation follows several steps (Camacho and Bordons, 2007). **Step 1.** The process model predicts the future outputs for a certain horizon N (prediction horizon) at each instant t. These predicted outputs $\{Y(t+k|t)\ \ k=1\ldots N\}$ depend on the known values up to instant t and on the future control signal $\{u=(t+k|t)\ \ k=0,\ldots,N-1\}$. **Step 2.** The set of future control signals is calculated by minimizing a criterion (cost function) to maintain the process near the reference trajectory $\{\mathrm{r}(t+k|t)\ \ k=1\ldots N\}$. **Step 3**. The control signal $u(t\mid t)$ is given to the process while the subsequent control signals computed are rejected because $\mathrm{Y}(t+1)$ is already known. Finally, step 1 is repeated, and the system is updated. The MPC strategy is represented in Figure 1.3.1.

The use of MPC considers 4 crucial elements: The prediction model (dY/dt) to describe de behavior of the system, the cost function (J) which indicates the optimization criterium, the constraints $(\lambda_i, \psi_i, u^L, u^U)$ to bound the evolution of the system, and the optimization method to minimize the cost function.

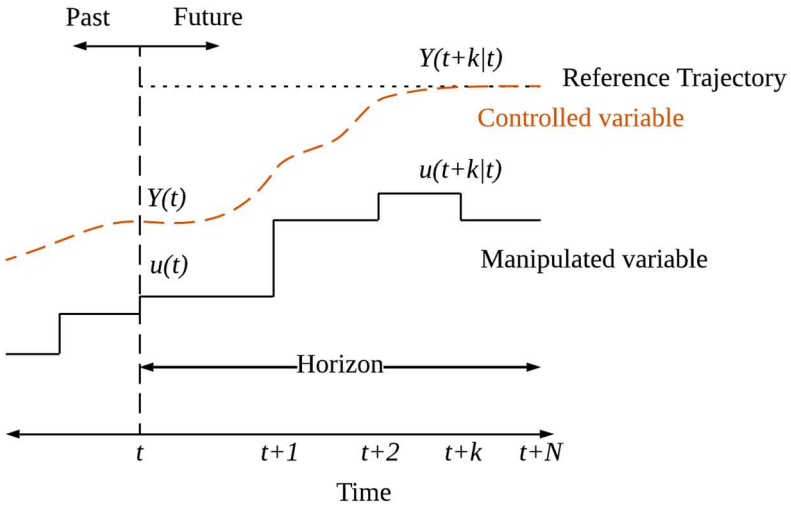


Figure 1.3.1. Model Predictive Control Strategy. Adapted from (García *et al.*, 1989).

If it is assumed that the objective function generated from the dynamic system (ξ), could be represented by Equation (1.3.4).

$$J = \varphi\big(\xi(Y(t), u(t), t)\big) \qquad (1.3.4)$$

Thus the optimization problem for the MPC is formulated as, Equation (1.3.5), subject to the constraints in Equation (1.3.6).

$$\min_{u(t)} J = \min_{u(t)} \varphi\big(\xi(Y(t), u(t), t)\big) \qquad (1.3.5)$$

$$Subject\ to \begin{cases} \dfrac{dY}{dt} = \xi(Y,\ u, \theta, t) & t \in [0, t_f] \\ \lambda_i(Y,\ u, \theta, t) \leq 0 & i = 1,2, \dots, n_\lambda \\ \psi_i(Y,\ u, \theta, t) = 0 & i = 1,2, \dots,\ n_\psi \\ u^L \leq u \leq u^U \end{cases} \qquad (1.3.6)$$

MPC has been widely applied for bioprocess optimization (Rathore *et al.*, 2021), *e.g.*, to control full-scale biogas production according to the demand (Mauky *et al.*, 2016), to adjust glucose and lactose feed in a fed-batch reactor producing a green fluorescent protein with *E.coli* (Ulonska *et al.*, 2018), to control variables in the AD such as CH_4 production rate, using manipulated variables such as glucose flow rate (Ahmed and Rodríguez, 2020). It has also been

applied in the pharmaceutical industry to control the excessive lactate production in a fed-batch reactor to cultivate Chinese hamster ovary cells (Schmitt *et al.*, 2019).

There are some advantages of using MPC (García *et al.*, 1989; Yamashita *et al.*, 2016; Rossiter, 2017):

- The incorporation of an explicit model for the process calculations. This means the consideration of the dynamic characteristics of the process
- It is possible to know the effect that the disturbances caused on the process
- The possibility of incorporating constraints on the system (physical limitations of the processes)
- High acceptance at the industrial level

However, the MPC feasibility could be affected by its strong dependence on the model, making it less effective when the system dynamics are not accurately captured or when unexpected disturbances arise that are not accounted in the process model (Chinea-Herranz & Rodríguez, 2012; Schwenzer *et al*., 2021).

3.1.5 Economic Model Predictive Control

The current challenge of MPC is the integration of dynamic market-driven operations, including more efficient and agile operations (Ellis *et al.*, 2017). The solution is to consider the economic objective (concerned management, scheduling, or involving the multivariable loop controls) directly in the cost function of the control system and to redefine the MPC towards a new approach known as Economic Model Predictive Control (EMPC). The EMPC controller is developed to optimize the economic process performance rather than tracking a set point (Rawlings *et al.*, 2012; Zhang *et al.*, 2014).

The tracking cost function usually uses a quadratic cost that penalizes the deviation of state and inputs from the corresponding reference trajectory. Nonetheless, the EMPC cost function could use any general stage cost that reflects the process/system economics (Ellis *et al.*, 2017).

Commonly, tracking MPC optimization problems takes the following general form (Limon *et al.*, 2014):

Optimize: Tracking cost function (constant or changing set-points)

Subject to:

- Dynamic model initialized with state measurement/estimate
- State/input constraints
- Stability constraints

However, the EMPC optimization problem takes the following general form (Limon *et al.*, 2014):

Optimize: Economic cost function

Subject to:

- Dynamic model initialized with state measurement/estimate
- State/input constraints
- Economic-oriented constraints
- Stability constraints

The paradigm between an MPC and EMPC relies on the operation of the processes. A MPC aims to maintain a feasible steady state, although the steady state is not necessarily the best economic operation. Moreover, EMPC aims to determine the optimal operating strategy based on the economic aspects respecting operational constraints. These economic aspects could be real-time energy, substrates pricing, or time-varying disturbances (Rawlings *et al.*, 2012; Ellis *et al.*, 2017).

Let us consider the following objective function generated from the dynamic system (ξ), Equation (1.3.7).

$$J = \omega_1 \cdot \varphi_1\big(\xi(Y(t), u(t), t)\big) + \omega_2 \cdot \varphi_2(u(t)) \qquad (\ 1.3.7\ )$$

The optimization problem for the EMPC is formulated as Equation (1.3.8), subject to the constraints in Equation (1.3.9).

$$\min_{u(t)} J = \big(\omega_1 \cdot \varphi_1\big(\xi(Y(t), u(t), t)\big) + \omega_2 \cdot \varphi_2(u(t))\big) \qquad (\ 1.3.8\ )$$

$$Subject\ to \begin{cases} \dfrac{dY}{dt} = \xi(Y,\ u, \theta, t) & t \in [0, t_f] \\ \lambda_i(Y(t),\ u(t), \theta, t) \leq 0 & i = 1,2, \dots, n_\lambda \\ \psi_i(Y(t),\ u(t), \theta, t) = 0 & i = 1,2, \dots,\ n_\psi \\ u^L \leq u(t) \leq u^U \end{cases} \quad (\ 1.3.9\ )$$

The first term $\omega_1 \cdot \varphi_1\big(\xi(Y(t), u(t)), t\big)$ represents the process performance criterium. The second term $\omega_2 \cdot \varphi_2(u(t))$ holds for the economic performance criterium, where $\varphi_2(u(t))$ is the cost related to the control inputs $u(t)$. ω_1 and ω_2 represent the relative weighting between the process and economic performance criteria and the relative weighting between the cost of the control inputs and economic performance criteria.

EMPC has been applied in different areas: optimization of the production of liquid oxygen and nitrogen from an air separation process with an integrated liquefication cycle and liquid assist operation (Caspari *et al.*, 2019), in drinking water networks (Limon *et al.*, 2014), in thermal energy storage (Touretzky and Baldea, 2014), in chemical processes (Santander *et al.*, 2016), in the aeration systems of a full-scale wastewater treatment plant (Nejjari *et al.*, 2017). In this paper, Nejjari and co-workers developed an EMPC strategy to control the dissolved oxygen concentrations in an aerated reactor of a wastewater treatment plant while optimizing the effluent quality and operating cost. The activated sludge models (ASM) and the Benchmark Simulation Model number 1 (BSM1) were used as dynamic models to represent the plant. The performance of the objective function considered the tracking term (J_{track}), the output water quality term (J_{qual}), the smooth set-points for equipment conservation (J_{smo}), and the economic cost (J_{eco}), Equation (1.3.10).

$$J = \sum_{k=0}^{Hp-1} J_{eco}(t) + \sum_{k=1}^{Hp} J_{track}(t) + \sum_{k=1}^{Hp} J_{qual}(t) + \sum_{k=0}^{Hp-1} J_{smo}(t) \quad (\ 1.3.10\ )$$

where $J_{eco}(t)$ was calculated using Equation (1.3.11).

$$J_{eco}(t) = W_e\big(\omega_1 \cdot u(t) + \omega_2(t) \cdot u(t)\big) \quad (\ 1.3.11\ )$$

herein, ω_1 corresponds to a known vector related to the economic costs of the water treatment, $\omega_2(t)$ holds to a vector associated with the economic cost of the flow through certain actuators (pumps only) and their control cost (pumping), in this case $\omega_2(t)$ was time-varying due to pumping effort having different values according to the time of the day (electricity costs). W_e refers to the weight matrix that expresses the relative priority of one objective concerning the others.

The use of EMPC presents advantages such as improvements in economic performance by integrating the process operation tasks of scheduling and the possibility of using it as a decision-making tool between control objectives and economic aspects (Angeli, 2013; Limon *et al.*, 2014). However, the EMPC also presents disadvantages relate to the MPC, such as dependency on the model accuracy (Caspari *et al.*, 2019).

According to Ellis *et al.* (2017), there are three challenges in working with EMPC:

- To establish correctly a mathematic representation of the economic terms for the process in the construction of the cost function, the process model, and the constraints
- To guarantee through the formulation of an EMPC control theory, the essential control properties, such as closed-loop stability
- To develop the numerical computational algorithms that will allow the application of the desired control actions in real-time operation

3.2 Dynamic & Multi-Objective Optimization

Dynamic optimizations have been applied successfully over biological processes. However, optimal decisions must be made to find trade-offs between two or more conflicting variables. Multi-objective Optimization (MOO) is a research field of multiple-criteria decision-making that involves the optimization of more than one objective simultaneously (Chang, 2015). Commonly, the m objective functions $(J_m(Y))$ are conflicting, and the number of solutions (trade-offs) might be infinite.

In biological process optimization, the solution could be of two types, local or global. Local solutions strongly depend on the initial value used for the optimization, while global solutions are computationally more expensive but provide a global solution to the process.

A MOO can be mathematically stated as in (Sawaragi *et al.*, 1985; Ahmadi *et al.*, 2016), Equation (1.3.12) subject to the constraints in Equation (1.3.13).

$$\min_{Y,u,\theta,t} \{J_1(Y,u,\theta), \dots, J_m(Y,u,\theta)\} \qquad (1.3.12)$$

$$Subject\ to \begin{cases} \dfrac{dY}{dt} = \xi(Y,\ u,\theta,t) & t \in [0, t_f] \\ \lambda_i(Y,u,\theta,t) \le 0 & i = 1,2,\dots,n_\lambda \\ \psi_i(Y,u,\theta,t) = 0 & i = 1,2,\dots,\ n_\psi \\ u^L \le u(t) \le u^U & \end{cases} \qquad (1.3.13)$$

where $J_1, \dots, J_m$ are the m objective functions to minimize, Y are the state variables (ODE system), λ_i and ψ_i indicate inequality and equality constraints on the states variables, u are the control variables, θ are the parameters, $u^L$and u^U correspond to the lower and upper bounds of the control variable.

Two approaches have been used to solve multi-objective optimization problems. The first method consists of weighting the objectives and then optimizing the weighted sum. However, the weights need to be predefined, and this choice could be ambiguous. The second approach consists of finding a set of optimal solutions (via Pareto fronts), including the trade-off between objectives (Coello Coello, 1999; Logist *et al.*, 2010; Bortz *et al.*, 2014; Mitsos *et al.*, 2018). Each solution is considered a Pareto Optimal Point (POP) if it is not dominated by any other solution in the solution space; all these solutions are the well-known Pareto optimal set (POS), also called the Pareto front (see Figure 1.3.2) (Bortz *et al.*, 2014). A solution is called non-dominated if none of the objective functions can be improved without degrading some other objective functions (Konak *et al.*, 2006; Dupont *et al.*, 2008; Ahmadi *et al.*, 2016).

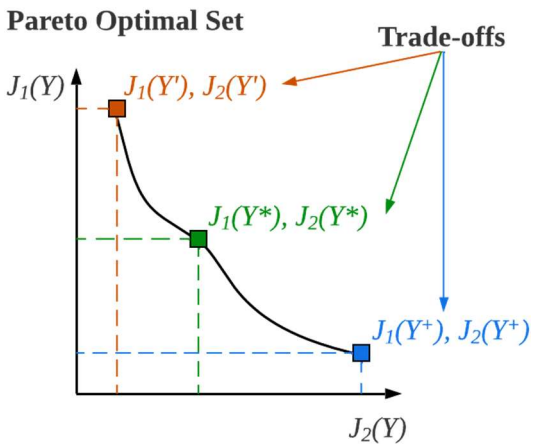


Figure 1.3.2. Pareto Optimal Set for two objective functions $J_1(Y)$ and $J_2(Y)$. Black continuous line (-) correspond to the Pareto Optimal Set. The colored squares (■, ■, ■) correspond to three different Pareto Optimal Points, representing the trade-offs between the two objective functions.

The algorithms for determining Pareto-optimal solutions can be classified into deterministic and stochastic. Deterministic algorithms are based on Scalarization techniques. These are time-efficient and yield accurate results for the optima; however, they usually find only the local optimum related to the starting point. Some examples are the Weighted Sum Approach, Goal Programming, Goal Attainment, and ε-Constraint Method. Stochastic algorithms, also known as Evolutionary Algorithms, try to find the global optimum by sampling large areas of the objective space. This means that more points are evaluated at a time, which lead to time-consuming and less accurate procedures. Some examples are Multiple Objective Genetic Algorithms, Non-Dominated Sorting Genetic Algorithms, Niched Pareto Genetic Algorithms, and Multi-objective Particle Swarm Optimization (Kennedy and Eberhart, 1995; Coello Coello, 1999; Bortz *et al.*, 2014; Reyes-Sierra and Coello, 2006; Mitsos *et al.*, 2018).

3.3 Conclusions Bioprocess Control and Optimization

There are several model-based control approaches with satisfactory results at the industrial level. However, the MPC becomes at replacing the PID controllers successfully at the laboratory and industrial levels (Ahmed and Rodríguez, 2020; Mauky *et al.*, 2016; Schmitt *et al.*, 2019; Ulonska *et al.*, 2018). The main advantages of MPC are the consideration of the

dynamic characteristics of the process, the possibility of knowing the effect of the disturbances caused on the process, and the incorporation of the physical limitations of the processes (García *et al.*, 1989; Yamashita *et al.*, 2016; Rossiter, 2017). Other control process approaches combine some of the presented controllers, *e.g.*, Jabarivelisdeh *et al.* (2020) presented an adaptive MPC control to consider the biological variability using model-based flux balance analysis to maximize ethanol production.

Another aspect to highlight in this section is the extension of MPC to the approach EMPC as intended to integrate the dynamic market-driven operations (Ellis *et al.*, 2017) by considering economic objectives that reflect the process/system economics (Zhang *et al.*, 2014). Nevertheless, applying model-based controls and approaches such as EMPC on biological methanation remains unexplored.

In the literature, only control strategies for biological methanation that have been reported implemented PI controllers. Bensmann *et al.* (2014) implemented two simple Proportional Integral (PI) controllers to limit the hydrogen added to the system to supply the demands of the product gas. The control strategy aims to detect the transfer limit by the accumulation of H_2 in the gas phase and the biological limit by depletion of CO_2. The controller uses the measure of the molar fractions of H_2 and CO_2 in the systems. If those variables are close to the limits (*i.e.*, the maximum and minimum feasible molar fractions of H_2 and CO_2, respectively), the H_2 flow rate is manipulated. The PI control was tested over three different qualitative cases regarding the violation of the biological limit (case I), the transfer limit (case III), or both (case II), all with satisfactory results in avoiding the accumulation of H_2 in the system. Nevertheless, if the case is to improve the process performance (optimize), it is necessary to consider more advanced techniques such as model-based controls.

This section permits us to elucidate some relevant questions that need to be considered in this thesis, especially related to biological methanation:

- How to implement a computationally feasible model-based control strategy for biological methanation?
- Can the multi-objective optimization approaches improve biological methanation?

- Could the multi-objective optimizations consider several objectives, such as the yields, the productivities, and other variables in economic terms (*e.g.*, substrates prices)?

3.4 References

Ahmadi, A., Tiruta-Barna, L., Capitanescu, F., Benetto, E., Marvuglia, A., 2016. An archive-based multi-objective evolutionary algorithm with adaptive search space partitioning to deal with expensive optimization problems: Application to process eco-design. Computers and Chemical Engineering 87, 95–110. https://doi.org/10.1016/j.compchemeng.2015.12.008

Ahmed, W., Rodríguez, J., 2020. A model predictive optimal control system for the practical automatic start-up of anaerobic digesters. Water Research 174, 115599. https://doi.org/10.1016/j.watres.2020.115599

Alford, J.S., 2006. Bioprocess control: Advances and challenges. Computers & Chemical Engineering 30, 1464–1475. https://doi.org/10.1016/j.compchemeng.2006.05.039

Angeli, D., 2013. Economic Model Predictive Control, in: Baillieul, J., Samad, T. (Eds.), Encyclopedia of Systems and Control. Springer London, London, pp. 1–9. https://doi.org/10.1007/978-1-4471-5102-9_6-1

Åström, K.J., Wittenmark, B., 2008. Adaptive control, 2nd ed. Dover Publ, Mineola, N.Y.

Bastin, G., Dochain, D., 1990. On-line Estimation and Adaptive Control of Bioreactors. Elsevier. https://doi.org/10.1016/C2009-0-12088-3

Bastin, G., Van Impe, J.F., 1995. Nonlinear and Adaptive Control in Biotechnology: A Tutorial. European Journal of Control 1, 37–53. https://doi.org/10.1016/S0947-3580(95)70006-1

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

Bertsekas, D.P., 2007. Dynamic Programming and Optimal Control, 3rd ed, Athena scientific optimization and computation series. Athena scientific, Belmont, Massachusetts.

Bortz, M., Burger, J., Asprion, N., Blagov, S., Böttcher, R., Nowak, U., Scheithauer, A., Welke, R., Küfer, K.H., Hasse, H., 2014. Multi-criteria optimization in chemical process design and decision support by navigation on Pareto sets. Computers and Chemical Engineering 60, 354–363. https://doi.org/10.1016/j.compchemeng.2013.09.015

Bryson, A.E., Ho, Y.-C., 1975. Applied optimal control: optimization, estimation, and control, 1st ed. Cambridge University Press, Cambridge.

Camacho, E.F., Bordons, C., 2007. Model Predictive control. Springer London. https://doi.org/10.1007/978-0-85729-398-5

Caspari, A., Offermanns, C., Schäfer, P., Mhamdi, A., Mitsos, A., 2019. A flexible air separation process: 2. Optimal operation using economic model predictive control. AIChE J 65. https://doi.org/10.1002/aic.16721

Chang, K.-H., 2015. Multiobjective Optimization and Advanced Topics, in: Design Theory and Methods Using CAD/CAE. Elsevier, pp. 325–406. https://doi.org/10.1016/B978-0-12-398512-5.00005-0

Chinea-Herranz, J. A., & Rodríguez, M. (2012). Control of integrated unit operations. In Computer Aided Chemical Engineering (Vol. 31, pp. 325–329). Elsevier. https://doi.org/10.1016/B978-0-444-59507-2.50057-3

Coello Coello, C.A., 1999. A Comprehensive Survey of Evolutionary-Based Multiobjective Optimization Techniques. Knowledge and Information Systems 1, 269–308. https://doi.org/10.1007/BF03325101

Doran, P.M., 2013. Bioprocess engineering principles, 2nd ed. Elsevier Academic Press, Amsterdam Heidelberg.

Dupont, G., Adam, S., Lecourtier, Y., Grilhère, B., Multi, B.G., Grilheres, B., 2008. Multi objective particle swarm optimization using enhanced dominance and guide selection, International Journal of Computational Intelligence Research. Research India Publications.

Ellis, M., Liu, J., Christofides, P.D., 2017. Economic Model Predictive Control, Advances in Industrial Control. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-41108-8

García, C.E., Prett, D.M., Morari, M., 1989. Model predictive control: Theory and practice—A survey. Automatica 25, 335–348. https://doi.org/10.1016/0005-1098(89)90002-2

Jabarivelisdeh, B., Carius, L., Findeisen, R., Waldherr, S., 2020. Adaptive predictive control of bioprocesses with constraint-based modeling and estimation. Computers & Chemical Engineering 135, 106744. https://doi.org/10.1016/j.compchemeng.2020.106744

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization, in: Proceedings of ICNN'95 - International Conference on Neural Networks. Presented at the ICNN'95 - International Conference on Neural Networks, IEEE, Perth, WA, Australia, pp. 1942–1948. https://doi.org/10.1109/ICNN.1995.488968

Konak, A., Coit, D.W., Smith, A.E., 2006. Multi-objective optimization using genetic algorithms: A tutorial. Reliability Engineering and System Safety 91, 992–1007. https://doi.org/10.1016/j.ress.2005.11.018

Lewis, F.L., Vrabie, D., Vamvoudakis, K.G., 2012a. Reinforcement Learning and Feedback Control: Using Natural Decision Methods to Design Optimal Adaptive Controllers. IEEE Control Syst. 32, 76–105. https://doi.org/10.1109/MCS.2012.2214134

Lewis, F.L., Vrabie, D.L., Syrmos, V.L., 2012b. Optimal control, 3rd ed. Wiley, Hoboken, NJ.

Limon, D., Pereira, M., Muñoz de la Peña, D., Alamo, T., Grosso, J.M., 2014. Single-layer economic model predictive control for periodic operation. Journal of Process Control 24, 1207–1224. https://doi.org/10.1016/j.jprocont.2014.03.013

Logist, F., Houska, B., Diehl, M., Impe, J.V., 2010. Fast Pareto set generation for nonlinear optimal control problems with multiple objectives. Structural and Multidisciplinary Optimization 42, 591–603. https://doi.org/10.1007/s00158-010-0506-x

Luo, Y., Kurian, V., Ogunnaike, B.A., 2021. Bioprocess systems analysis, modeling, estimation, and control. Current Opinion in Chemical Engineering 33, 100705. https://doi.org/10.1016/j.coche.2021.100705

Marlin, T.E., 1995. Process control: designing processes and control systems for dynamic performance : solutions manual to accompany. McGraw-Hill, New York.

Mauky, E., Weinrich, S., Nägele, H.-J., Jacobi, H.F., Liebetrau, J., Nelles, M., 2016. Model Predictive Control for Demand-Driven Biogas Production in Full Scale. Chem. Eng. Technol. 39, 652–664. https://doi.org/10.1002/ceat.201500412

Mitsos, A., Asprion, N., Floudas, C.A., Bortz, M., Baldea, M., Bonvin, D., Caspari, A., Schäfer, P., 2018. Challenges in process optimization for new feedstocks and energy sources. Computers & Chemical Engineering 113, 209–221. https://doi.org/10.1016/j.compchemeng.2018.03.013

Nejjari, F., Puig, V., Quevedo, J., Pascual, J., de Campos, S., 2017. Economic Model Predictive Control of Aeration Systems in a Full Scale Biological Wastewater Treatment Plant, in: Computer Aided Chemical Engineering. Elsevier, pp. 1579–1584. https://doi.org/10.1016/B978-0-444-63965-3.50265-8

Petre, E., Selişteanu, D., Şendrescu, D., 2013. Adaptive and robust-adaptive control strategies for anaerobic wastewater treatment bioprocesses. Chemical Engineering Journal 217, 363–378. https://doi.org/10.1016/j.cej.2012.11.129

Pind, Peter F., Angelidaki, Irini, Ahring, Birgitte K., Stamatelatou, Katerina, Lyberatos, Gerasimos, 2003. Monitoring and Control of Anaerobic Reactors. Biomethanation II, Advances in Biochemical Engineering/Biotechnology 82, 135–182. https://doi.org/10.1007/3-540-45838-7_4

Rathore, A.S., Mishra, S., Nikita, S., Priyanka, P., 2021. Bioprocess Control: Current Progress and Future Perspectives. Life 11, 557. https://doi.org/10.3390/life11060557

Rawlings, J.B., Angeli, D., Bates, C.N., 2012. Fundamentals of economic model predictive control, in: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC). Presented at the 2012 IEEE 51st Annual Conference on Decision and Control (CDC), IEEE, Maui, HI, USA, pp. 3851–3861. https://doi.org/10.1109/CDC.2012.6425822

Renard, P., Dochain, D., Bastin, G., Naveau, H., Nyns, E.-J., 1988. Adaptive control of anaerobic digestion processes?a pilot-scale application. Biotechnol. Bioeng. 31, 287–294. https://doi.org/10.1002/bit.260310402

Reyes-Sierra, M., Coello, C.A.C., 2006. Multi-objective particle swarm optimizers: A survey of the state-of-the-art. International Journal of Computational Intelligence Research 2, 287–308. https://doi.org/10.5019/j.ijcir.2006.68

Rossiter, J.A., 2017. Model-Based Predictive Control. CRC Press. https://doi.org/10.1201/9781315272610

Santander, O., Elkamel, A., Budman, H., 2016. Economic model predictive control of chemical processes with parameter uncertainty. Computers & Chemical Engineering 95, 10–20. https://doi.org/10.1016/j.compchemeng.2016.08.010

Sastry, S., Bodson, M., 1989. Adaptive control: stability, convergence, and robustness. Dover Publications, Mineola, N.Y.

Sawaragi, Y., Nakayama, H., Tanino, Tetsuzō., 1985. Theory of multiobjective optimization. Academic Press, Orlando.

Schmitt, J., Downey, B., Beller, J., Russell, B., Quach, A., Lyon, D., Curran, M., Mulukutla, B.C., Chu, C., 2019. Forecasting and control of lactate bifurcation in Chinese hamster ovary cell culture processes. Biotechnology and Bioengineering 116, 2223–2235. https://doi.org/10.1002/bit.27015

Schwenzer, M., Ay, M., Bergs, T., Abel, D., 2021. Review on model predictive control: an engineering perspective. Int J Adv Manuf Technol 117, 1327–1349. https://doi.org/10.1007/s00170-021-07682-3

Smets, I.Y., Claes, J.E., November, E.J., Bastin, G.P., Impe, J.F.V., 2004. Optimal adaptive control of (bio)chemical reactors: past, present and future. Journal of Process Control 14, 795–805. https://doi.org/10.1016/j.jprocont.2003.12.005

T. Nguyen, 2014. Bi-objective optimal control modification adaptive control for systems with input uncertainty. IEEE/CAA J. Autom. Sinica 1, 423–434. https://doi.org/10.1109/JAS.2014.7004669

Touretzky, C.R., Baldea, M., 2014. Integrating scheduling and control for economic MPC of buildings with energy storage. Journal of Process Control 24, 1292–1300. https://doi.org/10.1016/j.jprocont.2014.04.015

Ulonska, S., Kager, J., Herwig, C., 2018. Automatic controller failure detection with application in model based control of an E. coli fed-batch, in: Computer Aided Chemical Engineering. Elsevier, pp. 1673–1678. https://doi.org/10.1016/B978-0-444-64235-6.50291-6

Van Impe, J.F., Bastin, G., 1995. Optimal Adaptive Control of Fed-Batch Fermentation Processes. Advanced Instrumentation, Data Interpretation, and Control of Biotechnological Processes 401–435. https://doi.org/10.1007/978-94-015-9111-9_13

Van Impe, J.F.M., 1996. Power and limitations of model based bioprocess optimization. Mathematics and Computers in Simulation 42, 159–169. https://doi.org/10.1016/0378-4754(95)00128-X

Vrabie, D.L., Vamvoudakis, K.G., Lewis, F.L., 2013. Optimal adaptive control and differential games by reinforcement learning principles, IET control engineering series. Inst. of Engineering and Technology, Stevenage.

Yamashita, A.S., Zanin, A.C., Odloak, D., 2016. Tuning of Model Predictive Control with Multi-objective Optimization. Brazilian Journal of Chemical Engineering 33, 333–346. https://doi.org/10.1590/0104-6632.20160332s20140212

Zabczyk, J., 2008. Mathematical Control Theory. Birkhäuser Boston, Boston, MA. https://doi.org/10.1007/978-0-8176-4733-9

Zhang, J., Liu, S., Liu, J., 2014. Economic model predictive control with triggered evaluations: State and output feedback. Journal of Process Control 24, 1197–1206. https://doi.org/10.1016/j.jprocont.2014.03.009

Section 4 Soft Sensors and Fault Detection

Summary of Section 4

This section briefly introduces data-driven soft sensors as a valuable tool for monitoring, control, and optimization tasks in biological processes. Our attention is specially focused on soft sensors based on machine learning algorithms to process monitoring and fault detection. This section aims to present several supervised machine learning algorithms with particular detail on the Support Vector Machine (SVM) used in this thesis to process fault detection and classification. We aim to highlight that data-driven soft sensors have been widely applied to biological processes such as AD. However, from our knowledge, we have not found an application for biological methanation.

At the industrial level, physical sensors have been applied for monitoring, control, and optimization tasks. These sensors have allowed the acquisition of valuable information in the process. Regardless of the type of process to be developed, several types of sensors (*e.g.*, temperature, pH, pressure, and flow rates sensors (Kazemi *et al.*, 2020b) will always be necessary to correct processes during monitoring. Physical sensors have limitations, such as high purchase, installation, and maintenance costs. For example, if the maintenance is not performed periodically, the sensor could fault due to the constant exposition to extreme conditions inside the reactor (Doraiswami and Cheded, 2014). In recent years, soft sensors have become an alternative to the monitory process in modern industry (Yan *et al.*, 2021).

Soft sensors are mathematical models that estimate a hard-to-measure property using relatively easy measurements (Kazemi *et al.*, 2021; Sharma and Tambe, 2014; Zhu *et al.*, 2020). In our context, soft sensors are related with two words, "software" since the models are computer programs and "sensor" because these models provide similar information to the physical sensors (Kadlec *et al.*, 2009; Zhu *et al.*, 2020).

According to Fortuna *et al.* (2007) and Jiang *et al.* (2021), the use of soft sensors presents several advantages:

- Soft sensors are low-cost alternatives to hardware devices (physical sensors)
- They are not subject to physical constraints such as space installation and extreme working conditions
- This type of sensor can be implemented on existing hardware (embedded systems)
- Soft and physical sensors work in parallel, which allows them to obtain valuable information for fault detection tasks
- It allows the real-time estimation of data, improving the performance of the control strategies

Soft sensors can be classified in three standard classes: model-driven soft sensors, data-driven soft sensors, and hybrid soft sensors (Kadlec and Gabrys, 2009; Yan *et al.*, 2021).

Model-driven soft sensors (white-box models) can be subdivided into first principle models, Kalman filters, and adaptative observers. Model-driven soft sensors are based on the knowledge

of process phenomenology, *i.e.*, the use of mass and energy balances and constitutive physical-chemical equations to describe reaction kinetics and mass transfer in the process (Fortuna *et al.*, 2007). One of the limitations of this approach is that it might require an in-depth knowledge of the process. **Data-driven soft sensors** (black-box models) used the recorded, stored, and provided historical data of the process, describing the real process conditions and empirical observations. These models are considered more realistic than model-driven since they exploit real process information (Gopakumar *et al.*, 2018). **Hybrid soft sensors** (grey-box models) combine model-driven and data-driven soft sensors. Figure 1.4.1 summarizes the type of soft sensors and their applications (Kazemi *et al.*, 2020b; Wade, 2020).

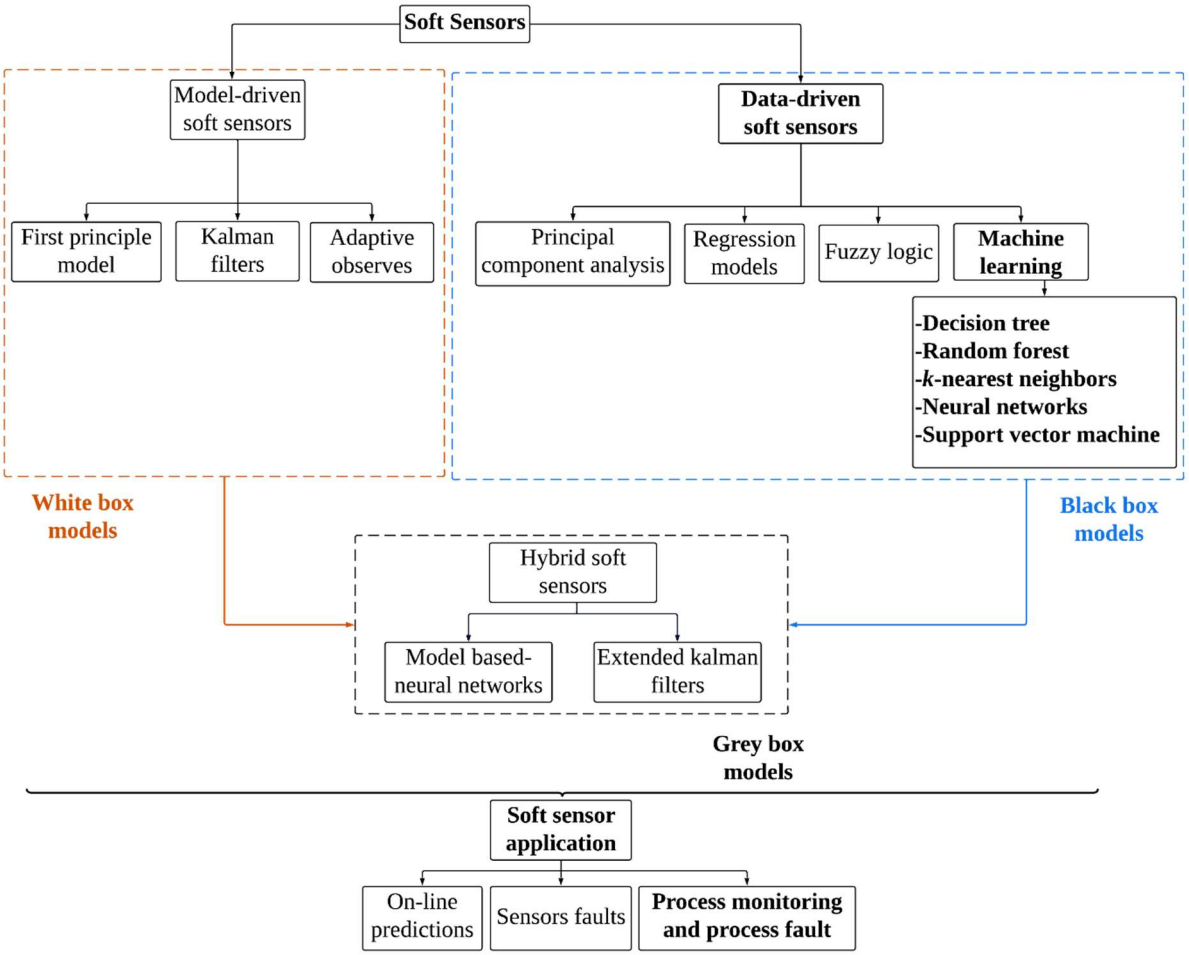


Figure 1.4.1. Soft sensor categories adapted from (Kadlec *et al.*, 2009; Kadlec and Gabrys, 2009; Jiang *et al.*, 2021).

Soft sensors present three main applications in the industry: on-line predictions, process monitoring, and process and sensor fault detection (Kadlec *et al.*, 2009; Sharma and Tambe, 2014). Soft sensors have been developed to estimate variables that cannot be measured directly through automated systems, *e.g.*, temperature, pH, or flow rates can be measured on-line. In contrast, some variables, such as concentrations, in particular cell concentrations, require off-line quantification using complex methodologies, and these concentrations are frequently related to process performance, yields or productivity. Sensor fault detection is another area explored with soft sensors. Many sensors could have faults during measurements of some critical variables. Soft sensors can double-check the system behavior and detect these faults (Fortuna *et al.*, 2007; Kadlec *et al.*, 2009). Another application of soft sensors is process monitoring and process fault detection. It is important to highlight that sensor and process fault detection are two different applications. In the framework of this thesis, we mainly work with process fault detection. These soft sensors can be trained to describe the expected process performance or to recognize possible fault detection, *i.e.*, deviations from the target trajectory (Kadlec *et al.*, 2011; Kazemi *et al.*, 2021).

Nowadays, the most popular data-driven techniques are the principle component analysis, regression models, fuzzy logic, and techniques from machine learning theory such as neural networks and support vector machines (Kadlec and Gabrys, 2009; Gopakumar *et al.*, 2018; Brunner *et al.*, 2021). Support vector machine (SVM) has gained acceptance in constructing data-driven soft sensors (Sbárbaro and del Villar, 2010; Sharma and Tambe, 2014) due to the theoretical background in the statistical learning theory, the simplicity in implementation, the capacity to work with high-dimensional and small datasets, and efficiency in avoiding the local minima and overlapping classes (overfitting) (Kadlec *et al.*, 2011; Yan *et al.*, 2021).

This thesis focuses on using data-driven soft sensors based on machine learning techniques with particular attention to support vector machines. Therefore, the following subsections present a definition of machine learning algorithms, their categories, a detailed explanation of support vector machines, and a brief explanation of other algorithms used in this work.

4.1 Machine Learning Algorithms Development

Machine learning involves designing algorithms that automatically detect and extract recurring patterns in a dataset (Deisenroth *et al.*, 2020; Muller and Guido, 2016) that can be numerical, textual, or visual (Yuxi (Hayden) Liu, 2020). Detecting these patterns leads the algorithms to learn, improve their accuracy, and make predictions on new input data.

There are four main aspects to consider in machine learning algorithms, as presented in Figure 1.4.2: dataset, model, training, and validation.

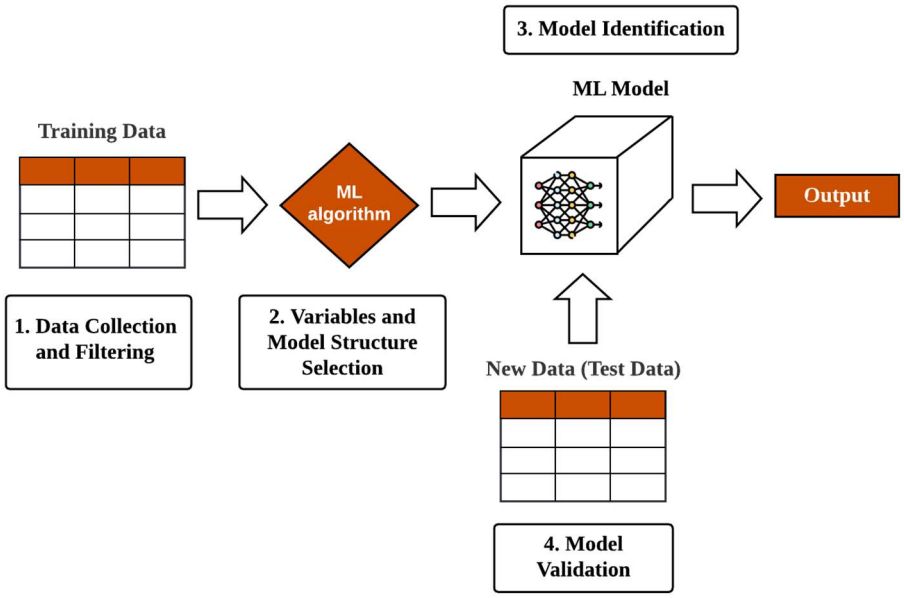


Figure 1.4.2. Steps in the development of Machine Learning algorithms. Illustration adapted from the webpage Techvidvan, June 2023.

Data collection and filtering imply the selection and preparation of the training dataset. The dataset must be carefully prepared, organized, and cleaned. Otherwise, the training of the machine learning model may be biased, and the results of the future predictions will be directly affected.

Variables and model structure selection entail the selection of the algorithm for training the model. The type of algorithm used depends on the nature, volume of training data, and the problem to be solved.

Model identification involves training and generating the best model. It is an iterative process in which the model is trained using the dataset. Then the model results are compared with the expected values using some statistical criteria, *e.g.*, Akaike, Bayes, *etc*. The weights and biases are adjusted to improve the accuracy of the result. Finally, the **trained model is validated** with new data. The origin of these data depends on the problem to be solved (Fortuna *et al.*, 2007). Generally, the dataset is partitioned 70/30 or 80/20, *i.e.*, 70 or 80% of the dataset is derived for training the algorithms, and the remaining 30 or 20% to evaluate the model's performance.

4.1.1 General Machine Learning Categories

There are three main categories in which machine learning can be formulated according to the type of learning: unsupervised, reinforcement, and supervised learning.

In unsupervised learning, the dataset to train the algorithms only contains indicative signals without any description or output assigned (unlabeled data). This type of machine learning aims to find similar characteristics in the instances included in the dataset and group them, followed by interpretation of the results (Watson, 2023).

In reinforcement learning, some supervised learning exists, but not in the usual form where each dataset output corresponds to an input. Reinforcement learning receives feedback after selecting an output for a given input or observation, and the system evaluates its performance based on that feedback response and reacts accordingly. The feedback indicates how the output (action in reinforcement learning) fulfills the learner's goals (Yuxi (Hayden) Liu, 2020).

In supervised learning, the dataset to train the algorithms contains pairs of input, a description, classes, or desired outputs besides indicative signals. The learning goal is to find a general rule that maps input to output (Simeone, 2018; Yuxi (Hayden) Liu, 2020). Figure 1.4.3 presents the three types of machine learning.

This thesis will focus on using supervised Machine Learning Algorithms, emphasizing classification. The following subsections will introduce some of the most popular learning algorithms.

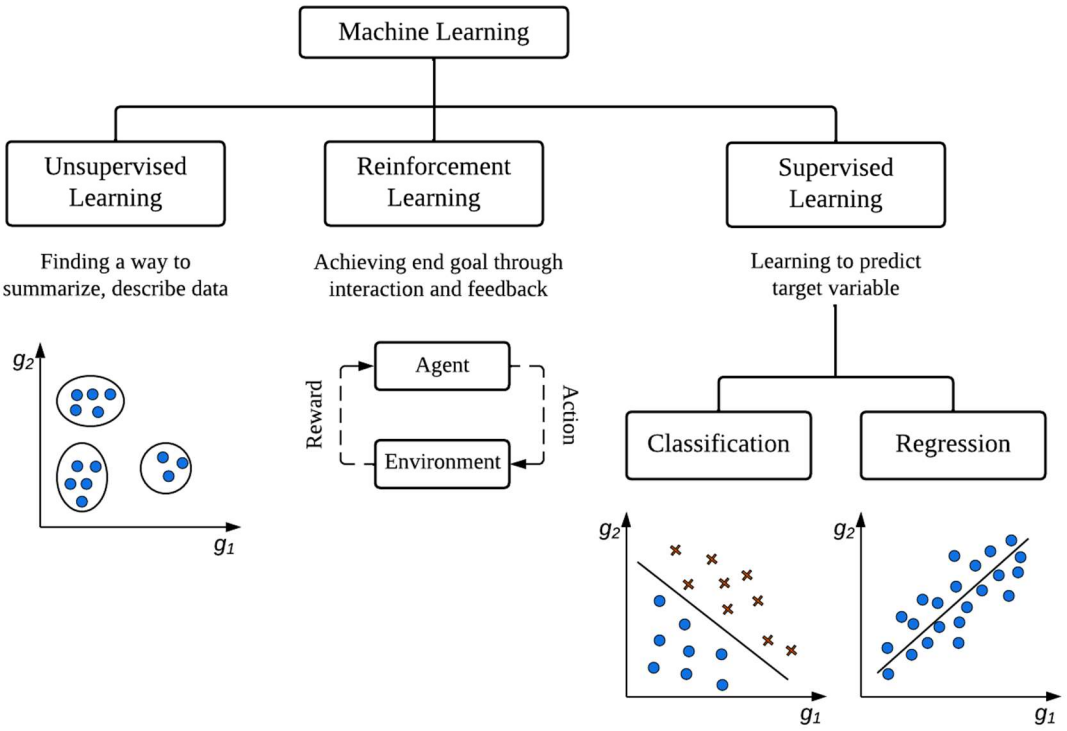


Figure 1.4.3. Categories of machine learning. Adapted from (Yuxi (Hayden) Liu, 2020).

4.1.2 Supervised Machine Learning Algorithms

The main goal of supervised machine algorithms is to build a model capable of predicting the class of a sample given a set of features (Nasteski, 2017).

In the context of this thesis, the machine learning inputs refer to the features and instances. Features correspond to the state variables of the dynamic models (see section 3 chapter1) and instances correspond to each one of the time samples

Supervised learning is divided into two categories, regression and classification. These algorithms could be explained mathematically as follows:

Considering a dataset $\mathcal{D}$ generated as,

$$\boldsymbol{X} = (\boldsymbol{x_i}, \boldsymbol{g_i}) \underset{i.i.d}{\sim} p(x, g) \qquad i = 1, \ldots, N \tag{1.4.1}$$

where $\boldsymbol{X}$ is the input space and $p(x, g)$ is the true joint distribution. Each sample pair (x_i, g_i) of $\boldsymbol{X}$ is generated using the same p function, which means that all the instances are independently identically distributed (*i.i.d.*).

$\mathcal{D}$ can be rewritten as,

$$\mathcal{D} = \{\boldsymbol{x_i}, \boldsymbol{g_i}\} \quad (1.4.2)$$

with $\boldsymbol{x_i} \in \mathcal{R}^{\mathrm{N}\times\mathrm{p}}$ consists of an input vector with N instances and p Features and the associated response ($\boldsymbol{g_i} \in \mathcal{R}^{\mathrm{N}\times 1}$).

The aim is to obtain a model $s(x)$ that generalizes the input-output mapping (Φ) in $\mathcal{D}$ to unknown inputs x. The quality of the prediction $s(x)$ for a test pair (x, g) is measured by a given loss function as $\ell(g, s(x))$.

The outputs g are discrete variables that take a finite number of possible values in classification algorithms, i.e., to predict a class label from a predefined list of possibilities (Muller and Guido, 2016; Yuxi (Hayden) Liu, 2020). The objective is to present an unclassified input (instances + features), evaluate it in the trained model and have the model perform a binary classification (● and ✕). Figure 1.4.4 illustrates the binary case, where the goal is to predict between two classes (● and ✕) from an unclassified input x. The most straightforward case to understand the binary classification could be defined as,

$$\ell(g, s(x)) = \begin{cases} +1 & if\ g\ = s(x) \\ -1 & if\ g\ \neq s(x) \end{cases} \quad (1.4.3)$$

where $+1$ and -1 correspond to correct and incorrect classification, respectively.

The most common supervised machine learning algorithms which differ in the level of interpretation are Decision Trees, Random Forests, k-Nearest Neighbor, Random Forests, Naïve Bayes, Quadratic Discriminant Analysis, Support Vector Machines, and Neural Networks (Fawagreh *et al.*, 2014; Muller and Guido, 2016; Rokach and Maimon, 2005; Vapnik *et al.*, 1997; Wu *et al.*, 2008).

4.1.3 Machine Learning Algorithms

4.1.3.1 Support Vector Machine

One of the most robust and accurate machine learning algorithms used in literature is the Support Vector Machine (SVM). It works well in high-dimensional spaces when there is a clear margin of separation between classes. It has been successfully applied in several scientific and engineering areas (Batuwita and Palade, 2013; Cervantes *et al.*, 2020; López Cabrera and Pereira-Toledo, 2018; Panup *et al.*, 2022; Wu *et al.*, 2008; Xiao *et al.*, 2022)

The idea of SVM relies on mapping the input data or features $x_i \in \mathcal{R}^{N \times p}$ into a nonlinear space in order to predict a desired vector of outputs $g_i \in \mathcal{R}^{N \times 1}$ (Zhu *et al.*, 2020) (Figure 1.4.4). It solves pattern recognition (classification) and regression problems (Cristianini and Shawe-Taylor, 2000). This section presents an SVM with two classes as an example. However, these algorithms could be applied to multiple classes.

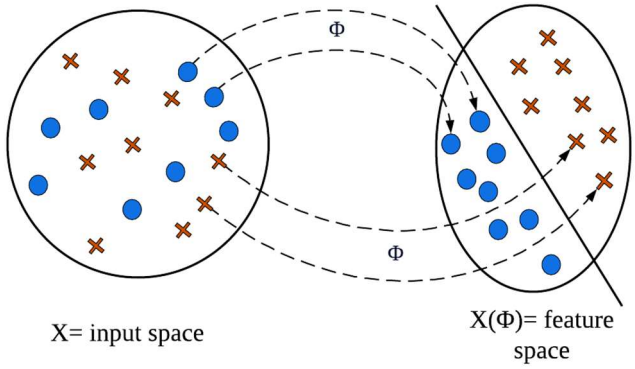


Figure 1.4.4. SVM basis idea with two classes. ***X***: Input space associated with labels (● and ✕). ***Φ***: transformations based on the hypothesis space of linear functions.

The algorithms are trained with a dataset $\mathcal{D}$ described in Equation (1.4.2). Usually, the data are in the form of attribute vectors or matrices. Therefore, the input space is a subset $\epsilon \mathcal{R}^{N \times p}$. Once the attribute vectors are available, a number of sets of hypotheses could be chosen for the problem to predict desired outputs (Cristianini and Shawe-Taylor, 2000).

The SVM Optimization Problem

SVM algorithms need to solve an optimization problem, *i.e.*, maximizing a particular mathematical function concerning a given $\mathcal{D}$. This mathematical function is commonly referred to as a discriminant function. It establishes the rules for determining the class label of unknown inputs (x) (Tharwat, 2016).

Defining the optimization problem requires knowing its nature:

- Linear discriminant, in which the cases linearly separable or non-separable occurs, or
- Kernel type, in which the non-linear discriminant is possible.

The two cases are approached through four concepts: (i) the separating hyperplane, (ii) the maximum-margin hyperplane, (iii) the soft margin, and (iv) the kernel function (Noble, 2006).

(i) The Separating Hyperplane

Let us consider again a dataset $\mathcal{D} = \{\boldsymbol{x_i}, \boldsymbol{g_i}\}$ defined in Equation (1.4.2), where each point $\boldsymbol{x_i} \in \mathcal{R}^{\mathrm{N}\times\mathrm{p}}$ consists of an input vector and the associated response ($\boldsymbol{g_i} \in \mathcal{R}^{\mathrm{N}\times 1}$). Each one with features and whose response variable has two levels (for example, +1 and -1). Hyperplanes could be used to build a classifier that allows predicting to which class a sample belongs based on its features. The point falls on one side or the other of the hyperplane. Thus, a hyperplane can be understood to divide a *p*-dimensional space into two halves.

The simplest case corresponds to the linearly separable case (Cervantes *et al.*, 2020), where points of two classes (+1 and -1) are perfectly separable by various hyperplanes. In two dimensions (Figure 1.4.5-A), a straight line divides the space in half; in three dimensions, a plane divides the space (Figure 1.4.5-B).

(ii) The Maximum-Margin Hyperplane

It is necessary to calculate the sign of the discriminant function to find out which side of the hyperplane a given point falls, which is linear if it can be written as Equation (1.4.4).

$$s(x) = W^T x + W_0 \quad if \begin{cases} s(x) > 0 \rightarrow x \in \text{class 1 } (+1) \\ s(x) < 0 \rightarrow x \in \text{class 2 } (-1) \end{cases} \qquad (1.4.4)$$

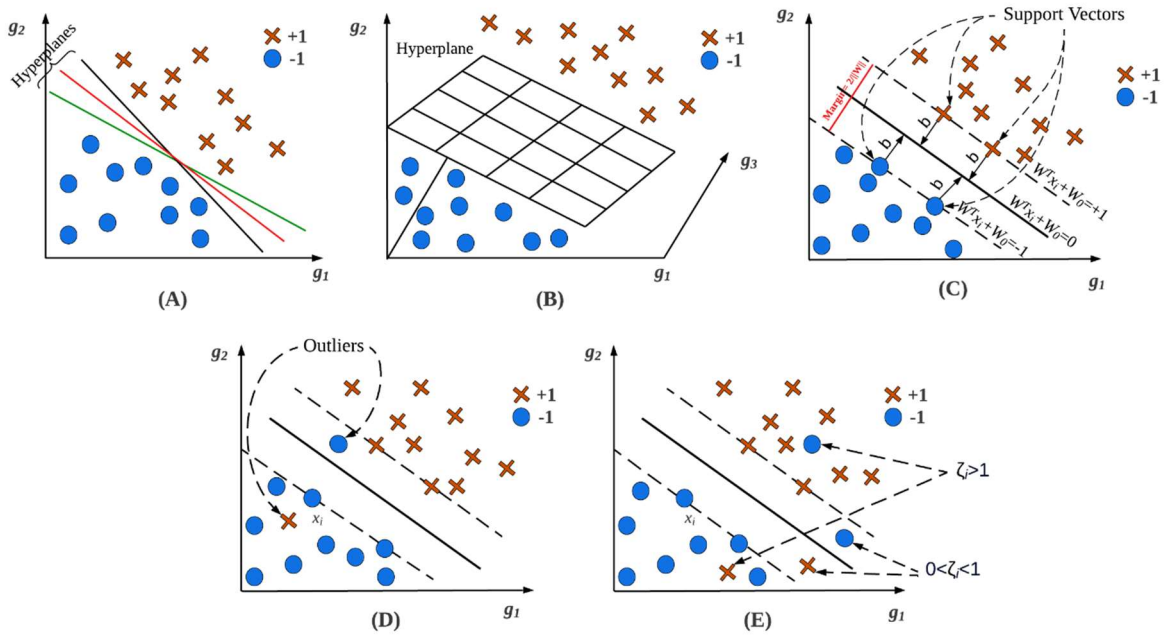


Figure 1.4.5. Support vector machine (SVM) graphical representation. (A) linearly separable case with two labels, (B) optimal hyperplane defined by support vectors, (C) Support vectors definition (Margin hyperplane), (D) data defined on the wrong side of the hyperplane (outliers), (E) deviation of ideal data (slack variables).

The definition of hyperplane for perfectly linearly separable cases results in infinite possibilities, lines red, green, and black in Figure 1.4.5-A. SVM tries to find the maximal margin separation between the hyperplane and the data, *i.e.*, to find the hyperplane furthest from all the data; it is called the maximal margin hyperplane or optimal separation hyperplane (Cervantes *et al.*, 2020; Deisenroth *et al.*, 2020; Kowalczyk, 2017; Panup *et al.*, 2022). To do that, it is necessary to calculate the perpendicular distance of each point to a given hyperplane. The smallest of these distances (known as the margin) (Cervantes *et al.*, 2020; Noble, 2006) determines how far the hyperplane is from the training data (x_i); it means that the optimal hyperplane distance to the closest negative data is equal to the distance to the nearest positive data.

The SVM optimization problem is reduced to maximize the margin by determining the support vectors, which are the data closest to the separating hyperplane. They are the most complex patterns to classify. Support vectors completely define the optimal hyperplane.

The SVM finds the hyperplane (continuous line) with the broadest margin $2b$ in Figure 1.4.5-C. The points inside the margin's edge (discontinuous lines) are called support vectors. Margin is twice the absolute distance b from the closest data to the separating hyperplane, and it can be calculated as Equation (1.4.5). The absolute distance between x_i and the boundary $g(x) = 0$ (Deisenroth *et al.*, 2020; Yuxi (Hayden) Liu, 2020).

$$Margin = \frac{|W^T xi + W_0|}{\|W\|} \quad (1.4.5)$$

For a unique hyperplane $|W^T x_i + W_0| = 1$ (See Figure 1.4.5-C), the distance between any data x_i and the decision hyperplane $s(x) = 0$ is defined as,

$$\frac{|W^T xi + W_0|}{\|W\|} = \frac{1}{\|W\|} \quad (1.4.6)$$

Since the margin is defined as twice the absolute distance from the closest training data to the separating hyperplane (Equation (1.4.5)) and in the case of a unique hyperplane the absolute distance is given by Equation (1.4.6), thus, the margin is,

$$Margin = 2\frac{1}{\|W\|} = \frac{2}{\|W\|} \quad (1.4.7)$$

The goal is to maximize the margin:

$$Subject\ to \begin{cases} W^T x_i + W_0 \geq 1 & if\ x_i, is\ positive\ data \\ W^T x_i + W_0 \leq -1 & if\ x_i, is\ negative\ data \end{cases}$$

Let:

$$\begin{cases} z_i = 1 & if\ x_i, is\ positive\ data \\ z_i = -1 & if\ x_i, is\ negative\ data \end{cases}$$

It is possible to convert the problem as,

$$\min_{W, W_0} J(W) = \frac{1}{2}\|W\|^2 \qquad (1.4.8)$$

$$Subject\ to\ \{z_i(W^T \boldsymbol{x_i} + W_0) \geq +1 \quad \forall i$$

where $J(W)$ is a quadratic function. Thus, there is a single global minimum.

(iii) The Soft Margin

It is impossible to separate all the outputs perfectly in several datasets (Kowalczyk, 2017) (Figure 1.4.5-D). It is possible to find data close to the x_i on the wrong side of the hyperplane. These points are called outliers. However, a linear classifier may still be appropriate (Bzdok *et al.*, 2018).

Applying SVM in non-linearly separable cases is possible to obtain good performance, but the data must be "almost" linearly separable. The use of slack variables $\zeta_1,\ldots, \zeta_n$, one for each data can solve this (Figure 1.4.5-E) and change the constrains:

$$z_i(W^T \boldsymbol{x_i} + W_0) \geq 1 \quad \forall i \quad \text{to} \quad z_i(W^T \boldsymbol{x_i} + W_0) \geq 1 - \xi_i \quad \forall i$$

where ξ_i is a measured of deviation from the ideal for data i, and is classified as:

- If $\zeta_i > 1$ data i is on the wrong side of the hyperplane
- If $0 < \zeta_i < 1$ sample i is on the right side of the separating hyperplane but within the region of the maximum margin
- If $\zeta_i < 0$ is the ideal case for sample i

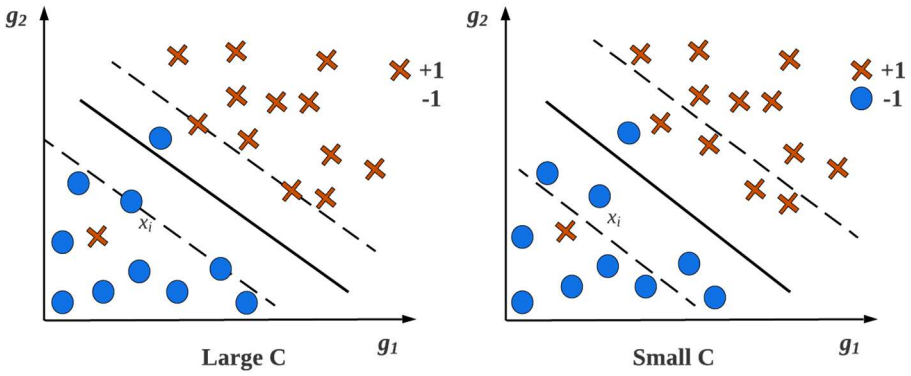


Figure 1.4.6. Hyperplane with large C, few data not in ideal position (left), and small C, several data not in ideal position (right).

Then, Equation (1.4.8) is transformed as,

$$\min_{W,\,W_0,\,\zeta_i} J(\boldsymbol{W},\zeta_1,\ldots,\zeta_n) = \frac{1}{2}\|\boldsymbol{W}\|^2 + C\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) \qquad (1.4.9)$$

$$Subject\ to \begin{cases} z_i - W^T\Phi(\boldsymbol{xi}) - W_0 \le \varepsilon - \xi_i & \forall i \\ W^T\Phi(\boldsymbol{xi}) + W_0 - z_i \le \varepsilon - \xi_i & \forall i \\ \xi_i,\ \xi_i^* \ge 0 \quad \forall i \end{cases}$$

The term $\sum_{i=1}^{n}(\zeta_i + \zeta_i^*)$ corresponds to the number of misclassified data, ε displays the loss function variable, ξ_i and ξ_i^* are slack variables that allow certain points to fall on the incorrect side of the hyperplane, and C is a constant that measures the relative weight of the first and second terms.

If C is small, it allows many instances in a non-ideal position (loose segregation); if C is large, it is possible to have and accept just a few data in a non-ideal position (strict segregation), Figure 1.4.6.

The problem foundation of Equation (1.4.9) relies on convex quadratic programming. The Lagrangian function is utilized to integrate the constrains into the cost function, and the dual representation, may be solved as in Xiao *et al.* (2022), Equation (1.4.10).

$$max\left[-\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\Phi(x_i)\Phi(x_j)+\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)y_i-\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)y_i\varepsilon\right] \quad (1.4.10)$$

$$Subject\ to\ \{\textstyle\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)=0.0 \quad 0.0\le\alpha_i,\alpha_i^*\le C$$

where α_i represent the Lagrangian multiplier.

(iv) The Kernel Function

In some cases, data are not linearly separable. Therefore, the machine learning model obtained could not give correct results even if the hyperplane is optimally determined (Cervantes *et al.*, 2020); an example is presented in Figure 1.4.7-A.

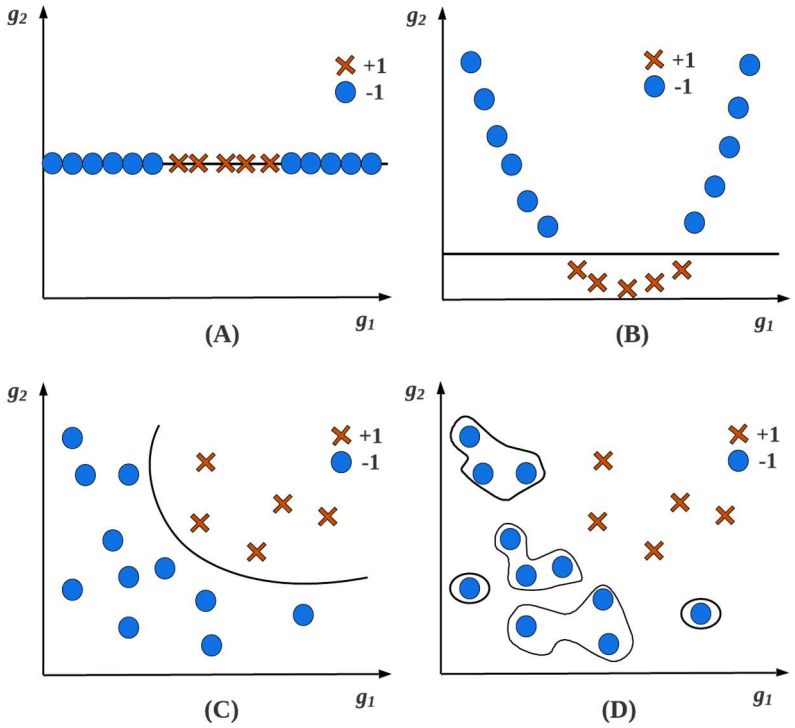


Figure 1.4.7. Graphic representation of the non-linear dataset. Adapted from (Noble, 2006). (A) a non-linear one-dimensional dataset, (B) the dataset was squared, (C) a non-linearly two-dimensional dataset, which is linearly separable in four dimensions, (D) An SVM that has overfit a two-dimensional dataset.

Learning non-linear relations with a linear machine is necessary to select a set of non-linear features and rewrite the data in the new representation, *i.e.*, apply a fixed non-linear mapping

of data to a feature space used by the linear machine. In Figure 1.4.7-B, the data were squared; in Figure 1.4.7-C, a non-linear mapping projects the data from the two-dimensional space to four dimensions (corresponding to the products of all pairs of features), allowing the data to be linearly separated. Finally, Figure 1.4.7-D presents the case where the dataset is projected into a space with too many dimensions; the projected hyperplane comes from an SVM that uses a very high-dimensional kernel function (Noble, 2006).

It is possible to represent this mathematically using Equation (1.4.11).

$$s(x) = \sum_{i=1}^{N} W^T \Phi_i(x) + b \qquad (1.4.11)$$

where $\Phi: X \to F$ is a non-linear map from the input space to some feature space.

Non-linear machines can be built in two steps: first, a fixed non-linear mapping transforms the data into a feature space F, and then a linear machine is used to classify them in the feature space (Cristianini and Shawe-Taylor, 2000), Figure 1.4.8.

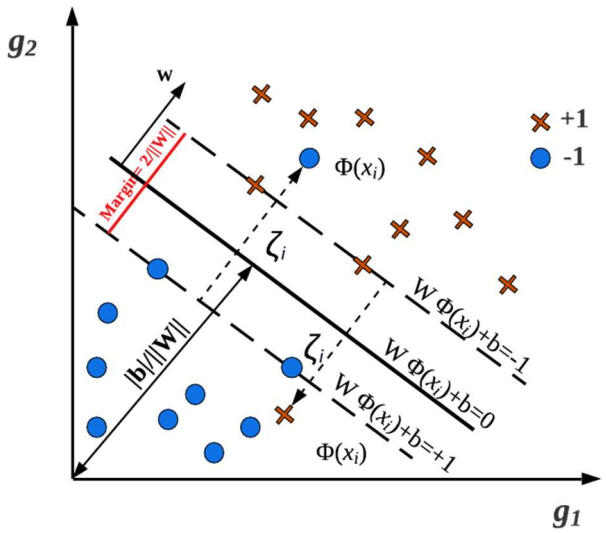


Figure 1.4.8. Support vector machine (SVM) graphical representation for the non-linear case.

One important property of linear learning machines is that they can be expressed in a dual representation. The hypothesis or decision function can be described as a linear combination of

the training data. Therefore, the decision rule can be evaluated using just inner products between the test point and the training data (Cristianini and Shawe-Taylor, 2000):

$$s(x) = \sum_{i=1}^{N} \alpha_i g_i\big(\Phi(x_i) \cdot \Phi(x)\big) + b \qquad (1.4.12)$$

If we have a way of computing the inner product $(\Phi(x_i) \cdot \Phi(x))$ in feature space as a function of the original input points, merging the two steps needed to build a non-linear learning machine becomes possible. This direct computation method is called a kernel function. The kernel function is a mathematical transformation that allows to project data from a low-dimensional space to a higher dimension space.

Mathematically a kernel is a function K, such that for all $x, x_i \in X$, Equation (1.4.13).

$$K(x, x_i) = \big(\Phi(x) \cdot \Phi(x_i)\big) \qquad (1.4.13)$$

where Φ is a mapping from X to an (inner product) feature space F.

The name kernel is derived from integral operator theory, which supports much of the theory of the relation between kernels and their corresponding feature spaces. An essential consequence of the dual representation is that the dimension of the feature space does not have to affect the calculation. Since the feature vectors are not represented explicitly, the number of operations required to compute the inner product by evaluating the kernel function is not necessarily proportional to the number of features. The only information used about the training data is their matrix Gram (K) in the feature space, it is also called the kernel matrix. The key to this approach is finding a kernel function that can be evaluated efficiently. Then, this function is evaluated by at least l evaluations as the decision rule, Equation (1.4.14).

$$s(x) = \sum_{i=1}^{l} \alpha_i K(x, x_i) + b \qquad (1.4.14)$$

With the use of kernels, it is possible to operate in the feature space without knowing the details of how the input data was transformed.

Some of the Kernel functions that will be used in this work are represented by Equations (1.4.15)-(1.4.17) (Cervantes *et al.*, 2020):

Linear Kernel :

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \boldsymbol{x_i} \cdot \boldsymbol{x_j} \qquad (1.4.15)$$

Polynomic Kernel :

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i} \cdot \boldsymbol{x_j} + c)^d \qquad (1.4.16)$$

with $c = 0$ and $d = 1$ it is a linear kernel.

Gaussian Kernel or Radial Basis Function (RBF):

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = e^{\frac{-\|x - x_i\|^2}{2\sigma^2}} \qquad (1.4.17)$$

The value of $1/2\sigma^2$ controls the Kernel's behavior; when it is very small, the final model is equivalent to that obtained with a linear kernel as its value increases (flexibility of the model). σ represents the width of the RBF.

4.1.3.2 Other Machine Learning Algorithms

Decision trees

The decision trees (DT) algorithm follows a methodology where the classification process is performed using a hierarchical decision on the feature's variables, similar to a tree structure. Each decision node corresponds to a feature test, which is referred to as the split, and each leaf node refers to the attributes. DT algorithms are generally a recursive process, *i.e.*, a sequence of splits is performed from the top (root node) to the bottom (leaf nodes) over a dataset. Each decision node corresponds to a split of the dataset into subsets, where each subset will be used as the dataset of the next decision node. The challenge with DT is how to perform the partition.

Figure 1.4.9 shows a common structure for a DT (Aggarwal, 2015; Rokach and Maimon, 2005; Uddin *et al.*, 2019; Zhou, 2012).

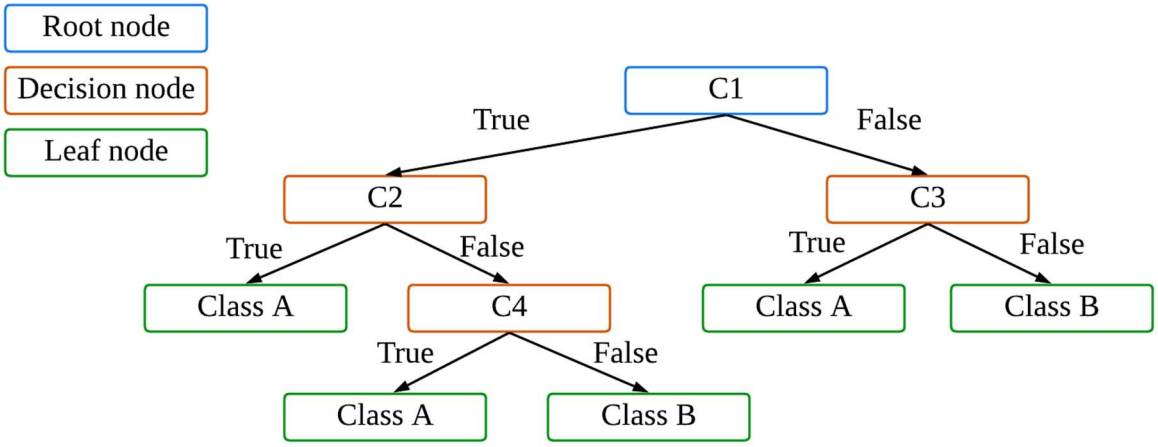


Figure 1.4.9. Decision tree algorithm scheme. Root node (C1), decision nodes (C2, C4, and C3), and leaf nodes (class A and B).

Depending on the information gain criterion used to perform the partition, *i.e.*, based on the measure used to evaluate the quality of a split at a particular node in the tree; The decision trees can be classified into (i) ID3, (ii) C4.5, and (iii) CART (Charbuty and Abdulazeez, 2021; Rokach and Maimon, 2005):

(i) The ID3 algorithm uses as information gain criterion the entropy $\left(E(\mathcal{D})\right)$ of the dataset $\mathcal{D}$, defined as,

$$E(\mathcal{D}) = -\sum_{g} P(g|\mathcal{D}) log P(g|\mathcal{D}) \qquad (1.4.18)$$

herein, $g \in \{g_1, g_2, \ldots, g_h\}$ is the vector of possible classes. $P(g|\mathcal{D})$ holds for the probability of instances in $\mathcal{D}$ that belong to class g.

The training set $\mathcal{D}$ is partitioned into subsets $\mathcal{D}_1$, $\mathcal{D}_2$, …, $\mathcal{D}_k$, with weights $|\mathcal{D}_k|$, the overall entropy is calculated and compared with the weighted average of the subset entropies; The amount of the reduction is the information gain (G), Equation (1.4.19).

$$G(\mathcal{D};\ \mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_k) = E(\mathcal{D}) - \sum_{i=1}^{k} \frac{|\mathcal{D}_k|}{|\mathcal{D}|} E(\mathcal{D}_k) \quad (1.4.19)$$

The split is performed based on the feature-value pair, which causes the largest information gain. The entropy measures the uncertainty of the information in the dataset. A high value of entropy means high uncertainty and more information is necessary to develop the model (Charbuty and Abdulazeez, 2021). The disadvantage of using the information gain criterion is that features with a large number of instances will be encouraged, regardless of their relevance for classification (Priyanka and Kumar, 2020).

(ii) The C 4.5 algorithm addresses the problem mentioned for the ID3, using a variant of the information gain criterion by performing the gain ratio, Equation (1.4.20).

$$P(\mathcal{D};\ \mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_k) = G(\mathcal{D};\ \mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_k) \cdot \left(- \sum_{i=1}^{k} \frac{|\mathcal{D}_k|}{|\mathcal{D}|}\ log \frac{|\mathcal{D}_k|}{|\mathcal{D}|} \right)^{-1} \quad (1.4.20)$$

This variant of the information gain criterion normalizes the number of features values, using a correction factor that penalizes the number of subsets k and the size of each subset $|\mathcal{D}_k|$. The feature with the highest gain ratio, among the features with better than average information gains is selected for the split.

(iii) In the CART algorithm, the information gain criterion corresponds with the overall Gini index $(Gini)$. For a k subset into $\mathcal{D}$, the $Gini$ can be calculated as the weighted average of the $Gini$ index values of each $\mathcal{D}_k$, where the weight of $\mathcal{D}_k$ is $|\mathcal{D}_k|$.

$$Gini(\mathcal{D};\ \mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_k) = I(\mathcal{D}) - \sum_{i=1}^{k} \frac{|\mathcal{D}_k|}{|\mathcal{D}|}\ I(\mathcal{D}_k) \quad (1.4.21)$$

where $I(\mathcal{D})$ is the dataset purity,

$$I(\mathcal{D}) = 1 - \sum_{g} P(g|\mathcal{D})^2 \quad (1.4.22)$$

CART algorithm presents some advantages over the other trees: the algorithm itself identifies the most significant variables and eliminates the non-significant ones. Furthermore, it is non-parametric and can easily handle outliers (Priyanka and Kumar, 2020; Singh and Giri, 2014).

An aspect to consider in the use of DT is overfitting. DT with a perfect training result may have a poor ability to generalize concerning other DT with an acceptable training result, *i.e.*, perfect training results do not mean perfect test results. The cause of this is the noise in the dataset collection. Pruning is used to reduce the risk of overfitting, *i.e.*, cutting off some branches of the tree caused by this noise. There are two options for pruning, pre-pruning and post-pruning. The first tries to prune the branches when the tree is grown, while the second re-examines fully grown trees to decide which branches should be removed (Zhou, 2012).

Random Forest

Random Forest (RF) is a classifier defined as an ensemble classifier. RF uses several basic DT to build a forest (Biau and Scornet, 2016) (Figure 1.4.10).

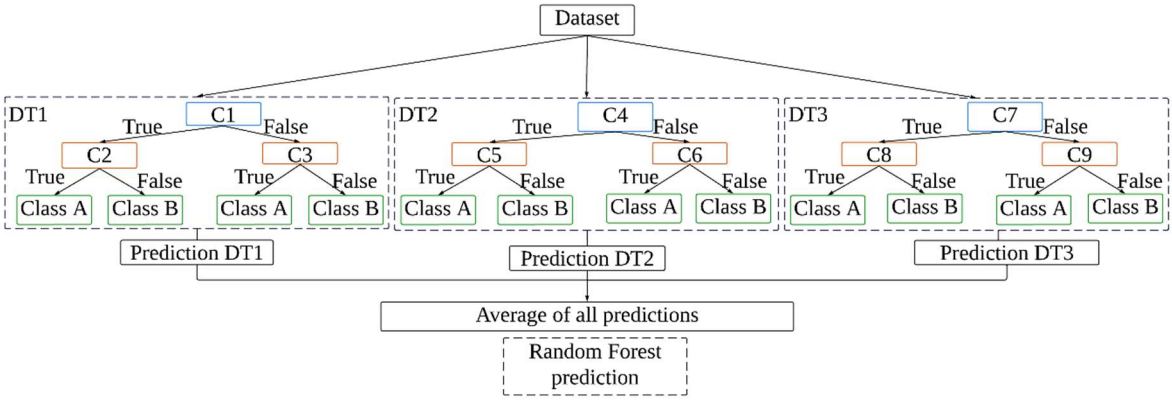


Figure 1.4.10. Random Forest scheme with three different decision trees. Root node (C1, C4, C7), decision nodes (C2, C3, C5, C6, C8, C9), and leaf nodes (class A and B).

The DT in the RF are trained with different parts of the training dataset. To classify a new sample (x, g) the inputs x of that sample must pass through each DT, which considers a different part of x to give a classification. Finally, the RF selects the average prediction of these DT, *i.e.*, the classification with more votes (Uddin *et al.*, 2019).

Naïve Bayes

Naïve Bayes (NB) algorithm predicts a sample's probability of belonging to a specific class. Consider again the dataset ($\mathcal{D} = \{\boldsymbol{x_i}, \boldsymbol{g_i}\}$) defined in Equation (1.4.2). NB calculates a posterior class probability to a sample using the Bayes theorem (Mehmood *et al.*, 2018),

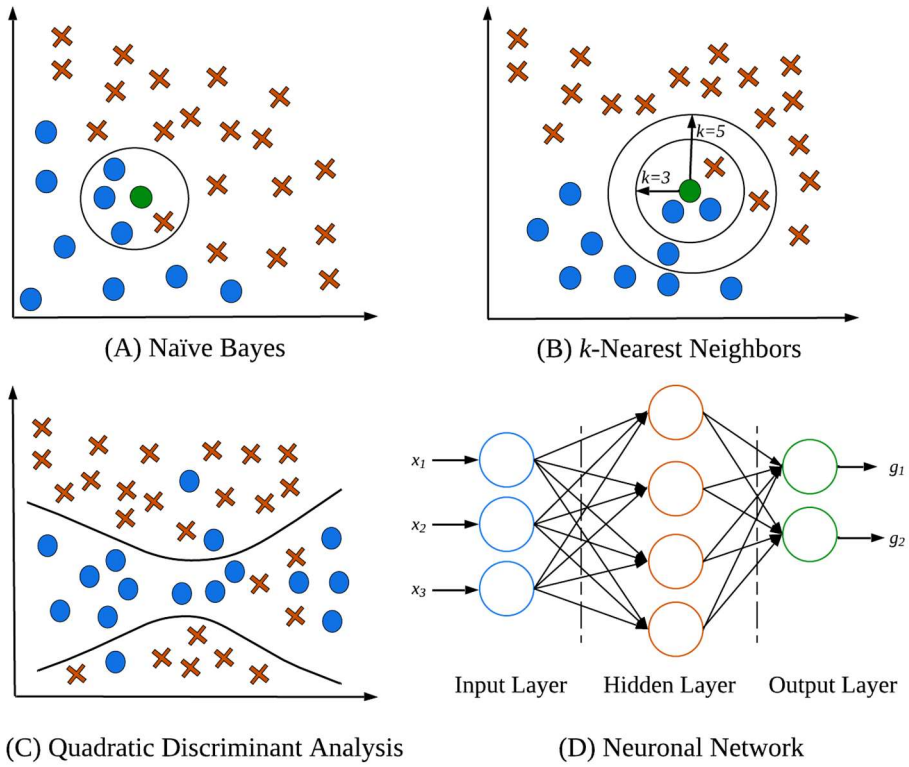
$$P(g|\boldsymbol{x_i}) = \frac{P(\boldsymbol{x_i}|g)P(g)}{P(\boldsymbol{x_i})} \quad (1.4.23)$$

NB states as hypothesis, the individuality between each pair of features, *i.e.*, each sample is assumed to belong to one class $g \in \{g_1, g_2, \dots, g_h\}$ (Ranganathan *et al.*, 2019; Uddin *et al.*, 2019) as,

$$P(g|x) = P(g)\prod_{i=1}^{h} P(x_i|g) \quad (1.4.24)$$

The example presented by Uddin *et al.* (2019) illustrates the idea in the NB algorithm, Figure 1.4.11-A.

Consider a trained NB, the green circle (🟢) in Figure 1.4.11-A must be classified into one of two classes blue circle (🔵) and orange x (✖). According to the prior probability, $20/30 = 0.67$ for orange X and $10/30 = 0.33$ for blue circle, which is more probable to fall in the orange x class. However, if it is considered the four points, three blue circles, and one orange X, the likelihood of the green circle in the blue circle is $3/10 = 0.3$ the likelihood for the orange x is $1/20 = 0.05$. The posterior probability is calculated for both classes. The blue circle is $0.33 \times 0.3 = 0.099$, and the orange x $0.67 \times 0.05 = 0.034$. Thus, according to the NB technique, the sample is classified as a blue circle.



(A) Naïve Bayes (B) *k*-Nearest Neighbors

(C) Quadratic Discriminant Analysis (D) Neuronal Network

Figure 1.4.11. Naïve Bayes, k-Nearest Neighbors, Quadratic discriminant analysis, and Neural Network with two classes blue circle (●) and orange x (✕).

k-Nearest Neighbors

k-Nearest Neighbors (k-NN) algorithm is considered a simpler version of the NB, where it is not required to compute the probability values. In this algorithm, the k nearest neighbors are considered to determine the classes. It follows a Memory-based Classification rather than having an explicit training process (Cunningham and Delany, 2022), Figure 1.4.11-B. k-NN is developed in 2 steps. **Step 1.** The nearest neighbors are determined. **Step 2.** With the neighbors previously calculated is the determined class. This algorithm could be defined as follows:

Consider the training dataset ($\mathcal{D} = \{\boldsymbol{x_i}, \boldsymbol{g_i}\}$) from Equation (1.4.2). The distance between the unknown sample q and x_i for each x_i is calculated as,

$$d(q, x_i) = \sum_{f \epsilon F} w_f \cdot \delta(q, x_i)_f \qquad (\ 1.4.25\ )$$

This equation represents the sum over all the features $F\epsilon\mathcal{R}^{1\times n}$, where w_f is the weight for each feature and $\delta(q,x_i)_f$ is the metric distance. Metric distance is zero If $q = x_i$ and one if $q \neq x_i$. This metric distance is used to select the nearest neighbors. Afterward, it is necessary to assign the class to q, which is commonly performed, considering the class of the closest neighbors. There are two ways to develop this idea. The first one is the distance-weighted voting, Equation (1.4.26).

$$Vote(g_j) = \sum_{c=1}^{h} \frac{1}{d(q,x_c)^p} 1(g_j, g_c) \qquad (1.4.26)$$

The neighbors vote and these votes are weighted by the inverse of their distance. The vote assigned by the neighbor x_c to class g_j is divided by the distance to their neighbor. The term $1(g_j, g_c)$ is one if the class labels match and zero in the opposite case.

The second is based on Shepard's work, using exponential function rather than the inverse distance, Equation (1.4.27).

$$Vote(g_j) = \sum_{c=1}^{k} e^{d(q,x_c)^p} 1(y_j, y_c) \qquad (1.4.27)$$

Quadratic Discriminant Analysis

Quadratic Discriminant Analysis (QDA) is an algorithm that assumes a quadratic decision boundary, assuming that the features follow a normal distribution and quadratic interaction between classes (Hastie *et al.*, 2009), Figure 1.4.11-C.

The QDA algorithm is derived from probabilistic approaches which model the conditional class distribution of the data $P(x|g)$ for each class h. QDA could be formulated for each sample $\boldsymbol{x_i} \in \mathcal{R}^{\mathrm{N}\times\mathrm{p}}$ as,

$$P(g_i|x_i) = \frac{P(x_i|g_i)P(g_i)}{P(x_i)} = \frac{P(x_i|g_i)P(g_i)}{\sum_h P(x_i|g)P(g)} \qquad (1.4.28)$$

and the class h, which maximizes the posterior probability is selected.

$$P(x|g) = \frac{1}{(2\pi)^{d/2}|\Sigma_h|^{1/2}} e^{\left(-\frac{1}{2}(x-\mu_h)^T \Sigma_k^{-1}(x-\mu_h)\right)} \quad (1.4.29)$$

Then, the predicted class is the one that maximizes the logarithmic of the posterior represented as follows, Equation (1.4.30).

$$\begin{aligned} logP(g|x) &= logP(x|g) + logP(g) + C_{st} \\ &= -\frac{1}{2} log|\Sigma_h| - \frac{1}{2}(x-\mu_h)^T \sum_h^{-1} (x-\mu_h) + logP(g) + C_{st} \end{aligned} \quad (1.4.30)$$

The term C_{st} accounts for the denominator of Equation (1.4.29) and the other constant terms from the Gaussian distribution.

Neural Network

Neuronal Network (NN) algorithm is based on the functioning of the neural networks in the brain. It can be described as an interconnected group of nodes, called neurons through axons, and thus form the network structure (Zhou, 2012). In NN, the output of one neuron is the input to another neuron, constructing signals that are multiplied by the respective connection weights (signal strengths), and the signals are aggregated and compared with a threshold called the bias of the neuron (Uddin *et al.*, 2019).

One of the most used NN is the multi-layer feed-forward network. In this NN, the neurons are related layer by layer, *i.e.*, there is an input layer that receives the input feature vector, the output vector where each neuron commonly corresponds to a possible label, and the hidden layer, which is the layer between input and output layer, Figure 1.4.11-D.

The Back-Propagation Neuronal Network (BPNN) is one of the most implemented in NN. BPNN consists of feeding forwarded the input layer, the information could be processed through the hidden layer to the output layer, and the error is calculated by comparing the output of the NN with the ground truth. Finally, this error is propagated to the hidden layer and returned to the input layer, correcting the weights and bias to reduce this error (Zhou, 2012).

4.2 Machine Learning Models in Biological Process

This subsection mainly presents examples of machine learning models used in biological processes. The above-mentioned black box machine learning approaches are of two types, (a) regression and (b) classification (Gupta *et al.*, 2023). The regression goal is to predict output variables using numerical or categorical predictor variables (Kazemi *et al.*, 2020b; Robles-Rodriguez *et al.*, 2022; Sharma and Tambe, 2014; Tufaner *et al.*, 2017). The classification objective addresses fault or anomalous detection (Kazemi *et al.,* 2020a) or categorizes a process according to the determinate substrate produced (Cinar *et al.*, 2022; Wang *et al.*, 2020).

Robles-Rodriguez *et al.* (2016) developed a soft sensor based on SVM coupled with a Particle Swarm Optimization (PSO) algorithm for monitoring lipid fermentation of *Yarrowia lipolytica* growing on glucose. The objective was to estimate the lipid, biomass, and citric acid concentrations. PSO was performed to estimate the SVM parameters $C, \sigma, and\ \varepsilon$ to avoid local minimum and high calculations times. SVM was trained and validated with different datasets of fed-batch fermentations using the on-line measurements: added base to control pH and partial oxygen pressure. The authors determined that SVM models coupled with PSO estimate the lipids and biomass concentration in fed-batch reactors satisfactory, concluding that the soft sensor in a PSO-SVM is an efficient alternative for monitoring fermentations.

Robles-Rodriguez *et al.* (2022) used a soft sensor based on SVM to monitor the production of proteins by *B. thuringiensis,* a microorganism with physiological changes during fermentation. The experiments were developed in a 3 L fermenter operating at 30°C and pH 6.8 with a synthetic medium (glucose) as a substrate. Quadratic, cubic, and Radial Basis Function (RBF) SVM models were constructed using seven datasets with 8 k-folds cross-validation and three datasets for the test. Twelve measurements (on-line and off-line measures) were used as features to train the model. A total of 8 predictor combinations were computed via the Root Mean Square Error (RMSE). The preliminary results indicated that a quadratic SVM with ten features achieved the best results. Nevertheless, a new SVM model was performed using only on-line measurements, where the RBF SVM obtained good results concerning the RMSE for the validation and training, concluding that SVM can accurately represent the process's non-linearities.

Sharma and Tambe (2014) used three different soft sensors. Multi-Layer Perceptron (MLP) Neural Network (NN), RBF Support Vector Regression (SVR), and Genetic programming (GP), for monitoring two biological processes: the extracellular production of lipase and the bacterial production of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) copolymer. In the first process, a GP-based soft sensor is used to predict the lipase activity from four inputs, soy oil, NH_4NO_3, corn steep liquor concentrations, and fermentation time. The results showed that the three model-driven methods achieved good results in the test in terms of $R^2 > 0.83$ and $RMSE < 1.54$. However, the GP presented better results compared with the others ($R^2 > 0.92$ and $RMSE < 0.96$). In the second case a GP based soft sensor is used to estimate the accumulation of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) from four inputs, acetate and propionate concentration, incubation period, and pH. The results showed that all the three model-driven achieved good results in the test in terms of $R^2 > 0.93$ and $RMSE < 3.8$. However, the GP and the RBF SVR have better results compared with the MLPNN (R^2>0.93 and $RMSE < 2.6$). In both cases it was concluded that soft sensors, especially the GP, could be applied to monitor biological processes with non-linearities

Tufaner *et al.* (2017) performed a Back Propagation Neural Network (BPNN) soft sensor to estimate the biogas production in the AD process. The experiments were performed in 6.15 L up-flow anaerobic sludge blanket (UASB) reactors operating at mesophilic conditions. The reactors treated cattle manure with the co-digestion of different substrates such as grass waste, household organic waste, industrial organic waste, and sludge. The BPNN considered ten inputs: working days, influent COD, influent pH, influent alkalinity, influent ammonia, total influent phosphorus, HRT, waste adding ratio, pretreatment waste sorts, and reactor number, while the biogas production was considered as output. The dataset comprised 180 experimental data points from five AD reactors, where one half was used for training, one quarter for validation, and the rest for the test sets. An interesting aspect is that the authors evaluated 11 algorithms to identify the back propagation in terms of the MSE. The authors concluded that BPNN could be used as a soft sensor to determine the biogas performance in the AD with coefficients of determination (R^2) of 0.89 and 0.75 for the training and test.

Yilmaz *et al.* (2010) used three different NN techniques to determine the output COD in the AD process: MLPNN, Radial Basis Neural Network (RBNN), and Generalized Regression neural network (GRNN). The experiments were performed in a 1.33 L up-flow anaerobic filter (UAF) reactor with an HRT of 24 h, operating at 35°C for 130 days. The UAF reactors treated a mixture of sludge and wastewater (30% v/v). The reactor was also continuously fed with increasing cyanide (CN) concentrations from 1 mg/L to 130 mg/L. The three NN techniques used as inputs: the inlet chemical oxygen demand, HRT, and inlet cyanide concentration. The dataset was composed of 134 experimental data, split in a proportion of 70/30 for the training and test. According to the test results, the MLP neural network was the best to predict the outlet COD in the process with a R^2>0.87 and $MSE < 98.3$, values significantly higher compared with the RBNN and GRNN (R^2>0.75 and $MSE < 157$).

Kazemi *et al.* (2020b) performed interesting studies of soft sensors over the AD process with different aims. In the first study, they evaluated several data-driven soft sensors to develop robust VFA monitoring using easy on-line measured variables in the AD process. The authors explored soft sensors such as BPNN, SVM, RF, extreme learning machines, and GP. The dataset was obtained using the BSM2 with different OLR as a plant model and the ASM1 and ADM1 models to describe the phenomena involved in transforming the activated sludge and AD reactor, respectively. The dataset was built using thirteen variables as inputs, which were measured at the input and output of the system for a total of 609 days every 15 minutes. The influence of these variables on the process was evaluated through a feature ranking method, determining that the variables with the strongest influence were pH, ammonia concertation, pressure, and CO_2 molar fraction. Therefore, these variables were used to build the soft sensors. The authors concluded that genetic programming achieved the best results in terms of $R^2 > 0.99$ and normalized root mean squared error $NRMSE < 0.0037$ for the test. Nevertheless, the NN, SVM, and extreme learning machine also achieved good results $R^2 > 0.99$ and $NRMSE < 0.0090$. Finally, it was demonstrated that the use of this type of soft sensors in AD processes is possible from the use of datasets generated from models such as BSM2.

Kazemi *et al.* (2020a) also applied the SVM algorithm to fault detection. The dataset was selected similarly using the BSM2 model and 13 input variables. To detect the faults, the VFA

soft sensor was used together with three statistical control charts (Square prediction error, Cumulative sum, and T^2 Hotelling's charts). These control charts were built using the residual determined between the simulated VFA and the predicted VFA using the SVM. The quality of the fault detection methods was evaluated using some statistics criteria such as precision, recall, and F1 scores. They disturbed the acetate concentration in the system by manipulating the maximum uptake rate of acetate in BSM2 from ±5% to ±15% concerning the default value. The disturbance was simulated from day 530 to the end of the simulation. All the control charts obtained interesting results. However, the VFA cumulative sum achieved the best F1 score, *i.e.*, it was considered the best control chart to determine small-magnitude faults.

Wang *et al.* (2020) performed an interesting study to predict CH_4 production and identify determinant operational parameters. Four machine learning algorithms were selected for regression and classification: RF, logistic regression multiclass, SVM, and k-NN. Their model used as inputs the total content of carbon and nitrogen, C/N ratio, cellulose, xylan, lignin, and glucan content, and temperature and as output the CH_4 production. The dataset was built using 17 instances from the literature with the same AD configuration. In both cases, the dataset was split randomly into training and test set: 15/2 and 14/3 for regression and classification, respectively. The instances were divided into three classes according to the CH_4 content to evaluate the classification, low (>300 $mL/L/d$), medium (300-400 $mL/L/d$), and high (>400 $mL/L/d$). All the machine learning algorithms demonstrated good results in predicting CH_4 production. RMSE of 65.1, 83.6, 83.6 and 36.9 were achieved in the RF, logistic regression multiclass, SVM, and k-NN during the training. During the validation, RMSE values of 81.5, 71.7,68.6, and 89.0 were obtained in the RF, logistic regression multiclass, SVM, and k-NN, respectively. On the other hand, the accuracy metric was used in the classification to measure the results of the algorithms with low, medium, and high CH_4 production. Values of 0.64, 0.73, 0.59, and 0.61 were determined for RF, logistic regression multiclass, SVM, and k-NN, respectively. This result shows that the logistic regression multiclass was suitable for classifying the dataset proposed.

Cinar *et al.* (2022) used various machine learning algorithms (linear regression, logistic regression, k-NN, DT, RF, SVM, and extreme gradient boosting) to study temperature changes

in the AD process. The experiments were performed in a 5.4 L $CSTR$ operating at 42°C and 55 rpm. Pellets (animal feed material) were used as a substrate for the AD. Four $CSTR$ with different temperatures and organic loading rates were evaluated. The dataset was built with ten features with possible impacts on the model, biogas production, nutrient solution usage, biogas production temperature, biogas pressure, waste vapor pressure, standard volume, reactor temperature, character, and feed. Then, the dataset was classified based on the volume of dry gas in the normal state, low class (9.91-901.82 NmL/d), medium class (901.82 - 1707.86 NmL/d), and high class (>1707.86 NmL/d). The regression models used the RMSE to identify the accuracy of the prediction, while classification models used the confusion matrix approach to classify the AD according to the standard CH_4 volume and compute the precision, recall, and F1-score. Only the best results were presented for the regression and classification models. In the regression, $RMSE$ of 246.96, 72.16, and 93.91 were obtained for the linear regression, DT and RF, respectively, while in classification, accuracy values of 0.93, 0.89, 0.88, and 0.86 were determined for SVM, RF, k-NN, and DT, respectively. The authors successfully implemented machine learning models capable of predicting changes in the temperature and feedings in the AD and performing efficient real-time monitoring.

4.3 Conclusions Soft Sensors and Fault Detection

Soft sensors have become a valuable tool for monitoring, control, and optimization tasks in biological processes due to the capacity to use on-line measurements to estimate unmeasurable variables in real time (Brunner *et al.*, 2021; Yan *et al.*, 2021). Soft sensors are a low-cost alternative concerning physical devices and it can be implemented on existing hardware (Fortuna *et al.*, 2007; Jiang *et al.*, 2021).

Data-driven approaches are based on readily available online data or historical recordings of the process (Cruz *et al.*, 2022), avoiding the need to build a mathematical description that requires an in-depth knowledge of the process (Gopakumar *et al.*, 2018; Kadlec and Gabrys, 2009; Wade, 2020).

Several data-driven soft sensors have been successfully applied to biological processes, especially in AD. Table 1.4.1 summarizes the previously described works comparing the type

of algorithm that was used and the application. The interesting aspect is the wide use of those soft sensors. As a regression type to predict variables such as VFA or CH_4 production (Kazemi *et al.*, 2020b; Tufaner *et al.*, 2017). As a classification type to describe the process based on the CH_4 content (Wang *et al.*, 2020) or a combination of both tasks (Cinar *et al.*, 2022).

There are no works in process fault detection over the biological methanation process. However, some studies have promising results in AD (Kazemi *et al.*,2020a) using Machine learning models. This leads us to think that it is feasible to perform similar studies on biological methanation, which has not been explored yet.

This section provides some questions that we could explore:

- Can machine learning be used as a data-driven soft sensor in biological methanation?
- Can these soft sensors be used to detect faults during the process?

Table 1.4.1. Soft sensor in biological process.

Process	Method applied	Application type	Inputs	Outputs	R^2	$RMSE$	Accuracy / F1 score	Comments	Reference
Lipid fermentations	Particle Swarm Optimization (PSO)-SVM	Process Monitoring	Added base to control pH, volume and partial oxygen pressure	Lipid, biomass, and citric acid concentration	--	--	--	The PSO was performed to estimate the SVM parameters C, $\sigma, and\ \varepsilon$ to avoid local minimum and high calculations times	(Robles-Rodriguez *et al.*, 2016)
Protein production	Quadratic, cubic, and RBF SVM	Process Monitoring	Partial pressure, agitation, pH, aereation, and strain number	Protein concentration	--	148.65 for training and 172.29 for validation	--	Exploration of SVM using quadratic, cubic, and RBF kernels	(Robles-Rodriguez *et al.*, 2022)
Extracellular production of lipase enzyme	Genetic programming (GP), neural network, RBF SVR	On-line prediction	Soy oil, NH_4NO_3, and corn concentration	Lipase activity	$R^2 > 0.83^*$	$RMSE < 1.54^*$	--	GP is an extension of the Genetic algorithm. Given the objective function, the GA searches and optimizes the values of the decision variables that would maximize/minimize the function	(Sharma and Tambe, 2014)
Bacterial production of poly(3-hydroxybutyrate-co-3-hydroxyvalerate)			Acetate, propionate concentration, incubation period, and pH	poly(3-hydroxybutyrate-co-3-hydroxyvalerate)	R^2>0.93*	$RMSE < 2.6^*$	--		

* for all methods applied

Table 1.4.1. Continuation soft sensor in biological process.

Process	Method applied	Application type	Inputs	Outputs	R^2	$RMSE$	Accuracy / F1 score	Comments	Reference
AD of cattle manure with the co-digestion of different substrates such as grass waste, house hold organic waste industrial organic waste and sludge	BPNN	on-line prediction	Time, inlet COD, pH, alkalinity, ammonia, and total phosphorus, HRT, waste adding ratio, pretreatment waste sorts, and reactor number	Biogas production	0.89 for training and 0.75 for test	--	--	Evaluation of 11 algorithms to identify the back propagation in terms of the MSE. The scaled conjugate gradient backpropagation with an MSE of 0.0322 was selected	(Tufaner *et al.*, 2017)
Activated sludge	BPNN, SVM, RF, Extreme learning machine, genetic program ming	Process Monitoring	pH, ammonia concertation, pressure and CO_2 molar fraction.	VFA	$R^2 > 0.99^{*}$	$NRMSE <$ $0.0090^{\$*}$	--	All the soft sensors are considered black box models. on the contrary, the genetic programming soft sensor is transparency, which makes it easy to integrate into process control systems without any further modifications	(Kazemi, Steyer *et al.*, 2020)
		Process fault detection			--	--	F1 score = $0.96^{\#}$	A control chart is a graphical technique wherein a value of specific statistics is presented over time and, for the normal operation of the process, the statistics must not pass a predetermined control limit	(Kazemi *et al.*, 2020a)

* for all methods applied; $^{+}$ the MLP NN; $^{\$}$ normalize RMSE $(NRMSE)$; $^{\#}$ VFA CUSUM with +5% disturbance in $\mu_{m,ac}$

Table 1.4.1. Continuation soft sensor in biological process.

Process	Method applied	Application type	Inputs	Outputs	R^2	$RMSE$	Accuracy / F1 score	Comments	Reference
AD with different substrates from the literature	RF logistic regression multiclass SVM k-NN	Process Monitoring and Classification	Total content of carbon and nitrogen, C/N ratio, cellulose, xylan, lignin, and glucan content, and temperature	CH_4 production	--	RMSE < 89 81.5* in test	With values of 0.64,0.73,0.59,0.61 for RF, logistic regression multiclass, SVM and k-NN, respectively£	Further improvements such as increasing the amount of data and algorithm optimization could benefit the application of machine learning in predicting digestion performance	(Wang *et al.*, 2020)
AD with OLR and temperature variations	linear regression, logistic regression, k-NN, DT, RF, SVM, and extreme gradient boosting	Process Monitoring and Classification	biogas production, nutrient, biogas production temperature and pressure, waste vapor pressure, standard volume, reactor temperature, character, and feed	Standard CH_4 volume	--	$RMSE <$ 246.96*	accuracy values of 0.93, 0.89, 0.88, and 0.86 were determined for SVM, RF, k-NN, and DT, respectively¢	Concerning the regression models, the linear regression model, decision tree and random forest obtained the best RME. In classification models, the SVM, RF, and k-NN algorithms	(Cinar *et al.*, 2022)
AD of a mixture of sludge and wastewater (30% v/v), and cyanide concentrations	MLP NN, RBNN, and Generalized Regression NN	On-line prediction	Inlet COD, HRT, and inlet cyanide concentration.	Outlet COD	R^2>0.87 +	$RMSE <$ 98.3+	--	--	(Yilmaz *et al.*, 2010)

£In classification for low, medium, and high CH_4 production; ¢Low class (9.91-901.82 NmL/d), medium class (901.82 -1707.86 NmL/d), and high class.

4.4 References

Aggarwal, C.C., 2015. Data Mining: The Textbook. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-14142-8

Batuwita, R., Palade, V., 2013. Class Imbalance Learning Methods for Support Vector Machines, Imbalanced Learning. John Wiley & Sons, Inc., Hoboken, NJ, USA. https://doi.org/10.1002/9781118646106.ch5

Biau, G., Scornet, E., 2016. A random forest guided tour. TEST 25, 197–227. https://doi.org/10.1007/s11749-016-0481-7

Brunner, V., Siegl, M., Geier, D., Becker, T., 2021. Challenges in the Development of Soft Sensors for Bioprocesses: A Critical Review. Front. Bioeng. Biotechnol. 9, 722202. https://doi.org/10.3389/fbioe.2021.722202

Bzdok, D., Krzywinski, M., Altman, N., 2018. Machine learning: supervised methods. Nature Methods 15, 5–6. https://doi.org/10.1038/nmeth.4551

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A., 2020. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing 408, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Charbuty, B., Abdulazeez, A., 2021. Classification Based on Decision Tree Algorithm for Machine Learning. JASTT 2, 20–28. https://doi.org/10.38094/jastt20165

Cinar, S.Ö., Cinar, S., Kuchta, K., 2022. Machine Learning Algorithms for Temperature Management in the Anaerobic Digestion Process. Fermentation 8, 65. https://doi.org/10.3390/fermentation8020065

Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge: Cambridge University Press, London, UK. https://doi.org/10.1017/CBO9780511801389

Cruz, I.A., Chuenchart, W., Long, F., Surendra, K.C., Andrade, L.R.S., Bilal, M., Liu, H., Figueiredo, R.T., Khanal, S.K., Ferreira, L.F.R., 2022. Application of machine learning in anaerobic digestion: Perspectives and challenges. Bioresource Technology 345, 126433. https://doi.org/10.1016/j.biortech.2021.126433

Cunningham, P., Delany, S.J., 2022. k-Nearest Neighbour Classifiers - A Tutorial. ACM Comput. Surv. 54, 1–25. https://doi.org/10.1145/3459665

Deisenroth, M.P., Faisal, A.A., Ong, C.S., 2020. Mathematics for Machine Learning, Springer Undergraduate Mathematics Series. Cambridge University Press, Cambridge. https://doi.org/10.1017/9781108679930

Doraiswami, R., Cheded, L., 2014. ROBUST MODEL-BASED SOFT SENSOR: DESIGN AND APPLICATION. IFAC Proceedings Volumes 47, 5491–5496. https://doi.org/10.3182/20140824-6-ZA-1003.00245

Fawagreh, K., Gaber, M.M., Elyan, E., 2014. Random forests: From early developments to recent advancements. Systems Science and Control Engineering 2, 602–609. https://doi.org/10.1080/21642583.2014.956265

Fortuna, L., Graziani, S., Rizzo, A., Xibilia, M.G. (Eds.), 2007. Soft sensors for monitoring and control of industrial processes, 1st ed, Advances in Industrial Control. Springer, London.

Gopakumar, V., Tiwari, S., Rahman, I., 2018. A deep learning based data driven soft sensor for bioprocesses. Biochemical Engineering Journal 136, 28–39. https://doi.org/10.1016/j.bej.2018.04.015

Gupta, R., Zhang, L., Hou, J., Zhang, Z., Liu, H., You, S., Sik Ok, Y., Li, W., 2023. Review of explainable machine learning for anaerobic digestion. Bioresource Technology 369, 128468. https://doi.org/10.1016/j.biortech.2022.128468

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Springer Series in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-84858-7

Jiang, Y., Yin, S., Dong, J., Kaynak, O., 2021. A Review on Soft Sensors for Monitoring, Control, and Optimization of Industrial Processes. IEEE Sensors J. 21, 12868–12881. https://doi.org/10.1109/JSEN.2020.3033153

Kadlec, P., Gabrys, B., 2009. Soft sensors: where are we and what are the current and future challenges? IFAC Proceedings Volumes 42, 572–577. https://doi.org/10.3182/20090921-3-TR-3005.00098

Kadlec, P., Gabrys, B., Strandt, S., 2009. Data-driven Soft Sensors in the process industry. Computers & Chemical Engineering 33, 795–814. https://doi.org/10.1016/j.compchemeng.2008.12.012

Kadlec, P., Grbić, R., Gabrys, B., 2011. Review of adaptation mechanisms for data-driven soft sensors. Computers & Chemical Engineering 35, 1–24. https://doi.org/10.1016/j.compchemeng.2010.07.034

Kazemi, P., Bengoa, C., Steyer, J.-P., Giralt, J., 2021. Data-driven techniques for fault detection in anaerobic digestion process. Process Safety and Environmental Protection 146, 905–915. https://doi.org/10.1016/j.psep.2020.12.016

Kazemi, P., Giralt, J., Bengoa, C., Steyer, J.P., 2020a. Data-driven fault detection methods for detecting small-magnitude faults in anaerobic digestion process. Water Science and Technology 81, 1740–1748. https://doi.org/10.2166/wst.2020.026

Kazemi, P., Steyer, J.P., Bengoa, C., Font, J., Giralt, J., 2020b. Robust data-driven soft sensors for online monitoring of volatile fatty acids in anaerobic digestion processes. Processes 8. https://doi.org/10.3390/pr8010067

Kowalczyk, A., 2017. Support Vector Machines Succintctly. Morrisville: Succintly E-Book Series, Morrisville.

López Cabrera, J.D., Pereira-Toledo, A., 2018. Análisis Del Comportamiento Del Algoritmo Svm Para Diferentes Kernel En Ambientes Controlados. Holos 5, 101–115. https://doi.org/10.15628/holos.2018.5563

Mehmood, A., Mukherjee, M., Ahmed, S.H., Song, H., Malik, K.M., 2018. NBC-MAIDS: Naïve Bayesian classification technique in multi-agent system-enriched IDS for securing IoT against DDoS attacks. J Supercomput 74, 5156–5170. https://doi.org/10.1007/s11227-018-2413-7

Muller, A.C., Guido, S., 2016. Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media, Inc, Sebastopol.

Nasteski, V., 2017. An overview of the supervised machine learning methods. HORIZONS 4, 51–62. https://doi.org/10.20544/HORIZONS.B.04.1.17.P05

Noble, W.S., 2006. What is a support vector machine? Nat Biotechnol 24, 1565–1567. https://doi.org/10.1038/nbt1206-1565

Panup, W., Ratipapongton, W., Wangkeeree, R., 2022. A Novel Twin Support Vector Machine with Generalized Pinball Loss Function for Pattern Classification. Symmetry 14, 1–26. https://doi.org/10.3390/sym14020289

Priyanka, N.A., Kumar, D., 2020. Decision tree classifier: a detailed survey. IJIDS 12, 246. https://doi.org/10.1504/IJIDS.2020.108141

Ranganathan, S., Gribskov, M.R., Nakai, K., Schönbach, C. (Eds.), 2019. Encyclopedia of bioinformatics and computational biology. Elsevier, Amsterdam, Netherlands.

Robles-Rodriguez, C.E., Abboud, J., Abdelmalek, N., Rouis, S., Bensaid, N., Kallassy, M., Cescut, J., Fillaudeau, L., Lara, C.A.A., 2022. Soft-Sensors for Monitoring B. Thuringiensis Bioproduction, in: Distributed Computing and Artificial Intelligence, Volume 1: 18th International Conference. Springer International Publishing, Cham, pp. 129–136.

Robles-Rodriguez, C.E., Bideaux, C., Roux, G., Molina-Jouve, C., Aceves-Lara, C.A., 2016. Soft-Sensors for Lipid Fermentation Variables Based on PSO Support Vector Machine (PSO-SVM), in: Distributed Computing and Artificial Intelligence, 13th International Conference, Advances in Intelligent Systems and Computing. Springer International Publishing, Cham, pp. 175–183. https://doi.org/10.1007/978-3-319-40162-1_19

Rokach, L., Maimon, O., 2005. Decision Trees, in: Springer (Ed.), Artificial Intelligence Methods in the Environmental Sciences. Springer Netherlands, Dordrecht, pp. 165–192. https://doi.org/10.1007/978-1-4020-9119-3_4

Sbárbaro, D., del Villar, R. (Eds.), 2010. Advanced Control and Supervision of Mineral Processing Plants, Advances in Industrial Control. Springer London, London. https://doi.org/10.1007/978-1-84996-106-6

Sharma, S., Tambe, S.S., 2014. Soft-sensor development for biochemical systems using genetic programming. Biochemical Engineering Journal 85, 89–100. https://doi.org/10.1016/j.bej.2014.02.007

Simeone, O., 2018. A Very Brief Introduction to Machine Learning with Applications to Communication Systems. IEEE Transactions on Cognitive Communications and Networking 4, 648–664. https://doi.org/10.1109/TCCN.2018.2881442

Singh, S., Giri, M., 2014. Comparative Study Id3, Cart And C4.5 Decision Tree Algorithm: A Survey. International Journal of Advanced Information Science and Technology.

Tharwat, A. (2016). Linear vs. quadratic discriminant analysis classifier: A tutorial. International Journal of Applied Pattern Recognition, 3(2), 145. https://doi.org/10.1504/IJAPR.2016.079050

Tufaner, F., Avşar, Y., Gönüllü, M.T., 2017. Modeling of biogas production from cattle manure with co-digestion of different organic wastes using an artificial neural network. Clean Techn Environ Policy 19, 2255–2264. https://doi.org/10.1007/s10098-017-1413-2

Uddin, S., Khan, A., Hossain, M.E., Moni, M.A., 2019. 2019. BMC Med Inform Decis Mak 19, 281. https://doi.org/10.1186/s12911-019-1004-8

Vapnik, V., Golowich, S.E., Smola, A., 1997. Support vector method for function approximation, regression estimation, and signal processing, in: 9th International Conference on Neural Information Processing Systems. pp. 281–287.

Wade, M.J., 2020. Not Just Numbers: Mathematical Modelling and Its Contribution to Anaerobic Digestion Processes. Processes 8, 888. https://doi.org/10.3390/pr8080888

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, 122495. https://doi.org/10.1016/j.biortech.2019.122495

Watson, D.S., 2023. On the Philosophy of Unsupervised Learning. Philos. Technol. 36, 28. https://doi.org/10.1007/s13347-023-00635-6

Wu, X., Kumar, V., Ross, Q.J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.H., Steinbach, M., Hand, D.J., Steinberg, D., 2008. Top 10

algorithms in data mining, Knowledge and Information Systems. https://doi.org/10.1007/s10115-007-0114-2

Xiao, B., Zhu, H., Zhang, S., OuYang, Z., Wang, T., Sarvazizi, S., 2022. Gray-Related Support Vector Machine Optimization Strategy and Its Implementation in Forecasting Photovoltaic Output Power. International Journal of Photoenergy 2022, 1–9. https://doi.org/10.1155/2022/3625541

Yan, P., Gai, M., Wang, Y., Gao, X., 2021. Review of Soft Sensors in Anaerobic Digestion Process. Processes 9, 1434. https://doi.org/10.3390/pr9081434

Yilmaz, T., Seckin, G., Yuceer, A., 2010. Modeling of effluent COD in UAF reactor treating cyanide containing wastewater using artificial neural network approaches. Advances in Engineering Software 41, 1005–1010. https://doi.org/10.1016/j.advengsoft.2010.04.002

Yuxi (Hayden) Liu, 2020. Python Machine Learning By Example, 3rd ed. Packt, Mumbai.

Zhou, Z.-H., 2012. Ensemble methods: foundations and algorithms, Chapman & Hall/CRC machine learning & pattern recognition series. Taylor & Francis, Boca Raton, FL.

Zhu, X., Rehman, K.U., Wang, B., Shahzad, M., 2020. Modern soft-sensing modeling methods for fermentation processes. Sensors (Switzerland) 20. https://doi.org/10.3390/s20061771

Chapter 2 Principal Contributions

Detailed Content Chapter 2

Summary of Chapter 2

The work developed during this thesis is presented as publications, where each section describes a scientific paper. The first section is an introduction to the publications, which include a summary, brief comments, and highlights of each work. In Section 2 is presented the Extended Anaerobic Digestion Model No.1 (ADM1_ME) for biological methanation, which is the base of all the optimization tools and data-driven models used in this thesis. In Sections 3-5 is applied the Mutlti-Objective Dynamic Optimization (MODO) strategy for different objectives. Section 3 is proposed for the simultaneous maximization of methane yield and productivity. In Section 4, the grade of complexity is increased, considering the maximization of yield and productivity of methane and acetate. Section 5 is developed to maximize two economic objectives, Gain and Profit Margin, simultaneously. In Sections 6 and 7, data-driven machine learning models are trained to fault detections in the biological methanation process. Those sections aimed to detect and classify deviations from the optimal biological methanation operations determined with the MODO strategy when disturbances of ±10, ±15, and ±20% occur in the inlet liquid flow rate. Section 6 addresses the use of several data-driven machine learning such as decision trees, random forest, quadratic discriminant analysis, neural networks, etc, while Section 7 is mainly oriented toward using a data-driven Support Vector Machine (SVM).

Section 1 Chapter Introduction

Some of the questions formulated in Section 1 of Chapter 1 were related to the experimental development of the biological methanation process, for example, which are the best conditions to carry out biological methanation in different types of reactors such as $CSTR$ or BCR and what is the effect of using different kinds of substrates varying GLR and OLR? This question could be answered from the bibliography review. Nevertheless, other questions in this thesis were addressed from a modeling and simulation perspective, *e.g.,* Which are the optimal operating conditions to improve the yields and productivities of biological methanation? Can we set them automatically? Can the variations in the composition of the syngas be used to upgrade the biological methanation and its effect on the mass transfer process?

As mentioned in Section 3 of Chapter 1, the dynamic modeling of biological processes is a tool that allows performing multiple scenarios without developing an experimental setup. However, to propose a model for biological methanation, it is necessary to review some of the questions proposed in Section 2 of Chapter 1: Can a mathematical model of biological methanation accurately reproduce multiple operational conditions with emphasis on using different kinds of substrates varying GLR and OLR? How can the transformation of CO into acetate and H_2 and their inhibitions be described in a model for biological methanation?

To address these questions, the article of **Section 2** proposes a model for biological methanation (ADM1_ME). The objective was to obtain a model allowing a global representation of the process. We initially extended the Anaerobic Digestion Model No. 1 (ADM1) to consider the addition of syngas (H_2, CO_2, and CO) as a substrate. We used equations that allowed us to analyze the variation of the volumetric transfer coefficient in relation to the reactor type, a bubble column reactor (BCR) and a Continuous Stirred Tank Reactor ($CSTR$). The parameters of this model were analyzed with sensitivity analysis to find the parameters that could significantly affect the system outputs. Afterwards, the ADM1_ME was calibrated by estimating the most sensitive model parameters identified from the sensitivity analysis and to fit the model outputs with the literature value. This parameter estimation was performed to minimize an adapted root mean of square errors. The ADM1_ME was validated by assessing the model performance against literature value to guarantee the model's reliability. In both cases, statistical analysis was performed using two criteria: the coefficient of determination (R^2) and the root mean squared error ($RMSE$).

The literature value consisted of two datasets: Operational Condition 1 (OP1) and Operational Condition 2 (OP2). Approximately 2/3 of both datasets were used in the calibration and the rest in the validation of the model. An exciting aspect of this study was that both datasets were completely different, *i.e.*, OP1 was generated in a mesophilic BCR that used glucose and syngas (H_2/CO 0.55/0.45) as substrate in the liquid and gas phases, respectively. On the contrary, OP2 consisted of a thermophilic $CSTR$ using primary sludge, activated ticked-disintegrated waste, and syngas (H_2/CO 0.5/0.5) as substrates in the liquid and gas phases, respectively.

Model simulations were accurate in the calibration step with $R^2 > 0.90$ and a $RMSE < 0.38$ for all outlet gas flow rates in OP1, and $R^2 > 0.91$ and a $RMSE < 2.52$ for all gas percent in OP2. In the validation step with OP1, $R^2 > 0.74$ and a $RMSE < 0.94$ were obtained for output variables such as CO and H_2 gas flow rates. With OP2 values of $R^2 > 0.82$ and $RMSE < 5.15$ were achieved for all gas percent composition at the output. It was concluded that ADM1_ME is a promising model that could be used to simulate, optimize, and control a wide range of operating conditions in biological methanation.

One of the common needs of biological processes is the simultaneous optimization of multiple variables, which are commonly conflicting. Therefore, in Section 3 of Chapter 1, some questions related to this topic were concluded: Can the multi-objective optimization approaches improve biological methanation? And how to implement a computationally feasible model-based control strategy for biological methanation?

With this on mind, a Multi-Objective Dynamic Optimization (MODO) was applied for biological methanation in the article in **Section 3**. A Model Predictive Control (MPC) schema was applied using the dynamic model ADM1_ME proposed in Section 2. The aim was to enhance the biological methanation process by maximizing two objective functions: methane yield and productivity, by using the inlet liquid and gas flow rates as manipulated variables.

The MODO strategy handled different trade-offs between the objective functions. We proposed five cases of study based on the selection of different Pareto Optimal Point (POP) from the Pareto Optimal Set (POS). Case 1 simulated literature value (without control). Cases 2-4 used the POP that maximized either methane productivity, the Euclidean length between them, or

methane yield, respectively. The Euclidean length refers to the maximum distance from the POS to the origin. Case 5 was the most interesting. It consisted in switching between the maximum productivity, Euclidean length, and yield.

The results evidenced the conflicting behavior between objective variables and improved yield and productivity to 1.02 and 3.67 times concerning case 1. Case 5 permitted us to elucidate the process's robustness and the well-accounted adaptations of the manipulated variables in simulation, especially for switching between objectives. We concluded that the MODO strategy could be considered a powerful tool to adapt the process for industry's requirements.

In the biological methanation process, besides methane, there are other value-added products, such as acetate, which can serve as a chemical platform in various industries. Therefore, in the article in **Section 4**, a MODO strategy with a more complexity level was applied to the biological methanation process to consider enhancing the biological methanation. Objectives were to maximize yields and productivities of methane and acetate, with the inlet liquid and gas flow rates as manipulated variables. Five case studies were proposed. Case 1 was used as a reference case without control. Case 2 consisted of using the POP directly in simulation. In contrast, cases 3 and 4 consisted of POP that allowed the maximization of the Euclidean distance, *i.e.*, simultaneous maximization of the yield and productivity of both methane (case 3) and acetate (case 4). Case 5 was used to demonstrate the robustness of the MODO strategy by switching the objectives between maximization of performances in the objective variables in terms of methane and acetate.

The results showed that the advantages of using the dynamic part in the MODO strategy are a reduction of approximately two days the time in which the steady state is reached once there is a stages changes and the reduction in the inlet gas flow rate. Additionally, the robustness of the strategy was demonstrated by the good adaptation of switching between products of interest, such as methane and acetate. It is concluded that the MODO strategy could allow the maximization between objective variables such as yield and productivity of methane and acetate in the biological methanation process. However, the potential scope is much broader, as it can consider adaptations to market requirements for methane and acetate.

The articles of the previous sections considered objective variables such as yields and productivities of methane and acetate and manipulated variables such as liquid and gas flow rates. To solve the highlighted part of the question concluded in Section 3 of Chapter 1: Could the multi-objective optimizations consider several objectives, such as the yields, the productivities, **and other variables in economic terms (*e.g.*, substrates prices)**? It is necessary to consider economic aspects, such as the substrate costs or product selling prices in the MODO strategy applied over the biological methanation process.

Therefore, in the article of **Section 5**, an Economic Multi-Objective Dynamic Optimization (EMODO) strategy was proposed based on the same principle as the MODO strategy but considering economic objective functions.

The objective function of the EMODO strategy deal with the maximization of $Gain$ and $Profit\ margin$ for methane and acetate by using the inlet liquid and gas flow rates as manipulated variables. The objective called $Gain$ was built based on the methane and acetate selling prices. This variable accounted for the global gain in EUR per liter of reactor per day. The $Profit\ margin$ was built based on the profitability of the process, considering the relation of the net incoming, *i.e.*, the difference between the revenue by selling the products (methane + acetate) and the cost of the substrates (glucose + H_2 and CO_2) over the total revenue.

The results showed that the EMODO strategy was a good alternative to improve the biological methanation regarding economic variables by manipulating the inlet liquid and gas flow rates. It was concluded that the EMODO strategy could be a good decision-making tool in selecting a profitable condition for the biological methanation process, even if there are fluctuations in the prices of the substrates and products.

As it was mentioned in Section 2 of Chapter 1, biological methanation is a complex process that can be affected by several factors, such as operating conditions or fluctuations in the liquid or gas flow rates, making the system susceptible to faults. Hence the importance of process monitoring using machine learning soft sensors described in Section 4 of Chapter 1. Machine learning soft sensors can be used as an alternative to detect the deviations of the biological methanation from an optimal operation or desired state. The questions that we tried to answer

were: Can machine learning be used as a data-driven soft sensor in biological methanation? And can these soft sensors be used to detect faults during the process?

Therefore **Section 6** presents the training of machine learning algorithms, such as Decision Trees (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB), k-Nearest Neighbors (k-NN), Quadratic Discriminant Analysis (QDA), Neural Networks (NN), and different Support Vector Machines (SVM). The objective of evaluating multiple algorithms was to explore alternatives with different levels of complexity and interpretability.

The results obtained in Section 5 using de MODO strategy were selected as an optimal operation point for the biological methanation. Then, the ADM1_ME was used to generate a dataset applying disturbances of ±10, ±15, and ±20% to the liquid and gas inlet flow rates with respect to their optimal values.

The results obtained are promising. DT, RF, and SVM reached the best results regarding statistic metrics with an average F1-score higher than 0.87 and accuracy values higher than 0.90 and 0.85 in the training and test. After training several types of machine learning soft sensors, we noticed that several algorithms could correctly classify faulty data in biological methanation. Computation times lower than 0.19 seconds were obtained in the training of models such as DT, while RF and SVM presented values of 13.75 and 4.77 seconds.

The results obtained in **Section 6** elucidated the advantages of using machine learning soft sensors to detect faults in the biological methanation process, especially with support vector machine models that showed shorter computation time with good accuracy. Although the results are promising, one of the objectives of this thesis is to provide simple, accurate, and fast detection of faults in the biological methanation process.

Consequently, in the article of **Section 7**, SVM was studied in more detail and applied for detecting faults in the biological methanation process. The selection of SVM algorithms was based on two aspects: (i) SVM presented one of the best results compared to the other algorithms, (ii) as mentioned in Section 4 of Chapter 1, SVM are a good alternative to construct soft sensors given its solid foundation in statistical learning theory, the capacity to work with high-dimensional feature space, small instances, and efficiency in avoiding overfitting.

Quadratic, cubic, and Radial Basis Function (RBF) SVM were trained using the same principle as the machine learning algorithms in **Section 6** to detect the optimal conditions and to classify disturbances.

The results in Section 3 were used as an optimal operation point for the biological methanation. Then, disturbances of ±10, ±15, and ±20% in the inlet and gas flow rates with respect to the optimal conditions were generated using the ADM1_ME to obtain a dataset. This dataset was then used to train the three SVM algorithms and to detect the optimal operation as well as the deviation from the optimal points subject to the aforementioned disturbances.

The results showed that in the test, a statistic metric accuracy higher than 0.88, 0.81, and 0.88 were obtained for the quadric, cubic, and RBF SVM, respectively. This study highlights that SVM models were trained using pairs of features to build 2D maps that indicate if the biological methanation process is operated in the optimal region or if a disturbance in the liquid flow rate causes a process deviation. We concluded that SVM presents promising results to classify data and can become a powerful tool at the industrial level in detecting and classifying faults in the biological methanation process, mainly if these faults occur in the inlet liquid and gas inlet flow rates.

Section 2 Dynamic Modeling of Biological Methanation for Different Reactor Configurations: An Extension of the Anaerobic Digestion Model No. 1

Juan C. Acosta-Pavas[a], Carlos. E Robles-Rodríguez[a], Jérôme Morchain[a], Claire Dumas[a], Arnaud Cockx[a], César A. Aceves-Lara[a]

[a]TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

Publications:

Acosta-Pavas, J. C., Robles-Rodríguez, Carlos. 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, 128106. https://doi.org/10.1016/j.fuel.2023.128106

Acosta-pavas, J. C., Morchain, J., Dumas, C., Ngu, V., Cockx, A., Aceves-Lara C.A. (2022). Towards Anaerobic Digestion (ADM No. 1) Model's Extensions and Reductions with In-situ Gas Injection for Biomethane Production. In A. Kugi, A. Körner, W. Kemmetmüller, A. Deutschmann-Olek, F. Breitenecker, I. Troch (eds.), 10th Vienna International Conference on Mathematical Modelling (Vol. 55(20), pp. 635-6409). IFAC-PapersOnLine. https://doi.org/10.1016/j.ifacol.2022.09.167

Abstract

In biological methanation, the methane produced by anaerobic digestion (AD) is upgraded with the addition of syngas. Several mathematical models have been developed to represent the AD process. However, the modeling of biological methanation is still under development. In this work, an extension of the anaerobic digestion model (ADM1_ME) was proposed to describe the dynamics of biological methanation. The model considered adding syngas flow rate (H_2:CO) and adapting the volumetric mass transfer coefficient for two different reactor configurations: bubble column reactor and continuous stirred tank reactor operating at mesophilic and thermophilic conditions. A sensitivity analysis using the Sobol' and Morris methods was performed for this model, where fourteen parameters were selected for model calibration. Simulation results showed an accurate fit for two experimental operating conditions from literature with a $RMSE$ <5.15. The results showed the feasibility of the ADM1_ME to describe the biological methanation process at different operational conditions and reactor configurations.

Keywords: Anaerobic Digestion, Biological Methanation, Sensitivity Analysis, Parameter Estimation

2.1 Introduction

The production of biogas (a mixture of CH_4 and CO_2) by anaerobic digestion (AD) is currently one of the most promising options in terms of bioenergy production (Brémond *et al.*, 2021). This biogas can be used locally to generate heat or electricity without any additional processing. However, it needs to be refined or purified to be used as vehicle fuels or for injection into the gas grid system (Gustafsson *et al.*, 2021; Zupančič *et al.*, 2022).

Biogas could endure post-treatment alternatives either to remove impurities (*e.g.*, H_2S, excess of water) or to be upgraded into biomethane (95 – 99 % CH_4) for a further injection in the gas grid. Biogas upgrading involves increasing the CH_4 concentration by removing CO_2 (Iglesias *et al.*, 2021; Rusmanis *et al.*, 2019). Some technologies employed in biogas upgrading include physical absorption (water, amine, and organic scrubbing), pressure swing adsorption, and membrane separation (Gustafsson *et al.*, 2021; Iglesias *et al.*, 2021). Another process that has gained interest and is currently under development is biological methanation, also called biomethanation (Rafrafi *et al.*, 2020).

In the biological methanation process, the CO_2 contained in the biogas is converted into CH_4 by using hydrogen coming from the addition of syngas, which is generally composed of CO_2, CO, and H_2 (Rusmanis *et al.*, 2019). Biological methanation involves a complex microbial consortium whose composition changes due to operational conditions, such as temperature, pH, hydraulic retention time, and syngas composition (Grimalt-Alemany *et al.*, 2020; Li *et al.*, 2020). Biological methanation can be performed either *in-situ* where syngas is introduced into the AD reactor, or *ex-situ,* where a microbial consortium, coming from the AD and adapted for H_2 consumption, is introduced into a second bioreactor to convert syngas into a high-purity CH_4 (Rusmanis *et al.*, 2019).

Until now, the development of dynamic models for AD has been focused on incorporating detailed knowledge of the process stages, the microorganisms involved, and the operating conditions. One of the first dynamic models to investigate AD was proposed b Mosey (1983), who developed a model to consider how the microorganisms managed to control the pH value

and the redox potential of their growth medium. A dynamic model describing AD from several types of wastes was developed by Angelidaki *et al.* (1999). Their work described the substrate by its composition in terms of essential organic components, *i.e.*, carbohydrates, lipids, proteins, volatile fatty acids, long-chain fatty acids, and inorganic components. This partition allowed the authors to simulate the dynamic changes during the process with different types of substrates. Batstone *et al.* (2002) proposed the well-known Anaerobic Digestion Model No. 1 (ADM1), which included multiple steps describing the biochemical and physicochemical reactions involved in AD. Regarding the biochemical reaction, the ADM1 considers the disintegration of components, such as carbohydrates, lipids, and proteins, into particulate constituents that are further hydrolyzed into soluble monomers precursors of CH_4 formation.

The physicochemical reactions describe ion associations, dissociations, and gas/liquid transfer phenomena. The ADM1 has been modified (Rosen and Jeppsson, 2006) to solve stiffness problems, mainly due to the mass transfer equations. Thus, these authors proposed a Benchmark Simulation Model (BSM). Several researchers have adapted the BSM to consider inhibition by free ammonia in high-solid sludge fermentation (Bai *et al.*, 2017), to design optimal continuous operation of experimental anaerobic digestion (Balde *et al.*, 2020), or to simulate the dynamic behavior of a pilot-scale process for two-stage anaerobic digestion of sewage sludge (Blumensaat and Keller, 2005).

Modeling of biological methanation has been scarcely studied. To our knowledge, only a few works have been reported. Grimalt-Alemany *et al.* (2020) proposed two structured models to describe the mesophilic and the thermophilic syngas biological methanation processes in batch mode. Each model presented a different structure based on catabolic routes as a function of the operating conditions. All biomass growth processes were made thermodynamically consistent by including a thermodynamic potential factor. Although these adaptations improved the predictive capacity of the models, the studied carbon source was only limited to the added gas, which could hamper any straightforward adaptation to other sources, such as agro-industrial residues, sludge, or sugars. The different volumetric mass transfer coefficients were calculated experimentally under specific mesophilic and thermophilic conditions. Although promising, the validation of results and the calculation of the mass transfer coefficient from experiments make

the adaptation of the model to other reactor configurations and conditions difficult. Sun *et al.* (2021) proposed an extension of the ADM1, considering biochemical reactions for the CO contained in the syngas under mesophilic conditions in continuous mode. In their work, a volumetric mass transfer equation was developed from the two-film theory to describe the mass transfer process. This consideration makes it difficult to apply the model to different process conditions and increases the number of parameters in the process.

Both models are a good basis for modeling the biological methanation process. However, they present deficiencies in considering the type of reactor, which directly affects how the mass transfer phenomenon is represented. This is highly important since H_2 transfer is one of the limiting factors of biological methanation (Ngu *et al.*, 2022). Additionally, the model from Sun *et al.* (2021) does not present the dynamic behavior of components, such as CO_2, which complicates the comprehensive analysis and the closing of the carbon balance in the biogas produced from AD.

This work aims to generate advances in understanding the dynamics of the biological methanation process by extending the Anaerobic Digestion Model No. 1 (ADM1_ME) to consider *in-situ* syngas addition. The model could be adapted to different substrates: agro-industrial waste, sludge, or sugars, and considers the addition of syngas at mesophilic and thermophilic conditions. The ADM1_ME represents the dynamic behavior of CH_4, H_2, CO, and CO_2 in liquid and gas phases, the inclusion of CO as a substrate of the process, and the adaptation of the volumetric mass transfer coefficient for two different configurations of bioreactor: bubble column reactor (BCR) and continuous stirred tank reactor ($CSTR$). This new model is based on the need to have a generic model for the biological methanation process with the capacity to be adapted to mesophilic and thermophilic conditions, to different substrates, and to allow the dynamic analysis of the volumetric mass transfer coefficients for different types of reactors.

2.2 Description of the Biological Methanation Process

Biological methanation is a process in which the biogas produced through the well-known AD is upgraded by the biological conversion of CO_2 and syngas to obtain high-purity CH_4 (Rafrafi *et al.*, 2020). In AD, the organic matter, such as agricultural residues, organic effluents from the food industry, animal manure, or waste/wastewater residues, are transformed through the synergistic work of a variety of microorganisms into a mixture of CH_4 and CO_2 through four steps: (i) hydrolysis, (ii) acidogenesis, (iii) acetogenesis and (iv) methanogenesis (Dar *et al.*, 2021). The biogas produced in the AD contains between 50 - 75% of CH_4, 25 – 50 % of CO_2, and 2–7% water vapor (Iglesias *et al.*, 2021; Laguillaumie *et al.*, 2022; Zupančič *et al.*, 2022). Hereby, the process is extended to biological methanation, which includes CH_4 production from a gas load, converting an inlet flow of H_2 and CO into high-purity CH_4 (Sun *et al.*, 2021).

2.2.1 Hydrolysis

In this step, the fermentative bacteria release enzymes that transform complex organic polymers (carbohydrates, proteins, and lipids) into soluble monomers, such as monosaccharides, amino acids, and long-chain fatty acids (LCFA). This process generally takes place on the surface of the acidogenic bacteria as it involves exo-enzymes secreted by hydrolytic bacteria, such as *Clostridia, Bacteroides, Fusobacterium, Butyrivibrio, Micrococci, Streptococcus*, and *Selenomonas* (Chandra *et al.*, 2012; Czatzkowska *et al.*, 2020).

2.2.2 Acidogenesis

Throughout the acidogenesis, the dissolved monomers or oligomers, amino acids, LCFA, in general, the components produced in the hydrolysis step undergo degradation reaction. These components are diffused into the acidogenic bacteria through the cell membrane and later fermented or anaerobically oxidized (Henze *et al.*, 2019) and produce mainly volatile fatty acids (VFA), such as propionate, butyrate, valerate, acetate, as well as new cell material. This step is carried out by the action of bacteria of the genera *Bacillus sp., Pseudomonas sp., Clostridium sp.,* and *Bifidobacterium sp.* (Czatzkowska *et al.*, 2020; Dar *et al.*, 2021).

2.2.3 Acetogenesis

The VFA produced in the acidogenesis are reduced and transformed into acetate, H_2, and CO_2, as well as in new cellular material by the action of bacteria of the genera *Clostridium, Syntrophomonas sp., Syntrophobacter sp.* (Chandra *et al.*, 2012; Czatzkowska *et al.*, 2020).

2.2.4 Methanogenesis

Acetate and H_2 are converted into CH_4 and CO_2, as well as in new cellular material in two types of processes, hydrogenotrophic methanogenesis, and acetoclastic methanogenesis, by the strictly anaerobic methanogens of the order Euryarchaeota: *Methanobacteriales, Methanococcales, Methanomicrobiales, Methanosarcinales,* and *Methanocellales*.

In hydrogenotrophic methanogenesis, the CO_2 is reduced into CH_4 using H_2 as a reduction agent (Ashraf *et al.*, 2020) by the action of hydrogenotrophic methanogens, *Methanobacterium, Methanogenium, Methanocorpusculum, Methanothermobacter* and *Methanosarcina.* In acetoclastic methanogenesis, the acetate is decarboxylated and converted into CH_4 by the action of acetoclastic methanogens, *Methanosaeta, Methanococcoides,* and *Methanosarcina* (Bharathiraja *et al.*, 2016; Czatzkowska *et al.*, 2020; Dar *et al.*, 2021; Dev *et al.*, 2019; Henze *et al.*, 2019; Saha *et al.*, 2020).

2.2.5 Biological Methanation

Biological methanation occurs in the last stage of AD. The syngas loading (commonly a combination of H_2:CO:CO_2) can be used to improve the process and convert the H_2 and CO_2 into CH_4 (Rusmanis *et al.*, 2019). The hydrogenotrophic methanogens with CO_2 consumption transform the H_2. Although this is a well-known route, CO consumption is still unclear (Sun *et al.*, 2021). The CO can be transformed indirectly into H_2 by carboxydotrophic hydrogenogenesis (water gas shift), then into acetate by CO-acetogenesis or CO-homoacetogenesis, and finally transformed into CH_4 through the hydrogenotrophic and acetoclastic methanogenesis (Guiot *et al.*, 2011). Table 2.2.1 presents a summary of the reactions involved in the biological methanation process.

Table 2.2.1. Reactions involved in biological methanation. Adapted from (Angelidaki *et al.*, 2011; Ashraf *et al.*, 2020; Liu *et al.*, 2016; Pan *et al.*, 2021; Rafrafi *et al.*, 2020; Sun *et al.*, 2021).

Reactions	$\Delta G°'(kJ/mol)$
Acidogenesis reactions:	
Acetate: $C_6H_{12}O_6 + 2H_2O \rightarrow 2\,CH_3COOH + 4H_2 + 2CO_2$	-206
Butyrate: $C_6H_{12}O_6 \rightarrow 2\,CH_3CH_2CH_2COOH + 2H_2 + 2CO_2$	-254
Propionate: $C_6H_{12}O_6 + 2H_2 \rightarrow 2\,CH_3CH_2COOH + 2H_2O$	-279.4
Acetogenesis reactions:	
Propionate: $CH_3CH_2COOH + 2H_2O \rightarrow CH_3COOH + 3H_2 + CO_2$	+76
Butyrate: $CH_3CH_2CH_2COOH + 2H_2O \rightarrow 2CH_3COOH + 2H_2$	+48.4
Methanogenesis reactions:	
Hydrogen: $4H_2 + CO_2 \rightarrow CH_4 + 2H_2O$	-130.7
Acetate: $CH_3COOH \rightarrow CH_4 + CO_2$	-31.0
Carbon monoxide reactions:	
$4CO + 2H_2O \rightarrow CH_3COOH + 2\,CO_2$	-176
$2CO + 2H_2 \rightarrow CH_3COOH$	-67
$CO + H_2O \rightarrow CO_2 + H_2$	-20

The biological methanation process can be performed in two ways: *in-situ* (directly in the AD reactor) or *ex-situ* (in a separate unit). The advantage of *in-situ* biological methanation is the reduction of infrastructure costs due to the use of only one reactor where syngas is directly added, allowing AD and biological methanation to occur simultaneously. The main drawback of *in-situ* biological methanation is that AD can be affected by the high concentration of gas. For instance, the H_2 added into the system can inhibit the early stages of the process. On the other hand, *ex-situ* biological methanation takes place in a separate external reactor, typically adapted to suit the hydrogenotrophic methanogens. Regardless of the type of configuration, microorganisms and metabolic pathways are similar (Mulat *et al.*, 2017; Rafrafi *et al.*, 2020; Rusmanis *et al.*, 2019).

2.3 Anaerobic Digestion Model Extension (ADM1_ME)

In this work, an extension of the Anaerobic Digestion Model No. 1 (ADM1_ME) is proposed based on the biological methanation scheme presented in Figure 2.2.1. The scheme includes acidogenesis, acetogenesis, methanogenesis, and its extension to consider CO and H_2 addition.

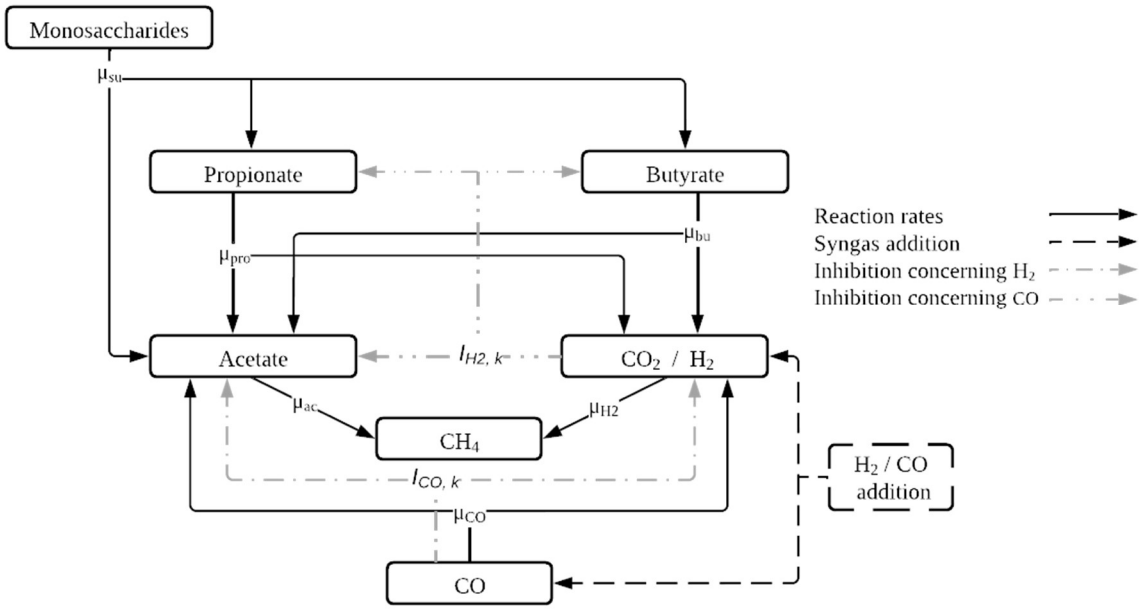


Figure 2.2.1. Biological methanation scheme implemented in the ADM1_ME including: (μ_{su}) acidogenesis from sugars, (μ_{pro}) acetogenesis from propionate, (μ_{bu}) acetogenesis from butyrate, (μ_{ac}) acetoclastic methanogenesis, (μ_{H2}) hydrogenotrophic methanogenesis, and (μ_{CO}) acetogenesis and carboxydotrophic hydrogenogenesis from carbon monoxide. $I_{H2,k}$ concerns inhibition with respect to H_2, and $I_{CO,k}$ concerns inhibition with respect to CO.

2.3.1 Mass Balances

Mass balances are derived from the ADM1 proposed by Batstone *et al.* (2002), which considers disintegration, hydrolysis, and uptake of the various components as well as biomass decay. Differently from ADM1, the model extension ADM1_ME considers the uptake of sugar, volatile fatty acids (*e.g.*, butyrate, propionate, and acetate), the uptake of H_2 and CO, and

biomass decay. The model describes three types of variables: soluble ($S_{liq,j}$), particulate (X_k) and gas ($S_{gas,i}$) components. Particulate components are considered to be part of the biomass as in the ADM1 (see Annexes Section 1 for more details). The model is here rewritten as Equations (2.2.1) – (2.2.3).

State variables in the liquid phase ($S_{liq,j}$):

$$\frac{dS_{liq,j}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(S_{liq,j}^{in} - S_{liq,j}\right) + \sum_{k} Y_k f_{j,k} \mu_k - N_i \qquad (2.2.1)$$

State variables in biomass (X_k):

$$\frac{dX_k}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(X_k^{in} - X_k\right) + Y_k \mu_k - \mu_{k,dec} \qquad (2.2.2)$$

State variables in the gas phase ($S_{gas,i}$):

$$\frac{dS_{gas,i}}{dt} = \frac{q_{gas}^{in}}{V_{gas}} S_{gas,i}^{in} + N_i \left(\frac{V_{liq}}{V_{gas}}\right) - \frac{q_{gas}}{V_{gas}} S_{gas,i} \qquad (2.2.3)$$

Sub-index j ϵ [1,8] denotes glucose, butyrate, propionate, acetate, H_2, CH_4, CO, and CO_2 in the liquid phase. The H_2, CH_4, and CO are expressed in $gCOD/L$, and CO_2 is expressed in mol/L. Chemical Oxygen Demand (COD) is the amount of oxygen needed to degrade the organic matter into CO_2 and H_2O. It is important to mention that CO_2 would be expressed in $moles$ instead of COD, as suggested in (Batstone *et al.*, 2002). Sub-index k ϵ [1,6] reads for the biomass that degrade glucose, butyrate, propionate, acetate, H_2, and CO, respectively. For the gas phase, the sub-index i ϵ [1,4] corresponds to H_2, CH_4, CO, and CO_2. The inlet flow rates of liquid and gas are represented by q_{liq}^{in} and q_{gas}^{in}, respectively, while q_{gas} denotes the outlet gas flow rate. V_{liq} and V_{gas} are the liquid and gas volumes, respectively, $S_{liq,j}^{in}$, $S_{gas,i}^{in}$, and X_k^{in} hold for the inlet concentration of the component j in the liquid phase, the inlet concentration of component i in gas phase, and the inlet concentration of biomass k in the liquid phase. Y_k is the yield of biomass k, $f_{j,k}$ the stoichiometric coefficients; μ_k and $\mu_{k,dec}$ the growth and decay rate of biomass k, and N_i the mass transfer rate of component i.

Table 2.2.2 Petersen matrix for soluble component in the ADM1_ME.

	Component	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**Reaction rate**
	Process	$S_{liq,su}$	$S_{liq,bu}$	$S_{liq,pro}$	$S_{liq,ac}$	$S_{liq,co}$	$S_{liq,H2}$	$S_{liq,CH4}$	$S_{liq,co2}$	X_{su}	X_{bu}	X_{pro}	X_{ac}	X_{co}	X_{H2}	
1	Uptake of sugar	-1	$(1-Y_{su})f_{bu,su}$	$(1-Y_{su})f_{pro}$	$(1-Y_{su})f_{ac,su}$		$(1-Y_{su})f_{H2,su}$		$-\sum_{k=1}^{6}\left(\sum_{j=1}^{9} C_j v_{j,su}\right)$	Y_{su}						μ_{su}
2	Uptake of butyrate		-1		$(1-Y_{bu})f_{ac,bu}$		$(1-Y_{bu})f_{H2,bu}$		$-\sum_{k=1}^{6}\left(\sum_{j=1}^{9} C_j v_{j,bu}\right)$		Y_{bu}					μ_{bu}
3	Uptake of propionate			-1	$(1-Y_{pro})f_{ac,pro}$		$(1-Y_{pro})f_{H2,pro}$		$-\sum_{k=1}^{6}\left(\sum_{j=1}^{9} C_j v_{j,pro}\right)$			Y_{pro}				μ_{pro}
4	Uptake of acetate				-1			$(1-Y_{ac})$	$-\sum_{k=1}^{6}\left(\sum_{j=1}^{9} C_j v_{j,ac}\right)$				Y_{ac}			μ_{ac}
5	Uptake of CO				$(1-Y_{CO})f_{ac,CO}$	-1	$(1-Y_{CO})f_{CO}$							Y_{CO}		μ_{CO}
6	Uptake of H_2						-1	$(1-Y_{H2})$	$-\sum_{k=1}^{6}\left(\sum_{j=1}^{9} C_j v_{j,H2}\right)$						Y_{H2}	μ_{H2}
7	Decay of X_{su}									-1						$\mu_{su,dec}$
8	Decay of X_{bu}										-1					$\mu_{bu,dec}$
9	Decay of X_{pro}											-1				$\mu_{pro,dec}$
10	Decay of X_{ac}												-1			$\mu_{ac,dec}$
11	Decay of X_{CO}													-1		$\mu_{CO,dec}$
12	Decay of X_{H2}														-1	$\mu_{H2,dec}$

In Equation (2.2.1), the term $-\sum_{k=1}^{6}\left(\sum_{j=1}^{7} C_j v_{j,k} \mu_k\right)$ is introduced in the CO_2 mass balance for the liquid phase (Annexes Section 1.2, Equation (4.1.9)), based on the original version of the ADM1 (Batstone *et al.*, 2002) (without considering the components associated with the hydrolysis). This term describes the fractionation of inorganic carbon, the composition of various species, and a standard biomass composition (Rosen and Jeppsson, 2006). Table 2.2.2 summarizes the stoichiometric coefficients associated with the variables in liquid phase and biomass growth.

2.3.2 Mass Transfer Rate Definition

The gas-liquid mass transfer rates are expressed as in Equation (2.2.4) to relate the liquid and gas balances.

$$N_i = k_L a_{,i}\left(S_{liq,j} - \gamma_{COD,i} H_i P_{gas,i}\right) \qquad (2.2.4)$$

where N_i is the flux of species H_2, CH_4, and CO, expressed as $gCOD/L/d$, and CO_2, expressed as $mol/L/d$. $k_L a_{,i}$ is the volumetric mass transfer coefficient of component i, and $\left(S_{liq,j} - \gamma_{COD,i} H_i P_{gas,i}\right)$ is the driving force. H_i, and $P_{gas,i}$ are Henry's law equilibrium constant and partial pressure of component i, respectively. $\gamma_{COD,i}$ is a conversion factor between the $moles$ and $gCOD$ of a component i, *e.g.*, 16 $gCOD/mol$ for H_2 and CO, and 64 $gCOD/mol$ for CH_4.

The volumetric mass transfer coefficient ($k_L a$) was set constant (200 $1/d$) in the ADM1 (Batstone *et al.*, 2002). However, for the biological methanation process, it is necessary to integrate the effect of the addition of gas on mass transfer. Depending on the reactor configuration, well-established correlations for BCR Equation (2.2.5) and $CSTR$ Equation (2.2.6) have been used to estimate $k_L a$. Both rely on the superficial gas velocity defined as the inlet gas flow rate divided by the cross-sectional area of the vessel ($U_G = q_{gas}^{in}/A$). If there is no addition of gas to the system, U_G is zero. In a BCR, the sparger type is crucial to ensure efficient gas distribution (Nauman, 2008). For a $CSTR$, the energy injected in the system controls the bubble size and turbulence, thus correlating with the mass transfer coefficient. This

energy dissipation per unit of volume in the presence of gas, the so-called gassed power input ($Power$), depends on the impeller type, rotation speed and aeration number (Gary, B. Tatterson, 1991; Liu *et al.*, 2019).

2.3.2.1 Bubble Column Reactor (BCR) Volumetric Mass Transfer Coefficient

The volumetric mass transfer coefficient in a BCR can be defined as,

$$k_L a_{,i} = b_{0,OP1} U_G^{b_{1,OP1}} \quad (2.2.5)$$

U_G is the superficial gas velocity, $b_{0,OP1}$ and $b_{1,OP}$ are parameters that can be affected by the liquid phase properties and type of sparger. Values of 0.467 and 0.82 were respectively proposed by Deckwer *et al.* (1983) for tap water and salt solutions with U_G ranging between 0.002-0.08 m/s.

2.3.2.2 Continuous Stirred Tank Reactor ($CSTR$) Volumetric Mass Transfer Coefficient

The volumetric mass transfer coefficient in a $CSTR$ can be defined as,

$$k_L a_{,i} = b_{0,OP2} \left(\frac{Power}{V_{liq}}\right)^{b_{1,OP2}} U_G^{b_{2,OP2}} \quad (2.2.6)$$

$b_{0,OP2}$, $b_{1,OP}$, and $b_{2,OP2}$ are constant parameters. The values of $b_{1,OP2}$ and $b_{2,OP2}$ present variations: $0.4 < b_{1,OP2} < 1$ and $0 < b_{2,OP2} < 0.7$. The value of $b_{0,OP2}$, however, is not reported in most cases, but it is highly correlated to $b_{1,OP2}$ because $(Power/V_{li})$ range from 1×10^3 - 1×10^4 (Van't, 1979). This means that if small variations are introduced into $b_{1,OP}$, $b_{0,OP}$ would show large changes.

The mass transfer coefficient k_L is proportional to the square root of the diffusivity in the liquid phase (Higbie, 1935). In a biological methanation process, some substrates (H_2, CO, and CO_2) are initially present in a gaseous form and must thus transfer from the gas to the liquid phase before getting involved in the biochemical reaction. As a result, the biochemical reaction rates might be, in the end, limited by the mass transfer rate. When comparing different gases,

considering both the $k_L a$ and the solubility it appears that the lowest value is that of $k_L a_{H2}$. For this reason, it can be claimed that if mass transfer becomes the limiting phenomena, it will be the mass transfer rate of H_2 which limit the bioreaction rate (Jensen *et al.*, 2021; Ngu *et al.*, 2022). Therefore, only $k_L a_{H2}$ was calculated using Equation (2.2.5) and Equation (2.2.6). The rest were calculated as,

$$k_L a_{,i} = k_L a_{H2} \sqrt{\frac{D_i}{D_{H2}}} \qquad (\ 2.2.7\ )$$

In this equation, the sub-index $i \in [1,4]$ corresponds to H_2, CH_4, CO, and CO_2. D_{H2} is the diffusion of H_2 and D_i is the diffusion of CH_4, CO, and CO_2.

An interesting analysis can be performed by increasing U_G in the BCR, and the $Power/V_{liq}$ in the $CSTR$. For the case of the BCR, it was assumed that the organic loading and gas loading rates were the same as proposed in the operational condition one (OP1) (Section 2.5, Table 2.2.4). Figure 2.2.2-A displays the variation of $k_L a_{,i}$ with the change of U_G in a range from 173 to 6912 m/d, *i.e.*, the boundaries indicated by Deckwer *et al.* (1983) ranged between 0.002 and 0.08 m/s. All the $k_L a_{,i}$ increased with a different rate, *i.e.*, $k_L a_{,H2}$ was the higher, followed by $k_L a_{,CO}$, $k_L a_{,CO2}$, and $k_L a_{,CH4}$. For the case of the $CSTR$, the organic loading and gas loading rates were assumed to be the same as proposed in operational condition two (OP2) (Section 2.5, Table 2.2.4). The q_{gas}^{in} was fixed at the maximum value proposed for OP2. Figure 2.2.2-B shows the variation of the $k_L a_{,i}$ concerning $Power/V_{liq}$, which was varied from 1 to 1×10^5 W/L, *i.e.*, the boundaries indicated by Van't, (1979) ranged between 1×10^3 and 1×10^5 W/m^3 m/s. Similar to the BCR, all the $k_L a_{,i}$ increased at a different rate, *i.e.*, $k_L a_{,H2}$ displayed the largest value, followed by $k_L a_{,CO}$, $k_L a_{,CO}$, and $k_L a_{,CH}$.

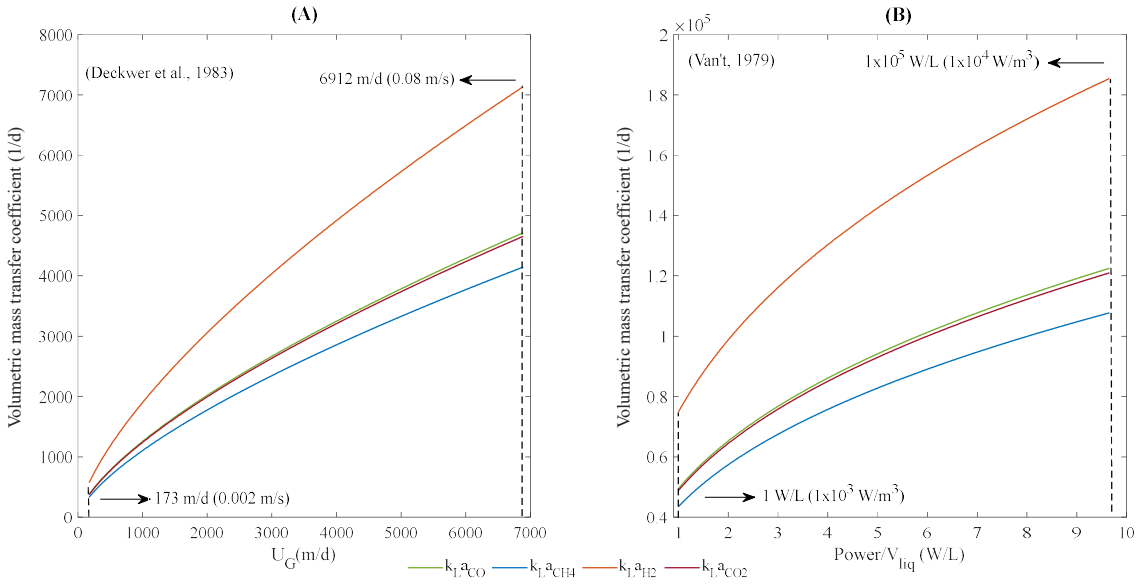


Figure 2.2.2. Variation of the volumetric mass transfer coefficient: (A) concerning the superficial gas velocity U_G for a BCR (Equation (2.2.5)), and (B) with respect to the $Power/V_{liq}$ for a $CSTR$ (Equation (2.2.6)).

2.3.3 Constitutive Equations

The total (q_{gas}) and specific $(q_{gas,i})$ outlet gas flow rates can be calculated with Equations (2.2.8) and (2.2.9) (Rosen and Jeppsson, 2006).

$$q_{gas} = K_P(P_{gas} - P_{atm}) \qquad (2.2.8)$$

$$q_{gas,i} = q_{gas}\left(\frac{P_{gas,i}}{P_{gas}}\right) \qquad (2.2.9)$$

In these equations, the sub-index i ϵ [1,4] corresponds to H_2, CH_4, CO, and CO_2. K_P is a parameter related to the friction in the gas outlet, *i.e.*, this parameter fixes the pressure drop at the gas outlet. P_{atm}, P_{gas} and $P_{gas,i}$, are the atmospheric, total, and partial pressure of component i, respectively. The term $(P_{gas,i}/P_{gas})$ is the molar fraction of component i, in which the partial $(P_{gas,\,i})$ and total pressure (P_{gas}) are calculated by using Equations (2.2.10)-(2.2.11).

$$P_{gas,i} = \frac{S_{gas,i} RT}{\gamma_{COD,i}} \quad (2.2.10)$$

$$P_{gas} = \sum P_{gas,i} \quad (2.2.11)$$

where R and T are the Universal gas constant and the temperature of the process, respectively.

For each component i in the gas phase, the percent composition $(p_{gas,\,i})$ was calculated based on the molar fraction, Equations (2.2.12).

$$p_{gas,i} = \left(\frac{P_{gas,i}}{P_{gas}}\right) \cdot 100\% \quad (2.2.12)$$

To model the conversion rates process, the biochemical reaction rates (μ_k) are considered by Monod kinetics for substrate consumption with inhibition, while biomass decay $(\mu_{k,dec})$ follows first-order kinetics. This allows the representation of the cell growth and death associated with each biomass involved in the process. Their mathematical representation is expressed as,

$$\mu_k = \frac{\mu_{m,k} S_{liq,j}}{Ks_k + S_{liq,j}} X_k I_{H2,k} I_{CO,k} \quad (2.2.13)$$

$$\mu_{k,dec} = K_{k,dec}\, X_k \quad (2.2.14)$$

where $\mu_{m,k}$ is the maximum specific growth rate, Ks_k the saturation constant, and $K_{k,dec}$ the decay biomass constant. The term $I_{H2,k}$ in Equation (2.2.13) concerns inhibition respect to H_2, which affects the reaction rates of butyrate, propionate, and acetate. The term $I_{CO,k}$ concerns inhibition respect to CO, which occurs for the reaction rates of acetate and H_2. Equations (2.2.15) – (2.2.16) represent these inhibitions.

$$I_{H2,k} = \frac{1}{1 + S_{liq,H2}/KI_{H2,k}} \quad (2.2.15)$$

$$I_{CO,k} = \frac{1}{1 + S_{liq,CO}/KI_{CO,k}} \quad (2.2.16)$$

$KI_{H2,k}$ are the inhibition constants over butyrate, propionate, and acetate due to H_2, meanwhile $KI_{CO,k}$ are the inhibition constants of the effect of CO over acetate and H_2.

2.4 Global Sensitivity Analysis Techniques

Biological kinetic models are constituted of several parameters, *e.g.*, stoichiometric, biochemical, and physicochemical parameters, which present a certain degree of uncertainty. Sensitivity Analysis (SA) is a powerful alternative to determine how the uncertainty of the model inputs or parameters influence the outputs (Damblin *et al.*, 2013; Sepulveda *et al.*, 2013; Sohier *et al.*, 2014; Tosin *et al.*, 2020). Sensitivity Analysis was developed to identify the contribution of each uncertainty of the inputs on the outputs (Feng *et al.*, 2019; Kucherenko *et al.*, 2015; Zhang *et al.*, 2015). Techniques for SA can be classified into local and global (Morio, 2011; Ochoa *et al.*, 2016). In the context of this work, we will focus on Global sensitivity analysis (GSA).

In GSA, the model is studied from a probabilistic point of view. The exploration of the entire range of variation of the model parameters is considered using a probability density function associated with each input parameter and repeated simulations of the model (Iooss and Lemaître, 2015; Ochoa *et al.*, 2016; Tosin *et al.*, 2020). GSA could employ regression, screening, and variance-based methods (Sepúlveda *et al.*, 2014).

2.4.1 Sobol's Method

The Sobol method (Sobol′, 2001) is an interesting variance-based method in which the variance of the model output can be decomposed into partial variances that represent the contribution of

the inputs over the overall uncertainty of the model output (Morio, 2011; Ochoa *et al.*, 2016; Sepulveda *et al.*, 2013; Sobol′, 2001; Tosin *et al.*, 2020).

Consider the model define by ξ, Equation (2.2.17).

$$Y = \left[S_{liq,j}, \mathrm{X}_k, S_{gas,i}\right] = \xi(\boldsymbol{\theta}) \qquad (\ 2.2.17\ )$$

where $Y \in \mathcal{R}^m$ is the model output of interest, and $\boldsymbol{\theta} \in \mathcal{R}^n$ is a n-dimensional parameter vector defined as $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)$ and characterized by a probability density function (PDF).

The function $\xi(\boldsymbol{\theta})$ can be decomposed into summands of different dimensions, Equation (2.2.18).

$$Y = \xi_0 + \sum_{i=1}^{n} \xi_i(\theta_i) + \sum_{1 \le i < j \le n} \xi_{ij}(\theta_i, \theta_j) + \cdots + \xi_{1\dots n}(\theta_i, \dots, \theta_n) = \sum_{\mathbf{u} \subseteq \{1 \dots p\}} \xi_{\mathbf{u}}(\theta_{\mathbf{u}}) \qquad (\ 2.2.18\ )$$

where:

$$\begin{cases} \xi_0 = \mathbb{E}[Y] \\ \xi_i(\theta_i) = \mathbb{E}[Y|\theta_i] - \xi_0 \\ \xi_{ij}(\theta_i, \theta_j) = \mathbb{E}\left[Y\middle|\theta_i, \theta_j\right] - \xi_0 - \xi_i - \xi_j \end{cases} \qquad (\ 2.2.19\ )$$

ξ_0 is the mean of the function, $\xi_i(\theta_i)$ and $\xi_{ij}(\theta_i, \theta_j)$ are the expectation terms of increasing order and the conditional expectations defined recursively. This decomposition is unique, provided the inputs are independent, and the individual terms are square integrable. Squaring Equation (2.2.18) and integrating, we can get the so-called ANOVA decomposition as,

$$V(Y) = \sum_{i=1}^{n} V_i(\theta_i) + \sum_{1 \le i < j \le n} V_{ij}(\theta_i, \theta_j) + \cdots + V_{1\dots n}(\theta_i, \dots, \theta_n) \sum_{\mathbf{u}} V(\xi_{\mathbf{u}}(\theta_{\mathbf{u}})) \quad \text{for } \mathbf{u} \subset \{1, \dots n\} \qquad (\ 1.2.20\ )$$

where $V(\xi_{\mathbf{u}}(\theta_{\mathbf{u}}))$ express the conditional variance for the subvector $\theta_{\mathbf{u}}$, containing the variables whose indices are indicated by the subset $\mathbf{u}$.

The variance of the output can be decomposed into terms depending on the parameters and their interactions. The variance-based sensitivity index (SI) associated with the subset $\mathbf{u}$ is defined as the ratio between the contribution given by the interaction among the components of $\mathbf{u}$ for the model variance and the total variance, Equation (2.2.21).

$$SI_{\mathbf{u}} = \frac{V\left(\xi_{\mathbf{u}}(\theta_{\mathbf{u}})\right)}{V(Y)} \qquad (2.2.21)$$

Based on this, for $\mathbf{u} \subset \{1, \dots n\}$, and $\mathbf{u} \neq \mathbf{0}$,

$$\sum_{\mathbf{u}} SI_{\mathbf{u}} = \sum_{i=1}^{n} SI_i + \sum_{1 \le i < j \le n} SI_{ij} + \cdots + SI_{1 \dots n} = 1 \qquad (2.2.22)$$

The term SI_i is the first-order sensitivity index, which measures the fraction of the total output variance explained by the parameter θ_i alone as,

$$SI_i = \frac{V\left(\xi_i(\theta_i)\right)}{V(Y)} \qquad i = 1, \dots, n \qquad (2.2.23)$$

Similarly, SI_{ij} is the second order-sensitivity index that measures the amount of variance caused by the interaction between the parameters θ_i and θ_j as,

$$SI_{ij} = \frac{V\left(\xi_{ij}(\theta_{ij})\right)}{V(Y)} \quad 1 \le i < j \le n \qquad (2.2.24)$$

It is possible to construct the *SI* for all orders until the n^{th} order index $SI_{1 \dots n}$, which represents the contribution of the interactions between all the parameters in $\boldsymbol{\theta}$. To measure the full contribution of the i^{th} random parameter θ_i for the total variance either by its single effect or by its interaction with others, we use the total Sobol' indices,

$$SI_i^T = \sum_{\substack{\mathbf{u}\subset\{1,\dots,n\} \\ i\in\mathbf{u}}} SI_{\mathbf{u}} \qquad i = 1,\dots,n \tag{2.2.25}$$

Equation (2.2.25) indicates that the total sensitivity index does not only include the marginal contribution of θ_i to the variance of the output, but it also contains its cooperative contribution with all the other inputs. The total sensitivity index removes the parameter θ_i from the analysis and allocates the resulting variance reduction to that parameter.

2.4.2 Morris Method

The Morris method is the most well-known (Morris, 1991) screening method which performs SA by analyzing one-factor-at-a-time (OAT). This is generally used when the number of model parameters is large, and the computation of model simulations is expensive. This method provides qualitative sensitivity measures, ranking the factors by their importance. Nevertheless, it does not quantify the importance of one factor concerning another (Saltelli, 2004). The Morris method applied to parameter sensitivity discretizes the space of each parameter and performs a given number of OAT designs. These designs and variation directions are randomly chosen from the parameter space. The repetition of these steps allows the estimation of elementary effects $\left(EE_i^j\right)$ for each parameter i, which represents the relative difference between the outputs and the j^{th} parameter disturbance (Feng *et al.*, 2019; Iooss and Lemaître, 2015; Morio, 2011; Morris, 1991; Saltelli, 2008).

Consider a trajectory in the parameter space as,

$$\theta_i^{j+1} = \theta_i^j + e^j \Delta^j \qquad \mathrm{j} = 1,\dots,\mathrm{r} \tag{2.2.26}$$

where $\mathrm{j} = 1,\dots,\mathrm{r}$ corresponds to the number of repetititons and $\boldsymbol{\theta} \in \mathcal{R}^n$ is an n-dimensional parameter vector defined as $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)$.

The effect of parameter variation can be evaluated by estimating the difference between the model output with the actual parameter θ_i^j and the updated parameter $\theta_i^j + e^j\Delta^j$ over a given increment Δ^j. e^j is a vector of zeros but with a unit as its j^{th} component (canonical base). This variation is referred as elementary effects, which can be calculated as follows,

$$EE_i^j = \frac{\xi\left(\theta_i^j + e^j\Delta^j\right) - \xi\left(\theta_i^j\right)}{\Delta^j} \quad (2.2.27)$$

where θ_i^j is a sample of input $\boldsymbol{\theta}$ and $\xi\left(\theta_i^j\right)$ is the corresponding model output. Δ^j is a step between two consecutive input space points of the trajectory. The term $\left(\theta_i^j + e^j\Delta^j\right)$ represents a new sample by moving the i^{th} parameter input from θ_i^j to $\theta_i^j + \Delta^j$, with the respective model output $\xi\left(\theta_i^j + e^j\Delta^j\right)$.

The index j of EE_i^j expresses the ratio of the change of the output Y when the i^{th} parameter θ_i^j is given a particular change Δ^j. Then, EE_i^j can measure the effect of θ_i^j in a given scope of output Y. The sensitivity measures are expressed in terms of means μ_i^*, and standard deviations σ_i^j are defined as Equations (2.2.28) and (2.2.29).

$$\mu_i^* = \frac{\sum_{j=1}^{r}\left|EE_i^j\right|}{r} \quad (2.2.30)$$

$$\sigma_i^j = \sqrt{\frac{\sum_{j=1}^{r}\left(EE_i^j - \mu_i^*\right)^2}{\mathrm{r} - 1}} \quad (2.2.31)$$

where EE_i^j is the elementary effect of the i^{th} parameter obtained at the j^{th} repetition. The sensitivity measures μ_i^j and σ_i^j are the mean of the absolute value and standard deviation of the distribution of the elementary effects, respectively. μ_i^* measures the influence of the i^{th} parameter on the output. σ_i^j is a measure of non-linear and interaction effects of the i^{th} parameter. A high value of μ_i^* indicates that the parameter θ_i^j has a more important effect on

the output. A high value of σ_i^j indicates that the elementary effect of θ_i^j varies significantly from one to another, which shows that the value of EE_i^j is strongly influenced by the selected sample points.

2.5 Bioreactors Operating Conditions

This study used two experimental datasets from the literature to derive the model extension. The first dataset was taken from (Sun *et al.*, 2021), where the experiments were developed at operational condition one (OP1). The second dataset was obtained from (Andreides *et al.*, 2022), whose experiment was carried out at operational condition two (OP2). Both operating conditions are reported in Table 2.2.3.

The measured outputs were different for each operating condition. For OP1, these corresponded to the outlet flow rates of CH_4, H_2, and CO. For OP2, the measured outputs were the percent of CH_4, H_2, CO, and CO_2 in the gas phase.

Table 2.2.3. Operational conditions from OP1 (Sun *et al.*, 2021) and OP2 (Andreides *et al.*, 2022).

Operational conditions	OP1	OP2
Reactor type	BCR	$CSTR$
Temperature $(°C)$	37	55
Working volume (L)	37.5	10.5
HRT (d)	20	21
Experimental time (d)	207	150
Inlet liquid flow rate (L/d)	1.9	0.5

HRT: hydraulic retention time.

In OP1, the loading consisted of two additions: glucose and syngas. The organic loading rate (OLR) of glucose was kept at 0.5 $g/L_r/\,d$, where L_r represents the volume of the reactor. Syngas containing 50% v/v of H_2 and CO ($H_2/CO \approx 1$) was added into the reactor with a continuous flow but at different rates after the first reference stage. For OP2, the liquid fraction

of the OLR was a mixture of primary sludge and activated ticked-disintegrated waste (volume ratio of 3:1), which was fed with a flow rate varying in time. Syngas was also added into the reactor at diverse rates after a reference stage. In this case, however, the syngas contained 55% v/v of H_2 and 45% v/v CO ($H_2/CO \approx 0.55/0.45$). Table 2.2.4 reports the stages in which the substrates were added to the reactors.

Table 2.2.4. Syngas flow rate, gas loading rate, and organic loading rate from OP1 (Sun *et al.*, 2021) and OP2 (Andreides *et al.*, 2022).

Stage	Time (d)	Syngas flow rate (L/d)	Recirculation flow rate (L/h)	Gas loading rate ($L/\ L_r\ /d$)	Organic loading rate ($gCOD/L_r\ /\ d$)
			OP1		
Reference	1-32	0.0	0.0	0.0	0.53
I	33-64	7.5	3.75	0.2	
II	65-101	7.5	60	0.2	
III	102-135	15.0	120	0.4	
IV	136-171	37.5	120	1.0	
V	172-207	37.5	240	1.0	
			OP2		
Reference	1-36	0.0	-	0.0	3.08
I	36-51	3.15	-	0.3	3.72
II	51-81	7.35	-	0.7	3.24
III	81-118	10.5	-	1.0	3.09
IV	118-130	15.75	-	1.5	2.84

To standardize the units for both operating conditions, the OLR is expressed as $gCOD/L_r\ /d$ and the gas loading rate is expressed as $L/\ L_r\ /d$. In OP1, the OLR was expressed in $gCOD/L_r\ /d$ by multiplying the value 0.5 $gGlu/L_r\ /d$, by a factor of 1.07 $gCOD/gGlu$. In OP2, a mixture of different substrates was used. Those mixtures are usually expressed in terms of volatile solids (VS). Therefore, an equivalence between COD and VS must be estimated. Some authors mentioned 1.42 $gCOD/gVS$ for activated sludge (Ahnert *et al.*, 2021), while others indicated values between 1.6 – 1.7 $gCOD/gVS$ (Batstone *et al.*, 2010). In this case, an

estimation using the function *fmincon* from MATLAB® was performed by minimizing the *RMSE* with a lower and upper value of 1.42 and 1.7 $gCOD/gVS$, respectively. A value of 1.62 $gCOD/gVS$ was obtained. Additionally, a value of 0.025 $mol/gCOD$ was used for the CO_2 balance (lower than the one proposed for sugar 0.0313 $mol/gCOD$). To model the mass transfer for OP2, a value of 0.69 was used for $b_{1,OP1}$ in Equation (2.2.5) since the BCR structure is not known in detail.

2.6 Model Calibration and Validation

In this study, model simulations were implemented in MATLAB® and run using a computer with Intel® Core i7 8665U 2.11 GHz and 16 GB RAM. The ADM1_ME was calibrated using the measured outputs mentioned in the previous section, CH_4, H_2, and CO outlet flow rates for OP1, and CH_4 H_2, CO, and CO_2 percent in the gas phase for OP2.

Sobol's method was implemented using the toolbox: Global sensitivity and uncertainty analysis (GSUA) (Velez S. Carlos M., 2022), whereas the Morris method was performed by the toolbox: Sensitivity analysis-Morris method (advanced) (Mr, 2022). The two abovementioned SA methods were performed to identify the effects of a change in the parameters on the model outputs.

The ADM1_ME is described by 60 parameters comprising stoichiometric and kinetic parameters (see Annexes Section 2, Table 4.2.1). Among those parameters, a group of $n = 26$ parameters was selected to analyze their impact on the gas flow rates for OP1 and gases percent for OP2 (see Annexes Section 2, Table 4.2.1). Those comprised mostly kinetic parameters and the parameters where the estimation is more uncertain. The Morris method performs analysis upon several simulation runs, *e.g.*, $(n + 1) \cdot 10$. In this case, however, a larger number of simulation runs were executed $(n + 1) \cdot 100$, *i.e.*, 2700, to guarantee a good sampling in the distribution of the parameter domain. Since the Sobol' method is computationally expensive compared to the Morris method, only 200 simulation runs were evaluated per output.

The toolbox GSUA for Sobol' method allowed the computation of the first-order sensitivity indices with a scalar characteristic (SI_{is}) as,

$$SI_{is} = \xi(SSE) = \xi\left(\sum (Y(\theta_i) - Y(\theta_{i,nom}))^2\right) \quad (2.2.32)$$

where the sum squared error (SSE) was the scalar characteristic, measured between the output variable ($Y_m(\theta_i)$) calculated with the varied parameter θ_i, sampled with a uniform distribution, and the output variable ($Y_m(\theta_{i,nom})$) calculated with the nominal value of the parameter $\theta_{i,nom}$.

The most sensitive parameters found with SA were selected to recalibrate the model, while the rest of the parameters were fixed to the nominal values reported in the ADM1 (Batstone *et al.*, 2002).

Parameter estimation was performed to minimize the adapted root mean square error ($RMSE_{est}$) reported in Equation (2.2.33). The outputs for OP1 were the outlet gas flow rates, $q_{gas,\mathrm{CH4}}$, $q_{gas,\mathrm{H2}}$, and $q_{gas,\mathrm{CO}}$. Nevertheless, for OP2, the outputs were the gases percent, $p_{gas,CH4}$, $p_{gas,H2}$, $p_{gas,CO}$, and $p_{gas,CO2}$. Therefore, Equation (2.2.33) considers $q_{gas,i}$ for OP1 and $p_{gas,i}$ for OP2.

$$RMSE_{est} = \sqrt{\frac{W_{OP1}}{n_{OP1}}\sum_{i=1}^{3}\left(\frac{\left|\hat{q}_{gas,i} - q_{gas,i}\right|}{max\ (q_{gas,i})}\right)^2 + \frac{W_{OP2}}{n_{OP2}}\sum_{i=1}^{4}\left(\frac{\left|\hat{p}_{gas,i} - p_{gas,i}\right|}{max\ (p_{gas,i})}\right)^2} \quad (2.2.33)$$

Sub-index $i \in [1,4]$ corresponds to CH_4, H_2, CO, and CO_2. Note that only the first three are used for OP1. n_{OP1} and n_{OP2} are the number of observations in OP1 and OP2. W_{OP1} and W_{OP2} are the weights to trade-off the estimation of OP1 and OP2.

Model validation was carried out with the results from OP1 (stages IV and V) and OP2 (stages III and IV). The coefficient of determination (R^2), and the root mean squared error ($RMSE$) were used as criteria to qualify parameter estimation, Equations (2.2.32) – (2.2.33).

$$R_i^2 = 1 - \frac{\sum(\hat{Y} - Y)^2}{\sum(\hat{Y} - \hat{Y}_{mean})^2} \quad (2.2.34)$$

$$RMSE_i = \sqrt{\frac{1}{n_{OP1/OP2}} \sum(\hat{Y} - Y)^2} \quad (2.2.35)$$

In the former equations, i denotes CH_4, H_2, and CO for OP1 and CH_4, H_2, CO, and CO_2 for OP2. $n_{OP1/OP2}$ is the number of observations for the operating condition OP1 or OP2, $\hat{Y}$, $\hat{Y}_{mean}$, and Y are the experimental, mean, and model data, respectively.

Confidence intervals were determined for each estimated parameter by computing the Global Sensitivity Information Matrix ($GSIM$). This matrix is based on the first-order sensitivity indices (SI_i) of the Sobol' method and is analogous to the Fisher Information Matrix (FIM) (Asprey and Macchietto, 2000; Rodriguez-Fernandez *et al.*, 2007). The $GSIM$ is calculated as,

$$GSIM = \sum_{t=1}^{n_{OP1/OP2}} [Q^T(t_t)\, W_t^{-1}\, Q(t_t)] \quad (2.2.36)$$

where W_i^{-1} is a weighting matrix usually chosen as the measurement error covariance matrix, and $Q(t_t)$ is defined as,

$$Q(t_t) = \begin{bmatrix} SI_1^1(t_t) & SI_2^1(t_t) & \dots & SI_n^1(t_t) \\ SI_1^2(t_t) & SI_2^2(t_t) & \dots & SI_n^2(t_t) \\ \vdots & \vdots & \ddots & \vdots \\ SI_1^m(t_t) & SI_2^m(t_t) & \dots & SI_n^m(t_t) \end{bmatrix} \quad (2.2.37)$$

In this case, $SI_n^m(t_t)$ measures the sensitivity of the state Y_m concerning the parameter θ_n at the time t_t. Then the variance of each parameter θ_i can be approximated by $\sigma^2(\theta_i) \approx GSIM_{ii}^{-1}$ and used to evaluate the 95% confidence intervals as: $\theta_i \pm 1.96 \cdot \sigma(\theta_i)$ where $\sigma(\theta_i)$ is the standard deviation.

2.7 Results and Discussion

2.7.1 Sensitivity Analysis

The first-order sensitivity indices of Sobol' Method were computed for the 26 selected parameters and evaluated for each output and operating condition. Table 2.2.5 summarizes the results. The sum of the variance is considered to be around 100%.

Table 2.2.5. First-order sensitivity index with a scalar characteristic (SI_{is}) with the Sobol' method from OP1 (Sun *et al.*, 2021) and OP2 (Andreides *et al.*, 2022).

Parameter / Model outputs	**SI_{is} for OP1**				**SI_{is} for OP2**			
	$q_{gas,CH4}$	$q_{gas,H2}$	$q_{gas,CO}$	$q_{gas,CO2}$	$p_{\mathrm{gas,CH4}}$	$p_{\mathrm{gas,H2}}$	$p_{\mathrm{gas,CO}}$	$p_{\mathrm{gas,CO2}}$
Y_{su}	2.8	3.0	6.2	2.5	0.7	8.4	3.6	4.1
Y_{bu}	4.1	5.9	6.7	8.0	1.3	2.9	2.5	1.7
Y_{pro}	5.6	6.4	3.5	2.6	1.4	1.5	3.5	3.2
Y_{ac}	4.4	7.2	1.8	2.4	1.4	5.3	4.7	2.2
Y_{CO}	3.0	2.5	5.7	2.9	5.1	1.7	5.0	1.4
Y_{H2}	3.9	4.7	1.8	3.1	8.7	2.7	2.8	2.5
$\mu_{m,su}$	3.1	9.0	2.4	10.0	3.3	2.7	5.7	8.4
$\mu_{m,bu}$	3.2	3.6	3.9	9.6	0.7	3.1	2.1	2.5
$\mu_{m,pro}$	3.0	3.4	3.1	1.7	5.3	11.8	8.8	0.3
$\mu_{m,ac}$	4.0	2.2	3.0	4.4	1.2	2.5	2.8	6.0
$\mu_{m,CO}$	3.1	4.0	3.0	2.7	1.2	4.4	11.5	3.1
$\mu_{m,H2}$	3.3	5.7	4.7	2.7	0.6	3.2	4.0	11.3
$f_{ac,CO}$	3.7	2.5	4.3	3.7	1.3	1.8	1.3	5.4
$KI_{H2,\,ac}$	4.0	3.3	2.6	2.1	44.2	2.0	3.9	14.6
$KI_{H2,\,bu}$	4.9	3.9	3.3	6.7	0.8	3.4	4.8	0.4
$KI_{H2,\,pro}$	3.9	3.3	3.7	6.2	1.3	7.0	3.3	2.7
$KI_{CO,\,ac}$	4.4	3.0	4.5	2.5	3.3	2.5	3.8	3.6

$KI_{CO,\,H2}$	3.8	1.5	5.2	3.1	0.4	1.4	3.3	3.0
Ks_{su}	3.5	2.6	4.0	3.4	1.7	8.6	2.5	2.7
Ks_{bu}	4.2	1.7	3.3	2.1	2.1	3.6	2.0	0.9
Ks_{pro}	4.7	4.9	3.4	2.3	0.8	2.2	3.1	2.2
Ks_{ac}	4.9	2.4	3.0	2.8	0.7	2.1	1.6	2.7
Ks_{CO}	2.4	2.4	3.9	3.9	4.8	3.1	3.2	2.7
Ks_{H2}	3.0	2.9	5.1	3.0	1.8	3.2	3.7	5.4
$k_L a_{H2}$	4.8	5.9	2.6	2.9	3.6	1.5	3.7	5.4
K_P	4.2	2.3	5.1	2.8	2.3	7.5	2.9	1.5
Σ	99.9	100.2	99.8	100.1	100.0	100.1	100.1	99.9

This normalization helps to identify which parameters contribute the most to the total variance, *i.e.*, the most sensitive parameters (see Annexes Section 0-2.2, Figure 4.2.1–Figure 4.2.2).

Sobol' method allowed the determination of the first-order sensitivity index given the scalar characteristic. A threshold value of 5% was proposed to consider which parameters were sensitive.

Concerning OP1, a value of 5.6 was obtained for the first-order sensitivity index of Y_{pro} compared to $q_{gas,CH4}$. For $q_{gas,H2}$, the parameters $\mu_{m,su}$, Y_{ac}, Y_{pro}, Y_{bu}, $k_L a_{H2}$, and $\mu_{m,H2}$ were found to be the most sensitive. Regarding $q_{gas,CO}$, the parameters reporting the highest values were Y_{bu}, Y_{su}, Y_{CO}, $KI_{CO,H2}$, Ks_{H2}, and K_P. Finally, the parameters $\mu_{m,su}$, $\mu_{m,bu}$, Y_{bu}, $KI_{H2,\,bu}$, and $KI_{H2,\,pro}$ were found to be the most sensitive for $q_{gas,CO2}$.

For OP2, the parameters $KI_{H2,ac}$, $\mu_{m,pro}$, Y_{CO}, and Y_{H2} were reported as the most sensitivities for the $p_{gas,CH4}$. Regarding $p_{gas,H2}$, the parameters $\mu_{m,pro}$, Ks_{su}, Y_{su}, K_P, $KI_{H2,\,pro}$, and Y_{ac} were found to be the most sensitive, while the parameters reporting the highest values for $p_{gas,CO}$ were $\mu_{m,CO}$, $\mu_{m,pro}$, Ks_{su}, $\mu_{m,su}$, and Y_{CO}. Concerning $q_{gas,CO2}$, the most sensitive parameters were $KI_{H2,\,ac}$, $\mu_{m,H2}$, $\mu_{m,su}$, $\mu_{m,ac}$, $f_{ac,CO}$, Ks_{H2}, and $k_L a_{H2}$.

In OP2, Ks_{su} presented an effect on the model outputs $q_{gas,CH}$ and $q_{gas,CO}$, while $KI_{H2,ac}$ affected $q_{gas,H2}$ and $q_{gas,CO}$.The growth rate $\mu_{m,pro}$ presented an effect over $q_{gas,CO}$ and $q_{gas,CO2}$, while $\mu_{m,su}$ did it over $q_{gas,CH4}$ and $q_{gas,H2}$. The model outputs $q_{gas,H2}$ and $q_{gas,CO2}$ were also affected by Y_{CO}. The other high-influence parameters reported an effect in just one output. Therefore, the selection of unique candidates for parameter estimation was not straightforward. Nevertheless, parameters that appeared as sensitive for most of the model outputs were considered good candidates for parameter estimation, *e.g.*, Y_{bu}, Y_{pro},Y_{CO}, $\mu_{m,su}$, $\mu_{m,pro}$, $KI_{H2,\,ac}$, Ks_{su}, and K_P.

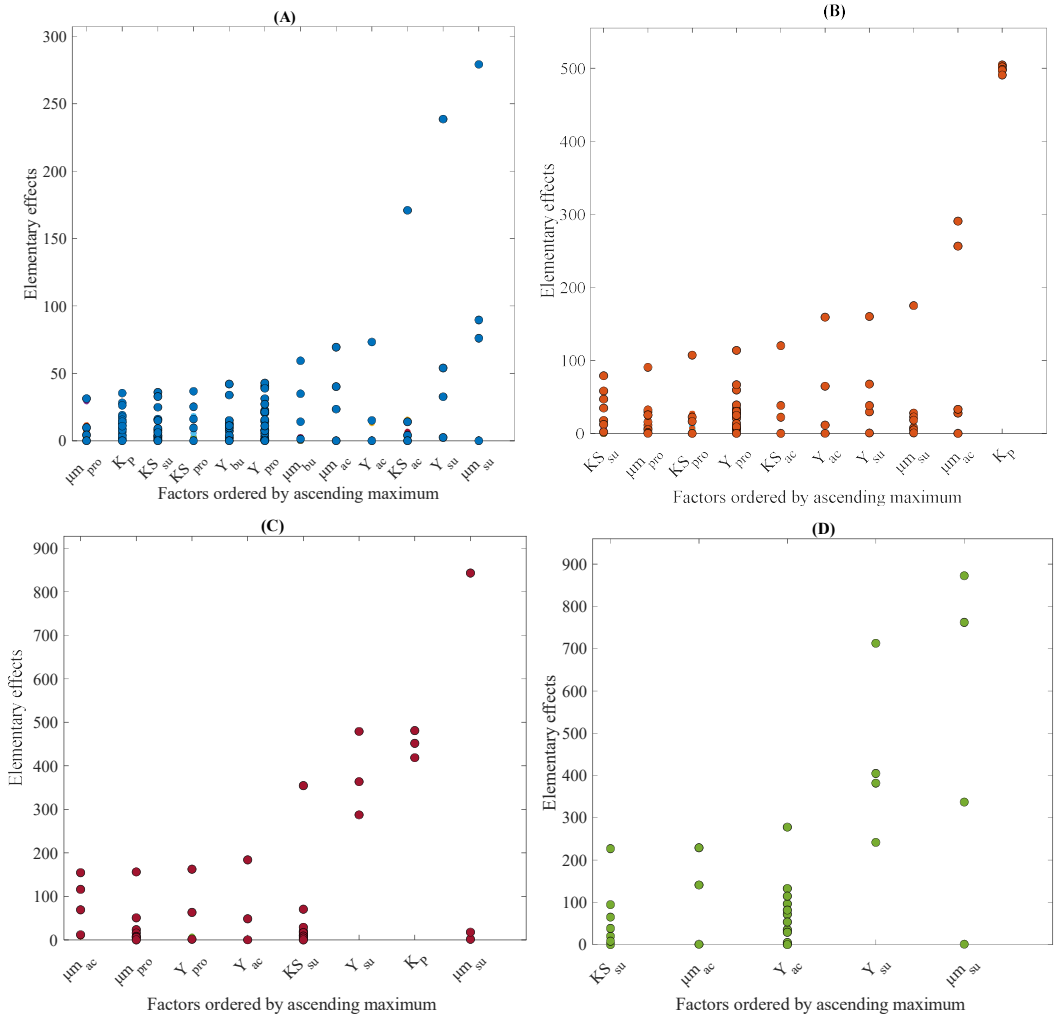


Figure 2.2.3. Morris sensitivity analysis with OP1 over the model outputs: (A) $q_{gas,CH4}$, (B) $q_{gas,H2}$, (C) $q_{gas,CO}$, and (D) $q_{gas,CO2}$.

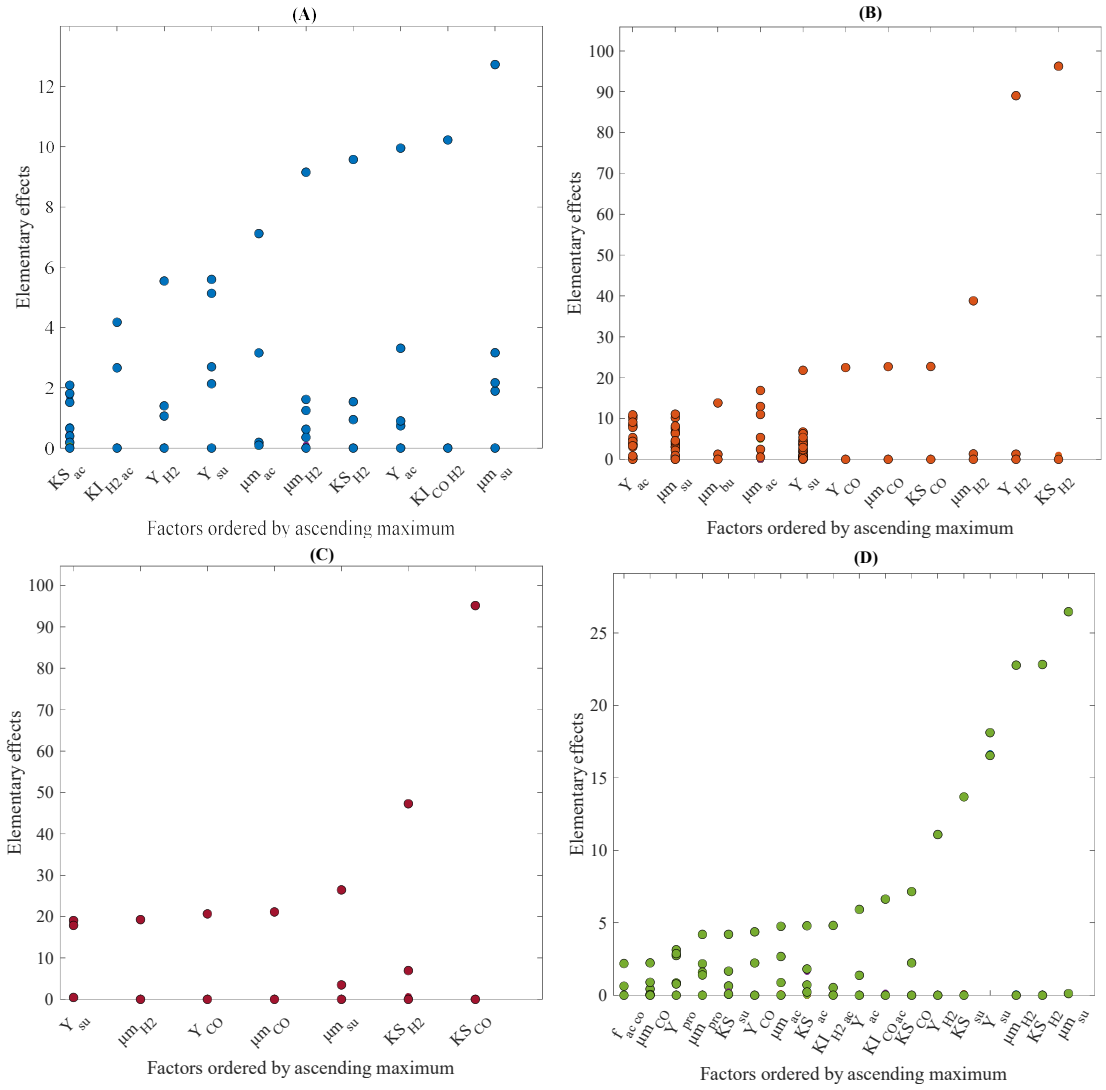


Figure 2.2.4. Morris sensitivity analysis with OP2 over the model outputs: (A) $p_{gas,CH}$, (B) $p_{gas,H2}$, (C) $p_{gas,CO}$, and (D) $p_{gas,CO2}$.

The elementary effects of the Morris method were computed for all 26 parameters. Figure 2.2.3 and Figure 2.2.4 show only the high influence factors identified for each output ordered by ascending maximum (see Annexes Section 2, Table 2.2.5 for more details). For OP1, 12, 10, 8, and 5 parameters were determined as high-influence factors over $q_{gas,CH4}$, $q_{gas,H2}$, $q_{gas,CO}$, and $q_{gas,CO2}$, respectively. Differently, for OP2, 10, 11, 7, and 18 parameters were determined as high influence factors over $p_{gas,CH4}$, $p_{gas,H2}$, $p_{gas,CO}$, and $p_{gas,CO2}$, respectively.

As both operating conditions represented different output variables (OP1: outlet gas flow rates and OP2: gases percent), the sensitivity concerning each factor could be different. Additionally, the operating conditions are not the same. OP1 was carried out under mesophilic temperatures, while OP2 was developed under thermophilic conditions. These conditions could affect the kinetics at which microorganisms are converting the substrates. Therefore, the best trade-off was pursued with both sets of parameters. In OP1 parameters, such as $\mu_{m,su}$, Y_{su}, Y_{ac}, $\mu_{m,ac}$, and Ks_{su} presented an influence over all the model outputs, and Y_{pro}, K_P, and $\mu_{m,pro}$ affected the model outputs $q_{gas,CH4}$, $q_{gas,H2}$, and $q_{gas,CO}$. Ks_{ac} and Ks_{pro} only influenced $q_{gas,CH4}$ and $q_{gas,H2}$. The other high-influence parameters presented an effect just in one output. In OP2, $\mu_{m,su}$, Ks_{H2}, $\mu_{m,H2}$, Y_{su} influence all the model outputs. Y_{ac}, $\mu_{m,ac}$, and Y_{H2} reported an influence on $q_{gas,CH4}$, $q_{gas,H2}$, and $q_{gas,CO2}$, and Ks_{CO}, $\mu_{m,CO}$, and Y_{CO} affected $q_{gas,H2}$, $q_{gas,CO}$, and $q_{gas,CO2}$. $KI_{H2,ac}$, and Ks_{ac} presented an influence on $q_{gas,CH4}$ and $q_{gas,CO2}$. The other high influence parameters presented an effect just in one output. Therefore, good candidates for parameter estimation were Y_{su}, Y_{pro}, Y_{ac}, $\mu_{m,su}$, $\mu_{m,pro}$, $\mu_{m,ac}$, $\mu_{m,H2}$, Ks_{su}, Ks_{ac}, Ks_{H2}, and K_P.

Both methods were similar in measuring the contribution of the parameters to each model output, either by the first-order sensitivity index or the elementary effects. However, Sobol' method was more computationally expensive. For instance, it takes more than one hour to run 200 simulations per output, while the Morris method ran 2700 simulations per output in less than one hour. Comparing the two sensitivity methods, selecting a group of parameters to be estimated was possible.

Table 2.2.6 reports the 14 most sensitive parameters of ADM1_ME that affect the model outputs of both operating conditions.

2.7.2 Parameter Estimation and Model Validation

The 14 most sensitive parameters were estimated using the *fmincon* function from MATLAB® to solve an optimization problem whose objective function was given by Equation (2.2.33). The weights W_{OP1} and W_{OP} were manually adjusted to values of 1 and 1×10^3. To guarantee

that the optimum was a global optimum, simulations were run ten times, adding 10% noise to each obtained parameter after the third iteration. The $RMSE_{est}$ was minimized to a value of 10.67. Table 2.2.6 shows the estimated parameters and their respective confidence intervals. Parameters such as Y_{su}, Y_{bu}, Y_{pro}, Y_{ac}, Y_{CO}, $\mu_{m,H2}$, Ks_{su}, Ks_{ac} and K_P have confidence intervals according to their magnitude, and contain the initial value used in the estimation. Parameters such as $\mu_{m,su}$, $KI_{H2,ac}$, and Ks_{H2} have smaller confidence intervals, indicating that their values cannot change significantly, *i.e.*, there is a 95% probability of finding their precise value in the given interval. The confidence intervals for $\mu_{m,ac}$ and $\mu_{m,pro}$ are comparable with the mentioned work (Blumensaat and Keller, 2005) and the values reported in the original ADM1 (Batstone *et al.*, 2002).

Table 2.2.6. Estimated parameters for the ADM1_ME.

Parameter	Initial value	Estimated value value
Y_{su}	0.1¶	0.0814
Y_{bu}	0.06¶	0.0605
Y_{pro}	0.06¶	0.0281
Y_{ac}	0.05¶	0.0429
Y_{CO}	0.025*	0.0226
$\mu_{m,su}$	45*	31.59
$\mu_{m,pro}$	13¶	10.40
$\mu_{m,ac}$	12.5*	8.79
$\mu_{m,H2}$	90*	109.47
$KI_{H2,\,ac}$	1.00×10^{-6}*	9.75×10^{-7}
Ks_{su}	0.02*	0.0211
Ks_{ac}	0.05*	0.0496
Ks_{H2}	1.00×10^{-6}*	1.03×10^{-6}
K_P	5.00×10^{4}¶	4.99×10^{4}

Reference of the initial value used in parametric estimation: ¶(Batstone *et al.*, 2002) * (Sun *et al.*, 2021)

Table 2.2.7 reports the statistical evaluation of the calibration of the ADM1_ME with the two operating conditions. Values higher than 0.90 and 0.91 were determined for R^2 with OP1 and OP2.

The variables $q_{gas,CH}$ and $p_{\mathrm{gas,CH4}}$ reported the highest values corresponding to 0.95 and 0.98, respectively. In general, the ADM1_ME showed a good fitting to the results of both operating conditions. Concerning the $RMSE$, values lower than 0.38 and 2.52 were obtained with OP1 and OP2, where the variables $q_{gas,CH4}$, $q_{gas,CO}$, $p_{\mathrm{gas,CH4}}$ and $p_{\mathrm{gas,CH4}}$ displayed the best results.

The model showed better accuracy on these variables, *i.e.*, the variation in error when the model results and the operating conditions are compared is lower for these variables. The difference in the magnitude between the $RMSE$ for OP1 and OP2 is related to the magnitude of $\Delta q_{gas,i}$ and $\Delta p_{gas,i}$ (numerators in Equation (2.2.33)) , and the use of the weights W_{OP} and W_{OP2} previously proposed ($W_{OP2} = 1 \times 10^3 W_{OP1}$) helps to counterbalance the differences in variable magnitude when the optimizer is applied.

Table 2.2.7. Statistical analysis for ADM1_ME calibration with OP1 and OP2.

Criteria	OP1				OP2			
	$q_{gas,CH4}$	$q_{gas,H2}$	$q_{gas,CO}$	$q_{gas,CO2}$	$p_{\mathrm{gas,CH4}}$	$p_{\mathrm{gas,H2}}$	$p_{\mathrm{gas,CO}}$	$p_{\mathrm{gas,CO2}}$
R^2	0.95	0.90	0.90	--	0.98	0.97	0.96	0.91
$RMSE$	0.29	0.38	0.27	--	1.06	1.94	1.36	2.52

Figure 2.2.5 displays the results of the simulations of the ADM1_ME with the estimated parameters against the experimental values. Regarding the experimental values, it is important to note that the behavior of the experimental data obtained from the two operational conditions differs. For OP1, there is an increase in the CH_4 content with a decrease in the CO_2 content, demonstrating the conversion of CO_2 into CH_4. However, for OP2, a decrease in both was observed. This decrease represents the negative effect of increasing the syngas addition in the biological methanation process (Andreides *et al.*, 2022). Even if different, the model fitted both datasets correctly. For OP1 (Figure 2.2.5-A), the model correctly fitted the experimental behavior of $q_{gas,CH4}$ and $q_{gas,CO}$. However, it presents difficulties in reproducing the experimental behavior of $q_{gas,H2}$. This could be due to unreported changes in the syngas composition, which was assumed to be the same for the simulations. Although there is no experimental data available for $q_{gas,CO}$, the model simulates $q_{gas,CO2}$ whose dynamic is reducing progressively to compensate for the increase in methane content. On the other hand, Figure 2.2.5-B shows the ADM1_ME adjustment with the OP2. The model reproduces in a better way the behavior of $p_{gas,CH4}$, $p_{gas,H2}$, $p_{gas,CO}$, and $p_{gas,CO}$.

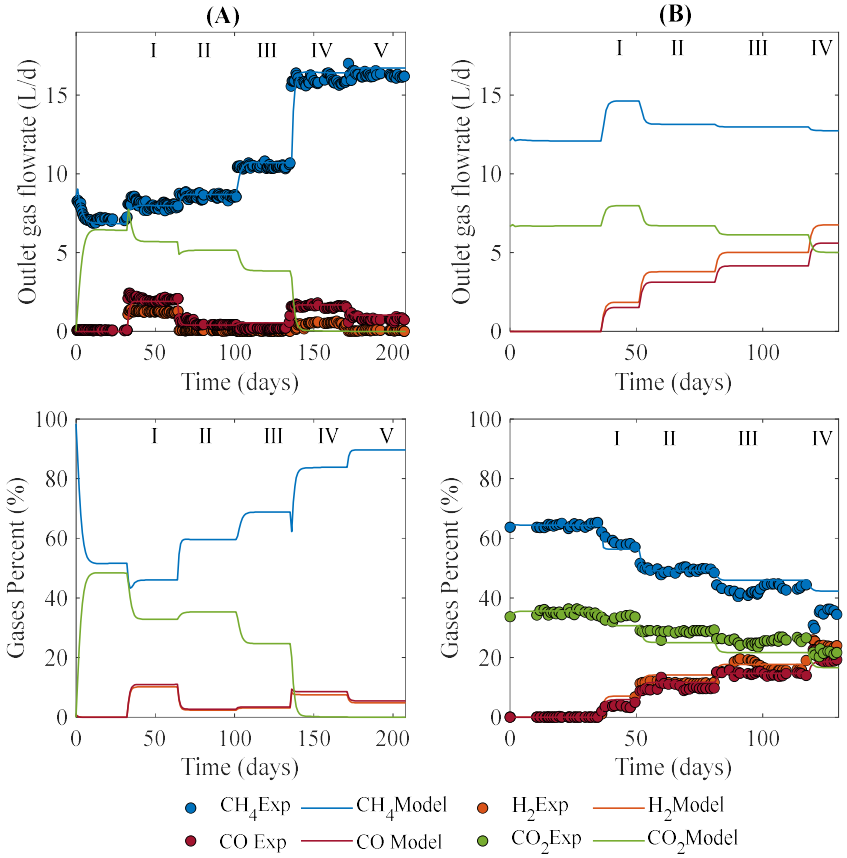


Figure 2.2.5. Outlet gas flow rate $(q_{gas,i})$ and gas percent $(p_{gas,i})$ with the ADM1_ME. (A) OP1 and (B) OP2.

Table 2.2.8 reports the statistical evaluation of the ADM1_ME validation with the two operating conditions. Values higher than 0.74 and 0.82 were obtained in the R^2 with OP1 and OP2, except for $q_{gas,CH}$. Concerning the $RMSE$, values less than 0.94 and 5.15 were exhibited with OP1 and OP2. The best-predicted variables were $q_{gas,H2}$ and $q_{gas,CO}$ ($R^2 > 0.74$ and $RMSE < 0.94$) with OP1. Concerning OP2, all the variables presented similar fitting ($R^2 > 0.82$ and $RMSE < 5.15$), resulting in a better model prediction.

Table 2.2.8. Statistical analysis for ADM1_ME validation with OP1 and OP2.

Criteria	OP1				OP2			
	$q_{gas,CH4}$	$q_{gas,H2}$	$q_{gas,CO}$	$q_{gas,CO2}$	$p_{gas,CH4}$	$p_{gas,H2}$	$p_{gas,CO}$	$p_{gas,CO2}$
R^2	0.39	0.74	0.81	--	0.84	0.87	0.82	0.83
$RMSE$	0.94	0.94	0.22	--	5.15	1.79	1.30	4.19

2.7.3 Model Analysis

Figure 2.2.6 presents the concentration of $S_{liq,su}$, $S_{liq,bu}$, $S_{liq,pro}$, $S_{liq,ac}$, $S_{liq,CO}$, $S_{liq,H2}$, $S_{liq,CH4}$, and $S_{liq,CO2}$ with both operating conditions.

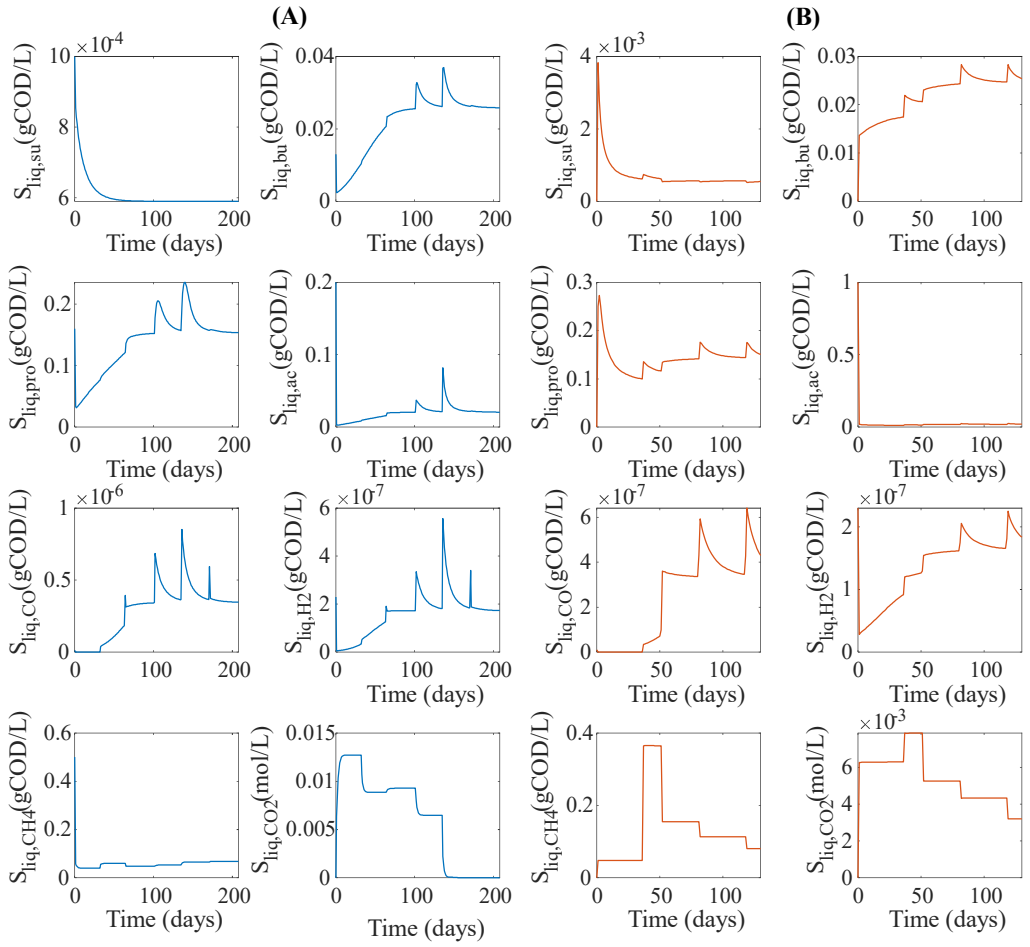


Figure 2.2.6. Concentrations in the liquid phase: sugar $(S_{liq,su})$, butyrate $(S_{liq,bu})$, propionate $(S_{liq,pro})$, acetate $(S_{liq,ac})$, CO $(S_{liq,CO})$, H_2 $(S_{liq,H2})$ CH_4 $(S_{liq,CH4})$ and CO_2 $(S_{liq,CO2})$. (A) OP1 and (B) OP2.

The results were slightly similar in both cases, *e.g.*, a final concentration of 5.9×10^{-4} and 5.6×10^{-4} $gCOD/L$ were obtained for $S_{liq,su}$ with OP1 and OP2. Concentrations of 2.6×10^{-2} and 2.5×10^{-2} $gCOD/L$ were obtained for $S_{liq,bu}$, while values of 0.15 and 0.16 $gCOD/L$ were obtained for $S_{liq,pro}$. Steady-state concentrations of 0.21 and 0.23 $gCOD/L$ were obtained for $S_{liq,ac}$,

3.47×10^{-7} and 4.89×10^{-7} $gCOD/L$ for $S_{liq,CO}$, whereas values of 1.74×10^{-7} and 1.92×10^{-7} $gCOD/L$ were reached for $S_{liq,H2}$.Concentrations of 6.84×10^{-2} and 8.1×10^{-2} $gCOD/L$ were obtained for $S_{liq,CH}$. However, values of 1.11×10^{-5} and 3.20×10^{-3} mol/L were achieved for $S_{liq,CO2}$ with the OP1 and OP2. This is evident from Figure 2.2.5 in which CO_2 is almost exhausted for OP1 at stages IV-V, but there is still some CO_2 for OP2 at stage IV.

Figure 2.2.7 presents the biomass concentrations of X_{su}, X_{bu}, X_{pro}, X_{ac}, X_{CO}, and X_{H2} which were simulated by the ADM1_ME with OP1 and OP2.

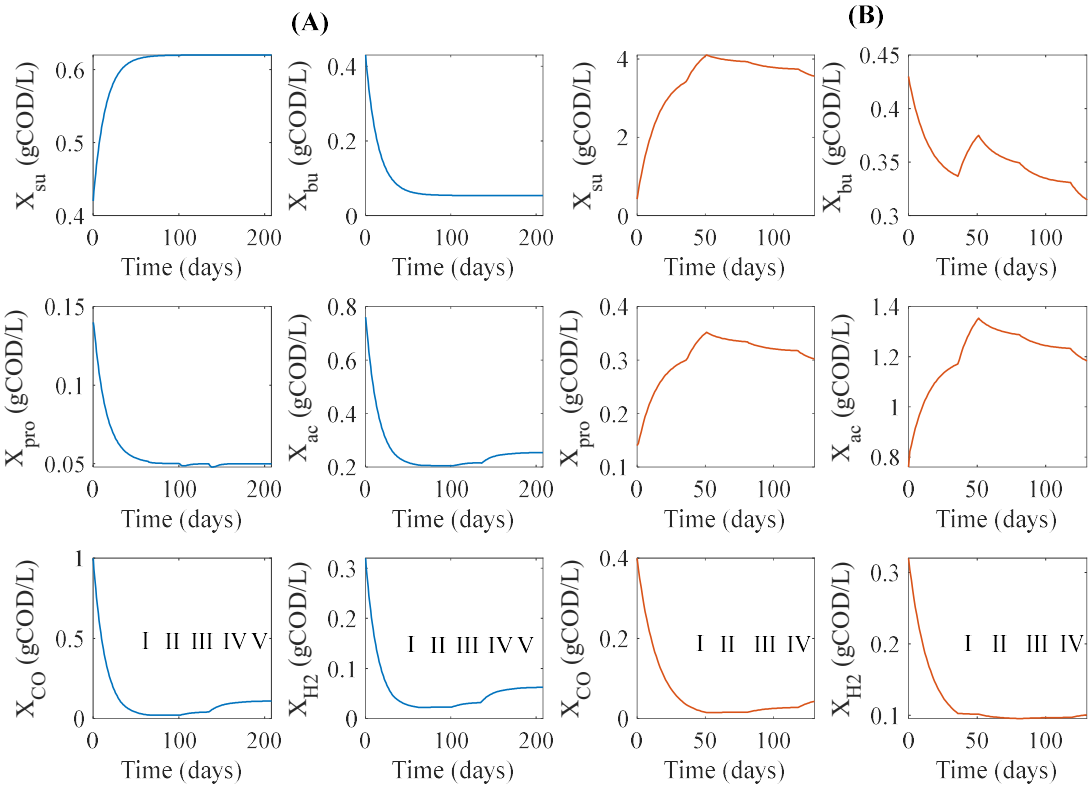


Figure 2.2.7. Biomass concentrations that degrade the components: sugar (X_{su}), butyrate (X_{bu}), propionate (X_{pro}), acetate (X_{ac}), CO (X_{CO}), H_2 (X_{H2}). (A) OP1 and (B) OP2.

The final biomass concentrations of X_{su} were 0.62 and 3.56 $gCOD/L$ for OP1 and OP2, respectively. Concentrations of 0.05 and 0.31 $gCOD/L$ around were obtained for X_{bu} and X_{pro}, whereas values of 0.25 and 1.18 $gCOD/L$ were obtained for X_{ac}. At the end of the

simulation, concentrations of 0.11 and 0.04 $gCOD/L$ were obtained for X_{CO}, and 0.06 and 0.10 $gCOD/L$ for X_{H2} with OP1 and OP2. Regarding X_{CO} and X_{H2}, it was observed that its concentration decreased drastically. However, it increased as the gas was added. This behavior illustrates the correlation between the added gas flow rate and the biomass concentration throughout the simulation of both OP1 and OP2. Results from OP1 indicate that the concentration of X_{H2} fluctuated between 0.023 and 0.063 $gCOD/L$ between stages II to V. Similarly, values ranged between 0.02 and 0.11 $gCOD/L$ were found for X_{CO} across the same stages. In contrast, OP2 displayed a consistent X_{H2} concentration of approximately 0.10 $gCOD/L$ throughout all stages while X_{CO} values varied between 0.024 and 0.037 $gCOD/L$ in stages I-V.

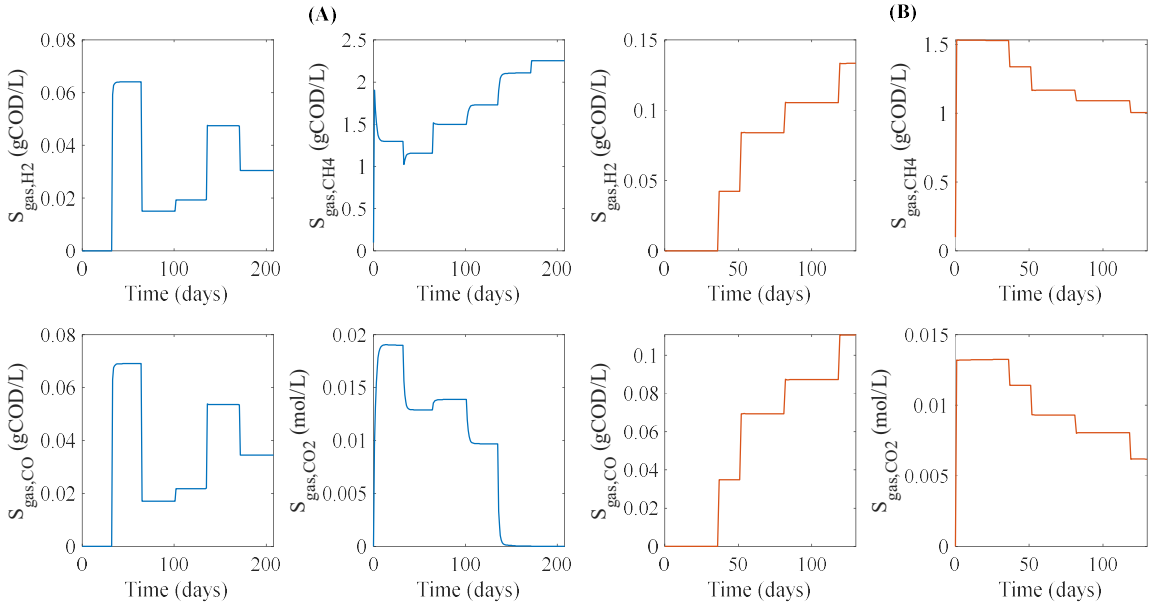


Figure 2.2.8. Concentrations in the gas phase of H_2 $(S_{gas,H2})$ CH_4 $(S_{gas,CH})$, CO $(S_{gas,CO})$ and CO_2 $(S_{gas,CO})$. (A) OP1 and (B) OP2.

Figure 2.2.8 displays the variables $S_{gas,H2}$, $S_{gas,CH4}$, $S_{gas,CO}$, and $S_{gas,CO}$ simulated by the ADM1_ME with OP1 and OP2. Gas concentrations of 0.03 and 0.13 $gCOD/L$ were obtained for $S_{gas,H2}$ with OP1 and OP2. The obtained gas concentrations at the end of the simulation for

both operating conditions corresponded to 2.25 and 1.00 $gCOD/L$ for $S_{gas,CH4}$, 0.03 and 0.10 $gCOD/L$ for $S_{gas,CO}$, and 6.18×10^{-3} and 0.00 $gCOD/L$ for $S_{gas,CO}$.

The behavior for $S_{liq,j}$, X_k, and $S_{gas,i}$ between OP1 and OP2 are not directly comparable. They differ due to the dependence on either OLR or GLR. However, it is observed that the ADM1_ME reproduces different operational conditions and provides information about components such as X_k that are not easy to measure.

Additionally, the model respects the COD balance, which was calculated as proposed by Paudel *et al.*(2015).

$$COD_{liq}^{in} + COD_{gas}^{in} + COD_{bio}^{in} = COD_{liq}^{out} + COD_{gas}^{out} + COD_{bio}^{out} \quad (1.2.38)$$

where COD_{liq}^{in} and COD_{gas}^{in} are the loading in the liquid and gas streams; COD_{bio}^{in} is the initial COD biomass; $COD_{liq,out}$, $COD_{gas,out}$, and $COD_{bio,out}$ are the COD output in the liquid phase, the COD converted to produce biogas, and the assimilated COD for biomass growth, respectively.

For OP1, errors of 1.34, 2.88, 2.89, 2.45, 1.84, and 1.84% were obtained for the COD balance of the stages: reference, I, II, III, IV, and V, respectively. For OP2, the errors of the COD balance were 4.07, 3.90, 2.48, 2.07, and 1.14 for stages: reference, I, II, III, and IV. The small errors could be due to the initial concentrations of biomass in the reactor.

Concerning the mass transfer, to improve the biological methanation process by increasing the syngas added, it is necessary to maintain the mass transfer capacity of the system. In a BCR, the system must be kept in a homogeneous bubbly flow regime, where the bubble size is nearly constant and dictated by the sparger design and the system properties (Hissanaga *et al.*, 2020). In a $CSTR$, the impeller agitation speed must increase as the gas flow rates increases, ensuring high gas retention and complete dispersion (avoid flooding). In addition, in the case of aerated systems, the power consumption is lower than the non-aerated system due to the presence of cavities behind the agitator blade (Gabelle *et al.*, 2011). It is necessary to mention that this

model is based on the hypothesis where both operational conditions are carried out satisfactorily: The BCR is operated in a homogeneous bubbly flow regime, and the power consumption in a $CSTR$ is constant, which allows maintaining the mass transfer.

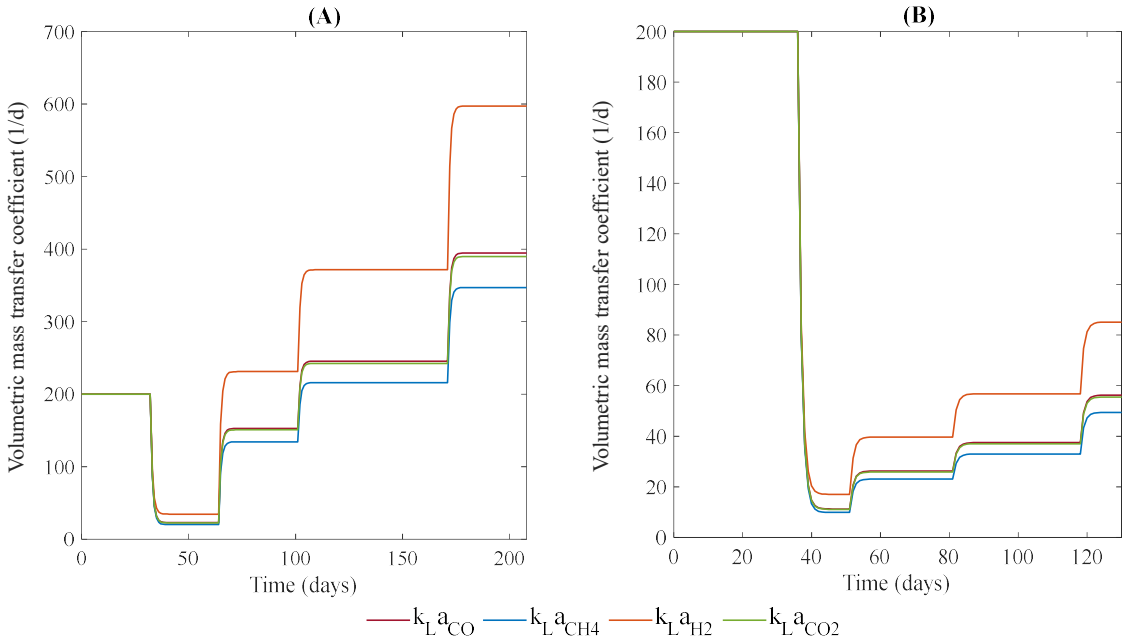


Figure 2.2.9. Volumetric mass transfer coefficient of CH_4 (kLa_{CH}), H_2 (kLa_{H2}), CO (kLa_{CO}), and CO_2 (kLa_{CO}). (A) OP1 and (B) OP2.

As previously mentioned, one of the limiting factors of the performance of biological methanation is the mass transfer process (Guiot *et al.*, 2011; Luo *et al.*, 2013; Rafrafi *et al.*, 2020). The behavior is explored for both operating conditions, OP1 and OP2. Figure 2.2.9 displays the dynamic behavior of the volumetric mass transfer coefficients for OP1 and OP2. In both cases, the values of $k_La_{,i}$ in the reference stage were 200 $1/d$, as proposed in Batstone *et al.* (2002). However, $k_La_{,i}$ in OP1 (Figure 2.2.9-A) depends only on the added gas flow rate. Therefore, their values decrease in stage I and then increased progressively up to 394.6, 347.1, 597.3, and 389.7 $1/d$ for k_La_{CO}, k_La_{CH} , k_La_{H2}, and k_La_{CO} in stage V. In contrast, the $k_La_{,i}$ in OP2 (Figure 2.2.9-B) decreased in stage I below the value in the reference stage and increased progressively until reaching values of 56.22, 49.46, 85.08, and 55.54 $1/d$ for k_La_{CO}, k_La_{CH4}, k_La_{H2}, and k_La_{CO} , in stage V given the contribution provided by the gassed power.

2.8 Conclusions

An extension of the Anaerobic Digestion Model No. 1 (ADM1_ME) was proposed to represent the biological methanation process. The model extension was derived and assessed based on two operating conditions from the literature using two different bioreactor configurations: bubble column reactor and continuous stirred tank reactor, with two different substrates: glucose and a mixture of primary sludge and ticked-disintegrated waste activated, respectively. Sensitivity analysis was performed by the Sobol' and Morris method to identify the candidate parameters to be estimated. In this case, 14 of the 26 previously selected parameters (mainly kinetic ones) strongly influenced the model outputs. Model validation was accurately performed where the variables $q_{gas,CO}$ and $q_{gas,H2}$ reported the best fitting ($R^2 > 0.74$ and $RMSE < 0.94$) for OP1, while for OP2 all the variables presented similar fitting ($R^2 > 0.82$ and $RMSE < 5.15$). Simulation results demonstrated that the proposed model could reproduce the gas outlet flow rates of the biological methanation process for BCR and $CSTR$ while providing information about the dynamics of the biomasses involved in the process. Additionally, the model was able to simulate different operating conditions and the use of various substrates, where an increase in CH_4 and a decrease in CO_2 content is expected. Further work will explore the application of the model in model-based optimization to maximize yields and productivities of CH_4. Moreover, the model could be simulated for different conditions to generate data that could be used for machine-learning-based fault detection techniques.

2.9 References

Ahnert, M., Schalk, T., Brückner, H., Effenberger, J., Kuehn, V., Krebs, P., 2021. Organic matter parameters in WWTP – a critical review and recommendations for application in activated sludge modelling. Water Science and Technology 84, 2093–2112. https://doi.org/10.2166/wst.2021.419

Andreides, D., Pokorna, D., Zabranska, J., 2022. Assessing the syngas biomethanation in anaerobic sludge digestion under different syngas loading rates and homogenisation. Fuel 320, 123929. https://doi.org/10.1016/j.fuel.2022.123929

Angelidaki, I., Ellegaard, L., Ahring, B.K., 1999. A comprehensive model of anaerobic bioconversion of complex substrates to biogas. Biotechnol. Bioeng. 63, 363–372.

https://doi.org/10.1002/(SICI)1097-0290(19990505)63:3<363::AID-BIT13>3.0.CO;2-Z

Angelidaki, I., Karakashev, D., Batstone, D.J., Plugge, C.M., Stams, A.J.M., 2011. Biomethanation and its potential, 1st ed, Methods in Enzymology. Elsevier Inc. https://doi.org/10.1016/B978-0-12-385112-3.00016-0

Ashraf, M.T., Sieborg, M.U., Yde, L., Rhee, C., Shin, S.G., Triolo, J.M., 2020. Biomethanation in a thermophilic biotrickling filter — pH control and lessons from long-term operation. Bioresource Technology Reports 11, 100525. https://doi.org/10.1016/j.biteb.2020.100525

Asprey, S.P., Macchietto, S., 2000. Statistical tools for optimal dynamic model building. Computers & Chemical Engineering 24, 1261–1267. https://doi.org/10.1016/S0098-1354(00)00328-8

Bai, J., Liu, H., Yin, B., Ma, H., Chen, X., 2017. Modified ADM1 for modeling free ammonia inhibition in anaerobic acidogenic fermentation with high-solid sludge. Journal of Environmental Sciences 52, 58–65. https://doi.org/10.1016/j.jes.2016.03.004

Balde, Y.M., Diop, S., Tebbani, S., Kante, C., 2020. Modeling of a Continuous Anaerobic Digestion of Wastes, in: 2020 24th International Conference on System Theory, Control and Computing (ICSTCC). Presented at the 2020 24th International Conference on System Theory, Control and Computing (ICSTCC), IEEE, Sinaia, Romania, pp. 596–601. https://doi.org/10.1109/ICSTCC50638.2020.9259735

Batstone, D.J., Balthes, C., Barr, K., 2010. Model assisted startup of anaerobic digesters fed with thermally hydrolysed activated sludge. Water Science 6.

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). Water Science and Technology 45, 65–73. https://doi.org/10.2166/wst.2002.0292

Bharathiraja, B., Sudharsanaa, T., Bharghavi, A., Jayamuthunagai, J., Praveenkumar, R., 2016. Biohydrogen and Biogas – An overview on feedstocks and enhancement process. Fuel 185, 810–828. https://doi.org/10.1016/j.fuel.2016.08.030

Blumensaat, F., Keller, J., 2005. Modelling of two-stage anaerobic digestion using the IWA Anaerobic Digestion Model No. 1 (ADM1). Water Research 39, 171–183. https://doi.org/10.1016/j.watres.2004.07.024

Brémond, U., Bertrandias, A., Steyer, J.-P., Bernet, N., Carrere, H., 2021. A vision of European biogas sector development towards 2030: Trends and challenges. Journal of Cleaner Production 287, 125065. https://doi.org/10.1016/j.jclepro.2020.125065

Chandra, R., Takeuchi, H., Hasegawa, T., 2012. Methane production from lignocellulosic agricultural crop wastes: A review in context to second generation of biofuel production. Renewable and Sustainable Energy Reviews 16, 1462–1476. https://doi.org/10.1016/j.rser.2011.11.035

Czatzkowska, M., Harnisz, M., Korzeniewska, E., Koniuszewska, I., 2020. Inhibitors of the methane fermentation process with particular emphasis on the microbiological aspect: A review. Energy Science and Engineering 8, 1880–1897. https://doi.org/10.1002/ese3.609

Damblin, G., Couplet, M., Iooss, B., 2013. Numerical studies of space-filling designs: optimization of Latin Hypercube Samples and subprojection properties. Journal of Simulation 7, 276–289. https://doi.org/10.1057/jos.2013.16

Dar, R.A., Parmar, M., Dar, E.A., Sani, R.K., Phutela, U.G., 2021. Biomethanation of agricultural residues: Potential, limitations and possible solutions. Renewable and Sustainable Energy Reviews 135, 110217. https://doi.org/10.1016/j.rser.2020.110217

Deckwer, W.D., Nguyen-Tien, K., Kelkar, B.G., Shah, Y.T., 1983. Applicability of axial dispersion model to analyze mass transfer measurements in bubble columns. AIChE Journal 29, 915–922. https://doi.org/10.1002/aic.690290607

Dev, S., Saha, S., Kurade, M.B., Salama, E.S., El-Dalatony, M.M., Ha, G.S., Chang, S.W., Jeon, B.H., 2019. Perspective on anaerobic digestion for biomethanation in cold environments. Renewable and Sustainable Energy Reviews 103, 85–95. https://doi.org/10.1016/j.rser.2018.12.034

Feng, K., Lu, Z., Yang, C., 2019. Enhanced Morris method for global sensitivity analysis: good proxy of Sobol' index. Struct Multidisc Optim 59, 373–387. https://doi.org/10.1007/s00158-018-2071-7

Gabelle, J.-C., Augier, F., Carvalho, A., Rousset, R., Morchain, J., 2011. Effect of tank size on kLa and mixing time in aerated stirred reactors with non-newtonian. Can J Chem Eng 89, 1139–1153. https://doi.org/10.1002/cjce.20571

Gary, B. Tatterson, 1991. Fluid mixing and gas dispersion in agitated tanks. https://doi.org/10.5860/choice.29-0927

Grimalt-Alemany, A., Asimakopoulos, K., Skiadas, I.V., Gavala, H.N., 2020. Modeling of syngas biomethanation and catabolic route control in mesophilic and thermophilic mixed microbial consortia. Applied Energy 262, 114502. https://doi.org/10.1016/j.apenergy.2020.114502

Guiot, S.R., Cimpoia, R., Carayon, G., 2011. Potential of Wastewater-Treating Anaerobic Granules for Biomethanation of Synthesis Gas. Environ. Sci. Technol. 45, 2006–2012. https://doi.org/10.1021/es102728m

Gustafsson, M., Svensson, N., Eklund, M., Fredriksson Möller, B., 2021. Well-to-wheel climate performance of gas and electric vehicles in Europe. Transportation Research Part D: Transport and Environment 97, 102911. https://doi.org/10.1016/j.trd.2021.102911

Henze, M., van Loosdrecht, M.C.M., Ekama, G.A., Brdjanovic, Damir., 2019. Biological Wastewater Treatment: Principles, Modeling and Design. IWA Publishing. https://doi.org/10.2166/9781780408613

Higbie, R., 1935. The rate of absorption of a pure gas into still liquid during short periods of exposure. Transaction of the American Institute of Chemical Engineers (AIChe).

Hissanaga, A.M., Padoin, N., Paladino, E.E., 2020. Mass transfer modeling and simulation of a transient homogeneous bubbly flow in a bubble column. Chemical Engineering Science 218, 115531. https://doi.org/10.1016/j.ces.2020.115531

Iglesias, R., Muñoz, R., Polanco, M., Díaz, I., Susmozas, A., Moreno, A.D., Guirado, M., Carreras, N., Ballesteros, M., 2021. Biogas from Anaerobic Digestion as an Energy Vector: Current Upgrading Development. Energies 14, 2742. https://doi.org/10.3390/en14102742

Iooss, B., Lemaître, P., 2015. A Review on Global Sensitivity Analysis Methods, in: Dellino, G., Meloni, C. (Eds.), Uncertainty Management in Simulation-Optimization of Complex Systems, Operations Research/Computer Science Interfaces Series. Springer US, Boston, MA, pp. 101–122. https://doi.org/10.1007/978-1-4899-7547-8_5

Jensen, M.B., Ottosen, L.D.M., Kofoed, M.V.W., 2021. H2 gas-liquid mass transfer: A key element in biological Power-to-Gas methanation. Renewable and Sustainable Energy Reviews 147, 111209. https://doi.org/10.1016/j.rser.2021.111209

Kucherenko, S., Delpuech, B., Iooss, B., Tarantola, S., 2015. Application of the control variate technique to estimation of total sensitivity indices. Reliability Engineering & System Safety 134, 251–259. https://doi.org/10.1016/j.ress.2014.07.008

Laguillaumie, L., Rafrafi, Y., Moya-Leclair, E., Delagnes, D., Dubos, S., Spérandio, M., Paul, E., Dumas, C., 2022. Stability of ex situ biological methanation of H2/CO2 with a mixed microbial culture in a pilot scale bubble column reactor. Bioresource Technology 354, 127180. https://doi.org/10.1016/j.biortech.2022.127180

Li, C., Zhu, X., Angelidaki, I., 2020. Carbon monoxide conversion and syngas biomethanation mediated by different microbial consortia. Bioresource Technology 314, 123739. https://doi.org/10.1016/j.biortech.2020.123739

Liu, K., Phillips, J.R., Sun, X., Mohammad, S., Huhnke, R.L., Atiyeh, H.K., 2019. Investigation and modeling of gas-liquid mass transfer in a sparged and non-sparged continuous stirred tank reactor with potential application in syngas fermentation. Fermentation 5. https://doi.org/10.3390/fermentation5030075

Liu, R., Hao, X., Wei, J., 2016. Function of homoacetogenesis on the heterotrophic methane production with exogenous H2/CO2 involved. Chemical Engineering Journal 284, 1196–1203. https://doi.org/10.1016/j.cej.2015.09.081

Luo, G., Wang, W., Angelidaki, I., 2013. Anaerobic Digestion for Simultaneous Sewage Sludge Treatment and CO Biomethanation: Process Performance and Microbial Ecology. Environ. Sci. Technol. 130904143045005. https://doi.org/10.1021/es401018d

Morio, J., 2011. Global and local sensitivity analysis methods for a physical system. Eur. J. Phys. 32, 1577–1583. https://doi.org/10.1088/0143-0807/32/6/011

Morris, M.D., 1991. Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics 33, 161–174. https://doi.org/10.1080/00401706.1991.10484804

Mosey, F.E., 1983. Mathematical Modelling of the Anaerobic Digestion Process: Regulatory Mechanisms for the Formation of Short-Chain Volatile Acids from Glucose. Water Science and Technology 15, 209–232. https://doi.org/10.2166/wst.1983.0168

Mr, 2022. Sensitivity Analysis - Morris method (advanced) (https://www.mathworks.com/matlabcentral/fileexchange/48884-sensitivity-analysis-morris-method-advanced), MATLAB Central File Exchange. Retrieved October 5, 2022.

Mulat, D.G., Mosbæk, F., Ward, A.J., Polag, D., Greule, M., Keppler, F., Nielsen, J.L., Feilberg, A., 2017. Exogenous addition of H 2 for an in situ biogas upgrading through biological reduction of carbon dioxide into methane. Waste Management 68, 146–156. https://doi.org/10.1016/j.wasman.2017.05.054

Nauman, E.B., 2008. Chemical Reactor Design, Optimization, and Scaleup. John Wiley & Sons, Inc., Hoboken, NJ, USA. https://doi.org/10.1002/9780470282076

Ngu, V., Morchain, J., Cockx, A., 2022. Spatio-temporal 1D gas–liquid model for biological methanation in lab scale and industrial bubble column. Chemical Engineering Science 251, 117478. https://doi.org/10.1016/j.ces.2022.117478

Ochoa, M.P., Estrada, V., Di Maggio, J., Hoch, P.M., 2016. Dynamic global sensitivity analysis in bioreactor networks for bioethanol production. Bioresource Technology 200, 666–679. https://doi.org/10.1016/j.biortech.2015.10.069

Pan, X., Zhao, L., Li, C., Angelidaki, I., Lv, N., Ning, J., Cai, G., Zhu, G., 2021. Deep insights into the network of acetate metabolism in anaerobic digestion: focusing on syntrophic acetate oxidation and homoacetogenesis. Water Research 190, 116774. https://doi.org/10.1016/j.watres.2020.116774

Paudel, S., Kang, Y., Yoo, Y.-S., Seo, G.T., 2015. Hydrogen Production in the Anaerobic Treatment of Domestic-Grade Synthetic Wastewater. Sustainability 7, 16260–16272. https://doi.org/10.3390/su71215814

Rafrafi, Y., Laguillaumie, L., Dumas, C., 2020. Biological Methanation of H2 and CO2 with Mixed Cultures: Current Advances, Hurdles and Challenges. Waste and Biomass Valorization. https://doi.org/10.1007/s12649-020-01283-z

Rodriguez-Fernandez, M., Kucherenko, S., Pantelides, C., Shah, N., 2007. Optimal experimental design based on global sensitivity analysis, in: Computer Aided Chemical Engineering. Elsevier, pp. 63–68. https://doi.org/10.1016/S1570-7946(07)80034-4

Rosen, C., Jeppsson, U., 2006. Aspects on ADM1 Implementation within the BSM2 Framework, Technical report. Department of Industrial Electrical Engineering and Automation (IEA), Lund University, Lund, Sweden.

Rusmanis, D., O'Shea, R., Wall, D.M., Murphy, J.D., 2019. Biological hydrogen methanation systems – an overview of design and efficiency. Bioengineered 10, 604–634. https://doi.org/10.1080/21655979.2019.1684607

Saha, S., Basak, B., Hwang, J.H., Salama, E.S., Chatterjee, P.K., Jeon, B.H., 2020. Microbial Symbiosis: A Network towards Biomethanation. Trends in Microbiology 1–17. https://doi.org/10.1016/j.tim.2020.03.012

Saltelli, A. (Ed.), 2008. Global sensitivity analysis: the primer. Wiley, Chichester, West Sussex.

Saltelli, A. (Ed.), 2004. Sensitivity analysis in practice: a guide to assessing scientific models, Reprinted. ed. Wiley, Chichester Weinheim.

Sepúlveda, F.D., Cisternas, L.A., Gálvez, E.D., 2014. The use of global sensitivity analysis for improving processes: Applications to mineral processing. Computers & Chemical Engineering 66, 221–232. https://doi.org/10.1016/j.compchemeng.2014.01.008

Sepulveda, F.D., Cisternas, L.A., Gálvez, E.D., 2013. Global sensitivity analysis of a mineral processing flowsheet, in: Computer Aided Chemical Engineering. Elsevier, pp. 913–918. https://doi.org/10.1016/B978-0-444-63234-0.50153-6

Sobol′, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation 55, 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6

Sohier, H., Farges, J.-L., Piet-Lahanier, H., 2014. Improvement of the Representativity of the Morris Method for Air-Launch-to-Orbit Separation. IFAC Proceedings Volumes 47, 7954–7959. https://doi.org/10.3182/20140824-6-ZA-1003.01968

Sun, H., Yang, Z., Zhao, Q., Kurbonova, M., Zhang, R., Liu, G., Wang, W., 2021. Modification and extension of anaerobic digestion model No.1 (ADM1) for syngas biomethanation simulation: From lab-scale to pilot-scale. Chemical Engineering Journal 403, 126177. https://doi.org/10.1016/j.cej.2020.126177

Tosin, M., Côrtes, A.M.A., Cunha, A., 2020. A Tutorial on Sobol' Global Sensitivity Analysis Applied to Biological Models, in: da Silva, F.A.B., Carels, N., Trindade dos Santos, M., Lopes, F.J.P. (Eds.), Networks in Systems Biology, Computational Biology. Springer International Publishing, Cham, pp. 93–118. https://doi.org/10.1007/978-3-030-51862-2_6

Van't, R.K., 1979. Review of Measuring Methods and Results in Nonviscous Gas-Liquid Mass Transfer in Stirred Vessels 18, 357–364.

Velez S. Carlos M., 2022. Global sensitivity and uncertainty analysis (GSUA)(https://www.mathworks.com/matlabcentral/fileexchange/47758-global-sensitivity-and-uncertainty-analysis-gsua), MATLAB Central File Exchange. Retrieved October 5, 2022.

Zhang, X., Trame, M., Lesko, L., Schmidt, S., 2015. Sobol Sensitivity Analysis: A Tool to Guide the Development and Evaluation of Systems Pharmacology Models. CPT Pharmacometrics Syst. Pharmacol. 4, 69–79. https://doi.org/10.1002/psp4.6

Zupančič, M., Možic, V., Može, M., Cimerman, F., Golobič, I., 2022. Current Status and Review of Waste-to-Biogas Conversion for Selected European Countries and Worldwide. Sustainability 14, 1823. https://doi.org/10.3390/su14031823

Section 3 Multi-Objective Dynamic Optimization Applied to Biological Methanation Process

Juan C. Acosta-Pavas[a], Carlos E. Robles-Rodríguez[a], Camilo A. Suarez Méndez[b], Jérôme Morchain[a], Claire Dumas[a], Arnaud Cockx[a], César A. Aceves-Lara[a]

[a]TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France
[b]Department of Processes and Energy, Universidad Nacional de Colombia, Carrera 80 No. 65-223, Medellin, Colombia

Publications:

Acosta-Pavas, J. C., Robles-Rodríguez, C. E., Méndez Suarez, C. A., Morchain, J., Dumas, C., Cockx, A., & Aceves-Lara, C. A. (2022). Dynamic Multi-Objective Optimization Applied to Biomethanation Process. Chemical Engineering Transactions, 96, 319–324. https://doi.org/10.3303/CET2296054

Abstract

Dynamic mathematical models could be beneficial for understanding and simulating processes to achieve an optimal operation. The optimum, however, could depend on several variables that can be conflicting. In this regard, Multi-objective Dynamic Optimization (MODO) is necessary for the trade-off between several objectives. This work proposes a MODO as a control strategy integrating two optimization problems. The objective is to maximize the methane yield and productivity of the biological methanation processes by modifying the inlet liquid and gas flow rates. First, multi-objective optimization was applied. Three Pareto optimal points were selected to develop five cases in dynamic optimization. Case 1 corresponded to the literature value. Cases 2, 3, and 4 were considered as objectives: the maximum methane productivity, maximum methane productivity and yield, and maximum methane yield, respectively. Case 5 was performed to assess a switch between objectives. For case 3, the yield decreased to 0.97 times, while the productivity increased 3.26 times concerning case 1. The added gas flow rate ranged from 2.69 to 8.43 m^3/d, and the inlet liquid flow rate reached an approximate value of 7.0×10^{-3} m^3/d. These results showed the feasibility and good efficiency of the proposed methodology.

Keywords: Biological Methanation, Muti-Objective Optimization, Pareto Optimal Set, Dynamic Optimization, Model Predictive Control

3.1 Introduction

The use of dynamic models allows us to gain a better understanding of different biological processes. One of those is biological methanation. In this process, the organic matter, such as agricultural residues, organic effluents from the food industry, animal manure, or waste/wastewater residues, are transformed through the synergistic work of a variety of microorganisms into a mixture of CH_4 and CO_2 (Dar *et al.*, 2021). This process was first modeled using the Anaerobic Digestion Model No. 1 (ADM1) (Batstone *et al.*, 2002). This model has been adapted to solve stiffness problems (Rosen and Jeppsson, 2006), variation of pH (Czatzkowska *et al.*, 2020), and the inclusion of gas addition to obtain high-purity methane (Sun *et al.*, 2021). However, managing the biological methanation process is still an arduous task due to the multiple molecules and different microorganisms involved. As a result, obtaining desired objectives, such as high yields, high productivity, low processing times, or low flow rates, remains difficult at an industrial scale, especially when it is necessary to optimize several of them simultaneously.

The use of dynamic models plays a crucial role in the design of control strategies, *e.g.*, optimal control, adaptive control, or model predictive control (MPC) (Luna *et al.*, 2021; Morales-Rodelo *et al.*, 2020; Smets *et al.*, 2004) to maintain the value of the variables of interest during the process or to optimize several variables. In other words, a multi-objective optimization (MOO). When we talk about MPC, we refer to optimal controllers, *i.e.*, the control action responds to the optimization of a criterion (cost function) related to the system's future behavior determined from the dynamic model (Camacho and Bordons, 2007).

MOO involves multiple criteria decision-making. It implies optimizing problems where there are more than one variable to be optimized simultaneously, and those variables are usually conflictive (Chang, 2015; Vertovec *et al.*, 2021). In this context, an optimal solution set that fulfills the desired conditions of the conflicting variables is established and selected as Pareto optimal set (POS). If another solution does not dominate a solution point, it is considered a Pareto Optimal Point (POP). Therefore, it is ideal to have the highest number of Pareto optimal solutions (Deb *et al.*, 2002).

This work aims at proposing a Multi-Objective Dynamic Optimization (MODO) to find the trade-off between the maxima methane yield and productivity of the biological methanation process through a Pareto Optimal Set (POS). Afterward, a POP is selected and used as the optimal reference trajectory. Then, a dynamic optimization is formulated in terms of a MPC to modify the inlet liquid and gas flow rates to achieve the optimal values of yield and productivity obtained from the POP.

3.2 Biological Methanation Model Extension Proposal

The Anaerobic Digestion (AD) process can be divided into four phases: hydrolysis, acidogenesis, acetogenesis, and methanogenesis (Dar *et al.*, 2021). In the first phase, the fermentative bacteria excrete enzymes that hydrolyze complex organic polymers (carbohydrates, proteins, and lipids) into soluble monomers, such as monosaccharides, amino acids, and long-chain fatty acids. In the second phase, these monomers are transformed into volatile fatty acids (VFA), such as acetate, propionate, and butyrate. In the third phase, all the VFA are transformed into acetate, H_2, and CO_2. The fourth phase involves the conversion of these components by methanogenic archaea into biogas, *i.e.*, a mixture of CH_4 and CO_2. Finally, this process is extended to biological methanation, including methane production by the biological activity of methanogenic bacteria converting the added H_2 and CO.

The model was based on experimental data from the literature (Sun *et al.*, 2021). The entire experiment was carried out in a bioreactor with a working volume (V_{liq}) of 37.5 L and hydraulic retention time (HRT) of 20 days operating at 37°C for 207 days. The organic loading rate (OLR) was 0.53 $gCOD/L/d$ of glucose with an inlet liquid flow rate $\left(q_{liq}^{in}\right)$ of 1.9 L/d. The gas addition was performed in five stages, in which the inlet gas flow rate $\left(q_{gas}^{in}\right)$ and the gas loading rate (GLR) were varied in time. These values are reported in Table 2.3.1.

Table 2.3.1. Experimental conditions from literature (Sun *et al.*, 2021).

Stage	Time (Day)	q_{gas}^{in} (L/d)	GLR $(L/Lr/d)$
Reference	1-32	-	-
I	33-64	7.5	0.2
II	65-101	7.5	0.2
III	102-135	15	0.4
IV	136-171	37.5	1.0
V	172-207	37.5	1.0

An extension of the Anaerobic Digestion Model No. 1 (ADM1_ME) to consider the addition of H_2 and CO to improve CH_4 production was proposed in our previous work (Acosta-Pavas *et al.*, 2023). Here, it is rewritten as,

$$\frac{dS_{gas,i}}{dt} = \frac{q_{gas}^{in}}{V_{gas}} S_{gas,i}^{in} + N_i \left(\frac{V_{liq}}{V_{gas}}\right) - \frac{q_{gas}}{V_{gas}} S_{gas,i} \quad (2.3.1)$$

$$\frac{dS_{liq,j}}{dt} = \frac{q_{liq}^{in}}{V_{liq}} \left(S_{liq,j}^{in} - S_{liq,j}\right) + \sum_{k} f_{j,k} \mu_k X_k - N_i \quad (2.3.2)$$

$$\frac{dX_k}{dt} = \frac{q_{liq}^{in}}{V_{liq}} \left(X_k^{in} - X_k\right) + Y_k \mu_k X_k - K_{k,dec} X_k \quad (2.3.3)$$

where sub-index $i \in [1,4]$ corresponds to H_2, CH_4, CO, and CO_2; sub-index $j \in [1,8]$ denotes glucose, butyrate, propionate, acetate, H_2, CH_4, CO, and CO_2 in the liquid phase; and sub-index $k \in [1,6]$ reads for the biomass that degrade glucose, butyrate, propionate, acetate, H_2, and CO, respectively. V_{gas} is the molar fraction volume, $S_{liq,j}^{in}$ is the inlet concentration of component j in the liquid phase, q_{gas} is the outlet gas flow rate, $f_{j,k}$ are the stoichiometric coefficients, X_k^{in} is the inlet concentration of biomass k, μ_k and $K_{k,dec}$ are the growth rate and decay constant of biomass k, Y_k *is* the yield of biomass k, and N_i is the mass transfer rate of component i.

3.3 Multi-objective Dynamic Optimization Construction as Control Strategy

3.3.1 Multi-objective Optimization

Several variables can be optimized in biological processes, yields, productivities, process times, *etc*. Most of these variables are often conflicting. Therefore, it is necessary to find a trade-off between them; this is called a multi-objective optimization (MOO) problem. In this case, multiple optimal solutions that satisfy the desired conditions of both variables can be found. This is known as the POS. In general, a MOO can be formulated as follows,

$$\min_{Y,u,\theta,t} \{J_1^*(Y,u,\theta), \dots, J_m^*(Y,u,\theta)\} \tag{2.3.4}$$

$$Subject\ to \begin{cases} dY/dt = \xi(Y,u,\theta,t) & t \in [0,t_f] \\ \lambda_i(Y,u,\theta,t) \leq 0 & i = 1,2,\dots,n_\lambda \\ \psi_i(Y,u,\theta,t) = 0 & i = 1,2,\dots,n_\psi \\ u^L \leq u \leq u^U \end{cases}$$

where $J_1^*, \dots, J_m^*$ are the m objective functions, Y the state variables, λ_i and ψ_i indicate inequality and equality constraints on the variable states, u and θ denote the control variables and parameters, and u^L, u^U correspond to the lower and upper bounds of the control variables (Tsiantis *et al.*, 2018).

3.3.2 Dynamic Optimization as a Model Predictive Control

MPC is one of the most widely used control methods in the industry (Morales-Rodelo *et al.*, 2020; Yamashita *et al.*, 2016). A MPC is an advanced control strategy that solves an optimal control problem at every sampling time. The control uses an explicit model to predict the system's outputs at a future time by calculating the future control sequences to minimize a cost function (Giraldo *et al.*, 2022). The dynamic optimization (control problem) determines the future control value that minimizes a specified performance index, *i.e.*, the input variables, that minimizes the following objective function,

$$\min_u \left(\sum_{j=t}^{t+H_p} |J^* - J(t+j|t)|^2 + \sum_{j=t}^{t+H_c} W_u \, \Delta u(t+j|t)^2 \right) \tag{2.3.5}$$

$$Subject\ to \begin{cases} dY/dt = \xi(Y,u,\theta,t) \\ \lambda_i(Y,u,\theta,t) \leq 0 \qquad i = 1,2,\dots,n_\lambda \\ \psi_i(Y,u,\theta,t) = 0 \qquad i = 1,2,\dots,n_\psi \\ u^L \leq u \leq u^U \end{cases}$$

where u is the vector of the control variables, H_p and H_c are the prediction and control horizons, $J(t+j|t)$ refers to the output prediction calculated at time instant $t+j$ using the information available at time instant t, J^* holds for the reference trajectory, enables to reach the set point. These variables are determined by the MOO, $\Delta u(t+j|t)$ is the control move at time instant $t+j$ calculated using information available at time instant t.

A MODO strategy is proposed to determine the optimal values of the objective functions (Figure 2.3.1). This strategy entails five steps:

Step 1 - Model definition: Proposition of the dynamic model representative of the biological process.

Step 2 - Definition of the multi-objective optimization problem: Definition of the objective functions $J_1, \dots, J_m$ to be maximized/minimized by the MOO optimization, the vector of the control variables u, the constraints λ_i and ψ_i, and the bounds u^L and u^U of the control variables in the MOO optimization.

Step 3 - Selection of the Pareto optimal point (POP): Determination of the Pareto optimal set $J_1^*, \dots, J_m^*$ and selection of the POP to be used as the reference trajectory in the dynamic optimization.

Step 4 - Definition of the dynamic problem with a single weighted objective: Formulation of an objective function considering the previously identified POP in terms of a MPC problem. To indicate the initial guess values u_0, the constraints λ_i and ψ_i, and the bounds u^L and u^U of the control variables in the dynamic optimization.

Step 5 - Implementation of the optimization: Execution of the dynamic optimization and determination of the optimal values of the control and optimized variables at each time.

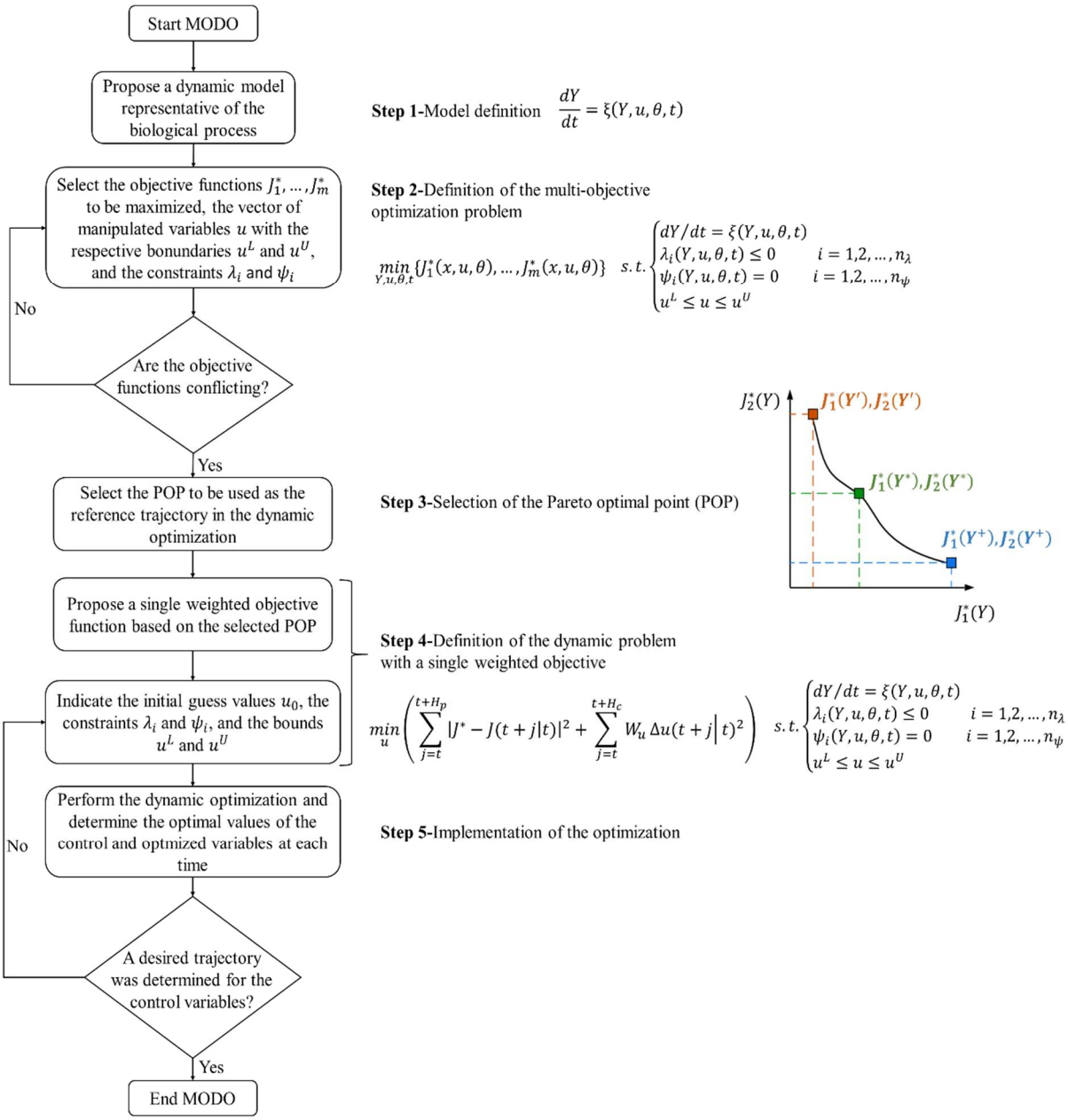


Figure 2.3.1. Multi-objective dynamic optimization strategy.

3.4 Case Study: Multi-Objective Dynamic Optimization in Biological Methanation Process

The main objective was to optimize the yield (Y_{CH}) and productivity (P_{CH}), obtaining larger values than those obtained in the literature (data without MODO). q_{gas}^{in} and q_{liq}^{in} were proposed as control variables. The yield Y_{CH} was defined as the CH_4 outlet flow rate $(q_{gas,CH4})$ produced over the total COD grams added per day, while productivity P_{CH} was the ratio between $q_{gas,CH4}$ and V_{liq}, expressed as,

$$Y_{CH4} = \frac{q_{gas,CH4}}{S_{liq,su}^{in}\, q_{liq}^{in} + S_{gas,H2}^{in}\, q_{gas}^{in} + S_{gas,CO}^{in}\, q_{gas}^{in}} \quad (2.3.6)$$

$$P_{CH4} = \frac{q_{gas,CH4}}{V_{liq}} \quad (2.3.7)$$

3.4.1 Multi-Objective Optimization

In this study, the simulations were run using an Intel® Core i7 8665U 2.11 GHz, 16 GB RAM computer. The *paretosearch* function from MATLAB® was used to obtain the POS for each stage. The MOO was proposed as,

$$\max_{\{q_{gas}^{in}, q_{liq}^{in}\}} (Y_{CH4}^{*}, P_{CH4}^{*}) \quad (2.3.8)$$

$$Subject\ to \begin{cases} Equations\ (2.3.1) - (2.3.3) \\ Y_{CH4} \leq 0.39\ L/gCOD_{added} \\ 1.0 \leq q_{gas,in} \leq 10.0 \times 10^{3}\ L/d \\ 1.0 \leq q_{liq,in} \leq 100\ L/d \end{cases}$$

The system is subject to the Equations (2.3.1)-(2.3.3), which correspond to the model dynamics (ADM1_ME), and 0.39 $L/gCOD_{added}$ represents the theoretical cumulative CH_4 volume at 37°C or 0.35 $L/gCOD_{added}$ at standard temperature and pressure conditions (Filer *et al.*, 2019).

The MOO was performed for stages I-V. 60 POP were computed for each stage. In which 20, 22, 20, 28, and 23 iterations were executed in 14.65, 19.01, 23.15, 38.29, 56.33 min, respectively. Figure 2.3.2-A displays the POS for each stage. In stages I and IV, the POS is far from the literature value, which indicates that optimization can perform a representative change in both optimum variables. For the other stages, the literature point is near the POS, denoting that the experiment was performed to maximize yield.

Three POP were selected to analyze different cases of biological methanation improvement. The first POP considered the maximization of P_{CH4}. The second POP maximized the Euclidean length, which was performed by normalizing the POS [0,1] and determining the maximum Euclidean length (d_{max}) from the origin on the normalized coordinates. The third POP involved the maximization of Y_{CH4} (orange, yellow, and purple squares in Figure 2.3.2-A). Table 2.3.2 summarizes the selected POP at each stage.

Table 2.3.2. Multi-objective optimization results.

Stage	**POP for maximum P_{CH4}**		**POP for maximum d_{max}**		**POP for maximum Y_{CH4}**	
	Y_{CH4}×10^{-1} $(L/gCOD_{added})$	P_{CH4}×10^{-1} $(L/Lr/d)$	Y_{CH4}×10^{-1} $(L/gCOD_{added})$	P_{CH4} ×10^{-1} $(L/Lr/d)$	Y_{CH4}×10^{-1} $(L/gCOD_{added})$	P_{CH4}×10^{-1} $(L/Lr/d)$
I	3.155	8.011	3.341	7.080	3.443	2.984
II	3.156	8.012	3.340	7.107	3.443	2.984
III	3.159	8.471	3.330	7.599	3.417	4.042
IV	3.149	9.847	3.308	9.042	3.371	5.940
V	3.167	9.844	3.313	8.943	3.371	5.940

3.4.2 Multi-Objective Dynamic Optimization

To perform the dynamic optimization, the *patternsearch* function from MATLAB® was used. The dynamic optimization problem was proposed as,

$$\min_{\{q_{gas}^{in},q_{liq}^{in}\}}\left(\sum_{j=t}^{t+H_p}\left(\frac{|Y_{CH4}^{*}-Y_{CH4}(t)|}{Y_{CH4}^{*}}\right)^2+\left(\frac{|P_{CH4}^{*}-P_{CH4}(t)|}{P_{CH4}^{*}}\right)^2+\sum_{j=t}^{t+H_c}W_{u,1}\,\Delta q_{gas}^{in}(t)^2\;+W_{u,2}\,\Delta q_{liq}^{in}(t)^2\right) \quad (2.3.9)$$

$$Subject\ to \begin{cases} Equations\ (2.3.1) - (2.3.3) \\ Y_{CH4} \leq 0.39\ L/gCOD_{added} \\ 1.0 \leq q_{gas}^{in} \leq 10.0 \times 10^3\ L/d \\ 1.0 \leq q_{liq}^{in} \leq 100\ L/d \end{cases}$$

where $Y_{CH}\ (t)$ and $P_{CH}\ (t)$ were calculated by using (2.3.6) and (2.3.7). Y_{CH4}^{*}, and P_{CH4}^{*} are the POP values for yield and productivity computed by the MOO, $\Delta q_{gas}^{in}(t)^2$ and $\Delta q_{liq}^{in}(t)^2$ represent the differences between the inlet gas and liquid flow rates before and after each step in the dynamic optimization. $W_{u,1}$ and $W_{u,2}$ are the parameters that weight the importance of the control effort for each input in the optimization.

Five cases were studied to assess the dynamic optimization: **Case 1:** ADM1_ME without MODO (literature value). **Case 2:** ADM1_ME with MODO (POP for maximum P_{CH}). **Case 3:** ADM1_ME with MODO (POP for maximum Euclidean length). **Case 4:** ADM1_ME with MODO (POP for maximum Y_{CH4}). **Case 5:** ADM1_ME with MODO switching between the maximum P_{CH} (stages I, V), maximum Euclidean length (stages II, III), and maximum yield (stage IV). In all cases, the initial guess (u_0) was 1.0 L/d for both control variables. The lower and upper bounds of the objective variables Y_{CH4} and P_{CH4}, and the constraints were the same as presented in the MOO. H_p and H_c were considered equal with values corresponding to the time of each stage (see Table 2.3.1).

For cases 2, 3, 4, and 5, the simulation times were 2.12, 3.35, 3.37, and 2.89 min, respectively, in which the weights $W_{u,1}$ and $W_{u,2}$ were manually adjusted to values of 1×10^{-6}. Figure 2.3.2-B shows the dynamical behavior of optimum and control variables.

With regard to case 2 in stage V, the P_{CH} was maximized from 4.9×10^{-1} without MODO to 9.84×10^{-1} $L/Lr\ /d$ with a slight decrease in Y_{CH4} from 3.34×10^{-1} without MODO to 3.17×10^{-1} $L/gCOD_{added}$ with MODO. The q_{liq}^{in} increased from 1.9 L/d without MODO to 8.4 L/d with MODO and remained constant for all stages. The q_{gas}^{in} increased from 37.5 L/d without MODO to 10.0×10^3 L/d with MODO in stage V.

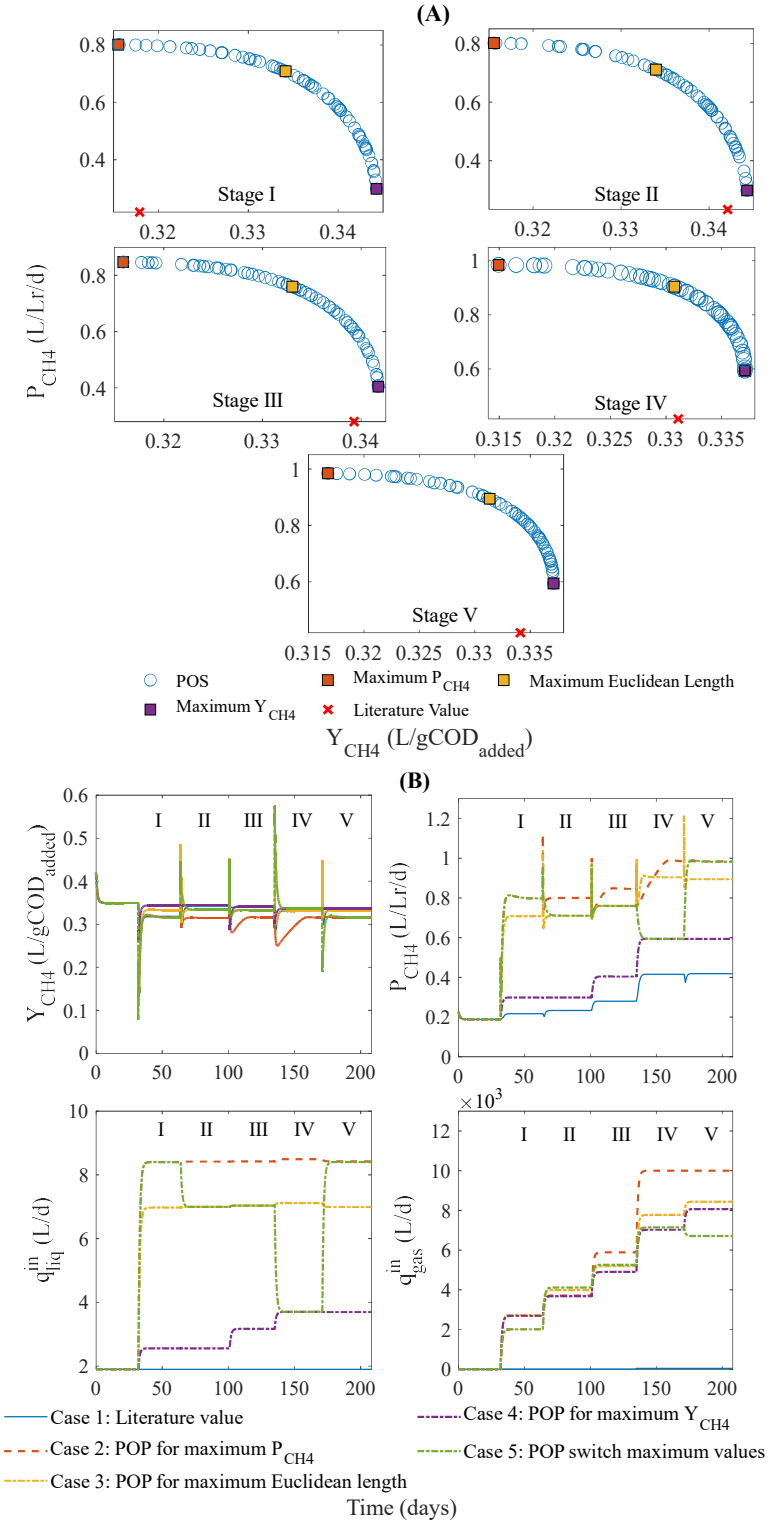


Figure 2.3.2. (A) Pareto optimal set for stages I-V. (B) Methane yield, methane productivity, and inlet liquid and gas flow rates in the MODO strategy.

In case 3, the Y_{CH4} increased from 3.18×10^{-1} to 3.34×10^{-1} $L/gCOD_{added}$ in stage I, but decreased compared to case 1 in stage V. The P_{CH4} increased from 4.19×10^{-1} without MODO to 9.82×10^{-1} $L/Lr\ /d$ in stage V. The q_{liq}^{in} increased from 1.9 L/d without MODO to 7.0 L/d with DMO in stage V. The q_{gas}^{in} increased from 37.5 L/d without MODO to 8.40×10^{3} L/d with DMO in stage V.

In case 4, the Y_{CH4} achieved a value of 3.44×10^{-1} $L/gCOD_{added}$ in stage I, but decreased slightly to 3.37×10^{-1} $L/gCOD_{added}$ in stage V. The P_{CH4} reached a value of 5.94×10^{-1} $L/Lr\ /d$ in stage V. The q_{liq}^{in} increased up to 3.69 L/d in stage V. The q_{gas}^{in} increased from 37.5 L/d in case 1 stage V to 8.01×10^{3} L/d in stage V.

In case 5, the Y_{CH4}, P_{CH4}, and the q_{liq}^{in} follow the behavior of case 2 in stage I, case 3 in stages II and III, case 4 in stage IV, and case 2 in stage V. However, the q_{gas}^{in} differs for all cases and stages.

Table 2.2.4 reports the obtained values for Y_{CH4}, and P_{CH4}, as well as a ratio of their respective values concerning case 1 (literature value). Values larger than one show that the MODO is better than the literature value. For case 2, the Y_{CH4} was 0.99 times lower than that for case 1 in stage I and 0.95 times in stage V. On the other hand, the P_{CH4} increases 3.72 times and decreases to 2.34 times from stages I to V, respectively. Concerning case 3, the Y_{CH4} ratio varied between 1.05 and 0.99 times, while the P_{CH4} ratio changed between 3.26 and 2.13 times concerning without MODO (case 1) for stages I and V, respectively. For case 4, the Y_{CH} was 1.08 times higher than the Y_{CH} for case 1 in stage I and 1.01 times in stage V. On the other hand, the P_{CH} increased between 1.28 and 1.44 for the different stages.

For case 2, the q_{liq}^{in} increased slightly, ranging from 1.9 L/d (case 1) to a value up to 8.4 L/d in stage V. However, the q_{gas}^{in} needed was higher, ranging from the values reported in Table 2.3.1 (case 1) to values between 2.69×10^{3} and 9.99×10^{3} L/d between stages I and V. In case 3, to maintain the maximum Y_{CH4} and P_{CH} , the MODO suggested keeping the q_{liq}^{in} at 7.0 L/d during all stages while changing the q_{gas}^{in} between 2.73×10^{3} and 8.44×10^{3} L/d from stage I to V. For case 4, an q_{liq}^{in} between 2.56 L/d and 3.70 L/d from stage I to V was used to maintain

the maximum Y_{CH4}. The q_{gas}^{in} ranged between 2.69×10^3 and 8.07×10^3 L/d from stage I to V. Case 5 followed the same behavior for the q_{liq}^{in}, whereas the q_{gas}^{in} behavior was different for all cases.

Table 2.3.3. Methane yield and productivity ratio with MODO.

Stage	Case 1	Case 2		Case 3		Case 4		Case 5	
	Y_{CH4} $(L/gCOD_{added})$								
	Value ×10^{-1}	Value ×10^{-1}	Ratio	Value ×10^{-1}	Ratio	Value ×10^{-1}	Ratio	Value ×10^{-1}	Ratio
I	3.18	3.15	0.99	3.33	1.05	3.43	1.08	3.15	0.99
II	3.42	3.15	0.92	3.33	0.97	3.44	1.00	3.33	0.97
III	3.39	3.16	0.93	3.32	0.98	3.41	1.01	3.32	0.98
IV	3.31	3.15	0.95	3.30	1.00	3.36	1.02	3.36	1.02
V	3.34	3.16	0.95	3.31	0.99	3.37	1.01	3.16	0.95
	P_{CH4} $(L/Lr/d)$								
I	2.17	8.07	3.72	7.08	3.26	2.98	1.37	7.98	3.67
II	2.34	8.00	3.42	7.10	3.04	2.98	1.28	7.11	3.04
III	2.80	8.46	3.02	7.60	2.71	4.04	1.44	7.60	2.71
IV	4.15	9.85	2.37	9.04	2.18	5.94	1.43	5.94	1.43
V	4.19	9.82	2.34	8.94	2.13	5.94	1.42	9.82	2.34

3.5 Conclusions

This work presented a MODO strategy for the biological methanation process based on the dynamic model ADM1_ME. Optimizations for two objectives were performed: maximization of Y_{CH4} and P_{CH} by modifying the inlet liquid and gas flow rates. The proposed strategy showed the conflicting behavior of both objectives. Five case studies were compared, it was observed that the maximization of P_{CH4} lowers the Y_{CH4} ratio and vice versa. Case 5 reported a switching strategy between objectives, which allows us to demonstrate the robustness of the process and the well-accounted adaptations of the input variables in simulations. Additionally, it was demonstrated that both input variables have a role in MODO. For instance, the variable inlet gas flow rate made a higher effort than the inlet liquid flow rate. This was observed in case 5, where the behavior of the inlet gas flow rate differed in all cases. These results show the feasibility of the MODO strategy and its use for multiple control objectives.

3.6 References

Acosta-Pavas, J.C., Robles-Rodríguez, Carlos.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, 128106. https://doi.org/10.1016/j.fuel.2023.128106

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). Water Science and Technology 45, 65–73. https://doi.org/10.2166/wst.2002.0292

Camacho, E.F., Bordons, C., 2007. Model Predictive control, Advanced Textbooks in Control and Signal Processing. Springer London, London. https://doi.org/10.1007/978-0-85729-398-5

Chang, K.-H., 2015. Multi-objective Optimization and Advanced Topics, in: E-Design. Elsevier, pp. 1105–1173. https://doi.org/10.1016/B978-0-12-382038-9.00019-3

Czatzkowska, M., Harnisz, M., Korzeniewska, E., Koniuszewska, I., 2020. Inhibitors of the methane fermentation process with particular emphasis on the microbiological aspect: A review. Energy Science and Engineering 8, 1880–1897. https://doi.org/10.1002/ese3.609

Dar, R.A., Parmar, M., Dar, E.A., Sani, R.K., Phutela, U.G., 2021. Biomethanation of agricultural residues: Potential, limitations and possible solutions. Renewable and Sustainable Energy Reviews 135, 110217. https://doi.org/10.1016/j.rser.2020.110217

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multi-objective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6, 182–197. https://doi.org/10.1109/4235.996017

Filer, J., Ding, H.H., Chang, S., 2019. Biochemical Methane Potential (BMP) Assay Method for Anaerobic Digestion Research. Water 11, 921. https://doi.org/10.3390/w11050921

Giraldo, S.A.C., Melo, P.A., Secchi, A.R., 2022. Tuning of Model Predictive Controllers Based on Hybrid Optimization. Processes 10, 351. https://doi.org/10.3390/pr10020351

Luna, R., Lima, B.M., Cuevas-Valenzuela, J., Normey-Rico, J.E., Pérez-Correa, J.R., 2021. Optimal Control Applied to Oenological Management of Red Wine Fermentative Macerations. Fermentation 7, 94. https://doi.org/10.3390/fermentation7020094

Morales-Rodelo, K., Francisco, M., Alvarez, H., Vega, P., Revollar, S., 2020. Collaborative control applied to bsm1 for wastewater treatment plants. Processes 8, 1–22. https://doi.org/10.3390/pr8111465

Rosen, C., Jeppsson, U., 2006. Aspects on ADM1 Implementation within the BSM2 Framework, Technical report. Department of Industrial Electrical Engineering and Automation (IEA), Lund University, Lund, Sweden.

Smets, I.Y., Claes, J.E., November, E.J., Bastin, G.P., Van Impe, J.F., 2004. Optimal adaptive control of (bio)chemical reactors: past, present and future. Journal of Process Control 14, 795–805. https://doi.org/10.1016/j.jprocont.2003.12.005

Sun, H., Yang, Z., Zhao, Q., Kurbonova, M., Zhang, R., Liu, G., Wang, W., 2021. Modification and extension of anaerobic digestion model No.1 (ADM1) for syngas biomethanation simulation: From lab-scale to pilot-scale. Chemical Engineering Journal 403, 126177. https://doi.org/10.1016/j.cej.2020.126177

Tsiantis, N., Balsa-Canto, E., Banga, J.R., 2018. Optimality and identification of dynamic models in systems biology: An inverse optimal control framework. Bioinformatics 34, 2433–2440. https://doi.org/10.1093/bioinformatics/bty139

Vertovec, N., Ober-Blöbaum, S., Margellos, K., 2021. Multi-objective minimum time optimal control for low-thrust trajectory design.

Yamashita, A.S., Zanin, A.C., Odloak, D., 2016. Tuning of Model Predictive Control with Multi-objective Optimization. Brazilian Journal of Chemical Engineering 33, 333–346. https://doi.org/10.1590/0104-6632.20160332s20140212

Section 4 Switching Multi-Objective Dynamic Optimization (MODO) for the Production of Value-Added Products

Juan C. Acosta-Pavas[a], Carlos. E Robles-Rodríguez[a], Jérôme Morchain[a], David Camilo Corrales[b], Claire Dumas[a], Arnaud Cockx[a], César A. Aceves-Lara[a]

[a]TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France
[b]INRAE, UMS (1337) TWB, 135 Avenue de Rangueil, 31077 Toulouse, France

Publications:

Acosta-Pavas, J. C., Robles-Rodríguez, C. E., Morchain, J., Corrales, D. C., Dumas, C., Cockx, A., & Aceves-Lara, C. A. (2023). Switching Multi-Objective Dynamic Optimization (MODO) for the Production of Value-Added Products. In A. C. Kokossis, M. C. Georgiadis, E. Pistikopoulos (eds.), 33rd European Symposium on Computer-Aided Process Engineering (Vol. 52, pp. 583-588). Elsevier. https://doi.org/10.1016/B978-0-443-15274-0.50092-5

Abstract

A Multi-Objective Dynamic Optimization (MODO) applied in the biological methanation process is developed. The MODO strategy was designed to find the trade-off between the maxima yield and productivity of methane and acetate, modifying the inlet liquid and gas flow rates. First, a multi-objective optimization was applied to find the Pareto Optimal Set (POS) between productivity and yield of methane and acetate independently. Then, Pareto optimal points (POP) were selected to develop five cases in dynamic optimization, which approach used a MPC. Cases 1-2 corresponded to the literature value and the use of the Pareto results directly in simulation. Cases 3-4 addressed POP to maximize the Euclidean length for methane and acetate. Case 5 was performed to assess a switch between these objectives. The ability of the MODO strategy to perform the switch was demonstrated. Additionally, the dynamic optimization reduced the inlet gas flow rate to $1.5\times10^3\ L/d$.

Keywords: Biological Methanation, Multi-Objective Optimization, Dynamic Optimization, Value-Added Products

4.1 Introduction

In Anaerobic Digestion (AD), organic matter is transformed by the synergistic work of different microorganisms into CH_4 and CO_2 through four steps: hydrolysis, acidogenesis, acetogenesis, and methanogenesis. The biogas contains 50-75% CH_4 and 25–50 % CO_2 (Iglesias *et al.*, 2021). Biological methanation uses microorganisms to convert the CO_2 in the biogas from AD and syngas (a mixture of H_2, CO, and CO_2) to obtain high-purity CH_4. Nevertheless, this process could also be used to produce value-added products such as acetate (Chaikitkaew *et al.*, 2021). Acetate serves as a chemical platform in the textile, polymer, pharmaceutical, and food industries (Martín-Espejo *et al.*, 2022). However, controlling this type of process is an arduous task due to the multiple reactions and microorganisms involved. As a result, obtaining desired performances of yields or productivities at an industrial scale remains difficult, mainly when it is necessary to optimize several of them simultaneously. Dynamic models play a crucial role in the design of control strategies. For instance, Model Predictive Control (MPC) (Morales-Rodelo *et al.*, 2020) is implemented to maintain or optimize several variables. MPC refers to control actions that respond to the optimization of a criterion related to the system's future behavior determined by the dynamic model (Camacho and Bordons, 2007). Multi-Objective Optimization (MOO) implies optimizing problems where there is more than one objective to be optimized simultaneously. These objectives are usually conflictive (Vertovec *et al.*, 2021).

This work aims at proposing a Multi-objective Dynamic Optimization (MODO) to maximize yield and productivity of the biological methanation process regarding two potential products: CH_4 and acetate. The proposed dynamic optimization approach used a MPC with two control variables corresponding to the inlet liquid and gas flow rates. MPC uses the Pareto Optimal Set (POS), where each solution is considered a Pareto Optimal Point (POP). The MODO strategy's robustness is analyzed by switching between the optimized values for CH_4 and acetate.

4.2 Multi-Objective Dynamic Optimization as Control Strategy

A MODO strategy was proposed in previous work (Acosta-Pavas *et al.*, 2023) to determine the optimal values of the objectives. This strategy entails the following steps:

Step 1-Model definition: Proposition of the dynamic model to represent the biological process.

Step 2 - Definition of the multi-objective optimization problem: Definition of the objective functions $J_1^*, \dots, J_m^*$ to be maximized/minimized by the MOO. A MOO can be formulated as a minimization problem, Equation (2.4.1).

$$\min_{Y,u,\theta,t} \{J_1^*(Y,u,\theta), \dots, J_m^*(Y,u,\theta)\} \quad (2.4.1)$$

$$Subject\ to \begin{cases} dY/dt = \xi(Y,u,\theta,t) \\ \lambda_i\,(Y,u,\theta,t) \le 0 \qquad i = 1,2,\dots,n_\lambda \\ \psi_i(Y,u,\theta,t) = 0 \qquad i = 1,2,\dots,n_\psi \\ u^L \le u \le u^U \end{cases} \quad (2.4.2)$$

where $J_1^*, \dots, J_m^*$ are the m objective functions; Y the state variables; u and θ the control variables and parameters, respectively. The dynamic model is represented by dY/dt; λ_i and ψ_i indicate inequality and equality constraints. u^L and u^U corresponds to the lower and upper bounds of the control variables.

Step 3 - Selection of the Pareto optimal point (POP): Determination of the Pareto optimal set $J_1^*, \dots, J_m^*$ and selection of the POP to be used as the reference trajectory in the dynamic optimization.

Step 4-Definition of the dynamic problem with a single weighted objective: Formulation of an objective function considering the previously identified POP in terms of a MPC problem. Dynamic optimization determines the input variables that minimize the following objective function,

$$\min_u \left(\sum_{j=t}^{t+H_p} \left(\frac{|J^* - J(t+j|t)|}{J^*} \right)^2 + \sum_{j=t}^{t+H_c} W_u\, \Delta u(t+j|t)^2 \right) \quad (2.4.3)$$

where u is the vector of the control variables; H_p and H_c are the prediction and control horizons; $J(t+j|t)$ is the output prediction calculated at time instant $t+j$ using the information available at time instant t. J^* is a reference trajectory that enables to reach the set

point and is determined by the MOO. The term $\Delta u(t+j|t)$ is the control move at time instant $t+j$ calculated using information available at time instant t. The problem in Equation (2.4.3) is also subject to Equation (2.4.2).

Step 5 - Implementation of the optimization: Execution of the dynamic optimization and determination of the optimal values of the control and optimized variables at each time.

4.3 Multi-objective Dynamic Optimization in Biological Methanation Process

The main goal was to optimize yields (Y_{CH} , Y_{ac}) and productivities (P_{CH4}, P_{ac}) of two value-added products, CH_4 and acetate, to demonstrate that a control strategy could help to improve the biological methanation process. Two manipulated variables were proposed for the optimization: the inlet gas $\left(q_{gas}^{in}\right)$ and liquid $\left(q_{gas}^{in}\right)$ flow rates.

Step 1: The model employed for the simulation corresponds to an extension of the Anaerobic Digestion Model No. 1 (ADM1_ME) to consider the conversion of H_2 and CO to improve CH_4 production, which was proposed in previous work (Acosta-Pavas *et al.*, 2022). The model was built upon experimental data from the literature (Sun *et al.*, 2021). The experiment was carried out in a bubble column reactor with a working volume of 37.5 L and a hydraulic retention time (HRT) of 20 days operating at 37°C for 207 days. The organic loading rate (OLR) was 0.53 $gCOD/L/d$ with q_{liq}^{in} of 1.9 L/d. The gas addition was carried out in five stages (I -V), in which q_{gas}^{in} and the gas loading rate (GLR) were varied in time. The model could be rewritten as,

$$\frac{dS_{gas,i}}{dt} = \frac{q_{gas}^{in}}{V_{gas}} S_{gas,i}^{in} + N_i \left(\frac{V_{liq}}{V_{gas}}\right) - \frac{q_{gas}}{V_{gas}} S_{gas,i} \qquad (2.4.4)$$

$$\frac{dS_{liq,j}}{dt} = \frac{q_{liq}^{in}}{V_{liq}} \left(S_{liq,j}^{in} - S_{liq,j}\right) + \sum_k Y_k f_{j,k} \mu_k - N_i \qquad (2.4.5)$$

$$\frac{dX_k}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(X_k^{in} - X_k\right) + Y_k \mu_k - \mu_{k,dec} \quad (2.4.6)$$

Sub-index $j \in [1,8]$ denotes glucose, butyrate, propionate, acetate, H_2, CH_4, CO, and CO_2 in the liquid phase. Sub-index $k \in [1,6]$ reads for the biomass that degrade glucose, butyrate, propionate, acetate, H_2, and CO, respectively. For the gas phase, the sub-index $i \in [1,4]$ corresponds to H_2, CH_4, CO, and CO_2. q_{gas} is the outlet gas flow rate, V_{liq} and V_{gas} the liquid and molar fraction volume; $S_{liq,j}^{in}$, $S_{gas,i}^{in}$, and X_k^{in} are the inlet concentrations of components j in the liquid phase, the inlet concentration of components i in the gas phase, and the inlet concentration of biomass k in the liquid phase; Y_k is the yield of biomass k, $f_{j,k}$ refers to the stoichiometric coefficients; μ_k and $K_{k,dec}$ are the growth rate and decay constant of biomass k, and N_i is the mass transfer rate of component i.

Step 2: Two objectives were considered: the yields Y_{CH} , Y_{ac} and productivities P_{CH4}, P_{ac}, for CH_4 and acetate. These are defined as,

$$Y_{CH4} = \frac{q_{gas,CH4}}{gCOD_{added}} \; ; \; P_{CH4} = \frac{q_{gas,CH4}}{V_{liq}} \quad (2.4.7)$$

$$Y_{ac} = \frac{q_{liq}^{in} \cdot S_{liq,ac}}{gCOD_{added} \cdot V_{liq}} ; \; P_{ac} = \frac{q_{liq}^{in} \cdot S_{liq,ac}}{V_{liq}} \quad (2.4.8)$$

The MOO for Y_h^* and P_h^* maximization was proposed as,

$$\max_{\left\{q_{gas}^{in},\, q_{liq}^{in}\right\}} \left(P_h^*, Y_h^*\right) \quad (2.4.9)$$

$$Subject\ to \begin{cases} Equations\ (2.3.4) - (2.3.6) \\ Y_{CH} \leq 0.39\ L/gCOD_{added} \\ 1.0 \leq q_{gas}^{in} \leq 10.0 \times 10^3\ L/d \\ 1.0 \leq q_{liq}^{in} \leq 100.0\ L/d \end{cases} \quad (2.4.10)$$

where $h = CH_4, acetate$, the value 0.39 $L/gCOD_{added}$ represents the maximum theoretical cumulative CH_4 volume at 37°C or 0.35 $L/gCOD_{added}$ at standard temperature and pressure conditions (only for P_{CH4}, and Y_{CH} maximization).

The *paretosearch* function from MATLAB® was used to obtain the POS for each stage. In this study, the simulations were run using an Intel® Core i7 8665U 2.11 GHz, 16 GB RAM computer.

Step 3: For the selection of the POPs, the POS was computed by the MOO for each of the V stages. In the MOO for CH_4 maximization, 60 POP were obtained for each stage. On the other hand, In the MOO for acetate maximization, 60 POP points were computed for stages I, II, and V, while 29 and 35 POP were computed in stages III and IV.

Figure 2.4.1 shows the POS for CH_4 and acetate at each stage. A progressive increase in P_{CH} was observed while Y_{CH} decreased slightly. The Y_{ac} and P_{ac} increased between stages I and II, then the Y_{ac} decreased in all stages. However, the P_{ac} increased until stage IV and then decreased in stage V.

At each stage, the POPs were selected to maximize the Euclidean length (d_{max}) for CH_4 (red squares in Figure 2.4.1-A) and for acetate (red squares in Figure 2.4.1-B). In both cases d_{max} is calculated from the origin using a normalization as in Equation (2.4.11).

$$d_{max} = max\left(\sqrt{\left(\frac{Y^*_{CH} - min(Y^*_{CH})}{max(Y^*_{CH4}) - min(Y^*_{CH4})}\right)^2 + \left(\frac{P^*_{CH} - min(P^*_{CH})}{max(P^*_{CH4}) - min(P^*_{CH4})}\right)^2}\right) \qquad (2.4.11)$$

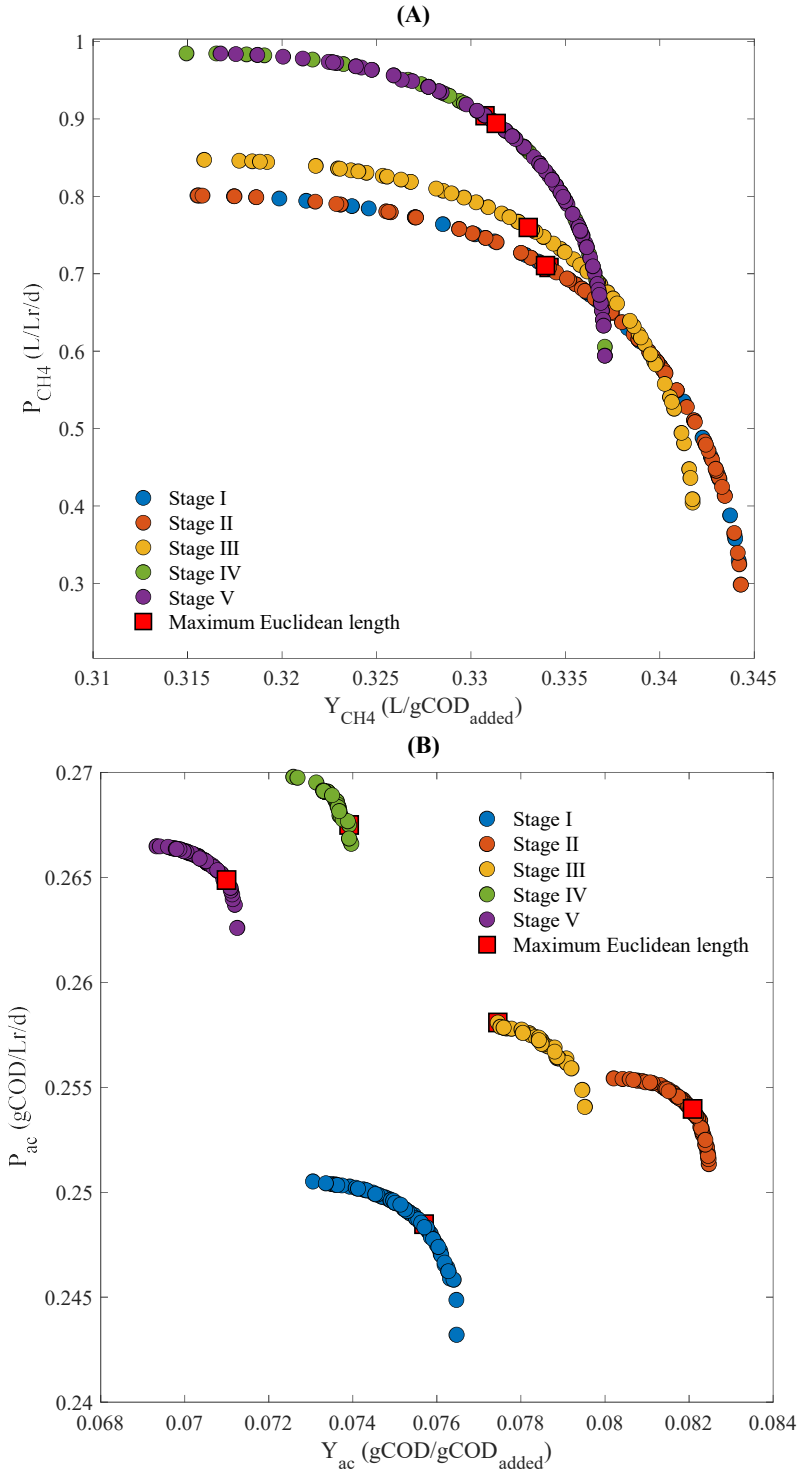


Figure 2.4.1. Pareto optimal sets for CH_4 (A), and acetate (B) at each stage.

Step 4: The dynamic optimization was performed with the *patternsearch* function from MATLAB® was used. The dynamic optimization for Y_h and P_h maximization was proposed as follows,

$$\min_{\{q_{gas}^{in},\, q_{liq}^{in}\}} \left(\sum_{j=t}^{t+H_p} \left(\frac{|Y_h^* - Y_h(t)|}{Y_h^*} \right)^2 + \left(\frac{|P_h^* - P_h(t)|}{P_h^*} \right)^2 + \sum_{j=t}^{t+H_c} W_{u,1}\, \Delta q_{gas}^{in}(t)^2 \; + W_{u,2}\, \Delta q_{liq}^{in}(t)^2 \right) \quad (2.4.12)$$

where $h = CH_4$ and acetate. The optimization was also subject to the constraints in Equation (2.4.10). Y_i^* and P_i^* are the POP values for yield and productivity computed by the MOO, $\Delta q_{gas}^{in}(t)^2$ and $\Delta q_{liq}^{in}(t)^2$ the differences between the inlet gas and liquid flow rates before and after each step in the dynamic optimization. $W_{u,1}$ and $W_{u,2}$ are the parameters that weight the importance of the control effort term in the optimization. In all cases, the initial values for both manipulated variables were 1.0 L/d. H_p and H_c were considered to have equal values and were equivalent to the final time of each stage.

Step 5: Five cases were studied to assess the dynamic optimization. Case 1 was the literature value (data without MODO). For the remaining cases, we wanted to demonstrate first a comparison between the direct use of the POP in simulation (case 2) and the inclusion of POP into a MODO (case 3). Then, a switching strategy was proposed between the maximization for d_{max} CH_4 (case 3) and d_{max} acetate (case 4). Case 5 verifies the robustness of the MODO, switching between the CH_4 maximization in stages I-III and acetate maximization in stages IV-V. For cases 2-5 the weights $W_{u,1}$ and $W_{u,2}$ were manually adjusted to values of 1×10^{-6}.

Figure 2.4.2-A presents the results of cases 1 to 3. Case 1 is the base case regarding the results obtained from the literature. The advantages of using dynamic optimization are observed when cases 2 and 3 are compared. Both of them achieved similar results for P_{CH4}. However, the behavior at each stage change is smoother and faster in case 3 due to the dynamic part of the MODO (zoom in Figure 2.4.2-A). For instance, between stages II and III, the time to reach 95% of the steady state decreased from 104 days in case 2 to 102 days in case 3. Y_{CH4} was similar in all cases, q_{liq}^{in} varied from 7.1 to 6.9 L/d at stage V. For q_{gas}^{in}, a value of 10.0×10^3 L/d was

obtained in case 2 for all stages. Nonetheless, this value was reduced in case 3, ranging from 2.73×10^3 to 8.44×10^3 L/d.

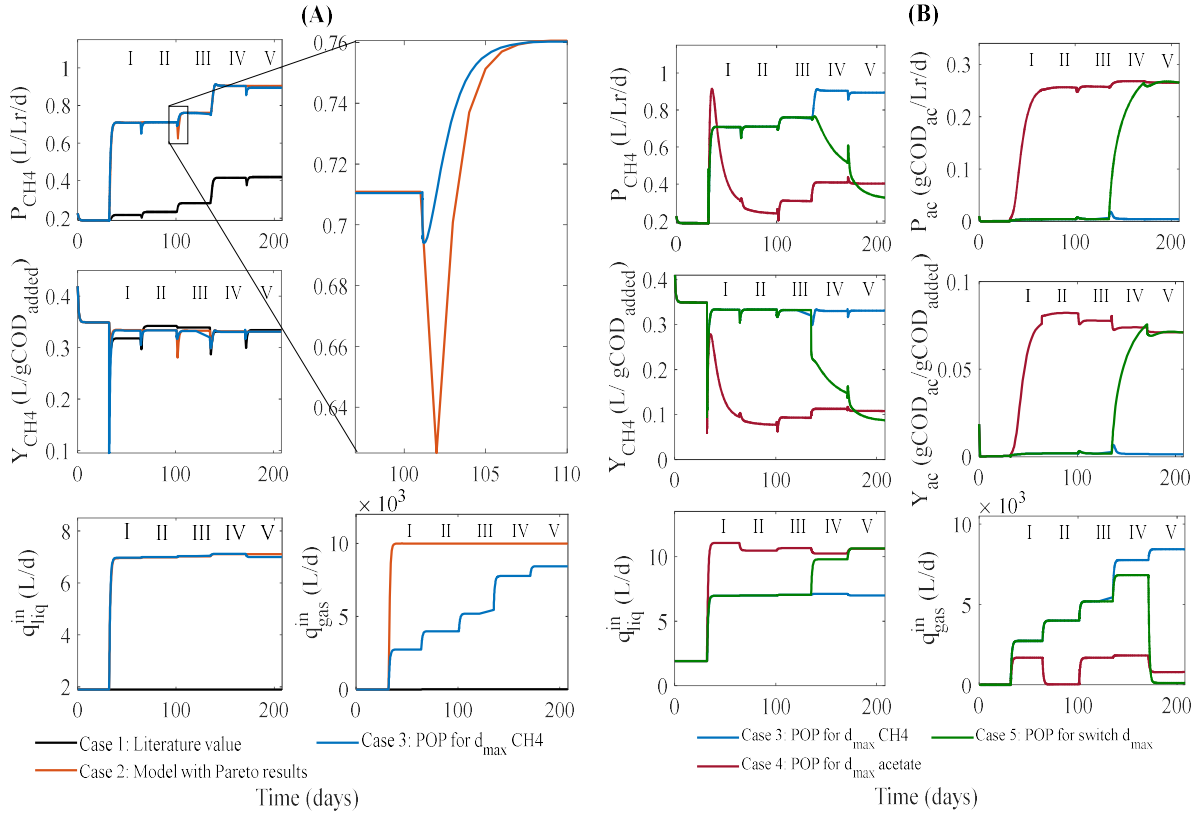


Figure 2.4.2. Inlet liquid and gas flow rates, yields, and productivities in the MODO. (A) Comparison of cases 1-3; (B) comparison of cases 3-5.

Figure 2.4.2-B displays cases 3 to 5. In case 3 stage V, a value of 0.89 $L/Lr/d$ was obtained for P_{CH4}, while 0.33 $L/gCOD_{added}$ was obtained for Y_{CH} . On the other hand, in case 4 stage V a value of 0.27 $gCOD_{ac}/Lr/d$ and 0.07 $gCOD_{ac}/gCOD_{added}$ were achieved for P_{ac} and Y_{ac}, respectively. In stage IV, the switch was applied. Therefore, the reference values for P_{ac} and Y_{ac} were achieved at the end of the stage. Finally, values of 0.27 $gCOD_{ac}/Lr/d$ and 0.07 $gCOD_{ac}/gCOD_{added}$ were obtained in stage V for P_{ac} and Y_{ac}, the same at case 4. The control variables were adapted to each case to maintain similar values. It means that the MODO strategy is robust, permitting the definition of multi-objectives of different types and preserving the reference values determined in the MOO.

When d_{max} for CH_4 was maximized, P_{CH} increased between 3.26 to 2.13 times from stage I to V concerning case 1, while Y_{CH} was maintained similarly. On the other side, when d_{max} for acetate was maximized, P_{ac} is increased between 669 to 680 times from stage I to V concerning case 1, while Y_{ac} is increased until 138 to 228 times from stage I to V.

4.4 Conclusion

A MODO strategy was successfully applied over a biological methanation process based on the dynamic model ADM1_ME. The feasibility of using Paretos to find the trade-off between objective functions such as yields and productivities of CH_4 and acetate was demonstrated, and the subsequent application of a dynamic optimization that allows an improvement in the response by reducing approximately two days the time in which the steady state is reached in the stage changes. Additionally, a reduction in gas flow rates up to $1.5\times10^3\ L/d$ was achieved with dynamic optimization. It demonstrated the robustness of the MODO strategy to switch between products of interest, CH_4 and acetate. It was evidenced the key role of the inlet liquid and gas flow rates as control variables due to its ability to adapt well to each case and stage. Although these results show the feasibility of the proposed strategy, it is important to note that these are simulation results, and the microorganism's adaptation to the proposed switching conditions might differ. However, this work showed the possibility of optimizing the production of these two products more smoothly and the change of objectives.

4.5 References

Acosta-Pavas, J.C., Robles-Rodríguez, Carlos.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, 128106. https://doi.org/10.1016/j.fuel.2023.128106

Acosta-Pavas, J.C., Robles-Rodríguez, C.E., Méndez Suarez, C.A., Morchain, J., Dumas, C., Cockx, A., Aceves-Lara, C.A., 2022. Dynamic Multi-Objective Optimization Applied to Biomethanation Process. Chemical Engineering Transactions 96, 319–324. https://doi.org/10.3303/CET2296054

Camacho, E.F., Bordons, C., 2007. Model Predictive control, Advanced Textbooks in Control and Signal Processing. Springer London, London. https://doi.org/10.1007/978-0-85729-398-5

Chaikitkaew, S., Seengenyoung, J., Mamimin, C., Birkeland, N.-K., Reungsang, A., O-Thong, S., 2021. Simultaneous biogas upgrading and acetic acid production by homoacetogens consortium enriched from peatland soil. Bioresource Technology Reports 15, 100701. https://doi.org/10.1016/j.biteb.2021.100701

Iglesias, R., Muñoz, R., Polanco, M., Díaz, I., Susmozas, A., Moreno, A.D., Guirado, M., Carreras, N., Ballesteros, M., 2021. Biogas from Anaerobic Digestion as an Energy Vector: Current Upgrading Development. Energies 14, 2742. https://doi.org/10.3390/en14102742

Martín-Espejo, J.L., Gandara-Loe, J., Odriozola, J.A., Reina, T.R., Pastor-Pérez, L., 2022. Sustainable routes for acetic acid production: Traditional processes vs a low-carbon, biogas-based strategy. Science of The Total Environment 840, 156663. https://doi.org/10.1016/j.scitotenv.2022.156663

Morales-Rodelo, K., Francisco, M., Alvarez, H., Vega, P., Revollar, S., 2020. Collaborative control applied to bsm1 for wastewater treatment plants. Processes 8, 1–22. https://doi.org/10.3390/pr8111465

Sun, H., Yang, Z., Zhao, Q., Kurbonova, M., Zhang, R., Liu, G., Wang, W., 2021. Modification and extension of anaerobic digestion model No.1 (ADM1) for syngas biomethanation simulation: From lab-scale to pilot-scale. Chemical Engineering Journal 403, 126177. https://doi.org/10.1016/j.cej.2020.126177

Vertovec, N., Ober-Blöbaum, S., Margellos, K., 2021. Multi-objective minimum time optimal control for low-thrust trajectory design, in: European Control Conference (ECC). Delft, Netherland, pp. 1975–1980. https://doi.org/10.23919/ECC54610.2021.9654919

Section 5 Economic Multi-Objective Dynamic Optimization (EMODO) as a Decision-Making tool in Biological Methanation Process

Juan C. Acosta-Pavas[a], Carlos E. Robles-Rodríguez[a], Jérôme Morchain[a], David Camilo Corrales[b], Claire Dumas[a], Arnaud Cockx[a], César A. Aceves-Lara[a]

[a]TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France
[b]INRAE, UMS (1337) TWB, 135 Avenue de Rangueil, 31077 Toulouse, France

Publications:

Acosta-Pavas, J. C., Robles-Rodríguez, C. E., Morchain, J., Corrales, C., Dumas, C., Cockx, A., & Aceves-Lara, C. A. (2023). Economic Multi-Objective Dynamic Optimization (EMODO) as a Decision-Making tool in Biomethanation Process. Chemical Engineering Transactions (Accepted).

Abstract

In biological methanation, the methane produced by anaerobic digestion is upgraded with the addition of syngas. The successful implementation of biological methanation requires optimizing the production to be economically competitive against chemical processes. Optimization is an arduous task, especially when it is desired to optimize multiple objectives that can be conflicting, such as yields, productivities, process times, and profit gains, among others. In this context, this work aims to implement an Economic Multi-Objective Dynamic Optimization (EMODO) approach as a decision-making tool for adequately operating the biological methanation process. The proposed EMODO strategy was based on a previously developed dynamic model for biological methanation. This strategy effectively optimized the $Gain$ and the $Profit\ margin$ by manipulating the inlet flow rates of gas (q_{gas}^{in}) and liquid (q_{liq}^{in}). The strategy also highlights the conflicting behavior of economic objectives and the dependence on substrates. The dynamic optimization improves the response time of the model smoothing the transitions between stages and achieving well adaptation to disturbances regarding the substrates' cost and the products' selling prices.

Keywords: Biological Methanation, Economic Model Predictive Control, Economic Multi-Objective Optimization, Dynamic Optimization, Market evolution

5.1 Introduction

The successful implementation of biological processes requires optimization to be competitive against chemical processes in economic terms. Emerging bioprocesses such as biological methanation can benefit from multi-objective optimization by maximizing or minimizing multiple variables of interest simultaneously.

Biological methanation or Biomethanation is a process in which the biogas produced through the Anaerobic Digestion (AD) is upgraded by the biological conversion of CO_2 using syngas (a combination of H_2, CO, and CO_2) to obtain high-purity CH_4 (Rafrafi *et al.*, 2020). The biogas produced in the AD contains between 50 - 75% of CH_4, 25 – 50 % of CO_2, and 2–7% water vapor (Laguillaumie *et al.*, 2022). Through biological methanation, the biogas can be upgraded into biomethane (95 – 99 %) while removing CO_2 with the addition of H_2 or syngas (CO/H_2) (Sun *et al.*, 2021). The hydrogenotrophic methanogens with CO_2 consumption transform the H_2. The CO can be transformed indirectly into H_2 by carboxydotrophic hydrogenogenesis, then into acetate by CO-acetogenesis and CO-homoacetogenesis, and finally transformed into CH_4 through hydrogenotrophic and acetoclastic methanogenesis (Guiot *et al.*, 2011). Other works have shown that biological methanation can also be used to produce acetate (Laguillaumie *et al.*, 2022), a molecule of interest that could help make this process more economically profitable. Based on this complex biological system, managing the biological methanation process is still an arduous task. Therefore, achieving desired objectives such as high productivities, high-profit margins, or low flow rates remains difficult at an industrial scale, especially when it is desired to optimize several variables simultaneously.

Multi-Objective Optimization (MOO) involves optimizing problems where there is more than one objective to be optimized simultaneously, and these objectives are usually conflictive. The use of dynamic models plays a crucial role in designing control strategies. For instance, Model Predictive Control (MPC) (Morales-Rodelo *et al.*, 2020) is implemented to maintain or optimize several variables simultaneously (*e.g.*, productivities and yields). MPC refers to control actions that optimize a criterion in the system's future behavior, which is determined by the dynamic model (Camacho & Bordons, 2007). Economic MPC (EMPC) has recently been

proposed incorporating a general cost function or performance index in its formulation to consider economic criteria in process optimization (Ellis *et al.*, 2017).

MOO has been applied in bioprocess to find the trade-off between yields and productivities (Nimmegeers *et al.*, 2018). In the AD considering the determination of Pareto Optimal Sets (POS) to find the trade-off between the green degree as environmental impact and net present value as an economic aspect (Li *et al.*, 2018). In biological methanation, MOO has been applied to minimize energy consumption and maximize the green degree and CH_4 production (Yan *et al.*, 2016). However, these works not consider the dynamic optimization of the process, improving the performance of economic objectives.

This work aims at implementing an Economic Multi-Objective Dynamic Optimization (EMODO) strategy as a decision-making tool for the biological methanation process to guarantee the maximization of the $Gain$ and $Profit\ margin$. The $Profit\ margin$ was calculated based on changes in market prices using glucose, H_2, and CO as substrates and CH_4 and acetate as products. The $Gain$ was calculated with the price of CH_4 and acetate production.

POS associated with three process stages were determined through MOO. Each POS solution was considered a Pareto Optimal Point (POP). The POS is considered the first part of the decision-making tool, where it is necessary to select the best POP that maximizes the $Gain$ and $Profit\ margin$. In dynamic optimization is used a MPC, which is referred to as the second part of the decision-making tool that optimizes the performance of economic objectives with two control variables corresponding to q_{gas}^{in} and q_{liq}^{in}. To verify the efficacy of the EMODO strategy, the biological methanation process is simulated considering disturbances of ±20 % in the substrates, sugar, H_2, and CO cost, and the selling price of the products CH_4 and acetate.

5.2 Economic Multi-Objective Dynamic Optimization (EMODO)

Several variables can be optimized in the biological methanation process: yields, productivities, process times, *etc*. Most of these variables are often conflicting. A Multi-Objective Dynamic Optimization (MODO) strategy was proposed in previous work (Acosta-Pavas *et al.*, 2022) address the mentioned problem. However, this methodology does not consider any information

about market evolution. The formulation of a cost function could directly or indirectly reflect the process economy to consider economic optimization. Therefore, in this study, the MODO strategy is modified to consider economic aspects such as substrates costs or prices market through the Economic Multi-Objective Dynamic Optimization (EMODO) as the following five steps.

Step 1 - Model definition: Biological methanation was modeled by a dynamic model based on an extension of the Anaerobic Digestion Model No. 1 (ADM1_ME) (Acosta-Pavas *et al.*, 2023). This model considers the uptake of sugar, volatile fatty acids, such as butyrate, propionate, and acetate, the uptake of H_2 and CO, and the decay of biomass and *in-situ* syngas addition. The ADM1_ME describes three types of variables: soluble ($S_{liq,j}$), particulated biomass (X_k) and gas ($S_{gas,i}$) components. The ADM1_ME is summarized as Equations (2.5.1)-(2.5.3).

$$\frac{dS_{liq,j}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(S_{liq,j}^{in} - S_{liq,j}\right) + \sum_{k} Y_k f_{j,k} \mu_k - N_i \quad (2.5.1)$$

$$\frac{dX_k}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(X_k^{in} - X_k\right) + Y_k \mu_k - \mu_{k,dec} \quad (2.5.2)$$

$$\frac{dS_{gas,i}}{dt} = \frac{q_{gas}^{in}}{V_{gas}} S_{gas,i}^{in} + N_i \left(\frac{V_{liq}}{V_{gas}}\right) - \frac{q_{gas}}{V_{gas}} S_{gas,i} \quad (2.5.3)$$

Sub-index $j \in [1,8]$ represents glucose, butyrate, propionate, acetate, H_2, CH_4, CO, and CO_2 in the liquid phase. The H_2, CH_4, and CO are expressed in $gCOD/L$, and CO_2 is expressed in mol/L. Chemical Oxygen Demand (COD) is the amount of oxygen needed to degrade the organic matter into CO_2 and H_2O. It is important to mention that CO_2 is expressed in mol instead of COD, as suggested by Batstone *et al.* (2002). Sub-index $k \in [1,6]$ denotes for the biomass that degrade glucose, butyrate, propionate, acetate, H_2, and CO, respectively. For the gas phase, the sub-index $i \in [1,4]$ corresponds to H_2, CH_4, CO, and CO_2. The inlet flow rates of liquid and gas are represented by q_{liq}^{in} and q_{gas}^{in}, respectively, while q_{gas} denotes the outlet gas flow rate. V_{liq} and V_{gas} are the liquid and gas volumes, respectively. $S_{liq,j}^{in}$, $S_{gas,i}^{in}$ and X_k^{in}

represent the inlet concentration of the component j in the liquid phase, the inlet concentration of component i in gas phase, and the inlet concentration of biomass k in the liquid phase, respectively. Y_k is the yield of biomass k, $f_{j,k}$ corresponds the stoichiometric coefficients; μ_k and $\mu_{k,dec}$ refer to the growth and decay rate of biomass k, and N_i to the mass transfer rate of component i.

The simulations of the biological methanation process were carried out using the ADM1_ME considering a bubble column reactor (BCR) with a working volume of 37.5 L and a hydraulic retention time (HRT) of 20 days operating at 37°C for 330 days. The organic loading rate (OLR) was varied over time in all stages, according to Table 2.5.1. The reference stage corresponded to the simulation without gas addition, with a q_{liq}^{in} of 1.88 L/d. The flow rates q_{liq}^{in}, q_{gas}^{in}, and the gas loading rate (GLR) will be optimized by the EMODO strategy for stages I – III.

Table 2.5.1. Stages and OLR simulated with the ADM1_ME.

Stage	Time (Day)	$\boldsymbol{OLR}$ $(\boldsymbol{gCOD/L/d})$
Reference	1-30	0.53
I	30-130	1.07
II	130-230	1.60
III	230-330	2.13

To propose economic variables, literature values of 3.40×10^{-4},1.63×10^{-4}, 5.96×10^{-4}, and 1.63×10^{-3} $EUR/gCOD$ were suggested for the cost of sugar, syngas, the selling price of CH_4, and selling price of acetate, respectively (see Annexes Section 3). Then, to verify the efficacy of the EMODO strategy, selling prices were simulated, considering disturbances in the price. First, an increase of 20% (+20%) in the selling price of CH_4 and a reduction of 20% (-20%) in the selling price of acetate were considered from 70-100 days (Disturb 1). Then, an increase of 20% in the cost of syngas was simulated from 190-210 days (Disturb 2). Finally, a decrease of 20% in the selling price of CH_4 and an increase of 20% in the cost of syngas were considered from 260-290 days (Disturb 3).

Step 2 - Definition of the multi-objective optimization problem: The definition of economic optimization corresponds to the maximization of the gain of CH_4 and acetate ($Gain$), and the profit margin of CH_4 and acetate ($Profit\ margin$) by modifying the q_{gas}^{in} and q_{liq}^{in}. The economic multi-objective optimization to find the POS was proposed as,

$$\max_{\{q_{gas}^{in},\, q_{liq}^{in}\}} (Gain, profit\ margin) \qquad (2.5.4)$$

$$Subject\ to \begin{cases} Equations\ (2.4.1) - (2.2.3) \\ 1 \leq q_{gas}^{in} \leq 100\ L/d \\ 1.88 \leq q_{liq}^{in} \leq 10\ L/d \end{cases} \qquad (2.5.5)$$

The objective variables are,

$$Gain = \frac{acetate\ selling\ price \cdot S_{liq,ac}}{HRT} + \frac{CH_4\ selling\ price \cdot q_{gas,CH4} \cdot 64}{22.4 \cdot V_{liq}} \qquad (2.5.6)$$

$$Profit\ margin = \frac{(CH_4\ sales + acetate\ sales) - Substrates\ cost}{CH_4\ sales + acetate\ sales} \cdot 100\% \qquad (2.5.7)$$

where 64 $gCOD/mol$ is the COD for CH_4 and 22.4 L/mol is the molar volume of an ideal gas at standard conditions. $Substrates\ cost$ refers to the cost of glucose and syngas (Equation (2.5.8)). $CH_4\ sales$ and $acetate\ sales$ are the gains in EUR for selling all the CH_4 and acetate produced, Equations (2.5.9)-(2.5.10).

$$Substrates\ cost = \big((Sugar\ cost \cdot OLR) + (Syngas\ cost \cdot GLR)\big) \cdot V_{liq} \qquad (2.5.8)$$

$$CH_4\ sales = \frac{CH_4\ selling\ price \cdot\ q_{gas,CH4} \cdot 64}{22.4} \qquad (2.5.9)$$

$$acetate\ sales = \frac{acetate\ selling\ price \cdot S_{liq,ac}\ \cdot V_{liq}}{HRT} \qquad (2.5.10)$$

Step 3 - Selection of the Pareto optimal point (POP): In this study, the simulations were run using an Intel® Core i7 8665U 2.11 GHz, 16 GB RAM computer. The *paretosearch* function

from MATLAB® was used to obtain the POS for each stage. Figure 2.5.1 presents the three Pareto fronts computed for each stage, where 60 POP were calculated.

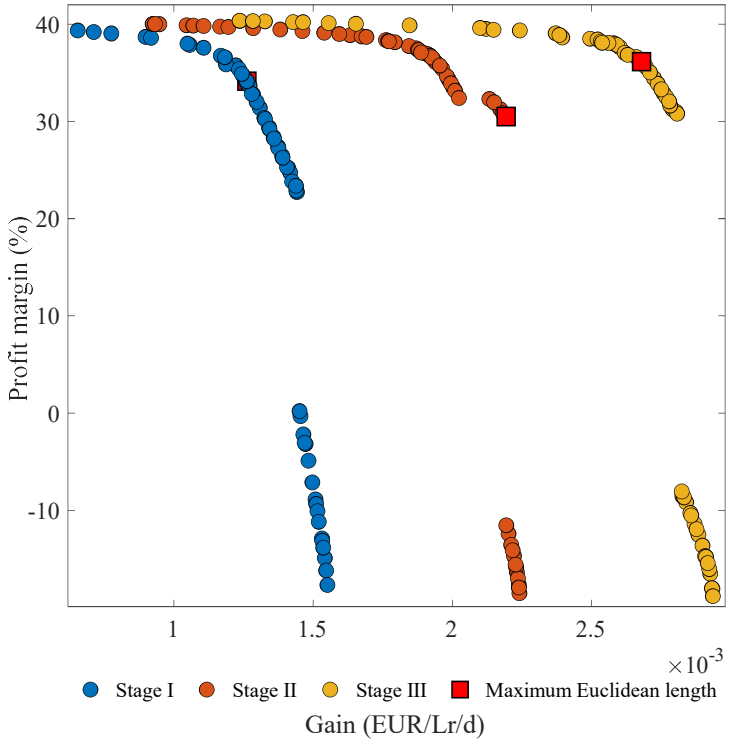


Figure 2.5.1. Pareto optimal sets for stages I-III and maximum Euclidean length.

At each stage, one POP was selected, which corresponded to the maximization of the Euclidean length (d_{max}) for the $Gain$ and the $Profit\ margin$ (red squares in Figure 2.5.1). For all the stages, d_{max} was calculated as the distance from the origin, using a normalization as in Equation (2.5.11).

$$d_{max} = max\left(\sqrt{\left(\frac{Gain^* - min(Gain^*)}{max(Gain^*) - min(Gain^*)}\right)^2 + \left(\frac{Profit\ margin^* - min(Profit\ margin^*)}{max(Profit\ margin^*) - min(Profit\ margin^*)}\right)^2}\right) \quad (2.5.11)$$

Step 4 - Definition of the dynamic problem with a single weighted objective: To consider a dynamic optimization, the two previously defined objectives and its POP were merged into one objective function and solved based on an MPC problem. The proposed dynamic optimization determines the input variables that minimize the following objective function,

$$\min_{\{q_{gas}^{in},\, q_{liq}^{in}\}} \left(\sum_{j=t}^{t+H_p} \left(\frac{|Gain^* - Gain(t)|}{Gain^*} \right)^2 + \left(\frac{|Profit\ margin^* - Profit\ gain(t)|}{Profit\ gain^*} \right)^2 + \sum_{j=t}^{t+H_c} W_{u,1}\, \Delta q_{gas}^{in}(t)^2 + W_{u,2}\, \Delta q_{liq}^{in}(t)^2 \right) \quad (2.5.12)$$

Equation (2.5.12) is subject to the constraints in Equation (2.5.5). $Gain^*$ and $Profit\ margin^*$ denote the POP values for $Gain$ and $Profit\ margin$ computed by the MOO, $\Delta q_{gas}^{in}(t)^2$ and $\Delta q_{liq}^{in}(t)^2$ are the differences between q_{gas}^{in} and q_{liq}^{in}, respectively, before and after each step of the dynamic optimization. $W_{u,1}$ and $W_{u,2}$ are the parameters that weigh the importance of the control effort term in the optimization. The initial values for both manipulated variables, q_{gas}^{in} and q_{liq}^{in} were 1 L/d and 1.88 L/d, respectively.

Step 5 - Implementation of the optimization: Two cases were analyzed. Case 1 corresponded to the use of the POP identified in step 3 and applied directly in the simulation with the ADM1_ME (Pareto results). Case 2 referred to dynamic optimization as a control strategy (Dynamic opt). The weights $W_{u,1}$ and $W_{u,2}$ were manually adjusted to values of 1×10^{-7}. The prediction (H_p) and control (H_c) horizons were considered to have equal values and were equivalent to the final time of each stage (Table 2.5.1). Optimization was performed with the *patternsearch* algorithm in MATLAB®.

The results of the optimization are displayed in Figure 2.5.2. In both cases, the $Gain$ increased at each stage change, while the $Profit\ margin$ varied between 30 and 35% (Figure 2.5.2-C). For both economic variables, it is observed that the dynamic optimization improved the model's response, smoothing the transition between stages, which is ideal in this type of biological process to avoid additional disturbances.

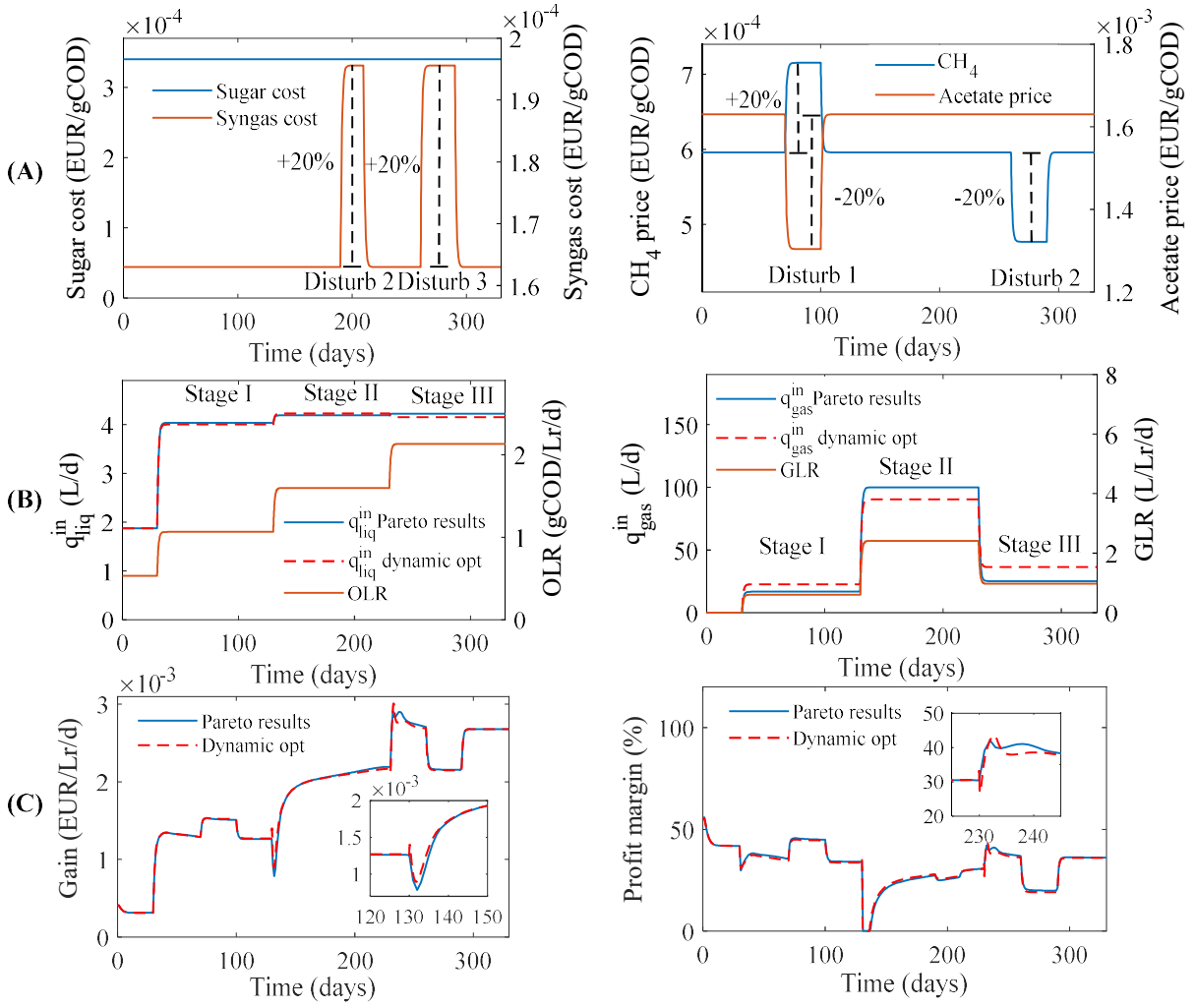


Figure 2.5.2. ADM1_ME inputs and outputs. (A) ADM1_ME Economic inputs (B) ADM1_ME inputs (C) ADM1_ME Economic outputs. Case 1: Pareto results, case 2: dynamic optimization as a control strategy (Dynamic opt). Disturbance 1-3 (Disturb 1-3).

Additionally, the EMODO strategy responds satisfactorily to the three proposed disturbances regarding the cost of substrates and the selling price of products, especially with the disturbance presented between 190-210 days, where there was a 20% increase in syngas cost and subsequent transition between stages II and III (Figure 2.5.2-A).

In Figure 2.5.2-B, comparing case 2 to case 1 in terms of control variables, a slight reduction of 3.3×10^{-2} and 7.5×10^{-2} L/d was observed for q_{liq}^{in} in stages I and III, respectively. A slight

increase of 3.6 $\times 10^{-2}$ L/d was observed in stage II. In contrast, q_{gas}^{in} showed an increase of 5.85 and 11.15 L/d in stages I and III, respectively, and a reduction of 9.49 L/d in stage II.

If the EMODO strategy is considered as a decision-making tool in the biological methanation process, it is necessary to refer to the ADM1_ME inputs (Figure 2.5.2-B) and the ADM1_ME outputs at a steady state (Figure 2.5.2-C). For stages I and III, there were slight decreases in q_{liq}^{in}, while the OLR doubled, and the q_{gas}^{in} increased from 22 to 25 L/d, respectively, resulting in an increase in GLR from 0.60 to 0.97 $L/L_r/d$. This led to an increase in $Gain$ from 1.27×10^{-3} to 2.68×10^{-3} $EUR/L_r/d$, while the $Profit\ margin$ slightly increased from 33.8% to 36.0%.

From a $Gain$ point of view, it can be increased by maintaining similar $Profit\ margin$. However, it should be noted that a significant increase in q_{gas}^{in} is needed to achieve these changes, as in stage II, where values of 91 L/d were obtained.

5.3 Conclusions

The EMODO strategy demonstrates to be a good alternative to obtain the best $Gain$ and $Profit\ margin$ by manipulating q_{gas}^{in}, and q_{liq}^{in}. These variables played a key role and ranged between optimal values of 22 - 91 L/d and 4.00 and 4.22 L/d through all stages. The proposed strategy shows the conflicting behavior of both economic objectives and the high dependence of the substrates added to the process (the three POS, clearly differentiated for each stage). The application of dynamic optimization improves the response, smoothing the transitions between stages. The efficacy of the EMODO strategy is demonstrated with a successful adaptation to three disturbances in the substrate's cost and the product's selling price. These results show the feasibility of the proposed methodology as a decision-making tool and its use for multiple control objectives.

5.4 References

Acosta-Pavas, J.C., Robles-Rodríguez, Carlos.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, 128106. https://doi.org/10.1016/j.fuel.2023.128106

Acosta-Pavas, J.C., Robles-Rodríguez, C.E., Méndez Suarez, C.A., Morchain, J., Dumas, C., Cockx, A., Aceves-Lara, C.A., 2022. Dynamic Multi-Objective Optimization Applied to Biomethanation Process. Chemical Engineering Transactions 96, 319–324. https://doi.org/10.3303/CET2296054

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). Water Science and Technology 45, 65–73. https://doi.org/10.2166/wst.2002.0292

Camacho, E.F., Bordons, C., 2007. Model Predictive control, Advanced Textbooks in Control and Signal Processing. Springer London, London. https://doi.org/10.1007/978-0-85729-398-5

Ellis, M., Liu, J., Christofides, P.D., 2017. Economic Model Predictive Control, Advances in Industrial Control. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-41108-8

Guiot, S.R., Cimpoia, R., Carayon, G., 2011. Potential of Wastewater-Treating Anaerobic Granules for Biomethanation of Synthesis Gas. Environ. Sci. Technol. 45, 2006–2012. https://doi.org/10.1021/es102728m

Laguillaumie, L., Rafrafi, Y., Moya-Leclair, E., Delagnes, D., Dubos, S., Spérandio, M., Paul, E., Dumas, C., 2022. Stability of ex situ biological methanation of H2/CO2 with a mixed microbial culture in a pilot scale bubble column reactor. Bioresource Technology 354, 127180. https://doi.org/10.1016/j.biortech.2022.127180

Li, W., Huusom, J.K., Zhou, Z., Nie, Y., Xu, Y., Zhang, X., 2018. Multi-objective optimization of methane production system from biomass through anaerobic digestion. Chinese Journal of Chemical Engineering 26, 2084–2092. https://doi.org/10.1016/j.cjche.2018.01.001

Morales-Rodelo, K., Francisco, M., Alvarez, H., Vega, P., Revollar, S., 2020. Collaborative control applied to bsm1 for wastewater treatment plants. Processes 8, 1–22. https://doi.org/10.3390/pr8111465

Nimmegeers, P., Vallerio, M., Telen, D., Impe, J., Logist, F., 2018. Interactive Multi-objective Dynamic Optimization of Bioreactors under Parametric Uncertainty. Chemie Ingenieur Technik cite.201800082. https://doi.org/10.1002/cite.201800082

Rafrafi, Y., Laguillaumie, L., Dumas, C., 2020. Biological Methanation of H2 and CO2 with Mixed Cultures: Current Advances, Hurdles and Challenges. Waste and Biomass Valorization. https://doi.org/10.1007/s12649-020-01283-z

Sun, H., Yang, Z., Zhao, Q., Kurbonova, M., Zhang, R., Liu, G., Wang, W., 2021. Modification and extension of anaerobic digestion model No.1 (ADM1) for syngas biomethanation simulation: From lab-scale to pilot-scale. Chemical Engineering Journal 403, 126177. https://doi.org/10.1016/j.cej.2020.126177

Yan, N., Ren, B., Wu, B., Bao, D., Zhang, X., Wang, J., 2016. Multi-objective optimization of biomass to biomethane system. Green Energy & Environment 1, 156–165. https://doi.org/10.1016/j.gee.2016.05.001

Section 6 Fault Detection in Biological Methanation Process using Machine Learning: A Comparative Study of Different Algorithms

J.C. Acosta-Pavas[a], David G[b], Zoraida Callejas[b], David Camilo Corrales[c], C.E. Robles-Rodríguez[a], J. Morchain[a], and C.A. Aceves-Lara[a]

[a]TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France
[b]University of Granada, CITIC-UGR, Dept. Software Engineering, Periodista Daniel Saucedo Aranda sn, 18071 Granada, Spain
[c]INRAE, UMS (1337) TWB, 135 Avenue de Rangueil, 31077 Toulouse, France

Publications:

Acosta-Pavas, J. C., Griol, D., Callejas, Z., Corrales, D. C., Robles-Rodríguez, C. E., Morchain, J., & Aceves-Lara, C. A. (2023). Fault Detection in Biological Methanation Process Using Machine Learning: A Comparative Study of Different Algorithms. In P. García Bringas, et al. (eds.), 18th International Conference on Soft Computing Models in Industrial and Environmental Applications. Lecture Notes in Networks and Systems (Vol. 749, pp. 132-142). Springer, Cham. https://doi.org/10.1007/978-3-031-42529-5_13

Abstract

In this paper we present a study that evaluates different machine learning models for fault detection based on the optimal operation of the biological methanation process. The optimal operation has been obtained from a multi-objective dynamic optimization based on an extended model of the anaerobic digestion model (ADM1 ME). Two datasets have been generated for the ADM1 ME model by generating disturbances on the inlet liquid flow rate (dataset 1) and the inlet gas flow rate (dataset 2). Variations of ±10, ±15, and ±20% of both optimal inlets have been assumed. These datasets have been used to train several algorithms: decision tree CART, Random Forest (RF), Gaussian Naïve Bayes (GNB), k-Nearest Neighbors (k-NN), Quadratic Discriminant Analysis (QDA), Support Vector Machine (SVM), and Neural Network (NN). In dataset 1, CART, RF, and Radial Basis Function (RBF) SVM have achieved accuracies higher than 0.90 and 0.85 in the training and test, respectively. In dataset 2, accuracies higher than 0.90 and 0.87 have been obtained for the RF, QDA, and RBF SVM models in the training and test, respectively.

Keywords: Biological Methanation, Multi-Objective Dynamic Optimization, Soft Sensors, Machine Learning Algorithms, Fault Detection

6.1 Introduction

Biological methanation is a bioprocess recognized for its potential to produce methane (Rusmanis *et al.*, 2019). In this process, the biogas produced through the Anaerobic Digestion (AD) is upgraded by the biological conversion of CO_2 and syngas (a combination of H_2, CO, and CO_2) to obtain high-purity CH_4 (Rafrafi *et al.*, 2020). The biogas produced in the AD contains between 50 – 75% of CH_4, 25 – 50 % of CO_2, and 2–7% water vapor. The hydrogenotrophic methanogens transform the H_2 with CO_2 consumption. The CO can be converted indirectly into H_2 by carboxydotrophic hydrogenogenesis, then into acetate by CO-acetogenesis and CO-homoacetogenesis, and finally transformed into CH_4 through hydrogenotrophic and acetoclastic methanogenesis (Guiot *et al.*, 2011).

Biological methanation involves a large number of microorganisms interacting simultaneously. Small variations in the inlet flow rates, or variations in operational conditions, such as changes in the temperature or pH, can generate significant variations in process outputs, resulting in a product with undesired specifications (biogas with low CH_4 content), which at the industrial level imply high operational costs.

Soft sensors are models that estimate a hard-to-measure property using relatively easy measurements (Kazemi *et al.*, 2020). Soft sensors have been recently proposed based on Machine Learning techniques to study these types of problems in AD.

Kazemi *et al.* (2020) explored machine learning methods such as back-propagation neural network, Support Vector Machine (SVM), Random Forest (RF), extreme learning machines, and genetic programming to monitor volatile fatty acids (VFA) using on-line measured variables. Cinar *et al.* (2022) used seven different machine learning algorithms: linear regression, logistic regression, nearest neighbors, decision trees, random forest, SVM, and XGBoost to define and predict the possible impacts of wide-range temperature fluctuations on process stability in the AD process compared to experimental data. Wang *et al.* (2020) performed a study to predict CH_4 production in AD using seven operation parameters: temperature, C/N ratio, total nitrogen and carbon, glucan, lignin, xylan, and cellulose content.

Four machine learning algorithms were selected for regression and classification: RF, logistic regression multiclass, SVM, and k-NN. The k-NN algorithm demonstrated better prediction in the regression models, while the logistic regression multiclass algorithm showed higher accuracy in classification models. Kazemi *et al.* (2021) used several data-driven approaches to detect faults in the evolution of the total volatile fatty acids (VFA) concentrations of the AD process. VFA concentration was used as a state indicator of the AD process since they are highly susceptible to system input variations. The soft sensors were trained with a dataset Benchmark Simulation Model No.2 (BSM2) developed by the International Water Association (IWA). Three VFA soft sensors were tested and compared: support vector machine (SVM), extreme learning machine (ELM), and the ensemble of neural network (ENN). SVM presented the best results in terms of accuracy and robustness.

Although several of these machine learning models have been successfully applied over the AD process, the application to model biological methanation process is still unexplored.

The present study aims to compare several machine learning models to detect small disturbances respect to deviations from the optimal operation of biological methanation. The optimal operation refers to the optimal values of the inlet liquid and gas flow rates determined by the Multi-objective Dynamic Optimization (MODO) strategy. Additionally, disturbances were assumed to occur from variations in the inlet liquid and gas flow rates.

The remainder of the paper is as follows. Section 6.2 presents the dynamical biological methanation model used to generate the datasets and explains the optimal operation and training dataset generations. Section 6.3 describes the experimental setup and the discussion of the results. Finally, Section 0 presents the main conclusions of the study and future research lines.

6.2 Biological Methanation Model and Optimization

This section describes the dynamic model and the optimization approach used to find the optimal operation of the biological methanation process. Finally, the explanation of disturbances in inlet liquid and gas flow rates is presented to generate the datasets.

6.2.1 Extended Anaerobic Digestion Model (ADM1 ME)

An extension of the Anaerobic Digestion Model No. 1 (ADM1_ME) to consider *in-situ* syngas addition to the biological methanation process was proposed in our previous work (Acosta-Pavas *et al.*, 2023). The ADM1 ME model considers the uptake of sugar, VFA (such as butyrate, propionate, and acetate), H_2 and CO, and biomass decay. The model describes three types of variables: soluble $S_{liq,j}$, particulate biomass X_k and gas $S_{gas,i}$ components. The dynamic model is summarized in Equations (2.6.1)-(2.6.3).

$$\frac{dS_{liq,j}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(S_{liq,j}^{in} - S_{liq,j}\right) + \sum_{k} Y_k f_{j,k} \mu_k - N_i \quad (2.6.1)$$

$$\frac{dX_k}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(X_k^{in} - X_k\right) + Y_k \mu_k - \mu_{k,dec} \quad (2.6.2)$$

$$\frac{dS_{gas,i}}{dt} = \frac{q_{gas}^{in}}{V_{gas}} S_{gas,i}^{in} + N_i \left(\frac{V_{liq}}{V_{gas}}\right) - \frac{q_{gas}}{V_{gas}} S_{gas,i} \quad (2.6.3)$$

Sub-index $j \in [1,8]$ denotes glucose, butyrate, propionate, acetate, H_2, CH_4, CO, and CO_2 in the liquid phase. The H_2, CH_4, and CO are expressed in $gCOD/L$, and CO_2 is expressed in mol/L. Chemical Oxygen Demand (COD) is the amount of oxygen needed to degrade the organic matter into CO_2 and H_2O. Sub-index $k \in [1,6]$ reads for the biomass that degrades glucose, butyrate, propionate, acetate, H_2, and CO, respectively. For the gas phase, the sub-index $i \in [1,4]$ corresponds to H_2, CH_4, CO, and CO_2. The inlet flow rates of liquid and gas are represented by q_{liq}^{in} and q_{gas}^{in}, respectively, while q_{gas} denotes the outlet gas flow rate. V_{liq} and V_{gas} are the liquid and gas volumes, respectively, $S_{liq,j}^{in}$, $S_{gas,i}$, and X_k^{in} hold for the inlet concentration of the component j in the liquid phase, the inlet concentration of component i in the gas phase, and the inlet concentration of biomass k in the liquid phase. Y_k is the yield of biomass k, $f_{j,k}$ the stoichiometric coefficients; μ_k and $\mu_{k,dec}$ the growth and decay rate of biomass k, and N_i the mass transfer rate of component i.

6.2.2 Optimal Operation

All the simulations and optimizations developed for the ADM1_ME model have been run using an Intel®Core i7 8665U 2.11 GHz, 16 GB RAM computer, and MATLAB software.

The optimal operation was determined by a MODO strategy developed in previous work (Acosta-Pavas *et al.*, 2022). The MODO intended to discriminate in terms of economic performances between the production of two possible products of the biological methanation process: CH_4 and acetate by manipulating the inputs q_{gas}^{in} and q_{liq}^{in}. The two explored objectives concerned the generation of a product $Gain$ and the use of substrates $Profit\ margin$.

The Pareto optimal sets (POS) associated with the process stages were determined through Multi-Objective Optimization (MOO). Each solution is considered a Pareto Optimal Point (POP). Dynamic optimization uses a model Predictive Control (MPC) approach to optimize the performance of economic objectives with two control variables corresponding to q_{gas}^{in} and q_{liq}^{in}, *i.e.*, the dynamic optimization determines the input variables that minimize the following objective function,

$$\min_{\{q_{gas}^{in},\, q_{liq}^{in}\}} \sum_{j=t}^{t+H_p} \left(\frac{|Gain^* - Gain(t)|}{Gain^*}\right)^2 + \left(\frac{|Profit\ margin^* - Profit\ margin(t)|}{Profit\ margin^*}\right)^2 + \sum_{j=t}^{t+H_c} W_{u,1}\Delta q_{gas}^{in}(t)^2 + W_{u,2}\Delta q_{liq}^{in}(t)^2 \qquad (2.6.4)$$

$$subject\ to \begin{cases} Equations\ (2.6.1) - (2.6.3) \\ 1 \leq q_{gas}^{in} \leq 100\ L/d \\ 1.88 \leq q_{liq}^{in} \leq 10\ L/d \end{cases}$$

where $Gain^*$ and $Profit\ margin^*$ denote the POP values for $Gain$ and $Profit\ margin$ computed by the MOO, $\Delta q_{gas}^{in}(t)^2$ and $\Delta q_{liq}^{in}(t)^2$ are the differences between q_{gas}^{in} and q_{liq}^{in}, respectively, before and after each step of the dynamic optimization. H_p and H_c are the prediction and control horizons. $W_{u,1}$ and $W_{u,2}$ are the weights to balance the importance of the control effort term in the optimization.

The simulations of the biological methanation process were carried out using the ADM1_ME. It was considered a bubble column reactor (BCR) with a working volume of 37.5 L and a hydraulic retention time (HRT) of 20 days operating at 37°C for 330 days. The organic loading rate (OLR) varied over time in all stages. The reference stage corresponded to the simulation without gas addition, with a q_{liq}^{in} of 1.88 L/d. q_{liq}^{in}, q_{gas}^{in}, and the gas loading rate (GLR) were optimized by the MODO strategy for stages I–III. Table 2.6.1 summarizes the values of each variable at each stage.

Table 2.6.1. Optimal conditions used in the simulation with the ADM1 ME.

Stage	Time (days)	q_{liq}^{in} (L/d)	q_{gas}^{in} (L/d)	OLR $(gCOD/Lr/d)$	GLR $(L/Lr/d)$	$Gain$ $(EUR/Lr/d)$	$Profit\ margin$ (%)
Reference	0-30	1.88	-	0.53	-	3.11×10^{-4}	40.91
I	30-130	3.99	22.38	1.07	0.60	1.27×10^{-3}	33.77
II	130230	4.22	90.96	1.60	2.43	2.17×10^{-3}	30.59
III	230-330	4.14	36.88	2.13	0.98	2.68×10^{-3}	35.95

In the MODO strategy, the weights $W_{u,1}$ and $W_{u,2}$ were manually adjusted to values of 1×10^{-7}. H_p and H_c were considered to have equal values and were equivalent to the final time of each stage (Table 2.6.1).

6.2.3 ADM1_ME Disturbances and Dataset Generation

To train the supervised learning algorithms for fault detection in the biological methanation process, two datasets were constructed using the ADM1_ME. In the first dataset (dataset 1), disturbances of ±10% (10% HL, 10% LL), ±15% (15% HL, 15% LL), and ±20% (20% HL, 20% LL) concerning the optimal operation were performed in q_{liq}^{in}. LL and HL refer to the lower and higher liquid disturbances concerning the operational operation. Similarly, in the second dataset (dataset 2), disturbances of ±10% (10% HG, 10% LG), ±15% (15% HG, 15% LG), and ±20% (20% HG, 20% LG) concerning the optimal value were performed in q_{gas}^{in}. LG and HG refer to the lower and higher gas disturbances concerning the operational operation.

Figure 2.6.1-A shows that all disturbances in q_{liq}^{in} decreased the $Gain$ with respect to the optimal operation, especially with 20% LL disturbance. On the other hand, increases in q_{liq}^{in} significantly decreased the $Profit\ margin$ with respect to the optimal operation while

decreasing the maintained values close to the optimum. On the other side, disturbances in q_{gas}^{in}. (Figure 2.6.1-B) generate a slight variation in the $Gain$ and the $Profit\ margin$.

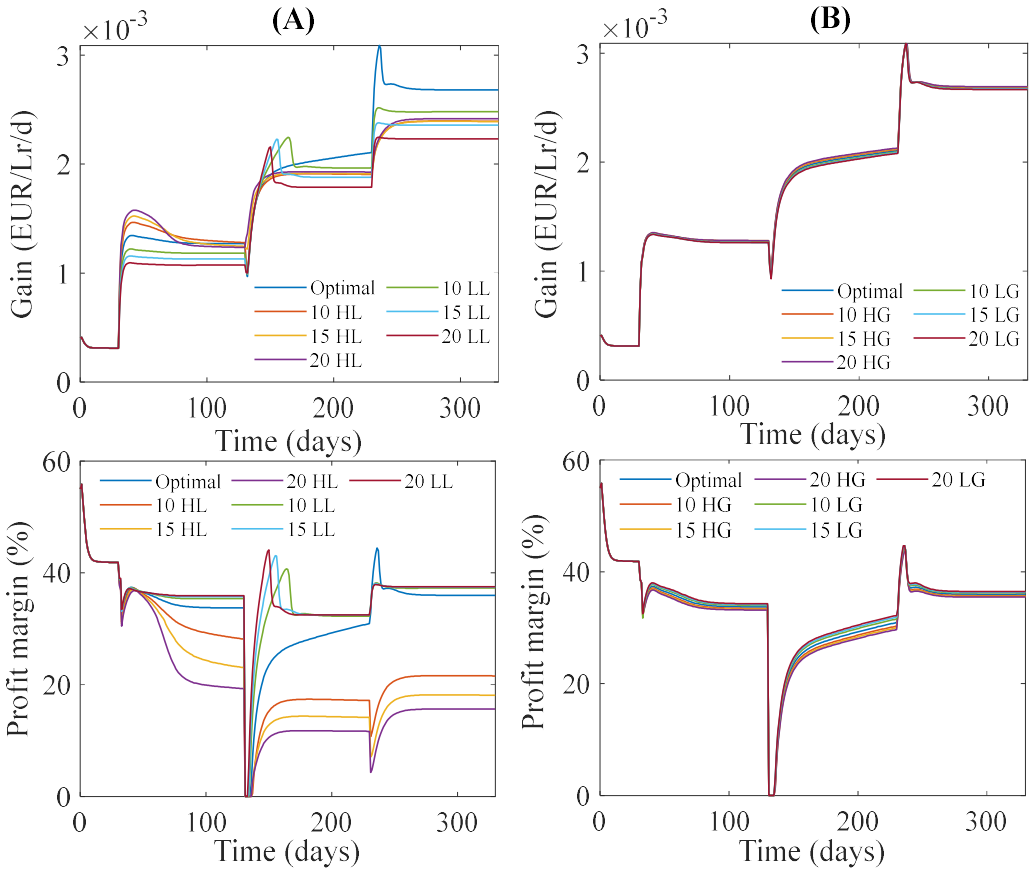


Figure 2.6.1. Effect of disturbances over the optimized economic objectives $Gain$ and $Profit\ margin$. (A) Disturbances in Dataset 1. (B) disturbances in Dataset 2.

The ADM1 ME model was developed to describe the dynamics of biological methanation, *i.e.*, concentrations, flow rates, pressures, and other variables over time. Table 2.6.2 summarizes the variables that are directly related to the inputs and outputs of the process and can be measured easily in the process. Variables such as biomass concentration were omitted due to the difficulty in their measure.

Finally, two datasets were built using the ADM1 ME simulations and the features in Table 2.6.2. Dataset 1 corresponds to the simulation of the optimal operation and the six disturbances (10% HL, 10% LL, 15% HL, 15% LL, 20% HL, and 20% LL) over q_{liq}^{in} (Dimension of dataset

1: 2317×28). Similarly, Dataset 2 corresponds to the simulation of the optimal operation and the six disturbances (10% HG, 10% LG, 15% HG, 15% LG, 20% HG, and 20% LG) over q_{liq}^{in} (Dimension of dataset 2: 2317×28).

Table 2.6.2. Variables used to train the supervised learning algorithms.

Variable		Description	Variable		Description
1	Time	Process time	**15**	$q_{gas,CH4}$	CH_4 outlet gas flow rate
2	$S_{liq,su}$	Sugar concentration in liquid phase	**16**	$q_{gas,H2}$	H_2 outlet gas flow rate
3	$S_{liq,bu}$	Butyrate concentration in liquid phase	**17**	$q_{gas,CO}$	CO outlet gas flow rate
4	$S_{liq,pro}$	Propionate concentration in liquid phase	**18**	$q_{gas,CO2}$	CO_2 outlet gas flow rate
5	$S_{liq,ac}$	Acetate concentration in liquid phase	**19**	$P_{gas,H2}$	H_2 partial pressure
6	$S_{liq,CO}$	CO concentration in liquid phase	**20**	$P_{gas,CH4}$	CH_4 partial pressure
7	$S_{liq,H2}$	H_2 concentration in liquid phase	**21**	$P_{gas,CO}$	CO partial pressure
8	$S_{liq,CH4}$	CH_4 concentration in liquid phase	**22**	$P_{gas,CO2}$	CO_2 partial pressure
9	$S_{liq,CO2}$	CO_2 concentration in liquid phase	**23**	P_{gas}	Total Pressure
10	$S_{gas,H2}$	H_2 concentration in gas phase	**24**	$p_{gas,H2}$	H_2 percent composition
11	$S_{gas,CH4}$	CH_4 concentration in gas phase	**25**	$p_{gas,CH4}$	CH_4 percent composition
12	$S_{gas,CO}$	CO concentration in gas phase	**26**	$p_{gas,CO}$	CO percent composition
13	$S_{gas,CO2}$	CO_2 concentration in gas phase	**27**	$p_{gas,CO2}$	CO_2 percent composition
14	q_{gas}	Total outlet gas flow rate	**28**	pH	pH of the system

6.3 Results and Discussion

The two datasets described in the previous section have been used to assess the performance of the machine learning algorithms to model the biological methanation process:

- Decision Tree CART
- Random Forest (RF)
- Gaussian Naïve Bayes (GNB)
- k-Nearest Neighbors (k-NN)
- Quadratic Discriminant Analysis (QDA)
- Linear SVM
- Quadratic SVM
- Cubic SVM
- Radial Basis Function (RBF) SVM
- Back Propagation Neural Network (BPNN)

All the algorithms were trained using the Scikit-Learn Python module for machine learning. The two datasets have been generated ensuring an equilibrated data proportion at each

disturbance. Therefore, a random 80/20 split was performed in both datasets, 80% (1853 ×28) of the dataset was used for training and 20% (464 × 28) for testing. A fivefold cross-validation with three repetitions was performed to guarantee the correct distribution of the instances (class-balanced). It is important to highlight that both datasets were normalized to train the SVM models.

An optimization was performed to determine the best optimizer and model hyperparameters to evaluate the best SVM between the linear, quadratic, cubic, and RBF SVM. In particular, the parameter C ranged from 15 to 25 with a step size of 1.0, the bias term in the linear SVM ranged from 0.001 to 3 with a step size of 0.5, and the gamma value ranged between the options auto and scale. The accuracy and F1-score measures have been used as model performance metrics.

Accuracy is the ratio of correct predictions across the total population size. The F_1 score is a weighted combination of precision and recall, with values between 0 and 1. A value of 1 indicates a perfect performance in terms of precision and recall, while a value of 0 indicates otherwise (Pezoulas et al., 2020).

Table 2.6.3 presents the accuracy results of seven machine learning models for disturbances in q_{liq}^{in} and q_{gas}^{in} (dataset 1 and dataset 2). Only the SVM with the best results in the hyperparameter optimization is shown. The three machine learning models with the best accuracy results were selected to study the faults in detail.

Table 2.6.3. Accuracy of training and test process with liquid and gas disturbances (best results are presented in bold).

Machine Learning model	**Accuracy dataset 1**			**Accuracy dataset 2**		
	Training	**Test**	**Computation time (seg)**	**Training**	**Test**	**Computation time (seg)**
CART	**0.92**	**0.89**	**0.38**	0.57	0.54	0.74
RF	**0.92**	**0.89**	**12.53**	**0.92**	**0.87**	**13.75**
GNB	0.29	0.27	0.11	0.14	0.14	0.13
NN	0.88	0.78	1.20	0.89	0.78	1.45
QDA	0.82	0.82	0.17	**0.91**	**0.88**	**0.19**
RBF SVM	**0.90**	**0.85**	**4.77**	**0.90**	**0.87**	**3.36**
BPNN	0.82	0.77	59.83	0.62	0.58	23.95

With regard to dataset 1, the best results have been obtained by the CART and RF models, with accuracy values of 0.92 and 0.82 for training and testing, respectively. Additionally, the RBF SVM model also provide satisfactory accuracy results with values of 0.90 and 0.85 for training and test, respectively. With regard to dataset 2, the RF, QDA, and RBF SVM models provide the best results in accuracy, with values in the training of 0.92, 0.91, and 0.90, respectively, and values in the test of 0.87, 0.88, and 0.87, respectively.

Both datasets were generated using the ADM1_ME model. The effect of each disturbance differed significantly on both objective variables (see Figure 2.6.1). Disturbances in (q_{liq}^{in}) have a higher effect on the output variables since the microorganisms are more sensitive to liquid substrate variations, which was reflected in the training of the different algorithms. Nevertheless, some of these algorithms, such as RF and RBF SVM obtained similar results with both datasets. The computation times obtained with both datasets for CART and QDA presented values less than 0.19 seconds, while RF and SVM presented values of 13.75 and 4.77 seconds. SVM models showed shorter computation time with good accuracy (Table 2.6.3), which could be associated with its capacity to work with high-dimensional feature space, small instances, and efficiency in avoiding overfitting (Yan *et al.*, 2021).

Figure 2.6.2 presents the confusion matrix for the test of the best three machine learning models. With dataset 1, some classes proved difficult to differentiate. For instance, in the decision tree CART, on average 11% of the instances in each class were classified incorrectly, and the majority of the classes were incorrectly classified as 10% HL disturbance. In the RF, on average 11% of the instances in each class were classified incorrectly, and the most conflicting classes were 10% LL and 15% HL. In RBF SVM, on average 15% of the instances in each class were classified incorrectly, and the most conflicting classes were associated with the HL disturbances, especially the 15% HL. With dataset 2, in the RF, 13% of the instances in each class were classified incorrectly, and most classes were incorrectly classified as 10% HG disturbance. In the quadratic discriminant analysis, on average 12% of the instances in each class were classified incorrectly, and the most conflicting classes were the low disturbances 10% HG and 10% LG. In the RBF SVM, on average 13% of the instances in each class were

classified incorrectly, and the most conflicting classes were the low disturbances 10% HG and 20% LG.

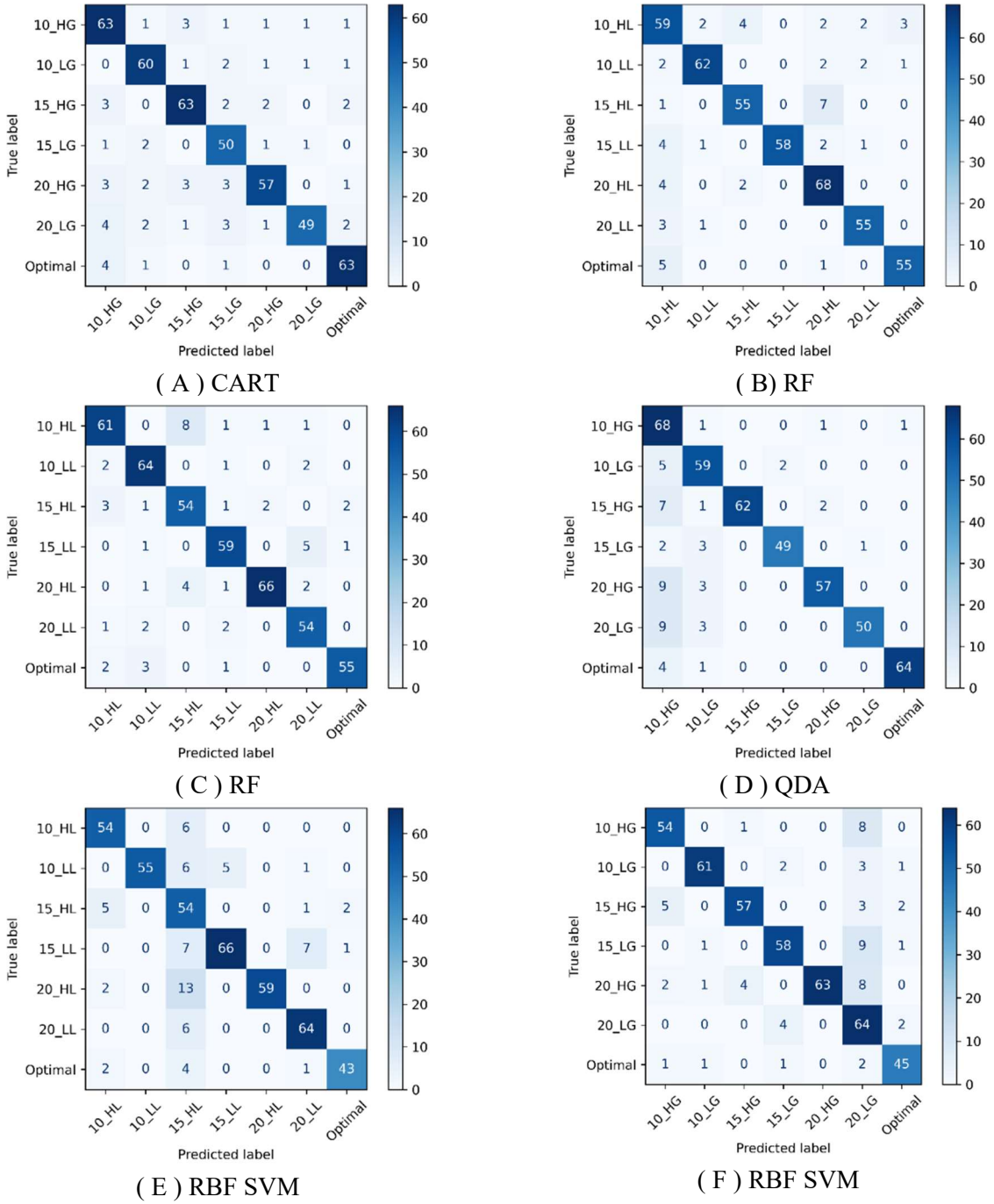


Figure 2.6.2. Confusion matrix of best machine learning models. Dataset 1: (A) CART, (C) RF, (E) RBF SVM. Dataset 2: (B) RF, (D) QDA, (F) RBF SVM.

Table 2.6.4 shows a detailed analysis based on the F1-score. For disturbances in the liquid phase, values of 0.89 were obtained in the macro average for the decision tree CART and the RF, while a value of 0.90 was obtained for RBF SVM. The three machine learning models presented similar results. However, each of them showed different F1-scores for each class, *i.e.*, CART presented the lower and highest F1-scores in 10% HL and 15% LL, with values of 0.79 and 0.94, respectively. Nevertheless, with RF, a value of 0.84 was obtained as a lower F1-score in 15% HL disturbance, while values of 0.92 were obtained as a higher F1-score in an optimal operation and 20% HL. Finally, with RBF SVM, values in the F1-score of 0.68 and 0.90 were obtained with 15% HL and 10% LL disturbances, respectively.

For disturbances in the gas phase, values of 0.87, 0.89, and 0.87 were obtained in the macro average for the decision tree CART, quadratic discriminant analysis, and RBF SVM, respectively. The machine learning models displayed similar results. However, the RF presented the lower and higher F1-scores in 10% HG and optimal operation, with values of 0.85 and 0.91, respectively. The QDA presented values of 0.78 and 0.96 in 10% HG and optimal operation, respectively. The RBF SVM showed values of 0.77 and 0.93 in 20% LG and 10% LG, respectively.

Table 2.6.4. F1 score results with both datasets.

	F1- score dataset 1				**F1- score dataset 2**		
Classes	**Decision tree CART**	**RF**	**RBF SVM**	**Classes**	**RF**	**QDA**	**RBF SVM**
10 HL	0.79	0.87	0.88	10 HG	0.85	0.78	0.86
10 LL	0.92	0.91	0.90	10 LG	0.90	0.86	0.93
15 HL	0.89	0.84	0.68	15 HG	0.88	0.93	0.88
15 LL	0.94	0.89	0.87	15 LG	0.85	0.92	0.87
20 HL	0.87	0.92	0.89	20 HG	0.86	0.88	0.89
20 LL	0.92	0.88	0.89	20 LG	0.86	0.88	0.77
Optimal	0.92	0.92	0.89	Optimal	0.91	0.96	0.89
Macro Average	0.89	0.89	0.90	**Macro Average**	0.87	0.89	0.87

6.4 Conclusions

In this paper, we have presented an approach to fault detection in the biological methanation process based on the practical assessment of machine learning models in the biological methanation process. The best results to detect disturbances in the inlet liquid flow rate concerning the optimal operation have been obtained using decision trees and RF, which achieved accuracies of 0.92 and 0.89 in the training and test, and 0.89 in F1-scores. In the case of gas inlet flow rate disturbances, the best results have been obtained using RF, QDA, and RBF SVM, which achieved accuracies of 0.92, 0.91, and 0.90% in training and test, and F1-scores of 0.87, 0.89, and 0.87, respectively.

As future work, we propose to: (i) study the effect of using more complex datasets based on the combination of disturbances in q_{liq}^{in} and q_{gas}^{in} in order to train machine learning models for fault detection when disturbances in gas and liquid flow rates occur at the same time, (ii): explore the use of homogeneous ensemble methods such as Forest, Bagging, and Voting classifiers or heterogeneous ensemble methods such as AdaBoost and Gradient Boosting Classifiers to try to improve the results obtained with the different methods presented in the current study, (iii) compare the obtained results with traditional methodologies for fault detection in bioprocesses (Principal Component Analysis) or deep learning methods.

6.5 References

Acosta-Pavas, J.C., Robles-Rodríguez, Carlos.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, 128106. https://doi.org/10.1016/j.fuel.2023.128106

Acosta-Pavas, J.C., Robles-Rodríguez, C.E., Méndez Suarez, C.A., Morchain, J., Dumas, C., Cockx, A., Aceves-Lara, C.A., 2022. Dynamic Multi-Objective Optimization Applied to Biomethanation Process. Chemical Engineering Transactions 96, 319–324. https://doi.org/10.3303/CET2296054

Cinar, S.Ö., Cinar, S., Kuchta, K., 2022. Machine Learning Algorithms for Temperature Management in the Anaerobic Digestion Process. Fermentation 8, 65. https://doi.org/10.3390/fermentation8020065

Guiot, S.R., Cimpoia, R., Carayon, G., 2011. Potential of Wastewater-Treating Anaerobic Granules for Biomethanation of Synthesis Gas. Environ. Sci. Technol. 45, 2006–2012. https://doi.org/10.1021/es102728m

Kazemi, P., Bengoa, C., Steyer, J.-P., Giralt, J., 2021. Data-driven techniques for fault detection in anaerobic digestion process. Process Safety and Environmental Protection 146, 905–915. https://doi.org/10.1016/j.psep.2020.12.016

Kazemi, P., Steyer, J.P., Bengoa, C., Font, J., Giralt, J., 2020. Robust data-driven soft sensors for online monitoring of volatile fatty acids in anaerobic digestion processes. Processes 8. https://doi.org/10.3390/pr8010067

Pezoulas, V. C., Exarchos, T. P., & Fotiadis, D. I. (2020). Machine learning and data analytics. In Medical Data Sharing, Harmonization and Analytics (pp. 227–309). Elsevier. https://doi.org/10.1016/B978-0-12-816507-2.00007-4

Rafrafi, Y., Laguillaumie, L., Dumas, C., 2020. Biological Methanation of H2 and CO2 with Mixed Cultures: Current Advances, Hurdles and Challenges. Waste and Biomass Valorization. https://doi.org/10.1007/s12649-020-01283-z

Rusmanis, D., O'Shea, R., Wall, D.M., Murphy, J.D., 2019. Biological hydrogen methanation systems – an overview of design and efficiency. Bioengineered 10, 604–634. https://doi.org/10.1080/21655979.2019.1684607

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, 122495. https://doi.org/10.1016/j.biortech.2019.122495

Yan, P., Gai, M., Wang, Y., Gao, X., 2021. Review of Soft Sensors in Anaerobic Digestion Process. Processes 9, 1434. https://doi.org/10.3390/pr9081434

Section 7 Use of Support Vector Machine to Fault Detection in Biological Methanation Process

J.C. Acosta-Pavas[a], C.E. Robles-Rodríguez[a], C. Dumas[a], A. Cockx[a], J. Morchain[a], and C.A. Aceves-Lara[a]

[a]TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

Publications:

Acosta-Pavas, J.C., Robles-Rodríguez, C.E., Dumas, C., Cockx, A., Morchain, J., Aceves-Lara, C.A. (2022). Use of Support Vector Machine to Fault Detection in Biomethanation Process. In S. Omatu, R. Mehmood, P. Sitek, S. Cicerone, S. Rodríguez (eds), Distributed Computing and Artificial Intelligence, 19th International Conference. Lecture Notes in Networks and Systems (Vol. 583, pp. 176–186). Springer, Cham. https://doi.org/10.1007/978-3-031-20859-1_18

Abstract

Biological methanation processes are complex due to the interaction of multiple molecules and different microorganisms. Hence, changes in the system inputs or operational conditions make them susceptible to faults, *e.g.*, deviation from steady states or optimal operation points. Support Vector Machine (SVM) is a relatively simple technique that can be used to identify those deviations. In this study, SVM have been applied for fault detection in the biological methanation process. Data obtained from a model-based Multi-Objective Dynamic Optimization (MODO) have been considered as the optimal operating point. Disturbances of ±10%, ±15%, and ±20% in the inlet liquid flow rate with respect to the optimal were generated by simulation with an extended Anaerobic Digestion Model No. 1 (ADM1_ME). Three SVM models, quadratic, cubic, and Radial Basis Function (RBF) were trained and validated with a dataset of 449 points (449×15) and 80 (449×15) points, respectively. The aim was to classify the regions of each disturbance and identify the percentage of the disturbance. Accuracies higher than 0.96 and 0.81 were achieved for all SVM models in the training and test, respectively.

Keywords: Biological Methanation, Multi-Objective Dynamic Optimization, Soft Sensors, Support Vector Machine, Fault Detection

7.1 Introduction

Biological methanation is a process in which the biogas produced through Anaerobic Digestion (AD) is upgraded by the biological conversion of CO_2 and syngas to obtain high-purity CH_4 (Rafrafi *et al.*, 2020). In AD, the organic matter such as lignocellulosic and amylaceous materials (agricultural residues), food residues (organic effluents from food industry), animal manure, and human excreta (waste or wastewater residues), are transformed into a mixture of methane, and carbon dioxide by a microbial complex consortium (Dar *et al.*, 2021). It is a complex process which entails four steps: (i) hydrolysis, (ii) acidogenesis, (ii) acetogenesis and (iv) methanogenesis. In the first step, fermentative bacteria excrete enzymes that transform complex organic polymers (*i.e.*, carbohydrates, proteins, and lipids) into soluble monomers, such as monosaccharides, amino acids, and long-chain fatty acids. In the second step, these monomers are converted into volatile fatty acids (VFA), such as acetate, propionate, and butyrate. In the third step, all the VFA are transformed into acetate, hydrogen, and carbon dioxide. Finally, the fourth step involves the conversion of these components into biogas, (*i.e.*, mixture of methane, carbon dioxide, and carbon monoxide). New advance technologies such as biological methanation has been proposed including the addition of gases to improve methane formation. In this paper, the used dataset was generated from a model (ADM1 _ME) (Acosta-Pavas *et al.*, 2023) considering also the input of hydrogen and carbon monoxide.

Given the multiple molecules and different microorganisms involved, small disturbances in the system inputs or operational conditions make the biological methanation process susceptible to deviations from the desired values, *e.g.*, deviation from steady states or optimal operation points, which at the industrial level imply high operational costs. Soft-Sensors were recently proposed based on Machine Learning techniques to study this type of problem in different processes. Cinar *et al.* (2022) used seven different machine learning algorithms, linear regression, logistic regression, k-nearest neighbors, decision trees, random forest, SVM, and XGBoost to define and predict the possible impacts of wide-range temperature fluctuations on process stability in the anaerobic digestion process compared to experimental data. Kazemi *et*

al. (2020) used SVM soft-Sensors to detect small magnitudes faults in VFA concentrations with pH, ammonia concentration, pressure, and CO_2 molar fraction as features.

In this paper, the goal is to use SVM models to detect deviations from an optimal region at which biological methanation can operate. For this study, the optimal conditions have been computed by Multi-objective Dynamic Optimization (MODO) and disturbances were assumed to occur from variations in the inlet liquid flow rates.

7.2 Support Vector Machine

SVM is a machine learning method for regression and classification (Kazemi *et al.*, 2020). The advantages of SVM are its simplicity, easy implementation, and the theoretical proof that it will find a hyperplane that separates the data (Kowalczyk, 2017). The use of SVM relies on the training with a given dataset (or inputs) having associated classes (or output values) for solving pattern recognition (Vapnik *et al.*, 1997). This dataset is in the form of vectors or matrices. It means that the input space X is a subset $\epsilon\, \mathcal{R}^{m\times n}$. Then, this dataset is mapped into a non-linear space Φ. By applying this mapping procedure, non-linear problems could be solved in linear space (Cristianini and Shawe-Taylor, 2000; Kazemi *et al.*, 2020), which allows to obtain a high dimensional feature space $\Phi(X)$ where the prediction of the desired vector of outputs s in $\mathcal{R}$ is possible (Bzdok *et al.*, 2018; Cervantes *et al.*, 2020).

The SVM optimization problem is to find the optimal hyperplane that separates the data correctly. Then, hyperplanes can be used to build a classifier that allows to predict class to which an observation belongs based on its features (Bzdok *et al.*, 2018). Thus, a hyperplane can be understood to divide a *p*-dimensional space into two halves (Kowalczyk, 2017).

The output of SVM can be represented as,

$$s(x) = W^T \Phi_i(x) + b \qquad (2.7.1)$$

$\Phi(x)$ is a non-linear map function that can be represented by the use of Kernel functions $K(\boldsymbol{x_i}, \boldsymbol{x_j})$, b and W are the offsets and weight vector (support vectors), respectively (Xiao *et*

al., 2022). The determination of the values of the weight vector follows an optimization problem that can be formulated as,

$$\min_{W, W_0, \xi_i} J(\boldsymbol{W}, \zeta_1, \dots, \zeta_n) = \frac{1}{2}\|\boldsymbol{W}\|^2 + C\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) \qquad (2.7.2)$$

$$Subject\ to \begin{cases} si - W^T\Phi(\boldsymbol{xi}) - W_0 \leq \varepsilon - \zeta_i & \forall i \\ W^T\Phi(\boldsymbol{xi}) + W_0 - yi \leq \varepsilon - \zeta_i & \forall i \\ \zeta_i, \quad \zeta_i^* \geq 0 & \forall i \end{cases}$$

where $\sum_{i=1}^{n}(\zeta_i + \zeta_i^*)$ is a measure of the number of misclassified data, C is a constant which measures the relative weight of the first and second terms, ε displays the loss function variable, and ζ_i, ζ_i^* are slack variables that allow certain points to fall on the incorrect side of the hyperplane.

The problem foundation of Equation (2.7.2) relies on convex quadratic programming. The Lagrangian function is utilized to integrate the constraints into the cost function, and the dual representation could be solved as in (Xiao *et al.*, 2022),

$$max\left[-\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\Phi(x_i)\Phi(x_j) + \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)y_i - \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)y_i\varepsilon\right] \qquad (2.7.3)$$

$$Subject\ to\left\{\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0.0 \quad 0.0 \leq \alpha_i, \alpha_i^* \leq C\right.$$

where α_i represent the Lagrangian multiplier.

Linear learning machines could be expressed in a dual representation, *i.e*, it can be expressed as a linear combination of the training data. Therefore, the decision rule can be evaluated using just inner products between the test point and the training points (Cervantes *et al.*, 2020; Cinar *et al.*, 2022),

$$s(x) = \sum_{i=1}^{n} \alpha_i y_i K(\boldsymbol{x_i}, \boldsymbol{x_j}) + b \quad (2.7.4)$$

with the Kernel function K, the inner product can be computed. The key to this approach is finding a Kernel function that can be evaluated efficiently. Some of the most used Kernel functions are (Cervantes *et al.*, 2020): Linear, polynomial, and RBF Kernels.

Linear Kernel :

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \boldsymbol{x_i} \cdot \boldsymbol{x_j} \quad (2.7.5)$$

Polynomic Kernel :

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i} \cdot \boldsymbol{x_j} + c)^d \quad (2.7.6)$$

Gaussian Kernel or Radial Basis Function (RBF):

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = e^{\frac{-\|x_i - x_j\|^2}{2\sigma^2}} \quad (2.7.7)$$

7.3 Biological Methanation Process Model and Multi-Objective Dynamic Control Strategy

The biological methanation process was simulated by the ADM1_ME prosed in a previous work (Acosta-Pavas *et al.*, 2023) to include gas. The operation time was 207 days with a working volume of 37.5 L at 37°C. The organic loading rate (OLR) was 0.53 $gCOD/Lr/d$ of glucose and the inlet liquid flow rate $\left(q_{liq}^{in}\right)$ was 1.9 L/d. For the first 32 days there was no gas addition, afterwards, it was carried out in 5 stages. The ADM1_ME is rewritten as,

$$\frac{dS_{gas,i}}{dt} = \frac{q_{gas}^{in}}{V_{gas}} S_{gas,i}^{in} + N_i \left(\frac{V_{liq}}{V_{gas}}\right) - \frac{q_{gas}}{V_{gas}} S_{gas,i} \quad (2.7.8)$$

$$\frac{dS_{liq,j}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(S_{liq,j}^{in} - S_{liq,j}\right) + \sum_{k} Y_k f_{j,k} \mu_k - N_i \quad (2.7.9)$$

$$\frac{dX_k}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(X_k^{in} - X_k\right) + Y_k \mu_k - \mu_{k,dec} \quad (2.7.10)$$

Sub-index $j \in [1,8]$ denotes glucose, butyrate, propionate, acetate, H_2, CH_4, CO, and CO_2 in the liquid phase. Sub-index $k \in [1,6]$ reads for the biomass that degrade glucose, butyrate, propionate, acetate, H_2, and CO, respectively. For the gas phase, the sub-index $i \in [1,4]$ corresponds to H_2, CH_4, CO, and CO_2. V_{liq} and V_{gas} are the working and gas molar fraction volume, $f_{j,k}$ are the stoichiometric coefficients, $S_{gas,i}^{in}$ and $S_{liq,j}^{in}$ are the inlet concentration of components i and j. X_k^{in} is the inlet concentration of biomass k, μ_k is the growth rate of biomass k, Y_k is the yield of biomass k, N_i is the mass transfer rate of component i, q_{liq}^{in} is the inlet liquid flow rate, q_{gas} is the outlet gas flow rate.

The ADM1_ME model has been employed in a MODO strategy prosed in previous work (Acosta-Pavas *et al.*, 2022). The objective was to find the optimal operating point, *i.e.*, the inlet gas $\left(q_{gas}^{in}\right)$ and liquid $\left(q_{liq}^{in}\right)$ flow rates that maximize optimize the biological methanation process. The MODO intended to maximize the methane yield and productivity along the process. The yield $Y_{CH4}(L/gCOD_{added})$ is the ratio between the CH_4 outlet flow rate and the total COD grams added ($gCOD_{added}$) per day, while productivity P_{CH4} $(L/Lr/d)$ is the ratio between the flow rate of the CH_4 formed and the volume of the reactor. The mathematical formulation of methane yield and productivity is reported in Equation (2.7.11) – (2.7.12).

$$Y_{CH4} = \frac{q_{gas,CH4}}{S_{liq,su}^{in}\, q_{liq}^{in} + S_{gas,H2}^{in}\, q_{gas}^{in} + S_{gas,CO}^{in}\, q_{gas}^{in}} \quad (2.7.11)$$

$$P_{CH4} = \frac{q_{gas,CH4}}{V_{liq}} \quad (2.7.12)$$

To consider both objectives and a good trade-off with the input variables, the optimization control problem was defined as,

$$\min_{\{q_{gas}^{in},q_{liq}^{in}\}}\left(\sum_{j=t}^{t+H_p}\left(\frac{|Y_{CH4}^{*}-Y_{CH4}(t)|}{Y_{CH4}^{*}}\right)^2+\left(\frac{|P_{CH4}^{*}-P_{CH4}(t)|}{P_{CH4}^{*}}\right)^2+\sum_{j=t}^{t+H_c}W_{u,1}\,\Delta q_{gas}^{in}(t)^2+W_{u,2}\,\Delta q_{liq}^{in}(t)^2\right) \quad (2.7.13)$$

$$Subject\ to\begin{cases}Equations\ (2.6.8)-(2.6.10)\\ Y_{CH4}\leq 0.39\ L/gCOD_{added}\\ 1.0\leq q_{gas}^{in}\leq 10.0\times 10^3\ L/d\\ 1.0\leq q_{liq}^{in}\leq 100\ L/d\end{cases}$$

Both objectives were normalized into a single objective function with trajectory corrections. The parameters $W_{u,1}$ and $W_{u,2}$ weights the importance of the control effort term in the optimization, $\Delta q_{gas}^{in}(t)^2$ and $\Delta q_{liq}^{in}(t)^2$ are the differences between the injected gas and inlet flow rates before and after each control step. The values of Y_{CH}^{*} and P_{CH4}^{*} are the optimal values of yield and productivity determined by the Multi-Objective Optimization (MOO) through the Pareto Optimal Set (POS) with the same definitions of Equations (2.7.11) and (2.7.12). Table 2.7.1 summarizes the MODO strategy results for maximizing the Euclidean length (Case 3 in Section 3). It reported the optimal profiles of the control variables, q_{gas}^{in} and q_{liq}^{in}, and the respective values of the variables, gas loading rate (GLR), and the OLR for each stage of the process.

Table 2.7.1. Input data in ADM1_ME for each stage obtained in MODO (Acosta-Pavas *et al.*, 2022).

Stage	Time (days)	$q_{liq}^{in}(L/d)$	OLR $(gCOD/L/d)$	$q_{gas}^{in}\times 10^3$ (L/d)	GLR $(L/Lr/d)$
I	33-64	7.00	0.53	2.73	0.07
II	65-101			3.99	0.11
III	102-135			5.26	0.14
IV	136-171			7.77	0.21
V	172-207			8.44	0.23

7.4 Methodology

7.4.1 Biological Methanation Disturbance Analysis

Disturbances of ±10%, ±15%, and ±20% in $q_{gas,in}$ and $q_{liq,in}$ with respect to the optimal value were performed, Table 2.7.2.

Table 2.7.2. q_{gas}^{in} and q_{liq}^{in} disturbances in ADM1_ME.

Liquid disturbance		Gas disturbance	
10% less liquid disturbance	10% LL	10% less gas disturbance	10% LG
15% less liquid disturbance	15% LL	15% less gas disturbance	15% LG
20% less liquid disturbance	20% LL	20% less gas disturbance	20% LG
10% higher liquid disturbance	10% HL	10% higher gas disturbance	10% HG
15% higher liquid disturbance	15% HL	15% higher gas disturbance	15% HG
20% higher liquid disturbance	20% HL	20% higher gas disturbance	20% HG

Figure 2.7.1 presents the effect of disturbances in q_{gas}^{in} on the $S_{gas,CH}$, q_{gas}, and the biomasses X_{ac}, X_{CO} and X_{H2}. For $S_{gas,CH4}$ in stage V, values of 2.27, 2.29, and 2.30 $gCOD/L$ were reached for 20% LG disturbance, optimal point, and 20% HG disturbance, respectively. Regarding q_{gas}, a value of 34 L/d was reached in stage V for disturbances of 20% LG, optimal point, and disturbance of 20% HG. The biomasses X_{ac}, X_{CO} and X_{H2} did not display any variations with respect to changes in the optimal value of q_{gas}^{in}. Figure 2.7.2 presents the effect of disturbances in q_{liq}^{in} on the $S_{gas,CH}$, q_{gas}, and the biomasses X_{ac}, X_{CO} y X_{H2}. For $S_{gas,CH4}$, no changes were observed in comparison with the optimal value except for the disturbance of 20% HL. For q_{gas} in stage V, values of 29, 34, and 37 L/d for disturbance of 20% LL, the optimal point and disturbance of 20% HL were reached. Regarding X_{ac} in stage V, values of 0.31, 0.31, and 0.30 $gCOD/L$ were reached for 20% LL disturbance, optimal point, and 20% HL disturbance, respectively. For X_{CO} in stage V, values of 0.67, 0.55, and 0.48 $gCOD/L$ for disturbance of 20% LL, optimal point, and disturbance of 20% HL were reached, respectively. For X_{H2} in stage V, values closed to 0.03 $gCOD/L$ were reached for disturbance of 20% LL, optimal point, and disturbance of 20% HL, respectively. It is concluded that the process presents a higher sensitivity to disturbances in the inlet liquid flow rate. Therefore, q_{liq}^{in} is proposed as a disturbance variable to generate data for the SVM models.

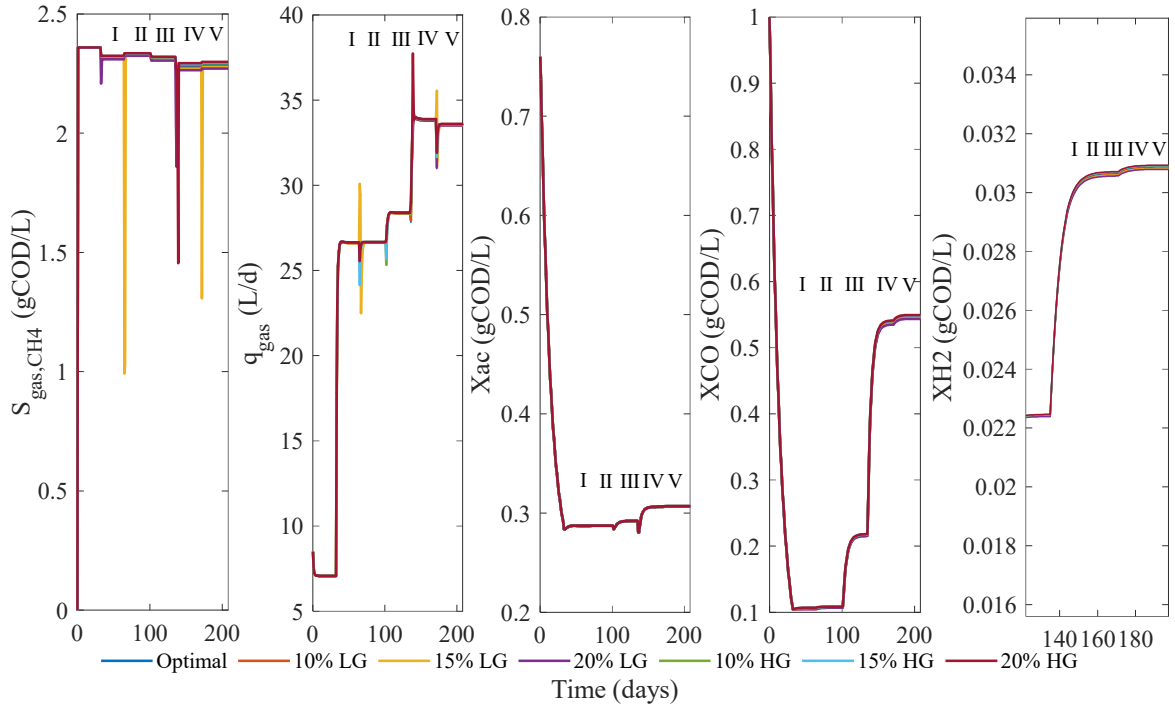


Figure 2.7.1. $S_{gas,CH}$, q_{gas}, X_{ac}, X_{CO}, and X_{H2}, with added gas flow rate disturbances.

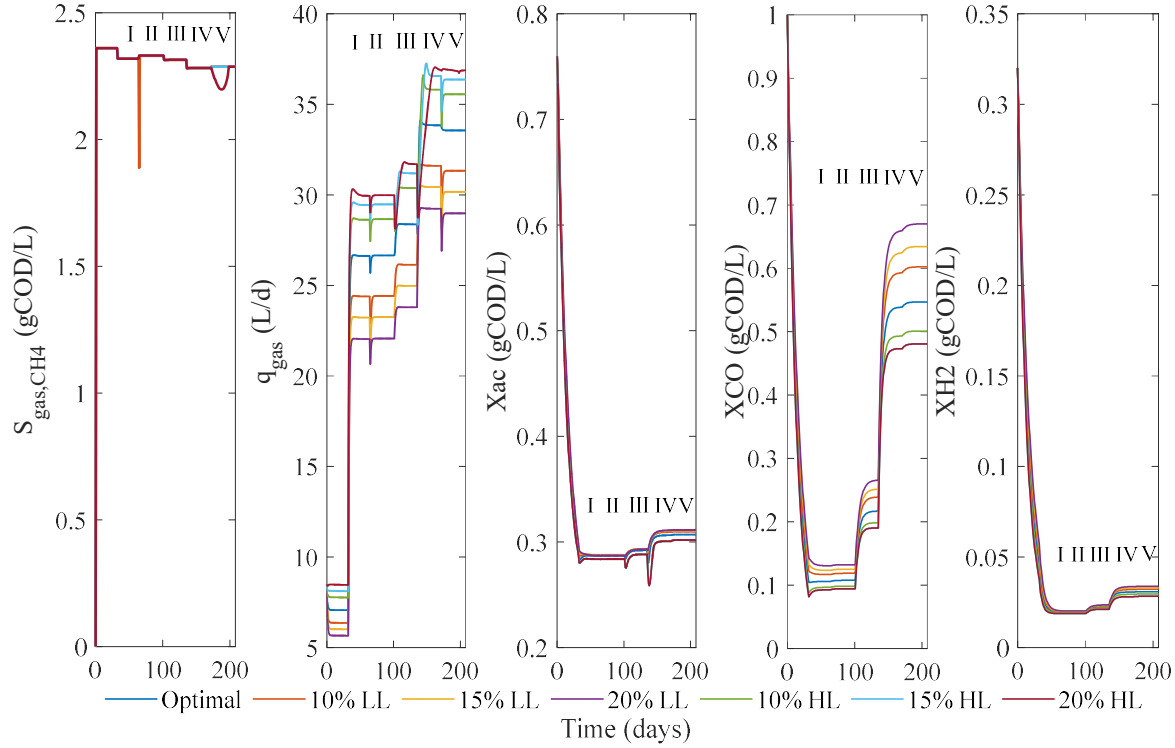


Figure 2.7.2. $S_{gas,CH}$, q_{gas}, X_{ac}, X_{CO}, and X_{H2}, with inlet liquid flow rate disturbances.

7.4.2 Training of SVM Models

Three types of SVM models were trained: quadratic, cubic, and RBF. A total of 449 points and 15 features were used as training dataset (449×15). The interest in these features is due to the direct relation with the inputs and outputs of the process, such as sugar concentration, biomasses, outlet concentrations, and gas flow rates (see Table 2.7.3). Time was also selected since each of the stages was analyzed concerning it. This dataset was obtained from the MODO strategy applied over the ADM1_ME. The points were selected between each stage change until reaching the steady state. Each one of the features was normalized to train the model as,

$$Pred_{i,N} = \frac{Pred_i - \overline{Pred_i}}{\sigma_i} \quad (2.7.14)$$

$Pred_{i,N}$ is the i normalized predictor, $\overline{Pred_i}$, and σ_i are the average and standard deviation values of i predictor. The performance evaluation for fault detection is assessed by the accuracy metric formulated as,

$$Accuracy = \frac{TP}{TP + FN} \quad (2.7.15)$$

where TP are the true positive values, and FN are the false negative values.

Table 2.7.3. Variables selected to train the SVM models.

	Variable	Description	Variable		Description
1	Time	Time	**9**	$S_{gas,H2}$	H_2 gas concentration
2	$S_{liq,su}$	Sugar concentration	**10**	$S_{gas,CH4}$	CH_4 gas concentration
3	X_{su}	Sugar biomass	**11**	$S_{gas,CO}$	CO gas concentration
4	X_{bu}	Butyrate biomass	**12**	q_{gas}	Total Outlet gas flow rate
5	X_{pro}	Propionate biomass	**13**	$q_{gas,CH4}$	CH_4 outlet gas flow rate
6	X_{ac}	Acetate biomass	**14**	$q_{gas,H2}$	H_2 outlet gas flow rate
7	X_{CO}	CO biomass	**15**	$q_{gas,CO}$	CO outlet gas flow rate
8	X_{CO}	H_2 biomass			

A dataset of 80 points was selected as the test dataset (80×15). This dataset was built using 20 points from the optimal region and 10 points from each region of the disturbance in the inlet liquid flow rates. These points came from random regions which differ from those selected for training.

7.5 Results and Discussion

7.5.1 Training and Test of SVM Models

The *fitcecoc* function from MATLAB® was used to train the SVM models. Gaussian kernels for the RBF model and Polynomial kernels with orders 2 and 3 for quadratic and cubic models were used for each training set. Finally, cross-validation was performed with 5-Folds. Accuracies of 0.96, 0.96, and 0.90 for the quadratic, cubic, and RBF models were obtained in the training of the SVM models. Afterward, the test evaluation was performed for each SVM model. Table 2.7.4 summarizes the results.

Table 2.7.4. Evaluation test of SVM models.

Disturbance	Points	Quadratic SVM	Cubic SVM	RBF SVM
Optimal	20	Optimal: 15 10% LL: 5	Optimal: 15 10% LL: 5	Optimal: 20
10% LL	10	10% LL: 10	10% LL: 5 20%LL: 5	10% LL: 5 Optimal : 5
15% LL	10	15% LL:10	15% LL: 10	15% LL: 10
20% LL	10	20% LL: 10	20% LL: 10	20% LL: 10
10% HL	10	10% HL: 10	10% HL: 10	10% HL: 10
15% HL	10	15% HL: 10	15% HL: 10	15% HL: 10
20% HL	10	20% HL: 5 10% LL: 5	20% HL: 5 20%LL: 5	20% HL: 5 Optimal: 5
Accuracy		0.88	0.81	0.88

The quadratic, cubic, and RBF SVM models obtained accuracies of 0.88, 0.81, and 0.88. For quadratic SVM, 5 of the 20 optimal points were predicted as 10% LL and 5 of the 10 points in the 20% HL disturbance were predicted as 10% LL. For the cubic SVM model, 5 of the 20 points at the optimal were predicted as 10% LL and 5 of the 10 points in the 20% HL disturbance

were predicted as 20% LL. For RBF SVM, 5 of the 10 points in the 10% LL disturbance and 5 of the 10 points in the 20% HL were predicted as optimal.

7.5.2 Fault Detection Based on SVM Models

Once the SVM models have been trained and tested, we analyzed pairs of features using the cubic SVM model to determine possible features to include or exclude. 2D visualizations were computed for each disturbance. Figure 2.7.3 presents these 2D visualizations and the training points for different pairs of features.

The goal was to find pairs of features that allow faster detection of faults in the biological methanation process. In the case of X_{ac} and X_{CO} the regions in the 2D visualization are not separable, which makes it challenging to use for predictions. In the case of X_{CO} and X_{H2}, there is a separation with minor differences between regions. For cases where $S_{liq,su}$ is used with $q_{gas,CH4}$ or X_{su}, the graphics show a remarkable separation. To highlight the implementation, let us consider the following two points from the test dataset (white points in Figure 2.7.3):

(i) Features $S_{liq,su}$, X_{su}

- Data from the test dataset: $S_{liq,su} = 1.42 \times 10^{-3}\ gCOD/L$ and $X_{su} = 0.57\ gCOD/L$
- Normalized data: (-0.53, 0.13)
- Expected prediction: The point falls within the 15 %LL region

(ii) Features $S_{liq,su}$, $q_{gas,CH4}$

- Data from the test dataset: $S_{liq,su} = 1.42 \times 10^{-3}\ gCOD/L$ and $q_{gas,CH4} = 3.02 \times 10^{-2}\ L/d$
- Normalized data: (-0.53, 0.46)
- Expected prediction: The point falls within the 15 %LL region

In these cases, with the evaluation of the 2D visualizations, it is observed that the points fall in the 15% LL region, which is consistent with the expected predictions. Nevertheless, it must be careful when performing fast evaluations, *e.g.*, in the 2D visualizations for pairs of features $S_{liq,su}$ and X_{su}, it could be assumed that the test point corresponds to a 10% LL disturbance. Nevertheless, it is clearly observed (zoom in Figure 2.7.3) that it corresponds to a 15% LL disturbance.

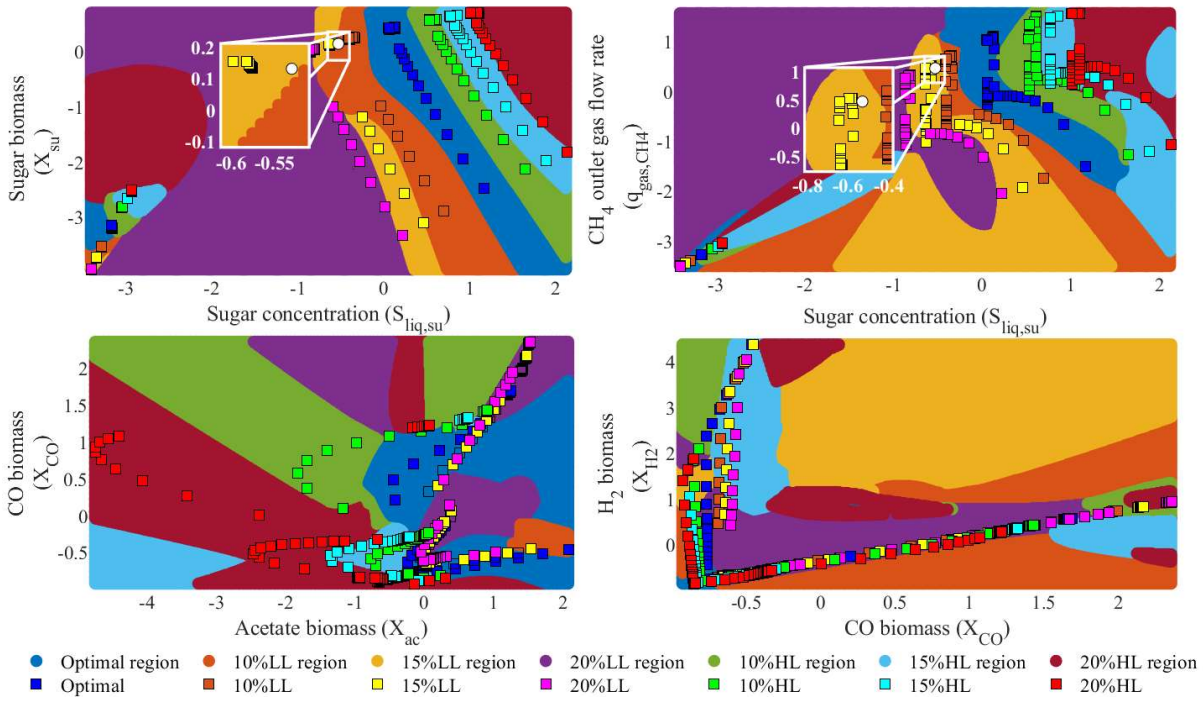


Figure 2.7.3. 2D visualization using pair of features.

It is evident that combinations of features, such as $S_{liq,su}$, $q_{gas,CH4}$, and X_{su} allow adequate separation of the optimal and disturbances regions. Therefore, those features can be considered in reducing the number of features for training cubic SVM models. Those results can be used to propose a tool for monitoring the biological methanation process and detecting deviations in $q_{liq,in}$ and $q_{gas,in}$, using features such as $S_{liq,su}$ and X_{su}, which can be measured online.

7.6 Conclusions

The detection of deviations from the optimal operation points in the biological methanation process using three SVM models was determined. The quadratic and cubic SVM achieved accuracies higher than 0.96 for the training. The RBF SVM obtained accuracies higher than 0.89. In the test evaluation, all SVM achieved accuracies higher than 0.81. A 2D visualization considering pairs of features in the cubic SVM training for fast fault detection when disturbances of q_{gas}^{in} and q_{liq}^{in} occurred showed that $S_{liq,su}$, $q_{gas,CH4}$, and X_{su} could provide an

efficient identification (fault detection) of the regions that represent a percentage of deviation from the optimal points.

7.7 References

Acosta-Pavas, J.C., Robles-Rodríguez, Carlos.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, 128106. https://doi.org/10.1016/j.fuel.2023.128106

Acosta-Pavas, J.C., Robles-Rodríguez, C.E., Méndez Suarez, C.A., Morchain, J., Dumas, C., Cockx, A., Aceves-Lara, C.A., 2022. Dynamic Multi-Objective Optimization Applied to Biomethanation Process. Chemical Engineering Transactions 96, 319–324. https://doi.org/10.3303/CET2296054

Cinar, S.Ö., Cinar, S., Kuchta, K., 2022. Machine Learning Algorithms for Temperature Management in the Anaerobic Digestion Process. Fermentation 8, 65. https://doi.org/10.3390/fermentation8020065

Guiot, S.R., Cimpoia, R., Carayon, G., 2011. Potential of Wastewater-Treating Anaerobic Granules for Biomethanation of Synthesis Gas. Environ. Sci. Technol. 45, 2006–2012. https://doi.org/10.1021/es102728m

Kazemi, P., Bengoa, C., Steyer, J.-P., Giralt, J., 2021. Data-driven techniques for fault detection in anaerobic digestion process. Process Safety and Environmental Protection 146, 905–915. https://doi.org/10.1016/j.psep.2020.12.016

Kazemi, P., Steyer, J.P., Bengoa, C., Font, J., Giralt, J., 2020. Robust data-driven soft sensors for online monitoring of volatile fatty acids in anaerobic digestion processes. Processes 8. https://doi.org/10.3390/pr8010067

Rafrafi, Y., Laguillaumie, L., Dumas, C., 2020. Biological Methanation of H2 and CO2 with Mixed Cultures: Current Advances, Hurdles and Challenges. Waste and Biomass Valorization. https://doi.org/10.1007/s12649-020-01283-z

Rusmanis, D., O'Shea, R., Wall, D.M., Murphy, J.D., 2019. Biological hydrogen methanation systems – an overview of design and efficiency. Bioengineered 10, 604–634. https://doi.org/10.1080/21655979.2019.1684607

Chapter 3 Conclusions and Perspectives

Section 1 Conclusions & Perspectives

The general objective of this thesis was to develop a model for biological methanation (biomethanation) that can be used to optimize process operation, especially for producing value-added products such as methane and acetate at the industrial scale. The approach was based on model-based control. During this work, a dynamic model was proposed to implement control strategies and develop data-driven soft sensors to detect faults.

In literature, most of the models and optimization strategies have been developed and applied over the Anaerobic Digestion (AD) process, leaving the biological methanation unexplored. Just a limited number of models were found for biological methanation (Grimalt-Alemany *et al.*, 2020; Santus *et al.*, 2022; Sun *et al.*, 2021; Tsapekos *et al.*, 2022), while model-based control has not been applied to biological methanation yet.

Biomethanation models did not consider some essential aspects, such as the transformation of CO and CO_2, the use of different substrates, and the effect of syngas addition on the gas-liquid mass transfer process.

In this context, Chapter 1 presented a literature review that addressed aspects of the modeling, optimization, and monitoring of biological processes, emphasizing biological methanation. From Section 1, it was possible to have a detailed explanation of the biological methanation, presenting some of the microorganisms involved, factors influencing its behavior, and different configurations reported in the literature. As dynamic modeling was the core of this thesis, Section 2 studied different dynamic models for AD and biological methanation. To implement the optimization of the process, Section 3 presented model-based controls as a powerful tool to maintain optimal profiles in non-linear systems, among which the MPC was highlighted. The principle of the MPC relies on tracking a reference trajectory based on optimizing a criterium (Camacho and Bordons, 2007). In our case, this criterium was based on the simultaneous maximization of yields, productivities, or economic variables. Therefore, this section also showed strategies such as Pareto Optimal Set to calculate these reference trajectories considering multiple objectives.

Another aspect that has not yet been studied is the detection of faults in the biomethanation process. Consequently, we addressed in Section 4 the use of soft sensors as fast tools for detecting faults in the biological methanation process. Specifically, we focused on presenting

data-driven soft sensors as a tool based on previously obtained data, which can complement dynamic models to represent unknown phenomena that dynamic models cannot still describe. In this Ph.D. thesis, machine learning algorithms were presented as tools for (i) estimation and monitoring of variables of interest, (ii) process classification based on the amount of any of the compounds generated, and (iii) detection of faults in bioprocesses.

The results of this thesis were presented in Chapter 2 through a series of sections.

From a simulation point of view, Section 2 of Chapter 2 provided positive answers to the following question from Chapter 1: **Can a mathematical model of biological methanation accurately reproduce multiple operational conditions with emphasis on different liquid OLR, syngas addition, and varying GLR? How can the transformation of CO into acetate and H_2 and their inhibitions be described in a model for biological methanation?**

An extension of the anaerobic digestion model (ADM1_ME) was proposed to describe the dynamics of biological methanation. The advantage of the ADM1_ME over the models proposed in the literature was the generalization of the operating conditions by adapting the volumetric mass transfer coefficient for two different reactor configurations: (i) a mesophilic bubble column reactor (BCR), using glucose, and (ii) a thermophilic continuous stirred tank reactor ($CSTR$), using primary sludge and activated ticked-disintegrated waste. The ADM1_ME was built to consider the biological transformation of CO into acetate and H_2 by carboxydrotrophic acetogens and carboxydrotrophic hydrogenogens.

The model was calibrated and validated using two experimental operating conditions from the literature, Operational Condition 1 (OP1) and Operational Condition 2 (OP2), developed with a varying OLR and GLR. According to the statistical evaluation, a coefficient of determination $R^2 > 0.74$ and a Root Mean Squared Error $RMSE < 5.15$ were obtained in the model validation with both operational conditions, which allowed us to highlight the feasibility of the ADM1_ME to describe the biological methanation process at different operational conditions and reactor configurations.

Several variables can be optimized in biological processes, yields, productivities, process times, *etc*. Most of these variables are often conflicting. Multi-objective Optimization (MOO), which

is based on multiple-criteria decision-making, involves more than one objective to be optimized simultaneously (Chang, 2015).

Different model-based optimization techniques have been used to analyze and improve the biological process. However, this kind of optimization begins with experimentation that enables an understanding of biological processes. Unfortunately, it implies that the optimization of biological processes are time-consuming and economically expensive. Modeling and simulation are a good manner to optimize these biological processes without the need to develop an experimental setup (Mitsos *et al.*, 2018).

Therefore, a Multi-Objective Dynamic Optimization (MODO) was built to develop the optimization of the biological methanation process, where some questions of Section 3 were answered:

- **Can the multi-objective optimization approaches improve biological methanation?** This was addressed for the first time with a MODO strategy to maximize simultaneously two objective functions, yields and productivities. It was computed using a Multi-Objective Optimization (MOO) based on Pareto Optimal Set. MODO strategy use as control variables the inlet liquid $\left(q_{liq}^{in}\right)$ and gas $\left(q_{gas}^{in}\right)$ flow rates.
- **How to implement a computationally feasible model-based control strategy for biological methanation?** The MOO was couplished with a model predictive control (MPC) schema using the dynamic model ADM1_ME.

In Section 3 was computed the maximization of yield (Y_{CH}) and productivity (P_{CH}) of methane. Section 4 showed the maximization of (Y_{CH4}, P_{CH4}) and it was complemented by switching between the maximization of yields (Y_{ac}) and productivities (P_{ac}) of acetate. Then, to answer the last part of the question: **Could the multi-objective optimizations consider several objectives, such as the yields, the productivities, and other variables in economic terms (*e.g.*, substrates prices)?** Section 5 accomplished the Economic Multi-Objective Dynamic Optimization (EMODO) for the maximization of economic variables, $(Gain)$ and $(Profit\ margin)$.

The results showed the feasibility of the MODO strategy to switch between products of interest using two control variables, q_{liq}^{in} and q_{gas}^{in}. It is important to note that these are simulation results and that the adaptation of the microorganisms may vary, especially in the switching between products such as methane and acetate. However, this work generated advances in the biological methanation process optimization through multi-objective optimizations and model-based control tools.

Questions of Section 4 of Chapter 1 were addressed: **Can machine learning be used as a data-driven soft sensor in biological methanation?** And **can these soft sensors be used to detect faults during the process?** An approach to detect faults in the biological methanation process was studied in sections 6 and 7 to answer these questions.

The development of several machine learning models led to the detection and classification of deviations from the optimal operation points when disturbances occurred in the q_{liq}^{in} and q_{gas}^{in}. In Section 6, Machine learning algorithms, such as decision trees, random forest, and radial basis function Support Vector Machine (SVM), obtained the best statistical metrics accuracy and F1-score results. Then in Section 7, SVM were applied to construct 2D visualization formed by training pair of features. SVM were used due to: the good results that could be obtained, the simplicity of implementation, and the efficiency in avoiding overfitting.

It is important to note that this is the first study to use machine learning soft sensors in fault detection on the biological methanation process. Although it is an area that has been scarcely studied, this work generated insights towards the easy and fast application of machine learning soft sensors on biological processes, mainly in the industrial application of 2D visualization considering pairs of features (variables that can be measured on-line), which provides an easy and fast reading of the deviations of certain variables in the process.

The results of this thesis lead us to think about some perspectives:

An interesting approach that could be explored is coupling metabolic and dynamic models to obtain deep insights into the biological methanation process, especially in the methanogenesis step where the syngas is added. The metabolic modeling uses constraint-based methods, which implement cellular limitations on biological networks such as physicochemical, genotypical,

environmental, thermodynamic, *etc*. (Perez-Garcia *et al.*, 2016). Here, the challenge is the metabolic construction of some populations that could be used to determine the biochemical properties of key components and analyze methanogenic metabolism (Feist *et al.,* 2006). This task could be addressed using hybrid-cybernetic models, which integrate intracellular kinetics with a description of metabolic regulation (Robles-Rodriguez *et al.*, 2017; Shuler and Varner, 2011). On this basis, a more robust study could be performed in the biological methanation process, analyzing the behavior of the different microorganisms in implementing the MODO strategy.

The MODO strategy could be used to consider energetic issues to improve biological methanation, *e.g.*, the minimization of the power input required for mixing a $CSTR$ and the simultaneous maximization of the performance or the maximization of the volumetric mass transfer coefficient. Therefore, we can propose energy balances and then design multi-objective optimizations to maximize the volumetric mass transfer coefficient and minimize the power consumption in a bioreactor (Krasławski *et al.*, 1991).

Concerning the machine learning soft sensors, one of those perspectives is to evaluate more complex datasets that integrate diverse combinations of disturbances into the training of machine learning algorithms to detect and classify faults in the biological methanation process. In this case, homogeneous ensemble methods such as AdaBoost and Gradient Boosting Classifiers could be explored to improve the current results.

It is possible to assess the coupling of regression and classification machine learning models to improve the performance of the biological methanation process. The regression machine learning models could be employed as soft sensors for monitoring variables that are difficult to measure, such as biomass or substrates. This could be complemented by selecting the most important features to train the models. On the other hand, classification machine learning models could be proposed as previously presented in this thesis, in the detection of faults in the process but considering simultaneous disturbances in the manipulated variables, or could be used to classify the performance of the process in low, medium, and high, according to the methane content in the outlet gas flow rate (Cinar *et al.*, 2022; Wang *et al.*, 2020).

Section 2 Towards Digital Twins Perspectives

Through this thesis, the aspects of dynamic modelisation, optimization, and faults detections are adressed. This is how one of the perspectives of this work arises, the exploration of a way to merge all the individual tools into a global and uniform system (Neubauer *et al.*, 2020) to be applied to biological methanation. The first approach is the formulation of digital twins.

Digital twins can be defined in many ways (Glaessgen and Stargel, 2012; Stark and Damerau, 2019; Jiang *et al.*, 2021). In our context, "a virtual representation of a physical system (and its associated environment and processes) that is updated through the exchange of information between the physical and virtual systems" (VanDerHorn and Mahadevan, 2021). Thelen *et al.* (2022, 2023) proposed a five-dimensional digital twin model as $Digital\ twin = \mathbb{F}(PS, DS, P2V, V2P, OPT)$ where PS and DS consisted in the physical and digital system, $P2V$ and $V2P$ referred to the updating and prediction engine, and OPT held for the optimization dimension. Figure 3.2.1 summarizes the five-dimensional digital twin adapted to this work.

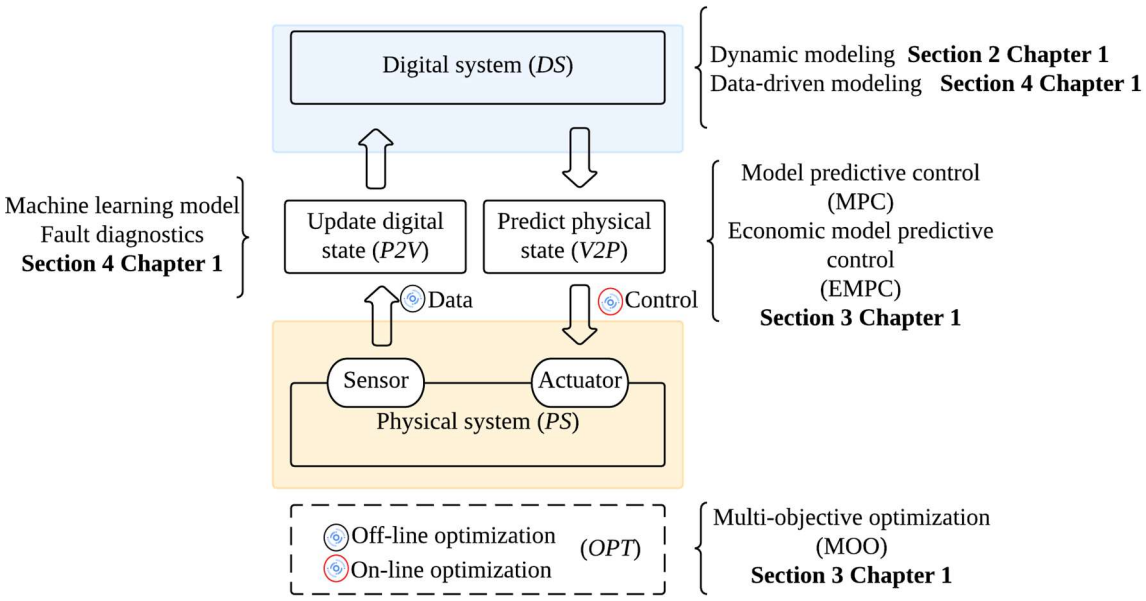


Figure 3.2.1. Five-dimensional digital twin. Adapted from (Juarez *et al.*, 2021; Thelen *et al.*, 2022).

In the context of biological processes, especially in biological methanation, PS can be related to the systems associated with online data acquisition during the experimental process,

generating a data set. $P2V$ updates the state of the Digital Twin based on the data obtained. In this part, process faults can be included. DS refers to the model used to represent the obtained data set. In our case, it can be represented with the mathematical models: dynamic models or soft sensors. $V2P$ is in charge of the prediction and optimization of the optimal trajectories of the process. Then, this information will be sent to the system in real-time. Here, the system is designed to optimize multiple variables, *e.g.*, yields and productivities of the process, using strategies such as MOO or EMPC.

At this moment there are no reports of digital twins applied to biological methanation. However, there are some researches on AD. Moretta *et al.* (2021) studied the AD process by improving the ADM1 model, called Anaerobic Digestion Enhancement (ADE), in which the authors considered a digital twin. The objective was to improve the ADM1 model, *i.e.*, the CH_4 content by increasing its application at the industrial level. The authors considered that the ADE could process different feedstocks (*i.e.*, animal manure, silage, sludge) through co-digestion to produce biogas. However, to have a more realistic representation of the process, the authors included the production of hydrogen sulfide in the ADE, which can be found in significant concentrations in these feedstocks.

Additionally, the authors included the kinetic reactions and mass balances equations to produce hydrogen sulphide from the sulphur by Sulphur-Oxidizing-Bacteria. They optimized the kinetic parameters for different configurations and feedstocks using as measured variables the biogas flow rate. Finally, the results obtained from the simulation were used to optimize a plant, and the plant simulation results were adjusted to typical operating ranges, demonstrating the reliability and flexibility of using the ADE. The same authors, Moretta *et al.* (2022), developed the conceptual design of a digital twin for producing bio-methanol as a value-added product of the AD process using microalgae. This conceptual design used the ADE model for biogas production and the subsequent implementation of a biorefinery to convert biogas into bioethanol. Finally, the authors performed an economic analysis to investigate the feasibility of producing methanol from biogas produced using microalgae. Although the results are interesting, other aspects must be considered and integrated in real-time to consider a functional digital twin (see Figure 3.2.1).

2.1 References

Camacho, E.F., Bordons, C., 2007. Model Predictive control. Springer London. https://doi.org/10.1007/978-0-85729-398-5

Chang, K.-H., 2015. Multiobjective Optimization and Advanced Topics, in: Design Theory and Methods Using CAD/CAE. Elsevier, pp. 325–406. https://doi.org/10.1016/B978-0-12-398512-5.00005-0

Cinar, S.Ö., Cinar, S., Kuchta, K., 2022. Machine Learning Algorithms for Temperature Management in the Anaerobic Digestion Process. Fermentation 8, 65. https://doi.org/10.3390/fermentation8020065

Feist, A.M., Scholten, J.C.M., Palsson, B.Ø., Brockman, F.J., Ideker, T., 2006. Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barker. Molecular Systems Biology 2, 2006.0004. https://doi.org/10.1038/msb4100046

Glaessgen, E., Stargel, D., 2012. The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles, https://doi.org/10.2514/6.2012-1818

Grimalt-Alemany, A., Asimakopoulos, K., Skiadas, I.V., Gavala, H.N., 2020. Modeling of syngas biomethanation and catabolic route control in mesophilic and thermophilic mixed microbial consortia. Applied Energy 262, 114502. https://doi.org/10.1016/j.apenergy.2020.114502

Jiang, Y., Yin, S., Li, K., Luo, H., Kaynak, O., 2021. Industrial applications of digital twins. Phil. Trans. R. Soc. A. 379, 20200360. https://doi.org/10.1098/rsta.2020.0360

Juarez, M.G., Botti, V.J., Giret, A.S., 2021. Digital Twins: Review and Challenges. Journal of Computing and Information Science in Engineering 21, 030802. https://doi.org/10.1115/1.4050244

Krasławski, A., Rzyski, E., Stelmach, J., 1991. Application of fuzzy multiobjective optimization for self-sucking impellers in a bioreactor. Bioprocess Engineering 6, 109–116. https://doi.org/10.1007/BF00369063

Mitsos, A., Asprion, N., Floudas, C.A., Bortz, M., Baldea, M., Bonvin, D., Caspari, A., Schäfer, P., 2018. Challenges in process optimization for new feedstocks and energy sources. Computers & Chemical Engineering 113, 209–221. https://doi.org/10.1016/j.compchemeng.2018.03.013

Moretta, F., Rizzo, E., Manenti, F., Bozzano, G., 2021. Enhancement of anaerobic digestion digital twin through aerobic simulation and kinetic optimization for co-digestion scenarios. Bioresource Technology 341, 125845. https://doi.org/10.1016/j.biortech.2021.125845

Moretta Federico, Fedeli Matteo, Manenti Flavio, Bozzano Giulia Luisa, 2022. Conceptual Design of Digital Twin for Bio-methanol Production from Microalgae. Chemical Engineering Transactions 92, 253–258. https://doi.org/10.3303/CET2292043

Neubauer, P., Anane, E., Junne, S., Cruz Bournazou, M.N., 2020. Potential of Integrating Model-Based Design of Experiments Approaches and Process Analytical Technologies for Bioprocess Scale-Down, in: Herwig, C., Pörtner, R., Möller, J. (Eds.), Digital Twins, Advances in Biochemical

Engineering/Biotechnology. Springer International Publishing, Cham, pp. 1–28. https://doi.org/10.1007/10_2020_154

Perez-Garcia, O., Lear, G., Singhal, N., 2016. Metabolic Network Modeling of Microbial Interactions in Natural and Engineered Environmental Systems. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.00673

Robles-Rodriguez, C.E., Bideaux, C., Guillouet, S.E., Gorret, N., Cescut, J., Uribelarrea, J.-L., Molina-Jouve, C., Roux, G., Aceves-Lara, C.A., 2017. Dynamic metabolic modeling of lipid accumulation and citric acid production by Yarrowia lipolytica. Computers & Chemical Engineering 100, 139–152. https://doi.org/10.1016/j.compchemeng.2017.02.013

Santus, A., Corbellini, V., Trionfini, M., Malpei, F., Ferretti, G., 2022. Modelling and Parameter Identification of Ex-Situ Biological Biogas Upgrading. IFAC-PapersOnLine 55, 624–629. https://doi.org/10.1016/j.ifacol.2022.09.165

Shuler, M.L., Varner, J.D., 2011. Cell Growth Dynamics, in: Comprehensive Biotechnology. Elsevier, pp. 25–32. https://doi.org/10.1016/B978-0-08-088504-9.00021-0

Stark, R., Damerau, T., 2019. Digital Twin, in: The International Academy for Production Engineering, Chatti, S., Tolio, T. (Eds.), CIRP Encyclopedia of Production Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–8. https://doi.org/10.1007/978-3-642-35950-7_16870-1

Sun, H., Yang, Z., Zhao, Q., Kurbonova, M., Zhang, R., Liu, G., Wang, W., 2021. Modification and extension of anaerobic digestion model No.1 (ADM1) for syngas biomethanation simulation: From lab-scale to pilot-scale. Chemical Engineering Journal 403, 126177. https://doi.org/10.1016/j.cej.2020.126177

Thelen, A., Zhang, X., Fink, O., Lu, Y., Ghosh, S., Youn, B.D., Todd, M.D., Mahadevan, S., Hu, C., Hu, Z., 2023. A comprehensive review of digital twin—part 2: roles of uncertainty quantification and optimization, a battery digital twin, and perspectives. Struct Multidisc Optim 66, 1. https://doi.org/10.1007/s00158-022-03410-x

Thelen, A., Zhang, X., Fink, O., Lu, Y., Ghosh, S., Youn, B.D., Todd, M.D., Mahadevan, S., Hu, C., Hu, Z., 2022. A comprehensive review of digital twin — part 1: modeling and twinning enabling technologies. Struct Multidisc Optim 65, 354. https://doi.org/10.1007/s00158-022-03425-4

Tsapekos, P., Alvarado-Morales, M., Angelidaki, I., 2022. H2 competition between homoacetogenic bacteria and methanogenic archaea during biomethanation from a combined experimental-modelling approach. Journal of Environmental Chemical Engineering 10, 107281. https://doi.org/10.1016/j.jece.2022.107281

VanDerHorn, E., Mahadevan, S., 2021. Digital Twin: Generalization, characterization and implementation. Decision Support Systems 145, 113524. https://doi.org/10.1016/j.dss.2021.113524

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, 122495. https://doi.org/10.1016/j.biortech.2019.122495

Annexes

Section 1 Biological Methanation Model Development

1.1 Chemical Oxygen Demand (COD) calculation

Consider the COD calculation for CH_4, according to the balanced Equation (4.1.1):

$$\begin{array}{ccccc} CH_4 & + & 2O_2 \rightarrow & CO_2 + & 2H_2O \\ 1mol\ CH_4 & & 64gO_2 & & \end{array} \quad (4.1.1)$$

To degrade one mole of CH_4 to CO_2, and H_2O, $64g$ of oxygen are needed; in other words, $64\frac{gCOD}{mol}$ (Henze *et al.*, 2019):

$$\frac{2\ mol\ O_2}{1mol\ CH_4}\frac{32\ g\ O_2}{1\ mol\ O_2} = \frac{64gO_2}{1mol\ CH_4} = 64\frac{gCOD}{mol}$$

1.2 Anaerobic Digestion Model Extension (ADM1_ME)

The ADM1_Me model is a derivative of the ADM1 model, which is a well-established framework for simulating anaerobic digestion. This model is based on a series of bioreactions that are responsible for breaking down complex organic matter into methane. In order to enhance the performance of the ADM1 model, the ADM1_Me model introduces the use of H_2 and CO in the gaseous phase as substrates for biomethanation. This modification alters the traditional model to take on the form of $dS/dt = inputs - outputs + r$ with $r = M\mu_k$, where M represents the stoichiometric matrix and μ_k represents the reaction rate of each associated microorganism.

1.2.1 Mass balances

State variables in the liquid phase $(S_{liq,j})$:

$$\frac{dS_{liq,su}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(S_{liq,su}^{in} - S_{liq,su}\right) - \mu_{su} \quad (4.1.2)$$

$$\frac{dS_{liq,bu}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(S_{liq,bu}^{in} - S_{liq,bu}\right) + (1 - Y_{su})f_{bu,su}\mu_{su} - \mu_{bu} \quad (4.1.3)$$

$$\frac{dS_{liq,pro}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(S_{liq,pro}^{in} - S_{liq,pro}\right) + (1 - Y_{su})f_{pro,su}\mu_{su} - \mu_{pro} \quad (4.1.4)$$

$$\frac{dS_{liq,ac}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(S_{liq,ac}^{in} - S_{liq,ac}\right) + (1 - Y_{su})f_{ac,su}\mu_{su} + (1 - Y_{bu})f_{ac,bu}\mu_{bu} + \left(1 - Y_{pro}\right)f_{ac,pro}\mu_{pro} - \mu_{ac} + (1 - Y_{CO})f_{ac,co}\mu_{CO} \quad (4.1.5)$$

$$\frac{dS_{liq,co}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(S_{liq,CO}^{in} - S_{liq,CO}\right) - \mu_{CO} - N_{co} \quad (4.1.6)$$

$$\frac{dS_{liq,H2}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(S_{liq,H2}^{in} - S_{liq,H2}\right) + (1 - Y_{su})f_{H2,su}\mu_{su} + (1 - Y_{bu})f_{H2,bu}\mu_{bu} + \left(1 - Y_{pro}\right)f_{H2,pro}\mu_{pro} + (1 - Y_{co})f_{H2,co}\mu_{CO} - \mu_{H2} - N_{H2} \quad (4.1.7)$$

$$\frac{dS_{liq,CH4}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(S_{liq,CH4}^{in} - S_{liq,CH4}\right) + (1 - Y_{ac})\mu_{ac} + (1 - Y_{H2})\mu_{H2} - N_{CH4} \quad (4.1.8)$$

$$\frac{dS_{liq,co2}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}\left(S_{liq,CO2}^{in} - S_{liq,CO2}\right) - \sum_{k=1}^{6}\left(\sum_{j=1}^{7} C_j v_{j,k}\mu_k\right) - N_{co2} \quad (4.1.9)$$

The term $\left(-\sum_{k=1}^{6}\left(\sum_{j=1}^{7} C_j v_{j,k}\mu_k\right)\right)$ refers to the carbon balance from compound j on uptake of k, which is derived for the formation of inorganic carbon, *i.e.*, the stoichiometric balance of inorganic carbon in each k processes (Batstone *et al*., 2002). In this definition C_j is the carbon content of component j, and $v_{j,k}$ the rates coefficients for component j on uptake of k.

As an example, considers the glucose uptake derived from inorganic carbon:

$$s_1 = -C_{su} + (1 - Y_{su})\left(f_{bu,su}C_{bu} + f_{pro,su}C_{pro} + f_{ac,su}C_{ac}\right) + Y_{su}C_{bac}$$

This is equivalent to the difference between the total available from sugar (C_{su}), which is derived from the compounds butyrate $(1 - Y_{su})\left(f_{bu,su}C_{bu}\right)$, propionate $(1 - Y_{su})\left(f_{pro,su}C_{pro}\right)$ and acetate $(1 - Y_{su})\left(f_{ac,su}C_{ac}\right)$ and the formation of biomass $(Y_{su}C_{bac})$, resulting in the theoretical amount of residual carbon that will be derived to inorganic carbon.

Following the idea, the terms s_2- s_5 represent the resulting amount of carbon derived to CO_2 formation from sugar, butyrate, propionate, acetate, and H_2, while s_6 represents the carbon derived to CO_2 formation from all biomass degradation.

$$s_2 = -C_{bu} + (1 - Y_{bu})(f_{ac,bu}C_{ac}) + Y_{ac}C_{bac}$$

$$s_3 = -C_{pro} + (1 - Y_{pro})(f_{ac,pro}C_{ac}) + Y_{pro}C_{bac}$$

$$s_4 = -C_{ac} + (1 - Y_{ac})C_{CH4} + Y_{ac}C_{bac}$$

$$s_5 = (1 - Y_{H2})C_{CH4} + Y_{H2}C_{bac}$$

$$s_6 = -C_{bac}$$

State variables in biomass (X_k):

$$\frac{dX_{su}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}(X_{su}^{in} - X_{su}) + Y_{su}\mu_{su} - \mu_{su,dec} \quad (4.1.10)$$

$$\frac{dX_{bu}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}(X_{bu}^{in} - X_{bu}) + Y_{bu}\mu_{bu} - \mu_{bu,dec} \quad (4.1.11)$$

$$\frac{dX_{pro}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}(X_{pro}^{in} - X_{pro}) + Y_{pro}\mu_{pro} - \mu_{pro,dec} \quad (4.1.12)$$

$$\frac{dX_{ac}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}(X_{ac}^{in} - X_{ac}) + Y_{ac}\mu_{ac} - \mu_{ac,dec} \quad (4.1.13)$$

$$\frac{dX_{co}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}(X_{CO}^{in} - X_{co}) + Y_{co}\mu_{co} - \mu_{CO,dec} \quad (4.1.14)$$

$$\frac{dX_{H2}}{dt} = \frac{q_{liq}^{in}}{V_{liq}}(X_{H2}^{in} - X_{H2}) + Y_{H2}\mu_{H2} - \mu_{H2,dec} \quad (4.1.15)$$

State variables in gas phase $(S_{gas,i})$:

$$\frac{dS_{gas,H2}}{dt} = \frac{q_{gas}^{in}}{V_{gas}} S_{gas,H2}^{in} + \frac{V_{liq}}{V_{gas}} N_{H2} - \frac{q_{gas}}{V_{gas}} S_{gas,H2} \quad (4.1.16)$$

$$\frac{dS_{gas,CH4}}{dt} = \frac{q_{gas}^{in}}{V_{gas}} S_{gas,CH4}^{in} + \frac{V_{liq}}{V_{gas}} N_{CH4} - \frac{q_{gas}}{V_{gas}} S_{gas,CH4} \quad (4.1.17)$$

$$\frac{dS_{gas,CO}}{dt} = \frac{q_{gas}^{in}}{V_{gas}} S_{gas,CO}^{in} + \frac{V_{liq}}{V_{gas}} N_{co} - \frac{q_{gas}}{V_{gas}} S_{gas,co} \quad (4.1.18)$$

$$\frac{dS_{gas,CO2}}{dt} = \frac{q_{gas}^{in}}{V_{gas}} S_{gas,CO2}^{in} + \frac{V_{liq}}{V_{gas}} N_{co2} - \frac{q_{gas}}{V_{gas}} S_{gas,co2} \quad (4.1.19)$$

1.2.2 Constitutive equations

Biochemical reactions (μ_k):

$$\mu_{su} = \frac{\mu_{m,su}\, S_{liq,su}}{Ks_{su} + S_{liq,su}} X_{su} \quad (4.1.20)$$

$$\mu_{bu} = \frac{\mu_{m,bu}\, S_{liq,bu}}{Ks_{bu} + S_{liq,bu}} X_{bu} I_{H2,bu} I_{CO,H2} \quad (4.1.21)$$

$$\mu_{pro} = \frac{\mu_{m,pro}\, S_{liq,pro}}{Ks_{pro} + S_{liq,pro}} X_{pro} I_{H2,pro} I_{CO,H2} \quad (4.1.22)$$

$$\mu_{ac} = \frac{\mu_{m,ac}\, S_{liq,ac}}{Ks_{ac} + S_{liq,ac}} X_{ac} I_{H2,ac} I_{CO,ac} \quad (4.1.23)$$

$$\mu_{H2} = \frac{\mu_{m,H2}\, S_{liq,H2}}{Ks_{H2} + S_{liq,H2}} X_{H2} I_{CO,H2} \quad (4.1.24)$$

$$\mu_{CO} = \frac{\mu_{m,CO}\, S_{liq,CO}}{Ks_{CO} + S_{liq,CO}} X_{CO} \quad (4.1.25)$$

Decay biomass ($\mu_{k,dec}$) :

$$\mu_{su,\,dec} = K_{su,\,dec}\, X_{su} \quad (4.1.26)$$

$$\mu_{bu,\,dec} = K_{bu,\,dec}\, X_{bu} \quad (4.1.27)$$

$$\mu_{pro,\,dec} = K_{pro,\,dec}\, X_{pro} \quad (4.1.28)$$

$$\mu_{ac,\,dec} = K_{ac,\,dec}\, X_{ac} \quad (4.1.29)$$

$$\mu_{CO,\,dec} = K_{CO,\,dec}\, X_{CO} \quad (4.1.30)$$

$$\mu_{H2,\,dec} = K_{H2,\,dec}\, X_{H2} \quad (4.1.31)$$

1.2.3 Mass transfer rates

$$N_{H2} = k_L a_{,H2} \left(S_{liq,H2} - \gamma_{COD,\,H2}\, H_{H2}\, P_{gas,H2}\right) \quad (4.1.32)$$

$$N_{CH4} = k_L a_{,CH4} \left(S_{liq,CH4} - \gamma_{COD,\,CH4}\, H_{CH4}\, P_{gas,CH4}\right) \quad (4.1.33)$$

$$N_{CO} = k_L a_{,CO} \left(S_{liq,CO} - \gamma_{COD,\,CO}\, H_{CO}\, P_{gas,CO}\right) \quad (4.1.34)$$

$$N_{CO2} = k_L a_{,CO2} \left(S_{liq,CO2} - H_{CO2}\, P_{gas,CO2}\right) \quad (4.1.35)$$

where N_i is the flux of species H_2, CH_4, CO and CO_2 expressed as COD (Chemical oxygen demand: amount of oxygen needed to degrade the organic matter into CO_2 and H_2O). $k_L a_{,i}$ is the volumetric mass transfer coefficient of component i, and $\left(S_{liq,j} - \gamma_{COD,\,i}\, H_i\, P_{gas,i}\right)$ is the driving force. H_i, and $P_{gas,i}$ are the, Henry's law equilibrium constant, and partial pressure of component i, respectively. $\gamma_{COD,\,i}$ is the COD, *e.g.*, 16 $gCOD/mol$ for H_2 and CO, and 64 $gCOD/mol$ for CH_4, it permits the conversion between the $moles$ and $gCOD$ of a component i. It is important to mention that the CO_2 do not present a COD, therefore, through this paper it will be expressed in mol instead of COD, as suggested in (Batstone *et al.*, 2002).

Section 2 Sensitivity analysis

A total of 60 parameters were considered in the ADM1_ME, 22 stoichiometric, 23 biochemical, and 15 physiochemical parameters (Table 4.2.1). In which 26 parameters were considered in the SA. All stoichiometric coefficients $f_{j,k}$ and C_j were fixed for the sensitivity analysis, except $f_{ac,\mathrm{CO}}$ which was an unknown parameter. All Y_k were used in the SA. Concerning biochemical parameters, the $K_{k,\,dec}$ and all $\mu_{m,k}$, Ks_k, $KI_{H2,\,k}$, and $KI_{CO,\,k}$ were considered in the SA. With respect to physiochemical parameters, all were fixed except K_P, $k_L a_{H2}$, which were used in the SA.

Table 4.2.1. Parameters considered in the ADM1_ME.

Parameter	Units	Value	Reference	SA	Estimated	Description
Stoichiometric parameters (22 parameters)						
$f_{ac,\mathrm{CO}}$		0.3	-	X		Stoichiometric conversion of CO to acetate
$f_{H2,\mathrm{CO}}$		0.7	-			Stoichiometric conversion of CO to H_2. The sum $f_{ac,\mathrm{CO}} + f_{H2,\mathrm{CO}} = 1$
$f_{H2,su}$		0.19				Stoichiometric conversion of sugar to H_2
$f_{bu,su}$		0.13				Stoichiometric conversion of sugar to butyrate
$f_{pro,su}$	_	0.27				Stoichiometric conversion of sugar to propionate
$f_{ac,su}$		0.41				Stoichiometric conversion of sugar to acetate
$f_{ac,pro}$		0.57	(Rosen and Jeppsson, 2006)			Stoichiometric conversion of propionate to acetate
$f_{H2,pro}$		0.43				Stoichiometric conversion of propionate to H_2
$f_{ac,bu}$		0.8				Stoichiometric conversion of butyrate to acetate
$f_{H2,bu}$		0.2				Stoichiometric conversion of butyrate to H_2
$C_{su,DS1}$		0.0313				Stoichiometric conversion of sugar to inorganic carbon
$C_{su,DS2}$		0.025	-			Stoichiometric conversion of sugar to inorganic carbon
C_{pro}	$mol_c/gCOD$	0.0268				Stoichiometric conversion of propionate to inorganic carbon
C_{ac}		0.0313	(Rosen and Jeppsson, 2006)			Stoichiometric conversion of acetate to inorganic carbon
C_{bac}		0.0313				Stoichiometric conversion of biomass to inorganic carbon

C_{ch4}		0.0156				Stoichiometric conversion of methane to inorganic carbon
Y_{su}	$gCOD_X/gCOD_S$	0.1	(Sun *et al*., 2021)	X	X	Biomass yield from sugar
Y_{bu}		0.06		X	X	Biomass yield from butyrate
Y_{pro}		0.04		X	X	Biomass yield from propionate
Y_{ac}		0.05		X	X	Biomass yield from acetate
Y_{CO}		0.025		X	X	Biomass yield from CO
Y_{H2}		0.06	(Rosen and Jeppsson, 2006)	X		Biomass yield from H_2
Biochemical parameters (23 parameters)						
$\mu_{m,su}$	$1/d$	45	(Sun *et al*., 2021)	X	X	Maximum growth rate of sugar biomass
$\mu_{m,bu}$		20		X		Maximum growth rate of butyrate biomass
$\mu_{m,pro}$		13		X	X	Maximum growth rate of propionate biomass
$\mu_{m,ac}$		12.5		X	X	Maximum growth rate of acetate biomass
$\mu_{m,CO}$		75		X		Maximum growth rate of CO biomass
$\mu_{m,H2}$		90		X	X	Maximum growth rate of hydrogen biomass
Ks_{su}	$gCOD/L$	0.02		X	X	Substrate saturation constant of sugar biomass
Ks_{bu}		0.3		X		Substrate saturation constant of butyrate biomass
Ks_{pro}		0.3		X		Substrate saturation constant of propionate biomass
Ks_{ac}		0.05		X	X	Substrate saturation constant of acetate biomass
Ks_{CO}		8×10^{-6}		X		Substrate saturation constant of CO biomass
Ks_{H2}		1×10^{-6}		X	X	Substrate saturation constant of hydrogen biomass
$KI_{H2,\,ac}$	$gCOD/L$	1×10^{-6}	-	X	X	Inhibition constant of H_2 to acetate
$KI_{H2,\,bu}$		1×10^{-5}	(Rosen and Jeppsson, 2006)	X		Inhibition constant of H_2 to butyrate
$KI_{H2,\,pro}$		3.5×10^{-6}		X		Inhibition constant of H_2 to propionate
$KI_{CO,\,ac}$		1×10^{-6}	-	X		Inhibition constant of CO to acetate
$KI_{CO,H2}$		1×10^{-6}	-	X		Inhibition constant of CO to H_2
$K_{su,\,dec}$	$1/d$	0.02	(Rosen and			Sugar biomass decay

			Jeppsson, 2006)			constant
$K_{bu,\,dec}$		0.02				Butyrate biomass decay constant
$K_{pro,\,dec}$		0.02				Propionate biomass decay constant
$K_{ac,\,dec}$		0.02				Acetate biomass decay constant
$K_{CO,\,dec}$		0.02				CO biomass decay constant
$K_{H2,\,dec}$		0.02				H_2 biomass decay constant
Physiochemical parameters (15 parameters)						
K_P	$m^3/bar-d$	5×10^4	(Rosen and Jeppsson, 2006)	X	X	Parameter related to the friction in the gas outlet
k_La_{H2}	$1/d$	200		X		Volumetric mass transfer coefficient of H_2
$H_{H2,OPi}$	$mol/L-bar$	*7.3074×10^{-4} **6.6857×10^{-4}	(Sander, 2015)			Henry's law constants.The Van't Hoff equation was used to correct by temperature: $H_i = H_i^0 e^{-\frac{\Delta_{sol}\mathcal{H}_i}{8.314}\left(\frac{1}{298.15}-\frac{1}{T_{DSi}}\right)}$ $i= H_2, CO_2, CH_4$,* T_{DS} = 310.15 K, ** T_{DS2}=328.15 K, $\Delta_{sol}\mathcal{H}_i$=Enthalpy of dissolution, H_i^0= Henry's constant at reference temperature
$H_{CO2,OPi}$		*0.0259 **0.0171				
$H_{CH4,DSi}$		*0.0011 **8.2805×10^{-4}				
H_{CO}		9.7×10^{-4}				Henry's law constants
D_{CH4}	m^2/s	1.57×10^{-9}	(Cussler, 2011)			Diffusivity constant of CH_4
D_{H2}		4.65×10^{-9}				Diffusivity constant of H_2
D_{CO2}		1.98×10^{-9}				Diffusivity constant of CO_2
D_{CO}		2.03×10^{-9}				Diffusivity constant of CO
$b_{0,OP1}$	–	0.467	(Deckwer *et al.*, 1983)			Parameter in k_La_i calculation with DS1
$b_{1,OP}$	–	0.82				Parameter in k_La_i calculation with DS1.Value adjusted to 0.62
$b_{0,OP2}$	–	2.6×10^{-2}	(Van't, 1979)			-Parameter in k_La_i calculation with DS2
$b_{1,OP2}$	–	0.4				Parameter in k_La_i calculation with DS2
$b_{2,OP2}$	–	0.5				Parameter in k_La_i calculation with DS2

2.1 Sobol' Method Results

From Sobol' method (Figure 4.2.1-Figure 4.2.2), Sobol' sensitivy analysis with OP1 and OP2 over outputs $q_{gas,i}$ and $p_{gas,i}$ were computed. It was observed that parameters such as Y_{bu}, Y_{pro}, and $\mu_{m,su}$, presented a representative effect in at least two of the model outputs with OP1. On the other hand, Y_{CO}, $\mu_{m,su}$, $\mu_{m,pro}$, $KI_{H2,\,ac}$, and Ks_{su} presented a representative effect in at least two of the model outputs with OP2.

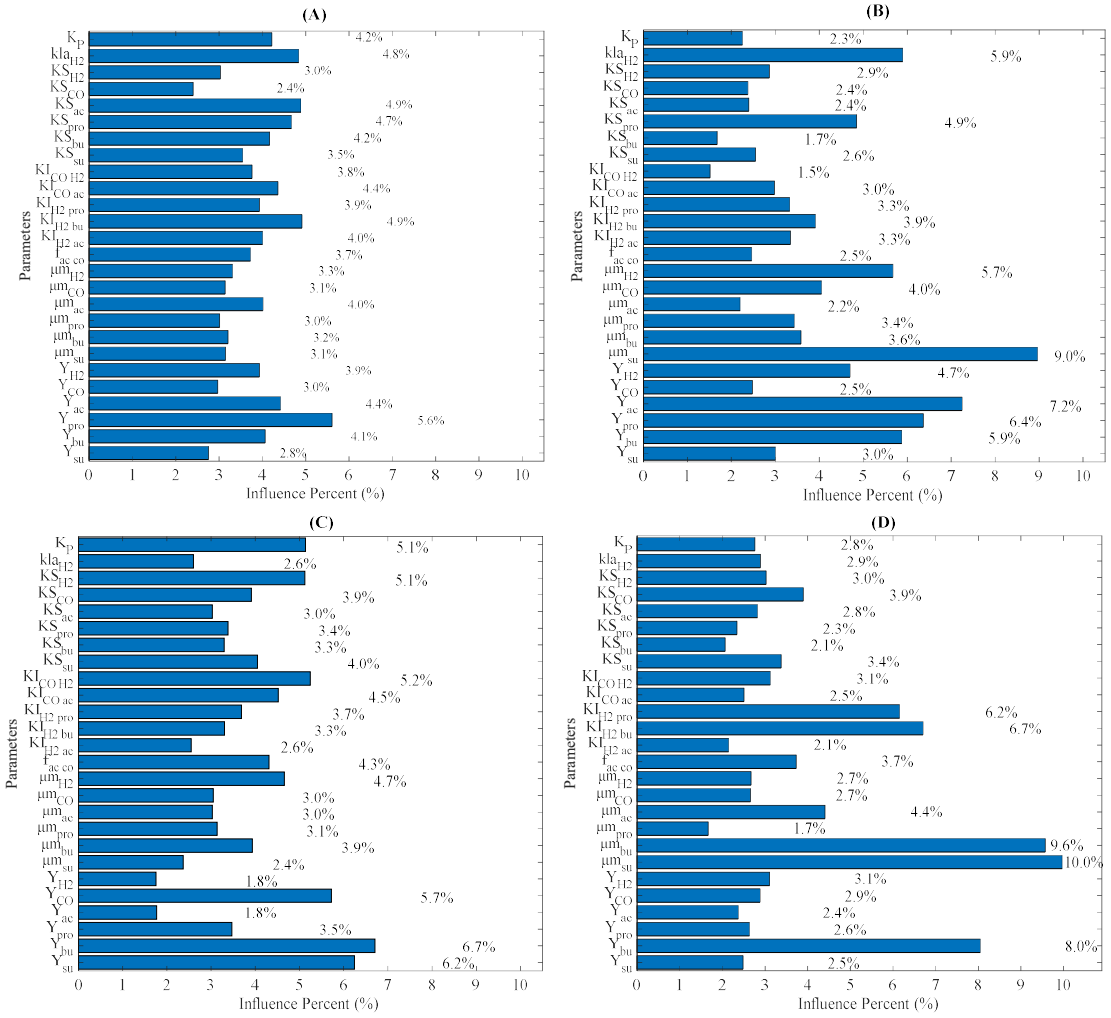


Figure 4.2.1. First order sensitivity index with a scalar characteristic (SI_{is}) with OP1 over outputs: (A) $q_{gas,CH4}$, (B) $q_{gas,H2}$, (C) $q_{gas,CO}$, and (D) $q_{gas,CO}$. The influence is calculated based on scalar characteristic $SI_{is} = \xi(SSE) = \xi\left(\sum(Y_m(\theta_i) - Y_m(\theta_{i,nom}))^2\right)$.

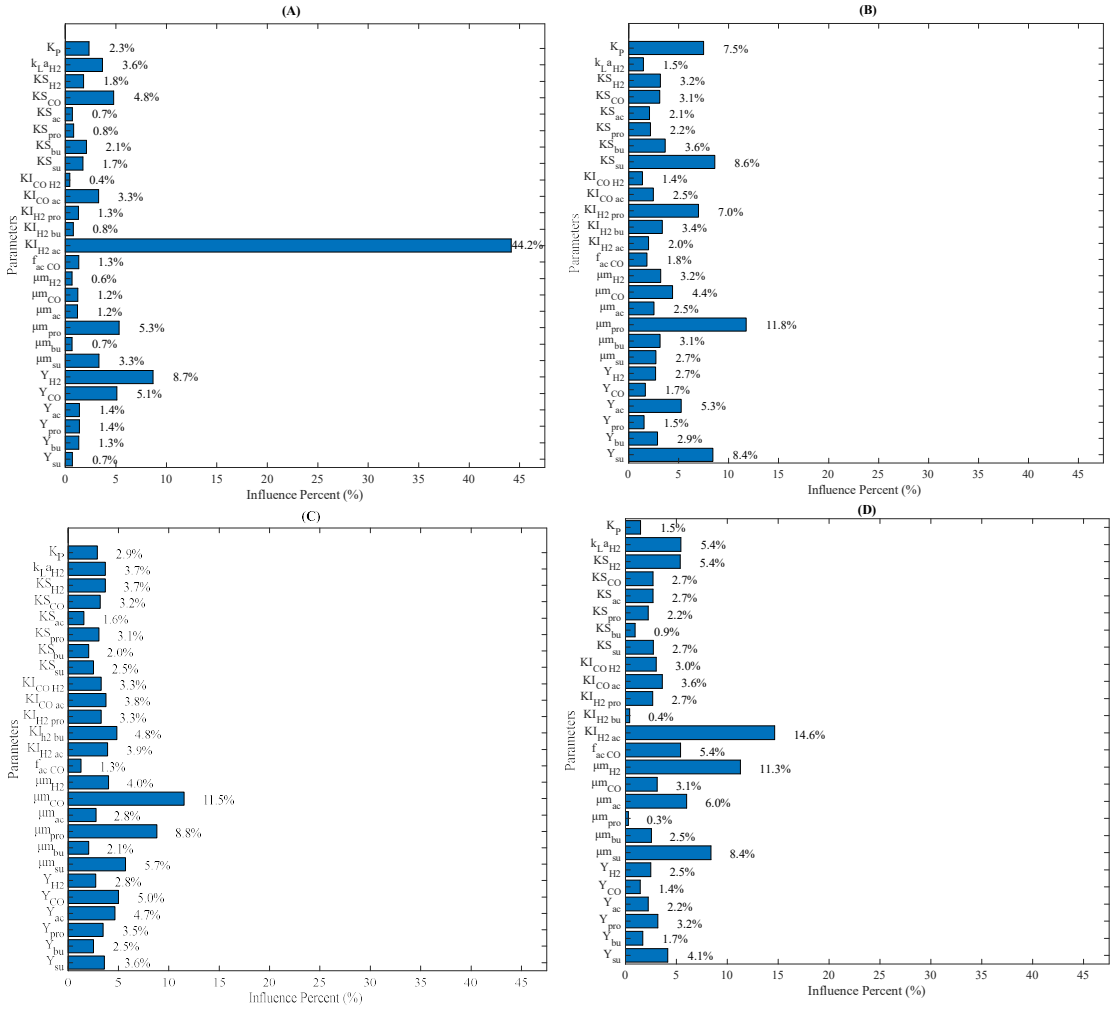


Figure 4.2.2. First-order sensitivity index with a scalar characteristic (SI_{is}) with OP2 over outputs: (A) $p_{gas,CH4}$, (B) $p_{gas,H2}$, (C) $p_{gas,CO}$, and (D) $p_{gas,CO2}$. The influence is calculated based on scalar characteristic $SI_{is} = \xi(\text{SSE}) = \xi\left(\sum(Y_m(\theta_i) - Y_m(\theta_{i,nom}))^2\right)$.

2.2 Morris Method Results

With the Morris method (Table 4.2.2), parameters such as Y_{su}, Y_{pro}, Y_{ac}, $\mu_{m,su}$, $\mu_{m,pro}$, $\mu_{m,ac}$, Ks_{su}, Ks_{pro}, Ks_{ac}, and K_P presented a representative effect in at least two of the model outputs with DS1. On the other side, Y_{su}, Y_{ac}, Y_{H2}, Y_{CO}, $\mu_{m,su}$, $\mu_{m,ac}$, $\mu_{m,H2}$, $\mu_{m,CO}$, $KI_{H2,ac}$, Ks_{ac}, Ks_{H2}, and Ks_{CO} presented a representative effect on at least two of the model outputs with OP2. From the sensitivity analysis using both methods, 14 parameters

are selected for estimation: Y_{su}, Y_{bu}, Y_{pro}, Y_{ac}, Y_{CO}, $\mu_{m,su}$, $\mu_{m,pro}$, $\mu_{m,ac}$, $\mu_{m,H2}$, $KI_{H2,\,ac}$, Ks_{su}, Ks_{ac}, Ks_{H2}, K_P.

Table 4.2.2. Most influence parameters in Morris Method.

Model outputs		Parameters
OP1	$q_{gas,CH4}$	$\mu_{m,su}$, Y_{su}, Ks_{ac}, Y_{ac}, $\mu_{m,ac}$, $\mu_{m,bu}$, Y_{pro}, Y_{bu}, Ks_{pro}, Ks_{su}, K_P, $\mu_{m,pro}$
	$q_{gas,H2}$	K_P, $\mu_{m,ac}$, $\mu_{m,su}$, Y_{su}, Y_{ac}, Ks_{ac}, Y_{pro}, Ks_{pro}, $\mu_{m,pro}$, Ks_{su}
	$q_{gas,CO}$	$\mu_{m,su}$, K_P, Y_{su}, Ks_{su}, Y_{ac}, Y_{pro}, $\mu_{m,pro}$, $\mu_{m,ac}$
	$q_{gas,CO2}$	$\mu_{m,su}$, Y_{su}, Y_{ac}, $\mu_{m,ac}$, Ks_{su}
OP2	$p_{\mathrm{gas,CH4}}$	$\mu_{m,su}$, $KI_{CO,\,H2}$, Y_{ac}, Ks_{H2}, $\mu_{m,H2}$, $\mu_{m,ac}$, Y_{su}, Y_{H2}, $KI_{H2,\,ac}$, Ks_{ac}
	$p_{\mathrm{gas,H2}}$	Ks_{H2}, Y_{H2}, $\mu_{m,H2}$, Ks_{CO}, $\mu_{m,CO}$, Y_{CO}, Y_{su}, $\mu_{m,ac}$, $\mu_{m,bu}$, $\mu_{m,su}$, Y_{ac}
	$p_{\mathrm{gas,CO}}$	Ks_{CO}, Ks_{H2}, $\mu_{m,su}$, $\mu_{m,CO}$, Y_{CO}, $\mu_{m,H2}$, Y_{su}
	$p_{\mathrm{gas,CO2}}$	Y_{CO}, Ks_{su}, $\mu_{m,pro}$, Y_{pro}, $\mu_{m,CO}$, $f_{ac,CO}$

2.3 References

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). Water Science and Technology 45, 65–73. https://doi.org/10.2166/wst.2002.0292

Cussler, E.L., 2011. Diffusion: mass transfer in fluid systems, 3. ed., 4. printing. ed. Cambridge Univ. Press, Cambridge.

Deckwer, W.D., Nguyen-Tien, K., Kelkar, B.G., Shah, Y.T., 1983. Applicability of axial dispersion model to analyze mass transfer measurements in bubble columns. AIChE Journal 29, 915–922. https://doi.org/10.1002/aic.690290607

Henze, M., Loosdrecht, M.C.M. van, Ekama, G.A., Brdjanovic, D., 2019. Biological Wastewater Treatment: Principles, Modeling and Design. IWA Publishing. https://doi.org/10.2166/9781780408613

Rosen, C., Jeppsson, U., 2006. Aspects on ADM1 Implementation within the BSM2 Framework, Technical report. Department of Industrial Electrical Engineering and Automation (IEA), Lund University.

Sander, R., 2015. Compilation of Henry's law constants (version 4.0) for water as solvent. Atmospheric Chemistry and Physics 15, 4399–4981. https://doi.org/10.5194/acp-15-4399-2015

Sun, H., Yang, Z., Zhao, Q., Kurbonova, M., Zhang, R., Liu, G., Wang, W., 2021. Modification and extension of anaerobic digestion model No.1 (ADM1) for syngas biomethanation simulation: From lab-scale to pilot-scale. Chemical Engineering Journal 403, 126177. https://doi.org/10.1016/j.cej.2020.126177

Van't, R.K., 1979. Review of Measuring Methods and Results in Nonviscous Gas-Liquid Mass Transfer in Stirred Vessels 18, 357–364.

Section 3 Market Prices

Table 4.3.1 summarizes the cost production and selling prices used in the Economic Multi-Objective Optimization.

Table 4.3.1. Cost production of sugar and syngas and selling prices of biogas and acetate.

Component	Value	Unit	Conversion ($EUR/gCOD$)	Reference*
Sugar Cost production	0.321	USD/kg	2.86×10^{-4}	European Union Sugar Import Price
	420	EUR/Ton	3.94×10^{-4}	World Sugar Market
Mean Value			$3.40\times10^{-4}\pm7.63\times10^{-5}$	
Syngas Cost production	90	USD/TMC	1.20×10^{-4}	(Pei *et al*., 2016)
	1.73	$USD/kgH2$	2.05×10^{-4}	International Energy Agency (IEA)
Mean Value			$3.40\times10^{-4}\pm6.06\times10^{-5}$	
Bio-gas (biomethane) selling price:	0.1645	EUR/KWH	6.30×10^{-4}	Selectra
	0.1325	EUR/KWH	5.07×10^{-4}	2_fournisseurs-electricite.com
	0.1645	EUR/KWH	6.30×10^{-4}	3_fournisseurs-electricite.com
	0.1615	EUR/KWH	6.18×10^{-4}	4_fournisseurs-electricite.com
Mean Value			$5.96\times10^{-4}\pm5\text{-}96\times10^{-5}$	
Acetate Selling Price	1805	USD/Ton	1.61×10^{-3}	Chemanalyst
	1845	USD/Ton	1.64×10^{-3}	(Vidra & Németh, 2017)

Mean Value	$1.63\times10^{-3}\pm2.52\times10^{-5}$

*References consulted on December 2022.

All the market prices were converted on units of $EUR/gCOD$ as following,

- **Sugar Cost Production**

$$\frac{USD}{kg}\frac{180.15g}{1mol}\frac{1kg}{1x103g}\frac{1mol}{192gCOD}\frac{0.95EUR}{1USD}[=]\ \frac{EUR}{gCOD}$$

$$\frac{EUR}{TON}\frac{1TON}{1x106g}\frac{180.15g}{1mol}\frac{1mol}{192gCOD}[=]\ \frac{EUR}{gCOD}$$

- **Syngas Cost Production :**

$$\frac{USD}{TCM}\frac{TCM}{1\times10^{6}L}\frac{22.4L}{1mol}\frac{1mol}{16gCOD}\frac{0.95EUR}{1USD}[=]\ \frac{EUR}{gCOD}$$

$$\frac{EUR}{KWH}\frac{39.17KWH}{kg}\frac{1kg}{1\times10^{3}g}\frac{2g}{mol}\frac{1mol}{16gCOD}[=]\ \frac{EUR}{gCOD}$$

$$\frac{USD}{kg_{H2}}\frac{2g}{mol}\frac{1kg}{1\times10^{3}g}\frac{1mol}{16gCOD}\frac{0.95EUR}{1USD}[=]\ \frac{EUR}{gCOD}$$

-Bio-gas (biomethane) Selling price:

$$\frac{EUR}{KWH}\frac{1KWH}{3.6MJ}\frac{55MJ}{kg}\frac{1kg}{1\times10^{3}g}\frac{16.04g}{1mol}\frac{1mol}{64gCOD}[=]\ \frac{EUR}{gCOD}$$

$$\frac{EUR}{MWH}\frac{1MWH}{1\times10^{3}KWH}\frac{1KWH}{3.6MJ}\frac{55MJ}{kg}\frac{1kg}{1\times10^{3}g}\frac{16.04g}{1mol}\frac{1mol}{64gCOD}[=]\ \frac{EUR}{gCOD}$$

- **Acetate Selling Price :**

$$\frac{USD}{TON}\frac{1TON}{1x106g}\frac{60.052g}{1mol}\frac{1mol}{64gCOD}\frac{0.95EUR}{1USD}[=]\ \frac{EUR}{gCOD}$$

3.1 References

Pei, P., Korom, S. F., Ling, K., & Nasah, J. (2016). Cost comparison of syngas production from natural gas conversion and underground coal gasification. *Mitigation and Adaptation Strategies for Global Change*, *21*(4), 629–643. https://doi.org/10.1007/s11027-014-9588-x

Vidra, A., & Németh, Á. (2017). Bio-produced Acetic Acid: A Review. *Periodica Polytechnica Chemical Engineering*, *62*(3), 245–256. https://doi.org/10.3311/PPch.11004