Modeling and Optimization of Anaerobic Digestion: A Review

Fatima Walid1-∗, Sanaa El Fkihi1-∗∗, Houda Benbrahim1-∗∗∗, and Hicham Tagemouati2-∗∗∗∗

1IRDA Team, ENSIAS, Mohammed V University, Rabat, Morocco
2Moroccan Foundation for Advanced Science, Innovation and Research (MAScIR), Madinat Al Irfane, Rabat, Morocco

Abstract. Anaerobic digestion is recognized as being an advantageous waste management technique representing a source of clean and renewable energy. However, biogas production through such practice is complex and it relies on the interaction of several factors including changes in operating and monitoring parameters. Enormous researchers have focused and gave their full attention to mathematical modeling of anaerobic digestion to get good insights about process dynamics, aiming to optimize its efficiency. This paper gives an overview of the different approaches applied to tackle this challenge including mechanistic and data-driven models. This review has led us to conclude that neural networks combined with metaheuristic techniques has the potential to outperform mechanistic and classical machine learning models.

1 Introduction

Demographic pressure, industrial development, and changes in consumption patterns are leading to an increase in waste volume. It is becoming necessary to involve innovative sustainable methods of waste treatment and management to overcome the challenges of climate change besides air and water pollution. One of the powerful techniques of waste recovery is the anaerobic digestion (AD) process also known as methanization. It is among the economical and effective techniques applied in organic waste treatment and recovery, that is widely used in the agricultural and industrial domains.

AD is a natural treatment performed by a community of microorganisms in the absence of air, that involves the disintegration of complex organic wastes and generates biogas which is mainly composed of methane \((CH_4)\) (60-70%) and carbon dioxide \((CO_2)\) (30-40%) [1]. Subsequently, the produced biogas is reevaluated to produce electricity, heat, or can be used as biofuel.

AD is an advantageous waste management practice since it reduces the polluting organic load in waste and allows the recovery of various types of substrates including sewage sludge, urban, livestock, agricultural, and industrial waste [2]. It represents a source of clean and renewable energy, unlike the conventional energy sources which have negative implications on the environmental balance. Furthermore, AD produces digestate that is potentially fit for use as bedding for livestock or as an organic amendment (i.e. fertilizer) for the benefit of farms and agricultural communities.

The selection and monitoring of indicators that influence biogas production play a vital role in process optimization as they provide information on the condition and functioning of the system. A growing body of literature has focused and put effort into mathematical modeling of the AD process because it is a powerful tool for a better understanding of the process dynamics, thus it provides optimization opportunities for enhancing digester’s performance [3].

This paper reviews AD modeling and optimization regarding process parameters as well as architectures built to predict process behavior. In Section 2, some backgrounds about AD are presented including main process factors and monitoring, whereas mechanistic and data-driven based models are discussed in Section 3. The fourth section is dedicated to discussing the advantages and limitations of the presented approaches.

2 Background Review

During the entire AD process, there is a group of anaerobic microorganisms for each step that works together to breakdown the organic substance and generate biogas. As presented in Fig. 1, the AD process occurs in four main stages namely: hydrolysis, acidogenesis, acetogenesis, and methanogenesis [4, 5].

At the hydrolysis stage, the complex compounds are broken into soluble components, it corresponds to the enzymatic transformation of high molecular weight components (organic polymers and lipids) into simple molecules (monomers: fatty acids, monosaccharides, amino acids), which can be assimilated by the microbial metabolism and thus used as a source of energy [1, 5]. Then comes fermentation which includes the acidogenesis and acetogenesis stages. Acidogenesis is the step where the components resulted from hydrolysis are converted into intermediate components with a lower molecular weight. During this step, simple volatile fatty acids
(VFAs) are formed from the sugar and amino acids. Whilst acetogenesis refers to the fermentation process of VFAs and alcohols to acetate, hydrogen, and carbon dioxide [1, 5].

Finally, a group of archaea organisms called methanogens carry out the bacterial conversion of intermediate components into various simpler end products, mainly \( \text{CH}_4 \) and \( \text{CO}_2 \) [1, 5].

2.1 Process key factors

The microorganism groups during AD are sensitive to several factors such as changes in some monitoring parameters including pH, alkalinity, VFAs as well as some operating parameters (temperature, hydraulic retention time, substrate composition, and organic loading rate (OLR)) [6]. As in Fig. 2, monitoring and operating parameters that have been taken into consideration by researchers take place in the liquid and gas phases.

Temperature influences the metabolic activities of the microorganisms during fermentation [7], which affects the rate of digestion and methane production [1]. There are three operating ranges considered for AD including psychrophilic (4-15°C), mesophilic (20-40°C), and thermophilic (45-70°C).

OLR is among the key operational parameters in AD systems. The biogas production rate is correlated to OLR in a way that maintenance of OLR at an appropriate range promotes biogas yield [6]. However, a significant increase in OLR beyond the suitable range is a major factor of VFAs accumulation, therefore decreasing biogas production [6].

pH is a key parameter of AD because it determines the predominant microorganisms at each stage of fermentation. During AD process, pH is generally within the range (0.6-0.8), its variation depends on the evolution of the fermentation of organic matter. pH is important for the fulfillment of the methanogenic phase since a decrease in pH results in an inhibition of the process [6].

The concentration of VFAs has been widely suggested for the control and monitoring of anaerobic digesters as it is the main methanogenic intermediate also its accumulation in reactors is reliable in indicating process imbalance [8]. It is mostly used to determine the stress level of the system because it can provide specific information for process diagnosis [6, 8].

The measurement of biogas flow and composition is very important because they indicate the overall performance of the digester [8]. It is reported that low \( \text{CH}_4 \) content in biogas may indicate inhibition of methanogenic bacteria. \( \text{CH}_4 \) concentration must be above 50% to ensure efficient operation. Moreover, the presence of an elevated concentration of hydrogen sulfide (\( \text{H}_2\text{S} \)) in biogas may cause inhibition in AD processes also serious emission and corrosion problems [9].

Furthermore, it has been reported that considering single indicators is not valid for all the digesting systems. Alternatively, the combination of several indicators such as VFA, the ratio of bicarbonate alkalinity to total alkalinity, and the ratio of VFA to total alkalinity can provide fast, reliable, and complementary information about the metabolism of the system; which encourage fast and effective early warning [19]. The most used indicators in previous works are summarized in Table 1.

Regular process monitoring helps in providing information about the state and process performance, therefore
detecting and reacting to major process imbalances and disturbances [20]. It is important to detect the process imbalance at the beginning therefore action can be taken in time to prevent any process failure. For this reason, it is crucial to monitor and regularly control AD parameters in order to ensure that the process is running smoothly thus maximizing biogas production [1].

3 AD Modeling and Optimization

Process modeling aims to design and describe a system for the purpose of a better understanding of the occurring operations and its optimal working conditions. It can also be used to control a system and forecast its behavior and outcomes. Several research studies took advantage of this powerful tool to guarantee efficient control and optimization of biogas production aiming to enhance digesters performance.

Available AD models can be classified into two main categories: mechanistic and data-driven models. Mechanistic models are based on biological, physical, and chemical laws; whereas empirical data-driven models are based on mathematical equations to determine the relationships between input and output variables using measured process data [27].

3.1 Mechanistic Models

Mechanistic models refer to the models using a defined set of differential equations that describe the biological and physicochemical laws of a process [28]. More specifically, they tend to characterize the growth and inhibition of bacteria according to the substrate.

The procedures for developing a mechanistic model require three main components. First, fundamental knowledge of the interactions between process variables. Second, determination of model parameters using experimental data. Last, data collection from the process, which is needed for validation; and as long as the designed model is not efficient, the process knowledge is reviewed and the model is restructured [29, 30].

Among widely used mechanistic models to describe AD processes there is Anaerobic Digestion Model No. 1 (ADM1) that was proposed by researchers from the international water association [30], their goal was to develop a generalized model allowing the best possible simulation of anaerobic digesters.

ADM1 includes multiple steps describing physicochemical reactions, namely: ion association and dissociation as well as gas-liquid transfer, besides biochemical interactions including degradation of organic materials of substrate together with bacteria’s growth [30].

ADM1 can be performed as a differential equations model, which contains 32 dynamic concentration state variables and 6 acid-base kinetic processes, or as a differential and algebraic equations model, which contains 26 dynamic state concentration variables and 8 implicit algebraic [30].

It is well known that ADM1 is one of the most comprehensive available AD models, it has been broadly applied in AD processes for methane production [31–36].

3.2 Data-driven models

In contrast to the models previously discussed, the objective of data-driven models is to pattern the behavior of the system without any prior knowledge of the occurring operations. It is an intersection of several domains including mathematical modeling, statistics, information theory, and data science. These powerful tools are useful for identifying the structure of a given process together with analyzing the correlations between its components without prior knowledge [37].

The learning process and the construction of knowledge require historical data (dataset) that is a set of examples, each is represented by a set of characteristics also called variables or attributes. Using the experience gained in the training phase, a testing phase comes immediately after which serves to evaluate the model formed by a classification, prediction, or clustering of new examples called test data. The evaluation of the model is calculated by a performance measure that improves during data training [37].
3.2.1 Classical machine learning

Sathish et al. [12] carried out anaerobic biodigester optimization by applying response surface methodology to evaluate the relationship between biogas yield and four input variables, namely: pH, temperature, agitation time, and substrate concentration. This study reported that response surface methodology could be used effectively for predicting the biogas yield of agricultural waste since its coefficient of determination ($R^2$) was 0.991. Moreover, temperature and substrate concentration showed the most significant effect in a way that increasing them can significantly boost the biogas output. Besides, the value of pH and agitation time shows significant interactive effects, which demonstrates that the inhibition is caused by a decrease in pH value.

Akbas et al. [10] have performed the decision tree algorithm on a dataset of 776 data points collected from a wastewater treatment plant in Turkey. The list of input variables was: temperature, pH, VS, sludge loading rate, and total suspended solids, while biogas flow was the output entity. They have reported that the decision tree algorithm gave the most precise prediction with a regression coefficient of 0.80 compared to the curvature and interaction curvature test.

Wang et al. [11] have carried out a comparison of the performance of random forest, support vector machine, and K-nearest neighbors in regression and classification models. In the classification models, the samples were classified into 3 groups based on methane production which were low, medium, and high. The process parameters selected as input entities to build the dataset were: total carbon, total nitrogen, the C/N ratio, xylan content, and temperature whereas the output variable was methane production rate.

The classification model using logistic regression along with the elastic net penalty gave the best precision of 0.73, and the regression model using K-nearest neighbors gave the best precision of 26.7 that was evaluated by the root mean square error. Moreover, they have reported that the total carbon load was considered the most important characteristic.

In another work, Kazemi and co-authors [16] implemented different data-driven methods including random forest, support vector machine, and genetic programming in order to predict the effluent VFAs. The used data consist of 5 input variables (COD, alkalinity, total soluble solids, biological oxygen demand, and gas flow). Feature ranking was performed using the fscaret method along with the support vector machine algorithm, and the best subset of input variables selected was: pH, pressure, $CO_2$, and ammonia concentration. All models except random forest achieved precise VFAs prediction and genetic programming performance was the best in terms of accuracy, robustness, and transparency.

Qdais et al. [13] built a multi-layer ANN with two hidden layers, each with 25 neurons and a sigmoid activation function in order to determine the nonlinear impact of digester pH, temperature, TS, and VS on the quantity of biomethane produced. The daily operational data collected contain 227 records and about 78% of examples were used for training the model whereas the remaining records were used for the test. As a result, the neural network was able to predict the methane output with a correlation coefficient of 0.87%.

Dalmau et al. [15] applied a wrapper approach to select the most significant and relevant features for AD imbalances, they have used ANNs as learning algorithms having an input layer, a hidden layer with a sigmoid transfer function, and an output layer with a linear transfer function. In each step, the wrapper method drops one of the features and evaluates the accuracy of the learning algorithm using root square mean error. Thereafter, it removes the feature causing the least reduction in the accuracy. The measurement of parameters was performed using online sensors both in the liquid phase (pH, flow rates, VFAs, total organic carbon, and COD concentrations) and in the gas phase ($CO_2$, $CH_4$, $H_2$). Acidogenic state results showed that pH, VFAs, and entry rates were the most relevant variables whereas in the foaming state incoming flow and organic carbon were among the relevant variables that were related to the feedstock of the digester.

Similarly, Sathish et al. [12] carried out anaerobic biodigester optimization by applying ANN and considered four input variables (pH, temperature, agitation time, and substrate concentration). The network consists of an input layer, a hidden layer with a sigmoid activation function, and an output layer with a gaussian linear transfer function. Results showed that ANN was effective for predicting the biogas yield of agricultural waste ($R^2 = 0.998$).

However, Güçlü et al. [14] modeled the anaerobic sludge digester of the Ankara central wastewater treatment plant by implementing ANN using two different backpropagation methods: the gradient descent with adaptive learning rate and levenberg–marquardt algorithm. Temperature, pH, gas flow rate, VFAs, alkalinity, dry matter, and organic matter were the inputs to predict effluent VS concentration and methane yield. Data were collected during 245 days and more than 100 000 neural networks were trained to determine the best ANN model by tuning several hyperparameters such as iteration numbers, transfer functions, and hidden nodes. The best backpropagation algorithm was the gradient descent with an adaptive learning rate; the $R^2$ correlations were 0.89 and 0.71 for VS concentration and methane yield respectively.

Beltramo et al. [22] performed also ANNs using 15 variables including the concentration of VFAs, TS, VS, ammoniacal nitrogen, hydraulic retention time, organic charge rate, and biogas yield. The Measurements were taken over ten months with a frequency of one sampling per week. The data of measured variables were used as input neurons, while the rate of biogas production was calculated by the output neuron. They reported a prediction with an error of 13.08% and a precision of $R^2 = 0.76$. 

3.2.2 Neural networks

Much work on the potential of artificial neural network (ANN) in the AD field has been carried out [12–15, 22].
Besides the wide application of ANN in the AD field, other works explored the performance of other types of neural networks. For example, Yesileveki et al. [38] studied the possibility of building an AD adaptive control system in order to improve biogas production efficiency. The system was based on the recursive neural network (LSTM) and reinforcement learning. The experiment showed that the learning convergence could be reached after 600 epochs.

McCormick et al. [17] studied predictive monitoring of the AD process using LSTM and one-dimensional convolutional neural network. They performed three dimensional reduction methods to identify the significant contributors to process stability. The principal component analysis showed that TS, carbon, waste loading rate, nitrogen content of the feed, and the TS concentration in the digester make the greatest contributions to the overall variance of the system. Results of the independent component analysis showed that VFA concentration and $H_2$ content in the biogas were the most important independent variables of the data. According to partial least squares analysis, OLR and the nitrogen content of the feed had the greatest magnitudes. Authors selected total mass of feed, $pH$, ammonium concentration, and VFA concentration as input features and biogas flow rate as a target variable. The one-dimensional convolutional model was best able to accurately make biogas yield predictions on unseen data.

### 3.2.3 Neural networks combined with metaheuristic algorithms

A growing body of literature has explored ANN application in predicting biogas through the AD process. With the purpose to improve its performance, experiments on combining ANN with metaheuristic methods were conducted and reported prominent results as they helped to identify the optimal combination of process parameters which consequently improve the prediction of biogas yield.

Qdais et al. [13] used the designed neural network as a fitness function for the genetic algorithm (GA) in order to optimize the methane output and it provided the opportunity to improve methane production by $6.9\%$.

Ilamathi et al. [39] explored simulated annealing for predictive modeling and optimization of nitrogen oxides emission, it was used to determine the optimum level of input operating parameters that provide minimum nitrogen oxides emission. They reported that the optimized results agree well with the experimental experience, leading to low nitrogen oxide emission.

Beltramo et al. [22] implemented an ant colony and GA optimization to select and identify the significant process variables which reduced the size of the model and improved the ability to identify predicates from ANN models. The GA and ant colony combination provided the best prediction of the biogas yield with a coefficient of determination of $R^2 = 0.9$ and a prediction error of $6.24\%$.

### 4 Discussion

The identification and selection of indicators influencing the AD process are key factors for process optimization because they provide information on the condition and functioning of the system. Monitoring such indicators is an essential practice as it ensures the process smooth running and also gives the ability to detect process malfunctions at an early stage.

We can notice that each of the previous research has focused on a limited set of indicators and did not capitalize on the whole set including those which are related to the substrate, biodegradation, as well as the ones involved in the gas phase. It can thus be conceivably hypothesized that considering the combination of these parameters may improve biogas productivity, namely: temperature, $pH$, redox potential, alkalinity, ammonium, and VFA concentrations, C/N ratio together with pressure, gas flow, and production rate of the different gas components including $CH_4$, $CO_2$, $H_2S$, and $H_2$. This suggestion is given based on the direct influence of each of the mentioned parameters on biogas production along with their complementary interaction.

Moreover, this review has highlighted that although mechanistic models are complete patterns in terms of describing AD steps, some potential weak points need to be considered. First, the requirement of having a good knowledge of the kinetic and stoichiometric parameters of bioreactors. Second, the incomplete understanding of AD microbial and physicochemical processes as well as the complexity of implementing it digitally. These limitations are considered as a major obstacle for the optimization and synthesis of control laws in fermentation through such models.

However, data-driven based models have been found to be more effective in providing information on the behavior and the relationships between process factors compared to mechanistic methods. This strength lies in the fast estimation of optimal conditions in the absence of explicit knowledge of the system’s dynamic, also the precise analysis of the relationships among process state variables with respect to its nonlinear and complex behavior.

Additionally, we have seen that the combination of metaheuristic algorithms with artificial neural networks is very efficient in biogas production optimization compared to classical machine learning methods. On the other hand, machine learning-based techniques do not require a lot of data and perform predictions in less time.

In fact, neural networks-based techniques are costly consuming as they propose high computational prerequisites including processing ability, memory capacity, and long training time. This leads us to take advantage of metaheuristic and neural networks algorithms and put an effort to find a trade-off in robustness, cost, and performance.

### 5 Conclusion

The complex interactive behavior of the various factors in the AD process, necessitate adopting efficient strategies that seek better modeling of the biogas production...
procedure. Research has tended to focus on data-driven based models because they allow rapid estimation of optimal conditions without explicit knowledge of the system’s physical reactions, as well as precise analysis of the relationships among process state variables with respect to its nonlinear and complex behavior. This literature review shows us that the combination of metaheuristic algorithms with artificial neural networks is very efficient in biogas production optimization compared to classical machine learning methods. Nevertheless, neural networks is costly consuming in terms of performance and time complexity, it also requires vast and huge data sets to build accurate models. This guides us to take advantage of metaheuristic and neural networks algorithms and put an effort to find a trade-off in robustness, cost, and performance.

6 Acknowledgments

The authors gratefully acknowledge the financial support of the Research Institute for Solar and Energy and New Energies (IRESEN) and the Moroccan Foundation for Advanced Science, Innovation and Research (MAScIR).

References

[1] K.B. Cantrell, T. Ducey, K.S. Ro, P.G. Hunt, Bioresource technology 99, 7941 (2008)
[2] J.B. Holm-Nielsen, T. Al Seadi, P. Oleskowicz-Popiel, Bioresource technology 100, 5478 (2009)
[3] J. Lauwers, L. Appels, I.P. Thompson, J. Degrèvre, J.F. Van Impe, R. Dewil, Progress in Energy and Combustion Science 39, 383 (2013)
[4] G. Merlin, H. Boileau, Anaerobic digestion: types, processes and environmental impact. New York: Nova Science Publishers, Inc (2013)
[5] K.R. Manchala, Y. Sun, D. Zhang, Z.W. Wang, in Advances in Bioenergy (Elsevier, 2017), Vol. 2, pp. 69–141
[6] C.O. Onwosi, I.E. Eke, V.C. Igboke, J.N. Odimba, J.K. Nwagu, Renewable and Sustainable Energy Reviews 116, 109424 (2019)
[7] A.A. Issah, T. Kabera, F. Kemausuor, Biomass and Bioenergy 133, 105449 (2020)
[8] K. Boe, D.J. Batstone, J.P. Steyer, I. Angelidaki, Water research 44, 5973 (2010)
[9] C. Wolf, S. McLoone, M. Bongards, Automatisierungstechnik 57, 638 (2009)
[10] H. AKBAS, G. OZDEMIR, The Eurasia Proceedings of Science Technology Engineering and Mathematics pp. 212–217 (2018)
[11] L. Wang, F. Long, W. Liao, H. Liu, Bioresource Technology 298, 122495 (2020)
[12] S. Sathish, S. Vivekanandan, Alexandria Engineering Journal 55, 3297 (2016)
[13] H.A. Qdais, K.B. Hani, N. Shatnawi, Resources, Conservation and Recycling 54, 359 (2010)
[14] D. Güçlü, N. Yılmaz, U.G. Ozkan-Yucel, Journal of Chemical Technology & Biotechnology 86, 691 (2011)
[15] J. Dalmau, J. Comas, I. Rodríguez-Roda, E. Latrille, J.P. Steyer, Water environment research 82, 492 (2010)
[16] P. Kazemi, J.P. Steyer, C. Bengoa, J. Font, J. Giralt, Processes 8, 67 (2020)
[17] M. McCormick, A.E. Villa, LSTM and 1-D Convolutional Neural Networks for Predictive Monitoring of the Anaerobic Digestion Process, in International Conference on Artificial Neural Networks (Springer, 2019), pp. 725–736
[18] A.J. Ward, P.J. Hobbs, P.J. Holliman, D.I. Jones, Bioresource technology 99, 7928 (2008)
[19] L. Li, Q. He, Y. Wei, Q. He, X. Peng, Bioresource technology 171, 491 (2014)
[20] S. Theuerl, C. Herrmann, M. Heiermann, P. Grundmann, N. Landwehr, U. Kreidenweis, A. Prochnow, Energies 12, 396 (2019)
[21] W. Musingarimbi, B.I. Okeleye, V.I. Okudoh, S.K.O. Ntwampe (2019)
[22] T. Beltramo, M. Klocke, B. Hintzmann, Information Processing in Agriculture 6, 349 (2019)
[23] A. Menert, M. Liiders, T. Kurissoo, R. Vilu, Journal of thermal analysis and calorimetry 64, 281 (2001)
[24] A. Domnanovich, D. Strik, L. Zani, B. Pfieffer, M. Karlovits, R. Braun, P. Holubar, Commun. Agric. Appl. Biol. Sci 68, 215 (2003)
[25] B. Chowdhury, S.B. Masgi, H.N.J. Ting, B.R. Dhar, Processes 8, 555 (2020)
[26] E.G. Kana, J. Oloke, A. Lateef, M. Adesiyen, Renewable energy 46, 276 (2012)
[27] A.M. Enitan, J. Adeyemo, F.M. Swalah, S. Kumari, F. Bux, Reviews in Chemical Engineering 33, 309 (2017)
[28] E. Barampouti, S. Mai, A. Vlyssides, Chemical Engineering Journal 106, 53 (2005)
[29] A. Donoso-Bravo, J. Mailier, C. Martin, J. Rodríguez, C.A. Aceves-Lara, A.V. Wouwer, Water research 45, 5347 (2011)
[30] D.J. Batstone, J. Keller, I. Angelidaki, S. Kalyuzhnyi, S. Pavloustathis, A. Rozzi, W. Sanders, H. Siegrist, V. Vavilin, Water Science and technology 45, 65 (2002)
[31] F. Blumensaat, J. Keller (2002)
[32] H.S. Jeong, C.W. Suh, J.L. Lim, S.H. Lee, H.S. Shin, Bioprocess and biosystems engineering 27, 81 (2005)
[33] S. Magnet, J.S. Blanchard, Chemical reviews 105, 477 (2005)
[34] S. Mu, Y. Zeng, P. Wu, S. Lou, B. Tartakovsky, Bioresource technology 99, 3665 (2008)
[35] K. Derbal, M. Bencheikh-Lehocine, F. Cecchi, A.H. Meniai, P. Pavan, Bioresourcer Technology 100, 1539 (2009)
[36] H. Zhou, D. Löffler, M. Kranert, Bioresource Technology 102, 10819 (2011)
[37] K.G. Liakos, P. Busato, D. Moshou, S. Pearson, D. Bochtis, Sensors 18, 2674 (2018)
[38] V. Yesilevskyi, S. Dyadun, V. Kuznetsov, Scientific Bulletin of National Mining University (2019)
[39] P. Ilamathi, V. Selladurai, K. Balamurugan, IAES International Journal of Artificial Intelligence 1, 11 (2012)