POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

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

TESI DI LAUREA MAGISTRALE IN
CHEMICAL ENGINEERING

Author: **Filippo Alliata**

Student ID: 10580870
Advisor: Giulia Luisa Bozzano
Co-advisor: Federico Moretta
Academic Year: 2023-24

Abstract

Anaerobic digestion is a reliable technology that, in absence of oxygen and through the action of a specific bacterial community, allows to produce biogas starting from biomasses of different nature. The performances of this process are strongly affected by synergy and antagonism effects formed within the reactor. Their generation depends on the interactions established between the macromolecules of the biomasses and the microorganisms of the system. Because of their importance, the aim of this dissertation is to reach a greater knowledge about the dependence of the anaerobic digestion with respect to these two effects. In such a way as to achieve the previous goal, this work is characterized by two steps: the first one consists of carrying out detailed analysis of two stability parameters to understand their real influence on the performances of anaerobic digestion. The first indicator is the FOS/TAC, that permits to monitor the process stability by focusing on the alkalinity level of the system. The second parameter is the OLR, that defines the amount of volatile solids that must be fed to the reactor every day. Regarding the second step, this dissertation is characterized by the creation of a machine learning to identify 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, while the second one is an algorithm that allows to find the optimal composition of the mixture that must be fed to the reactor. 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. However, there are still several weaknesses to be solved to reach a complete understanding of this topic.

Key-words: co-digestion, synergy effects, antagonism effects, stability parameters, FOS/TAC, OLR, neural network, conjugated substrate.

Abstract in italiano

La digestione anaerobica è una tecnologia affidabile che, in assenza di ossigeno e attraverso l'intervento di una specifica comunità batterica, permette di produrre biogas partendo da biomasse di diversa natura. Le performance di questo processo sono fortemente influenzate dagli effetti di sinergia e antagonismo che si formano all'interno del reattore. La loro generazione dipende dalle interazioni che si instaurano tra le macromolecole delle biomasse e i microorganismi dell'ambiente di reazione. A causa della loro importanza, l'obbiettivo di questa tesi è quello di acquisire una migliore conoscenza riguardo alla dipendenza della digestione anaerobica rispetto a questi due effetti. In maniera tale da poter raggiungere il precedente obbiettivo, questo lavoro è caratterizzato da due step: il primo consiste nell'effettuare analisi dettagliate di due parametri di stabilità per poter comprendere la loro reale influenza sulle performance della digestione anaerobica. Il primo indicatore è il FOS/TAC, che permette di monitorare la stabilità del processo focalizzandosi sul livello di alcalinità del sistema. Il secondo parametro è l'OLR, che quantifica la quantità di solidi volatili che devono essere alimentati al reattore ogni giorno. Per quanto riguarda il secondo step, questa tesi è caratterizzata dalla creazione di una machine learning per poter identificare il corretto substrato coniugato che, in funzione di una biomassa predefinita, è in grado di massimizzare la resa di biometano attraverso un processo di co-digestione. Questa machine learning sfrutta due strumenti: il primo è una rete neurale, mentre il secondo è un algoritmo che permette di trovare la composizione ottimale della miscela che deve essere alimentata al reattore. Grazie a questo lavoro, è stato possibile ottenere risultati significativi per quanto riguarda il ruolo svolto dagli effetti di sinergia e antagonismo durante la digestione anaerobica. Tuttavia, ci sono ancora diverse criticità da risolvere per poter raggiungere una completa comprensione di questo argomento.

Parole chiave: co-digestione, effetti di sinergia, effetti di antagonismo, parametri di stabilità, FOS/TAC, OLR, rete neurale, substrato coniugato.

Contents

1 Introduction

1.1. Anaerobic digestion

Nowadays the society is distinguished by two problems that are strictly connected: energy requirements and waste disposal. Combining these two aspects with a significant demographic increase, there is the risk to cause an irreversible scenario for the planet. Indeed, with the passing of the years, it is increasingly necessary to find a correct strategy to valorise these wastes reducing their environmental impact. Furthermore, regarding environmental preservation, it is fundamental to increase the ability to produce renewable energy by exploiting alternative sources in a more efficient way. These sources must be accumulable and available in large amounts but, at the same time, they must be able to be used through economic and simple technologies. For the reasons explained previously, a sector that is playing a crucial role is the one of the anaerobic digestion that produce biogas through the degradation of waste biomasses. This process is very important because it allows to reach the two goals mentioned above, i.e. obtain sustainable energy by a reliable technology for the wastes' valorisation. Among the biomasses of interest, apart from FORSU and agricultural wastes, it is significant to point out the zootechnical wastes that are affected by some issues. Indeed, they are used to fertilize the fields but, at the same time, it is not possible to use them in a large amount to avoid an excessive release of nitrates, that are contained within them, in the environment. For this reason, the farms can employ only a specific amount of zootechnical wastes and only in specific times of the year by law. Thereby, there is a surplus of wastes that remains unused, and it represents a significant cost for small and medium corporations. However, nowadays there is a solution to this problem, i.e. to use these residues in biogas production plants. This is a classic example of circular economy; in fact, it is possible to convert wastes of different nature into more resources, such as biomethane and compost.

The anaerobic digestion falls within those processes that permit to obtain the so-called bio energies, that are a specific category of renewable energy obtained starting from biomasses. They are assuming a meaningful importance in Italy and this can be observed in the figure 1. It is obtained by exploiting the data provided by Terna on their web site:

In the plot reported above, it is possible to highlight the following aspects:
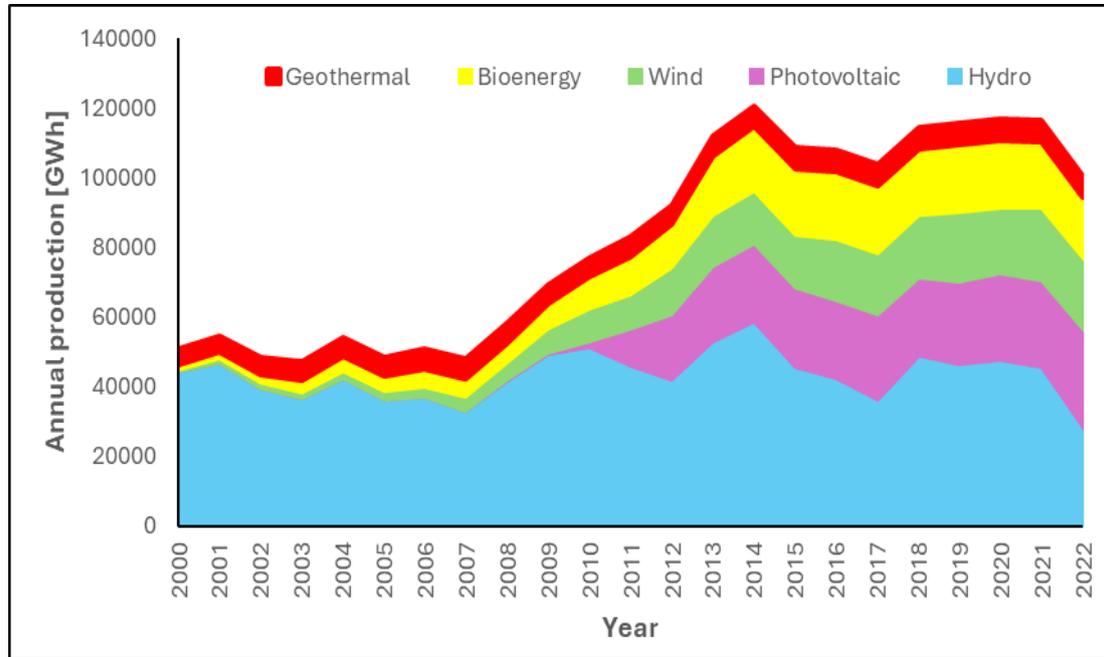

Figure 1: Annual production of renewable energies.

- The hydro is always the more used alternative source to obtain renewable energy;

- In the last decays, the bio energies production is significantly increased, and it is destined to play a very important role in the future. Indeed, it is useful to point out the biomethane contribution within the energy consumption in Italy. The existent plants allow to produce 773'000'000 Sm^3 of biomethane. Considering that, for example, in the 2021 in Italy it has been consumed a total amount of gas equal to 76'000'000'000 Sm^3, it is possible to argue that the produced biomethane is relatively small if compared to the total energy consumption; also because a fraction of biomethane is used to self-feeding the plant. On the other hand, it is important to highlight that, with the help of considerable investments, in Italy there is the aim to reach a production of biomethane of 8'000'000'000 Sm^3 by 2030. This purpose is ambitious because it could represent the 10% of the Italian gas requirement. These considerations need to bring the attention to the fact that, effectively, the bio energies are destined to play a very important role also in the future.

Going back to the anaerobic digestion, it represents a biological process that degrades and transforms the organic substance into biogas through the involvement of complex bacterial communities and in absence of oxygen. The

biogas is a mixture of gas composed by the 45% - 70% of methane and the remaining part by CO_2, neglecting other components that are present in lower amounts. It is possible to observe the fundamental steps of a biogas production plant through the simplified illustration shown in figure 2:

Figure 2: Simplified illustration of a biogas production plant.

Observing the figure reported above, the steps that characterize the process are as follows:

- Once that the biomasses are transported into the plant of interest thanks to tractors and trucks, they are subjected to some pre-treatments in order to favor the digestion. According to the nature of the substrate, there are different typologies of pre-treatments: for example, the mechanical one in which it is realized a trituration of the biomass, another one can be the separation of organic matter from all that is not organic, like the plastic;
- After these pre-treatments, the substrates are fed to the digester within which happens the anaerobic digestion. The anaerobic digestion consists of four phases characterized by specific typologies of microorganisms. These phases are strictly connected to each other since a single step uses the degraded molecules coming from the previous one. The production of biomethane occurs through the achievement of products increasingly less complex. The four steps of interest are reported below:

1. Hydrolysis: the macromolecules fed to the reactor, such as proteins and polysaccharides, are converted to simpler units, i.e. monosaccharides, fatty acids and amino acids. It is a very slow process; indeed, it is considered as the limiting step of the anaerobic digestion. [1];

2. Acidogenesis: the molecule produced by hydrolysis are transformed at first into pyruvate and subsequently converted into butyric acid, propionic acid, acetic acid, alcohols, aldehydes, and gas, such as NH_3, CO_2 and H_2;

3. Acetogenesis: the substances of the previous step are converted into H_2, CO_2, acetic acid and formic acid. As a function of the fatty acids that are used for the degradation, the conversion can happen through different mechanisms [2]. In fact, there are the long chain fatty acids, with more than 5 atoms of carbon, or the short chain fatty acids.

4. Methanogenesis: thanks to the acetogenesis, it has been possible to obtain all those molecules thar are required to produce methane through two possible pathways, that will be analyzed in the chapter three. The aspect that is important to remember for the time being is that, through this step, it occurs the production of CH_4 and CO_2. The former will pass only in the gas phase because it has a low solubility in water (0.024 g/L in standard condition [3]). The latter will tend to split itself between gas phase and liquid phase since it has a higher solubility (1.688 g/L in standard condition [3]). In the liquid phase, the CO_2 will be present in the form of bicarbonate ion HCO_3^-.

- After the methanogenesis, the anaerobic digestion can be considered as completed. The most important aspect to highlight is that, through this process, it is possible to obtain two useful resources, i.e. biogas and digestate. They will have different uses:

   1. Regarding the digestate, it represents the residual part of the anaerobic digestion once the biomasses have released methane and carbon dioxide. Thanks to a solid-liquid separation, it is obtained one of the final products that is the compost; it is the solid part of the digestate. At this point, it can be exploited as fertilizer;

2. Regarding the biogas, it can be fed to a cogeneration process in order to produce electric and thermal energy. Another option is to submit the biogas to the upgrading, that permits to recover biomethane in the form of CNG (Compressed Natural Gas) or LNG (Liquefied Natural Gas), that represents a valid fuel for trucks and ships.

After this general introduction related to the importance and the properties of anaerobic digestion, at this point, it is necessary to define the goal of this dissertation highlighting the starting considerations and results extrapolated from the literature.

1.2. State of the art

One of the goals of this dissertation is to define, with a specific reliability, two parameters thar are able to guarantee a stable functioning of the anaerobic digestor to optimize its performance. For this reason, connected with the achievement of a more efficient process, it is important to point out the motivations that permit to explain why, nowadays, it is preferred the co-digestion, AcoD, with respect to the simple digestion, AD. The first represents the digestion of a mixture of two or more substrates, while the second refers to all the plants that are projected in order to manage a single biomass. It is correct to consider the simple anerobic digestion as a reliable technology to produce renewable energy through the wastes valorization but, at the same time, it is affected by some limits. These limits 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 drawbacks associated to the AD thanks to the combination of the features of the different substrates. Indeed, the advantages that is possible to reach through the co-digestion are as follows ([4], [5], [6], [7], [8], [9], [10], [11]):

- The main benefit is the one represented by the possibility to reach a higher biomethane yield. In fact, the co-digestion can guarantee a production of biogas greater than 25% up to 400% with respect to the one obtainable with the digestion of only one substrate ([9], [12]);
- It permits to achieve a better nutrients balance. The single biomass is affected by the presence of specific macromolecules such as proteins, lipids, sugars whose concentrations can vary. Therefore, it is advisable to adopt substrates with different macromolecules concentrations in order to combine their effects ensuring the achievement of better synergies. As a function of the predominant

macromolecule, the biomass of interest will cause specific effects inside the anaerobic digester, indeed:

1. High concentration of carbohydrates: these substrates, such as fruit and vegetable wastes and food wastes, are affected by a huge concentration of sugars whose degradation leads to the formation of a critical amount of volatile fatty acids. Obviously, a high concentration of VFA is undesirable because it can cause a dramatic reduction of pH system. This leads to the microorganisms deactivation of the methanogenesis step stopping the biogas production [13];

2. High concentration of proteins: this feature is distinctive of the manures. These substrates, also alone, are able to produce an high amount of methane within the biogas ([14], [15]). The problem is associated to the huge amount of N within them. Indeed, subsequently their degradation, there is a significant release of ammonia ions whose critical content can inhibit the microorganisms growth [13];

3. High concentration of fats: these are the substrates that, for example, came from the dairy products industry. They are easily degradable but, on the other hand, a high amount of lipids can cause a malfunction of the process. In fact, the degradation of triglycerides leads to the formation of long chain fatty acids (LCFA) and glycerol. The real problem is associated with a high concentration of LCFA. Indeed, their conversion results to be very complex because within them there is a lot of carbon atoms and this can lead to a strong inhibition of microorganisms [13];

- It permits to improve the dilution of toxic compounds such as the heavy metals [16];
- It allows to reach a better process stability and biogas production thanks to the achievement of a great synergy between the different microorganisms that carry out the anaerobic digestion. Indeed, within each step of this biological process, there are various groups of microorganisms and each of them with its own specific features in terms of nutritional needs, operating conditions, and kinetic growth. These aspects make arduous the goal of finding those operating conditions that favor the survival of all the species involved in the process. In addition, the purpose to choose the correct biomasses that guarantee the

formation of good synergistic effects is very challenging. Indeed, it is necessary to monitor the interaction between the nutrients and the microorganisms. This topic will be better deepened in the chapter three.

The co-digestion is also affected by some drawbacks; Indeed, combining in a not very efficient way two or more substrates, it is possible to cause a critical acidification of the system or an organic overloading [17]. But, because of the advantages mentioned above, it is undoubted that the co-digestion is more worthwhile with respect to a digestion of a single substrate. What emerges is that, during the organic process, the environment inside the reactor is complex to control because of the several factors that exert a specific influence on it. Indeed, beyond the roles played by the different microorganisms and nutrients, it is fundamental to consider the dependency of the process performance with respect to the operating temperature and the behavior of pH value. In fact:

- The anaerobic digestion, typically, can operate in two different regimes as a function of the adopted range of operating temperature. Each of them is affected by advantages and drawbacks [13]:

    1. Mesophilic regime: the T is around a value close to $35\ °C$. In this case, the process is more stable and it is affected by a higher microbial diversity but, on the other hand, the microbial growth is slow;
    2. Thermophilic regime: the T is around a value close to $55\ °C$. It's true that in this case the digestion process is faster, but it presents different disadvantages. Indeed, it is more difficult to control, it is more unstable in presence of a high ammonia concentration, and it requires a higher amount of energy. For these reasons, typically, the mesophilic regime is the favorite one.

- Regarding the behavior of pH value, its influence on process performance is much more complex. The four phases of anaerobic digestion and their microorganisms are characterized by different optimal pH values. Indeed, it is required an alkalinity level between 5.5 and 6.5 for the hydrolysis and acidogenesis, while the methanogenesis requires a pH close to 7. Therefore, in conclusion, it is recommended that, in order to guarantee a good performance of the process, the pH should stay in a range of alkalinity level between 6.8 and 7.2 during the anaerobic digestion [13].

Thereby, because of what has been said until now, it is obvious that the anaerobic digestion is complex to control not only to optimize its performance, but also to avoid a possible inhibition of the process. In such a way as to reach the previous two goals, a crucial factor is defined by the presence of effects of synergism or antagonism formed during the co-digestion; in fact, it will be a prevalence of the first that allow the achievement of an optimized process. In order to monitor these effects, it is possible to exploit specific parameters that are calculated considering the properties of the single biomass. As a function of their value, it is feasible to verify if the co-digestion is defined by a stable function thanks to a correct choice of the substates. This choice has allowed the formation of significant synergistic effects between the several nutrients and microorganisms.

This dissertation starts from a paper [18] within which there is also the analysis of the stability of anaerobic digestion as a function of two indicators:

- The first parameter is represented by the C/N. Therefore, in this case, it is important the composition of the substrate in terms of C, H, O, N, S. In the study of interest, it has been demonstrated that the C/N should stay in the range of values between 20 and 40 to guarantee good performance of the process. There are other papers that show a stable functioning of the reactor also for C/N values close to 10-15 ([19], [20], [21]). For this reason, in this dissertation, it has been considered a more extended C/N stability range, so the lower boundary is shifted towards a value of 10 while the upper boundary remains 40. This means that, for C/N higher than 40 there is a critical level of volatile fatty acids that can cause the bacteria deactivation, while for C/N lower than 10, there is a too high ammonia concentration that can inhibit the microorganisms growth;

- The second parameter is represented by the biodegradability BD. The biodegradability is defined as the property of a biomass to be biologically degraded through the action of microorganisms. The BD of a substrate is influenced by several chemical and physical factors such as the temperature, the moisture, and the pH [13]. In the paper of interest, the biodegradability is defined through the equation 1.1. This parameter assumes values between 0 and 1. In fact, the numerator indicates the real biomethane yield (EBMP expressed in terms of [mL/gVS]), while the denominator is the theorical one calculated through the modified relation of Buswell that does not consider the non-degradable fraction of the biomass. Indeed, defining the hypothetical biomass composition with the following chemical formula: $C_n H_a O_b N_c S_d$, the TBMP can be calculated through equation 1.2:

$$BD = \frac{EBMP}{TBMP} \quad [\text{-}] \tag{1.1}$$

$$TBMP = \frac{\left(\frac{n}{2} + \frac{a}{8} - \frac{b}{4} - \frac{3c}{8} - \frac{d}{4}\right)*22415}{12n+a+16b+14c+32d} \quad [\text{mL/gVS}] \tag{1.2}$$

1.3. Objective of the dissertation

This dissertation is characterized by two goals:

1. To identify two new parameters that are able to guarantee a stable performance of the digestor as a function of the biomasses fed to the reactor. If possible, it could be better to find new indicators that are connected to different substrate properties with respect to those related to C/N and biodegradability. This permits to reach a better control on the global process. Indeed, through detailed analyses, it will be possible to obtain a greater reliability regarding the real influence of two stability parameters on a process that is complicated to monitor;

2. To create an algorithm that is characterized by a specific purpose. Its purpose is to find a conjugated substrate that, as a function of a well-known biomass fed to the reactor, is able to maximize the yield of biomethane thanks to the synergistic effects created by the interactions of the two substrates. This goal can be achieved through the creations of two tools, in which one is a neural network, while the other is a particular algorithm called as FBO2. FBO stands for "feedstock blending optimization", while the 2 is because it started from an old model based only on the stability parameters of the C/N and of the biodegradability. The features of these tools will be better described and analyzed respectively in chapter 4 and in chapter 5. The fundamental aspect that is important to point out for the time being is that, through the achievement of this second goal, it is possible to foresee the biomass that can create the best synergistic effects with a given substrate. This is a very tough challenge, because it is difficult to find clear correlations between the properties of different biomasses inside a complex environment such as the one of the anaerobic digestor.

1.4. Available database

Before starting with the analysis performed in this dissertation, it is important to point out the nature of the data that are used in this work to reach the results. It has been created a database containing information related to substrates of different nature such as: zootechnical wastes, agricultural wastes, organic wastes and sludge wastes. These data have been extracted from different papers in which each contains the performance associated with a co-digestion. These analysis permit to highlight the features of the biomass in terms of: C, H, O, N, S composition, macromolecules concentration, biodegradability, C/N and EBMP. It was possible that, for a substrate analyzed in a specific paper, there were some missing data. Then, through regression analysis performed over the available data of other biomasses, it has been possible to recover the different lacks with reliable results.

2 FOS/TAC

2.1. Meaning and achievement of iso-regions

The first goal of this dissertation is to define two new parameters that can guarantee a stable operation of the anaerobic digestion as well as the known C/N and biodegradability indicators. The first analyzed parameter is the one that 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), i.e. is the FOS/TAC.

This parameter is used to monitor process stability by focusing on the pH value. Since volatile fatty acids accumulate during the acidogenesis phase, 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, as said before, the optimal pH value for an anaerobic digester is around 7.

From the literature, it is found that the optimal range for FOS/TAC to maintain stable operation of the anaerobic digester is between 0.3 and 0.4 [22], [23], [24].

In fact, for values greater than 0.4, 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 0.3, the digester can support a higher organic volumetric load [25].

At this point, to estimate this new parameter, it is important to point out how the FOS and TAC terms can be calculated starting from the properties of the single biomass.

Alkalinity (TAC) is mainly characterized by the coexistence of ammonia, generated by protein degradation, and the dissolution of CO_2 in the liquid promoting the formation of bicarbonate. The resulting buffer system is called the calcium-acetate system leading to the formation of sodium bicarbonate, which is a salt whose dissolution ensures high alkalinity, thus counteracting the decrease in pH caused by the accumulation of volatile fatty acids [26]:

$$CO_2 + H_2O \leftrightarrow HCO_3^- + H^+ \tag{2.1}$$

$$HCO_3^- + NH_4^+ \leftrightarrow NH_4HCO_3 \tag{2.2}$$

Therefore, in the study carried out in the literature, to obtain an accurate estimate of alkalinity, the total amount of nitrogen obtained from proteins and the amount of CO_2 dissolved in the medium are evaluated for each substrate.

Regarding FOS, the literature presents a problematic issue as this parameter considers the acids in the system as if they were only acetic acid, neglecting all other types of acids formed during the acidogenesis phase [22].

Therefore, to obtain more accurate results, yield coefficients of the biomasses concerning these substances, taken from the literature, were used in this study allowing a simpler but effective estimation of acetic acid, butyric acid, propionic acid and valeric acid production from the proteins and sugars in each substrate, which are considered as the main sources of system acidification. Only proteins and sugars are considered since no significant contribution has been shown from lipids degradation to FOS components.

Parameter	Value	Ref. macromolecule	Ref. compound
fac,aa	0.4	Proteins	Acetic acid
fbu,aa	0.26	Proteins	Butyric acid
fpro,aa	0.05	Proteins	Propionic acid
fva,aa	0.23	Proteins	Valeric acid

Table 1: yield coefficients used to calculate FOS values from macromolecule concentration.

At this point, all the information needed to derive the FOS/TAC of each substrate are available. Once this parameter has been calculated, the biomethane yield trend to this stability indicator can be derived. An example can result in the trend shown in Figure 3:
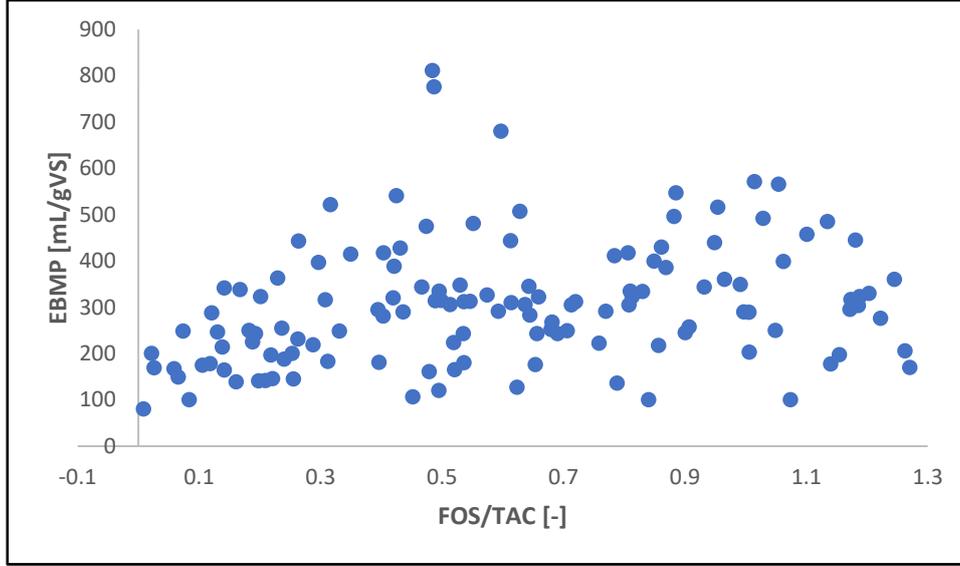
Figure 3: FOS&TAC calculation results with respect to BMP.

From the plot, 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. It is important to point out that the obtained results deviate slightly from what was expected from the literature due to more precise results thanks to the previously described fi coefficients. In fact, the maximum of FOS/TAC is obtained in correspondence of a value equal to 0.5 instead of 0.4.

At this point, the goal is to demonstrate the power of FOS/TAC to evaluate the stability of anaerobic digestion. Consequently, it has been used a two-parameters equation, that are C/N and the biodegradability, to estimate the biomethane yield for a mixture of two substrates: Chicken Manure (CM) as referenced biomasses, and Sugar-Beet Byproducts (SBB) and Exhaust Kitchen Oil (EKO) as blending agent. The possible synergistic effects between them are taken into consideration. Equations used are (Eq. 2.3-2.6):

$$BMP_{acod} = x_1 BMP_1 + x_2 BMP_2 + x_1 x_2 BMP_{mix} \tag{2.3}$$

$$BMP_{mix} = \beta_0 + \beta_1 \left(\frac{C}{N}\right)_{mix} + \beta_2 BD_{mix} + \beta_3 \left(\frac{C}{N}\right)^2_{mix} + \beta_4 BD^2_{mix} \tag{2.4}$$

$$\left(\frac{C}{N}\right)_{mix} = \Sigma_{i=i}^{NC} x_i \left(\frac{C}{N}\right) \tag{2.5}$$

$$BD_{mix} = \Sigma_{i=1}^{NC} x_i BD_i \tag{2.6}$$

19

BMP_{acod} is the biomethane potential associated with the mixture taken into consideration, while the BMP_{mix} term allows to quantify the goodness of the synergy formed between the biomasses. The β_i parameters are interaction coefficients obtained with a multidimensional regression analysis on all the biomasses [18]. Their values are reported below:

β_0	β_1	β_2	β_3	β_4
21.6613	1.2558	445.7076	-0.0223	-7.8201

Table 2: interaction coefficients values.

To assess whether FOS/TAC can be considered a stability parameter, a reference substrate (represented by the red dot in Figure 5) has been selected, and using the BMP_{mix} equation, synergy with all other substrates was evaluated one by one. To determine if synergy was present between these substrates, the biomethane yield for each pair was plotted against the mixture composition. The presence of a maximum point between the two pure components is the signal of synergy between these biomasses.

However, if the trend resembled a monotony curve, no synergy is present at all. The results of both cases are shown in Figure 4.
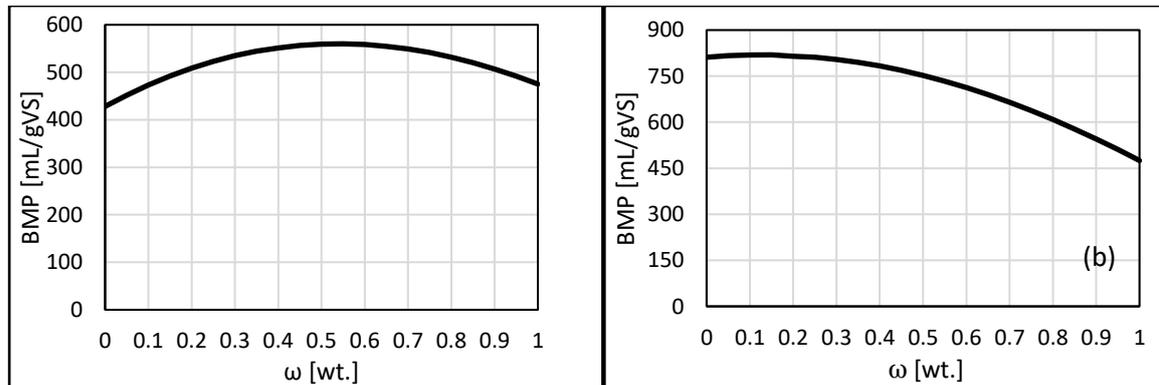

Figure 4: (a) synergies found between Chicken Manure (CM, 475 mL/gVS) and Sugar-Beet Byproducts (SBB, 428 mL/gVS); and (b) antagonism/absence of synergies found between Chicken Manure(CM, 475 mL/gVS) and Exhaust Kitchen Oil (EKO, 811 mL/gVS).

If observed from a C/N point of view, it is possible to note that the obtained results are coherent with the acceptable values of C/N for an anaerobic digester. In fact, in the case of synergism, we have a co-digestion between a substrate that is characterized by a high N content (CM, with N = 4.37%mol) and a substrate with a high C content (SBB,

with C = 41.62%mol) leading to an optimal C/N, while, in the case of antagonism, we have a co-digestion between two substrates that are both rich in terms of C. CM has a C content of 40.16%mol, while EKO has a C content of 73.52%mol. Therefore, this combination leads to a high value of C/N causing so a possible process instability.

The results of the study are shown in Figure 5, where the reference substrate is highlighted in red and substrates that showed synergy with it were colored green, while those without synergy were colored yellow.
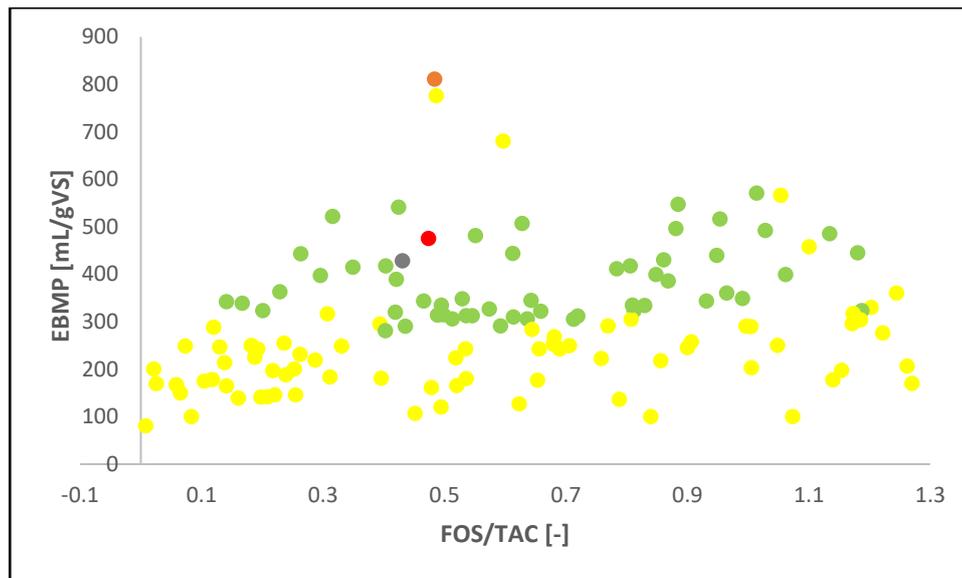

Figure 5: FOS/TAC results for Chicken Manure with highlighted regions of synergies (green dots) and regions of no synergies (yellow dots). In dark green, is indicated the Sugar-Beet Byproducts biomass, while in orange the Exhaust Kitchen Oil biomass.

It can be observed that there are different regions where the closer a substrate is to the reference substrate, the higher the probability of synergy. Moreover, to validate the synergies presence, there are indicated in dark green and orange the two biomasses that show synergies and antagonism studied before and reported in Figure 4, namely sugar-beet byproducts and exhaust kitchen oil, respectively.

As a confirmation, a second reference biomass has been analyzed, Sow Manure (SM), while the procedure for evaluating synergistic effects is the same as before. For this new case study, the trends obtained are shown in Figure 6.
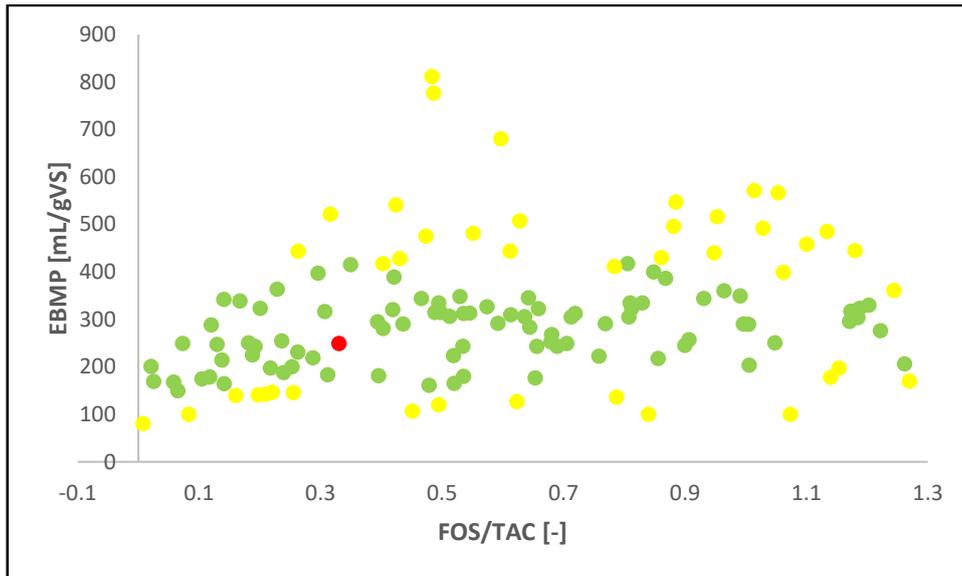
Figure 6: FOS/TAC results for Sow Manure with highlighted regions of synergies (green dots) and regions of no synergies (yellow dots).

It is possible to observe that, even under these new conditions, there are areas where it is possible to have a greater presence of conjugated biomass (i.e., biomasses that are in synergy with the referenced one) and thus a greater probability of having a stable and performant process; this occurs for regions closer and closer to the reference substrate. On the other hand, there are also areas where the probability of no synergy is high, and these emerge more and more as we move away from the red dot. So, iso-regions of synergies can be highlighted, where biomasses can be classified by their level of process stability and quality when blended with the referenced one in an optimal solution (Figure 7-8).
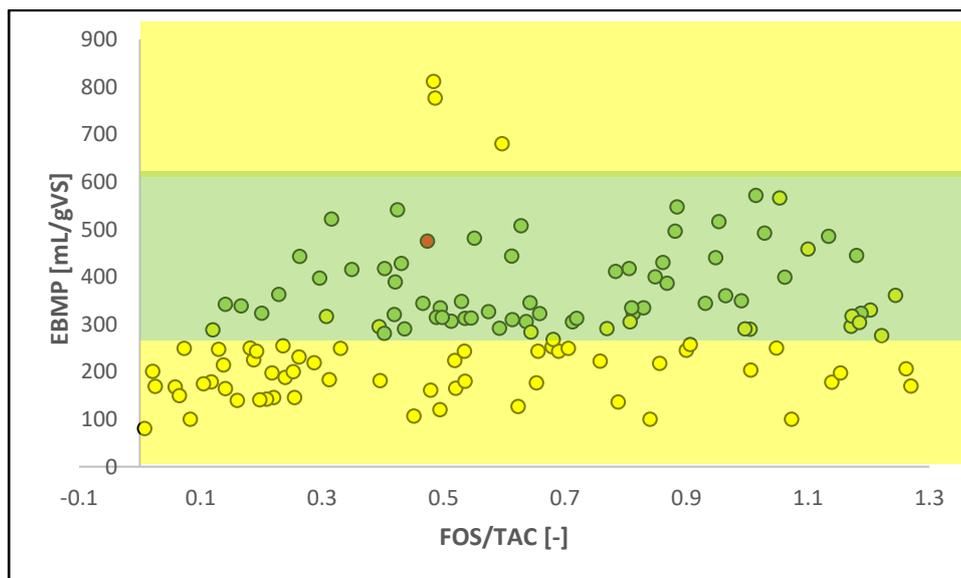

Figure 7: Iso-regions detection in FOS/TAC analysis for Chicken Manure.
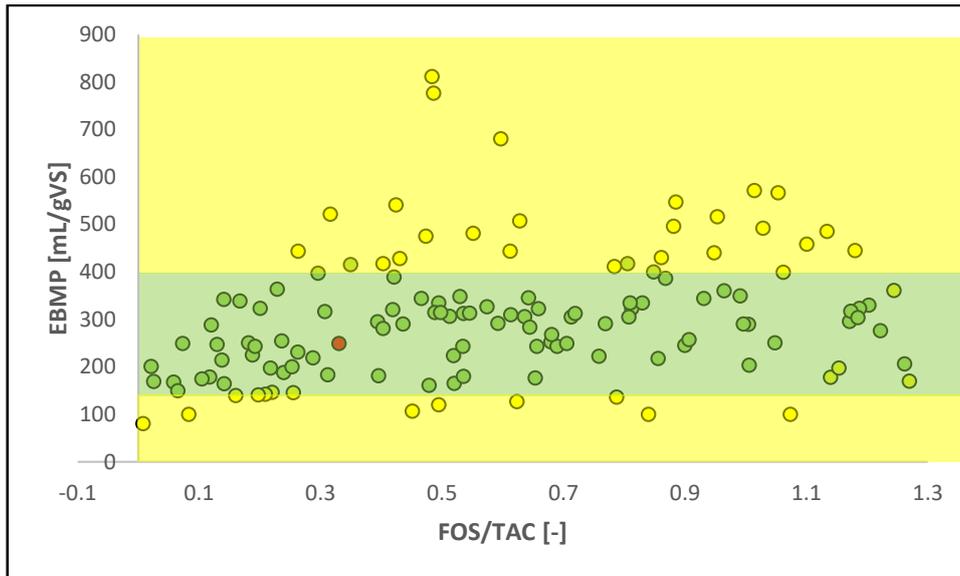
Figure 8: Iso-regions detection in FOS/TAC analysis for Sow Manure.

In conclusion, it can be argued that the FOS/TAC parameter can be used to assess the stability of the anaerobic digester but if it is considered within a specific influence zone (the red area in Figure 9-10), i.e., for FOS/TAC values less than 1.0. The reasons connected to this choice are the following two:

1. 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;
2. It's true that, if the considered reference substrate is affected by a small value of FOS/TAC, it is necessary to use a second biomass characterized by a high value of this parameter in order to reach an optimal FOS/TAC_{mix}; but, on the other hand, it makes no sense to take into consideration a substrate with a too high volatile fatty acids content.
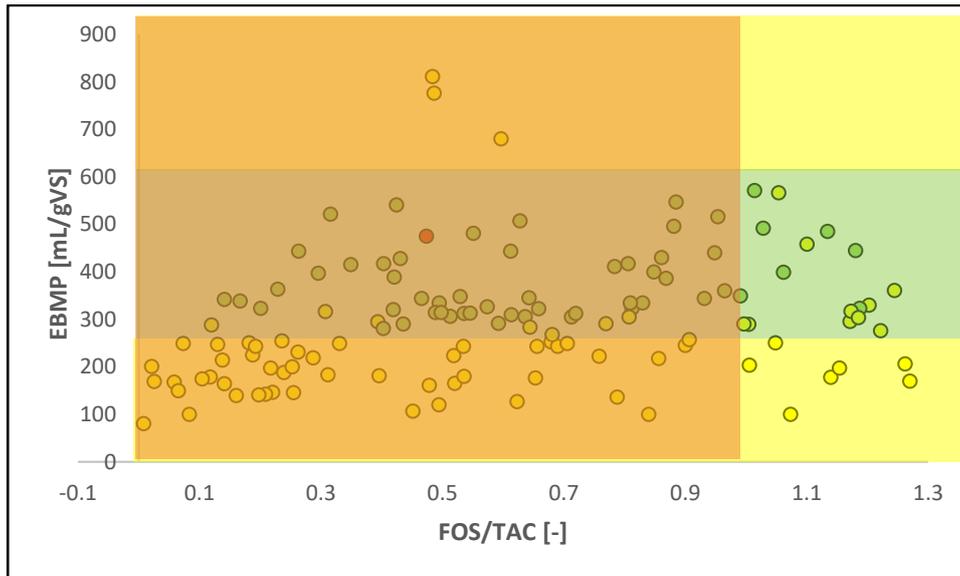
Figure 9: Iso-regions detection in FOS/TAC analysis for Chicken Manure with the area of interest.
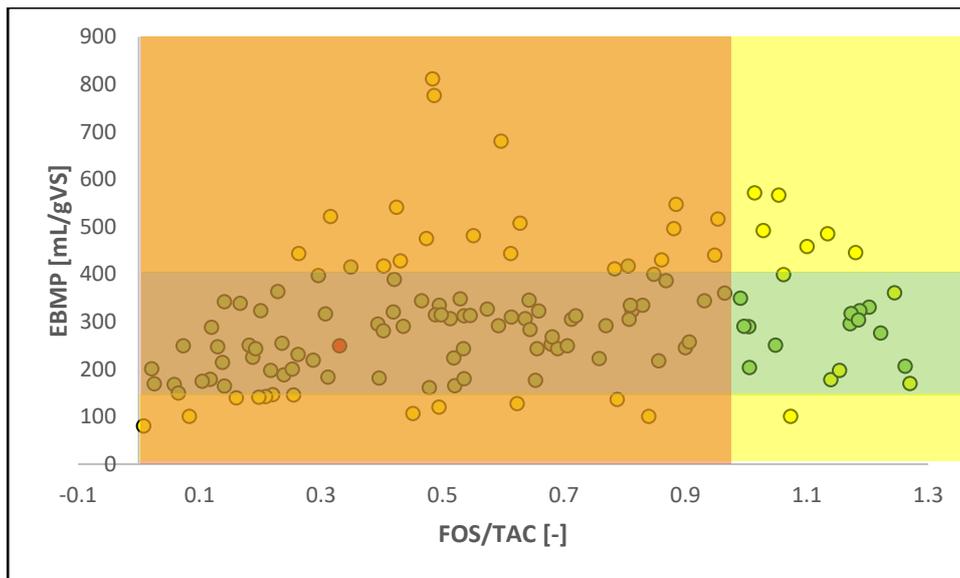

Figure 10: Iso-regions detection in FOS/TAC analysis for Sow Manure with the area of interest.

It is important to consider two substrates that have sufficiently similar predicted FOS/TAC values to each other to achieve synergistic effects, thus being within the same iso-regions. At the same time, it is noteworthy to make a consideration regarding the possible choice of two substrates that are close to each other but in regions where the FOS/TAC is too high or where it is too low:

- If the reference substrate and the conjugated one are characterized by synergistic effects but they are both affected by low values of FOS/TAC, this

means that the synergy is not connected with this stability indicator. Indeed, having two small values, the FOS/TAC_{mix} is outside the acceptable range. The same holds true for two high values of FOS/TAC.

In conclusion, it is possible to see that iso-regions are different for every biomass considered. However, by taking substrates that are near to the referenced one, there will be strong synergies effects. The biomasses that are in the green iso-regions are those that explicate this behavior and are called conjugated biomass. The ones in the yellow iso-regions do not show any synergies at all and won't significantly influence the digestion process but decrease the biomethane potential. Regarding the substrates that show synergies effects, but they are outside the area of interest, this can be explained by the fact that this synergy could be real, but the FOS/TAC is not the correct stability parameter that is able to describe these effects.

2.2. *Model derivation to calculate FOS/TAC_{mix}*

Once have been demonstrated the presence of iso-regions in the previous plots, now it is necessary to estimate a model that is able to calculate the FOS/TAC of the mixture fed to the reactor. Its aim is to understand if the obtained FOS/TAC_{mix} plays an important role over the stability of the process. In such a way as to derive the desired model, the following assumption has been made:

$$FOS/TAC_{mix} = x_1 FOS/TAC_1 + x_2 FOS/TAC_2 + x_3 FOS/TAC_3 + FOS/TAC_{syn} \quad (2.7)$$

The x_i represent the massive fractions of the substrates with their values of FOS/TAC. The fourth term permits to take into consideration the possible synergistic effects between the substrates and it is defined through the following expression:

$$FOS/TAC_{syn} = x_1 FOS/TAC_1 x_2 FOS/TAC_2 + x_1 FOS/TAC_1 x_3 FOS/TAC_3 + x_2 FOS/$$
$$TAC_2 x_3 FOS/TAC_3 + x_1 FOS/TAC_1 x_2 FOS/TAC_2 x_3 FOS/ TAC_3 \quad (2.8)$$

In order to carry out a first validation about the model proposed above, it has been done a multiple linear regression by exploiting the general expression reported below:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 x_3 \quad (2.9)$$

In the case of interest, it becomes:

$$FOS/TAC_{mix} = \beta_0 + \beta_1 x_1 FOS/TAC_1 + \beta_2 x_2 FOS/TAC_2 + \beta_3 x_3 FOS/TAC_3 + \beta_4 FOS/TAC_{syn} \qquad (2.10)$$

At this point, the experimental data obtained from this paper [27] were used to reach a first validation. This essay shows the time-dependent development of both the FOS/TAC of a mixture of three substrates and the FOS/TAC in the case of mono-digestion of both the same three biomasses. The three substrates are:

1. Banana pseudo-stem, (BPS);
2. Sugarcane baggage, (SCB);
3. Chicken manure, (CM).

The information related to the co-digestion and to the three mono-digestions allow to derive all the terms needed to realize the operation of regression. Indeed:

- FOS/TAC_{mix}, assumed as the y, has been obtained through the co-digestion taking its specific value for each day;
- FOS/TAC_1, FOS/TAC_2 and FOS/TAC_3 have been procured through the mono-digestions in the same way as done with FOS/TAC_{mix};
- x_1, x_2 and x_3 are fixed by the specific mixture of the co-digestion. This mixture is defined by the following values: 0.457 for BPS, 0.262 for SCB and 0.281 for CM.

After these premises, with a R_2 equal to 0.926, the regression has allowed the achievement of the following behaviors:
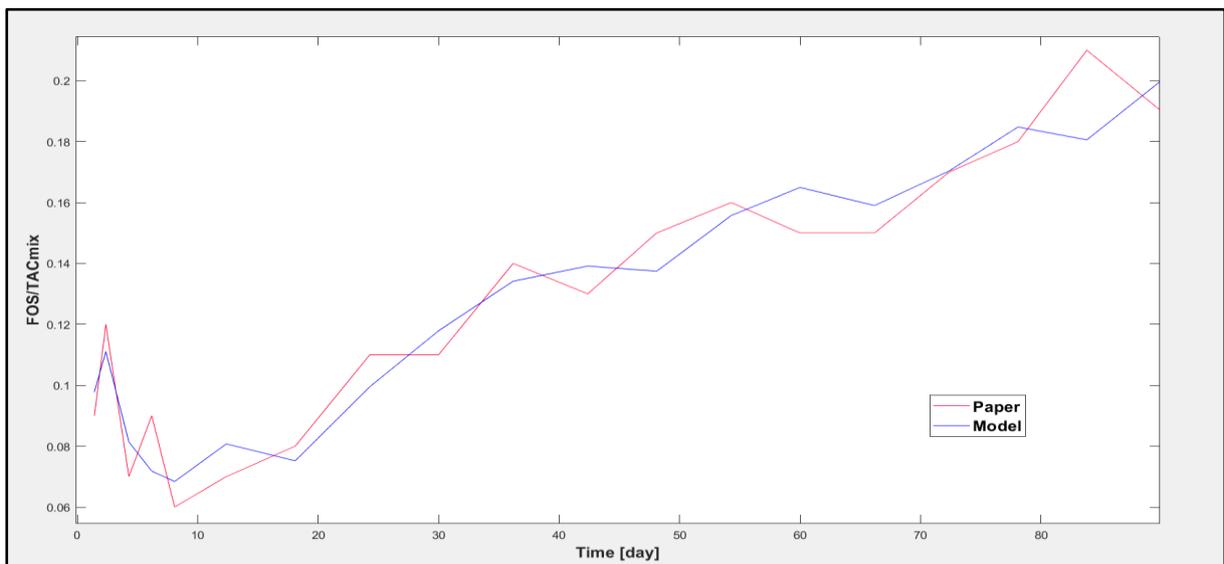

Figure 11: Plot related to the regression analysis.

26

The red trend represents the development of the experimental data, while the blue one is obtained through the multiple linear regression. It is possible to note that, the model expressed by the equation 2.10 is able to fit the experimental points in a very good way and this can be further confirmed by the previous value of R_2.

Therefore, it is possible to argue that the FOS/TAC_{mix} can be calculated thanks to the equation proposed previously. At this point, it is fundamental to obtain the specific value of β_0, β_1, β_2, β_3 and β_4 in order to use the FOS/TAC_{mix} equation for all the possible case studies. For this reason, what it has been done is to analyze different papers related to co-digestions of two and three substrates ([27], [28], [29], [30], [31], [32], [33]). Then, their relative FOS/TAC_{mix} values have been extracted with their specific compositions in order to know x_1, x_2 and x_3. Regarding FOS/TAC_1, FOS/TAC_2 and FOS/TAC_3, they have been taken from our database considering the substrates that characterize the specific co-digestion.

At this point, in order to obtain the values of the five β, it has been used Minitab; It is a software that is able to perform very accurate regression analysis thus obtaining the following results with a R_2 equal to 0.913:

$$FOS/TAC_{mix} = 0.0332 + 0.267S1 + 0.635S2 + 0.750S3 + 3.23S1S2S3 \qquad (2.11)$$

Therefore, the different β values are:

β_0	β_1	β_2	β_3	β_4
0.0332	0.267	0.635	0.75	3.23

Table 3: β$_i$ values.

Regarding the meaning of S1, S2, S3 and S1S2S3:

S1	S2	S3	S1S2S3
FOS/TAC1*X1	FOS/TAC2*X2	FOS/TAC3*X3	FOS/TAC,syn

Table 4: meaning of S1, S2, S3 and S1S2S3.

In such a way as to highlight the accuracy of this model, it is very important to also analyze the residual plots:
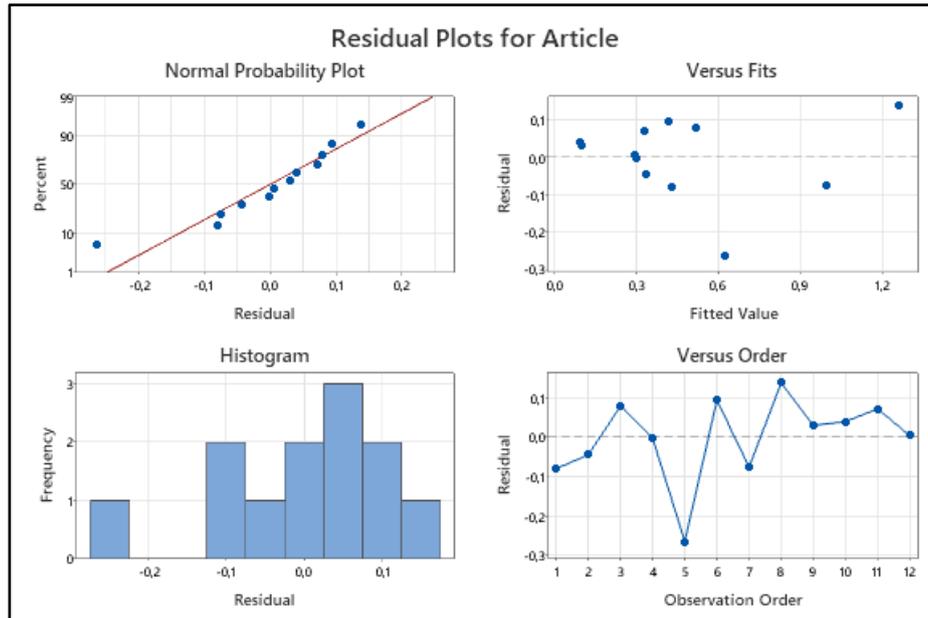
Figure 12: Residual plots related to the regression analysis.

The two graphs in the upper part are fundamental and they are characterized by correct and significant behaviors. The one on the left requires that the different points distribute themselves normally with respect to the diagonal. The second plot must have a random distribution of the points with respect to the zero without following specific trends. It is possible to observe that these required behaviors are present in the previous graphs except for the tails of the plot on the left, but this is common; in fact, very often the tails are not able to follow a normal distribution over the diagonal. This is an additional confirmation that the model is accurate and it permits to achieve reliable results. At this point, it could be interesting to evaluate the influence that the S1, S2, S3 and S1S2S3 terms have over the average value of FOS/TAC_{mix}. For this reason, it is reported the following plot:
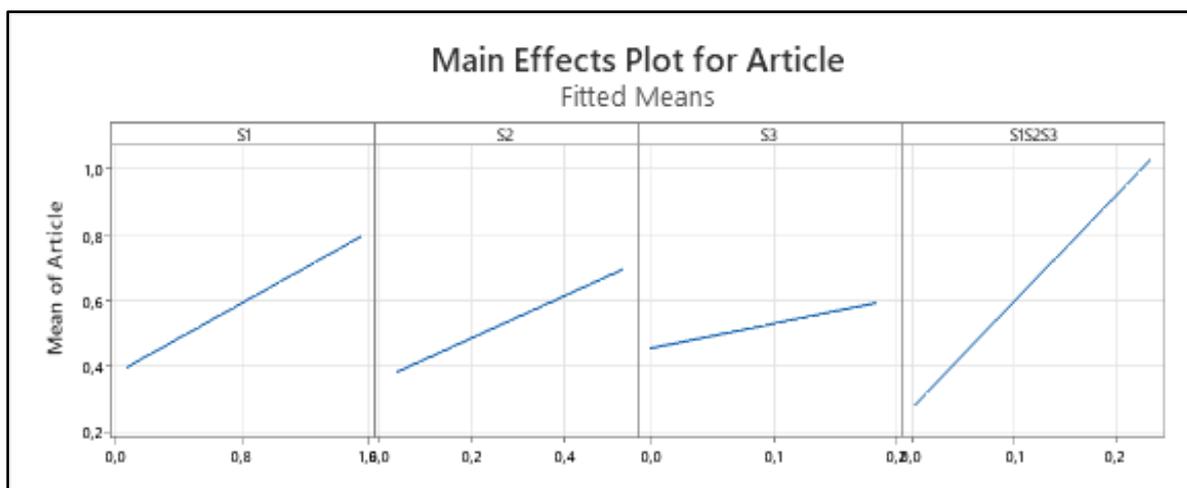

Figure 13: S1, S2, S3 and S1S2S3 influence over the average value of FOS/TACmix.

The term "Article" must be intended as the FOS/TAC_{mix}, that represents the experimental data. Returning to the plot, it is possible to observe that the S1S2S3 term is characterized by the highest influence since its straight line has a greater slope with respect to those related to S1, S2 and S3. This aspect is crucial because it permits to point out that the synergistic effects between the different substrates play an important role in the estimation of FOS/TAC_{mix}. Therefore, it is a mistake if these effects are neglected.

2.3. Industrial validation

One of the last things that remains to do regarding the FOS/TAC is related to the application of this parameter to a real industrial case-study. For this reason, we start from experimental data given by a specific company containing the information about the daily diet of an anaerobic digester from January 1, 2022 to January 31, 2022. This is a co-digestion of three substrates that are chicken manure, straw and fruit, vegetables and food waste. Two different situations are examined: a non-optimized condition and an optimized one. The first case reports, for each day, the experimental composition of the three substrates fed to the anaerobic digester and the BMP obtained from that mixture. The second one is characterized by the specific composition that is able to maximize the methane yield in every single day obtained by the proposed model. At this point, to evaluate the behavior of FOS/TAC during the process of optimization, the alkalinity parameter has been calculated through the model derived previously by exploiting the available FOS/TAC values of our database. Therefore, it has been achieved the following results:
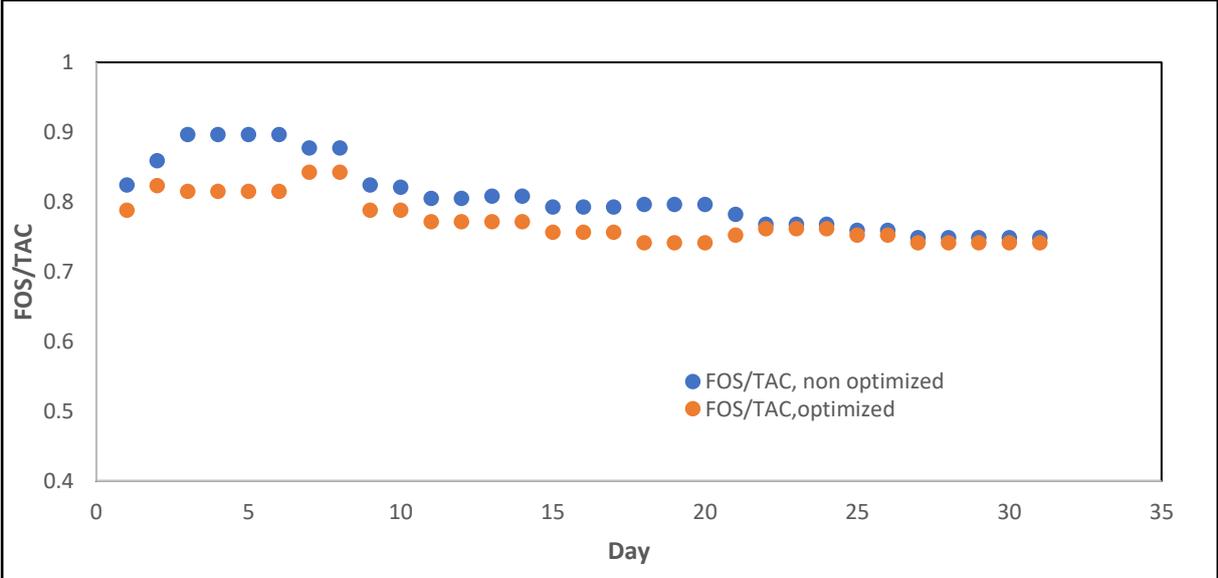

Figure 14: FOS/TAC behavior with the passing of the days in the non optimized condition and in the optimized one.

From the trends reported above, it is possible to observe that FOS/TAC plays an important role during the process of optimization. Indeed, in the non-optimized case study, the parameter reaches values in the order of 0.8-0.9, so values that can cause the accumulation of VFAs leading than to the inhibition of the process. On the other hand, in the optimized condition, the FOS/TAC values are always lower reaching in some days values that does not exceed 0.75. This means that, through the optimization, the FOS/TAC try to reach more acceptable values. Furthermore, also in the optimized case, the ideal value of 0.5 is not reached. This is because, during the anaerobic digestion, there is an important concatenation of different operating parameters and each of them influences in its own way the process. Therefore, in conclusion, it's not correct to optimize the anaerobic digestion only considering a single stability indicator. In fact, it's better to consider it as the result of an optimization realized over a complex combination of various operating parameters.

2.4. Validity of optimal range

The last aspect on which it is important to focus the attention is the range of FOS/TAC values used in this dissertation to guarantee a stable performance of the process. Coherently with what is generally provided by the literature and with what has been demonstrated until now, it's correct to consider as acceptable range the one between values of 0.3 and 0.5. On the other hand, there are some researchers that have demonstrated a stable performance of anaerobic digestion outside the range mentioned before, indeed:

- [34], [35], [36], [37]: These papers highlight a stability of the process for FOS/TAC values close to 0.2-0.25;
- [35], [36], [38], [39], [40]: These studies show that, also for high values between 0.6 and 0.7, it is possible to avoid the inhibition of the anaerobic digestion.

Therefore, it is correct to apply a relaxation to the stability range associated with FOS/TAC. The one used for this dissertation considers values between 0.2 and 0.7. This choice allows to obtain more reliable results in the machine learning described in the chapter 6. Indeed, exploiting the range of values between 0.3 and 0.5, the simulations are not able to return reasonable results because of the limited amount of acceptable FOS/TAC values. Thanks to this relaxation, it is possible to improve the reliability of the predictions provided by the machine learning.

3 OLR

3.1. Influence of the bacterial community on process stability

Once the stability of anaerobic digestion has been monitored through a parameter linked with its alkalinity level, at this point it could be interesting to obtain a stability indicator related to the bacterial community of the system. Indeed, by means of meticulous analysis, an exhaustive understanding of the evolution of the microbial community during the process could be fundamental to optimize the performance of the reactor. Unfortunately, this is not possible; In fact, as we will see in a while, the results from the literature are very variable reaching contradictory conclusions. These results show the fact that the field of microorganisms and their behaviors is complex to be studied to find some laws that can clearly describe the operation of anaerobic digestion. For this reason, it will be studied a different parameter which shows coherent behaviors in the literature and, at the same time, it will be able to influence the bacterial community itself.

The anaerobic digestion consists in a sequence of biochemical reactions in which there are different groups of microorganisms with the aim of degrading the organic matter fed to the reactor; In fact, in absence of oxygen, they are able to convert the starting macromolecules such as proteins, carbohydrates and lipids into CO_2 and CH_4. The anaerobic digestion is characterized by four steps: hydrolysis, acidogenesis, acetogenesis and methanogenesis. It is very important to point out that, depending on the considered phase, there will be different groups of microorganisms and each of them with its own specific features in terms of nutritional needs, operating conditions and kinetic growth. This is the aspect that makes so challenging the understanding of how to achieve a harmonious coexistence between the microorganisms of the 4 phases. In the first three phases, there are microorganisms related to the Bacteria domain in which it is possible to find the genres of Pseudomonas, Bacillus and Clostridium that oxidize the organic products coming from hydrolysis step to pyruvate. Then, it will be converted into volatile fatty acids, alcohols and ketones. On the contrary, the last phase is only characterized by microorganisms related to the Archaea domain. In particular, there are two types of methanogens distinguished by the specific metabolic pathway through which the methane production happens:

- acetoclastic methanogens, such as Methanosarcina and Methanosaeta genres, they use acetic acid as a precursor in order to obtain CH_4:

$$CH_3COOH \rightarrow CH_4 + CO_2 \tag{3.1}$$

- hydrogenotrophic methanogens, like Methanobacteriales and Methanomicrobiales orders, that start from hydrogen to produce methane:

$$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O \tag{3.2}$$

One of the aspects related to the influence of bacterial community over the stability of the process could be the microbial diversity. It can define the complexity of the environment through two characteristics: the number of the species inside the system and their relative abundance. The biodiversity can be analyzed through three level:

1. α-diversity,
2. β-diversity
3. γ-diversity.

Regarding microbial diversity, α-diversity is the most used in the literature and it is measured by means of several diversity indices such as:

- Shannon index: $H' = -\sum_{i=1}^{S} p_i \log p_i$ $\qquad$ (3.3)
  It assumes a value equal to zero when all the microorganisms belong to one species only. While the higher the value of H', the greater the microbial diversity achieving in the end a more uniform distribution of microorganisms in all the species involved;
- Simpson index: $\lambda' = \sum_{i=1}^{S} p_i^2$ $\qquad$ (3.4)
  Its value varies between 0 and 1, in which 1 means low diversity, so one species represents all the microorganisms, while a greater microbial diversity is reached with a value increasingly closer to zero.

In both the previous formulas, S denotes the species inside the bacterial community and p_i indicates the relative abundance of the single species i:

$$p_i = \frac{n_i}{N} \tag{3.5}$$

N is the total number of microorganisms of the system, while n_i is the number of microorganisms belonging to the specific species i.

In the literature, unfortunately, it is not possible to find out a specific correlation between the level of microbial diversity inside the process and the methane yield; in fact, in [41] and in [42] the condition of maximum BMP is reached with the highest value of α-diversity, reflecting that microbial diversity decreases at the collapse phase because the VFAs accumulation causes the loss of different typologies of bacteria and methanogens. On the other hand, in [43] and in [44] the most efficient case study is characterized by a lower microbial diversity with respect to the one related to a situation with a lower methane yield. Considering what has been said, this parameter cannot be used as reliable indicator in order to define a stable and efficient performance of anaerobic digestion since a higher α-diversity does not necessarily lead to a higher BMP. The inconsistency of the results related to the microbial diversity can be also observed by changing the operating temperature. In fact, in these papers ([42], [45], [46], [47]), it has been analyzed how the temperature is able to affect microbial community by considering two different regimes: mesophilic anaerobic digestion (35°C) and thermophilic anaerobic digestion (55°C). What emerges from these papers is that in some cases the microbial diversity is higher in the mesophilic case study while in others is the opposite. So, this means that, it's true that temperature is able to change the properties of the bacterial community, but without following regular behaviors.

A second aspect that the literature suggests as possible influence factor over the stability of the process is the ratio between the amount of acetoclastic methanogens and hydrogenotrophic methanogens. In fact, it is important to have a good combination of both methanogens' typologies. The former allows to consume the volatile fatty acids preventing their accumulation that could cause the inhibition of the process. The second is able to use hydrogen atoms coming from the bioconversion of VFAs that occur in the acetogenesis step. This aspect is fundamental because, if they are not consumed, these reactions are no longer thermodynamically favored due to an high hydrogen content, causing so the inhibition of the process that occurs for a H_2 concentration higher than 10^-4 [atm] [48]. Also under this point of view, the results seem to be very discordant with each other; in fact, in [43] the most efficient case study is the one characterized by an acetoclastic-hydrogenotrophic ratio equal to 7/3. This means that, in order to reach a stable performance of anaerobic digestion, it would be necessary to have a methanogenesis shifted more towards an acetoclastic pathway, so a predominance of acetoclastic methanogens.

This aspect is also highlighted in [41] and in [49], in which the inhibition of the process happens with the transition from an acetoclastic pathway to a one more hydrogenotrophic. The problem is that there are a lot of papers in which the phenomenon described previously does not happen; in fact, in [48], it is possible to note that, in the conditions in which the inhibition of the process is reached, the dominant methanogens continue to remain those that consume the volatile fatty acids.

This leads to the conclusion that, a more acetoclastic methanogenesis is not a sufficient condition in order to obtain a more efficient and stable performance of the process. Such approximative results with respect to this parameter may also be caused by low accuracy associated to the laboratory analysis. These analyses are performed in order to estimate the typologies of microorganisms present in the system, but, at the end of these tests, there is always a very high percentage of Bacteria and Archea that cannot be identified. Therefore, this inefficiency leads to inaccurate results, since it is not possible to have the absolute certainty of how many acetoclastic methanogens or hydrogenotrophic methanogens are present in the bacterial community of the system. Therefore, both microbial diversity and acetoclastic-hydrogenotrophic ratio are not so reliable, even if apparently seemed to be good indicators for process stability. From what has emerged so far, it is possible to understand the complexity of monitoring the microbial community inside the anaerobic digestion in order to ensure good performances of the process. For this reason, now the focus will be on another operational parameter, i.e. the organic loading rate, that plays a very important role in the literature and that, as we will see in a while, it will be able to also affect the bacterial community itself.

3.2. Meaning of OLR and its influence on pH level of the system

The organic loading rate (OLR) is a very important parameter in order to define the operating conditions that characterize the process of anaerobic digestion; in fact, OLR represents the amount of volatile solids that must be fed to the reactor every day [50]. The volatile solids are the portion of organic matter inside the substrate while the remaining part of solids is defined as fixed solids [50]. The amount of undegraded solids is given by the sum of fixed solids and a specific fraction of volatile solids since the substrate's biodegradability is never equal to one. OLR can be calculated through the following equation:

$$OLR = \frac{VS * Q}{V} \tag{3.6}$$

Where VS [kg/m3] is the concentration of volatile solids in the feeding; Q [m3/day] is the inlet flow rate to the anaerobic digester, and V [m3] is the volume of the reactor [51]. So OLR is obtained in terms of [$kg_{VS}/m^3/day$], but since the ratio between V and Q represents the hydraulic retention time [day], OLR can be also expressed as:

$$OLR = \frac{VS}{HRT} \tag{3.7}$$

What is found in the literature is that, by increasing the value of OLR, it is possible to reach a certain point in which the value of organic loading rate is so high that causes the inhibition of the process. It is very important to point out that the OLR value that induces inhibition changes from case to case, as a matter of fact it depends on several factors such as operating conditions and reactor configuration. Despite these numerous parameters that play an important role in the stability of the process, it was possible to note that in many papers the inhibition occurs for an OLR value equal to 6 [$kg_{VS}/m^3/day$] ([41], [48], [52], [53]). For this reason, it makes sense to consider 6 [$kg_{VS}/m^3/day$] as that organic loading rate value beyond which anaerobic digestion becomes unstable.

A very important aspect associated with OLR is that the negative effects caused by the reaching of its critical value can be observed in many useful parameters for defining the performance of the process, like the reduction in VS degraded percentage, that occurs for very high OLR value, and the amount of CO_2 produced during the anaerobic digestion [50]. In fact, the amount of biogas in terms of CO_2 and methane is used to analyze the efficiency of the process in which stable and optimal conditions are obtained for a CH_4 composition between 60% and 65% [50]. Once critical OLR values are obtained, the quantity of CO_2 in the biogas increases and this means that the acidifying microorganisms starts to prevail over the methanogens leading to a strong VFAs accumulation [50]. In fact, for high organic loading rate, the production rate of hydrolysis and acidogenesis steps is much higher with respect to the one related to methanogenesis step and so it results very difficult to consume in a correct way the volatile fatty acids ([48], [52], [53]). This inefficiency causes an imbalance between the volatile fatty acids produced and those consumed, leading to a critical accumulation of VFAs and subsequent inhibition of the process. The amount of the two families of microorganisms is strongly affected by the OLR, but not only for its high values as described before, but also in start-up conditions in which the value of organic loading rate is low. In fact, in the presence of low OLR, methanogens are in a starving condition, in which their required nutrients are much higher than those fed [53]. Since it is not possible to satisfy in an appropriate way the methanogens diet, this group of microorganisms tends to die thus preventing the production of biomethane; so, it is necessary to take advantage of a higher OLR to keep methanogens alive.

Returning to the case study of a high OLR, it is important to highlight that a critical level of accumulated VFAs is reached, but if this problem is seen under a microorganism's point of view, why does process inhibition occur? The correct answer is related to the optimal pH range for methanogens growth [53]; In fact, a strong VFAs accumulation causes a drastic reduction in pH coming out of the desired alkalinity level in the system and this leads to the methanogens death. The literature reports that the optimal pH range for methanogens growth is between 6.5 and 7.5 [53]. This means that, in order to obtain an efficient production of biomethane, it is necessary to use a

high organic loading rate until this feed causes a reduction of pH below a value of 6.5. Consistently with those stated previously, this happens for an OLR equal to 6 [kg$_{VS}$/m^3/day]. At this point, considering the aspects described formerly, it would be interesting to get a correlation that allows to describe the variations of pH level of the system as a function of the OLR fed to the reactor. So, through experimental data derived from different papers, it has been collected information about pH and the corresponding OLR values obtaining the following correlation:

Figure 15: Correlation between pH and OLR. Data have been gathered from literature ([41], [48], [53], [54], [55], [56], [57]).

The considered model, a second order polynomial, is able to fit very well the experimental points with a R$_2$ equal to 0.75:

$$pH = -0.1683 \cdot OLR^2 + 1.5303 \cdot OLR + 3.7355 \tag{3.8}$$

Focusing now on the physical meaning of the graph above, it's correct to say that the represented behaviors are coherent with those expected; in fact, increasing OLR, it's possible to reach the optimal pH range for methanogens growth, particularly for an OLR between 2.5 and 5.5 [kg$_{VS}$/m^3/day]. On the other hand, getting an OLR equal to 6 [kg$_{VS}$/m^3/day], it can be observed that the pH tends to decrease until causing the achievement of an acid environment for high OLR like 8 [kg$_{VS}$/m^3/day]. So, in conclusion, it's important to point out two aspects related to the previous correlation:

- It can be considered suitable to describe the variation of alkalinity level inside the reactor by changing the total amount of VS fed to the anaerobic digester;
- It cannot be considered right for all the case study; in fact, as said at the beginning, OLR is a parameter that depends on various factors, so it is possible that the process inhibition occurs for OLR lower or higher than 6 [kg$_{VS}$/m³/day].

3.3. Industrial validation

At this point, in the same way as done with the FOS/TAC, the last thing that remains to do about OLR analysis is the application of this stability indicator to a real case study. It has been used the same industrial case study examined with FOS/TAC. It is affected by the following features:

- co-digestion of three substrates;
- the availability of information related to the daily diet of the anerobic digestor;
- the same two different situations are taken into consideration:

  1. The first is characterized by the real diet;
  2. The second is defined by the optimized diet to maximize the biomethane yield every day.

The first analyzed scenario has been the one characterized by the values of flow rate and volume provided by the company, so: $Q = 290 \ m^3/day$ and $V = 4500 \ m^3$. Unfortunately, as we will see in a while, with these Q and V it has not been possible to reach an OLR equal to its value assumed as ideal, so $6 \ kg_{VS}/(m^3 day)$. For this reason, to verify if it was possible to achieve it, it has been realized a design problem changing, one at a time, one of the two data in a specific range of values.

Returning to the initial conditions, taking advantage of the information provided by the company, the following plots have been obtained:
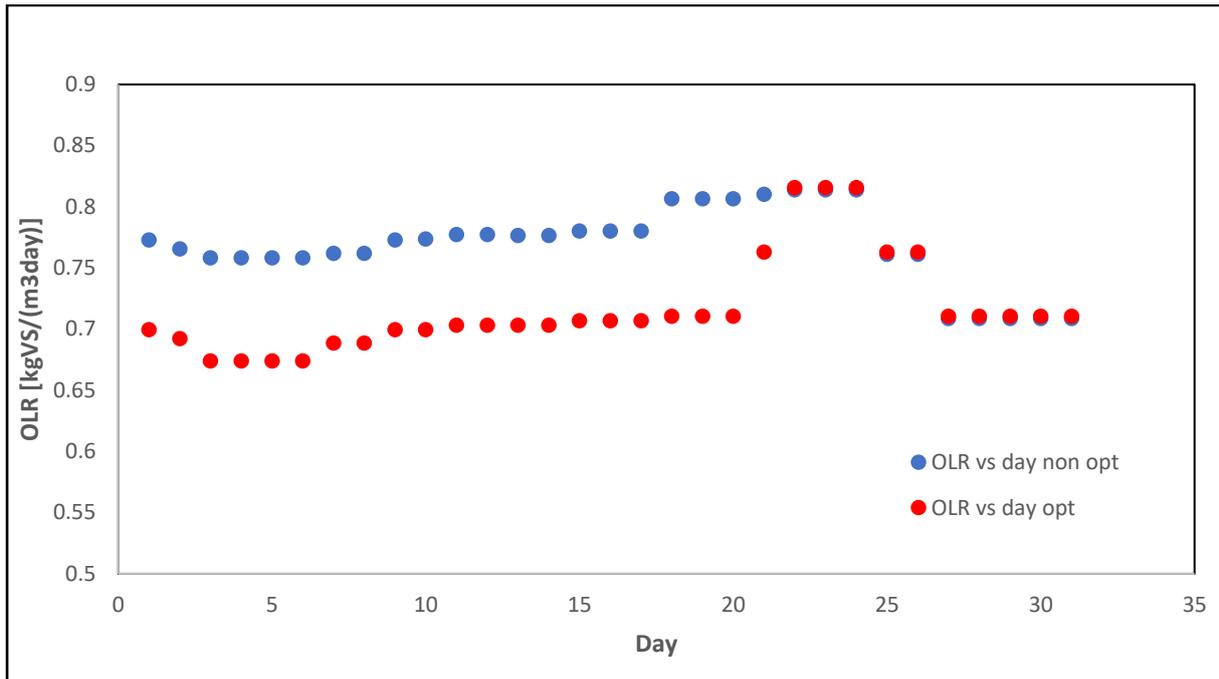
Figure 16: Behavior of the OLR with respect to the elapsed days in the non-optimized condition and in the optimized one: V = 4500 m3 and Q = 290 m3/day.

From the first plot, it is possible to extrapolate very significative and useful information; in fact, it is shown the behavior of OLR with respect to the passing of the days in the two different conditions. It is possible to note that, in the optimized case-study, the amount of volatile solids fed to the reactor is much lower than the one in the other situation. Therefore, this means that, the optimization of the process allows to achieve a higher BMP with a lower possibility to induce the inhibition of the process. Indeed, employing a lower amount of volatile solids, it is possible to accumulate a lower quantity of volatile fatty acids.

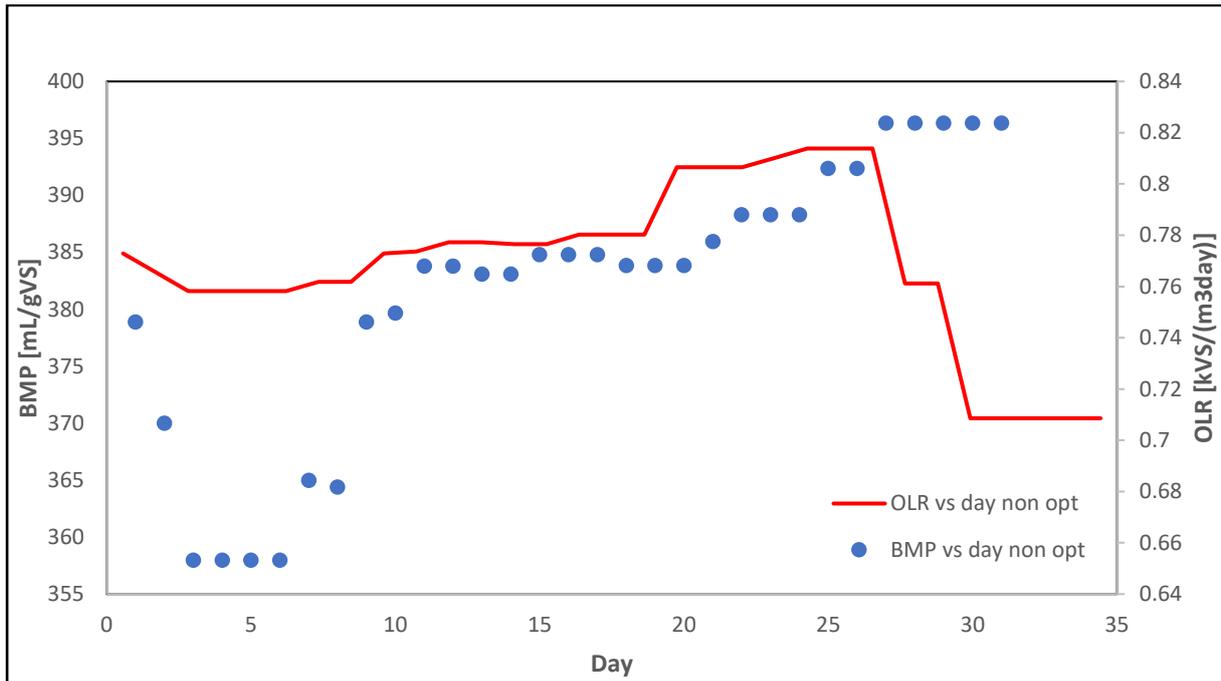Regarding the second and the third plot, it makes more sense if analyzed together:

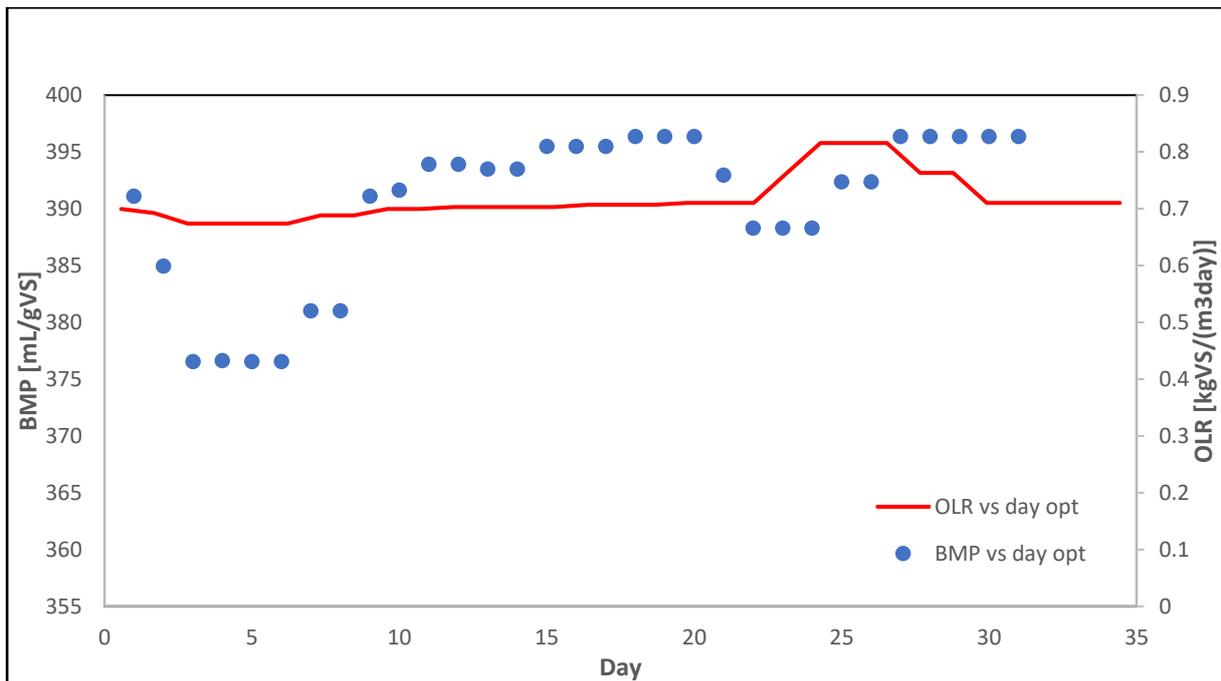Figure 17: Behavior of the OLR and BMP with respect to the elapsed days in the non-optimized condition: V = 4500 m3 and Q = 290 m3/day.

Figure 18: Behavior of the OLR and BMP with respect to the elapsed days in the optimized condition: V = 4500 m3 and Q = 290 m3/day.

In both, it is shown a double trend with respect to the elapsed days, so the one of OLR and the one of BMP, in order to observe the direct effect that the fed volatile solids have over the biomethane yield. The second graph represents the non-optimized case,

while the third is related to the optimized one. It is important to highlight how the BMP has been calculated in the optimized condition. In each day, it has been taken the specific diet and it is optimized with the blending optimization model described in this paper [18], allowing to find the optimal mixture composition to improve the biomethane potential. Returning to the graphs, what can be detected is that, thanks to the optimization, it is possible to obtain a feeding that allows to have a not very variable OLR from a day to another. This means that, with the passing of the days, there is a small variation of the fed volatile solids with respect to what happens in the non-optimized condition. The great advantage related to this aspect is that, thanks to the process optimization, it has been possible to define the specific VS concentrations of the different biomasses that permit to reach good synergies achieving in the end a higher BMP. Indeed, while the figure 17 has a minimum of BMP equal to 357 $[mL/g_{VS}]$, in the other situation the lowest BMP value is 376.53 $[mL/g_{VS}]$. Therefore, summarizing all the information extracted from these three graphs, what emerges is that:

- the optimized case study ensures an higher BMP;
- this higher BMP is achieved through a lower amount of volatile solids fed to the reactor and, at the same time, with a lower variation of them from one day to the next.

However, focusing now on the value achieved by OLR each day, the analysis leads to the conclusion that this parameter is far away from the ideal $6\ kg_{VS}/(m^3 day)$. Therefore, as anticipated previously, what has been done is to find the conditions that are able to reach the optimal organic loading rate by changing the flow rate and the volume with respect to the conditions provided by the company. The analyzed range for the volume is between 1000 m3 and 5000 m3; in fact, this is an acceptable range for the dimensions that the anaerobic digestor could have. Regarding the flow rate, it has been changed until reaching an hydraulic retention time without any physical sense. Each of the analyzed case-study is characterized by the same behaviors of those related to the initial conditions; in fact, as an example, the plots of V = 1500 m3 and Q = 290 m3/day are reported below:
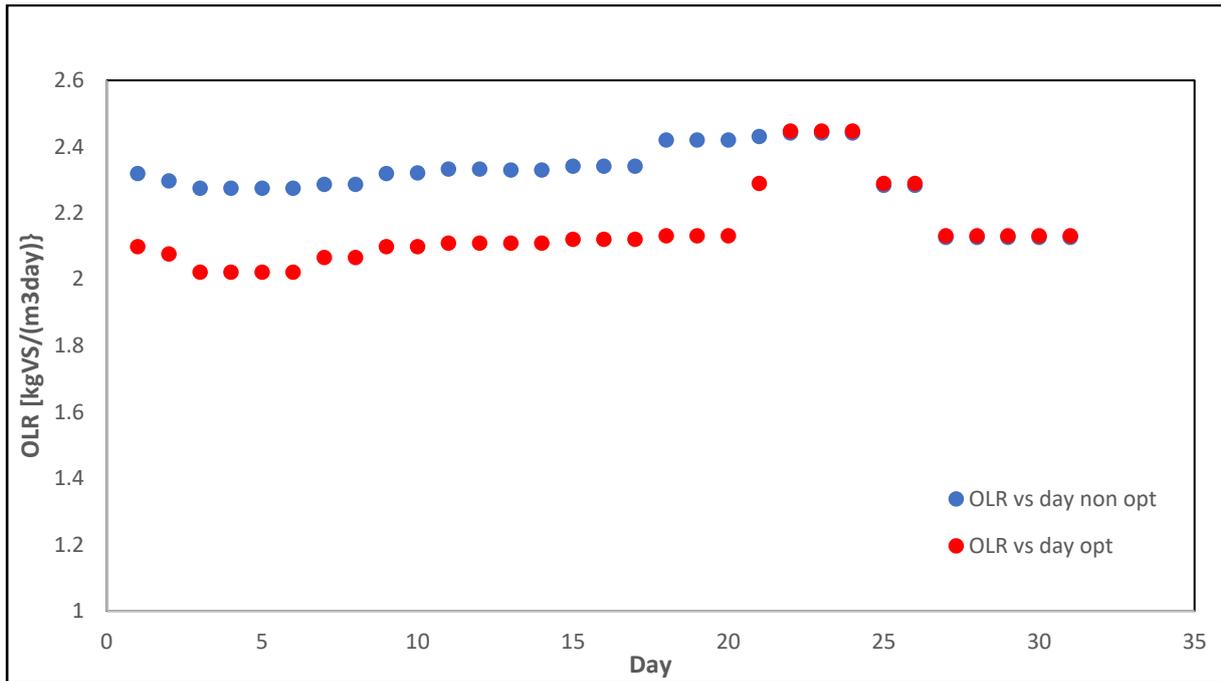
Figure 19: Behavior of the OLR with respect to the elapsed days in the non-optimized condition and in the optimized one for the case: V = 1500 m3 and Q = 290 m3/day.

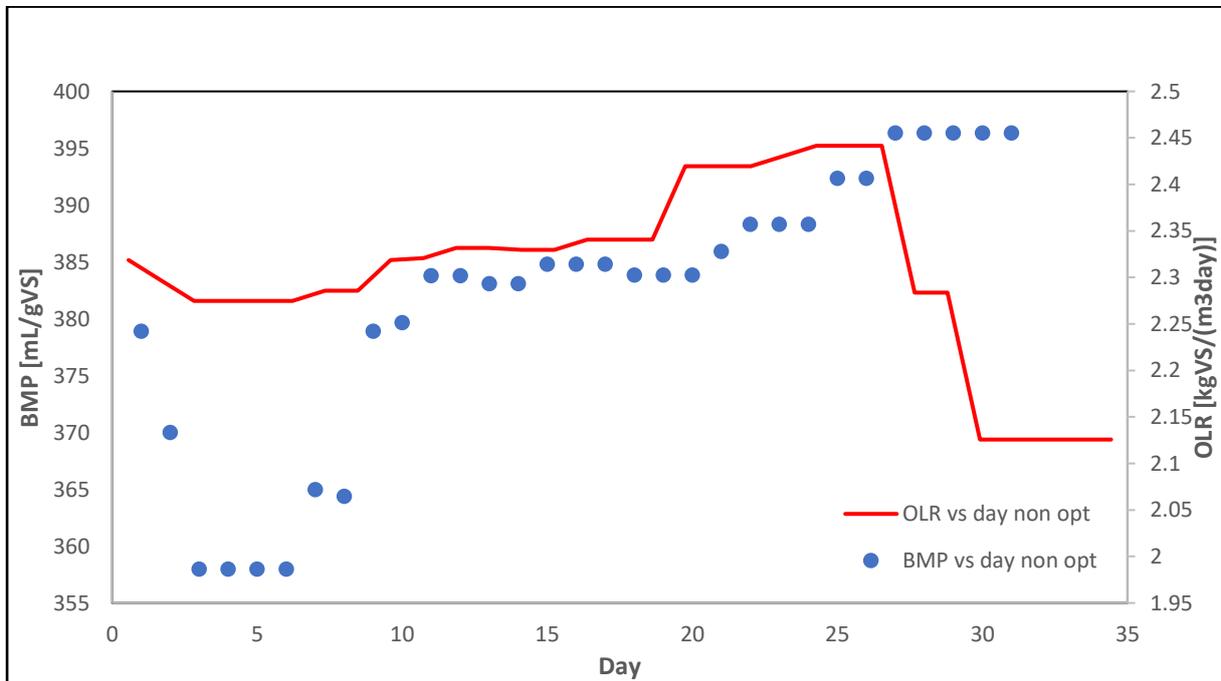Figure 20: Behavior of the OLR and BMP with respect to the elapsed days in the non-optimized condition for the case: V = 1500 m3 and Q = 290 m3/day.
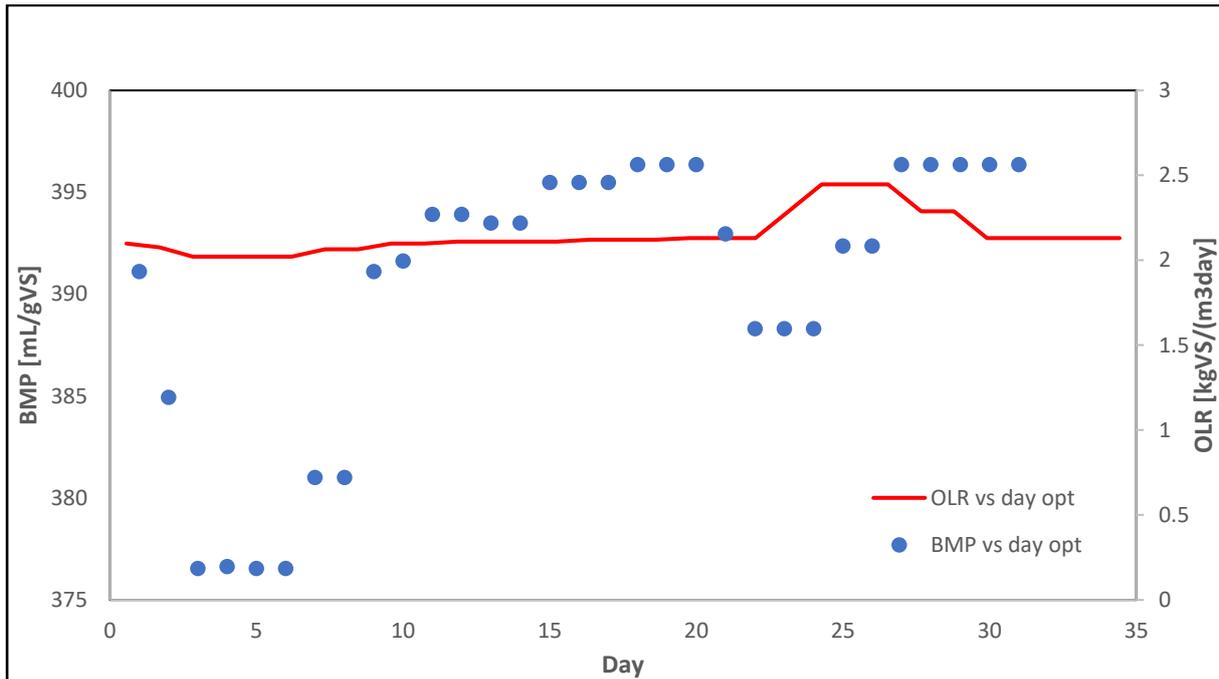
Figure 21: Behavior of the OLR and BMP with respect to the elapsed days in the optimized condition for the case: V = 1500 m3 and Q = 290 m3/day.

Indeed, also for this combination of data, the optimized condition guarantees a higher BMP through a lower amount of VS and through a more constant OLR. The problem is represented by the fact that, for the analyzed industrial case-study, there are two possible solutions to arrive close to the ideal organic loading rate value:

- using a volume equal to 580 m^3, that is small and outside the acceptable range;
- using a flow rate equal to 2300 m^3/day, that is a very high value causing the reaching of a small HRT without any physical sense, i.e. 1.96 $[day]$ with respect to the initial 15.5 $[day]$

Moreover, in order to highlight the difficulty in the achievement of an OLR of $6 \ kg_{VS}/(m^3 day)$, the trend that the flow rate should have each day to ensure an ideal organic loading rate is reported below:
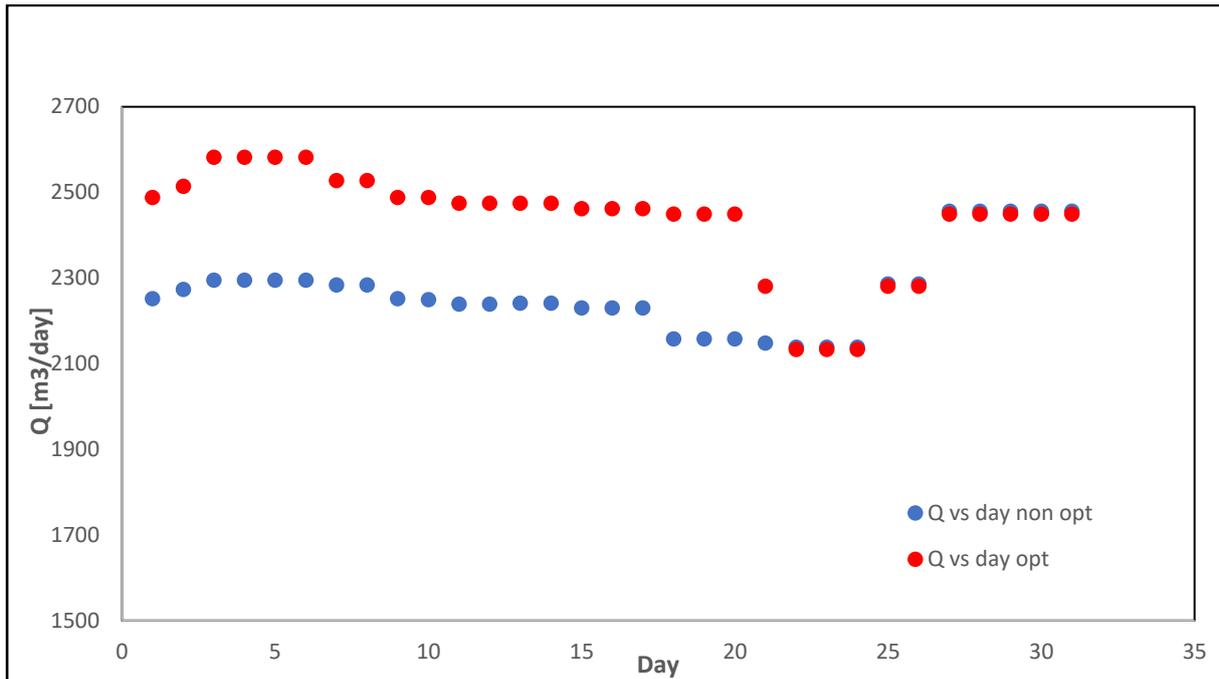
Figure 22: Behavior of the Q with respect to the elapsed days for an OLR = 6 kgVS/(m3day) in the non optimized condition and in the optimized one.

Therefore, it is possible to note that, in the two conditions, the flow rate should assume always too high values. They are of the same order of magnitude of the 2300 m^3/day reported previously. Despite this difficulty, this industrial case-study is very interesting because, as explained at the beginning, the optimal OLR value changes from case to case since there are a lot of factors that exert a specific influence over this stability indicator. For this reason, it is possible to conclude saying that, in this real case-study, it is not possible to reach the ideal value of OLR equal to 6 $kg_{VS}/(m^3 day)$ probably because the reactor configuration does not allow to achieve it. This indicates that the inhibition of the process will occur for a different value of organic loading rate.

4 Neural network

4.1. Introduction to neural networks

Thanks to the previous detailed analysis related to FOS/TAC and OLR, it has been achieved the first great goal of this dissertation: i.e. the definition of two new parameters required to guarantee a stable anaerobic digestion operation as a function of the features related to both the feeding and the conditions inside the reactor. At this point, it's possible to introduce the second great target of this work: the creation of an algorithm characterized by a specific purpose. Its aim is to define a conjugated substrate that can maximize the biomethane yield depending on a well-known substrate fed to the anaerobic digester. This algorithm must be able to also highlight the composition of the mixture that has formed. In such a way as to obtain this algorithm, it's necessary to create two different tools. Both are characterized by specific functionalities and peculiarities that will contribute to achieve this final goal. The first of these two tools is a neural network.

Also known as ANN (artificial neural network) or SNN (simulated neural network), the neural networks are a specific category that belong to the wide world of machine learning. They represent the central element of deep learning algorithms. Very often there is confusion between the concepts of neural networks and deep learning thinking that they are always the same thing, but there is a great difference between them: the depth of the algorithm. This feature refers to the number of layers that characterized the SNN; in fact, as shown in a while, in these typologies of machine learning, an important variable that permits to modify the architecture of the model is the number of the layers within it. The "deep" term refers to a higher amount of layers with respect to those that would characterize a simple ANN. A deep learning algorithm is when a number of three layers is exceeded, increasing not only the learning potentialities of the model, but also its computational costs. These concepts will be better explained later. The only thing that is important to remember for the time being is that the deep learning can be considered as a specific typology of neural networks.

Returning to SNN, let's describe how they can be trained. In such a way as to generate some predictions, a set of training data is used in which it is possible to find two typologies of examples:

1. input examples, x, characterized by specific features;
2. output examples, y, also known as classes or labels.

After given these x and y, with the passing of the iterations, the aim is to improve the accuracy of the model to guarantee increasingly accurate predictions. Once have been achieved this optimize version of neural network, what is obtained is a strong learning tool that permits to classify the data that will be given with a good reliability. One of the great advantages of ANN is that they can be applied to a wide range of problems, and they can analyze different input typologies, such as pictures, database, videos, and files. In this work, the input is represented by a database in which there are substrates of different nature, and their features are listed in terms of C, H, O, N, S and macromolecules. All these information were derived from various paper and subsequently carried over into the database of interest.

A very particular aspect associated with neural networks is that, very often, its functioning is compared to the one of human brain. Which are the similarities that makes this comparison valid? The small units that compose the algorithm are called artificial neurons that are mathematical models of biological neuron. The human neuron, if observed by a simplified point of view, is characterized by three components: the synapses, the nucleus and the axon. The signals pass from one neuron to the next through the synapses. The synapses are the contact point between two neurons whose thickness can change with the passing of time, causing so a strengthening or a weakening of the connection. Once these information enter inside the nucleus, they are processed in order to generate new signals. After that, these new signals pass through the axon moving towards another nucleus. The functioning of artificial neuron is exactly like this; in fact, the units of each single layer are in contact with those of the previous and following layer by means of links. In the same way as happens in the synapses, through these connections occurs both the information transfer but also the influence of the intensity with which the signal will be transmitted. Furthermore, inside these units, there will be a specific function that elaborates the inlet information generating the outgoing ones. For all these reasons, the SNN can be considered as a computational model inspired by the human brain functioning.

4.2. Theory underlying the functioning of neural network

At this point, in order to better understand the theory behind the neural networks, let's analyze the operation of a single neuron:
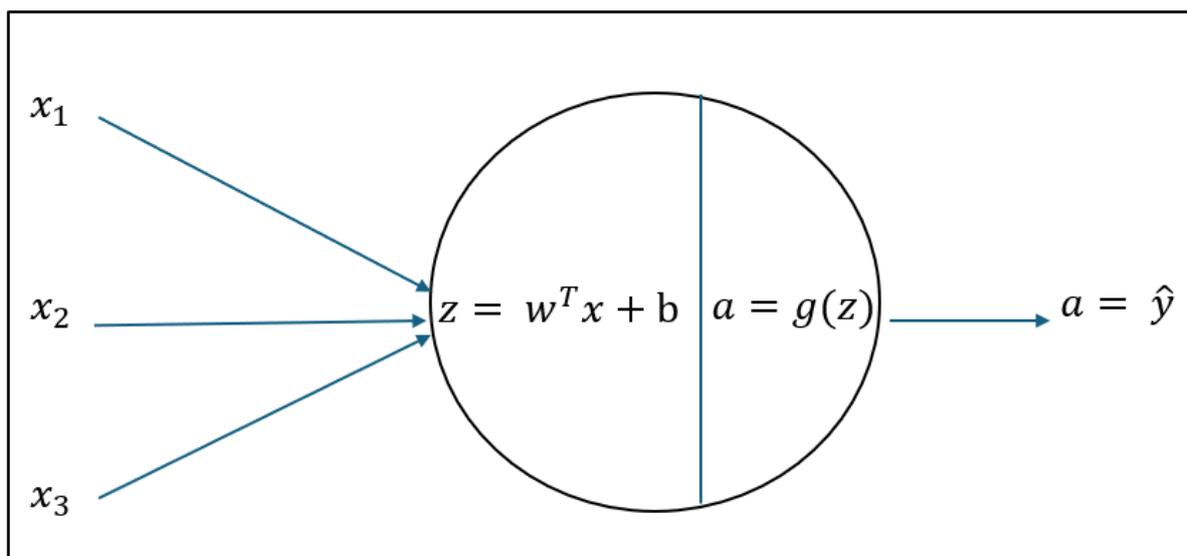
Figure 23: Simplified illustration of a single artificial neuron.

Assuming to feed to the unit a single sample, it's important to explain the meaning of the different terms represented in the figure above:

- x_i are the various features that characterize the sample. They will help the algorithm to realize the classification problem; in fact, there is a specific association between these features and a class y;
- w^T is the transpose of weights vector. It represents the first of the two parameters that can influence the intensity of the information transmitted between two units. Each single feature x_i is characterized by a specific weight that denotes its relative importance;
- b is called "bias" and it is the second of the two parameters described previously. It is very important because it allows a good adaptation of the model to the most complex data relationships;
- z is defined as "pre-activation parameter". It is defined in this way because it represents the input data given to the activation function a to obtain the prediction $\hat{y}$;
- The activation function a is a term of fundamental importance for the neural network; in fact, one of the most important advantages associated with the activation function is the introduction of non-linearity inside the model. Without a, the ability of the algorithm to learn complex data relationships would be very limited since it would be a simple linear combination of inlet information. In the same way as the choice associated to the number of the

layers, also the activation function adopted inside the ANN is an important variable that can influence the performance of the model. The different layers of the model require specific typologies of a.

- $\hat{y}$ is the prediction of the algorithm and it represents the belonging probability of a single sample with respect to a class of the classification problem.

After the description of all the previous terms, it is possible to note that a single artificial neuron is defined by two calculation steps:

- The first is the estimation of the pre-activation parameter through the following equation:
$$z = w^T x + b \tag{4.1}$$
- The second is the achievement of the prediction through the activation function:
$$\hat{y} = a = g(z) \tag{4.2}$$

Obviously the previous two stages are present inside each unit of the model. At this point, having understood how works a single neuron, it is possible to shift the focus to the global SNN:
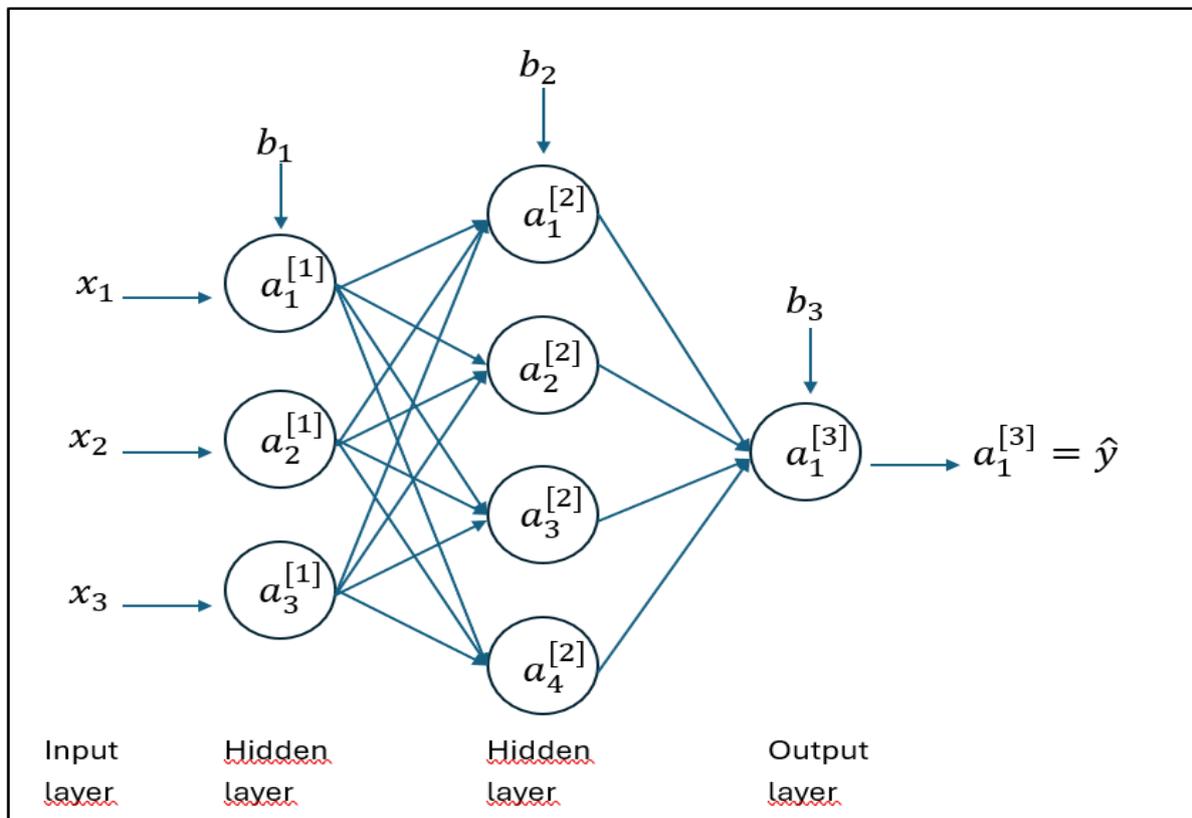

Figure 24: Simplified illustration of the global neural network.

All the aspects previously described for the single neuron remain valid for the global network. In addition, it is necessary to highlight further features. First of all, it is important to say that the ANN taken into consideration has 4 layers: the input layer, the two hidden layers and the output layer. This structure is called "three layers neural network"; in fact, in the counting of the layers, the input one does not play an important role. Before going to describe the properties of the different layers, it is important to make considerations concerning the b and w parameters. Regarding the first one, the bias is equal for all the units belonging to the same layer. This means that every single artificial neuron of the first layer is characterized by the same bias, defined as b_1, and the same holds true for b_2 and b_3. Considering now the second parameter, it is important to point out that every connection between the neuron of a layer and that of another will have a different weight with respect to those that characterize the other links. This means that, considering for example the signal coming from a single unit of the first hidden layer, this information will be sent to all the neurons of the second hidden layer with a different intensity depending on the relative importance of the specific connection. At this point, focusing on the three typologies of layers, it is important to highlight the following aspects:

- input layer is the first layer of the ANN and it accepts the data that must be analyzed still in a rough form. The number of its units is fixed; in fact, it must be equal to the number of features that characterized the analyzed sample. This means that, taking into consideration the figure 24, the sample has three features. The neurons of the input layer are not characterized by an activation function; in fact, their aim is to transfer the information to the first hidden layer without applying any transformation;

- hidden layer is the real typology of layer with which it is possible to play to reach the optimal architecture of the model. There are two variables that is possible to change. The first is represented by the number of hidden layers and it helps to define the depth of the neural network. The second is the number of units for every single layer and, through this variable, it is possible to create a big or small SNN according to the number of the artificial neuron that are used. The choice of these two variables depends on several factors such as: the complexity of the case study, the dimensions of the database and the number of the features of the single sample. A wrong choice can lead to the problem of overfitting. This problem occurs in correspondence of an exaggerated number of layers or units. The overfitting is an undesired behavior that happens when the algorithm of machine learning adapts itself too well to the set of the training data. On the other hand, it is not able to return the same level of accuracy in the

predictions obtained through new data. In this case, in order to be configured, the neurons of this layer require a specific activation function;

- output layer is the last layer of the ANN. Also in this case the number of its units is fixed. This amount is defined by the nature of the classification problem taken into consideration. Indeed, in the case of a binary classification, this layer requires only one unit. On the other hand, when the problem is characterized by more than two classes, also known as multiclass classification, the layer requires an amount of neurons equal to the number of the categories into which the output values are divided.

At this point, before going to analyze the neural network of this dissertation, the last aspect that needs to be examined is the optimization process. Through this step, with the passing of iterations, the SNN is able to generate some predictions $\hat{y}$ that are increasingly similar to the real labels y. The heart of this speech is related to the w and b parameters. Initially, the different weights and bias will have randomized values. This is because, at the beginning, it is not possible to know which w and b will be able to generate accurate predictions. Therefore, the various artificial neurons must learn to calibrate these two parameters to optimize the ANN performance. In order to achieve this goal, it is used a specific function and its name is "cost function". Below two cost function are introduced: one is used for the binary classification problems, while the second is used for the multiclass ones:

1. Binary Crossentropy:
$$C(y,\hat{y}) = -(y log(\hat{y}) + (1-y)\log(1-\hat{y})$$
$$C(y,\hat{y}) = -(y log(f(w,b)) + (1-y)\log(1-f(w,b)) \tag{4.3}$$

2. Categorical Crossentropy:
$$C(y,\hat{y}) = -\sum_i y_i \log(\hat{y}_i)$$
$$C(y,\hat{y}) = -\sum_i y_i \log(f(w,b)) \tag{4.4}$$
The subscript i indicates the different classes of the classification problem.

The cost function is defined with respect to a single training example. For this reason, in such a way as to monitor the performance of the model, it is used the loss function L that take into accounts the entire set of training data:

$$L(w,b) = \frac{\sum_{i=1}^{m} C(y^i,\widehat{y^i})}{m} \tag{4.5}$$

m indicates the number of the sample used for the training.

Therefore, the aim of this step is to find the value of w and b that are able to minimize L. The minimization of loss function permits to reduce the difference between $\hat{y}$ and y obtaining in the end an optimized neural network. This operation of optimization is realized through the backpropagation algorithm expressed in this way:

$$w_{i+1} = w_i - \delta_w a \qquad (4.6)$$

$$b_{i+1} = b_i - \delta_b a \qquad (4.7)$$

In which:

- w_{i+1} e b_{i+1} represent the w and b values of the next iteration;
- w_i e b_i are the parameters related to the current iteration;
- δ_w e δ_b are the w and b derivatives with respect to the loss function;
- a is defined as "learning rate"; it is an hyperparameter that control the rate at which the model adapts itself to the training data. For small value of a the convergence towards the minimum of L will be slow, while the larger a is, the faster the convergence will be. At the same time, if a is too large, there will be strong fluctuations.

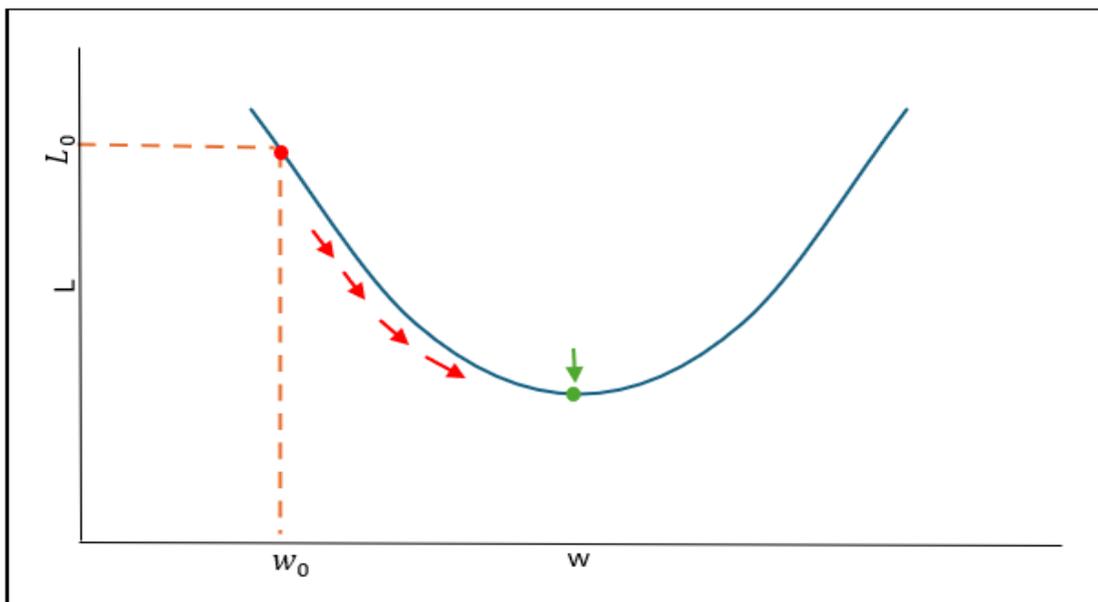Observing the behavior of the loss function with respect to w, it is possible to have the following trend:

Figure 25: Behavior of L with respect to w during the backpropagation algorithm.

The same chart can be obtained with the parameter b. L_0 and w_0 represent the loss function and the weights of the first iteration. It is possible to note that, due to the randomization of w_0 and b_0, L is characterized by an initial value that is high. This implies that the predictions will be very different with respect to the real labels; in fact, with the passing of iterations and through the backpropagation, it will be possible to be closer and closer to the minimum of L, represented by the green point. Once this minimum is achieved, the obtained values of w and b are those that allow the model to generate accurate predictions.

4.3. Neural network of interest

At this point, once the SNN theory has been analyzed, it is possible to describe the neural network created for this dissertation analyzing both its components and its performance.
The first important aspect is related to the database of interest. This database contains 216 samples, that are different typologies of substrates: manures of various nature (such as pig manure, chicken manure and goat manure), fruit and vegetable waste, food waste and so on. The properties of each substrate are defined by 19 features such as: the composition in terms of the C, H, O, N , S, and macromolecules and stable parameters like C/N. The most crucial information is the one related to the number of classes used to split the substrates. The labels of this case study are the following:

1. Zootechnical waste;
2. Agricultural waste;
3. Organic waste;
4. Sludge waste.

This means that this problem is characterized by a multiclass classification. After this introduction about the properties of the database, it is possible to analyze the most significant steps related to the creation of the neural network exploiting the software Python. Once all the libraries have been imported, such as TensorFlow that is fundamental for the SNN creation, and once the database has been uploaded, it is necessary to choose the criteria according to which the splitting between the set of training data and that of test data occurs. The first set is the one characterized by those substrates that allow the model to learn the correlations between the several features of the sample. On the other hand, the substrates of the second set will be used to evaluate the accuracy of the SNN with respect to data that are different with respect to those of training. The aim of this check is to evaluate the potential presence of an overfitting problem. Therefore, in order to realize this splitting, it has been chosen to

use for the set of test data only the 20% of all the substrates, while the remaining 80% will be used for the training as shown in the line code below:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20,
random_state = 42)
```

It has been selected this proportion because, after trying out several combinations, it was found that this splitting was able to guarantee better performance. Indeed, it is true that increasing the size of the set of test data it is possible to obtain a stronger model, but, on the other hand, if the other set becomes too small, the learn capability of the SNN decreases. This is a typical example of how, in the construction of an ANN, there are not variables value that can be considered generally as the best; in fact, the best solution depends only on the specific case study taken into consideration.

Once an appropriate splitting has been carried out, the next step is the one related to the creation of the model. The architecture shows the following properties:

```
model = keras.Sequential([

    layers.Dense(9, activation='relu', input_shape = [19]),

    layers.Dense(8, activation='relu', kernel_regularizer=l1(0.01)),

    layers.Dense(4, activation='softmax')

])
```

There are different aspects on which it is important to focus the attention because they significantly influence the accuracy of the model. The first one is represented by the nature of the layers; in fact, the option "layers.Dense" permits to adopt one of the most common layer typologies, i.e. the dense layers. This feature ensures that each neuron of the neural network would be connected to all the units that compose the previous layer. One of the great advantages associated to their use is represented by the easiness in the information propagation through the SNN; in fact, there is a strong connection between the various neurons. In addition, the dense layers are able to learn in an efficient way the complex data relationship. Regarding the ANN configuration, it has been chosen the one characterized by the input layer, two hidden layers and the output layer. This means that it is a three layers neural network since the input one is not taken into consideration. In the part of code reported above, it is possible to note the presence of some properties that are coherent with what has just been said. In fact, the input layer is characterized by a number of units that is equal to the amount of features of the single substrate; this is expressed through the option "input_shape". Furthermore, the other layer that contains a number of artificial neuron that is fixed by the problem is the output one. Indeed, it has 4 units like the 4 classes of the case study. The mathematical meaning of having 4 units is the following: the neural network

returns as output value the prediction $\hat{y}$. It represents the probability that the substrate has to belong to the different labels. This means that, for this problem, $\hat{y}$ will be a vector of 4 values and each of them indicates the sample belonging probability with respect to the single class. Therefore, adding these 4 probabilities, the result should return a probability of 100%. This is the reason thanks to which, in the multiclass classification, the output layer must be characterized by a number of units equal to the amount of the different classes. A very important role is also played by the choice of the activation functions. As explained previously, only the input layer is not characterized by an a, while the hidden and output layers require two different activation functions. For this problem, the former needs the so called "Relu activation function", while the latter is characterized by the "Softmax activation function". These $g(z)$ present the following properties:

- Relu: it stands for Rectified Linear Unit. It is the most common activation function to be used for the hidden layer. Compared to others, it guarantees a faster learning. In fact, using other activation functions like the sigmoid or the hyperbolic tangent, due to their definition, if the z of interest should became too big or too small, then the gradient of the function approaches zero causing a slowdown or even a termination of the backpropagation process. The gradient is represented by δ_w e δ_b as shown in the equations 4.6 and 4.7. This phenomenon does not occur with the Relu; in fact, it is defined in this way:

$$Relu(x) = \max(0, x) = \begin{cases} x \ if \ x > 0 \\ 0 \ otherwise \end{cases} \tag{4.8}$$

  Relu returns x if x is positive, otherwise it returns zero. Therefore, it is not possible to have the problem described before in the positive regions; in fact, when the input is positive, the gradient will be always equal to 1 avoiding to obstruct the learning process of the model. Associated to this a, there is the problem of "dead neuron" that occurs in correspondence of negative input; in fact, for this value, the neuron returns always zero. Despite this problem, Relu represents a good activation function for the hidden layers;
- Softmax is the typical activation function for the output layer when the case-study is characterized by a multiclass classification; in fact, it allows to convert the output value in a probability distribution over the 4 labels.

At this point, the last aspect of the model architecture that needs to be analyzed with attention is the choice of the amount of units within each hidden layer, subsequently defined as L1 and L2. In order to reach the best configuration, it has been necessary to

evaluate the performance obtained through different combinations of artificial neurons that characterize L1 and L2. The original configuration was composed by 3 hidden layers, in which the first contained 32 units, while the other two 16. Unfortunately, an overfitting problem has emerged since the amounts of layers and units were too large. Indeed, after a specific number of iterations, the accuracy of the SNN was able to reach an accuracy of 100% after the training step, so this meant that the model adapted itself perfectly to the given information. However, observing the performance obtained through the set of test data, the predictions reached a level of accuracy equal to 50%. This is the real definition of overfitting problem, so a great adaptation only in relation to the set of training data. At this point, it was necessary to modify the architecture of the model to obtain two similar accuracies even at the expense of losing something in terms of performance related to the training step. Going in order, the following changes have been made:

1. The first features subjected to a variation was the number of hidden layer, moving from 3 to 2. Only with this change, the accuracy related to the set of test data has increased to a value of 60%, but, at the same time, the other accuracy was close to 100%.

2. Therefore, at this point, it was necessary to play with the amount of artificial neurons inside each hidden layer in order to avoid the overfitting problem. Through different analysis, it was possible to note that, putting 5 units inside L1 and L2, both the two accuracies have been subjected to a significant decrease. In fact, with a too small number of units, the model develops a very limited learning capability. For this reason, as lower boundary for the number of neurons inside each hidden layer, it has been considered a value of 6. Regarding the upper boundary, the chosen value has been 16.

   For all the reasons explained above, the range of interest regarding the number of units within L1 and L2 was between 6 and 16. At this point, analyzing all the possible combinations inside this range, it has been possible to reach the best configuration: i.e. 9 artificial neurons in L1 and 8 in L2.

The motivations related to the previous choice are clearly shown in the figure 26. In order to understand this plot, it's correct to highlight the following aspects:

- The analyzed combinations are always characterized by a higher number of units in L1. This is due to the fact that, generally, the complexity of the information is higher in the first hidden layers; in fact, proceeding towards the

output layer, the information becomes simpler and so it can be used a lower amount of artificial neurons;

- Each combination is characterized by 10 implementations of the model in order to obtain the performance that the SNN was able to guarantee averagely with that architecture. At the end of these implementations, 4 parameters has been analyzed:

  1. Average train accuracy;
  2. Average test accuracy;
  3. Average difference between the train and test accuracy. This is the term that is most closely related to the potential overfitting problem;
  4. Average standard deviation with respect to the difference between the train and test accuracy. The goal is to achieve a model that allows to obtain similar results after every single implementation since the splitting of the database in the two set is randomized. This means that it is not possible to obtain every time the same identical results.

After these premises, it is possible to analyze the following plot:

Figure 26: Results of the neural network performances associated to the different combinations of artificial neuron inside of L1 and L2.

For greater clarity, below it is reported the table that contains the results associated to the chosen combination:

AVERAGE TR. ACC.	AVERAGE TEST ACC.	AVERAGE DIFF.	AVERAGE STD. DEV.
0.656	0.592	0.064	0.049

Table 5: Values of the previous four parameters related to the chosen configuration.

The most significant parameter is the third one; in fact, on average this combination can guarantee a small difference between the two accuracies. Thanks to this result, it is possible to say that now the model can reach a good adaptation to the two set of data. Observing the values associated to the first and second parameter, it is plausible to doubt that the model performances are not so optimal, because they are in the range of 60%-65%. On the other hand, thinking about the nature of the available data, these results are more reliable with respect to the ones reached with architecture characterized by an accuracy higher than 90%. In fact, these data are very variable in which it is difficult to find regular behaviors because they are information related to organic substances. In conclusion, also the value of the fourth parameter is significant; in fact, during the 10 implementations, it is possible to ensure a similar difference between the two accuracies. Among all the combinations, this architecture is the one characterized by the second smallest value regarding the fourth parameter. In fact, the model characterized by the smallest one is the combination in which L1 is characterized by 13 units and L2 by 11. But this last case has a very high value regarding the third parameter, i.e. 0.233. All these explanations permit to point out the fact that the structure with 9 neurons in L1 and 8 in L2 is a good choice to obtain reliable predictions. One last feature that is very important to highlight regarding the architecture of the model is represented by its amount of w and b parameter. This neural network is defined by 296 parameters. Therefore, the optimization process must be able to estimate the best values of this large amount of w and b to reach optimal performance of the model. This last consideration permits to underline the typical complexity of a SNN structure.

Once have been described all the aspects that characterize the architecture of the neural network, the next step is the one regarding its compilation. Its aim is to provide to the model the instructions on how to train and update:

```
model.compile(optimizer='Adam',    loss='categorical_crossentropy',    metrics
=['accuracy'])
```

The first term to describe is the one related to the optimizer. It has been used the "Adam optimizer", also known as "Adaptive Moment Estimation". It is the most used and it is a combination of two others optimizers: Momentum and RMSprop. Among the different advantages associated with Adam, it's correct to highlight that it can lead to a faster convergence since it is able to reach a good adaptation also when parameters with different orders of magnitude are used. The second term is related to the Loss function. Coherently with what has been said before, since this case study is characterized by a multiclass classification, it has been used the Categorical Crossentropy. The third term indicates the evaluation metrics used to monitor the performance of the model, such as the accuracy.

After the model compilation, the next step is represented by the implementation of the training stage of the ANN:

```
model.fit(X_train, y_train, epochs = 2000, batch_size = 64)
```

Obviously, it has been used the set of training data staring a sufficiently high number of iterations to ensure good performances. The third parameter is the most important one. For this reason, it is necessary to explain its meaning to understand its role played during this step and also to justify its value of 64. The batch size indicates the amount of examples coming from the set of training data that are used, at each iteration, to realize the learning of the SNN. This means that, during the optimization process, instead of take advantage of the entire training set simultaneously, only its subsets are used. The value given to the batch size can significantly influence the goal of reaching the minimum of L; in fact, the differences between two batch characterized by different dimensions are the following:

- Batch with big dimensions involves the achievement of the minimum of L with an higher accuracy even if with slower process; in fact, it is receiving an huge amount of data and so the computational costs are significant;
- Batch with small dimensions allow to speed up the process because they work with a lower amount of data. One disadvantage is that there are strong fluctuations along the backpropagation; in fact, providing less data, the neural network knows a lower amount of information regarding the database and so, in this case, it is common to say that the gradient has more variability. In addition, if the dimensions are too small, it becomes impossible to reach the minimum of the loss function.

The typical values regarding the batch size are: 32, 64, 128 and 256. Let's take as examples two batch, in which one has a size equal to 32 and the other one equal to 64.

Coherently with what has just been said, with the former the update of the model occurs after the calculation of the gradients over 32 samples, while with the latter over 64 examples. Therefore, this means that with the second there will be a slower and more stable process. On the other hand, compared to a batch with size equal to 128, the batch of interest is characterized by a faster optimization process but with greater fluctuations. Therefore, after the considerations reported above, for this neural network it has been chosen a batch size equal to 64 to reach a trade-off between the various advantages and disadvantages. It has been evaluated also the size equal to 128 but, for this case study, the fluctuation component does not change in a significant way moving from 128 to 64 samples. In addition, as will be seen shortly, 64 samples are enough to reach the minimum of L without any problems.

At this point, after the training of the SNN, the last step is represented by the evaluation of the accuracy that the model has with respect to the set of the test data. This stage is very important because it allows to check the potential presence of an overfitting problem by comparing this last accuracy with the accuracy related to the set of training data:

```
test_loss, test_acc = model.evaluate(X_test, y_test)

print(f'Test accuracy: {test_acc}')
```

After the evaluation of this second accuracy, the creation of the neural network is completed. Before going to analyze for which purposes this model can be used, it is better to assess if what has been shown until now has validity:

Figure 27: Behavior of the accuracy related to the set of training data as a function of the iterations.
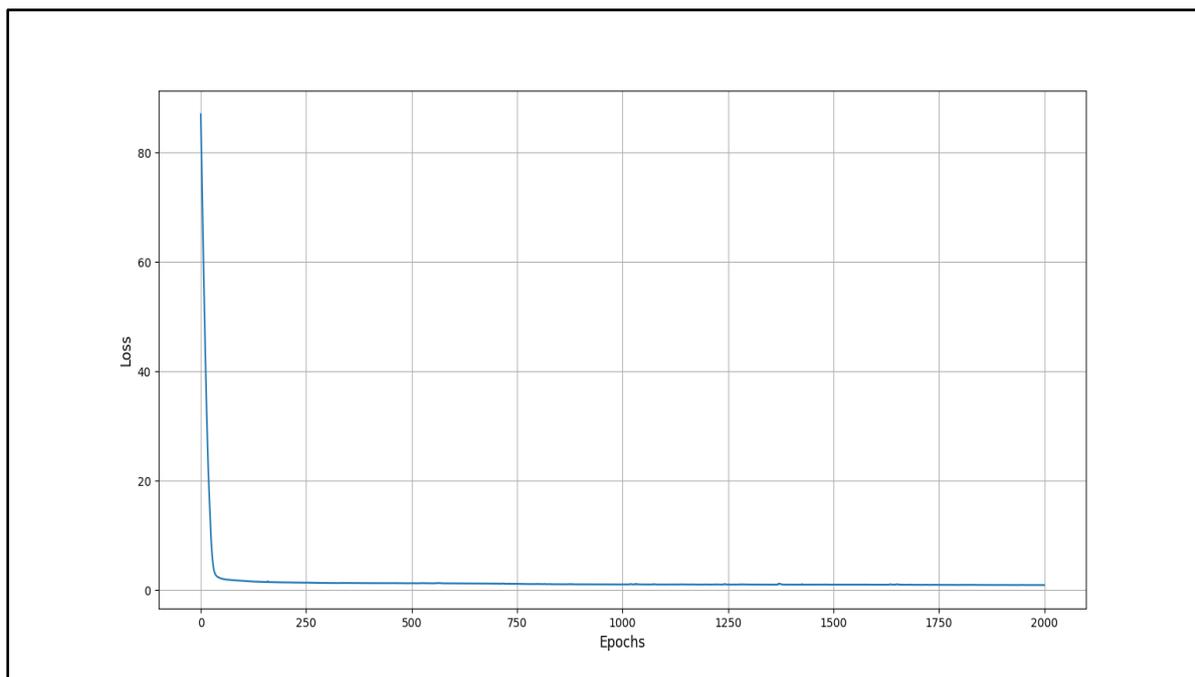
Figure 28: Behavior of the Loss function as a function of the iterations.

From the plots reported above, it is possible to argue that the adopted considerations have been correct; in fact, reviewing individually the two graphs, the following observations emerge:

- About the first figure, this implementation has reached a level of accuracy close to 70%. In the achievement of this final value, the displayed trend is affected by moderate fluctuations. Their presence is caused by the nature of the available data; in fact, they are information coming from organic substances in which it is arduous to find persistent correlations between the various properties of the substrates. It is possible to notice that a significant part of the learning happens in the first epochs;
- About the second figure, it is possible to observe that the Categorical Crossentropy can reach its minimum after few iterations without difficulty.

Therefore, because of the obtained results, it is correct to assume that this neural network is able to elaborate good predictions for whatever will be provided.

4.4. Neural network application

At this point, it could be interesting to highlight a possible application of this model, especially as a function of the role played by the SNN within the final algorithm of this dissertation. As explained before, considering a single sample, the $\hat{y}$ returned by the model will be a vector of four values that denote the substrate belonging probability to each of the four classes. This means that, providing a set of values associated to a single feature, such as C/N, it could be possible to identify the label characterized by the higher belonging probability associated to the specific sample. Considering for example a range of C/N values between 5 and 10 spaced by 0.5, so a range characterized by small values of this parameter, the neural network would elaborate the following output:

C/N	Prob.Cl1	Prob.Cl2	Prob.Cl3	Prob.Cl4
5	0.345	0.155	0.43	0.07
5.5	0.337	0.157	0.436	0.07
6	0.331	0.168	0.441	0.06
6.5	0.292	0.202	0.446	0.06
7	0.291	0.199	0.45	0.06
7.5	0.287	0.2	0.453	0.06
8	0.277	0.217	0.456	0.05
8.5	0.273	0.219	0.458	0.05
9	0.267	0.225	0.458	0.05
9.5	0.258	0.232	0.46	0.05
10	0.255	0.235	0.46	0.05

Table 6: Values regarding the belonging probability to the four classes as a function of the single C/N parameter.

From the table shown above, it is possible to note that the class 3 is characterized by the highest belonging probability going from values of 43% to 46%. This means that the substrates with a C/N between 5 and 10 have a great chance to be part of organic

waste. These results are quite coherent if compared to the properties that characterize the different substrates within each of these classes; in fact, for example, the class 2 has a lower probability since it is composed by wastes with a high percentage of lignin, such as straw. This implies a high content of C leading to an high C/N value. On the other hand, these predictions should not be taken as 100% true for the following two reasons:

1. They are probabilistic values coming from the learning of a SNN on the basis of variable data;

2. They are obtained by exploiting a single parameter such as C/N; in fact, it is possible to find substrates that, despite having a C/N between 5 and 10, in the reality they belong to a different class rather than the organic waste one. This is because, for example, that specific sample is characterized by a level of biodegradability or a percentage of macromolecules that are more typical of another label.

Despite these warnings, it is correct to evaluate this neural network as a strong and useful tool; in fact, in a complex world like the one of organic substances, this SNN is able to realize accurate correlations between the various properties of the substrate. At this point, in the next chapter, it will be described the second tool needed to reach the second goal of this dissertation.

5 Feedstock Blending Optimization FBO2

5.1. Introduction to FBO2 and its purposes

Once have been created the neural network, it is possible to analyze the second tool needed to reach the other great goal of this dissertation. As just been said in the previous chapter, this second objective consists in the creation of an algorithm that allows to define the specific conjugated substrate able to maximize the biomethane yield through the co-digestion with a predefined biomass.

The second created tool is called FBO2. FBO stands for "feedstock blending optimization", while the 2 is because it started from an old model based only on the stability parameters of the C/N and of the biodegradability. The FBO2 can achieve these two targets:

- After the definition of two substrates, subsequently named as S1 and S2, it is possible to obtain all the possible combinations of them that satisfy the conditions imposed by the following four stability indicators:

    1. C/N;
    2. Biodegradability;
    3. FOS/TAC;
    4. OLR.

- After the previous step, the model find the composition that permits to maximize the BMP.

In order to understand better what has been said before, it could be useful to observe and subsequently to analyze the scheme reported on the following page. It highlights the main steps that are performed by the FBO2:

Figure 29: Simplified illustration of the structure of FBO2.

Therefore, by observing the figure 29, it is possible to obtain a greater knowledge regarding what the model can do, i.e.: once have been defined the two substrates of interest, it is feasible to work by exploiting only the composition of S1 thanks to the conservation principle:

$$w_2 = 1 - w_1 \tag{5.1}$$

At this point, it occurs the definition of the initial vector x_0 that is characterized by all the possible compositions related to the first biomass:

$$x_0 = [0, 0.01, 0.02, \dots \dots, 1] \tag{5.2}$$

This x_0 represents the starting point towards the estimation of the optimal composition of the mixture capable to maximize the biomethane yield. It is possible to achieve this final goal through a series of filtration operations applied to the initial vector. In these filtration operations, the use of the different mixture equations takes place in order to calculate the previous four stability parameters:

$$C/N_{mix} = C/N_1 * w_1 + C/N_2 * (1 - w_1) \tag{5.3}$$

$$BD_{mix} = BD_1 * w_1 + BD_2 * (1 - w_1) \tag{5.4}$$

$$FOS/TAC_{mix} = 0.0332 + 0.267 * w_1 * FOS/TAC_1 + 0.635 * (1 - w_1) * FOS/TAC_2 + 3.23 * w_1 * (1 - w_1) * FOS/TAC_1 * FOS/TAC_2 \tag{5.5}$$

$$LOAD_{mix} = M * (w_1 * TS_1 * VS_1 + (1 - w_1) * TS_2 * VS_2)/100/100 \tag{5.6}$$

$$OLR_{mix} = LOAD_{mix}/HRT \tag{5.7}$$

Regarding the equations reported above, it is advisable to highlight the following considerations:

- All the parameters characterized by the subscript 1 are properties related to the biomass 1. The same holds true for the second substrate;
- The term $LOAD_{mix}$ represents the amount of the volatile solids of the two biomasses fed to the reactor. It is fundamental its estimation to achieve the organic loading rate of the mixture. In the equation 5.6 there is also the influence of the total solids because, in the database of interest, the volatile solids are expressed in terms of percentage with respect to the TS;
- In the last two equations, there are two important assumptions:

  1. The M indicates the kg of biomass that is used and, in this case, it is equal to 1000 kg;
  2. Regarding the hydraulic retention time, it is assumed a value equal to 35 days.

At this point, going back to the filtration operations, it is important to understand what they allow to do within the FBO2. Starting from the initial vector x_0, the first filtration operation is the one related to the C/N. In this case, it occurs the calculation of the different C/N_{mix} associated to all the w_1 contained within x_0. Once have been obtained all these C/N_{mix}, it takes place the real filtration operation. Indeed, considering a specific range of acceptable values for the C/N_{mix}, there will be the creation of the vector x_1 through the comparison between the C/N_{mix} calculated and those of the range mentioned before. This range is chosen by taking into consideration as lower and upper boundaries those value of C/N where the inhibition of the process can occur leading to an unstable operation of the anaerobic digestor. Therefore, inside x_1, there

will be only those compositions of S1 that respect the stability condition related to C/N_{mix}. This means that, considering for example an upper boundary equal to 40, if the w_1 causes the achievement of a C/N_{mix} of 45, then this w_1 will not be included within the new composition vector and consequently it will not be subjected to the next filtration operations. Indeed, as displayed in the figure 29, there are other three filtration operations, each associated to one of the three remaining stability parameters. These three operations will work in the same way as the one of C/N_{mix} but obviously with a different range of acceptable values. The last step of this process of filtrations is associated with 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 introduced previously.

At this point, once have been obtained this final composition vector, it occurs the second great aim of this model, i.e. the estimation of the specific mixture composition with which the maximum yield of biomethane is connected. Therefore, for each w_1 inside x_4, it is calculated the associated BMP by exploiting the following equation:

$$BMP_{mix} = w_1 * BMP_1 + (1 - w_1) * BMP_2 + w_1 * (1 - w_1) * BMP_{syn} \qquad (5.8)$$

In which:

$$BMP_{syn} = 21.6613 + 1.2558 * \left(\frac{C}{N}\right)_{mix} + 445.7076 * BD_{mix} - 0.0223 * \left(\frac{C}{N}\right)_{mix}^2 -$$
$$7.8201 * BD_{mix}^2 \qquad (5.9)$$

Therefore, after this step, it will be available a series of BMP_{mix} and, by identifying its maximum value, it will be possible to extrapolate the associated composition that can maximize the biomethane yield.

5.2. Analysis of the filtration operations implemented on Python

Regarding the creation of this model, it has been used Python as software. Below, the lines of code related to the filtration operations are reported. This could be useful to have a further explanation about the functioning of the model if observed by the point of view of the Python programming. The first reported part is the one related to the C/N filtration operation:

```python
x0 = np.linspace(0, 1, 101)

CNmix = np.empty(len(x0))

CN1 = S1['CN']

CN2 = S2['CN']

x1 = []

i = 0

for x_0 in x0:

    CNmix[i] = CN1*x_0 + CN2*(1-x_0)

    if 10 <= CNmix[i] <= 40:

        x1.append(x_0)

    i += 1
```

Coherently with what has been said before, it is possible to point out the following aspects:

- In the first row there is the creation of the initial vector x_0 characterized by all the possible compositions of the first substrate;
- Through the for loop, it is possible to calculate the C/N_{mix} for each value of w_1 inside x_0;
- In order to estimate C/N_{mix}, it has been used the equation reported previously;
- The range of stability is characterized by C/N_{mix} values between 10 e 40. Indeed, for values lower than 10 there is a very high content of ammonia leading to the inhibition of the microorganisms growing. On the other hand, for C/N_{mix} higher than 40, the high content of carbon can induce a critical increase in terms of volatile fatty acids production;
- The filtration operation is realized through the if loop. In fact, in this loop, for each value of x_0, there is the comparison between the stability range and the obtained C/N_{mix}. When this value is between 10 e 40, thanks to the function "append", the associated w_1 will be added to x_1. As said before, this term represents the new composition vector of S1 that contained all the w_1 that have respected the first stability condition. Thereby, it will be submitted to a new filtration operation.

Regarding the other filtration steps, they are characterized by the same structure of the previous one but obviously with a different stability range. Indeed:

```
BDmix = np.empty(len(x1))

BD1 = S1['BD']

BD2 = S2['BD']

x2 = []

i = 0

for x_1 in x1:

    BDmix[i] = BD1*x_1 + BD2*(1-x_1)

    if 0 <= BDmix[i] <= 1:

        x2.append(x_1)

    i +=1
```

- BD filtration operation: in this case the stability range is between 0 and 1. A biomass with a biodegradability higher than 1 has no physical meaning. In fact, BD is defined as the ratio between the real biomethane yield and the theorical one obtained through the Buswell relation in which it is not considered the non-degradable fraction of the substrate.

```
FTmix = np.empty(len(x2))

FT1 = S1['FOSTAC']

FT2 = S2['FOSTAC']

x3 = []

i = 0

for x_2 in x2:

    FTmix[i] = 0.0332 + 0.267*x_2*FT1 + 0.635*(1-x_2)*FT2 + 3.23*x_2*(1-x_2)*FT1*FT2

    if 0.2 <= FTmix[i] <= 0.7:

        x3.append(x_2)

    i +=1
```

- FOS/TAC filtration operation: in this case the stability range is between 0.2 and 0.7 for the reasons explained in the chapter 2.

```
M = 1000       #kg of biomass

HRT = 35       #days

Loadmix = np.empty(len(x3))
```

```
OLRmix = np.empty(len(x3))

TS1 = S1['TS']

TS2 = S2['TS']

VS1 = S1['VS']

VS2 = S2['VS']

x4 = []

i=0

for x_3 in x3:

    Loadmix[i] = M * (x_3*TS1*VS1 + (1-x_3)*TS2*VS2)/100/100

    OLRmix[i] = Loadmix[i]/HRT

    if 2.5 <= OLRmix[i] <= 6:

        x4.append(x_3)

    i +=1
```

- OLR filtration operation: in this case the stability range is between 2.5 and 6 for the reasons explained in the chapter 3.

As it is possible to observe in the rows of code related to the OLR filtration operation, it is obtained the final vector x_4. It contains all the possible compositions of S1 that are able to respect the four stability conditions. At this point, it is calculated the value of BMP_{mix} for each w_1 of x_4 thanks to the equation 5.8. Then, through specific instructions implemented on Pyhton, it is possible to get both the composition of the first biomass related to the maximum value of BMP_{mix} and the w_2 by means of the conservation equation. It is important to highlight that the results of FBO2 depends on the information extracted from a specific pad. Starting from the database used in the previous chapters, for each typology of substrate, the average value related to each property of the biomass is reported inside this pad. Therefore, considering e.g. the pig manure, in the starting database there are 15 samples of this substrate. Consequently, in the pad there are the average values of all those features that permits to represent averagely the properties of the pig manure. The same holds true for all the others type of biomass.

5.3. Analysis of the results provided by the FBO2 with respect to a specific co-digestion

Before moving to some industrial validations of the model, it could be interesting to observe and analyze the nature of the results that the FBO2 is able to return by

considering a specific co-digestion of two substrates. These two biomasses are simply chosen in order to have a practical example through which it is possible to evaluate the potentialities of the model.

The example taken into consideration is related to the co-digestion of fruit waste and chicken manure, therefore:

- S1: fruit waste;
- S2: chicken manure.

This co-digestion could be very noteworthy because it is characterized by two substrates that are affected by some properties that allow to achieve a good synergy and, consequently, an acceptable yield of biomethane. In fact, S1 presents a higher sugars concentration with respect to the one related to the proteins causing so a fast acidification that leads to the inhibition of the methanogenic activity. On the other hand, S2 is generally defined by a greater N concentration thanks to a high level of proteins. This aspect could counteract the possible inhibition of microorganisms' growth caused by the presence of S1. Therefore, thorough the FBO2, it is possible to achieve the following results:

Figure 30: BMP vs w1 plot associated to the co-digestion of fruit waste and chicken manure.

w1,max	w2,max	BMPmax [mL/gVS]
0.61	0.39	382.40

Table 7: Values of the optimal composition and the associated BMP for to the co-digestion of fruit waste and chicken manure.

C/Nmix,max	BDmix,max	FOS/TACmix,max	OLRmix,max [kgVS/m3/day]
12.687	0.678	0.258	3.511

Table 8: Values of the mixture parameters for the optimal composition in the co-digestion of fruit waste and chicken manure.

Starting from the figure 30 and the table 7, it is important to point out the following aspects:

- The maximum value of BMP obtained through this co-digestion is equal to 382.40 [mL/gVS];
- The optimal composition that allows to achieve the previous BMP is the one characterized by a 0.61 for the fruit waste and 0.39 for the chicken manure;
- The green arch represented in the figure 30 is related to the BMP values obtained starting from those compositions that have overcame the four stability conditions. The green point indicates the maximum amount of the biomethane yield;
- On the other hand, the arches dashed in red are related to BMP values calculated through the compositions that have been excluded from x_4 by means of the filtration operations.

Regarding the table 8, there are the mixture parameters related to the optimal composition. It is possible to note that all the four indicators are inside their respective stability range assuming values that are not close to the lower and upper boundaries. This is meaningful because shows a particular feature of the anaerobic digestion already introduced in the chapter of the FOS/TAC. In fact, considering its optimization process, it is very difficult to obtain a specific parameter that is able to achieve its optimal value, such as a FOS/TAC equal to 0.5 or an OLR of 6 kgVS/m3/day. This is because the anerobic digestion is not optimized with respect to a single indicator, but

the optimization occurs through a complex concatenation of various parameters that exert, in a different way, their own influence on the process.

5.4.   Application of FBO2 to real case-studies

Once have been observed the nature of the results that the FBO2 is able to return as a function of a specific co-digestion, now it is important to have a validation of this model if applied to real case-studies. Therefore, it has been analyzed the co-digestions studied by some researchers. The aim is to verify if the FBO2 is able to identify a different composition of the same real biomasses allowing to the achievement of a higher BMP with respect to the one found experimentally. Three different co-digestions have been analyzed: the first one is characterized by a feeding composed of sewage waste and straw waste, the second one is a co-digestion of pig manure and potato waste, while the last one exploits fruit waste and pig manure. The results associated with the analyzed case-studies are shown below:

Figure 31: BMP vs w1 plot associated to the first case-study: co-digestion of sewage waste and straw waste. Paper of interest: [58].

Figure 32: BMP vs w1 plot associated to the second case-study: co-digestion of pig manure and potato waste. Paper of interest:[59].

Figure 33: BMP vs w1 plot associated to the third case-study: co-digestion of fruit waste and pig manure. Paper of interest: [60].

	Paper			FBO2		
	w1	w2	BMP [mL/gVS]	w1	w2	BMP [mL/gVS]
case1	0.5	0.5	187	0.84	0.16	260.6
case2	0.8	0.2	330	0.49	0.51	434.58
case3	0.5	0.5	224.8	0.48	0.52	422.8

Table 9: Values of the optimal compositions and their associated BMP for the three case-studies.

In the table 9, for each of the three case-studies, there are both the composition of the mixture defined by the paper as the best one and the composition identified by the FBO2 with their respective BMP values. It is possible to note that, for these case-studies, the model has found a different optimal composition that has been able to achieve a higher yield of biomethane compared to the one reported by the paper. Therefore, after these results, it is important to make the following considerations:

- This model is very useful because it allows to consider a high number of possible mixtures. Indeed, in the laboratory analysis, it is not convenient to compare the results obtained from a huge amount of compositions because it could be very expensive. For this reason, in the analysis performed in the papers, the more common mixtures are taken into consideration, such as: 0.50-0.50, 0.25-0.75, 0.75-0.25, 0.80-0.20 and so on. On the other hand, the FBO2 allows to consider also mixtures characterized by compositions having for the different w_i a second decimal digit that is different with respect to the usual 0 and 5. Therefore, this shows the two great potentialities associated to this model:

  1. It is able to identify a composition characterized by a BMP that is higher than the one of the paper;
  2. It considers mixtures that would never been analyzed in the laboratory analysis.

- It is worth to focus on the reliability of these results. Indeed, considering the third case-study, it is possible to note that despite the small difference between the two compositions, their associated BMP vary significantly. Therefore, results of this nature could not be so consistent. At the same time, it is important to point out that the results provided by the FBO2 depend on information

extracted from paper and analysis similar to those of the previous three case-studies. This means that it is correct to associate to these results a specific level of reliability.

At this point, after the previous results and considerations, it is possible to consider as finished the creation of the second tool needed for the construction of the final algorithm.

6 Conjugated substrate

6.1. Description of the procedure to find the conjugated substrate.

Once have been analyzed the features and the performances of the two previous tools, it is finally arrived the moment to point out how the neural network and the FBO2 can be used to identify the conjugated substrate. The significant aspect of this algorithm is that it contains all the topics described within this dissertation. Indeed, the choice of the second biomass is done through the respect of the stability conditions of the indicators analyzed until now. For this reason, the feeding composed by the predefined biomass and the conjugated substrate can be seen as the ideal combination to favor the formation of meaningful synergistic effects. These effects allow to maximize the biomethane yield thanks to the interactions between the macromolecules of the two substrates and the microorganisms inside the digester.

At this point, it is necessary to highlight the steps performed within this model to define the conjugated substrate. In fact, the starting point of this research is represented by the knowledge of the predefined biomass (subsequently named as S1) and of its features in terms of:

- C/N_1;
- BD_1;
- FOS/TAC_1;
- TS_1 e VS_1 in order to obtain the value of the organic loading rate;
- BMP_1.

Therefore, it is important to understand how, by the knowledge of the previous features, it is possible to find the best biomass to combine with the S1 to realize an efficient co-digestion. In the same way as done in the previous chapter, it can be useful to observe the figure 34 in which are schematized the fundamental steps performed by the model to obtain:

- The conjugated substrate;
- The mixture composition that is able to maximize the biomethane yield.

Figure 34: Simplified illustration of the structure of the model.

Such as the neural network and the FBO2, the creation of this model has been done by exploiting Python. From the figure reported above it is possible to note that, once have been defined the typology of the predefined biomass, there are two pathways necessary to reach the first results: one is related to the C/N of the second substrate, while the other regards its BD. Below it will be carefully analyzed the former, but the latter is affected by the same procedure.

Thanks to the conservation principle, also in this case it is possible to work considering only the composition of the first substrate. Therefore, once have been created the composition vector w_1 and expressed the C/N_{mix} as follows:

$$w_1 = [0, 0.01, 0.02, \ldots\ldots, 1] \tag{6.1}$$

$$C/N_{mix} = C/N_1 * w_1 + C/N_2 * (1 - w_1) \tag{6.2}$$

In such a way as to obtain the first information regarding the conjugated substrate:

- It has been assumed a wide range of possible C/N_2;
- For each value of w_1, it has been considered only those C/N_2 that can respect the stability conditions of C/N_{mix}. Observing this step under the point of view of Python implementation:

```
CN1  = S_pred['CN']
CN2 = np.arange(5, 60.5, 0.5)


CN2_new = []


for x_1 in x1:
    CNmix = x_1*(CN1-CN2)+CN2
    CN2_acc = CN2[(CNmix >= 10) & (CNmix <= 40)]
    CN2_new.append(CN2_acc)
```

Therefore, taking into account as range of possible C/N_2 values the one that goes from 5 to 60, as final result it is possible to obtain a matrix with dimensions $n \times m$, in which:

- n: Each row is related to a specific value of w_1;
- m: Each column is related to a C/N_2 that, as function of the composition of the first substrate, is able to satisfy the stability conditions of C/N_{mix}.

Once have been obtained this matrix, it is possible to use SNN.

It is important to understand how the SNN can be used by exploiting as input data the matrix mentioned previously. For this reason, it can be useful to recall the application of the neural network shown in the paragraph 4.3. In the chapter 4, it has been considered as range of C/N values the one that goes from 5 to 10 reaching their belonging probability to each of the four classes. Its similarities and differences with respect to the application of this chapter are as follows:

- For both it is possible to obtain the predictions on the basis of a single feature with respect to the 19 with which the SNN has been trained;
- In this case, considering a specific value of w_1, it will not be achieved the belonging probability of all the C/N_2 for each class but, for each class, it will be extracted only the C/N_2 with highest belonging probability.

The new C/N_2 matrix can be represented through the table 10:

w1	CN2 max.prob.CL1	CN2 max.prob.CL2	CN2 max.prob.CL3	CN2 max.prob.CL4
0	$CN2_{1,1}$	$CN2_{1,2}$	$CN2_{1,3}$	$CN2_{1,4}$
0.01	$CN2_{2,1}$	-	-	-
0.02	$CN2_{3,1}$	-	-	-
-	$CN2_{n-2,1}$	-	-	-
-	$CN2_{n-1,1}$	-	-	-
1	$CN2_{n,1}$	$CN2_{n,2}$	$CN2_{n,3}$	$CN2_{n,4}$

Table 10: values of the CN_2 with the highest belonging probability as a function of the considered class and w_1 composition.

Therefore, it is possible to note that, for each composition of the predefined biomass, the table shows four different C/N_2. These are the stability indicator values with the highest belonging probability to the four labels of the classification problem. The achievement of the table 10 represents the last step of the C/N_2 pathway.

This means that, through the BD_2 pathway that is characterized by the same steps described before, it is possible to obtain the following table:

w1	BD2 max.prob.CL1	BD2 max.prob.CL2	BD2 max.prob.CL3	BD2 max.prob.CL4
0	$BD2_{1,1}$	$BD2_{1,2}$	$BD2_{1,3}$	$BD2_{1,4}$
0.01	$BD2_{2,1}$	-	-	-
0.02	$BD2_{3,1}$	-	-	-
-	$BD2_{n-2,1}$	-	-	-
-	$BD2_{n-1,1}$	-	-	-
1	$BD2_{n,1}$	$BD2_{n,2}$	$BD2_{n,3}$	$BD2_{n,4}$

Table 11: values of the BD_2 with the highest belonging probability as a function of the considered class and w_1 composition.

Like the table 10, also in this case, for each composition of the first substrate there are four different BD_2. They represent the biodegradability values affected by the highest belonging probability depending on the considered class.

Once have been completed these two pathways, the next goal consists in the identification of the class that allows to maximize the biomethane yield. The procedure that permits the achievement of this objective is as follows:

- Starting from the first class and from the first composition of the vector w_1, it occurs the calculation of the BMP_{mix} by exploiting the correct C/N_2 and BD_2 through the usual two-parameters equation:

$$BMP_{mix} = w_1 BMP_1 + (1 - w_1)BMP_2 + w_1(1 - w_1)BMP_{syn} \tag{6.3}$$

In which:

$$BMP_{syn} = \beta_0 + \beta_1 \left(\frac{C}{N}\right)_{mix} + \beta_2 BD_{mix} + \beta_3 \left(\frac{C}{N}\right)_{mix}^2 + \beta_4 BD_{mix}^2 \tag{6.4}$$

- The same operation takes place for all the values of the vector w_1 considering the correct C/N_2 and BD_2 related to the specific composition of the first substrate;

- At this point, it is available a series of BMP_{mix}, each one associated to a given composition. Consequently, it is possible to derive the maximum value of biomethane yield with respect to the first class;
- The same procedure occurs also for the other three classes;
- In the end four different values of BMP_{mix} are obtained. They represent the highest biomethane yield associated to the four classes. As a result, through the greater of these values, it is possible to reach the first important result of this model, i.e. the class to which probably belongs the second substrate to combine with the predefined biomass in order to optimize the process performance.

Therefore, thanks to the steps described previously, it has been achieved the first information regarding the possible conjugated substrate. This is very meaningful considering the amount of starting data. Indeed, having available only the properties of the predefined biomass, it has been possible to estimate, with a specific level of reliability, the belonging class of the second substrate that is able to maximize the BMP of the mixture.

At this point, it is necessary to identify the conjugated substrate. In such a way as to find it, the following instructions have been implemented on Python:

```python
class_condition = classe_max


PAD_filtration = PAD[PAD['category'] == class_condition]


PAD_filtration['diff_CN'] = np.abs(PAD['CN']-CN2_max)
PAD_filtration['diff_BD'] = np.abs(PAD['BD']-BD2_max)


PAD_filtration['diff_tot']                                      =
PAD_filtration['diff_CN']+PAD_filtration['diff_BD']


index_best_row = PAD_filtration.loc[PAD_filtration['diff_tot'].idxmin()]
conjugated_substrate = index_best_row['substrate']
```

In the part of code reported above, it is possible to note the following aspects:

- The first step is represented by a filtration operation applied to the PAD. It is the PAD already described in previous chapter, so it contains the values that

averagely have the properties of the substrates. This operation permits to consider only the biomasses that belong to the class identified by the ANN avoiding to select a conjugated substrate associated to a wrong label;

- Now, among all the available biomasses, it is fundamental to choose the one affected by the more similar C/N_2 and BD_2 with respect to those of the identified class. Indeed, "CN2_max" and "BD2_max" are the values of the two indicators that have allowed to maximize the biomethane yield. This operation is done by calculating, for each substrate, the difference between the parameters of the class and the C/N and BD related to the specific biomass;

- After the computation of the previous operations, it is possible to identify the conjugated substrate that it is the one affected by the smallest difference. Therefore, it is the biomass with the more similar properties with respect to the C/N_2 and BD_2 predicted by the neural network.

Thereby, after the previous procedure, it has been possible to obtain the most important result associated to this algorithm, i.e. the estimation of the conjugated substrate by means of only the knowledge of the features of the predefined biomass.

At this point, once it has been discovered the ideal feeding to optimize the performance of the process, it is necessary to achieve the mixture composition that is able to maximize the biomethane yield. For this reason, it is used the second of the two tools created in the previous chapters, i.e. the FBO2. Its functioning is the same of the one analyzed in the chapter 5. Therefore, through the implementation of the adopted feeding:

- S1 = predefined biomass;
- S2 = conjugated substrate;

It is possible to obtain their ideal composition to maximize the BMP. Also in this case, there are the same filtration operations to overcame satisfying the stability conditions of the four parameters: C/N, BD, FOS/TAC and OLR.

Subsequently the achievement of the conjugated substrate and of the ideal composition of the mixture, another significative result that this model can return is the CHONS of the mixture. Indeed, in collaboration with the CNR (Consiglio Nazionale delle Ricerche) of Naples, it has been performed a canonical correlation analysis in order to predict the CHONS of the biomass [61]. Therefore, it is possible to obtain the percentage of carbon, hydrogen, oxygen, nitrogen, and sulfur of the mixture by providing to the model the following input:

- The total solids in the biomass (TS);
- The volatile solids of biomass (VS);
- Sugar concentration in biomass (SU);
- Protein concentration in biomass (PR);
- Lipids concentration in biomass (LP);
- Lignin concentration in biomass (LG);
- Hemicellulose concentration in biomass (HCE);
- Biodegradability index of the biomass (BD).

Thereby, thanks to the information listed above, the elemental analysis of the biomass is carried out. This can have several potentialities such as: the estimation of the mixture TMBP through the Buswell relation or the prediction of the C/N that could form during the co-digestion. Since in the case of interest there is a mixture of two substrates, the different input values are calculated by means of a linear combination of the information related to the single biomasses. Therefore, by taking into account the TS of the mixture, it is calculated as follows:

$$TS_{mix} = TS_1 * w_{1,opt} + TS_2 * w_{2,opt}$$
(6.5)

The same holds true for the other input data.

6.2. Analysis of the results provided by the model

After the CHONS analysis, the description about the functioning of the model can be considered as completed and so now, it is possible to analyze the results that the algorithm is able to provide. The first considered simulation is the one having as predefined biomass the sheep manure:

class	CN2	BD2	w1	w2	BMP [mL/gVS]	Conjugated substrate
organic waste	15.5	0.5	0.26	0.74	385.152	fruit waste

Table 12: results provided by the neural network having the sheep manure as predefined biomass.

w1	w2	BMP [mL/gVS]	C/N	BD	FOS/TAC	OLR [kgVS/(m3*day)]
0.34	0.66	370.26	14.98	0.67	0.392	3.596

Table 13: results provided by the FBO2 for the co-digestion of sheep manure and fruit waste.

C [%mol]	H [%mol]	O [%mol]	N [%mol]	S [%mol]
40.566	6.057	48.542	3.654	1.181

Table 14: elemental analysis of the mixture: sheep manure-fruit waste.

Figure 35: BMP vs w1 plot associated to the co-digestion of sheep manure and fruit waste.

Analyzing individually the tables reported above:

- Table 12: It is composed of the results provided by the neural network. In fact, it is possible to note the following aspects

  1. The class at which belongs the conjugated substrate is the organic waste;
  2. It has been chosen this class and not the other three because, in correspondence of a composition of the predefined biomass equal to 0.26 and of a C/N_2 and BD_2 equal to 15.5 and 0.5 respectively, it has been able to maximize the biomethane yield reaching a value of 385.152 [mL/gVS].

It is important to underline that, the obtained composition is not the optimal one to feed to the reactor; in fact, it has been derived from the predications of the ANN on the basis of the average properties that characterize the specific class. The ideal composition is extracted from the FBO2;

3. 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 13: It is composed of the results provided by the FBO2, i.e.:

1. Through the co-digestion of sheep manure and fruit waste, it has been identified as optimal mixture the one characterized by a composition of 0.34 for the first substrate and of 0.66 for the second one. It is possible to note that effectively the composition found by the FBO2 is different with respect to the one estimated by the neural network, even if this difference is not very big. In fact, also in this case, the mixture is affected by a prevalence of fruit waste. The biomethane yield resulting from this co-digestion is equal to 370.26 [mL/gVS];

2. Observing the different stability parameters associated with the optimal mixture, it is possible to note that all the four indicators are inside their respective stability range favoring so the formation of significant synergistic effects.

- Table 14: It is characterized by the elemental analysis of the mixture. Considering for example the percentage of carbon and of nitrogen, it is possible to notice that the associated C/N is similar to the one provided by the FBO2 thus finding a good reliability in the results of the model.

In the figure 35, as already described in chapter 5, it is represented the behavior of the biomethane yield as a function of the mixture composition. Therefore, 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 biomethane yield calculated with the

compositions that have failed the different filtration operations. The green dot denotes the maximum value of the BMP.

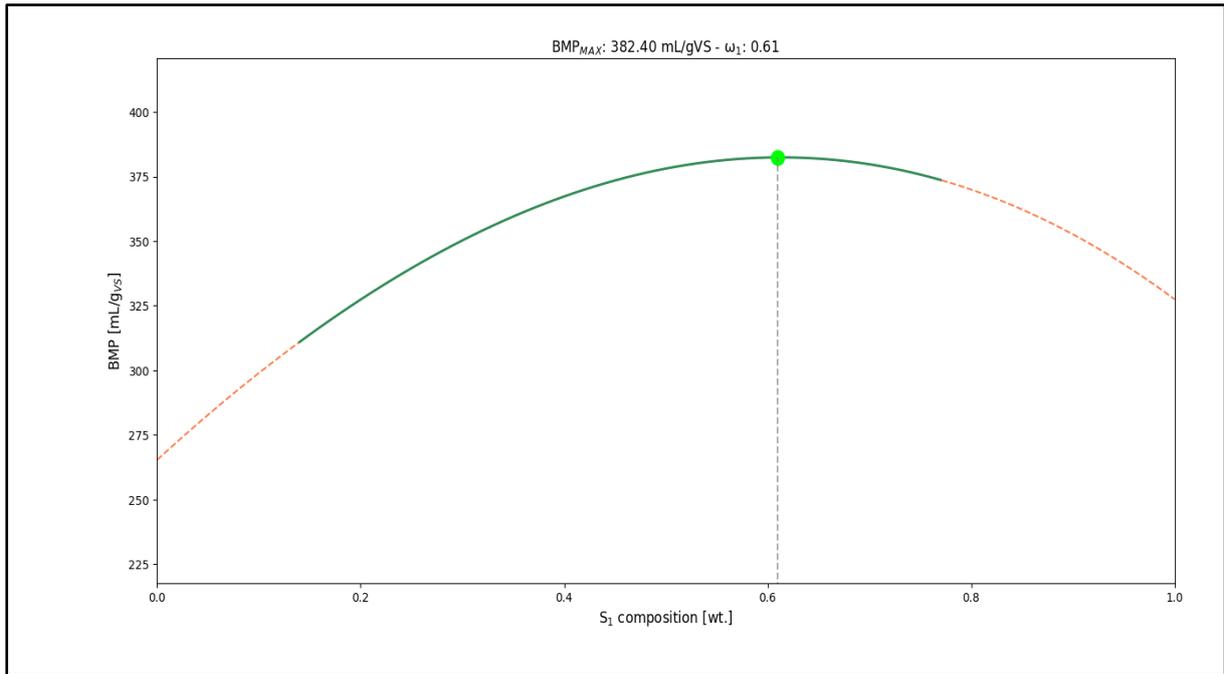Below are reported the plots and the table associated to other three different simulations:

Figure 36: BMP vs w1 plot associated to the co-digestion of dairy manure and chicken litter.

Figure 37: BMP vs w1 plot associated to the co-digestion of chicken manure and slaughter residue.

Figure 38: BMP vs w1 plot associated to the co-digestion of fish waste and fruit waste.

	case-study 1	case-study 2	case-study 3
Predefined biomass	dairy manure	chicken manure	fish waste
Conjugated substrate	chicken litter	slaughter residue	fruit waste
w1,NN	0.32	0.37	0.57
w2,NN	0.68	0.63	0.43
BMP,NN [mL/gVS]	393.479	393.202	455.891
w1,FBO2	0.34	0.39	0.58
w2,FBO2	0.66	0.61	0.42
BMP,FBO2 [mL/gVS]	400.48	380.94	447.62
C/N	16.139	12.45	17.161
BD	0.66	0.659	0.711
FOS/TAC	0.359	0.438	0.363
OLR [kgVS/(m3*day)]	3.441	5.93	5.691
C [%mol]	39.938	41.559	41.908
H [%mol]	5.971	6.134	6.258
O [%mol]	48.618	47.455	46.399
N [%mol]	4.204	3.699	4.136
S [%mol]	1.269	1.153	1.299

Table 15: results provided by the neural network and by the FBO2 for three different co-digestions: dairy manure-chicken litter, chicken manure-slaughter residue and fish waste-fruit waste.

Observing the results contained in the previous table, it is possible to conclude by saying that:

- For all the analyzed case studies, the ideal composition derived from the FBO2 is similar to the one predicted by the neural network. This consistency of results reflects a good level of reliability associated with the performances of the two tools;

- The stability conditions of the four indicators are always respected. This highlights the fact that, the research of the conjugated substrate is always performed not only with the aim of maximizing the biomethane yield, but also by taking into account the possible synergistic and antagonism effects formed within the anaerobic digester. The respect of the stability conditions favors the formation of the former;

- It is important to focus the attention on the choice of using as second biomass for the co-digestion the conjugated substrate identified by the model. Indeed, it represents the biomass that, with a high probability, permits to optimize the performance of the process. But, at the same time, it does not mean that it is the only that can be used or that the exploitation of a different substrate but with similar properties is a wrong choice. This is true for the following reasons:

  1. The result of this algorithm is obtained through predictions of a neural network based on information that are not characterized by regular behaviors. Therefore, it is possible that the identified conjugated substrate could be the correct one, but it is also probable that the one to be used is another. The significant aspects to extrapolate from this result are related to the features that the second biomass must have on average under the point of view of the macromolecules (i.e. proteins, sugars and lipids) and those associated with the possible mixture composition of the feeding;

  2. 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 FBO2 for each available substrate and, on the

basis of the considerations related to the results of the FBO2 and to the conjugated substrate provided by the algorithm, it is possible to identify the correct biomass for the co-digestion. An example of this application is shown in the following paragraph.

6.3. Application of the model to a real case-study

The example of interest is the one analyzed in paper [62]. The aim is to find the best substrate to combine with the fruit waste in a co-digestion process. The problem is that there are only two available wastes for the second biomass, i.e:

1. Dry grass;
2. Chicken manure.

By means of several analysis, it has been demonstrated that the best co-digestion was the one characterized by the feeding of fruit waste and chicken manure. At this point, it is very useful to observe the results provided by the two tools starting from the fruit waste as predefined biomass. The neural network returns the following results:

Conjugated substrate	w1,NN	w2,NN	BMP,NN [mL/gVS]
Sow manure	0.47	0.53	424.182

Table 16: results provided by the neural network having the fruit waste as predefined biomass.

Therefore, the conjugated substrate identified by the ANN is the sow manure. Observing the two available biomasses, it is possible to consider as the correct one the chicken manure because, coherently with what as been said before, it is the more similar to the sow manure; in fact, being two manures, both are affected by a high level of proteins. This could be a perfect choice because it is able to successfully counteract the significant amount of carbon coming from the sugars degradation of fruit waste. At this point, it is necessary to have also the FBO2 confirmation that the chicken manure is better than the dry grass. Thereby, the obtained results by means of the FBO2 for the two possible co-digestion with fruit waste are as follows:

Figure 39: BMP vs w1 plot associated to the co-digestion of fruit waste and chicken manure.

Figure 40: BMP vs w1 plot associated to the co-digestion of fruit waste and dry grass.

	w1,FBO2	w2,FBO2	BMP,FBO2 [mL/gVS]
Chicken manure	0.61	0.39	382.4
Dry grass	0.81	0.19	358.37

Table 17: results provided by the FBO2 for the co-digestions having as conjugated substrate chicken manure and dry grass.

From the table reported above, it is possible to note that also the FBO2 associates a higher biomethane yield to the co-digestion of fruit waste and chicken manure. Therefore, thanks to the paper taken into account, it has been demonstrated that the algorithm is able to provide reliable results also if applied to real case-studies. Indeed, by means of this model, it has been possible to identify as the most efficient co-digestion the same that the paper has found after a specific amount of analysis.

7 Conclusions

7.1. Summary of main results

This dissertation has been characterized by the following two goals:

1. Detailed analysis of two new stability parameters in order to evaluate their real influence on the performances of anaerobic digestion;
2. The creation of a machine learning to identify the specific conjugated substrate that, as a function of a predefined biomass, was able to maximize the biomethane yield.

Regarding the stability parameters, the starting point is represented by paper [18]. It has analyzed the dependency of the inhibition of the process with respect to the C/N ratio and to the biomass biodegradability BD. Therefore, in order to reach a higher knowledge about the functioning of the anaerobic digester, it has been performed the analysis of the new parameters, i.e. FOS/TAC and OLR. The first important aspect to point out is that they are two indicators influenced by different features of the biomass: in fact, the former is linked with its alkalinity level and with the volatile fatty acids formed after its degradation, while the second is connected to its total amount of volatile solids. The opportunity to monitor the process by considering different features of the substrate could be very useful; indeed, it allows to achieve a better understanding concerning the process behavior with respect to the several influence factors. For both parameters, it has been possible to obtain significant results, in fact:

- FOS/TAC: it has been demonstrated that, thanks to the use of specific yield coefficients, its value that allows to maximize the BMP is equal to 0.5, with respect to the 0.4 generally reported in the literature. Regarding the range of values to which is associated a stable functioning of the reactor, it has been considered the one that goes from 0.2 to 0.7. Indeed, for values higher than 0.7, the system alkalinity is not able to counteract the pH reduction caused by a critical accumulation of VFAs leading so to the inhibition of the process. Whereas, for values lower than 0.2, the digester can support a higher organic volumetric load.

Taking into account the database of interest and by representing the BMP behavior with respect to the FOS/TAC , it has been demonstrated the existence of iso-regions. Thanks to these regions, it is possible to predict the formation of synergistic and antagonism effects with respect to the substrates used to perform the co-digestion. Indeed, 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 inside a specific influence zone, i.e. for FOS/TAC lower than 1 and when the feeding is not 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_{mix} that is outside the acceptable range. Regarding the FOS/TAC_{mix}, it has been developed an accurate model in order to estimate the value of this mixture indicator as a function of the FOS/TAC of the single biomasses and of the possible synergistic effects formed by means of their interactions;

- OLR: 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 [kg$_{VS}$/m³/day] as that organic loading rate value beyond which anaerobic digestion becomes unstable. This is a reasonable assumption, even if there are also some cases in which the inhibition of the process occurs for values lower or higher than 6 [kg$_{VS}$/m³/day]. This is due to the complexity of factors that exert a specific influence on this indicator. It has been developed an accurate model that is able to predict the variation of the alkalinity level of the system as a function of the organic loading rate. Indeed, depending on the amount of volatile solids fed to the reactor, it is possible to modify the pH level of the anaerobic digester causing so the inhibition of the process or the achievement of ideal conditions to optimize the performance of the anaerobic digestion.

Regarding the second aim of this dissertation, i.e. the creation of a machine learning, two tools with specific functionalities have been developed. The first tool has been "a three layers neural network" applied to a multiclass classification problem characterized by the following labels: zootechnical waste, agricultural waste, organic waste and sludge waste. This ANN is able to identify the class to which belongs a substrate by means of a set of training data. This set contains a huge amount of substrate in which their properties are described through 19 features. Thanks to the values of this features, the neural network is able to develop some associations between the specific substrate and its respective class. The ANN can be used for

different applications, the one of interest consists in the capability to find the correct belonging class of a given biomass on the basis of a single feature.

The second tool is the so called FBO2; it represents an algorithm that, as a function of a specific feeding of two substrates, is able to define the ideal composition of the mixture that can maximize the biomethane yield. The achievement of this result occurs by the passing of four filtration operations. Each of them is associated to the compliance with the stability conditions imposed by the indicators: $C/N, BD, FOS/TAC$ and OLR. It has been demonstrated that, if compared to some real case studies analyzed in different papers, the FBO2 is able to identify a different composition of the same real biomasses guaranteeing the achievement of a higher BMP. The great advantage associated with this tool is represented by the huge amount of the considered mixtures with respect to the common ones tested in the laboratory analysis.

These two tools contribute to the creation the final machine learning. Indeed, as described previously, both are characterized by specific functionalities and peculiarities that allow to achieve this final goal. It has been shown that, starting from the knowledge of the properties of a predefined biomass, the machine learning is able to identify a reasonable conjugated substrate. In addition, it provides the ideal mixture composition to optimize the performance of the process by satisfying the different stability conditions. In the final part of chapter 6, it has been demonstrated the reliability of this machine learning if applied to a real case-study. This application consists in the running of the ANN and subsequently of the FBO2 for each available substrate of the plant; then, on the bases of the obtained results, it occurs the choice regarding the second biomass to use in the co-digestion. This choice depends on the properties that characterize the conjugated substrate predicted by the neural network, such as the macromolecules content.

In conclusion, this dissertation has allowed the achievement of a better understanding regarding the modalities with which the several factors influence the process. Particularly, thanks to this improvement, it has been possible to create a tool that was able to identify the best feeding features to optimize the performance of the process.

7.2. Limitations and recommendations for future researchers

Despite the significant results described above, it is also important to highlight the fact that this dissertation is affected by some limitations. They represent those obstacles that prevent to reach a complete knowledge about the real dependence of the anaerobic digestion on the synergism and antagonism effects formed within the reactor. First of all, one of the main weaknesses is directly attributable to the unreliability of the assumption of the OLR value that leads to the anaerobic digestion

inhibition, i.e. 6 [kg$_{VS}$/m^3/day]. It has been chosen this value because, consistently with that extrapolated from the literature, there are different papers that show an instable functioning of the process in correspondence of an organic loading rate equal to 6 [kg$_{VS}$/m^3/day]. Unfortunately, because of this parameter depends on several factors, it can happen that the inhibition of the process occurs for very different value of OLR, such as the industrial case-study analyzed in the chapter 3. This means that, the assumption related to the OLR can be useful in order to monitor the stability of the process through a different indicator but, at the same time, it is necessary to be aware of its limitations. Another problem found during this dissertation has been the inability to define a stability parameter associated with the bacterial community of the system. Indeed, despite the microbial diversity and acetoclastic-hydrogenotrophic ratio seemed to be good indicators for process stability, in reality they are not so reliable. The problems are connected to the complexity of the bacterial community but also to the low accuracy of the laboratory analysis. In spite of these difficulties, it seems that it is necessary to define a stability indicator associated with the microorganisms, because they play a fundamental role during the four steps of the anaerobic digestion. For this reason, it should be possible to define this new parameter, but through the analysis of bacterial community features that are different from the two considered in this dissertation. Other weaknesses could be the ones associated with the results provided by the neural network and the FBO2. Indeed, they depend on information that not necessarily follow regular behavior because they are organic substances, but, on the other hand, it is also important to point out that these data are extracted from papers that have analyzed specific co-digestions. For this reason, it is possible to associate to these results a good degree of reliability. A possible solution that can be performed to improve the accuracy of these predictions is represented by the increase of the information contained in the database. Therefore, through the analysis of further papers and the subsequent extrapolation of the biomasses properties that characterize the co-digestion of interest, it is possible to reach a better global overview of the connections between the different features of the substrates. Regarding future research, 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.

Another important point of reflection is the one related to the model to use to improve the reliability of the calculated BMP_{mix}. Indeed, particularly in the simulations performed in the chapter 5, it has emerged that, by means of the model used in this dissertation, it has been obtained a not very consistent biomethane potential with respect to the one extrapolated from the literature. For this reason, it is necessary to identify a new model that must be a trade-off between the complexity of the ADM1 and the simplicity of the method used in this dissertation.

Bibliography

[1] Cecchi F., Battistoni P., Pavan P., Bolzonella D., e Innocenti L., «Digestione anaerobica della frazione organica dei rifiuti soldi. Aspetti fondamentali, progettuali, gestionali, di impatto ambientale ed integrazione con la depurazione delle acque reflue», Università Ca' Foscari Venezia, 2005.

[2] Angelidaki I. *et al.*, «An automatic system for simultaneous monitoring of gas evolution in multiple closed vessels», *Journal of Microbilogical Methods*, vol. 33, pp. 93–100, 1998.

[3] Pavan Paolo, Busato Juliana, e Biasiolo Marco, «Digestione anaerobica a singola fase della frazione liquida ricavata da FORSU pressata, in mesofilia e in termofilia», Università Ca' Foscari Venezia, 2019.

[4] Mata-Alvarez J., Mace S., e Llabres P., «Anaerobic digestion of organic solid wastes. An overview of research achievements and perspectives», *Bioresource Technology*, vol. 74, pp. 3–16, 2000.

[5] Astals S., Ariso M., Galì A., e Mata-Alvarez J., «Co-digestion of pig manure and glycerine: experimental and modelling study», *Journal of Environmental Sciences*, vol. 92, pp. 1091–1096, 2011.

[6] Mata-Alvarez J., Dosta J., Mace S., e Astals S., «Codigestion of solid wastes: a review of its uses and perspectives including modelling», *Critical Reviews in Biotechnology*, vol. 31, 2011.

[7] Mata-Alvarez J., Dosta J., Romero-Guieza MS, Fonoll X., Peces M., e Astals S., «A critical review on anaerobic co-digestion achievements between 2010 and 2013.», *Renewable and Sustainable Energy Reviews*, vol. 36, pp. 412–427, 2014.

[8] Astals S., Batstone DJ., Mata-Alvarez J., e Jensen PD., «Identification of synergistic impacts during anaerobic co-digestion of organic wastes.», *Bioresource Technology*, vol. 169, pp. 421–427, 2014.

[9] Shah FA., Mahmood Q., Rashid N., Pervez A., Raja IA., e Shah MM., «Co-digestion, pretreatment and digester design for enhanced methanogenesis», *Renewable and Sustainable Energy Reviews*, vol. 42, pp. 627–642, 2015.

[10] Holm-Nielsen JB., Al Seadi T., e Oleskowicz-Popiel P., «The future of anaerobic digestion and biogas utilization.», *Bioresource Technology*, vol. 100, pp. 5478–5484, 2009.

[11] Jagadabhi PS., Lehtomaki A., e Rintala J., «Co-digestion of grass silage and cow manure in a cstr by re-circulation of alkali tretaed solids of the digestate.», *Environ Technol.*, vol. 29, pp. 1085–1093, 2008.

[12] Cavinato C., Fatone F., Bolzonella D., e Pavan P., «Thermophillic anaerobic co-digestion of cattle manure with agro-wastes and energy crops: comparison of pilot and full scale experiences.», *Bioresource Technology*, vol. 101, pp. 545–550, 2010.

[13] Kiros Hagos, Jianpeng Zong, Dongxue Li, Chang Liu, e Xiaohua Lu, «Anaerobic co-digestion process for biogas production: Progress, challenges and perspectives.», *Renewable and Sustainable Energy Reviews*, vol. 76, pp. 1485–1496, 2017.

[14] Wagner AO., Lins P., Malin C., Reitschuler C., e Illmer P., «Impact of protein-, lipid- and cellulose-containing complex substrates on biogas production and microbial communities in batch experiments.», *Sci Total Environ.*, pp. 458-460:256–266, 2013.

[15] Kallistova AY., Goel G., e Nozhevnikova AN., «Microbial diversity of methanogenic communities in the systems for anaerobic treatment of organic waste.», *Microbiology*, vol. 83, pp. 462–483, 2014.

[16] Ebner J.H., Labatut R.A., Lodge J.S., Williamson A.A., e Trabold T.A., «Anaerobic co-digestion of commercial food waste and dairy manure: characterizing biochemical parameters and synergistic effects.», *Waste Management*, vol. 52, pp. 286–294, 2016.

[17] Chow W.L. *et al.*, «Anaerobic co-digestion of wastewater sludge: a review of potential co-substrates and operating factors for improved methane yield .», *Processes*, vol. 8, p. 39, 2020.

[18] F. Moretta, A. Goracci, F. Manenti, e G. Bozzano, «Data-driven model for feedstock blending optimization of anaerobic co-digeston by BMP maximization.», *Journal of Cleaner Production*, vol. 375, 2022.

[19] David J. Hills, «Effects of carbon:nitrogen ratio on anaerobic digestion of dairy manure.», *Agricultural Wastes*, vol. 1, pp. 267–278, 1979.

[20] Romano R.T. e Zhang R.H., «Co-digestion of onion juice and wastewater sludge using an anaerobic mixed biofilm reactor.», *Bioresour Technol.*, vol. 99, pp. 631–637, 2008.

[21] Kumar M., Ou Y.L., e Lin J.G., «Co-composting of green waste and food waste at low C/N ratio.», *Waste Management*, vol. 30, pp. 602–609, 2010.

[22] B. Ruffino, M. Zanetti, G. Campo, e G. Gentilucci, «Digestione anaerobica del fango di supero in regime termofilo», Politecnico di Torino, 2020.

[23] G. Bucchi, S. M. Spampinato, D. Pinelli, D. Frascari, e L. Talia, «Studio di modelli matematici per la verifica e la progettazione di impianti di digestione anaerobica su scala di laboratorio e industriale.», Alma Mater Studiorum Università di Bologna, 2015.

[24] Rosa Cecilia Ruìz-Bastidas e Luz Stella Cadavid-Rodrìguez, «Effect of nutrients, inoculum and co-substrates on methane potential of cattle manure.», *Revista Facultad De Ingenieria Universidad De Antioquia*, vol. 108, pp. 41–53, 2022.

[25] Ulrich Lossie e Petra Putz, «Gestione ottimale degli impianti di biogas mediante FOS/TAC.», FOS/TAC.

[26] F. Cecchi, P. Battistoni, P. Pavan, D. Bolzonella, e L. Innocenti, *Digestione anaerobica della frazione organica dei rifiuti solidi*. 2005.

[27] Shiyou Pan, Hossain M. Zabed, Zhenchong Li, Q. Xianghui, e Yutuo Wei, «Enrichment and balancing of nutrients for improved methane production using three compositionally different agro-livestock wastes: Process performance and microbial community analysis», *Bioresource Technology*, vol. 357, 2022.

[28] Kun Li, Ronghou Liu, Shaofeng Cui, Qiong Yu, e Ruijie Ma, «Anaerobic co-digestion of animal manures with corn stover or apple pulp for enhanced biogas production», *Renewable Energy*, vol. 118, pp. 335–342, 2018.

[29] Xiaohu Dai, Nina Duan, Bin Dong, e Lingling Dai, «High-solids anaerobic co-digestion of sewage sludge and food waste in comparison with mono digestions: Stabilty and performance», *Waste Management*, vol. 33, pp. 308–316, 2013.

[30] Yangyang Li *et al.*, «Reactor performance and energy analysis of solid state anaerobic co-digestion of dairy manure with corn stover and tomato residues», *Waste Management*, vol. 73, pp. 130–139, 2018.

[31] M.J Kim, S.H Kim, G.W Kim, e I.H Bae, «Biogas production of ASBR (Anaerobic Sequencing Batch Reactor) for cheese whey», *ASABE-CSBE/SCGAB Annual International Meeting Paper*, 2014.

[32] Rongping Li, Shulin Chen, e Xiujiu Li, «Biogas production from Anaerobic Co-digestion of Food Waste with Dairy Manure in a Two-Phase Digestion System», *Appl Biochem Biotechnol*, vol. 160, pp. 643–654, 2010.

[33] Jia Lin *et al.*, «Effects of mixture ratio on anaerobic co-digestion with fruit and vegetable waste and food waste of China», *Journal of Environmental Sciences*, vol. 23, pp. 1403–1408, 2011.

[34] Furqan Muhayodin, Albrecht Fritze, Oliver Christopher Larsen, Marcel Spahr, e Vera Susanne Rotter, «Co-digestion of Rice Straw with Cow Manure in an Innovative Temperature Phased Anaerobic diestion technology: Performance Evaluation and Trace Elements», *Energies*, vol. 14, 2021.

[35] 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 phytotoxiciy», *Waste Management*, vol. 34, pp. 1603–1608, 2014.

[36] Chiara Tacconi, Mirko Cucina, Daniela Pezzolla, Claudia Zadra, e Giovanni Gigliotti, «Effect of the mycotoxin aflatoxin B1 on a semi-continuous anaerobic digestion process», *Waste Management*, vol. 78, pp. 467–473, 2018.

[37] Herald Wilson Ambrose, Ligy Philip, G.K. Suraishkumar, Anbarasu Karthikaichamy, e Tushar Kanti Sen, «Anaerobic co-digestion of activated sludge and fruit and vegetable waste: Evaluation of mixing ratio and impact of hybrid (microwave and hydrogen peroxide) sludge pre-treatment on two-stage digester stability and biogas yield», *Journal of Water Process Engineering*, vol. 37, 2020.

[38] Mirko Cucina, Daniela Pezzolla, Chiara Tacconi, e Giovanni Gigliotti, «Anaerobic co-digestion of a lignocellulosic residue with different organic wastes: Relationship between biomethane yield, soluble organic matter and process stability», *Biomass and Bioenergy*, vol. 153, 2021.

[39] Ahmad Reza Salehiyoun, Francesco Di maria, Mohammad Sharifi, Omid Norouzi, Hamid Zilouei, e Mortaza Aghbashlo, «Anaerobic co-digestion of sewage sludge and slaughterhouse waste existing wastewater digesters», *Renewable Energy*, vol. 145, pp. 2503–2509, 2020.

[40] C. Veluchamy, Brandon H. Gilroyed, e Ajay S. Kalamdhad, «Process performance and biogas production optimizing of mesophilic plug flow anaerobic digestion of corn silage», *Fuel*, vol. 253, pp. 1097–1103, 2019.

[41] Lin Chen *et al.*, «Effects of organic loading rates on the anaerobic co-digestion of fresh vinegar residue and pig manure: Focus on the performance and microbial communities», *Biochemical Engineering Journal*, vol. 183, 2022.

[42] Tianjie Ao · Zhijie Xie· Pan Zhou, · Xiaofeng Liu, · Liping Wan, e · Dong Li, «Comparison of microbial community structures between mesophilic and thermophilic anaerobic digestion of vegetable waste», *Bioprocess and Biosystems Engineering*, vol. 44, pp. 1201–1214, 2021.

[43] Shiyou Pa, Hossain M. Zabed, Zhenchong Li, Q. Xianghui, e Yutuo Wei, «Enrichment and balancing of nutrients for improved methane production using three compositionally different agro-livestock wastes: Process performance and microbial community analysis», *Bioresource Technology*, vol. 357, 2022.

[44] Yaya Wanga,b *et al.*, «Effects of co-digestion of cucumber residues to corn stover and pig manure ratio on methane production in solid state anaerobic digestion», *Bioresource Technology*, vol. 250, pp. 328–336, 2018.

[45] Qiang Lin, Guihua He, , Junpeng Rui, , Xiaoyu Fang, Yong Tao, e , Jiabao Li and Xiangzhen Li, «Microorganism-regulated mechanisms of temperature effects on the performance of anaerobic digestion», *Microb Cell Fact*, vol. 3, pp. 15–96, 2016.

[46] M. C. Gagliano & C. M. Braguglia & A. Gallipoli & e A. Gianico & S. Rossetti, «Microbial diversity in innovative mesophilic/thermophilic temperature-phased anaerobic digestion of sludge», *Envi ron Sci Pollut Res*, vol. 22, pp. 7339–7448, 2015.

[47] Dara S. M. Ghasimi1*†, Yu Tao1,2†, Merle de Kreuk1, , Marcel H. Zandvoort3, e and Jules B. van Lier, «Microbial population dynamics during long-term sludge adaptation of thermophilic and mesophilic sequencing batch digesters treating sewage fine sieved fraction at varying organic loading rates», *Biotechnol Biofuels*, pp. 8–171, 2015.

[48] Lei Li, Qin He, Yao Ma, Xiaoming Wang, e Xuya Peng, «Dynamics of microbial community in a mesophilic anaerobic digester treating food waste: Relationship between community structure and process stability», *Bioresource Technology*, vol. 189, pp. 113–120, 2015.

[49] S. Lerm a *et al.*, «Archaeal community composition affects the function of anaerobic co-digesters in response to organic overload», *Waste Management*, vol. 32, pp. 389–399, 2012.

[50] Azadeh Babaee e Jalal Shayegan, «Effect of organic loading rates (OLR) on production of methane from anaerobic digestion of vegetables waste», *Bioenergy Technology*, vol. 57:55, pp. 3428–3435, 2011.

[51] Rodrigo A. Labatut e Jennifer L. Pronto, «Sustainable Waste-to-Energy Technologies: Anaerobic Digestion», *Sustainable Food Waste-To-energy Systems*, pp. 47–67, 2018.

[52] Chengliu Gou, Zhaohui Yang, Jing Huang, Huiling Wang, Haiyin Xu, e Like Wang, «Effects of temperature and organic loading rate on the performance and microbial community of anaerobic co-digestion of waste activated sludge and food waste», *Chemosphere*, vol. 105, pp. 146–151, 2014.

[53] 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.

[54] Burak Demirel e Orhan Yenigun, «Anaerobic acidogenesis of dairy wastewater: the effects of variations in hydraulic retention time with no pH control», *Chem Technol Biotechnol*, vol. 79, pp. 755–760, 2004.

[55] M.A. Khan *et al.*, «Optimization of process parameters for production of volatile fatty acid, biohydrogen and methane from anaerobic digestion», *Bioresource Technology*, vol. 219, pp. 738–748, 2016.

[56] Ayhan Varol e Aysenur Ugurlu, «Comparative evaluation of biogas production from dairy manure and co-digestion with maize silage by CSTR and new anaerobic hybrid reactor», *Eng. Life Sci.*, vol. 17, pp. 402–412, 2017.

[57] Norio Nagao *et al.*, «Maximum organic loading rate for the single-stage wet anaerobic digestion of food waste», *Bioresource Technology*, vol. 118, pp. 210–218, 2012.

[58] Pengfei Li, Chao He, Ran Yu, Dekui Shen, e Youzhou Jiao, «Anaerobic co-digestion of urban sewage sludge with agricultural biomass», *Waste and Biomass Valorization*, vol. 11, pp. 6199–6209, 2020.

[59] Prasad Kaparaju e Jukka Rintala, «Anaerobic co-digestion of potato tuber and its industrial by-products with pig manure», *Resources, Conservation and Recycling*, vol. 43, pp. 175–188, 2005.

[60] Jian Shen, Chen Zhao, Yan Liu, Ruihong Zhang, Guangqing Liu, e Chang Chen, «Biogas production from anaerobic co-digestion of durian shell with chicken, dairy, and pig manures», *Energy Conversion and Management*, vol. 198, 2019.

[61] 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.

[62] Atul Navnath Muske e P. Venkateswara Rao, «Evaluation of Biogas Production Potential by Anaerobic Co-digestion with Substrate Mixture of Fruit Waste, Lawn Grass, and Manures», *Water Resources and Environmental Engineering II*, pp. 91–99, 2018.

[63] Neural Network and Deep Lerning - DeepLearning.AI

[64] Tony Chan, << Machine Learning con Python & Scikit-Learn TensorFlow2 Keras Pytorch e Fastai>>

[65] Terna Driving Energy - Terna spa

List of Figures

List of Tables