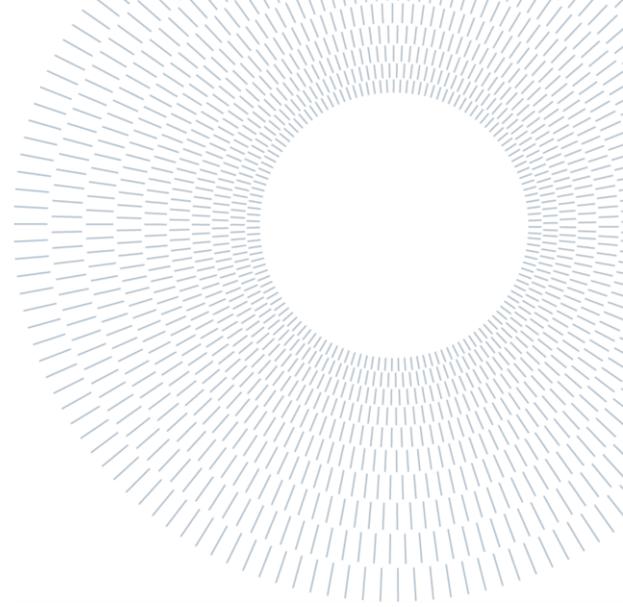




**POLITECNICO
MILANO 1863**

**SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE**



EXECUTIVE SUMMARY OF THE THESIS

Machine learning-based modelling of anaerobic digestion stability and productivity optimization by co-digestion

TESI MAGISTRALE IN CHEMICAL ENGINEERING – INGEGNERIA CHIMICA

AUTHOR: FILIPPO ALLIATA

ADVISOR: GIULIA LUISA BOZZANO

ACADEMIC YEAR: 2023-2024

1. Introduction

The anaerobic digestion is a biological process that allows to produce biogas through the degradation of organic substrates. This operation is performed by a specific bacterial community in absence of oxygen. Indeed, starting from the macromolecules that make up the biomass, these microorganisms can generate biomethane through a series of biochemical reactions. Nowadays, it is preferable a co-digestion with respect to a simple anaerobic digestion, i.e. the digestion of a single substrate, that is a reliable technology but, at the same time, it is affected by some limitations, as an inadequate nutrients balance, a lower biomethane yield and an inefficient dilution of toxic compounds. These are connected to the bio-chemical properties of the single biomass; in fact, they exert a significant influence over the performance of the reactor. The co-digestion, on the contrary, can overcome the previous drawbacks thanks to the combination of the features of the substrates. This combination permits to reach a better performance because it promotes the formation of important synergy effects within the anaerobic digester. Indeed,

through the interactions between the macromolecules of the biomass and the bacterial community of the system, it occurs the generation of synergy or antagonism effects. It is a prevalence of the former that allows the achievement of an optimized process, while the latter constitute one of the main causes leading to the inhibition of the system. The aim of this dissertation is strictly connected to these two typologies of effects. In fact, to perform an optimal co-digestion, it is fundamental to reach a complete knowledge about the real dependence of the anaerobic digestion on the synergy and antagonism effects formed within the reactor. In such a way as to achieve this goal, this work is characterized by two steps: the first consists of carrying out detailed analysis of two stability parameters, *FOS/TAC* and *OLR*. The second is represented by the creation of a machine learning.

2. State of the art

Depending on the stability parameters values, it is feasible to verify if the co-digestion is defined by a stable functioning. Two common indicators used to monitor the process performances are the *C/N* ratio and the biomass biodegradability *BD*. The

former depends on carbon and nitrogen percentage of the mixture, and its stability range considers all those values between 20 and 40. The latter is defined as the property of a biomass to be biologically degraded through the action of microorganisms in absence of oxygen. This parameter assumes values between 0 and 1. Therefore, according to their behaviors, it is possible to monitor the potential presence of synergy and antagonism effects that are forming inside the anaerobic digester. [1]

3. FOS/TAC

FOS/TAC is used to monitor process stability by focusing on the pH value. It is defined as the ratio of the amount of volatile fatty acids that accumulate during anaerobic digestion (FOS) to the alkalinity present in the system (TAC). Since volatile fatty acids accumulate during the process, they can lead to a consumption of alkalinity and consequently a decrease in pH. For this reason, it is of fundamental importance to have a buffer solution that can counteract the pH decrease bringing it back to neutral values, as the optimal pH value for an anaerobic digester is around 7. Representing the BMP behavior with respect to the *FOS/TAC*, it has been obtained the following trend:

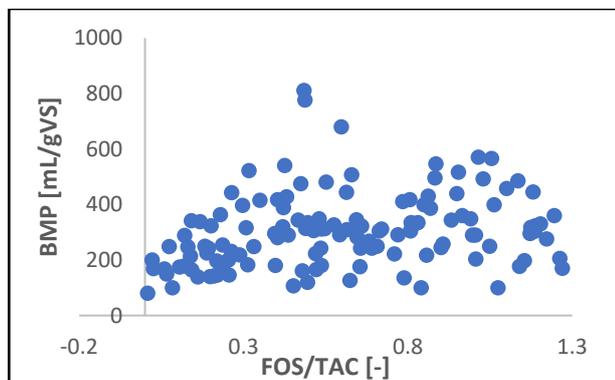


Figure 1: FOS&TAC calculation results with respect to BMP

It can be observed that the biomethane yield increases until reaching a maximum around the FOS/TAC value of 0.5, then it starts to decrease. Regarding its stability range, it goes from 0.2 to 0.7. This means that, for values greater than 0.7, the buffer solution’s alkalinity is unable to maintain the system’s pH on neutrality due to excessive production of volatile fatty acids; for values lower than 0.2, the digester can support a higher organic volumetric load [2].

Considering the plot reported before, it has been demonstrated the existence of iso-regions, that can vary as a function of the substrate used as the reference one. Thanks to these regions, it is possible to predict the formation of synergy and antagonism effects with respect to the substrates used to perform the co-digestion:

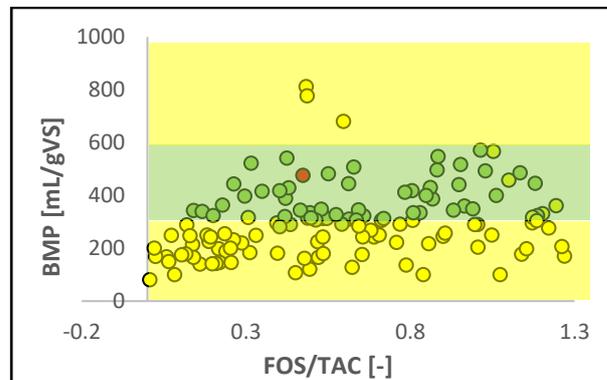


Figure 2: FOS/TAC with highlighted regions of synergies (green dots) and regions of no synergies (yellow dots).

The reference substrate is highlighted in red and substrates that showed synergy with it were colored green, while those without synergy were colored yellow. Indeed, it turns out that the achievement of significant synergistic effects occurs by considering substrates that have sufficiently similar predicted *FOS/TAC* values to each other, thus being within the same iso-regions. This is valid within a specific influence zone (the red area in Figure 3), i.e., for FOS/TAC values less than 1.0. The reasons connected to this choice are the following two: after a FOS/TAC close to 1.1, this parameter tends to stabilize meaning that probably it has a less significant influence on the biomethane yield. The other reason is that it makes no sense to take into consideration a substrate with a too high volatile fatty acids content.

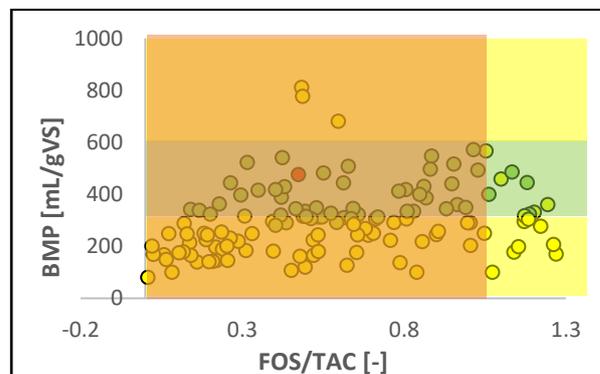


Figure 3: Iso-regions detection in FOS/TAC analysis with the influence zone

It is also important not to consider a feeding composed by substrates that are both affected by a low or a high value of this indicator. In fact, these combinations can lead to the achievement of a *FOS/TAC* of the mixture that is outside the acceptable range.

4. OLR

OLR (organic loading rate) represents the amount of volatile solids that must be fed to the reactor every day, in which the volatile solids are the portion of organic matter inside the substrate. OLR can be calculated through the following equation:

$$OLR = \frac{VS}{HRT} \quad (1)$$

Where VS [kg/m³] is the concentration of volatile solids in the feeding and HRT is the hydraulic retention time [day], so OLR is obtained in terms of [kgvs/m³/day].

Since this parameter depends on several factors such as operating conditions and reactor configuration, it has been more difficult to identify its specific value that can cause the inhibition of the process. Exploiting what emerges from the literature, it makes sense to consider 6 [kgvs/m³/day] as that organic loading rate value beyond which anaerobic digestion becomes unstable. A too high value of OLR leads to a critical accumulation of VFAs, because it causes a significant increase of the production rate of hydrolysis and acidogenesis, that are those steps in which occurs the production of volatile fatty acids. Analyzing the dependency of the bacterial community with respect to the OLR, it is important to point out that: for its low values, it occurs the death of methanogens because they are in a starving condition, since their required nutrients are much higher than those fed. On the other hand, for high values of OLR, due to the critical VFAs accumulation mentioned before, it happens the inhibition of microorganisms. Indeed, they require a pH close to the neutrality, but the accumulation of volatile fatty acids causes a drastic reduction of the alkalinity level leading so to the inhibition of the bacterial community [3].

Considering the aspects described formerly, it has been developed an accurate model that is able to predict the variation of the pH level of the system as a function of the amount of volatile solids fed to the reactor:

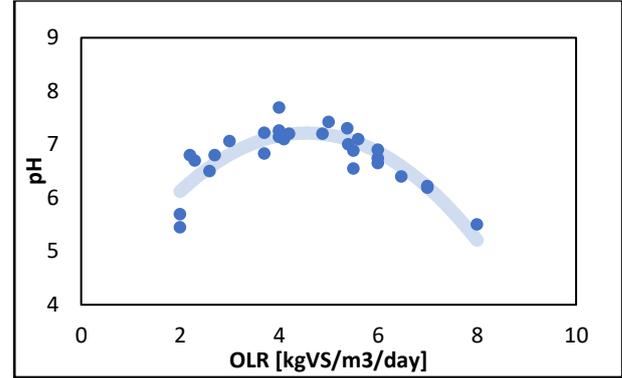


Figure 4: Correlation between pH and OLR.

$$pH = -0.1683 \cdot OLR^2 + 1.5303 \cdot OLR + 3.7355 \quad (2)$$

This model confirms those stated previously; in fact, for an OLR between 2.5 and 5.5 [kgvs/m³/day], it's possible to reach the optimal pH range for microorganisms growth. On the other hand, getting an OLR equal to 6 [kgvs/m³/day], it can be observed that the pH tends to decrease until causing the achievement of an acid environment for high OLR like 8 [kgvs/m³/day].

5. Machine learning

The machine learning of interest identifies the specific conjugated substrate that, as a function of a predefined biomass, can maximize the biomethane yield through a co-digestion process. This machine learning exploits two tools: the first one is a neural network, also known as ANN (artificial neural network) or SNN (simulated neural network), while the second one is an algorithm, named as "FBO2", that allows to find the optimal composition of the mixture that must be fed to the reactor.

5.1 Neural Network

This neural network is subjected to a classification problem. This means that, feeding a sample to this tool, it is possible to identify the class at which belongs. It can be done after a step of training in which it is used a dataset containing two types of information: the values of the features that characterize the single sample of training, and the class associated with it. After this step, the ANN develops a significant learning capability thanks to which it can classify the data that will be given with a good reliability. In this dissertation, it has been

performed a multiclass classification problem characterized by the following labels: zootechnical waste, agricultural waste, organic waste, and sludge waste. In the dataset of interest, each sample is defined by 19 features, that are the properties related to the substrate. Regarding its architecture, it has been selected a “three layers neural network”, so two hidden layers, 9 artificial neurons for the first and 8 for the second, and the output layer, with 4 units like the four classes that characterize the problem of interest. This ANN can be used for the different purposes; depending on its role played inside the final machine learning, the one of interest is the possibility to generate accurate predictions by means of only one features with respect to the 19 with which the SNN has been trained. This means that, providing a sample characterized by a single property such as the C/N , the ANN can predict the class at which belongs. An example is shown in the following table:

C/N	Pr.cl1	Pr.cl2	Pr.cl3	Pr.cl4
10	0.255	0.235	0.46	0.05

Table 1: belonging probability values to the four classes as a function of the single C/N parameter. Pr.cl1 indicates the probability that a biomass, in this case with a C/N equal to 10, has to belong to the first class, i.e. zootechnical waste. Regarding the other classes: cl2 = agricultural waste, cl3 = organic waste, cl4 = sludge waste

In the table reported before, there are the belonging probabilities to the four classes for a C/N equal to 10, so a low value of this indicator. These results are quite coherent if compared to the properties that characterize the different substrates within each of these classes; for example, the class 2 has a lower probability with respect to the one of class3 since it is composed by wastes with a high percentage of lignin, such as straw. This implies a high content of C leading to a high C/N value.

5.2 FBO2

FBO stands for “feedstock blending optimization”; in fact, as a function of a feeding of two biomasses, this algorithm can identify the best mixture composition to maximize the biomethane yield. It is important to highlight that the achievement of this result occurs by satisfying the stability conditions imposed by the four indicators: C/N , BD , FOS/TAC and OLR . Indeed, before finding the

optimal mixture, there are four filtration operations; each is associated to one of the parameters mentioned before. Thanks to the conservation principle, it is feasible to work by exploiting only the composition of the first substrate. Therefore, the starting point is represented by the definition of the vector x_0 that contains all the possible compositions related to the first biomass:

$$x_0 = [0, 0.01, 0.02, \dots, 1] \quad (3)$$

At this point, it occurs the first filtration operation, that is the one related to the C/N . It consists of calculating the C/N of the mixture that is expressed in this way:

$$C/N_{mix} = C/N_1 * w_1 + C/N_2 * (1 - w_1) \quad (4)$$

In which C/N_1 and C/N_2 are the C/N of the first and second biomass respectively. It occurs the calculation of the different C/N_{mix} associated to all the w_1 contained within x_0 . Now it takes place the real filtration operation. Indeed, considering the stability range of C/N_{mix} , it occurs the creation of a new composition vector, named as x_1 , that is composed of all those w_1 that allow the achievement of a C/N_{mix} within its stability range. Once this x_1 is obtained, it is subjected to the next filtration operation. It’s characterized by the same procedure described before exploiting its specific mixture equation and the associated stability range. The last step of this filtrations process is defined by the achievement of x_4 . It is the vector within which there are all the w_1 that were able to respect all the stability conditions fixed by the four parameters mentioned previously. At this point, by means of the w_1 contained in x_4 , it occurs the identification of the optimal mixture composition. It will be the one that permits to reach the maximum value of BMP.

These operations allow to consider in the end only those compositions that bring the mixture stability indicators within their acceptable ranges. This aspect points out the fact that, this algorithm has been created not only with the aim of maximizing the biomethane yield, but also by considering the possible synergy and antagonism effects formed within the reactor. The former are favored by the respect of the stability conditions.

In such a way as to analyze the nature of the results that the FBO2 can return, below it is considered the

co-digestion of fruit waste and chicken manure. These two biomasses are simply chosen to have a practical example through which it is possible to evaluate the potentialities of the model:

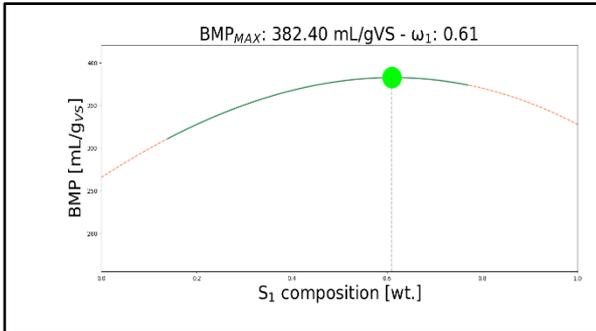


Figure 5: BMP vs w_1 plot associated to the co-digestion of fruit waste and chicken manure

$w_{1,max}$	$w_{2,max}$	BMPmax [mL/gVS]
0.61	0.39	382.40

Table 2: values of the optimal composition and the associated BMP for the co-digestion of fruit waste and chicken manure

In the figure 5, it is represented the behavior of the biomethane yield as a function of the composition of the first substrate, so fruit waste. The green arch refers to the BMP obtained through those w_i that satisfy all the four stability conditions, while those dashed in red are related to the biomethane yield calculated with the compositions that have failed the different filtration operations. The green dot denotes the maximum value of the BMP. The table 2 shows the optimal mixture composition that allows to maximize the biomethane yield by satisfying all the stability conditions. Therefore, the mixture characterized by 0.61 of fruit waste and 0.39 of chicken manure permits to achieve the maximum value of BMP that for this co-digestion is equal to 382.40 [mL/gVS].

5.3 Conjugated substrate

Once have been created the previous two tools, it is possible to move to the development of a machine learning to identify the correct conjugated substrate that, depending on a predefined biomass, is able to optimize the performance of the process. The starting point of this research is represented by the knowledge of only the properties of the predefined biomass in terms of: C/N , BD , $FOS/$

TAC , BMP , the total solids and the volatile solids in the biomass.

The first important result obtained from this model is represented by the class at which probably belongs the conjugated biomass based on C/N and BD indicators. In fact, thanks to the exploitation of specific instructions implemented on the ANN, it is possible to find the class that can maximize the BMP through a co-digestion with the predefined biomass. In addition, it is also extracted the associated mixture composition and mixture parameters values.

Once the belonging class is known, it occurs the identification of the conjugated substrate by selecting the biomass that allows to achieve the more similar mixture parameters with respect to the ones previously predicted by the ANN. Thereby, it has been possible to obtain the most important result associated to this machine learning, i.e. the estimation of the conjugated substrate by means of only the knowledge of the features of the predefined biomass.

At this point, it is necessary to define the ideal composition of the mixture to feed to the reactor. For this reason, it is used the FBO2 through which is possible to achieve the optimal compositions of the predefined biomass and of the conjugated substrate to maximize the BMP. In conclusion, starting from the previous mixture composition, an additional obtainable result from this model is the CHONS of the mixture. Indeed, thanks to a canonical correlation analysis performed in collaboration with the CNR (Consiglio Nazionale delle Ricerche) of Naples, it is possible to predict the percentage of carbon, hydrogen, oxygen, nitrogen, and sulfur of the mixture [4].

Considering as predefined biomass the sheep manure, following are reported the results obtained through this machine learning:

Class	Conjugated substrate
Organic waste	Fruit waste

Table 3: results provided by the neural network having the sheep manure as predefined biomass

$w_{1,max}$	$w_{2,max}$	BMPmax [mL/gVS]
0.34	0.66	370.26

Table 4: results provided by the FBO2 for the co-digestion of sheep manure and fruit waste

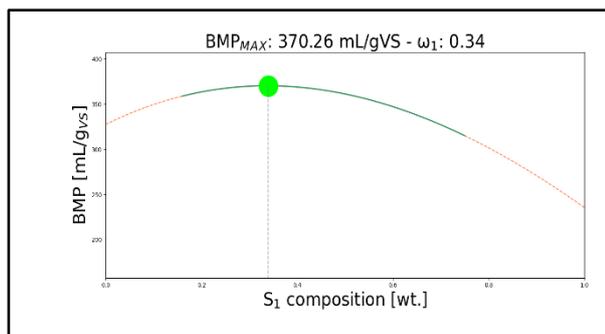


Figure 6: BMP vs w_1 plot associated to the co-digestion of sheep manure and fruit waste.

C	H	O	N	S
40.566	6.057	48.542	3.654	1.181

Table 5: elemental analysis of the mixture: sheep manure-fruit waste

Table 3 is composed of the results provided by the neural network: the class at which belongs the conjugated substrate is the organic waste, while the conjugated substrate identified by this simulation is the fruit waste. If observed by the macromolecules point of view, this solution can be considered as reasonable since it is possible to reach a good nutrients balance. Indeed, the sheep manure is much richer in terms of nitrogen, while the fruit waste is characterized by a significant level of sugars. Table 4 is composed of the results provided by the FBO2. Therefore, the mixture characterized by 0.34 of sheep manure and 0.66 of fruit waste permits to achieve the maximum value of BMP that for this co-digestion is equal to 370.26 [mL/gVS]. In conclusion, Table 5 reports the elemental analysis of the optimal mixture.

The production plant of biogas often does not have a great degree of freedom with respect to the choice of the wastes to be used in the co-digestion process; in fact, it depends on what comes from the surrounding farms. For this reason, it is possible to be in the situation of having to choose, among all the available substrates, the best one to combine with a predefined biomass. Therefore, the final choice can be achieved by running the ANN and subsequently the FBO2 for each available substrate. Then, based on considerations related to the FBO2 results and to the conjugated substrate provided by the algorithm, it is possible to identify the correct biomass for the co-digestion. The final choice depends on the properties that characterize the conjugated substrate predicted by the neural network, such as the macromolecules content,

using in the end a different biomass but with similar features.

6. Conclusions

Thanks to this work, it has been possible to obtain significant results regarding the role played by the synergy and antagonism effects during the anaerobic digestion but, at the same time, there are still several weaknesses to be solved to reach a complete understanding of this topic.

For future researchers, it could be very important to define a stability parameter connected to the rheological features of the system because, as a function of what emerges from different analysis, they are able to exert a significant influence on the possibility to avoid the process inhibition. At the same time, it is necessary to try with the bacterial community, because it plays a fundamental role during the process. Therefore, it could be very significant to define a stability parameter connected to the microorganisms.

References

- [1] F. Moretta, A. Goracci, F. Manenti, e G. Bozzano, «Data-driven model for feedstock blending optimization of anaerobic co-digestion by BMP maximization.», *Journal of Cleaner Production*, vol. 375, 2022.
- [2] Francesco Di maria, Alessio Sordi, Giuseppe Cirulli, Giovanni Gigliotti, Luisa Massaccesi, e Mirko Cucina, «Co-treatment of fruit and vegetable waste in sludge digesters. An analysis of the relationship among bio-methane generation, process stability and digestate phytotoxicity», *Waste Management*, vol. 34, pp. 1603–1608, 2014.
- [3] Qin Yu, Lei Feng, e Xiaofei Zhen, «Effects of organic loading rate and temperature fluctuation on the microbial community and performance of anaerobic digestion of food waste», *Environmental Science and Pollution Research*, vol. 28, pp. 13176–13187, 2021.
- [4] F. Moretta, V. Del Duca, G. Bozzano, A. Coppola, e F. Scala, «Canonical Correlation Analysis to Biomass CHONS Prediction», *Chemical Engineering Transactions*, vol. 105, 2023.