Deep deterministic policy gradient: applications in process control and integrated process design and control

by

Tannia Argelia Mendiola Rodriguez

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Chemical Engineering

Waterloo, Ontario, Canada, 2022

©Tannia Argelia Mendiola Rodriguez 2022

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

In recent years, the urgent need to develop sustainable processes to fight the negative effects of climate change has gained global attention and has led to the transition into renewable energies. As renewable sources present a complex dynamic behavior, this has motivated a search of new ways to simulate and optimize processes more efficiently. One emerging area that has recently been explored is Reinforcement learning (RL), which has shown promising results for different chemical engineering applications. Although recent studies on RL applied to chemical engineering applications have been performed in different areas such as process design, scheduling, and dynamic optimization, there is a need to explore further these applications to determine their technical feasibility and potential implementation in the chemical and manufacturing sectors. An emerging area of opportunity to consider is biological systems, such as Anaerobic Digestion Systems (AD). These systems are not only able to reduce waste from wastewater, but they can also produce biogas, which is an attractive source of renewable energy.

The aim of this work is to test the feasibility of a RL algorithm referred to as Deep Deterministic Policy Gradient (DDPG) to two typical areas of process operations in chemical engineering, i.e., process control and process design and control. Parametric uncertainty and disturbances are considered in both approaches (i.e., process control and integration of process and control design). The motivation in using this algorithm is due to its ability to consider stochastic features, which can be interpreted as plant-model mismatch, which is needed to represent realistic operations of processes.

In the first part of this work, the DDPG algorithm is used to seek for open-loop control actions that optimize an AD system treating Tequila vinasses under the effects of parametric uncertainty and

disturbances. To provide a further insight, two different AD configurations (i.e., a single-stage and a two-stage system) are considered and compared under different scenarios. The results showed that the proposed methodology was able to learn an optimal policy, i.e., the control actions to minimize the organic content of Tequila in the effluents while producing biogas. However, further improvements are necessary to implement this DDPG-based methodology for online large-scale applications, e.g., reduce the computational costs.

The second part of this study focuses on the development of a methodology to address the integration of process design and control for AD systems. The objective is to optimize an economic function with the aim of finding an optimal design while taking into account the controllability of the process. Some key aspects of this methodology are the consideration of stochastic disturbances and the ability to combine time-dependent and time-independent actions in the DDPG. The same two different reactor configurations considered in the optimal control study were explored and compared in this approach. To account for constraints, a penalty function was considered in the formulation of the economic function. The results showed that there are different advantages and limitations for each AD system. The two-stage system required a larger investment in capital costs in exchange of higher amounts of biogas being produced from this design. On the other hand, the single-stage AD system required less investment in capital costs in exchange of producing less biogas and therefore lower profits than the two-stage system. Overall, the DDPG was able to learn new control paths and optimal designs simultaneously thus making it an attractive method to address the integrated design and control of chemical systems subject to stochastic disturbances and parametric uncertainty.

ACKNOWLEDGEMENTS

Firstly, I would like to thank Prof. Luis A. Ricardez-Sandoval for giving me the opportunity to pursue my MASc degree and for being an attentive supervisor, who reviewed and advised my work and performance, guiding me throughout the realization of my master's degree.

I would like to thank my examining committee Prof. Qinqin Zhu and Prof. Eric Croiset. For taking the time to review my thesis and providing valuable feedback. I want to also thank Dr. Samuel Zapien, Dr. Carla Robledo and Eng. Jose Luis Vargas for their constant support to obtain the CONACyT scholarship.

I'm especially grateful to my queen, grandma Vita, who everyday answers my phone calls and keeps in touch. To my Mexican family, my supportive boyfriend, my Brazilian family and my friends, for their unconditional love and for supporting me and encouraging me to follow my goals, although that meant being far from them. Also, I want to thank Lima, for waiting for me until the last minute, you will always be in my heart.

I want to thank the financial assistance provided by the Consejo Nacional de Ciencia y Tecnologia (CONACyT) with the scholarship CONACYT-Regional Noreste 2020; MITACS fellowship (http://www.mitacs.ca/); and the Natural Sciences and Engineering Research Council of Canada (NSERC).

And the most important of all, thanks to GOD, for always showing me that His plans are better than mine.

What's the purpose of science if there is no humanity?

Olim olim deus accelere hoc saeculum splendidum accelere fiat venire olim

DEDICATION

Dedicated to God, my family, my angel Lima, and *meu céu*.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

RL	Reinforcement learning
AD	Anaerobic digestion
EMPC	Economic Model Predictive Controller
PPO	Proximal Policy Optimization
TRPO	Trust Region Policy Optimization
TD3	Twin-Delayed DDPG
PSO	Particle Swarm Optimization
MPC	Model predictive controller
DNN	Deep neural network
SAC	Soft actor-critical
A3C	Asynchronous Advantage Actor-Critic
NLP	Nonlinear programming
MINLP	Mixed integer nonlinear programming
PDF	Probability density function
PAROC	PARametric Optimization and Control
MIDO	Mixed-integer dynamic optimization
CSTR	Continuous stirred-tank reactor
PSE	Power series expansion
NMPC	Nonlinear Model Predictive Control
KKT	Karush–Kuhn–Tucker
NN	Neural network
MADRL	Multi-agent deep reinforcement learning
DOP	Dynamic optimization problem
SARSA	State-action-reward-state-action
MDP	Markov Decission Process
MSBE	Mean-squared Bellman error
TD	Temporal Difference
CRT	Consejo Regulador del Tequila

VFA	Volatile fatty acids
COD	Chemical oxygen demand
IPOPT	Interior Point OPTimizer
OU	Ornstein-Uhlenbeck noise

NOMENCLATURE

(Order of appearance)

w_i	Weights of neural network
b_i	Bias applied to ith neuron
$\boldsymbol{\theta}_j$	Set of time-independent realizations for the uncertain parameters
M	Number of uncertainty realizations
$\widehat{\boldsymbol{x}}_{t,j}$	Differential states for each uncertain realization *j* and every time step *t*,
$\boldsymbol{u}_t$	Control profile vector
u^l	Lower bound of control vector
u^h	Upper bound of control vector
$\widehat{\boldsymbol{y}}_{t,j}$	Controlled variables for each realization
$\widehat{\boldsymbol{Y}}_{t,j}$	algebraic variable vector
$\boldsymbol{x}_t$	State vector
f	Set of nonlinear differential equations representing the system dynamics
h	Set of algebraic equations
g	Set of inequality constraints
ω_j	Weights assigned for each uncertainty realization *j*
E	Environment of MDP
S	State of MDP
A	Action space of MDP
$\mathcal{P}(s_{t+1}\|s_t, a_t)$	Transition function of MDP
$r(s_t, a_t)$	Reward function of MDP
s_t	States from environment
a_t	Action
s_{t+1}	Next state

γ	Discount factor
μ	Policy of reinforcement learning
R_t	Expected reward
ϕ^Q	Weight parameters of critic network
ϕ^μ	Weight parameters of actor network
$\phi^{Q'}$	Weight parameters of critic target network
$\phi^{\mu'}$	Weight parameters of actor target network
Q^μ	Q-value function
L	Loss function of critic network
y_t	Target value
Q'	Output of target critic network
μ'	Output of target actor network
P	Maximum number of tuples in buffer
T	Minibatch size
$\mathcal{N}_t$	added before the action is returned to the environment
D	Dilution rate of single-stage AD system.
z_1	Concentration of acidogenic biomass (g/L)
z_2	Concentration of methanogenic biomass (mmol /L)
z_3	Substrate concentrations in terms of COD (g COD/L)
z_4	Substrate concentrations in terms of VFA (mmol VFA/L)
α	Biomass fraction that is suspended in liquid phase
μ_{1max}	Monod kinetics in acidogenic step $(1/d)$. Maximum growth rate of acidogenic bacteria
k_{s1}	Monod kinetics (g COD/L). Half velocity constant

μ_{2max}	Parameters of the Haldane kinetics involved in the methanogenic reaction$(1/d)$ Maximum growth rate of methanogenic bacteria
k_{s2}	Parameters of the Haldane kinetics involved in the methanogenic reaction
k_{I2}	Parameters of the Haldane kinetics involved in the methanogenic reaction
γ_1	yield coefficient of single-stage
$S_{1,in}$	Inlet concentration of COD (g COD/L) that enters into the reactor (single-stage)
$S_{2,in}$	Inlet concentration of VFA (mmol VFA/L) that enters into the reactor (single-stage)
x_1	Scaled acidogenic concentration of the biomass in the acidogenic reactor (g COD/L)
x_2	Substrate concentration of COD in acidogenic reactor (g COD/L)
x_3	Substrate concentration of VFA in acidogenic reactor (mmol VFA/L)
x_4	Scaled acidogenic concentration of the biomass in the methanogenic reactor (g COD/L)
x_5	Methanogenic concentration of the biomass in the methanogenic reactor (mmol VFA/L)
x_6	Substrate concentration of COD in methanogenic reactor (g COD/L)
x_7	Substrate concentration of VFA in methanogenic reactor (mmol VFA/L)
V_1	Acidogenic reactor capacity (L)
V_2	Methanogenic reactor capacity (L)
k_1	Yield coefficient acidogenic reactor

k_4	Yield coefficient acidogenic reactor
k_3	Yield coefficient methanogenic reactor
k_5	Yield coefficient methanogenic reactor
$S_{11,in}$	Inlet concentrations of COD (g COD/L) that enter to the acidogenic reactor
$S_{21,in}$	Inlet concentrations of VFA (mmol VFA/L) that enter to the acidogenic reactor
μ_{11max}	Parameters of the Monod kinetics (1/d)
k_{s11}	Parameters of the Monod kinetics (g COD/L)
μ_{12max}	Parameters of the Monod kinetics $(1/\mathrm{d})$
k_{s12}	Parameters of the Monod kinetics (g COD/L)
μ_{22max}	Parameters of the Haldane kinetics. Maximum growth of methanogenic bacteria (1/d)
k_{s22}	Parameters of the Haldane kinetics (mmol VFA/L)
k_{I2}	Parameters of the Haldane kinetics $(mmolVFA/L)^{1/2}$
S_{11,in_G}	Random inlet concentrations of tequila vinasses entering the system (g COD/L)
S_{21,in_G}	Random inlet concentrations of tequila vinasses entering the system (mmol VFA/L)
$\varepsilon_{t,1}$	Random Gaussian noises with respect to the nominal inlet concentrations (g COD/L)
$\varepsilon_{t,1}$	Random Gaussian noises with respect to the nominal inlet concentrations (mmol VFA/L)
OF	Economic function
$\boldsymbol{x}(t)$	System's states
$\dot{\boldsymbol{x}}(t)$	Derivatives of states
$\boldsymbol{u}$	Control profile vector
$\boldsymbol{u}^l$	Lower bound of control vector

$\boldsymbol{u}^h$	Upper bound of control vector
$\boldsymbol{\zeta}$	Vector of uncertain realizations
$\boldsymbol{des}$	Design variables
$\boldsymbol{d}$	Vector of disturbances
$\boldsymbol{\vartheta}$	Noise added to seasonal changing-disturbances
$\boldsymbol{\pi}$	Seasonal changes with respect to the nominal values in the disturbances
TS	final number of timesteps in each episode
χ	Inputs of the neuron
a_u	Activation function of the manipulated variables
a_{des}	Activation function of the design variables
Q_1	Volumetric flow of the acidogenic reactor (L/day)
Q_2	Volumetric flow of the methanogenic reactor (L/day)
β	Ratio between methanogenic and acidogenic reactor
CC	Capital costs
PP	Production profit costs
EC	Energy consumption costs
VC	Variability costs
C_{BMR1}	Bare-module cost for the single-stage system
C_{BMR21}	Bare-module cost for the two-stage system (acidogenic reactor)
C_{BMR22}	Bare-module cost for the two-stage system (methanogenic reactor)
$spCOD$	COD set-point (g COD/L)
pc	Penalty cost that accounts for set-point tracking errors

y_{CH_4}	Yield coefficient of methane
y_{mb}	Biomass factor
q	Selling price for methane
p	Tequila vinasses treatment price

CHAPTER 1
Introduction

Nowadays, the pollution levels combined with current global market demands have led industries to implement and improve sustainable processes to fight climate change. Therefore, renewable sources have been positioned as a key player in the transition into more sustainable processes. One of the clean technologies that has been extensively studied is anaerobic digestion (AD) systems, which are one of the leading biochemical conversion technologies commonly applied for organic waste valorization(Sikarwar et al., 2021). This process can simultaneously reduce organic matter content and generate biogas (a gas mixture mainly composed of methane), thus making it quite attractive for large-scale applications. To operate an AD process efficiently, it is necessary to tackle the problem of plant-model mismatch, which is common in AD systems since this process is subject to external disturbances and model parameter uncertainty. Plant-model mismatch may lead to inaccurate estimations of operating conditions and dynamically infeasible designs if the models assume perfect knowledge of key process parameters and inputs, i.e., it ignores the effects of external perturbations or parametric uncertainty. In AD systems, plant-model mismatch is often observed as disturbances in the substrate load whereas uncertainty in the kinetic parameters are key factors affecting the performance of these systems. Typically, disturbances are assumed to follow a deterministic profile; however, the disturbance profiles in a real setting are often subject to random realizations thus making these external variables stochastic in nature. One potential technique to consider these stochastic features is Reinforcement Learning, which has been recently gaining interest for different chemical engineering applications.

e.g., planning and scheduling, (Arai et al., 2000), real-time optimization (Powell et al., 2020), batch bioprocesses optimization (Petsagkourakis et al., 2020b), and dynamic optimization (MacHalek et

al., 2020).

RL is an iterative learning process where an agent decides a sequence of actions in order to maximize a reward. The purpose of RL is to learn policies, i.e., actions that lead optimal solutions. Recently, RL has presented several improvements, such as AlphaGo, a computer program that defeated a world champion player of Go, a complex Chinese game (Silver et al., 2016). Since then, multiple algorithms have applied the basis of AlphaGo to create new and efficient RL algorithms, such as PPO (Schulman et al., 2017) and TRPO (Schulman et al., 2015) and TD3((Dankwa & Zheng, 2019). Among the RL algorithms, the present study focuses on one method referred to as Deep Deterministic Policy Gradient (DDPG). This algorithm can describe high dimensional systems, it has an inherent stochastic nature, and can solve continuous action control problems (Yoo, Kim, et al., 2021). To the author's knowledge, the application of DDPG for chemical systems is very limited. Hence, there is an incentive to explore the application of this RL algorithm to systems that exhibit different complex dynamic behavior, such as the AD process considered in this study. As AD systems are prone to suffer from model uncertainty due to their kinetic parameters and external disturbances in the loading of the process, these stochastic features make RL a potential technique to deal with the controllability of AD systems. The two most common reactor configurations for AD systems are the single-stage system, where there is just one digester for the process, and the two-stage system, where two different reactors are used to separate the microorganisms to enhance their growth. Various studies have indicated that two-stage AD systems promote a better environment for microbial growth and thus, more biogas can be produced from those systems. However, other studies have indicated that the performance of an AD system is mostly influenced by the operational conditions and the types of substrates treated (Schievano et al., 2014). To provide further insight, the present study will discuss the performance between

these two typical reactor configurations for AD systems.

Typically, the operating conditions are strongly correlated to the design of a process, which is often obtained from steady state considerations. However, multiple studies have suggested that developing a simultaneous process design and control framework can improve process sustainability (Bahakim & Ricardez-Sandoval, 2014; Palma-Flores & Ricardez-Sandoval, 2022; Sharifzadeh, 2013) . Although different methodologies addressing an integrated process design and control have shown promising results, each study have addressed specific challenges; thus, a general methodology for integrated process design and control is not currently available (Rafiei & Ricardez-Sandoval, 2020a). As DDPG can deal with high dimensional systems under stochastic conditions, this work aims to explore the feasibility of this approach to perform the integration of process design and control for dynamic systems. To the author's knowledge, the study from (Sachio et al., 2021) is the only contribution that has proposed a RL-based methodology for process design and control applications. In that study, a bi-level mixed-integer nonlinear program was split into two problems: a design problem and a control problem; optimal control is computed by using RL, which is later embedded into the optimal design problem; parametric uncertainty was not considered in that study. Consequently, there is a need to provide further insights on the potential of this method to address the optimal design and control of large-scale dynamic systems. In addition, process design and control considerations have not been explored for AD applications, representing another gap in the existing literature.

1.1.- Research objectives

The aim of this study is to explore the feasibility of a model-free RL algorithm applied to two different AD system configurations to treat Tequila vinasses using a Deep Deterministic Policy Gradient (DDPG) (Lillicrap et al., 2015). Two studies will be developed; in the first study, the algorithm DDPG is used to search for optimal process control profiles for these systems. In the second study, the DDPG algorithm will be used for a simultaneous process design and control approach. Additionally, the present study compares the performance of the AD systems using a single-stage and a two-stage reactor configurations under different scenarios involving disturbances and model parameter uncertainty. The specific objectives of the present study are as follows:

- Develop a DDPG-based approach to search for optimal open-loop control actions that minimize the chemical oxygen demand (COD) of Tequila vinasses under disturbances and uncertainties in the system parameters.
- Develop a DDPG-based approach that simultaneously searches for optimal open-loop control actions and a process design that maximizes the annual profits of an AD process treating Tequila vinasses under stochastic disturbances and parametric uncertainty.

To the author's knowledge, the application of the algorithm DDPG for control or simultaneous design and control of AD systems for Tequila vinasses has not been explored in the literature. The aim of this work pursues to contribute to the current literature in the following aspects:

- Consideration of AD systems treating tequila vinasses in closed-loop using RL applications, and

- a new RL-based methodology to address the integration of process design and control of dynamic systems subject to stochastic disturbances and model uncertainty.

1.2.- Outline of this study

The thesis is structured as follows:

 Chapter 2: In this section, a discussion and review of the general state of the art of the literature using RL for process control and process design and control is provided. Necessary background concepts are described in this chapter to facilitate comprehension of the key topics covered in this research and to identify the key gaps in the existing literature related to the topics of this research study.

 Chapter 3: This chapter presents a robust control framework under parametric uncertainty and disturbances using an actor-critic model algorithm (i.e., DDPG). Different scenarios using two different AD configurations are tested and considered in the analysis. The results showed that the proposed DDPG framework was able to learn optimal policies to minimize the organic content matter while producing biogas. The outcomes of this chapter have been disseminated in a journal publication

 Chapter 4: This chapter presents a DDPG-based framework for integration of design and control subject to stochastic disturbances and parametric uncertainty. The combination of time-dependent and time-independent variables were considered for the actions of DDPG, which is a key feature introduced in this work. To add a more realistic representation of a chemical process, parametric uncertainty combined with stochastic disturbances were considered. To deal with process constraints, penalty functions were added to the reward function in the DDPG algorithm. Results showed different advantages and limitations between the two AD configurations. In general,

higher concentrations of biogas resulting in higher plant profits were observed for the two-stage AD system.

- Chapter 5: This chapter presents the main conclusions drawn from this study and suggestions for future lines of research.

CHAPTER 2

Literature review

This chapter presents the current state-of-the-art involving the application of reinforcement learning algorithms in process control and integration of process design and control. This literature review has been performed to clarify the scope of the present work and to better outline their contributions to this emerging area in process systems engineering. To have a better understanding of these topics, some conceptual framework is explained in this section. This chapter begins with section 2.1, where a description of the relationship between reinforcement learning and chemical engineering is provided; section 2.2 discusses the role of RL and the current state-of-the-art studies of RL-based methodologies in process control; section 2.3 provides a general description of neural networks (NNs); section 2.4 provides a description of AD systems and current advances in these systems in terms of process operations and management. A summary of this chapter is provided at the end.

2.1.- Reinforcement learning and chemical engineering

Due to the impacts of climate change, pollution prevention and waste minimization have become of utmost importance in the industrial operation of chemical processes. Likewise, renewable resources have become the centerpiece in fighting climate change, and several industries are considering the transition into cleaner technologies. However, turning into renewable sources requires more efficient and sustainable processes. In addition, designing and operating these types of processes is a challenge as these systems have a complex dynamic behavior that is often subject to external disturbances and uncertainty, which often makes the models computationally intractable. Similarly, the growing demand for globalized and customized products adds more

challenges to process operations, which is an area in engineering that plays a key role in developing sustainable processes. To develop sustainable processes, optimization is needed to specify the operating points that result in most profitable and environmentally friendly operation (Edgar et al., 2001); accordingly, accurate models are needed to provide truly implementable and attractive solutions for these processes. Throughout the history of chemical engineering, mathematical modelling has been of major influence on to design and control chemical processes. One emerging approach to modelling complex dynamic systems is machine learning (ML), which can be defined as a method of data analysis that uses statistical models that are able to learn from experience and explore new trajectories without the need to use classical dynamic programming methods (el Naqa et al., 2015). Although ML was not suitable for chemical engineering applications a few decades ago due to limitations, several improvements in ML have been recently made, thus enabling its potential for modelling complex systems in chemical engineering. Some reasons behind the previous limitations were the limited computational power, lack of data accessibility and lack of efficient programming environments (Schweidtmann et al., 2021). As the computing resources have become more compelling and there is free, open-source ML software publicly available (e.g., Pytorch and TensorFlow), a wide range of possibilities to explore possible applications in chemical engineering have emerged. Recently, ML algorithms have been rapidly developed with several advanced innovations in Artificial Intelligence, such as AlphaGo*, the first computer program that can outperform the most professional human players (Silver et al., 2016). The success behind AlphaGo relies on Reinforcement Learning, which is an area of machine learning that can be described as a sequential learning process, where a model interacts with an environment, executes actions, receives feedback, and executes a new action based on the observations from the feedback (Sutton and Barto, 2018). Another motivation for using RL is its stochastic nature, as it considers

noise while executing the actions. This feature makes it attractive for systems with complex dynamic behavior, i.e., when the model is subject to uncertainty or disturbances, as it could account for the plant-model mismatch.

RL can be considered as a sequential decision-making process, which is a type of problem that is commonly present in chemical process operations. The aim of process operations is the management and use of material, energy, human, capital, and information resources to produce selected products in a reliable, safe, flexible, and cost-efficient as rapidly as possible in an environmentally engaged manner (Edgar et al., 2001). Traditionally, process operations are optimized hierarchically and solved independently for each decision layer. As depicted in Figure 2.1, the hierarchy illustrates the information flow and the specific time scale for each layer.

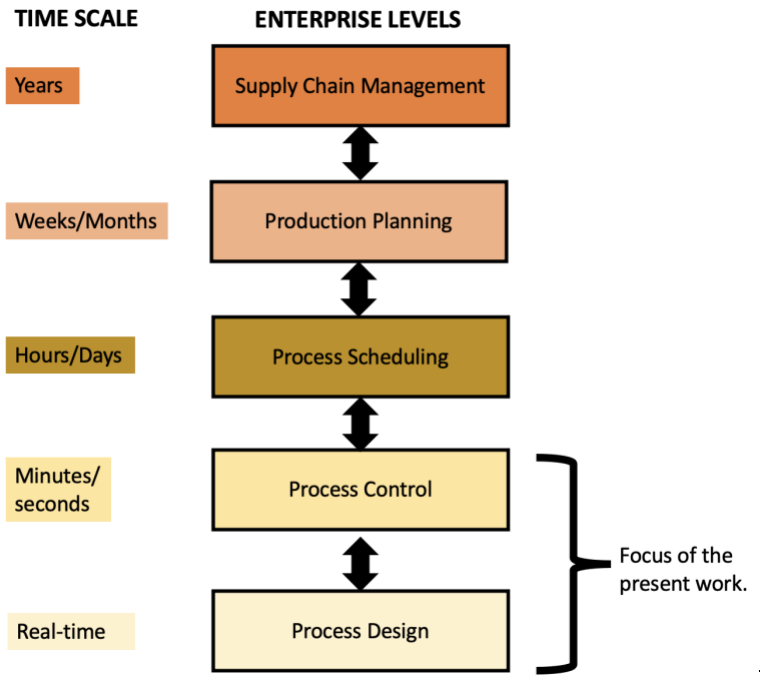


Fig. 2.1.-The five levels for process manufacturing.

Decision-making problems involving time and spatial decisions can be solved using dynamic

optimization methods. As there is a certain resemblance between RL behavior and decision-making problems in process operations, there is a need to explore the potential behind RL to address problems in chemical engineering. Some recent RL studies have shown promising results in different chemical engineering applications, such as process scheduling (Hubbs et al., 2020; Mowbray et al., 2022; Waschneck et al., 2018), dynamic optimization (Boulesnane & Meshoul, 2021; MacHalek et al., 2020; Petsagkourakis et al., 2020a), renewable energy systems (Mendiola-Rodriguez & Ricardez-Sandoval, 2022; Rangel-Martinez et al., 2021; Zhang et al., 2019), optimization of reactions (Lan & An, 2021; Neumann & Palkovits, 2022; Zhou et al., 2017), process design (Sachio et al., 2021), among others.

The aim of this work is to develop a DDPG-based methodology applied to two different areas of process operations, i.e., process control and integration of process design and control. Hence, this chapter will focus on these two aspects. A deep review of the current advances in each of these areas is described next.

2.2.-Reinforcement learning in process control

RL has recently gained interest in process control applications, and thusly, RL-based approaches for chemical engineering applications have been developed over the last years; different algorithms have been tested, such as DDPG (Lillicrap et al., 2015), PPO (Schulman et al., 2017) and TRPO (Schulman et al., 2015). As shown in Table 2.1, these frameworks achieved optimal policies and met the operational requirements. These promising results show that RL can play a key role in industrial process control in the coming years. Although the feasibility of RL has been explored in different applications of process control, e.g., hydraulic fracturing (Bangi & Kwon, 2021), zinc electrowinning processes (Shi et al., 2020) and semi-batch polymerization reactions (Yoo et

al.,2021), there is a gap for RL applications in Anaerobic Digestion systems; hence, this represents one contribution in the present study. Another gap identified is the lack of consideration of parametric uncertainty in the analysis of control systems using RL methods, which represents another contribution of the current work. Adding uncertainty to the problem increases its complexity; thus, more training is required by the RL algorithm. This point will be further discussed in chapter 3. Table 2.1 provides a list of some prominent studies that have presented RL-based methodologies for process control applications.

Table 2.1.- Reinforcement learning algorithms for process control.

Authors	Contribution
(Shi et al., 2020)	A DDPG learning controller was applied to zinc electrowinning processes, which exhibited better results than traditional controllers such as PI and MPC.
(Bangi & Kwon, 2021)	RL controller based on DDPG was implemented to handle hydraulic fracturing to obtain an adequate proppant concentration, and fast learning was observed.
(Yoo, Kim, et al., 2021)	A controller using DDPG for a semi-batch polymerization reaction. That study showed that the controller could learn even with random noise in the environment; parametric uncertainty was not considered in that study.
(Quah et al., 2020)	The algorithm PPO performed slightly better (profits and CPU time) when compared to the PSO algorithm. Nevertheless, PPO was easier to implement as it required less training data.
(Seo et al., 2021)	Developed a based predictive controls scheme by using a DNN and a PPO agent. When comparing this scheme to traditional approaches such as scheduling or the MPC method, the RL controller outperformed the conventional approaches.
(Zheng et al., 2021)	They proposed a RL control scheme to eliminate vortex-induced vibration of a cylinder. The framework was a soft actor-critical algorithm (SAC) and open-source software in the environment (OpenFOAM). The control requirements were met.
(Zhu et al., 2021)	A SAC-based controller was developed and met the specified target conditions. The controller obtained accurate results between the data obtained and the current optimal operating conditions.
(Dogru et al., 2021)	An A3C (Asynchronous Advantage Actor-Critic) controller was able to track an interface between two liquids under uncertainty for applications in the oil sands industry.

2.3.-Process and control design

As mentioned above, this work also focuses on process design and control. This section provides some background information for the integration of process and control design. A review of the main studies in process and control design is provided below.

Traditionally, the layers of process design and process control from the business manufacturing flowsheet depicted in Fig. 1 are solved sequentially. Although solving the layers separately may lead to promising economic solutions, there is no guarantee to achieve those optimal solutions during operation as some assumptions made during the stage design assume ideal conditions, e.g., transients are ignored in the sequential approach. Therefore, a key factor to consider in the early design stages is process dynamics. (LA-Ricardez-Sandoval, 2008; Rafiei & Ricardez-Sandoval, 2020b; Tian et al., 2021; Vega et al., 2014; Yuan et al., 2012)

An integrated design and control approach can be defined as the activity of incorporating both dynamic controllability and steady-state economics. This is particularly challenging as there are some inherent conflicts between dynamic performance (flexibility/disturbance rejection) and efficiency (steady-state economics). Nowadays, current research has justified the integrated framework by developing studies with potential designs with process dynamics, higher profits, and better operability of the process (Bernal et al., 2018; Flores-Tlacuahuac & Biegler, 2007a; Oyama & Durand, 2020; Palma-Flores & Ricardez-Sandoval, 2022b; Patilas & Kookos, 2021; Porru & Özkan, 2019; Tian et al., 2021; Toffolo & Ricardez-Sandoval, 2021). Table 2.2 provides a list of some prominent studies that have presented methodologies for the integration of design and control.

Table 2.2.- Studies with simultaneous design and control methodologies.

(Mohideen et al., 1996)	A unified framework using a robust stability criterion to solve a mixed-integer stochastic optimal control problem.
(Flores-Tlacuahuac & Biegler, 2007b)	A simultaneous dynamic optimization approach was implemented for a problem with relatively few integer variables and large NLP problems.
(Alvarado-Morales et al., 2010)	An integrated process design and controller design (IPDC) and a process-group contribution approach were presented.
(Sánchez-Sánchez & Ricardez-Sandoval, 2013)	Process and control design of dynamic systems under uncertainty is addressed through a simultaneous approach of dynamic feasibility and dynamic flexibility in a single optimization formulation.
(Trainor et al., 2013)	Stability and robust feasibility analysis are considered simultaneously for a framework to address convex mathematical problems. The methodology does not require the solution of an MINLP.
(Mansouri et al., 2016)	An integrated design and control method for reactive distillation processes is represented as a binary system. Different algorithms based on the element concept are considered in this method. An optimal design-control solution was obtained, and the study considered disturbances in the feed.
(Bansal et al., 2003)	A formulation based on generalized Benders' decomposition principles. This approach is independent of the solving method for the primal dynamic optimization problem.
(Bahakim & Ricardez-Sandoval, 2014)	A stochastic-based simultaneous design and control framework is developed. The dynamic variability of the system is determined using a stochastic-based worst-case variability index, which is computed from the PDF (assumed Gaussian) of the worst-case variability index. Flexibility in the design stage is considered a trade-off between attractive economical solutions and stable conventional designs.
(Diangelakis et al., 2017)	A PAROC framework that addresses design and control problems through multi-parametric programming.
(Meidanshahi & Adams, 2016)	An integrated design and control framework for a semicontinuous distillation system. The MIDO problem formulation is addressed via the deterministic outer approximation method by using a built-in optimization package of gPROMS.
(Koller et al., 2018)	A back-off methodology for integration of design and control, and scheduling under stochastic realizations in the disturbances of a multiproduct CSTR system. Using Monte Carlo sampling to generate random realizations. The effect of stochastic disturbances and uncertainties is approximated by a flexibility analysis incorporating back-off terms.
(Rafiei & Ricardez-Sandoval, 2018)	Methodology based on the back-off to address a simultaneous design and control problem. The main idea is to simulate the confidence interval of process constraints by the usage of power series expansion

	(PSE)-based functions, which are designed by Monte Carlo sampling. This approach considers stochastic descriptions in disturbances and parametric uncertainty in a wastewater treatment plant.
(de Carvalho & Alvarez, 2020)	Infinite horizon model predictive control is the basis of the methodology. The problem is divided into three stages that are solved by goal attainment and quadratic cost techniques. This methodology outperformed the traditional sequential configurations.
(Rafiei & Ricardez-Sandoval, 2020c)	The Trust region method is considered to address optimal process design for a large-scale system under uncertainty. PSE is used as a surrogate model to reduce the complexity of the problem.
(Sachio et al., 2021)	A bi-level MIDO is solved through a RL-based controller and mathematical programming. The outer optimization addresses the design problem while the inner optimization is solved through a Policy Gradient Algorithm.
(Palma-Flores & Ricardez-Sandoval, 2022)	An NMPC-based framework that results in a bilevel optimization. A classical KKT transformation is used to transform the problem into a single-level dynamic optimization problem.

As shown in Table 2.2., these frameworks presented different methodologies to address the design and control problem for different case studies. Nonetheless, very few studies have considered disturbances defined as time-varying random events. Some studies that have considered stochastic disturbances are (Bahakim & Ricardez-Sandoval, 2014; Koller et al., 2018; Rafiei & Ricardez-Sandoval, 2018). As these stochastic inputs can take random (unknown) values at any time *t*, the problem becomes a stochastic infinite-dimensional, which is prone to be computationally demanding. Therefore, there is a need to develop approaches that can take into consideration stochastic disturbances, which is a key motivation for the present research work. With this in mind, stochastic disturbances are considered in this work, contributing to this limited area in the literature for integration of design and control.

Although the area of simultaneous design and control has been studied since the 1960s, there is a lack of studies on its application in reinforcement learning. As shown in Table 2.2., the study from Sachio et al., (2021) has been the only work that has applied RL to address simultaneous design and control. The methodology proposed in that work used a policy gradient method for the outer

optimization and was tested in two small case studies obtaining promising results. Thus, there is a need to explore the feasibility of more RL algorithms, and this represents a novelty in the present study. Another contribution of the present research study is the fact that the methodology considers solving the entire process design and control problem through the DDPG-based framework; this is done by solving an economic function that enforces constraints within the same function. This will be discussed in detail in Chapter 3.

The Deep Deterministic Policy Gradient belongs to a class of RL algorithms called Actor-Critic (AC) models, where the main feature is the synergy of two different DNNs, one for action prediction (called actor-network) and another referred to as critic-network that aims to evaluate the predicted action through a Q-value. To improve stability to the algorithm, DDPG also includes a target DNN for the actor and a target DNN for the critic. A detailed description of the algorithm will be provided in chapter 3. As the architecture of the proposed algorithm DDPG consists of 4 Deep Neural Networks (DNNs), it is important to provide a general description of the performance of a neural network (NN). This will be discussed in the next section.

2.4.-Neural networks and their training

A neural network is a system approximator with interconnected nodes that aims to recognize relationships between the input data. It is inspired by the learning process of the human brain, imitating how the biological cells pass information by synapses. Neural networks have the advantage of being able to detect complex nonlinear relationships between variables (Tu, 1996); hence, they are good approximators. The main components of a neural network are the neuron and the activation function. The input layer receives the states (represented as x_1, $x_2 \ldots x_m$ in Fig. 2.2), where *m* is the total number of elements in the input layer, then the information is passed through

hidden layers to obtain an output layer. The hidden layers aim to determine the relationship between inputs and outputs (Morán et al., 2017).

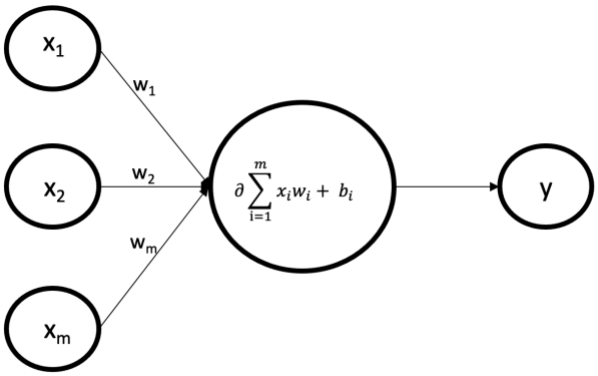


Fig. 2.2.- Schematic representation of a NN.

The values calculated in the hidden layers are passed to the next layer. The number of hidden layers depends on the complexity of the problem under consideration. As depicted in Fig 2.2, inside a neuron, the weighted sum of the weights w_1, w_2… w_m and biases b_i across the set "i", $\forall i \in \{1, \dots, m\}$ are calculated, and then it is multiplied by an activation function ∂, which is a function assigned to a neuron that decides the signal that will be passed as the output of the neuron (represented as *y*). The most used activation functions in chemical engineering applications are the rectified linear unit, sigmoidal transfer function, and the hyperbolic tangent function. The rectified linear unit function (also known as ReLu) thresholds values at 0, i.e., it outputs 0 when the inputs are negative values; this function is widely used in convolutional NN. The sigmoid function exists between the range [0,1], it is differentiable, and its output is interpreted as a probability function. Thus, this function is commonly used for NNs regression. Moreover, the hyperbolic tangent activation function (also known as tanh) takes inputs that are negative, neutral, and positive. However, this function requires a significant computational effort due to the exponential terms involved in the function (Cavalcanti et al., 2021).

The training of an artificial neural network consists of the adjustment of all the weights such that they provide accurate output predictions. The weights and the biases are the neural network parameters. From the same analogy of a NN imitating the human brain, the NN parameters could be seen as the assigned weights from the sensors of the synapses. When the weights and the biases are adjusted, this can be interpreted as how much the states (input layer) change. The adjustment of the parameters decides to what extent/proportion the signal is passed along the neuron. To train the weights and biases, an error calculation between the output value predicted by the NN and the target value is performed. For the updating process of the weights, the algorithm of backpropagation is considered (Rumelhart et al., 1986). Traditional backpropagation algorithms use gradient descent to minimize the error. To know to which extent the parameters should be modified, the computation of the gradients with respect to w_i and b_i are estimated. Generally, at the beginning of the training process, the weights and the biases of the NN are generated randomly, and once the training continues, w_i and b_i are constantly updated until the error is minimized.

A neural network that involves multiple layers is referred to as a deep neural network (DNN). As mentioned above, the specific performance of DDPG and how DNNs are used in this framework is discussed in chapter 3.

2.5.- Anaerobic digestion systems

As the world's population is rapidly increasing, the problem of wastewater management becomes a threat. Renewable energy systems are therefore of utmost importance to diminish the increasing levels of water pollution. Bioenergy is one class of renewable energy generated from living organisms or their bioproducts that plays an essential role in decarbonizing energy systems. An attractive alternative is biochemical conversion processes, which transform organic material into valuable fuels such as biogas and bioethanol using microorganisms (Singh & Olsen, 2011).

Among the different biochemical conversion technologies, the present work focuses on anaerobic digestion systems, which aim to reduce organic content in wastewater while producing biogas. Organic matter content in wastewater is usually measured by the COD, which estimates the concentration of dissolved oxygen required for the organic to be oxidated. A high COD concentration may lead to dangerous environmental consequences as it can consume dissolved oxygen from water bodies. A common consequence of high COD is the disturbance of aquatic life, affecting marine diversity and affecting human health via its consumption (Chukwu, 2008).

Biogas is a flexible renewable energy source that can potentially substitute fossil fuels such as natural gas. Moreover, the digestate is the waste stream from AD, mainly composed of a concentrated agricultural complex that can be potentially used as fertilizer (Nkoa, 2013). Some of the main operational and economic advantages of AD are lower amounts of nutrients required for microbial growth compared to aerobic systems, small reactor volumes and low energy inputs. Likewise, AD can be widely applicable to different substrates that emerge in industrial organic wastewater (Kleerebezem et al., 2003), agriculture (Merlin et al., 2021), food waste (Zhang et al., 2007), and municipal solid waste (Hartmann & Ahring, 2005).

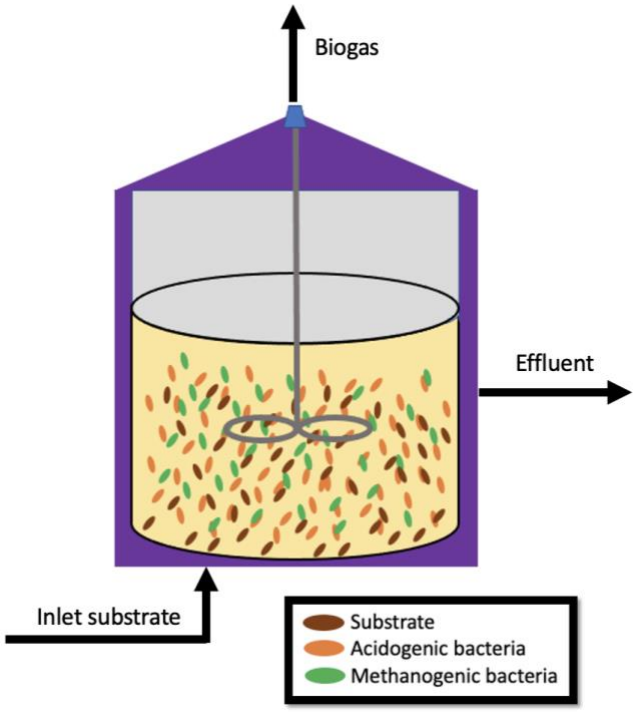


Fig. 2.3.- Scheme representation of an AD system.

The AD process often involves the following steps: hydrolysis, acidogenesis, acetogenesis and methanogenesis. The main groups of microorganisms involved in AD are acidogenic and methanogenic bacteria (as depicted in Fig. 2.3) The most conventional digester configuration consists of a single reactor where all the reactions involving these steps take place (Ahring et al., 2003). Méndez-Acosta et al., (2010) suggested a single-stage treatment for a lab-scale AD system to treat Tequila vinasses. Wan et al., (2013) also indicated a single-stage AD configuration to treat food waste. Several studies have also suggested that a two-stage structure (one reactor for the acidogenic stage and another one for the methanogenic stage) allows the development of appropriate conditions for the different microbial communities and increases the methane yield and stability of the process (Aslanzadeh et al., 2014; Bouallagui et al., 2004; Liu et al., 2002; Yang et al., 2003). Although the two-stage reactor configuration has been suggested to offer better

performance (i.e., higher biogas yields at low COD), some studies suggest that a two-stage structure may not be suitable for all types of substrates. Schievano et al., (2014) reported a similar performance regarding biomethane generation for the two-reactor configuration and the single-stage reactor treating maize silage, waste rice flour, olive pomace and waste fruit. Shen et al.,(2013) indicated a better performance from a single-stage model regarding methane production at a lower Organic Loading Rate (OLR). Lindner et al., (2016) suggested that the substrate composition influences the reduction of organic content matter and methane production.

As with any biological process, AD is subject to multiple operational conditions, e.g., disturbances in the loading rate, lack of knowledge (i.e., uncertainty) in the biochemical reaction kinetics and the conditions that generate an adequate bacterial growth (Ferenci, 1999). Likewise, AD is extremely sensitive to process disturbances that may result in low biogas yield, abrasive hydrogen sulphide production and unstable pH conditions. Hence, adequate and effective control systems for AD systems are needed to achieve a stable and dynamically feasible operation of the process while achieving high biogas yields (Nguyen et al., 2015).

In general, systems that exhibit different complex dynamic behaviour represent a great opportunity to explore the capacity of RL. That being the case, the aim of this study is the exploration of the feasibility of RL applied to AD systems. The feasibility of reinforcement learning has not been widely explored for AD applications, and thus, it represents a potential area for improvement. Chen et al., (2021) applied multi-agent deep reinforcement learning (MADRL) to optimize the dissolved oxygen and chemical dosage in a Wastewater treatment plant. The model was developed by Hydromantis GPS-X and was tested under different scenarios to identify the key factors that

influence the costs. The results suggested that more data is needed to apply this method to large-scale systems. Pettigrew & Delgado, (2016) developed a RL framework by using neural networks in a Java simulation environment to optimize the flow of a wastewater treatment plant, obtaining an improvement in methane yield in a two-stage AD system. Despite these efforts, more research is indeed necessary to explore the benefits and potential adoption of these technologies in bioprocess systems. In particular, those previous studies did not consider disturbances or uncertainty; the present study aims to address these gaps in the literature, as it will be discussed in detail in chapter 4.

2.6.-Summary

This chapter presented a review of relevant studies in process control and integrated process design and control using RL. In the first section, the relationship between chemical engineering and reinforcement learning was explained with the aim of showing the incentive behind the application of RL, and in particular the DDPG algorithm, in the present research study. The advantages of using RL for different fields in process operations were described, such as the capacity of considering stochastic disturbances and parametric uncertainty. Moreover, the literature review for process control and simultaneous design and control was presented with the purpose of showing the state-of-the-art studies in these areas as well as their corresponding challenges. Results from this review revealed that RL applied in process control is a relative new research area that has recently gained interest in different chemical engineering applications, with a very limited number of studies using RL-based methodologies applied to biological systems, such as AD systems. Similarly, there is a lack of studies addressing an integrated process design and control approach using RL, as currently just one work using a RL-based methodology has been proposed to address simultaneous design and control. To have a better understanding of the background information of

DDPG, the training process of a NN was discussed. Moreover, the introduction and background information on AD systems were presented, as well as the motivations and challenges of these types of systems. The next chapter describes the methodology of the DDPG approach and its application to process control.

CHAPTER 3

Robust control

This chapter presents the methodology of the DDPG applied to process control for AD systems treating Tequila vinasses. The first section describes the proposed methodology, where the relationship between dynamic optimization and reinforcement learning is described. Section 3.1.2. provides a description of the DDPG algorithm. In section 3.2, the case study of the process control approach is presented. Section 3.3 presents the problem statement related to the case study. Section 3.4 describes the performance of the AD systems using a single-stage and a two-stage reactor configuration under different scenarios involving disturbances and model uncertainty. In addition, the performance of the RL controller is compared to an optimal open-loop controller using a conventional multi-scenario dynamic optimization framework. Furthermore, a robust economic model predictive controller (EMPC) using the DDPG is tested under disturbances and parametric uncertainty. The outcomes of this chapter have been presented in a journal publication (Mendiola-Rodriguez & Ricardez-Sandoval, 2022)

3.1.-Methodology

This section describes the RL methodology from a Dynamic optimization framework. Also, the DDPG framework for optimization of discrete-time nonlinear processes is presented.

3.1.1.- Dynamic Optimization and Reinforcement Learning

One prominent feature of most chemical processes is their intrinsic dynamic nature to deal with disturbances and parametric uncertainties continuously during operation (Ricardez-Sandoval et al., 2009). Nowadays, among the increasing demand for customized products and globalization (Calvo

et al., 2022), the application of efficient process control schemes has become more challenging to meet product demands while achieving sustainable and economically viable processes. To circumvent this issue, dynamic optimization strategies are often considered to improve dynamic systems' controllability and process operations under uncertainty. A generic dynamic optimization problem (DOP) subject to parametric uncertainty can be formulated as follows:

$$\min_{u_t} \sum_{j=1}^{M} \omega_j \left(\mathbf{x}_{t,j},\ \boldsymbol{u}_t, \widehat{\boldsymbol{y}}_{t,j};\ \boldsymbol{\theta}_j\right) \tag{3.1}$$

$$f\left(\boldsymbol{x}_{t,j}, \boldsymbol{u}_t, \boldsymbol{\theta}_j\right) = \widehat{\boldsymbol{x}}_{t,j} \qquad \forall j \in \{1, \dots, M\}$$

$$h\left(\widehat{\boldsymbol{x}}_{t,j}, \boldsymbol{u}_t; \boldsymbol{\theta}_j\right) = \widehat{\boldsymbol{Y}}_{t,j} \qquad \forall j \in \{1, \dots, M\}$$

$$g\left(\widehat{\boldsymbol{x}}_{t,j}, \boldsymbol{u}_t; \boldsymbol{\theta}_j\right) \leq 0 \qquad \forall j \in \{1, \dots, M\}$$

$$\boldsymbol{x}_t = \boldsymbol{x}_0 \qquad \text{for t=0}$$

$$\boldsymbol{u^l} \leq \boldsymbol{u_t} \leq \boldsymbol{u^h}$$

where $\boldsymbol{\theta}_j \in \mathbb{R}^{n_\theta}$ is the set of time-independent realizations for the uncertain parameters across the set "j," $\forall j \in \{1, \dots, M\}$, where M represents the number of uncertainty realizations; $\widehat{\boldsymbol{x}}_{t,j} \in \mathbb{R}^{n_x}$ is the differential states for each uncertain realization *j* and every time step *t*, $\boldsymbol{u}_t \in \mathbb{R}^{n_u}$ is the control profile vector, u^l and u^h denote the lower and upper bounds for the control vector, $\widehat{\boldsymbol{y}}_{t,j} \in \mathbb{R}^{n_y}$ are the controlled variables for each realization, $\widehat{\boldsymbol{Y}}_{t,j}$ is the algebraic variable vector, $\boldsymbol{x}_t \in \mathbb{R}^{n_x}$ (*t=0*) is the state vector used to denote the initial conditions ($\boldsymbol{x}_0$); $f: \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\theta} \longrightarrow \mathbb{R}^{n_x}$ represents the set of nonlinear differential equations representing the system dynamics; $h: \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\theta} \longrightarrow \mathbb{R}^{n_Y}$ symbolizes the set of algebraic equations and $g: \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\theta} \longrightarrow \mathbb{R}^{n_Y}$ denotes the set of inequality constraints; ω_j are the weights assigned for each uncertainty realization *j*

considered in the formulation. Optimization problems in functional spaces are known as optimal open-loop control problems, which aim to search for the optimum of a specific function by manipulating a set of input (manipulated) variables available for control. While problem (3.1) is a conventional multi-scenario formulation to deal with parametric uncertainty, a fundamental assumption is that uncertainty remains static during the analysis. In real operations, systems may be subject to uncertain parameters that are continuously changing during operation. Also, the usual plant-model mismatch often found in chemical systems due to additional uncertainty not considered in the model (f) and process constraints (i.e., h and g) may impact the economics and operation of the process.

Recently, RL has emerged as an attractive technique to solve complex sequential decision-making problems like DOP. RL is made up of a set of agents (i.e., available variables for control) that iteratively interact with an environment (i.e., a process) to take decisions at every time-step *t*, seeking to find an optimal strategy referred to as a policy. In RL, an agent learns the best path through exploitation and exploration strategies to achieve a specific assignment (e.g., optimize an objective function). A key feature in RL is that it only requires input-output data instead of acquiring a fully dynamic model, which enables the application of this method to high-dimensional problems (Tang & Daoutidis, 2018). In addition, RL can predict data offline to optimize online computation time (Bemporad et al., 2002) and the possibility of using the same algorithm to learn different tasks (Nian et al., 2020). On the other hand, more complex systems also require larger data sets that would eventually impact the efficiency and convergence of this method (Shin et al., 2019). Nevertheless, exploring this method to offer optimal control solutions in reasonable turnaround times is still needed to fully assess their benefits and limitations.

Fig. 3.1 presents a potential adaptation of DOP in the context of the RL algorithm. The agent is the decision-maker; different algorithms can be used as agents in RL, e.g., State-action-reward-state-action (SARSA) (Rummery & Niranjan, 1994), REINFORCE (Williams, 1992), Asynchronous Advantage Actor-Critic (A3C) (Mnih et al., 2016), Deep Deterministic Policy Gradient (DDPG). In DOP, this concept would correspond to the numerical approach used to solve the optimization problem. The agent interacts sequentially over a discrete-time with the environment *E*, modelled as a Markov Decision Process (MDP) model. A MDP consists of a state-space S $\in \mathbb{R}^N$, an action space A $\in \mathbb{R}^N$, a transition function $\mathcal{P}(s_{t+1}|s_t, a_t)$, $\mathcal{P}$: S x A x S $\rightarrow \mathbb{R}$ that represents the probability of an action to be selected, and a reward function $r(s_t, a_t)$, which is given after an action was taken. In DOP, *E* would correspond to the mechanistic process model representing the system's transient behavior, which is often described by a set of nonlinear differential equations, algebraic equations, and constraints, i.e., *f*, *h* in Eq. (3.1). Penalty functions are usually added in RL to deal with process constraints g (Tessler et al., 2018; Yoo, Zavala, et al., 2021). At every time-step *t*, the agent interacts with the environment by observing the states $s_t \in$ S from the environment, which describes the current state of the process. In DOP, this would be depicted as the states presented of the DOP system $\hat{x}_{t,j}$. Based on the observations, the agent executes an action: each action $a_t \in$ A. In DOP, the variables used for control (i.e., u_t)would be interpreted as the actions taken to optimize the process. Once the action is executed in the environment, the environment changes and produces a new state s_{t+1}. Both RL and DOP aim to optimize a performance objective function referred to as the reward function in RL. In both cases, the actions selected influence the execution of the strategy. However, the objective in RL is to maximize the expected accumulated sum of discounted future rewards R_t using state-action pairs (s_t, a_t), i.e.,

$$R_t = \sum_{i=t}^{T} \gamma^{(i-t)} r(s_t, a_t) \tag{3.2}$$

where $\gamma \in [0,1]$ is a discount factor that describes the priority of immediate rewards over future rewards; usually, γ is set to a closer value to 1, i.e., it prioritizes the immediate rewards to emphasize the action of the state at time *t*. Note that γ must be within the bounds [0,1] to guarantee a finite convergence. Moreover, the reward in DOP would be indicated as the objective function in Eq. (3.1). The goal of the agent in RL is to find an optimal policy μ that maximizes R_t, i.e.,

$$Max_{\mu}[R_t|s_t, a_t] \tag{3.3}$$

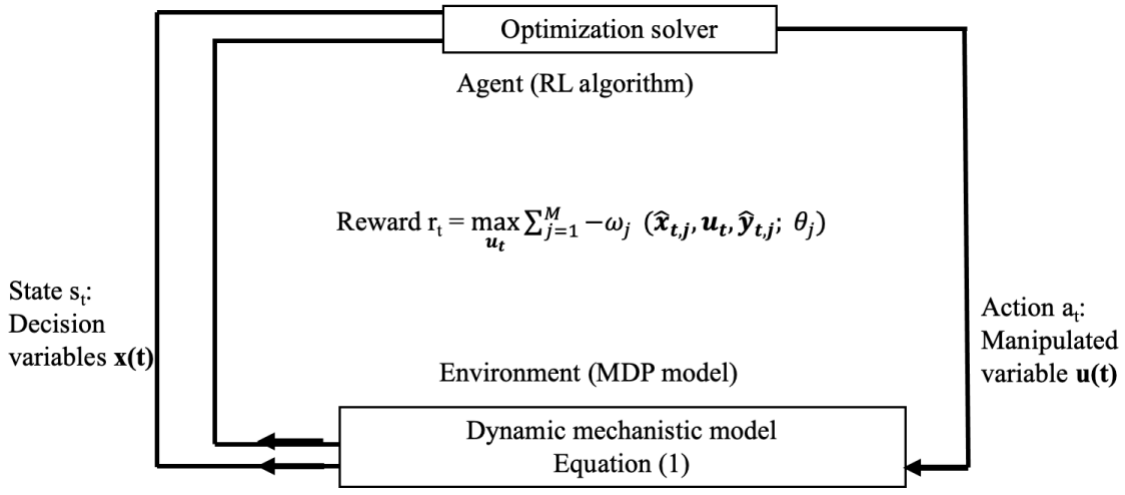


Fig. 3.1.- Schematic DOP through a RL framework.

The agent's actions are directly driven by a policy μ that depicts states to a probability distribution over the actions $\mu: SP \rightarrow (A)$, that is, the policy is the agent's strategy to find the optimal control trajectory. In the DOP shown in problem (1), the policy μ would be equivalent to the control actions u_t. The agent decides and executess actions to optimize a process. In this work, the DDPG algorithm is the RL agent. This is explained in detail next.

3.1.2.-Deep Deterministic Policy Gradient

This algorithm is a combination of Policy Gradient and Deep Q learning methods. As depicted in Fig. 3.2, the structure comprises two Deep-Q neural networks (DNN), two target networks, an environment (*E*), and a buffer memory. This algorithm includes a set of model parameters and hyperparameters. The former are the weights of the Deep-Q, neural networks (ϕ^{Q}, ϕ^{μ}, $\phi^{Q'}$, and $\phi^{\mu'}$), which work as function approximators, whereas the latter are variables that define the learning process, i.e., how the DNNs are trained. Hyperparameters have a strong influence in the algorithm's learning behaviour; hence, tunning of these parameters is necessary to develop good learning performance. These hyperparameters, together with the learning process of DDPG, are described below. Each element in the DDPG framework is explained next.

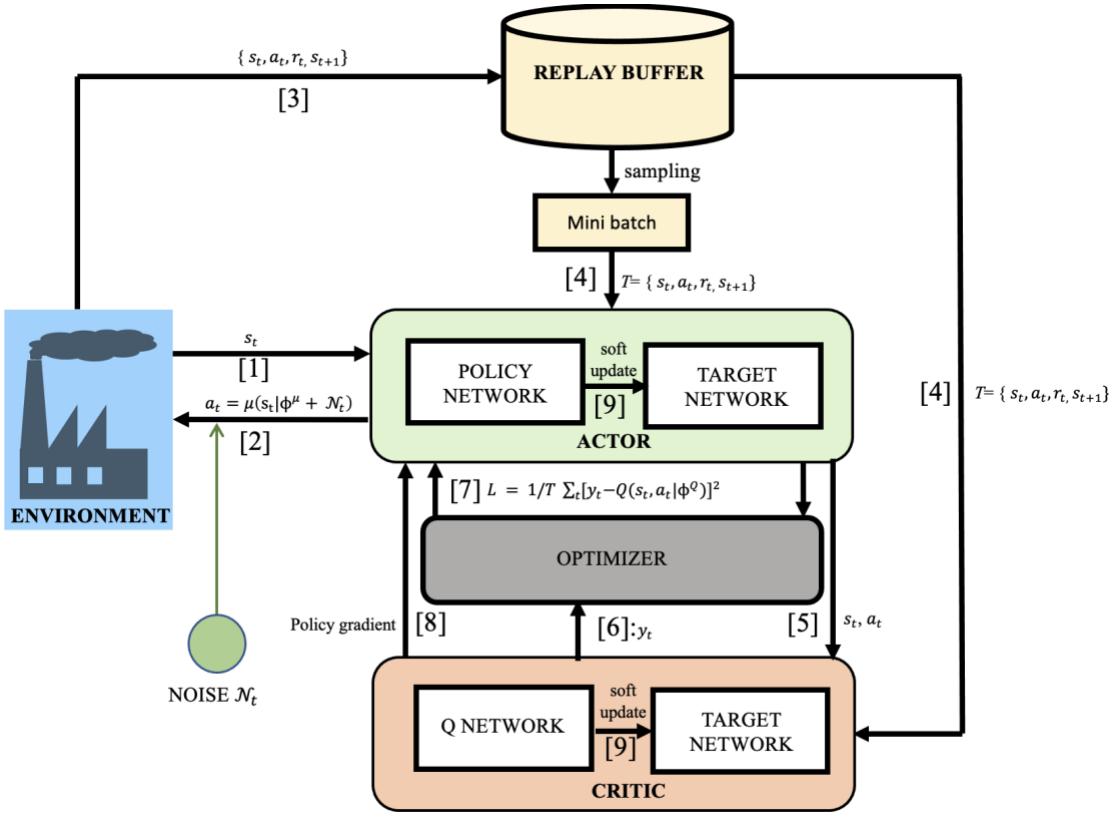


Fig. 3.2.- DDPG algorithm structure.

3.1.2.1.- Q network and target network.

The key concept behind the critic DNN is Q-learning, which is a value-based method (Watkins & Dayan, 1992). The critic DNN, also called Q-Network, is composed of a Q value function $Q(s_t, a_t|\phi^Q)$ with model parameters ϕ^Q. The critic calculates an optimal Q value function (Q^μ) to find an optimal action $a^\mu(s_t)$, i.e.,

$$a^\mu(s_t) = argmax Q^\mu(s_t, a_t) \tag{3.4}$$

A Q value can be defined as the expectation of the reward after one action was executed and can be seen as an indicator of the goodness of that action towards finding the optimal policy μ. Finding optimal actions for discrete spaces is straightforward. In this study, action spaces are considered

continuous and therefore, the optimization problem becomes more challenging since the computation of the max operator over the actions in Eq. (3.4) becomes intractable as there is in principle, an infinite action space to explore. To circumvent this issue, $Q^{\mu}(s_t, a_t)$ is assumed to be differentiable with respect to the actions a_t, which enables the use of a gradient-based policy $\mu(s_t)$ in Eq. (3.4), i.e.,

$$max Q^{\mu}(s_t, a_t) \approx Q\ (s_t, \mu) \tag{3.5}$$

This gradient-based policy $\mu(s_t)$ will be further discussed in the actor DNN process. The optimal-value function $Q^{\mu}(s_t, a_t)$ is described by the Bellman equation, i.e.,

$$Q^{\mu}(s_t, a_t) = \mathrm{E}[r(s_t, a_t) + \gamma max Q^{\mu}(s_{t+1}, a_{t+1})] \tag{3.6}$$

With this recursive equation, the agent starts by estimating an approximator to the optimal value function $Q^{\mu}(s_t, a_t)$. The critic network, i.e., Q network, is updated by calculating the loss function, which is represented by the mean-squared Bellman error (MSBE):

$$L\ =\ 1/T\ \sum_t [y_t - Q(s_t, a_t|\phi^Q)]^2 \tag{3.7}$$

where:

$$y_t = r(s_t, a_t) + \gamma Q'\left(s_{t+1}, \mu'\left(s_{t+1}\middle|\phi^{\mu'}\right)\middle|\phi^{Q'}\right) \tag{3.8}$$

where T is a minibatch of randomly sampled transitions from the replay buffer. The purpose of the replay buffer is to keep track of previous experiences (see next section). y_t represents the target value whereas $Q'$and μ'represent the outputs of the target critic network and actor-critic network, respectively. The use of target networks is a crucial innovation in the DDPG algorithm. Target DNNs are delayed copies of the actor and critic network to stabilize the process, i.e., to avoid divergence of the algorithm. As shown in Fig. 3.2, the model parameters of the Q-network and Policy network are copied to the Target DNNS (step 9 from Fig. 3.2), but to avoid the dependence

of the target NN on the same model parameters $\emptyset$, the target DNNs are periodically updated through a "soft update" (Eq. (3.9)), i.e., target DNN's parameters $\emptyset^{Q'}$, $\emptyset^{\mu'}$ are slowly copied over from the Q-network and Policy network.

$$\phi^{Q'} \leftarrow \tau\phi^{Q} + (1-\tau)\phi^{Q'} \tag{3.9}$$

$$\phi^{\mu'} \leftarrow \tau\phi^{\mu} + (1-\tau)\phi^{\mu'}$$

where τ is the hyperparameter that determines how often the weights of the target networks are updated and often ranges between 0.001 and 0.01. Soft update is important because if there were no time delay for the target DNNs in Eq. (3.7), the MSBE would be chasing a moving target, making this algorithm unstable. As shown in Eq. (3.8), the target DNNs μ' and Q' are necessary to calculate the target action value y_t. Therefore, by minimizing the loss function L in Eq. (3.7), the algorithm aims to make the Q-value function $Q(s_t, a_t|\phi^{Q})$ as close as possible to the target value y_t that accounts for the Q value for the following state (s_{t+1}); i.e., the agent seeks to minimize the error between the Q-value function of the current time step and the predicted Q-value function from the target DNNs; this is referred to as a Temporal Difference (TD) learning feature and hence, the error that DDPG seeks to minimize is referred to as the TD error, which is optimized through stochastic gradient descent to update the critic network. A key parameter for the update process is the learning rate, which determines how the model changes in response to the TD error. A small value of the learning rate could lead to divergence due to the lack of flexibility of the model, whereas large learning rate might result in an unstable training or a suboptimal set of parameters.

3.1.2.2.- Buffer memory

A replay buffer stores the agent's experiences more efficiently and has the purpose of tracking a *P* finite (user-defined) number of previous experiences in the form of tuples$\{s_t, a_t, r_t, s_{t+1}\}$; once the accumulated experiences reach *P* number tuples, the old experiences are discarded. For the training process, minibatches of size *T* of the experiences accumulated are sampled randomly to help developing a stable behaviour for the algorithm. The size of the buffer depends on the type of problem because few data might lead to overfitting, whereas a wide range of data might slow the learning process. To avoid divergence of the algorithm, the data used in the training process should be independently distributed. Therefore, at every time step, the DNNs are updated by stochastic sampling a minibatch from the buffer to reduce correlations between samples.

3.2.2.3.-Policy Network

The policy network, also referred to as actor network, observes a state from the environment and executes a continuous action every step time *t*. As with any RL algorithm, the agent's goal is to find the optimal policy μ; DDPG applies a policy gradient method to find the optimal policy where a parameterized function $\mu(s\,|\phi^{\mu})$ maps a state deterministically to a specific action. Using a deterministic policy avoids the high computational cost of calculating the gradient of the state distribution. It has been shown that the gradient of the policy's performance is equivalent to the deterministic policy gradient (Silver et al., 2014). By applying the chain rule shown in Eq. (3.10), the derivative of the policy can be estimated and, consequently, the actor-network and actor-target network parameters ϕ^{μ} and $\phi^{\mu'}$ can be updated through gradient ascent as the objective is to maximize the reward.

$$\nabla_{\emptyset^{\mu}} J \approx \mathbb{E}_{s_t}\left[\nabla_{\phi^{\mu}} Q(s,a|\phi^{Q})|_{s=s_t,a=\mu(s|\phi^{\mu})}\right] \tag{3.10}$$

$$= \mathbb{E}_{s_t}\left[\nabla_a Q(s,a|\phi^{Q})\,|_{s=s_t,a=\mu(s|\phi^{\mu})} \nabla_{\phi^{\mu}} \mu(s\,|\phi^{\mu})\,|_{s=s_t}\right]$$

To enable exploration, i.e., exploring new control paths, a Gaussian-type process, called Ornstein-Uhlenbeck (OU) noise (Uhlenbeck & Ornstein, 1930) (denoted by $\mathcal{N}_t$) is added before the action is returned to the environment (Eq. (3.11)). This noise can be interpreted as a to plant-model mismatch and samples noise from a correlated normal distribution:

$$a_t = \mu(\mathrm{s_t}|\phi^{\mu} + \mathcal{N}_t) \tag{3.11}$$

Zero-mean Gaussian noises have been reported as suitable for the DDPG (Liang et al., 2020). This feature is attractive from a process control perspective, as most of the problems are complex and nonlinear and can be computationally expensive with conventional software. A feature of the DDPG algorithm is its model-free structure, which means the transition probabilities are unknown, i.e., the agent does not have previous knowledge of how the environment works, which can be convenient for processes where no *a priori* information is available. As depicted in Fig. 3.2, the steps for the DDPG learning process per episode can be summarized as follows:

[1] The initial conditions of the states of the model are observed.

[2] The policy network receives the state and outputs an action. A Gaussian-type process, called Ornstein-Uhlenbeck (OU) noise, is added before the action is returned to the environment (Eq. (3.11)). This noise is interpreted as accounting for the plant-model mismatch.

[3] According to the action executed, the environment receives a reward r_t and produces a next state s_{t+1}. The MDP elements (s_t, a_t, r_t, s_{t+1}) are stored in the buffer memory.

[4] Randomly sample a minibatch of T transitions (s_t, a_t, r_t, s_{t+1}) from the buffer memory.

[5] Critic network receives the state s_t and action a_t

[6] The critic network computes the target y_t (Eq. (3.8))

[7] Update the critic through the minimization of the loss L (TD error, Eq. (3.7))

[8] Update the actor-network through the derivative of the policy (Eq. (3.10))

[9] Target networks update (Eq. (3.9))

This learning process is terminated when a user-defined number of episodes are reached, or a user-defined stopping criterion is satisfied, e.g., reward not improving over a large number of episodes.

3.2.-Case study: anaerobic digestion for tequila vinasses treatment

3.2.1.- Problem statement

One potential candidate for AD systems is Tequila vinasses. This waste is continuously increasing due to the high global demands of Tequila, which is one of the most traditional beverages in Mexico and the World (Colunga-GarcíaMarín & Zizumbo-Villarreal, 2006). According to the Mexican Tequila Regulatory Council, the production of Tequila was 374 million liters in 2020 (40 \% Alc. Vol) (*CRT*, 2021) representing a 6.3 % increase in production with respect to 2019.

Tequila vinasses are a by-product in the production of Tequila and are considered a pollutant due to their high temperature, low pH, and high organic. López-López et al., (2010) indicated that 10-12 liters of vinasses are produced for each liter of Tequila produced. Hence, 3740 million liters of vinasses were generated in 2020; this amount is equivalent to 48620 tons of organic material measured as biological oxygen demand. Due to inadequate technologies for residual water treatments, these highly polluted effluents are typically discharged into water bodies. Considering no post-treatment, the vinasses released in 2020 would have been equivalent to the annual pollution from a population of 2.46 million people, representing a major environmental

problem. Therefore, there is a great incentive to meet the environmental requirements and exploit the potential valorization of vinasses' high organic matter content (Arreola-Vargas et al., 2016). These characteristics make Tequila vinasses suitable for AD treatment. Thus, efficient operational and control strategies are needed to treat the Tequila vinasses and satisfy the environmental regulations that allow Tequila's clean and sustainable production. Studies on anaerobic digestion applied to Tequila vinasses treatment have proposed different control strategies. Méndez-Acosta et al., (2010) suggested a sampled delayed control for the chemical oxygen demand. Lizarraga-Palazuelos et al., (2013) proposed a saturated linear PI Output-Feedback controller, whereas Méndez-Acosta et al., (2016) designed a hybrid control scheme for the VFA and COD regulation. Recently, Piceno-Díaz et al., (2020) reported a multi-scenario robust NMPC for a two-stage AD process.

An emerging control strategy that has not been widely explored for AD treatment is the application of RL techniques. However, to the author's knowledge, the application of a reinforcement learning methodology applied to AD systems treating Tequila vinasses has not been proposed, thus, representing one contribution of the current research.

3.2.2.-Mathematical models of AD systems

The DDPG framework presented in the previous section was applied to control the organic matter content of the effluent of two different AD system configurations in mesophilic conditions to treat Tequila vinasses, i.e., a single-stage model in a CSTR (Fig. 3.3) and a two-stage model in two up-flow fixed bed reactors (Fig. 3.4). The inlet stream of the AD systems corresponds to the substrate, i.e., the amount of tequila vinasses, described by the inlet concentrations in terms of COD and volatile fatty acids (VFA). The selection of uncertain parameters and their range of values for both

systems are based on a preliminary sensitivity analysis reported by Piceno-Díaz et al., (2020). Each of these systems is described next.

3.2.2.1.- Single-stage AD system.

A single-stage AD unit system for the treatment of tequila vinasses was adopted (Zarate, 2013). The acidogenic step considers the reaction: $COD \rightarrow VFA + CO_2 + H_2$, where the objective is to convert the high organic substrate, i.e., tequila vinasses, into VFA with the by-products of CO_2 and H_2. In addition, the methanogenic step consists of the reaction $VFA \rightarrow CO_2 + CH_4$, where the VFA obtained in the acidogenic step is converted into methane and carbon dioxide. The system is perfectly mixed with the acidogenic and methanogenic microorganisms. The Monod kinetic growth model is used for acidogenic bacteria, whereas the Haldane kinetic growth model is for methanogenic bacteria. The manipulated variable is the dilution rate D (1/d) that regulates the inlet flow; the upper and lower bounds considered for this variable are 0.05 and 0.56 (1/d), respectively.

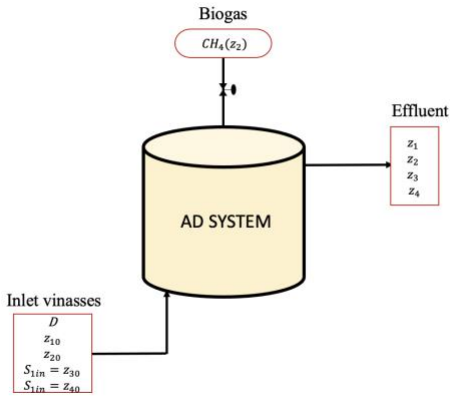


Fig. 3.3.- Schematic representation of the single-stage AD system.

The mathematical model for the single-stage system (Fig.3.3) is given as follows:

$$\frac{dz_1}{dt} = -D\alpha z_1 + \frac{\mu_{1max} S_1}{k_{s1} + S_1} z_1 \tag{3.12}$$

$$\frac{dz_2}{dt} = -D\alpha z_1 + \frac{\mu_{2max} S_2}{k_{s2} + S_2 + \left(\frac{S_2}{k_{I2}}\right)^2} z_2$$

$$\frac{dz_3}{dt} = \mathrm{D}\left(\mathrm{S}_{1,\mathrm{in}} - \mathrm{S}_1\right) - \frac{\mu_{1max} S_1}{k_{s1} + S_1} z_1$$

$$\frac{dz_4}{dt} = D\left(S_{2,in} - S_2\right) - \frac{\mu_{2max} S_2}{k_{s2} + S_2 + \left(\frac{S_2}{k_{I2}}\right)^2} z_2 + \gamma_1 \frac{\mu_{1max} S_1}{k_{s1} + S_1} z_1$$

where z_1 and z_2 are the concentration of the acidogenic biomass (g/L) and methanogenic biomass (g/L), respectively; z_3 and z_4 correspond to the substrate concentrations in terms of COD (g COD/L) and VFA (mmol VFA/L); α represents the biomass fraction that is suspended in liquid phase, μ_{1max} $(1/d)$ and $k_{s1}(gCOD/L)$ are parameters of the Monod kinetics in the acidogenic step whereas μ_{2max} $(1/d)$, $k_{s2}(mmol/L)$and, k_{I2} (mmol/L)$^{1/2}$ are parameters of the Haldane kinetics involved in the methanogenic reaction. $\gamma_1$represents the yield coefficient (mmol VFA/g z_1) whereas $S_{1,in}$ and $S_{2,in}$ are the inlet concentrations of COD (g COD/L) and VFA (mmol VFA/L) that enter the reactor, respectively. The nominal parameters and initial conditions for this system are depicted in Table 3.1.

Table 3.1.- Nominal model parameters and initial conditions of the single-stage model. Note that it is assumed that $z_{3,0}$ and $z_{4,0}$ correspond to the inlet concentrations of the system. (Zarate, 2013)

Parameters	Initial conditions
ks_2=36.468 $mmol/L$	$z_{1,0} = 3.002$ g /L (acidogenic biomass)

k_{i2}=16.773 (mmol/L)$^{1/2}$	$z_{2,0} = 143.496$ mmol /L (methanogenic biomass)
γ_1=2.6584 mmol VFA/g z_1	$z_{3,0} = S_{1in} = 16$ g COD /L
	$z_{4,0} = S_{2in} = 60$ mmol VFA/L

3.2.2.2.-Two-stage AD system

The two-stage AD system considered in this work is presented in Fig. 3.4. The model parameters used in this work were reported in the study of Piceno-Díaz et al. (2020). The acidogenic step takes place in the first reactor and considers the reaction: $S_{11,in} \rightarrow VFA + CO_2$, which transforms the inlet concentration of COD from vinasses ($S_{11,in}$) into VFAs and CO_2. It is assumed that the first-stage reactor only considers acidogenic bacteria. Likewise, the methanogenic stage in the second reactor features first, the reaction $x_{6,in} \rightarrow VFA + CO_2$ that transforms the organic concentration of COD of vinasses entering the second reactor into VFAs and CO_2. Later, the reaction $VFA \rightarrow CH_4 + CO_2$ takes place and converts the VFA produced by the acidogenic bacteria into the desired product (methane) and carbon dioxide. The manipulated variable is the dilution rate D_2 (1/d) that regulates the inlet flow to the methanogenic reactor, and its upper and lower bounds are 0.05 and 0.56 (1/d), respectively. It is assumed that the acidogenic reactor's inlet flow Q_1 is equal to the methanogenic reactor's Q_2 as it is a continuous system and to prevent biomass removal, i.e., washouts.

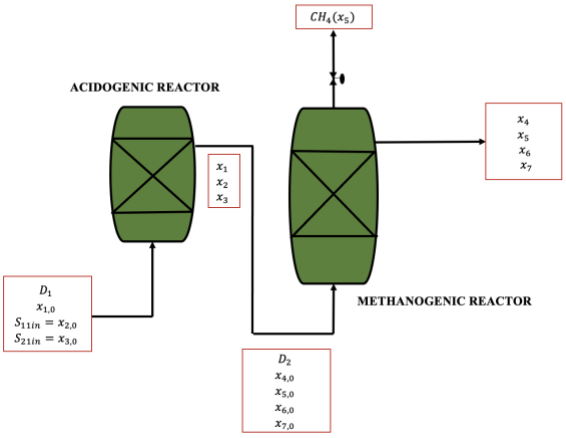


Fig. 3.4.- Schematic representation of the two-stage AD system.

The mathematical model considered for the present case study is as follows:

Acidogenic Reactor:

$$\frac{dx_1}{dt} = -\frac{Q_1}{V_1}\alpha_1 x_1 + \frac{\mu_{11max} x_2}{k_{s11} + x_2} x_1 \tag{3.13}$$

$$\frac{dx_2}{dt} = \frac{Q_1}{V_1}(S_{11,in} - x_2) - \frac{\mu_{11max} x_2}{k_{s11} + x_2} x_1$$

$$\frac{dx_3}{dt} = \frac{Q_1}{V_1}(S_{21,in} - x_3) + \frac{k_2}{k_1}\frac{\mu_{11max} x_2}{k_{s11} + x_2} x_1$$

Methanogenic Reactor:

$$\frac{dx_4}{dt} = -\frac{Q_2}{V_2}\alpha_2 x_4 + \frac{\mu_{12max} x_6}{k_{s12} + x_6} x_4$$

$$\frac{dx_5}{dt} = -\frac{Q_2}{V_2}\alpha_2 x_5 + \frac{\mu_{22max} x_7}{k_{s22} + x_7} x_5 \tag{3.14}$$

$$\frac{dx_6}{dt} = \frac{Q_2}{V_2}(x_2 - x_6) - \frac{\mu_{12max} x_6}{k_{s12} + x_6} x_4$$

$$\frac{dx_7}{dt} = \frac{Q_2}{V_2}(x_3 - x_7) - \frac{\mu_{22max} x_7}{k_{s22} + x_7 + \left(\frac{x_7}{k_{12}}\right)^2} x_5 + \frac{k_5}{k_3} \frac{\mu_{12max} x_6}{k_{s12} + x_6} x_4$$

where x_1 $(gCOD/L)$ and x_4 $(gCOD/L)$ are the scaled acidogenic concentration of the biomass in the acidogenic and methanogenic reactor; x_5 (mmol VFA/L) represents the methanogenic concentration of the biomass in the methanogenic reactor; x_2 (g COD/L) and x_3 (mmol VFA/L) are the substrate concentration of COD and VFA in the acidogenic reactor; x_6 (g COD/L) and x_7 (mmol VFA/L) represent the substrate concentration of COD and VFA in the methanogenic reactor; V_1 (8.7 L) and V_2 (4.5 L) are the reactors capacities; $k_1(gCOD/gX_{11})$, k_2 $(mmol\ VFA/\ g\ X_{11})$, k_3 $(g\ COD/\ g\ X_{12}), and\ k_5$ $(mmol\ VFA/\ g\ X_{12})$ are the yield coefficients of the system; α_1 and α_2 are the biomass fractions that leave the acidogenic and methanogenic reactor; $S_{11,in}$ (g COD/L) and $S_{21,in}(mmolVFA/L)$ are the inlet concentrations of COD (g COD/L) and VFA (mmol VFA/L) that enter to the acidogenic reactor; μ_{11max} (1/d), $k_{s11}(gCOD/L)$, $\mu_{12max}(1/\text{d})\text{and}k_{s12}(gCOD/L)$ are parameters of the Monod kinetics and μ_{22max} (1/d), $k_{s22}(mmolVFA/L)$, and $k_{I2}(mmolVFA/L)^{1/2}$ correspond to the Haldane kinetics. The nominal parameters and initial conditions used in this study are depicted in Table 3.2.

Table 3.2.- Nominal model parameters and initial conditions of the two-stage model. (Piceno-Díaz et al., 2020). Note that $x_{2,0}$ and $x_{3,0}$ correspond to the inlet concentrations of the system.

Parameters	Initial conditions
$k_{s11} = 24$ g COD/L	$x_{1,0} = 94.79$ g COD/L
$k_2/k_1 = 3.5$ mmol VFA/g COD	$x_{2,0} = S_{11,in} = 27$ g COD /L
$k_5/k_3 = 3.5$	$x_{3,0} = S_{21,in} = 50\ 50$ mmol VFA/L
$k_{s22} = 16$ mmol VFA/L	$x_{4,0} = 23.2$ g COD/L

$k_{I2} = 27$ (mmol VFA/L)$^{1/2}$	$x_{5,0} = 100$ mmol VFA/L
$V_2/V_1 = 1.9527$	$x_{6,0} = 10$ g COD/L
	$x_{7,0} = 30$ mmol VFA/L

3.3.-Results

In this section, the DDPG algorithm described in section 3.1.2 is used to search for optimal open-loop control profiles for the single-stage and two-stage AD systems presented in the previous section. Multiple scenarios involving ideal (nominal) conditions, parametric uncertainties as well as deterministic and random disturbances are considered. Each of these scenarios is explained next.

3.3.1.- DDPG structure

Most RL algorithms work through episodes, which is a series of interactions with the environment until a terminal stage is reached. For this study, the terminal stage for the DDPG algorithm was set to 2000 episodes. This number was chosen because the approximated amount of required episodes to converge was not known a priori. Thus, a large number of episodes was chosen to guarantee convergence in the AD systems. Also, we assumed that a plant model is available to simulate the dynamic operation of the AD process and used in this work as a digital twin plant for testing and development of the proposed DDPG algorithm. For every episode, the optimization was performed over 365 time steps; each time step represents one day of operation and is considered as the sampling time. These criteria were selected from prior trial-and-error tests. Note that implementing conventional feedback controllers for the present AD system may not necessarily result in significantly smaller closed-loop setting times (Oscar Méndez-Acosta et al., 2010.; Piceno-Díaz et al., 2020). As mentioned in section 2, the DDPG architecture consists of four deep neural networks

(DNNs). The parameters of DNNs are updated every step time of the optimization process. Each DNN consists of two hidden layers of 128 and 64 neurons. For the critic DNN, linear activation and rectified linear unit activation functions were used for the hidden and output layers, respectively. For the actor DNN, linear activation and sigmoid activation functions were used for the hidden and output layers, respectively. Layer normalization (Ba et al., 2016) was applied to normalize the different scales of the inputs. The parameters of the Ornstein-Uhlenbeck noise were set to $\sigma = 0.015$ and $\varphi = 0.15$. The selected size for the minibatch is T=64; the buffer memory size is P=10000; the architecture of the networks as well as the hyperparameters' values were chosen from prior trial-and-error experiments. Adam optimizer (Kingma & Ba, 2014) is used for the training process of the DNNs. The feasibility of DDPG was evaluated for the single-stage and the two-stage AD systems described above under different scenarios that are likely to occur during operation. This study was implemented in the framework Python Pytorch in Google Colab Pro; the calculations were performed on an Intel Core i7 CPU at 1.7 GHz and 8.00GB memory.

3.3.2.- Single-stage AD system optimization problem

The control objective for the following studies is the minimization of the COD z_3 of the effluent of the reactor. The optimization variable is the daily change in the dilution rate $D(t)$obtained from the DDPG algorithm. The optimization problem is the following:

$$\max_{D(t)} \sum_{j=1}^{M} -\omega_j \sum_{t_0}^{t_f} \left(z_{3_{t,j}} \right) \tag{3.15}$$

s.t.

$$\hat{\boldsymbol{z}}_{t,j} = f^I\left(\boldsymbol{z}_{t,j}, D(t), \boldsymbol{\theta}_j\right) \qquad \forall j \in \{1, \dots, M\}$$

$$\boldsymbol{z}_t = \boldsymbol{z}_0 \qquad \text{for t=0}$$

$0.05 \leq D(t) \leq 0.56$

t=[0, t_f]

where t_f represents the final integration time (365 days) whereas the sampling interval was set to 1 day. The single-stage AD model is represented by a function f^I and represents Eq. (3.12); the uncertain parameters considered in the present formulation are $\boldsymbol{\theta} = \{\mu_{1max}, \mu_{2max}, k_{s1}, \alpha\}$; the weights for each uncertainty realization (ω_j) are assumed to be equal for all the realizations. Note that the objective function in Eq. (3.15) is reformulated in the context of the DDPG framework, i.e., as a maximization problem. Since one of the objectives of AD is to produce biogas and the mathematical model does not account for a direct variable to approximate the biogas production, we assume the concentration of methanogenic microorganisms z_2 is equivalent to biogas production.

3.3.2.1.-Scenario 1: Single-stage AD system (nominal conditions)

To have a benchmark for the single-stage system, no disturbances or uncertainties were considered in this scenario. The nominal parameters and the initial conditions presented in Table 3.1 were adopted for this work. The hyperparameters in the DDPG are key factors that impact the performance of the algorithm. The selected configuration for the system is depicted in Table 3.3

Table 3.3.- Final hyperparameters configuration for single-stage.

Hyperparameters used in DDPG	Value
Minibatch size	64
Actor learning rate	0.0001
Critic learning rate	0.001

Discount factor	0.95
Buffer memory	10000

Fig. 3.5 depicts the learning curve of the algorithm for the single-stage AD system for this scenario, which shows the behavior of the final reward throughout the 2000 episodes. As shown in Fig. 3.5, the learning curve undergoes fluctuations at the beginning due to the random initialization of the model parameters, but once it reaches episode 240, the highest reward is achieved, and after that, the algorithm keeps oscillating near the highest value, resulting in a good performance.

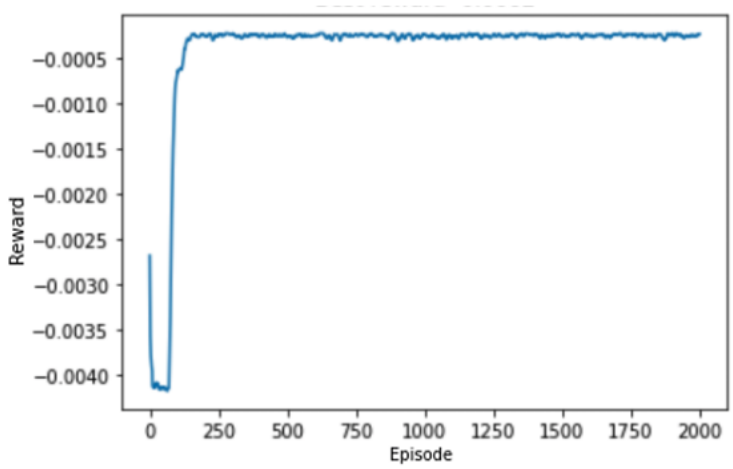


Fig. 3.5.- Learning curve of single-stage AD model of scenario 1 (moving window of 10 episodes).

The tequila vinasses have a high organic matter content; hence, the DDPG specified a rapid decrease in the dilution rate $D(t)$ at the beginning of the operation (Fig. 3.6(a)) thus causing a decrease in the COD of the effluent z_3, as depicted in Fig. 3.6(b). This behaviour prevents the reactor from overfeeding, which would cause the methanogenic bacteria's inhibition since the acidogenic bacteria may process the organic substrate faster than the methanogenic. After that

rapid change, the $D(t)$ kept progressively decreasing until reaching a saturation from the bottom. A similar performance in $D(t)$ was observed in Méndez-Acosta et al., (2016) where they suggested that the dilution rate should remain saturated to maintain appropriate control. This decreasing rate in *D* (t) is required to stabilize the process and avoid possible washouts of the biomass. Fig. 3.6(c) depicts the methanogenic biomass profile. As shown in this figure, the production of methanogenic biomass reacts to the rapid decrease in the dilution rate, resulting in a fast increase of this species at the beginning of the operation. As the manipulated variable $D(t)$ was adjusted, the acidogenic and methanogenic bacteria needed time to achieve a stable operation. The COD concentration of the effluent reached the value of 0.1534 g COD/L at day 365, and the methanogenic biomass reached a concentration of 220.4228 mmol/L. The expected optimal accumulated COD was 109.0748 g COD/L d and the CPU time needed to obtain the resulting optimal open-loop control profile *D(t)* was 5069.8 seconds.

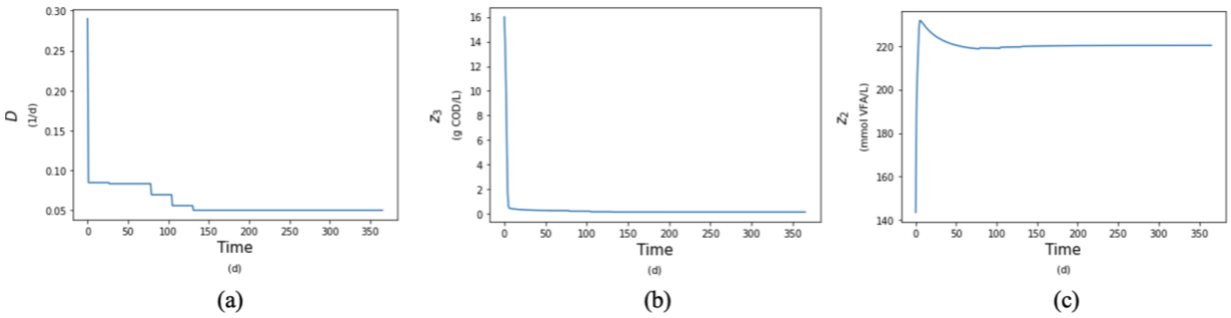


Fig. 3.6.- Scenario 1: Single-stage system over a 365-days optimization under nominal conditions. (a) The dilution rate profile of methanogenic reactor, which represents the manipulated variable of the system. In DDPG framework, it represents the optimal policy chosen by the algorithm. (b) The COD profile (c)Methanogenic biomass profile.

3.3.2.2.-Scenario 2: Single-stage AD system under disturbances and uncertainty

Considering the possible variations of the organic matter content of tequila vinasses reported in the literature, i.e., between 27-100 g COD/L (López-López et al., 2010; Méndez-Acosta et al., 2010; Piceno-Díaz et al., 2020), a disturbance profile was considered for the inlet concentrations ($S_{1,in}$ and $S_{2,in}$) as depicted in Fig. 3.7(a) and Fig. 3.7(b) in terms of COD and VFA, respectively. Step input disturbances were added at day 15, 101, 247, and 320 of -25%, -15%, +20%, and +25% with respect to the original concentrations, respectively. Note that disturbances are assumed to be known a priori. To test the robustness of DDPG, parametric uncertainty was considered for the following model parameters: $\boldsymbol{\theta} = \{\mu_{1max}, \mu_{2max}, k_{s1}, \alpha\}$. Table 3.4 depicts the realizations considered for these uncertain parameters. Note that realization number 5 corresponds to the nominal conditions considered in Scenario 1. The same DDPG hyperparameters from Scenario 1 were used for the present scenario.

Table 3.4.- Uncertainty parameters of scenario 2.

θ	1	2	3	4	5	6	7	8
μ_{1max},	0.63	0.95	1.2	0.48	0.7999	0.55	1.05	0.87
μ_{2max},	0.75	1.03	0.88	0.97	0.7357	0.65	0.81	0.92
k_{s1}	6.2	4.53	5.71	7.3	5.207	3.27	3.85	3.12
α	0.4	0.32	0.47	0.5	0.458	0.55	0.49	0.53

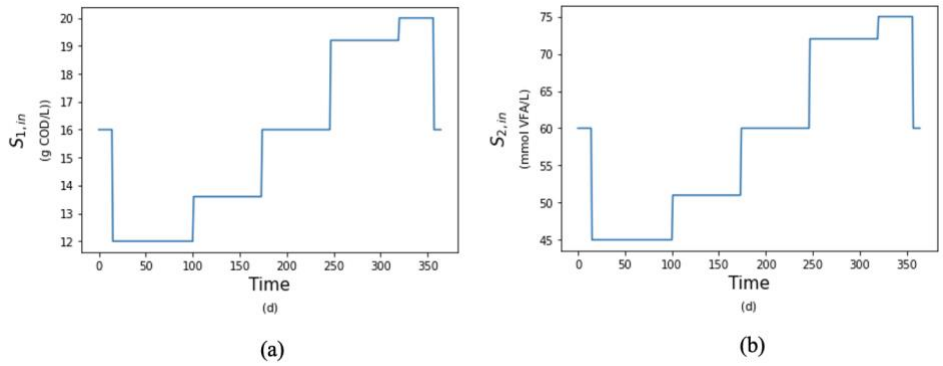
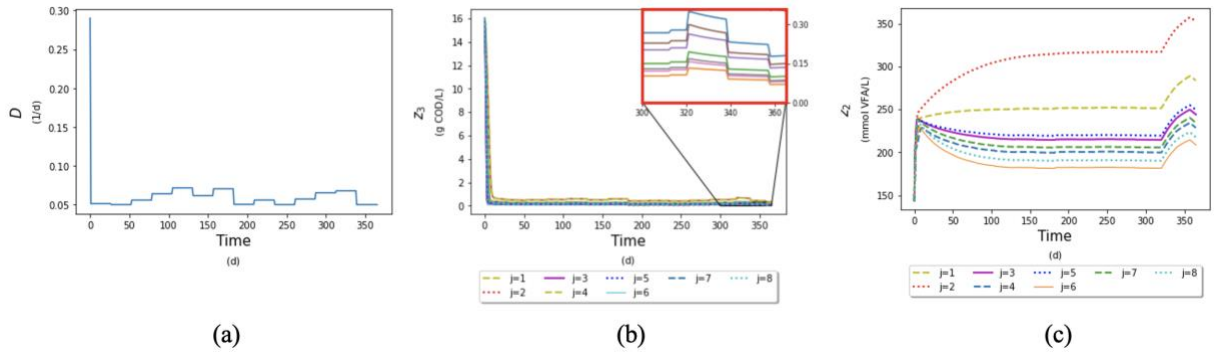


Fig. 3.7.- Scenario 2: Step-wise profile introduced to account for disturbances in the inlet concentrations of the single-stage AD system. (a) Disturbances in terms of COD, (b) Disturbances in terms of VFA.

As shown in Fig. 3.8(a), the dilution rate ($D(t)$) decreased to accommodate the effect of external perturbations and uncertainties affecting the operation of the system. The expected accumulated COD for this scenario was 115.8609 ± 63.9625 g COD/L d, which represents 6% more COD when compared to scenario 1. Note that the set of uncertainty realizations considered represent favorable and adverse conditions for microorganism growth. For example, for realizations j=3 and j=7 (Table 5), the effects of potential inhibition of the methanogenization stage are observed due to high and low realizations of μ_{1max}, and μ_{2max}, respectively; that is, as μ_{1max} becomes higher while μ_{2max} becomes lower than the nominal conditions, the acidogenic bacteria shows a rapid growth that does not allow the methanogenic bacteria to reach adequate conditions to grow; as a result, lower concentrations of methanogenic biomass were obtained for these realizations when compared to the nominal case scenario, as shown in Fig. 3.8(c). One realization that exhibited a strong influence in the robust approach was j=2, i.e., the fraction of biomass that leaves the reactor (α) has a small value; hence, it promotes the production of bacteria as less biomass is leaving the system, thereby increasing the production of biogas. Based on the above, the overall expected value and standard

deviation of COD and methanogenic biomass at day 365 is 0.1441 ± 0.0918g COD/L and 254.5420 ±45.1056 mmol VFA/L, respectively. Note that these are only the final concentrations of COD and VFA at the last time interval in the optimization horizon (i.e., day 365). This represents an improvement of 6% in COD removed and 15% more methanogenic biomass concentration achieved with respect to scenario 1. The CPU time of the optimization was 6192.8 seconds, representing an increase of 22% in CPU time when compared to scenario 1, respectively. As shown in Fig. 3.8(b), the DDPG can provide attractive control actions that can efficiently reject disturbances while considering parametric uncertainty.



(a) (b) (c)

Fig. 3.8.- Scenario 2: Single-stage scenario over a 365-days optimization under disturbances and parametric uncertainty. (a) Dilution rate profile of methanogenic reactor, (b) COD profile, (c)Methanogenic biomass profile.

3.3.3.-Two-stage AD system optimization problem

The control objective for the two-stage system is the same as in the single-stage AD system, i.e., minimization of the fraction of COD that was not processed in the system. The manipulated variable will be the daily change in the dilution rate of the methanogenic reactor ($D_2(t)$).As in the single-stage system, the optimization problem is shown through the DDPG framework as follows:

$$\max_{D_2(t)} \sum_{j=1}^{M} -\omega_j \sum_{t_0}^{t_f} \left(\frac{x_{6,t}}{x_{2,0}} \right)^2 \tag{3.16}$$

s.t.

$$\hat{\boldsymbol{x}}_{t,j} = f^{II}\left(\boldsymbol{x}_{t,j}, D_2(t), \boldsymbol{\theta}_j\right)$$

$$\boldsymbol{x}_t = \boldsymbol{x}_0$$

$$0.05 \leq D_2(t) \leq 0.56$$

t=[0, t_f]

where t_f corresponds to the final integration time (365 days); f^{II} represents the two-stage AD model described with Eq. (3.13) and Eq. (3.14); the uncertain parameters considered for this system are $\boldsymbol{\theta} = \{\mu_{11max}, \mu_{12max}, \mu_{22max}, k_{s12}, \alpha_1, \alpha_2\}$. All the weights ($\omega_j$) are assumed to be equal for all the realizations. As in the single-stage AD system, the concentration of methanogenic microorganisms x_5 is assumed to be equivalent to biogas production.

3.3.3.1.-Scenario 3: Two-stage AD system (nominal)

This scenario is the benchmark for the two-stage AD system. The nominal parameters and the initial conditions as shown in Table 3.2; these were adopted from Piceno-Díaz et al., (2020). The final configuration for the DDPG hyperparameters is depicted in Table 3.5.

Table 3.5- Final hyperparameters configuration for two-stage.

Hyperparameters used in DDPG	Value
Minibatch size	64
Actor learning rate	0.0001
Critic learning rate	0.001

Discount factor	0.99
Buffer memory	10000

Fig. 3.9(a) and Fig. 3.9(b) show that the COD (x_6) decreases as the dilution rate ($D_2(t)$) decreases for the present scenario. There is a rapid decrease of the dilution rate at the beginning, making the COD being reduced at a concentration of 0.2 g COD/L by day 3. Fig. 3.9(c) depicts the profile of the methanogenic concentration x_5; once the dilution rate reaches a steady operation, progressive growth is observed. The COD concentration of the effluent reached the value of 0.1094 g COD/L at day 365, whereas the methanogenic biomass reached a concentration of 365.8356 mmol/L. Compared to the results obtained for scenario 1, there was an improvement of 0.55% more COD removed and an increase of 65.9% in methanogenic concentration at the outlet. These results agree with the study of Luo et al., (2011) since they indicated that the main advantage of the two-stage AD system is a more efficient methanogenesis phase.

To compare the performance of DDPG to a conventional NLP algorithms, the optimization problem shown in Eq. (16) was solved using the interior-point optimization algorithm (IPOPT) (Wächter & Biegler, 2005). This optimization problem was discretized using orthogonal collocation on finite elements. As shown in Table 3.6, this problem was solved using different combinations of finite elements and collocation points. Moreover, Fig. 3.9(d), 3.9(e) and 3.9(f) show the results for the first numerical scheme (i.e., 14 finite elements and 4 collocation points). By day 365, the COD (x_6) was reduced to 0.2635 g COD/L and 358.2811 mmol VFA/L of methanogenic biomass x_5 were produced. This represents 0.57% less COD reduced and 2.84% less methanogenic biomass production. Nevertheless, the expected accumulated COD obtained for this numerical scheme is 22.8% more than that obtained from the DDPG (73.6323 g COD/L d).

On the other hand, numerical scheme 4 resulted in smooth control profiles (not shown for brevity) and returned an expected accumulated COD that is 1.63% lower than that obtained by the DDPG algorithm. In addition, numerical scheme 4 was solved in a CPU time that is an order of magnitude lower than that required by the DDPG algorithm (6015.5 s).

Table 3.6.- Different discretizations of IPOPT, their CPU time and their optimal solution. Note that after increasing the number of finite elements beyond numerical scheme number four, the optimal solution observes multiple oscillations.

Numerical scheme	No. finite elements	No. collocation points	CPU time (s)	Optimal accumulated COD (g COD/L d)
1	14	4	5.75	90.4273
2	30	4	28.5	79.2142
3	100	2	61.578	74.4868
4	100	4	222.985	72.4322

Table 3.6 shows that there is a trade-off between the accuracy of the solution and CPU time as the number of finite elements and collocation points change. As shown in this table, the solution improves as a larger number of finite elements are considered in the solution. Note that a larger number of elements resulted in non-smooth profiles thus indicating that the optimal solution may be numerically unstable. Although the expected accumulated COD obtained from the numerical scheme 4 is slightly smaller than that obtained by the DDPG algorithm, this is an expected result since the DDPG considers plant-model mismatch by its inherent stochastic nature, thus making it more suitable to accommodate plant uncertainty that is not accounted for by the process model.

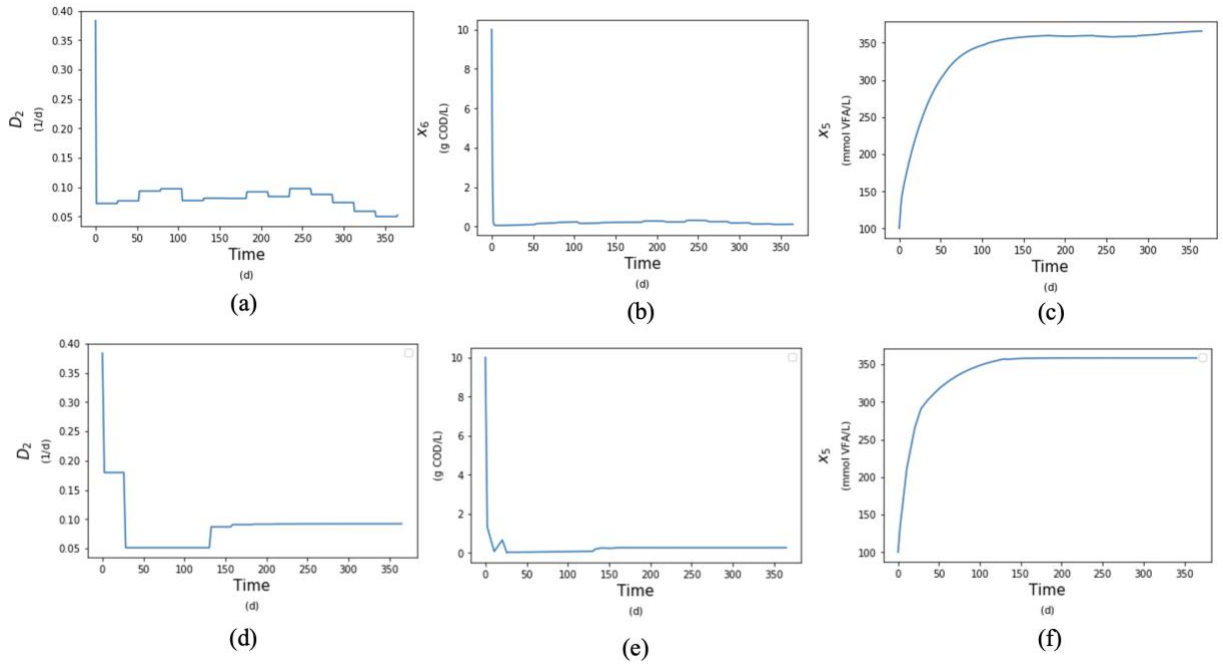


Fig. 3.9.- Scenario 3: Two-stage AD profiles applied to conventional DOP software (IPOPT) and DDPG under nominal conditions. (a) Dilution rate of methanogenic reactor (DDPG), (b) COD profile of effluent (DDPG), (c) Methanogenic profile (DDPG), (d) Dilution rate of methanogenic reactor (IPOPT), (e) COD profile of effluent (IPOPT), (f) Methanogenic profile (IPOPT)

3.3.3.2.-Scenario 4: Single-stage vs two-stage AD system

For comparison purposes, the inlet stream conditions of the single-stage system were used in the two-stage model ($S_{11,in} = 16$ g COD/L) and $S_{21,in} = 60$ mmol VFA/L). Both systems were designed using different experimental data, so a direct comparison between the two systems cannot be made; however, this comparison was performed with the aim to provide insight on the expected performance of these AD configurations using the same feedstream conditions.

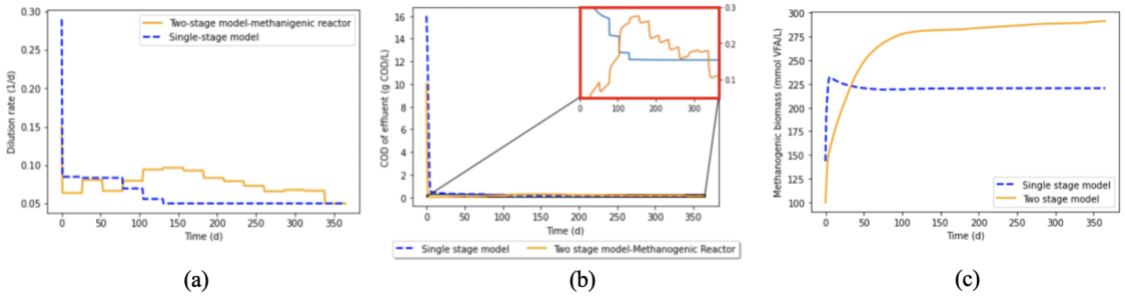


Fig. 3.10.- Scenario 4: Two-stage vs single-stage with the same inlet conditions and under nominal conditions. (a) Dilution rate of methanogenic reactor from two-stage and reactor of single-stage (b) COD profile, (c) Methanogenic profile.

Fig. 3.10(b) shows that the single-stage system reached a COD concentration of 0.1534 g COD/L (z_3) whereas the two-stage system reached 0.1044 g COD/L (x_6) at day 365. As a result, a higher concentration of methanogenic biomass x_5 (291.4750 mmol VFA/L) is obtained at day 365 in the two-stage configuration than in the single-stage system z_2 (220.4228 mmol VFA/L). As the s ingle-stage system deals with both microorganism groups simultaneously and the methanogenic microorganisms require a longer time to grow compared to the acidogenic bacteria, the dilution rate ($D(t)$) of the single system is set to low values to achieve suitable conditions (Fig. 3.10(a)). These changes in $D(t)$ are needed to ensure the adequate production of acidogenic bacteria; if overproduction of acidogenic bacteria occurs, an excess of VFA may be produced and result in a decrease in pH, causing an excess of H_2 and inducing methanogenic inhibition. The accumulated COD over the entire time horizon (365 days) obtained for the two-stage system was 68.1390 g COD/L d whereas the CPU time required to complete the 2000 episodes was 6682 seconds. When compared to scenario 1 (single-stage system), these represent an improvement of 37.5% less accumulated COD in the effluent and an increase of 31% in CPU costs. This difference in COD

performance is due to the changes observed for the two-stage AD system, as shown in the inner plot in Fig. 3.10(b).

3.3.3.3.-Scenario 5: Two-stage AD system with disturbances and uncertainty

From an operational perspective, a two-stage process becomes more challenging to control because it considers two digesters that operate in tandem. At the same time, both reactors are prone to disturbances and kinetic uncertainty. To test the performance of this system under such conditions, a step-wise profile was introduced as depicted in Fig. 3.11(a) and Fig. 3.11(b) combined with parametric uncertainty, i.e., $\boldsymbol{\theta} = \{\mu_{11max}, \mu_{12max}, \mu_{22max}, k_{s12}, \alpha_1, \alpha_2\}$. The step disturbances considered for this scenario are assumed to be known a priori and were added at day 15, 101, 247, and 320 of -25%, -15%, +20%, and +25% with respect to the original inlet COD and VFA concentrations, respectively. The uncertain realizations used for this scenario are shown in Table 3.7; note that realization j=5 corresponds to the nominal condition considered for scenario 3.

Table 3.7.- Uncertainty parameters scenario 5.

θ	1	2	3	4	5	6	7	8
μ_{11max},	0.15	0.351	0.183	0.284	0.27	0.34	0.234	0.41
μ_{12max},	0.58	0.635	0.4	0.439	0.5	0.6	0.575	0.7
μ_{22max},	0.44	0.248	0.26	0.228	0.29	0.28	0.395	0.37
k_{s12}	5	3.607	3	3.372	3.5	4.25	2.671	2.75
α_1	0.085	0.17	0.11	0.159	0.13	0.18	0.15	0.20
α_2	0.47	0.403	0.29	0.392	0.38	0.42	0.43	0.32

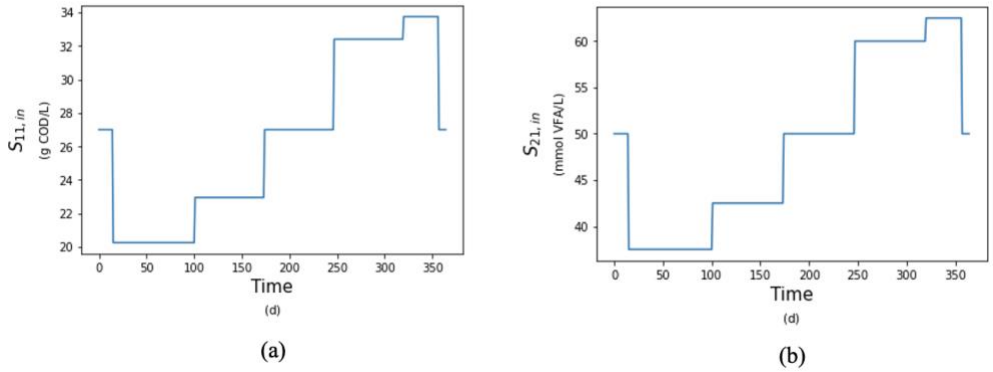


Fig. 3.11.- Scenario 5: Step-wise profile introduced to account for disturbances in the inlet concentrations of the two-stage AD system. (a) Disturbances in terms of COD, (b) Disturbances in terms of VFA.

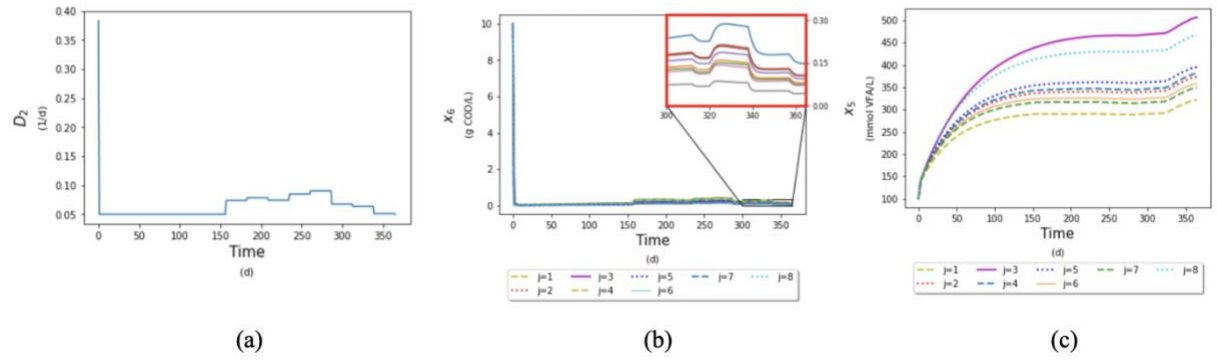


Fig. 3.12.- Scenario 5: Two-stage scenario over a 365-days optimization under disturbances and parametric uncertainty. (a) Dilution rate methanogenic reactor, (b) COD profile (c) Methanogenic biomass profile.

As shown in Fig. 3.12(a), the dilution rate ($D_2(t)$) is decreased to accommodate the disturbances and uncertainties considered in this scenario. Fig. 3.12(a) also shows how the disturbances of the inlet concentrations influence the dilution rate. Every time the inlet concentrations increase, the dilution rate ($D_2(t)$) performs a rapid change until it reaches a steady-state operation. Regarding

biogas production, the different realizations account for favorable (highest, j=3) and adverse conditions (lowest, j=1) for the AD system, as shown in Fig. 3.12(c). The expected value and standard deviation of COD and methanogenic biomass at day 365 is 0.0889± 0.0309 g COD/L and 393.5657±62.3238 mmol VFA/L, respectively; this represents an improvement of 18.7% more COD reduced and a 7.5% increase in the final methanogenic biomass concentration when compared to scenario 3, which shows a significant sensitivity of the system to disturbances and parametric uncertainty. The expected accumulated COD and standard deviation was 55.0791 ± 17.28 g COD/L d (Fig, 3.12(b)), which represent an improvement of 25.1% more COD reduced when compared to scenario 3. The CPU time reported for this scenario was 7580.6 seconds, which is 26% higher that the CPU time reported for scenario 3.

3.3.3.4.-Scenario 6: Two-stage with random disturbances

To provide further on the performance of the DDPG algorithm, random samples from a Gaussian distribution were added to the inlet concentrations in the two-stage AD system, i.e.,

$$S_{11,in_G} = S_{11,in} + \varepsilon_{1,t} \tag{3.17}$$

$$S_{21,in_G} = S_{21,in} + \varepsilon_{2,t}$$

where $\varepsilon_{t,1}$ and $\varepsilon_{t,2}$ are random Gaussian noises ($N[\varphi, \sigma]$) with a zero-mean ($\varphi = 0$) a standard deviation σ of 5% with respect to the nominal inlet concentrations. The frequency of the random samples was set to 1 day. S_{11,in_G} and S_{21,in_G} represent the random inlet concentrations of tequila vinasses entering the system whereas $S_{11,in}$ and $S_{21,in}$ are the nominal inlet concentrations considered in scenario 3. As this scenario considers a random disturbance for every step time, it becomes more computationally challenging. To reduce the computational costs, the following stopping criteria was selected for the present scenario: once every 100 episodes we evaluated the

expected reward and compared it to the previous 100 episodes. If the improvement in the expected rewards was below a user-defined tolerance criterion (i.e., 0.0001%), then the algorithm was terminated.

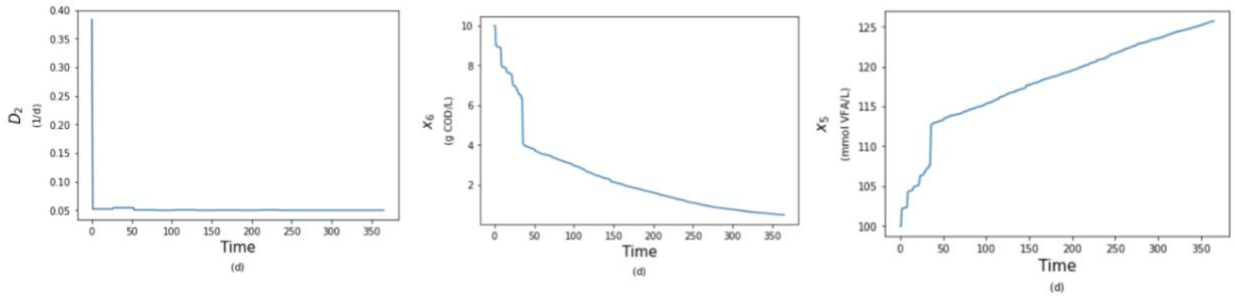


Fig. 3.13.- Scenario 6: Stochastic performance for random disturbances in a two-stage AD model. (a) Dilution rate profile of methanogenic reactor, (b) COD profile of effluent, (c) Methanogenic biomass profile.

As depicted in Fig. 3.13(b), the COD concentration x_6 decreases almost monotonically, reaching a concentration of 0.4934 g COD/L and a methanogens concentration x_5 of 125.715 mmol VFA/L at day 365 (Fig. 3.13(c)). This represents 3.5 times more COD and 65.6% less methanogenic concentration than that reported for scenario 3 (nominal conditions). The expected accumulated COD over the entire time horizon (365 days) for this scenario was 874.0724 g COD/L d (Fig. 3.13 (b)) representing an increase of one order of magnitude in COD to that obtained from scenario 3. On the other hand, the algorithm showed significant learning after 100 episodes (Fig. 3.14), demonstrating the potential of the DDPG algorithm to control processes under random external perturbations.

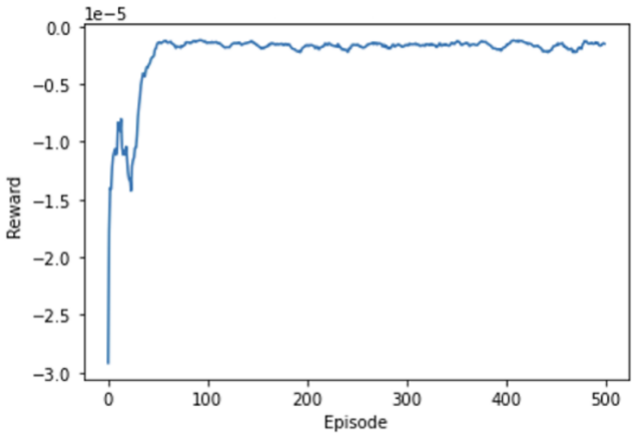


Fig. 3.14.- Scenario 6: Learning curve of the two-stage model with stochastic disturbances (moving window of 10 episodes).

The CPU time obtained for this scenario was 225371.3 s. Although scenario 6 might reflect a more realistic case for the inlet concentrations of tequila vinasses due to the random disturbances, the CPU time is at least two orders of magnitude higher than that reported for scenario 3, which is an indication that further improvements are needed to implement this method for online large-scale applications.

3.3.3.5.-Scenario 7: EMPC for Single-stage AD system

To further illustrate the benefits of the proposed DDPG algorithm, a robust EMPC strategy was implemented for the single-stage AD system. Typically, an EMPC uses a dynamic process model to obtain optimal control actions that minimize an economic function in the presence of constraints. Note that a feedback control strategy could compensate for parametric uncertainty; however, there is no guarantee that their control actions would result in optimal control actions unless uncertainty is explicitly considered in the design of the controller. To account for this condition, the DDPG algorithm presented in the previous section was embedded within a feedback strategy to search for

optimal control actions of the single-stage AD system under disturbances and parametric uncertainty. That is, the DDPG algorithm represents the EMPC strategy in the feedback strategy and is solved at each sampling interval. Parametric uncertainty was considered within the DDPG algorithm using the same multi-scenario approach considered in scenario 2 for this AD system. To simplify the analysis, the plant model depicted in problem (3.15) was assumed to be same used by the EMPC (DDPG) strategy; that is, problem (3.15) represents the environment in the DDPG algorithm that must be solved using the plant states assumed to be available at each sampling interval (10 days). Note that the environment in the DDPG (EMPC) framework accounts for all the realizations considered for the single-stage AD system depicted in Table 4 whereas the plant is assumed to operate at one of these uncertainty realizations (uncertainty set 5 in Table 4). In addition to parametric uncertainty, a series of measured disturbances in the inlet concentrations were considered (i.e., $S_{1,in}$ and $S_{2,in}$). As shown in Fig. 3.15, a series of step disturbances of -10%, -20%, +10% and +20% with respect to the inlet concentrations' nominal values enter the system at each sampling interval. Note that in the EMPC (DDPG) framework, the disturbance measurements for $S_{1,in}$ and $S_{2,in}$ at each sampling interval are considered constant throughout the time horizon in the DDPG (EMPC) formulation. For the present scenario, both the prediction and control horizons in the DDPG framework were set to 100 days. For the training process, 500 episodes were considered for each run of the DDPG (EMPC) framework.

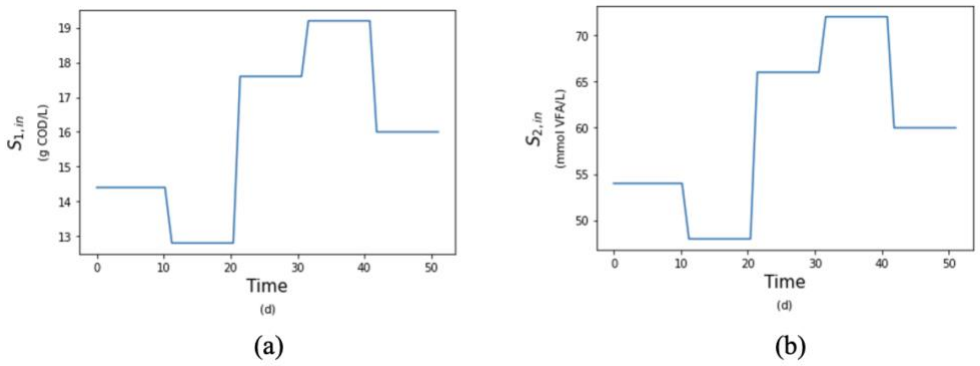


Fig. 3.15.- Scenario 7: Step-wise profile of disturbances in the inlet concentrations of the single-stage AD plant simulation. (a) Disturbances in terms of COD, (b) Disturbances in terms of VFA.

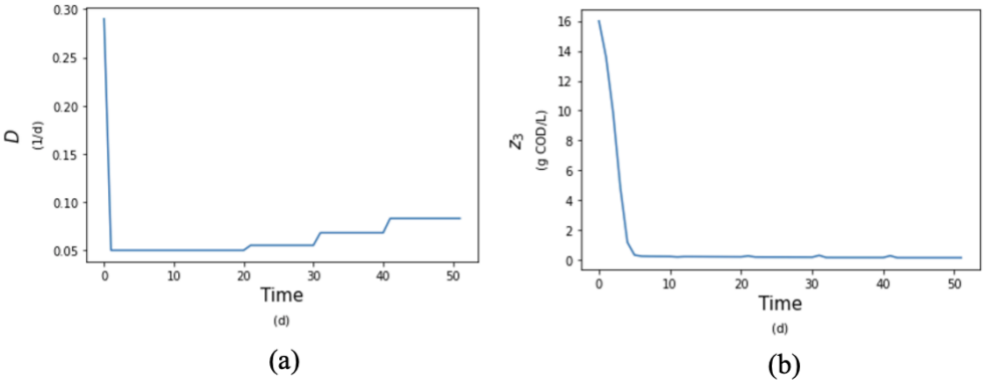


Fig. 3.16.- Scenario 7: Closed-loop simulation of the single-stage system over 50 days. (a) dilution rate profile of methanogenic reactor, (b) COD profile in the methanogenic reactor.

The expected optimal accumulated COD (z_3) was 46.7438 g COD/L d, whereas the averaged CPU time needed to obtain the optimal control actions in D(t) at each sampling interval was 1681.7 seconds. Fig. 3.16 (a) depicts a rapid decrease of the manipulated variable D during the first 20 days of operation. As a result, the plant responds to this rapid change in D to achieve a low organic concentration z_3, as shown in Fig. 16 (b). The final concentration of z_3 at day 50 was 0.1511 g

COD/L. Note that a somewhat similar performance was observed in scenario 2, (Fig. 3.8 (a)) regarding the rapid decrease of D during the first days of operation. Nevertheless, the control actions returned by the EMPC (DDPG) strategy can efficiently accommodate the changes in the disturbances such that a minimum COD is maintained in the process, as shown in Fig. 16. This scenario demonstrates the potential of the proposed DDPG strategy to operate slow systems involving large time-constants in closed-loop.

3.4.-Summary

The aim of this chapter was to present a DDPG-based methodology for process control under disturbances and parametric uncertainty. The selected case studies were two AD systems treating tequila vinasses, with the purpose of reducing COD while producing biogas. Two different configurations (single-stage and two-stage) were considered and compared using different scenarios. The results showed that the proposed methodology exhibited advantages and limitations with respect to conventional NLP solvers. Also, when comparing the two different reactor configurations under similar inlet conditions, the twos-stage system exhibited a more efficient performance on biogas production and COD reduction. One scenario considered stochastic disturbances, resulting in a more realistic representation of the process; however, the required CPU indicated that further improvements are needed before this methodology can be considered for large-scale online applications. The last scenario involved the application of a robust EMPC, showing the ability of the algorithm to maintain adequate COD concentrations while dealing with disturbances and parametric uncertainty.

CHAPTER 4

Integrated process design and control

This chapter presents the methodology of the DDPG applied to integrated process design and control for AD systems treating Tequila vinasses. This framework combines time-dependent and time independent variables, stochastic disturbances, and constraints, which are embedded in the formulation as penalty functions. As in the previous chapter, different scenarios involving the two different reactor's schemes are analyzed and compared in this study. This chapter begins with the description of the proposed methodology in section 4.1, where different key elements of the proposed framework are explained, such as the customed activation function, the objective function, and limitations. In section 4.2, three different scenarios are described, outlining the advantages and limitations of the current methodology applied to the two reactor configurations. A summary of this chapter is provided at the end.

4.1.-Methodology

In this section, the problem statement for a conceptual integration of design and control formulation is presented. In section 4.1.2, the analogy between integration of process and control design and the DDPG algorithm is explained. A brief explanation of the DDPG algorithm is also described. The cost function and the limitations of the proposed method are described at the end of this section.

4.1.1.-Problem statement: integrated process design and control

The purpose of simultaneous process and control design is to search for an optimal process design that can accommodate operating scenarios that are likely to occur during the lifetime of a plant while considering the process dynamics. Simultaneous approaches usually seek to solve complex mixed-integer dynamic optimization (MIDO) problems which are later discretized to become mixed-integer nonlinear programming (MINLP) problems. Often, those problems may become intractable for large-scale systems as they have increased complexity due to the addition of integer optimization variables. In the present work, integer decisions are not considered to simplify the analysis and alleviate the already taxing computational costs associated with these methodologies(Rafiei & Ricardez-Sandoval, 2020b). NLP approaches have been mostly formulated assuming perfect knowledge of the model parameters, user-defined profiles in the external perturbations, or discrete realizations in the uncertain parameters (i.e., deterministic approaches). Nevertheless, the implementation of design and control strategies under random realizations in the disturbances and parameter uncertainty are limited (Bahakim & Ricardez-Sandoval, 2014; Koller et al., 2018; Rafiei & Ricardez-Sandoval, 2018). Hence, there is a need to develop efficient methods that can deal with disturbances and parametric uncertainty that follow a stochastic representation to produce a more realistic representation of the systems during operation. Frameworks based on stochastic optimization are often prone to be computationally expensive and have converging problems as the dimensions of the problem increase (Chachuat et al., 2006). Similarly, stochastic global optimization methods tend to only find local solutions and are usually incapable of dealing with highly constrained problems (Sharifzadeh, 2013).

The promising potential of developing a DDPG-based simultaneous design and control methodology as a strategy for optimal open-loop control relies on the DDPG's ability to consider stochastic random variables that may account for plant-model mismatch (Lee & Lee, 2004) and its potential to deal with high dimensional systems, which are key challenges for NLP problems. A generic simultaneous design and control problem addressed via stochastic optimization can be formulated as follows:

$$\begin{aligned}
&\min_{\boldsymbol{u}(t),\boldsymbol{des}} \sum_{j=1}^{J} w_j \, OF(\dot{\boldsymbol{x}}(t), \boldsymbol{x}(t), \boldsymbol{y}(t), \boldsymbol{u}(t), \boldsymbol{d}(t), \boldsymbol{\zeta}, \boldsymbol{des}, t) \\
&f(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{d}(t), \boldsymbol{\zeta}, \boldsymbol{des}, t) = \dot{\boldsymbol{x}}(t) \\
&f_0(\boldsymbol{x}(t_0), \boldsymbol{u}(t_0), \boldsymbol{d}(t_0), \boldsymbol{\zeta}, \boldsymbol{des}, t_0) = \dot{\boldsymbol{x}}(t_0) \\
&h(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{d}(t), \boldsymbol{\zeta}, \boldsymbol{des}, t) = \boldsymbol{y}(t) \\
&g(\boldsymbol{x}(\boldsymbol{t}), \boldsymbol{u}(t), \boldsymbol{d}(t), \boldsymbol{\zeta}, \boldsymbol{des}, t) \leq 0 \\
&\boldsymbol{u^l} \leq \boldsymbol{u}(\boldsymbol{t}), \leq \boldsymbol{u^h}
\end{aligned} \tag{4.1}$$

where *OF* represents an economic function, $\boldsymbol{x}(t)$ and $\dot{\boldsymbol{x}}(t) \in \mathbb{R}^{n_x}$ are the system's states and their corresponding derivatives; $\boldsymbol{u} \in \mathbb{R}^{n_u}$ is the control profile vector, $\boldsymbol{u}^l$ and $\boldsymbol{u}^h$ denote the lower and upper bounds for the control vector whereas $\boldsymbol{y} \in \mathbb{R}^{n_y}$ represents algebraic variables. The formulation presented in Eq. (4.1) considers *j*th uncertain realizations given by the vector $\boldsymbol{\zeta}$. Hence, the present work assumes a multi-scenario approach where uncertainty realizations can be approximated using a finite set of discrete realizations defined *a priori*. A fundamental assumption is that uncertainty remains static during the analysis, i.e., time-invariant uncertainty. The design variables are denoted by $\boldsymbol{des} \in \mathbb{R}^{n_{des}}$, which are time-invariant variables; $f: \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_d} \times \mathbb{R}^{n_\zeta} \times \mathbb{R}^{n_{des}} \longrightarrow \mathbb{R}^{n_x}$ represents the set of nonlinear differential equations subject to their initial conditions f_0; $h: \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_d} \times \mathbb{R}^{n_\zeta} \times \mathbb{R}^{n_{des}} \longrightarrow \mathbb{R}^{n_x} \longrightarrow \mathbb{R}^{n_y}$ symbolizes the set of algebraic equations and $g: \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_d} \times \mathbb{R}^{n_\zeta} \times \mathbb{R}^{n_{des}} \longrightarrow \mathbb{R}^{n_Y}$ denotes the set of inequality constraints; ω_j are the weights assigned for each probability of occurrence of each uncertainty realization j considered in the formulation; $\boldsymbol{d} \in \mathbb{R}^{n_d}$ represents the vector accounting for

disturbances (Eq. (4.2)). The key novelties of this study are the application of DDPG for simultaneous process and control design under consideration of stochastic (random) time-varying noise in disturbances and discrete realizations in the uncertain parameters. In this work, we investigate the effect of noise $\boldsymbol{\vartheta}$ considered in the disturbances of the system $\boldsymbol{d}$, i.e.,

$$\boldsymbol{d}(t) = \boldsymbol{\pi}(t) + \boldsymbol{\vartheta}(t) \tag{4.2}$$

$$\vartheta(t) = \{\vartheta | \boldsymbol{\vartheta} \sim PDF(\varsigma)\} \tag{4.3}$$

where $\boldsymbol{\pi}$ denotes the seasonal changes with respect to the nominal values in the disturbances, these values π are assumed to be known *a priori*; $\boldsymbol{\vartheta}$ represents the noise added to seasonal changing-disturbances, such as sinusoidal functions with uncertain critical parameters (e.g., the study from Malcolm et al., (2007) or step changes with unknown magnitudes within a certain range as the study from Sakizlis et al., (2004). As shown in Eq. (4.3), the random noise follows a probability density function (PDF). The PDF parameters ς, such as mean and standard deviation for a Normal distribution, are user-defined parameters that can be obtained from heuristics or historical process data. In the present work, a normal PDF was selected; this type of PDF is widely used for engineering applications (DeCoursey, 2004). Note that other distributions can be considered, such as symmetric and non-symmetric probability distributions. In traditional approaches (e.g., sequential), overdesign factors are added to the process design to address disturbances and parametric uncertainty; these features may result in expensive process designs. Thus, the stochastic noise considered in the simultaneous process and control design makes it an attractive option for complex systems where stochastic perturbations plays a major role. However, adding these stochastic features to the system represents a challenge as they make the problem more difficult to converge due to the problem's stochastic NLP nature. A brief description of the RL-based DDPG algorithm is provided next.

4.1.2.-Simultaneous Design and Control using DDPG

DDPG belongs to a type of RL algorithms where an agent (a decision-maker) interacts with an environment (i.e., a process) iteratively by observing the states of the process, executing actions, and assign a positive or negative reward based on these actions. As most RL algorithms, DDPG works through episodes, which are a series of interactions with the environment until a terminal stage is reached. The terminal stage might be a user-defined number of episodes or a specific-termination criteria, such as specified operating conditions, concentrations, or a specific tolerance (e.g., the study of Bangi & Kwon, (2021)). For every episode, there is also a user-defined number of discrete time-steps $t_f \ \forall i \in \{1, \dots, t_f\}$. The architecture of this algorithm comprises two Deep-Q neural networks (DNN) and two target networks (one for the actor and another for the critic), an environment (*E*), and a buffer memory.

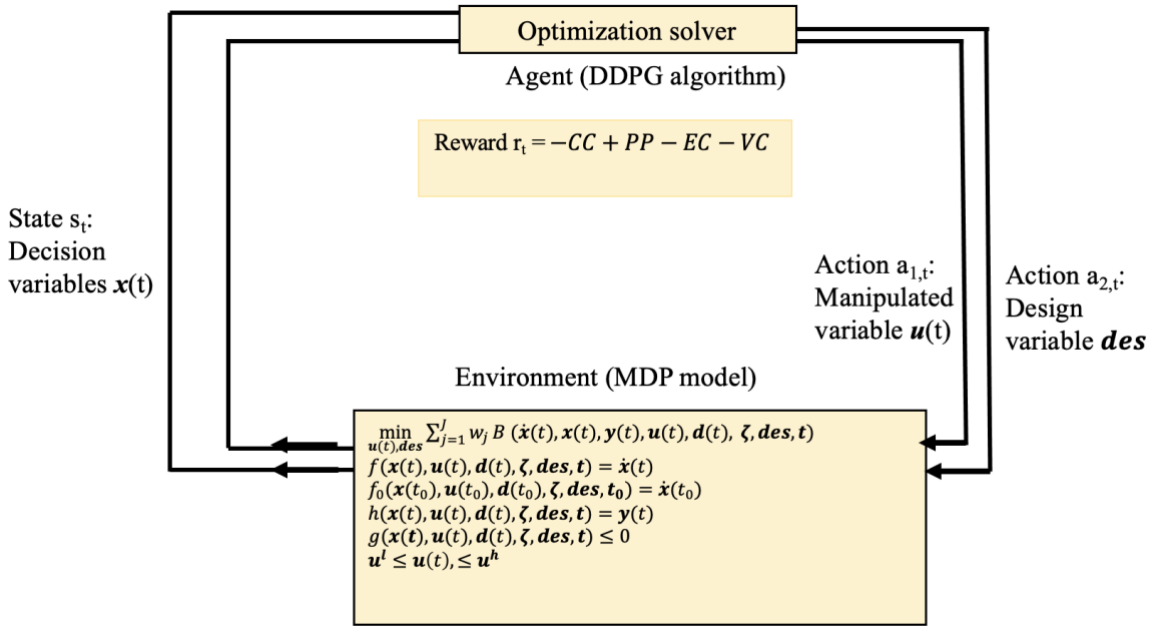


Fig. 4.1.- Schematic Design and control through a RL framework.

Fig. 4.1 presents the adaptation of the simultaneous design and control approach in the context of the RL algorithm proposed in this work. From a process design and control perspective, the agent (i.e., the decision-maker) would correspond to the numerical approach used to solve the problem statement. For the present study, the agent corresponds to the DDPG algorithm. The main feature in DDPG is the actor-critic architecture, where an actor network outputs a policy μ (which corresponds to the control decisions of the optimization problem) and the critic DNN evaluates the goodness of the policy through a Q-value (which helps in the convergence of the optimal solution). To avoid divergence of the algorithm, target DNNs are embedded in the DDPG framework. These are equivalent to delayed copies of the actor and critic networks that aim to stabilize the learning process by enforcing slow changes of the predicted values for the actor and critic DNNs. The performance of DDPG is highly sensitive to the hyperparameters, which are parameters that define the learning process in the algorithm. Thus, an adequate tuning of these DDPG parameters is relevant to achieve an acceptable performance. The learning process (also known as training process) of the actor and critic network consists in the improvement of the adjustment of the weights (ϕ^{Q}, ϕ^{μ}, $\phi^{Q'}$, and $\phi^{\mu'}$), which are the key parameters of the DNNs and are updated at every time step (Bouwmans et al., 2019). This iterative process is achieved through sequential interactions over a discrete time with the environment *E*, modelled as a Markov Decision Process (MDP).

In the case of simultaneous design and control, the environment corresponds to the mechanistic process model representing the system's transient behavior, which is often described by a set of nonlinear differential equations, algebraic equations, and constraints, i.e., *f, h* in Eq. (4.1). As with most RL algorithms, DDPG cannot explicitly handle process constraints; thus, penalty functions are usually added to deal with process constraints, i.e., h and g (Tessler et al., 2018; Yoo et al.,

2021). In the present work, the variability of the system is considered in the objective function as penalty functions that can account for process constraints. This will be further discussed on section 4.1.4. Note that stochastic time-dependent disturbances are explicitly considered in the present DDPG implementation. The purpose of the agent is to maximize a user-defined reward (objective) function r_t, which would correspond to the performance objective of the simultaneous design and control approach, i.e. *OF* in Eq (4.1).

At every time-step *t*, the agent observes the states $s_t \in$ S from the environment, which describes the current state of the design and control scheme. These states s_t would correspond to the system states $\boldsymbol{x}(t)$ presented in the design and control problem (Eq. (4.1)). Based on the observations, the agent (i.e., DDPG) executes an action $a_t \in$ A. The action is an n-dimensional action space to account for the different control decisions of the NLP. In the simultaneous design and control method, the manipulated variables available for control (i.e., u_t) and the design variables ***des*** would be interpreted as the actions taken by the agent to maximize a user-defined reward. To enable exploration in the DDPG algorithm, a user-defined noise ($\mathcal{N}_t$) is added to the action before it is sent back to the environment, i.e.,

$$a_t = \mu(\mathrm{s_t}|\phi^{\mu} + \mathcal{N}_t) \tag{4.4}$$

This noise can be interpreted as plant-model mismatch. For the current study, a Gaussian-type process referred to as Ornstein-Uhlenbeck (OU) noise (Uhlenbeck & Ornstein, 1930) is considered. Once the action is executed by the environment, a new state s_{t+1} is produced, and the transition generated in the time-step (s_t, a_t, r_t, s_{t+1}) is stored in the buffer memory to collect a finite number *P* of previous transitions; once this number of *P* transitions is reached, the old transitions are discarded. At every time step, the actor and critic DNNs are updated by randomly sampling minibatches of a user-defined finite number *T* of transitions from the buffer; this is

performed with the purpose of reducing correlations between samples. Simultaneously, the buffer is filled with *P* transitions. More details about the DDPG algorithm can be found in Lillicrap et al., (2015).

4.1.3.-Custom activation function

As shown in Eq. (4.1), the simultaneous design and control problem considers the interaction of time-dependent (manipulated variables $\boldsymbol{u}$) and time-independent (design variables ***des***) variables (e.g., equipment sizes). To accomplish this goal, we enforce that the actions referred to as the design variables remain fixed throughout the entire episode by implementing a custom activation function. In the present work, this function uses a sigmoid activation function in the output layer of the actor DNN to update the n-dimensional action vector at different frequencies of time-steps. The sigmoid function was chosen in this work as it can be interpreted as a probability and thus, they are commonly used in NN for regression. These types of functions can measure the relationship between independent and dependent variables, i.e., they can correlate complex nonlinear data (Gupta et al., 2020). To perform backpropagation (i.e., updating the weights of DNNs), the derivative of the activation function is needed, and the fact that sigmoid function is nonlinear, differentiable, and continuous everywhere, makes it a good candidate for this work (Ngah et al., 2016). For the manipulated variables, the updating process is at every timestep of the episode whilst the design variables are only updated at the first time-step of the episode and are kept constant for the remaining time-steps. This can be formulated as follows:

Custom activation function

for episode= 1, *MA* **do**
  for step t = 1, *TS* **do**
    if t=0:

$$a_u(t) = \frac{1}{1+e^{-\chi}}$$

$$a_{des}(t) = \frac{1}{1+e^{-\chi}}$$

    if t≥1:

$$a_u(t) = \frac{1}{1+e^{-\chi}}$$

$$a_{des}(t=0) = \frac{1}{1+e^{-\chi}}$$

  end for
end for

where *MA* is the final number of episodes, *TS* corresponds to the final number of timesteps in the each episode, χ corresponds to the inputs of the neuron, a_u is the activation function of the manipulated variables and a_{des} corresponds to the activation function of the design variables. Using this approach, the proposed method enforces that the output and input of the actor DNN would have the same value for the actions representing the design variables after the second time-step; hence, the design (time-independent) variables are restricted to a single value per episode, being adjusted through the realizations/executions of the episodes.

4.1.4.-Objective function

In the present study, the objective is the maximization of an economic function that combines the capital cost, the production profit and variability cost that is expected to account for the system's transients. This function is represented as the reward function in the RL algorithm and is defined as follows:

$$OF = -\mathrm{CC}(\boldsymbol{des}, \boldsymbol{\zeta}) + \mathrm{PP}\left(\boldsymbol{des}, \boldsymbol{\zeta}, \boldsymbol{u}(t), \boldsymbol{x}(t)\right) - EC\left(\boldsymbol{des}, \boldsymbol{\zeta}, \boldsymbol{u}(t), \boldsymbol{x}(t)\right) - \mathrm{VC}\left(\boldsymbol{des}, \boldsymbol{\zeta}, \boldsymbol{u}(t), \boldsymbol{x}(t)\right) \quad (4.5)$$

where CC are the capital cost, PP is the production profit, EC are the energy consumption costs and VC refers to the variability costs. The capital costs are related to the process design as they often represent equipment sizing. The production profit describes the economic gains of the product produced. Capital costs (CC) are typically calculated using steady-state information. The energy consumption costs describe the economic expenses needed to operate the process. The process variability costs are often defined as a function of those time-dependent variables that must be close to their corresponding target (desired) design values. As RL is known for its inability to satisfy process constraints, a way to include constraints is by adding a penalty function to the actual reward function (Pan et al., 2021). Hence, to accommodate for process constraints, the process variability will be implemented in this methodology as a penalty into the reward function; this function has the purpose of guiding the agent towards solutions that satisfy the constraints. In this work, the penalty function considers the difference between a set point and the predicted value obtained by the RL algorithm. Then, this function is weighted by a penalty factor, that all together will represent the deduced cost when violating the constraints, i.e., when the set point is not met, and consequently, will impact the plant profits. Note that in some cases, this function might also increase the profits, i.e., whenever the values are below the desired target (set-point) values). Some examples of VC are the final product specifications (e.g., product quality) which will be directly proportional to the deviations from a nominal or targeted design value. Note that process variability costs are also an implicit method to assign an economic cost to the process control performance (Ricardez-Sandoval et al., 2008, 2010).

4.1.5.-Limitations

Overall, the general economic function shown in Eq. (4.5) can accommodate the transient operation of the system while searching for an optimal process design. Nonetheless, one limitation

of this approach is that it relies on defining a coherent economic function that considers the objectives for the specific system, e.g., depending on the elements of the economic function (e.g., penalty costs, selling prices of products, etc.), there are possible scenarios where the agent might prefer to violate the constraints (Hua et al., (2019)) in exchange of producing more economical gains, as there will be a trade-off between performance (profits) and penalty costs. Therefore, this represents a limitation for safety-related constraints, i.e., problems where violating constraints represent a high-risk for the operation of the process. This will be further illustrated in the next section.

As mentioned above, an important feature for a good performance of the algorithm is the training of DNNs. This consists of two steps: forward and backpropagation. The first consists in the transformation of the input data through layers and activation functions to calculate a predicted output. The second step (backpropagation), aims to adjust the parameters (weights) of the DNNs. This is achieved using gradient descent to minimize the difference between the output value of the DNNs and its predicted values, causing the improvement of the accuracy of DNNs. Before adjusting these weights, it is necessary to compute the gradients with respect to the DNNs parameters to assess how much each weight needs to be adjusted to optimize the reward. To calculate these gradients, it is necessary the differentiation of the activation functions computed by the neurons, which will lead to several calculations as the derivates are calculated at each connection of the neurons. Therefore, calculating and storing the gradients for the backpropagation requires a longer computational time than the forward training step; this represents the main technical challenge in the training process (Chen et al., 2018).

Another challenge for the DDPG is the trade-off between the ability to deal with high dimensional problems and the computational costs. The higher the dimension of the problem, the higher the number of neurons needed in each layer of the DNN to perform as good approximators, and a higher number of episodes of training to converge. Therefore, larger the computational costs would be required. Despite these limitations, the DDPG algorithm is still able to provide economically attractive integrated design and control schemes in acceptable turnaround times.

4.2.-Results

In this section, the integrated process design and control approach described in section 4.1 is used to search for optimal process design and open-loop control profiles by maximizing an economic function for the single-stage and two-stage AD systems presented in section 3.3. For this case, the mathematical process models presented in chapter 3 were slightly modified. To address the integrated process design and control scheme, the volumes of the AD systems become design variables thus adding more complexity to the DDPG algorithm. Several scenarios describing a typical operation of these systems are considered. First, the modifications and considerations of the AD systems are explained. The DDPG configuration and the corresponding optimization problem for each AD system are described next. In section 4.2.3., a comparison between a sequential approach and an integrated process design and control approach for the single-stage AD system is presented. Section 4.2.4 presents the integrated process design and control approach implemented for the two-stage AD system. To compare both systems and provide further insight of advantages and limitations of each system, a comparison between the single-stage and the two-stage (section 4.2.5) is described at the end of this section.

4.2.1.-DDPG structure

It is assumed for this study the usage of a digital twin to simulate the dynamic operation of the AD processes. For each episode, 365 steps were performed, where each time-step represents one day of operation. The parameters of the DDPG algorithm are updated at every time-step; the four DNNs have two hidden layers of 128 and 64 neurons. For the actor DNN, linear activation and sigmoid activation functions were used for the hidden and output layers, respectively. For the critic DNN, linear activation and rectified linear unit activation functions were selected for the hidden and output layers, respectively. In this study, the Scipy ordinary differential equations solver was used to simulate the environment of DDPG, i.e., the mathematical model of the AD systems. For the exploration noise, Ornstein-Uhlenbeck (OU) noise (Uhlenbeck & Ornstein, 1930) is added (Eq. 4.4) with parameters of $\sigma = 0.015$ and $\varphi = 0.15$. The minibatch size is T=64; the replay buffer size is P=10000. Moreover, Adam optimizer (Kingma & Ba, 2014) is used for the training of the DNNs. The hyperparameters selected for the DDPG are presented on Table 4.1. This configuration setting was chosen from on our previous study presented in chapter 3. The present approach was implemented using Python Pytorch in Anaconda; the calculations were performed on an Intel Core i7 CPU at 1.7 GHz and 8.00GB memory.

Table 4.1.- Hyper-parameters configuration for AD systems.

HYPERPARAMETER	SINGLE-STAGE	TWO-STAGE
Learning rate for critic DNN	0.0001	0.001
Learning rate for actor DNN	0.001	0.001
Discount factor	0.95	0.99

4.2.2.-Modelling characteristics of AD systems

The AD system model presented in chapter 3 assumed fixed volumes for the reactors. To account for changes in the reactor's capacity, the following modifications are considered. For the single-stage AD-system, the following algebraic equation was added:

$$D = \frac{Q}{V} \tag{4.6}$$

where Q is the volumetric flow of the system (L/day), i.e., the vinasses in the AD system and V represents the reactor's capacity (L).

Similarly, for the two-stage AD system, the following algebraic equations are added:

$$\beta = \frac{V_2}{V_1} \tag{4.7}$$

$$D_2 = \frac{Q_1}{V_1} \tag{4.8}$$

$$D_1 = \beta D_2 \tag{4.9}$$

where $V_1$and V_2 correspond to the acidogenic and methanogenic reactor volume, respectively; β describes the ratio between the two reactors and Q_1 and Q_2 (L/day) are the volumetric flows of the acidogenic reactor and methanogenic reactor. The objective function considered for both single-stage and two-stage AD systems consists in the maximization of the economic functions presented in Eq. (4.10) and Eq. (4.11), where the subscripts 1 and 2 corresponds to the single-stage system and the two-stage system, respectively.

$$OF_1 = -\mathrm{CC}_1 + \mathrm{PP}_1 - EC_1 - \mathrm{VC}_1 \tag{4.10}$$

$$OF_2 = -\mathrm{CC}_2 + \mathrm{PP}_2 - EC_2 - \mathrm{VC}_2 \tag{4.11}$$

For the present work, the capacities of the digesters were used to specify the annualized capital costs of the economic function. Although the two-stage AD system considers two up-flow fixed

bed reactors and the single-stage a CSTR, all the reactors were assumed to be the same reactor-type when calculating the annualized costs. Both single-stage and two-stage AD systems were assumed to be vertical process vessels made of carbon steel. Accordingly, the bare-module cost in 2001 for each AD digester are as follows:

$$C_{BMR1} = 4.07\left[\left(10^{3.4974+0.1074\log_{10}V^2}\right)\left(V^{0.4485}\right)\right] \quad (4.12)$$

$$C_{BMR21} = 4.07\left[\left(10^{3.4974+0.1074\log_{10}V_1^2}\right)\left(V_1^{0.4485}\right)\right] \quad (4.13)$$

$$C_{BMR22} = 4.07\left[\left(10^{3.4974+0.1074\log_{10}V_2^2}\right)\left(V_2^{0.4485}\right)\right] \quad (4.14)$$

where C_{BMR1} (Eq. 4.12) corresponds to the bare-module cost for the single-stage system; C_{BMR21} and C_{BMR22} in Eq (4.13) and (4.14) correspond to the bare-module costs for the first and second reactor in two stage configuration, respectively. These costs were converted to 2020 USD using the Plant Cost Index from Chemical Engineering Magazine (The Chemical Engineering Plant Cost Index - Chemical Engineering, 2020.). Thus, the annualized capital costs for the single-stage AD (CC_1) system and the two-stage system (CC_2) are as follows:

$$CC_1 = r(C_{BMR1})(1/365) \quad (4.15)$$

$$CC_2 = r(C_{BMR21} + C_{BMR22})(1/365) \quad (4.16)$$

where *r* is the desired return on investment and under the assumption of 15%/year return (Mallon & Weersink, 2007). Note that the present study only considers changes in the capital costs due to the changes in the reactors' capacities. The bare-module costs were obtained from Turton et al., (2008). To define the bounds of the design variables, i.e., the capacities of the digesters, it is assumed that the acidogenic reactor from the two stage AD system and the reactor from the single stage have the same capacity limits (Ghanimeh et al., 2020).

To address the variability of the AD systems with respect to the set-point for COD, the following penalty functions are included in each system:

$$VC_1 = -pc \sum_{j=1}^{M} \omega_j \sum_{t_0}^{t_f} (z_3 - spCOD) Q * \Delta t \tag{4.17}$$

$$VC_2 = -pc \sum_{j=1}^{M} \omega_j \sum_{t_0}^{t_f} (x_6 - spCOD) Q_1 * \Delta t \tag{4.18}$$

where VC_1 in Eq. (4.17) represents the penalty function for the single-stage AD system and VC_2 in Eq. (4.18) refers to the penalty function for the two-stage AD system, respectively. As shown in both equations, a COD set-point ($spCOD$) of 2 g COD/L was assumed (Piceno-Diaz, 2018); t_f represents the final integration time (365 days) whereas the sampling interval Δt was set to 1 day; *pc* is the penalty cost that accounts for set-point tracking errors, with a user-defined value of -0.015 (USD/g COD) taken from the literature (Wastewater Rates :: East Bay Municipal Utility District, 2022) and (Kuo & Dow, 2017); ω_j represents the probability for each jth uncertain realization; z_3 and x_6 represent the corresponding COD concentrations of the single-stage and two-stage AD systems, respectively.

The production profits account for the methane production in the following equations:

$$PP_1 = \sum_{j=1}^{M} q\, y_{CH_4} y_{mb}\, \omega_j \sum_{t_0}^{t_f} z_2 Q \tag{4.19}$$

$$PP_2 = \sum_{j=1}^{M} q\, y_{CH_4} y_{mb}\, \omega_j \sum_{t_0}^{t_f} x_5 Q_1 \tag{4.20}$$

where PP_1 accounts for the methane production of biogas in the single-stage system, while PP_2 accounts for the methane production of biogas in the two-stage system y_{CH_4} has a value of 3.82 ($gCH4/g_{mb}$) and is the yield coefficient of methane production with respect to methanogenic biomas; g_{mb} represents the methanogenic biomass for each system (i.e., z_2 for the single-stage and x_5 for the two-stage AD system) ; y_{mb} (g_{mb} /mmol VFA) is the biomass factor with a value of 0.69 (Moguel-Castañeda et al., 2020). In the present study, the price of natural gas is assumed as the

selling price for methane (i.e., q in Eqs. (4.19) and (4.20)). Thus, the natural gas price of 2.71 USD/ft^3 from June 2020 is considered in the calculations of methane production in present work (U.S. Energy Information Administration, 2022).

The energy consumption costs in this chapter are assumed to be equivalent to the cost for tequila vinasses treatment; these costs can be estimated as follows:

$$EC_1 = -p \sum_{t_0}^{t_f} Q_{*\Delta t} \tag{4.21}$$

$$EC_2 = -p \sum_{t_0}^{t_f} Q_{1*\Delta t} \tag{4.22}$$

where EC_1 and EC_2 refer to the energy consumption costs from the single-stage and two-stage system, respectively; p is the tequila vinasses treatment price, which corresponds to 16 USD/m^3 (Martinez-Orozco et al., 2020).

Based on the above descriptions, the optimization problem for the single-AD system is as follows:

$$\max_{D(t), V_1} -\ \mathrm{CC}_1 + \mathrm{PP}_1 - EC_1 -\ \mathrm{VC}_1 \tag{4.23}$$

s.t.

$\dot{\boldsymbol{z}}(t) =\ \Omega^I(\boldsymbol{z}(t), \boldsymbol{D}(t), \boldsymbol{\xi})$

$\boldsymbol{z}(t) = \boldsymbol{z}(t_0)$

$0.1 \leq D(t) \leq 1.4174$

$5000 \leq V_1 \leq 10000$

t=[0, t_f]

Similarly, the optimization problem for the two-state AD system is as follows:

$$\max_{D_2(t), V_2, \beta} -\ \mathrm{CC}_2 + \mathrm{PP}_2 - EC_2 -\ \mathrm{VC}_2 \tag{4.24}$$

s.t.

$\dot{\boldsymbol{x}}(t) = \Omega^{II}(\boldsymbol{x}(t), \boldsymbol{D_2}(t), \boldsymbol{\xi})$

$\boldsymbol{x}(t) = \boldsymbol{x}(t_0)$

$0.05 \leq D_2(t) \leq 0.22$

$5000 \leq V_1 \leq 10000$

$1.6 \leq \beta \leq 5$

t=[0, t_f]

As for the capacities of the digesters, it is assumed that both acidogenic reactors have the same capacity limits (Ghanimeh et al., 2020). The digesters' capacities are within a range of 5m^3 to 50 m^3, which is common for Tequila factories (Méndez-Acosta et al., 2010); the ratio β between the two digesters in the two-stage AD system is set within a range based on different studies (Cohen et al., 1979, 1980; Saddoud et al., 2007; Vergara-Fernández et al., 2008). The single stage AD model is represented by function Ω^I and represents Eq. (3.12) and Eq. (4.6); the Ω^{II} represents the two-stage model (Eqs. (3.13), (3.14), (4.7), (4.8), and (4.9)). As in the optimal control approach presented in Chapter 3, parametric uncertainty was considered for the present case studies using the same uncertain parameters; however, different values were assumed for the set of uncertainty realizations. The weights (ω_j) of each realization are assumed to be 0.4 for the nominal realization (j=5), 0.05 for the "extreme" realizations (meaning that they are less likely to occur), and 0.1 for the rest of the realizations (see table 4.2, 4.3 below). Due to the sensitivity analysis from Piceno-Díaz et al., (2020) and the results from Chapter 3 of the present study, it is assumed that the extreme realizations are the cases with the highest and lowest biomass fractions. The fraction of the single-stage is represented by α whrereas for the two-stage, only the biomass fraction α_2 of the methanogenic reactor was considered.

As there was no previous knowledge of a specific number of episodes to guarantee convergence for an integrated process design and control framework using the DDPG-based approach, a study to determine the final number of episodes was performed. Each experiment of the training process of DDPG was performed five times using 500, 1000 and 1500 episodes under nominal/ideal

conditions, i.e., no disturbances or uncertainty were considered. The expected accumulated reward, i.e., the annualized costs (OF_1 in Eq. (4.10)) and (OF_2 in Eq. (4.11)), and CPU times were recorded together with their respective standard deviation.

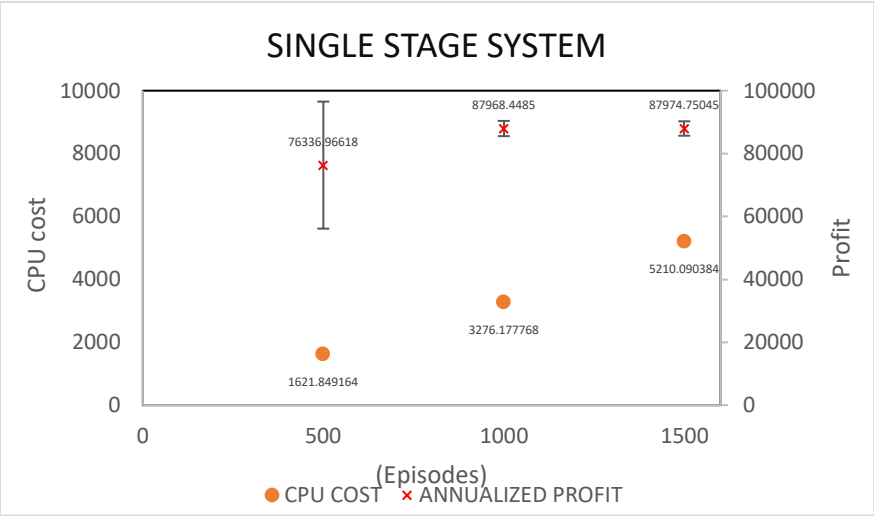


Fig. 4.2.- Terminal stage determination analysis. Annualized profit vs CPU cost for single-stage AD system

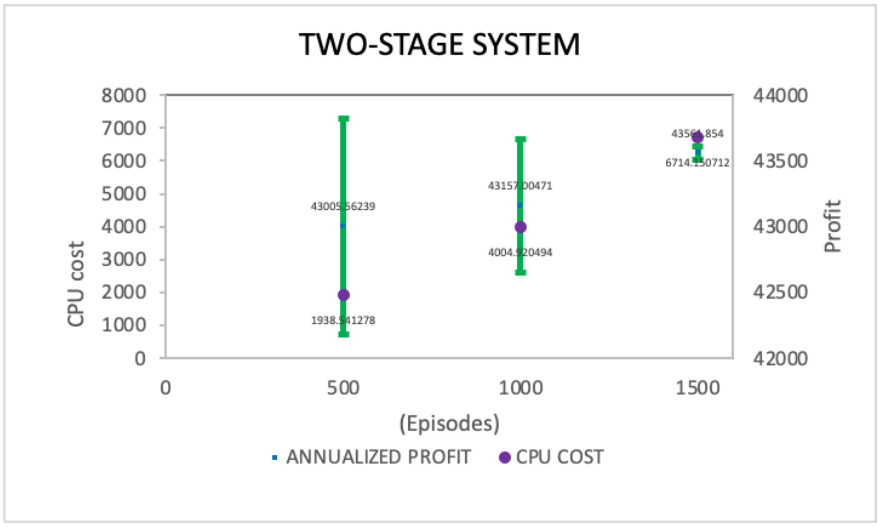


Fig. 4.3.- Terminal stage determination analysis. Annualized profit vs CPU cost for two-stage AD system.

As shown in Fig. 4.2 and 4.3, the criteria of 1500 episodes showed the highest annualized profit. However, they also exhibited the highest CPU costs for both systems. Hence, the criteria selected

for the terminal stage was 1000 episodes, as it showed high enough annualized costs and smaller CPU costs compared to those obtained in 1500 episodes.

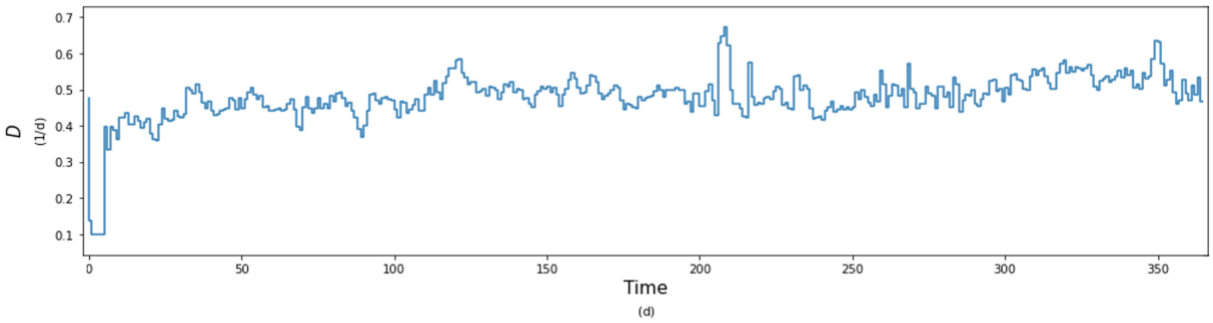
4.2.3.- Scenario 1: Integrated vs sequential approach

The aim of this section is to present a comparison between the integrated process design and control and the sequential method by implementing both techniques on the single-stage AD system. For the sequential approach, the optimization problem from Eq. (4.1) was implemented with the interior-point optimization algorithm (IPOPT) (Wächter & Biegler, 2005). Nominal conditions are assumed, i.e., no disturbances or parametric uncertainty were considered. To transform the problem into a steady-state optimization, the time domain in the process model was not considered. The resulting capacity of the digester from the steady state optimization was 5000 L with a dilution rate of 0.476 d^{-1}. Once these values were obtained, a dynamic optimization (Eq. (3.1)) was implemented using the DDPG algorithm. For this optimization, the capacity of the reactor obtained in the steady-state optimization is fixed, i.e., no design variables are considered, assuming the dilution rate $D(t)$ as the only control variable. To account for the possible variations of the organic matter content of tequila vinasses, stochastic disturbances and parametric uncertainty were considered for the dynamic optimization of the sequential approach and for the integrated design and control approach. Table 4.2 depicts the different values for the uncertain parameters used in this study, these values account for favorable and adverse conditions for microorganism growth. As the seasonal disturbances, steps with magnitudes $\boldsymbol{\pi}$ were added at day 15, 101, 247, and 320 of -25%, -15%, +20%, and +25% with respect to the nominal concentrations, i.e., the inlet concentrations of the substrates in terms of COD and VFA [27 g COD/L and 50 mmol VFA/L)]. respectively. Random Gaussian noises $\boldsymbol{\vartheta}$ (Eq. (4.3)) were added at a sampling interval of one

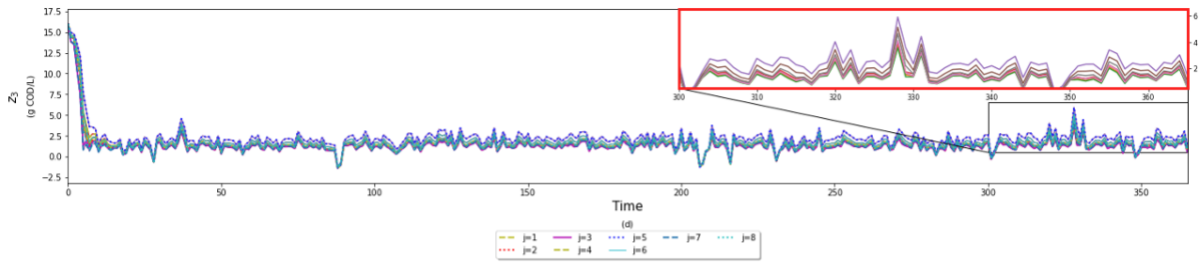
day. These Gaussian noises ($N[\varphi, \sigma\,]$) assume a zero-mean ($\varphi = 0$) with a standard deviation σ of 5% with respect to the nominal inlet concentrations.

Table 4.2.- Uncertainty parameters of scenario 1

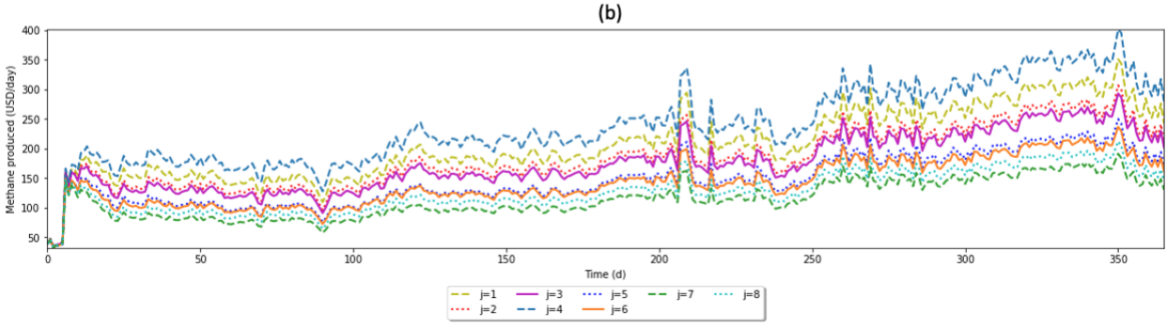
ζ	1	2	3	4	5	6	7	8
$\mu_{1max,}$	1.02	0.95	1.2	0.99	0.7999	0.83	1.05	0.87
$\mu_{2max,}$	0.7	1.03	0.88	0.97	0.7357	0.91	0.81	0.92
k_{s1}	6.2	4.53	5.71	7.3	5.207	4	3.35	3.12
α	0.35	0.41	0.42	0.32	0.458	0.5	0.58	0.55



(a)



(b)



(c)

Fig. 4.4.- Scenario 1: Comparison of sequential approach and integrated process design and control approach applied on single-stage AD system. (a) Dilution rate of the digester in sequential approach, (b) COD profile of effluent in sequential approach (c) Methane production profile in sequential approach.

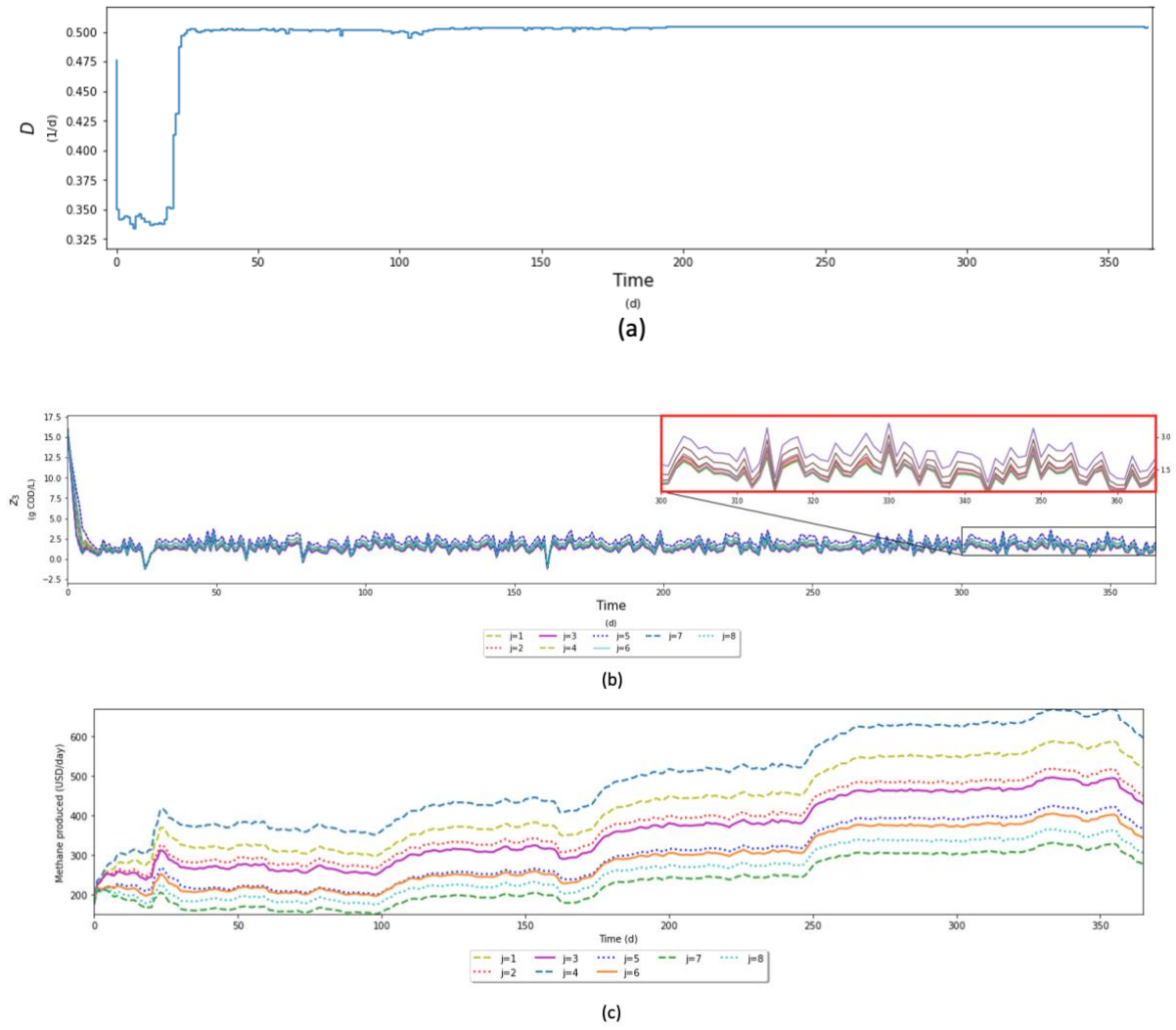


Fig 4.5.- Scenario 1: Comparison of sequential approach and integrated process design and control approach applied on single-stage AD system. (a) COD profile of effluent in integrated approach, (b) COD profile in sequential approach, (c) Methane production profile in integrated approach.

The overall profits of the integrated process design and control were $83,130.6 USD/year, representing an increase of at least twice of that observed in the sequential approach ($39,938.5 USD/year). The selected capacity for the reactor in the integrated approach was 10,000 L, which is the double the capacity of that specified by the sequential approach. Due to this difference of capacities, the capital costs in the integrated approach resulted 16% more expensive than the sequential. Similarly, the integrated approach treated double amount of vinasses than the sequential approach, thus resulting in higher profits. From an energy-point of view (Fig. 4.4 (c) and Fig. 4.5(c)), the methane produced in the integrated approach was approximately 2.04 times the methane produced in the sequential approach, i.e., the integrated approach produced an expected value of $119,307.8 USD/year. Similar results were reported by Kuo & Dow, (2017) where field data from a wastewater treatment plant employing AD was reported with methane production of the same order of magnitude as that one obtained by the present integrated approach for the single-stage AD system. From an environmental perspective, the integrated approach produced slightly a higher expected concentration of COD (z_3) in the effluent than the sequential (4% more). Nonetheless, the integrated approach process double the amount of liters of vinasses at an equivalent organic concentration. As depicted in Fig. 4.5 (a), the integrated DDPG-based approach specified a rapid decrease in the dilution rate at the beginning of the process; this was due to the reduction of the organic concentrations of the vinasses due to the stochastic disturbances. After the initial transients, the dilution rate progressively increased until reaching a steady-state; this increase aims to feed the system with more tequila vinasses and thus, produce more biogas. This agrees with the study of Poh et al., (2016) that showed that anaerobic digestors with higher organic concentrations shall be operated with high dilution rates. Also, by comparing both dilution rate profiles (Fig 4.4 (a) and 4.5(a)), it was observed a smoother operation for the integrated

approach. Fig. 4.4 (b) and Fig 4.5 (b) showed that both sequential and integrated approach reached COD concentrations above the set-point, suggesting that the trade-off between performance (profits) and the variability cost led to a set-point violation in exchange for more profits. From the last 100 days of the operation, the sequential approach violated the set-point almost 30% of the time, while the integrated design and control scheme did the same for 23%. Although these results are somewhat similar, the violation of the set-point was more notorious in the sequential approach, as the values were reaching concentrations around 6 g COD/L near day 330, as shown in Fig. 4.4 (b). For the integrated approach, the highest values were around 3 g COD/L near day 330 (Fig. 4.5 (b)). Hence, a more smooth and feasible operation was identified from the integrated approach. The fact that both techniques showed the highest COD concentrations at similar days is due to the stochastic disturbances, as it is shown on Figs. 4.4 (b) and 4.5 (b) that around day 330 is where the highest concentration of tequila vinasses entered the system. The production of methane represents the highest contribution to the economic function in both approaches, while the capital cost CC has the lowest contribution. Regarding computational costs, the integrated approach required a CPU time of 75,731 s while the sequential approach required a total of 75,878 s for the steady-state and dynamic optimization, which is an indication that the additional decisions (i.e., design variables) considered in the integrated approach did not affect considerably the CPU costs. Based on the above, the integrated approach showed multiple advantages with respect to the sequential approach, such as higher profits, similar environmental costs, and an increase in biogas production.

4.2.4- Scenario 2: integrated process design and control for a two-stage AD system

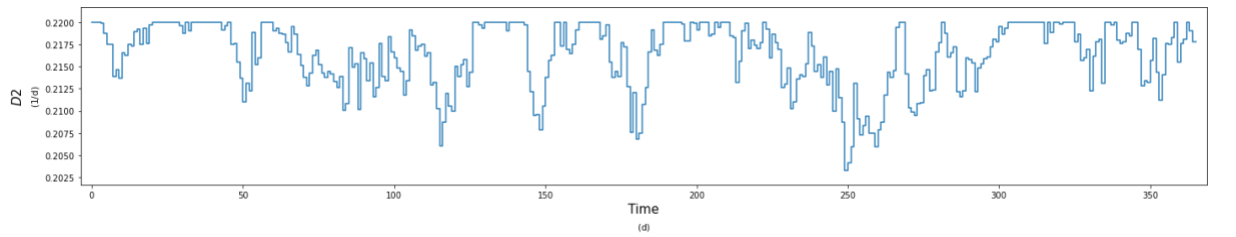
In contrast to the single-stage AD system, the two-stage system considers another design variable in the system, i.e., the ratio β (Eq. 4.7) between the two digesters. As in the previous scenario,

stochastic disturbances and parametric uncertainty were considered. Table 4.3 shows the uncertainty realizations for the two-stage AD system; these values account for the uncertain realizations consider favorable and adverse conditions for microorganism growth. The seasonal changing values in the disturbances also follow a step-wise profile complemented with random noises.

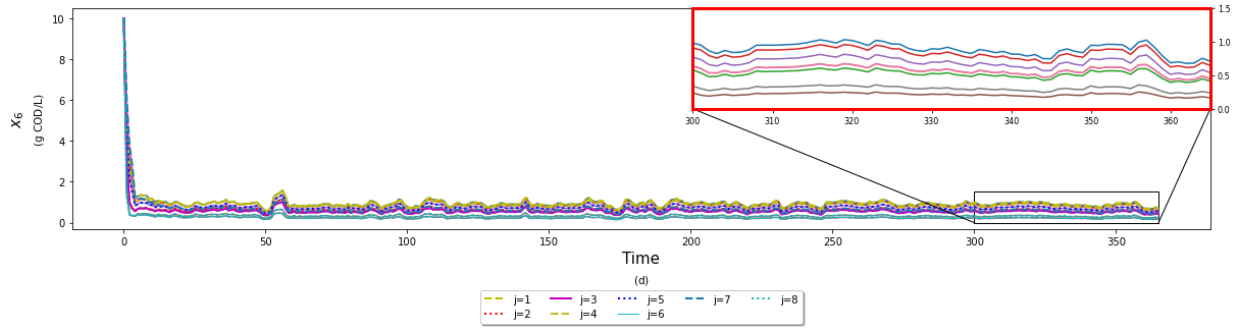
Table 4.3.- Uncertainty parameters for scenario 2: integrated process design and control for a two-stage AD system.

ζ	1	2	3	4	5	6	7	8
$\mu_{11max,}$	0.15	0.351	0.234	0.284	0.27	0.34	0.183	0.41
$\mu_{12max,}$	0.58	0.635	0.575	0.439	0.5	0.6	0.4	0.7
$\mu_{22max,}$	0.44	0.248	0.395	0.228	0.29	0.34	0.26	0.37
k_{s12}	5.0	3.607	2.671	3.372	3.5	2.0	3.0	2.75
α_1	0.085	0.17	0.15	0.159	0.13	0.18	0.11	0.20
α_2	0.4	0.403	0.43	0.392	0.38	0.26	0.29	0.32

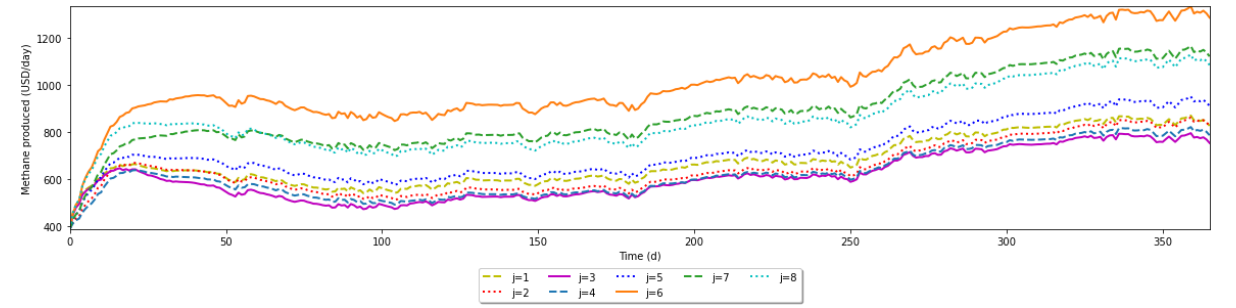
For this scenario, the DDPG resulted in an overall profit of $231,030.2 USD/year. The selected capacities were 10,000 L for the acidogenic reactor and 50,000 L for the methanogenic reactor, indicating a ratio β of value 5. The capital costs resulted in $55,270.9 USD/year; in addition, the system is able to process 3949720.1 liters of vinasses, resulting in EC of $ 63370.6 USD/year. In this scenario, the variability cost defined by the penalty function shown in Eq. (4.18) resulted in a positive value, i.e., the COD concentrations were more likely below their set-point thus less pollutants are diverted to the effluents thus making this process more environmentally-friendly. This variability cost represents 17% of the cost function.



(a)



(b)



(c)

Fig 4.6.-Scenario 2: integrated process design and control approach applied on a two-stage AD system. (a)Dilution rate profile, (b) COD profile (c) Methane production profile.

From an environmental perspective, the system did not exceed the COD limits over the last 100 days of the operation, as observed on Fig. 4.6 (b). This suggests that a two-stage reactor scheme promotes more adequate conditions for the microorganisms' growth, organic consumption and accordingly, more biogas production. From an energy point of view (Fig. 4.6 (c)), the methane produced was $267,549 USD/year, representing the highest contribution to the economic function

(58%), whereas the lowest contribution was the capital costs (11%). As depicted in Fig. 4.6 (a), the DDPG specified a dilution rate that tends to constantly saturate from the top, i.e., it reaches the upper bound of D_2; these high values aim to feed the system with more tequila vinasses and thus, produce more biogas. Nonetheless, it is also observed on Fig. 4.6 (a) that, to accommodate the stochastic disturbances and parametric uncertainty, the DDPG framework constantly changes the dilution rate with rapid movements that end up decreasing D_2 for short periods of time. This shows that the design and control scheme for this two-stage AD system has more capacity and flexibility to accommodate stochastic disturbances and parametric uncertainty. Regarding computational costs, a CPU time of 259,278 s was required, representing a change of one order of magnitude when compared to optimization under nominal conditions depicted in Fig. 4.3.

4.2.5.- Scenario 3: comparison between single-stage and two-stage AD system

As mentioned in Chapter 2, previous studies have shown that a two-stage AD configuration allows the development of more favorable growth conditions for microorganisms than the conventional approach of using a single digester. To provide a further insight, a comparison between these two systems was performed using the proposed DDPG methodology. For comparison purposes, the same inlet stream conditions of the two-stage AD system ($S_{11,in} = 27$ g COD/L) and $S_{21,in} =$ 50 mmol VFA/L) were used in the single-stage AD system as a nominal value for the seasonal values in the disturbances. As discussed in Chapter 3, both AD models were designed using different experimental conditions, so a direct comparison between the two AD systems cannot be made. Hence, parametric uncertainty was not considered, i.e., only stochastic disturbances were taken into account to perform the integration of design and control (Eq. (4.2)).

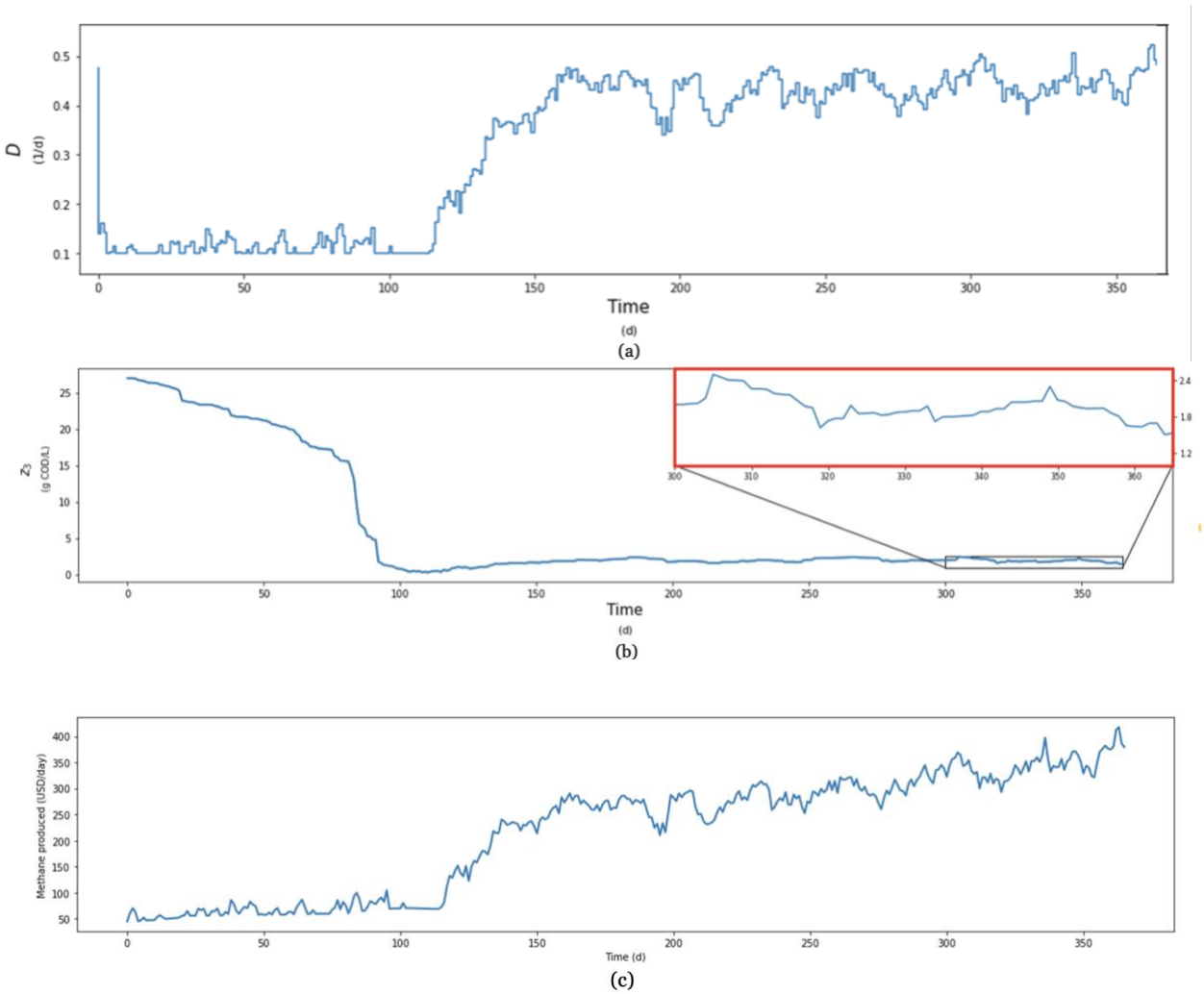


Fig.-4.7.-Scenario 3: comparison of single-stage AD system vs two-stage AD system. (a) COD profile of effluent in integrated approach in single-stage system, (b) COD profile in integrated approach in single-stage system, (c) Methane production profile in integrated approach in single-stage system

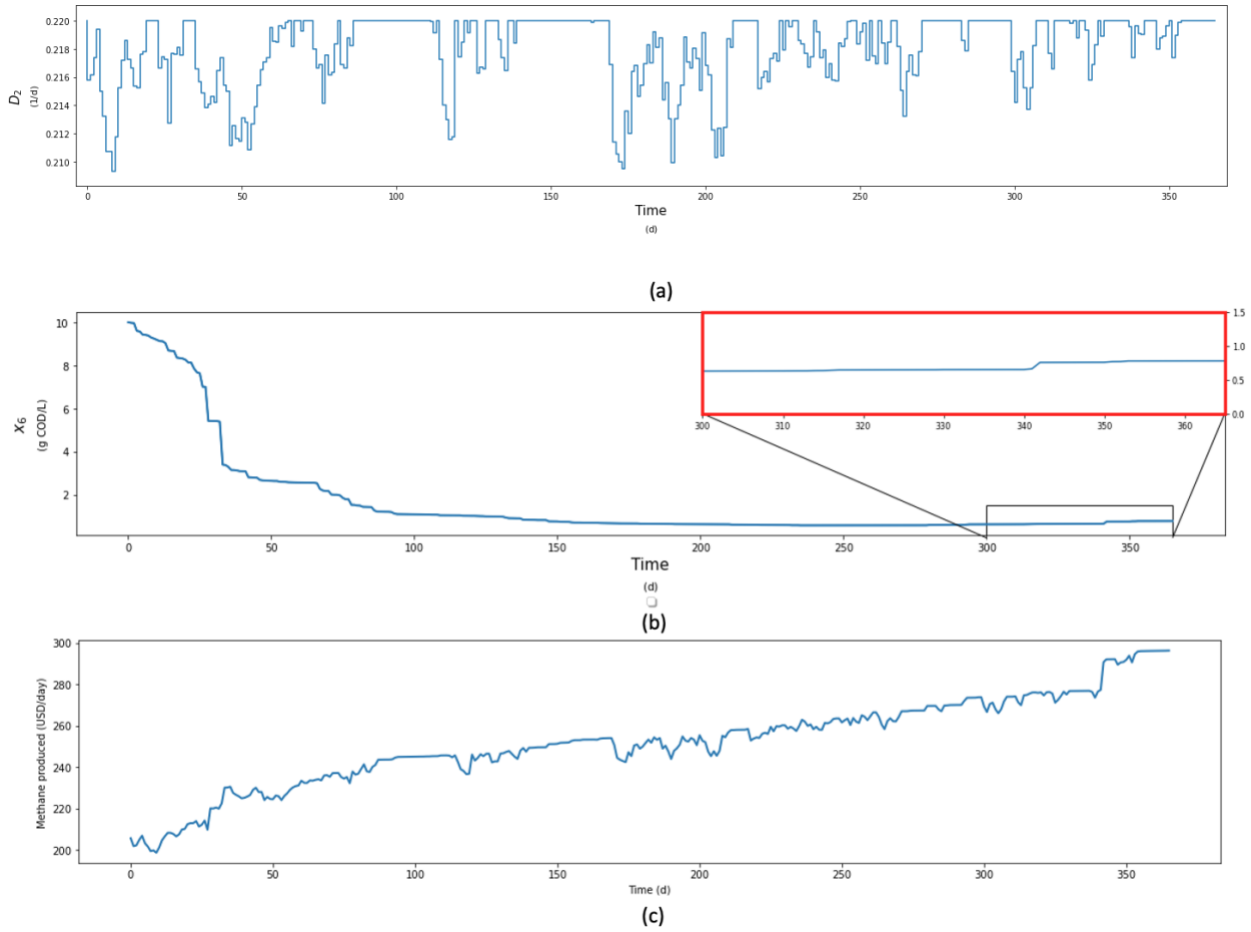


Fig.-4.8.- Scenario 3: comparison of single-stage AD system vs two-stage AD system. (a) COD profile of effluent in integrated approach in two-stage system, (b) COD profile in integrated approach in two-stage system, (c) Methane production profile in integrated approach in two-stage system

The overall profits obtained for the two-stage AD system were $30,718.6 USD/year, representing an increase in profits of 11% with respect to the single-stage system. For the two-stage AD system, the capacities were 9959.3 L for the acidogenic reactor and 26448.8 for the methanogenic, representing a ratio β of 2.6. On the other hand, the reactor capacity for the single-stage was 8567.8 L. Due to the difference in capacities, the volume of vinasses in the two-stage was equivalent to 2.1 times the volume of vinasses treated by the single-stage. The capital costs in the two-stage system were twice the costs obtained for the single-stage system. Nevertheless, the two-stage

produced 17% more methane than the single-stage, resulting in higher profits for the former. From an environmental perspective, the two-stage was able to accommodate the stochastic disturbances while respecting the specified set-point of COD and showed a smoother operation regarding COD reduction. As depicted in Figs. 4.7 (b) and Fig. 4.8 (b), a continuous decrease in COD was obtained from both approaches until reaching a steady state. However, the COD obtained in the two-stage returned a more steady operation with lower concentrations of COD than the single-stage system. In the single-stage, the proposed integrated scheme starts decreasing COD at a slower pace that results in less production of biogas. Fig. 4.7 (b) shows that the concentration z_3 of COD in single-stage AD system achieves values that are higher than their corresponding set-point. The value for the VC in the two-stage resulted in a positive value representing 7% of the economic function, while the VC for the single-stage resulted in negative profits as the set-point was violated multiple times during operation (see fig. 4.7 (b)); this suggests that the trade-off between performance (profits) and the variability cost led to a set-point violation in exchange for an increase in profits. From the last 100 days of the optimization, the single-stage exceeded the set-point 36% of the time, while the two-stage always met the set-point during this period of operation. The production of methane represents the highest contribution of the economic function in both approaches. The lowest contribution in the two-stage was the variability cost (7%), and in the single-stage, the capital costs (9%). Regarding computational costs, the two-stage system required a CPU time of 55,881 s, while the single-stage system consumed a total of 27,410 s. This shows how the CPU time increases considerably when the AD model becomes more complex. This supports the importance of the acidogenic microorganisms' growth, as it is known that they are able to convert substrate at a faster pace that the methanogenic bacteria ((López Velarde Santos et al., 2020; Solera et al., 2001). If this criterion is not carefully considered, treating more vinasses could lead to

accumulation of VFA, causing potential methanogenic inhibition and as a consequence, more biomass would be washed as the microorganisms would not be able to transform the substrate into methane. This analysis has shown that the two-stage AD system performs better than the single-stage AD system. Nonetheless, the user-defined criteria and parameters considered in this study may play a key role. For instance, the DDPG algorithm implemented for the single-stage returned a design and control scheme that sacrificed environmental aspects such as the set-point violation in COD, resulting in lower profits. On the other hand, due to the strengthen microbial growth achieved by the two-reactor configuration, the two-stage can accommodate better the variability of the process by obtaining concentrations below the COD set-point, which conducts to a more environmentally-attractive operation of the system. However, more investment in CC is required with a two-stage configuration but this allows the system to handle larger quantities of vinasses that result in higher plant profits.

4.3.-Summary

This chapter presented a methodology for an integrated process design and control based on the algorithm Deep Deterministic Policy Gradient. The case study involved AD systems under two different configurations (single-stage and two-stage) treating Tequila vinasses subject to stochastic disturbances and parametric uncertainty. A key aspect in this study was the consideration of stochastic disturbances, which are more challenging to handle but they represent a more realistic operation of the systems. The results showed that the proposed methodology exhibited improvements in performance with respect to a traditional approach, i.e., a sequential approach. Also, the integrated process design and control approach in the two-stage AD system showed higher profits in exchange of more expensive designs (i.e. higher capital costs). When comparing the two different reactor configurations under similar inlet conditions, several trade-offs were

observed, such as the ability to respect the imposed constraint on the COD, the investment on capital costs and methane production costs. Overall, the two-stage AD system exhibited a more stable and smooth operation with higher profits, while the single-stage resulted in lower profits than those obtained by the two-stage AD system.

CHAPTER 5

Conclusions and future work

5.1.-Conclusions

The objective of this thesis was to test the feasibility of a RL algorithm called Deep Deterministic Policy Gradient applied for process control and integrated process design and control. In chapter 3, the feasibility of the DDPG algorithm applied to AD systems was explored under different scenarios to optimize the reduction of COD while producing biogas. The results showed that the DDPG was able to abstract the states from the AD system and learn successfully through an adequate tuning of the hyper-parameters. When compared to conventional NLP solvers such as IPOPT, a crucial factor for the NLP solver was the trade-off between computational time and accuracy of the control performance by modifying the number of finite elements and collocation points. While the NLP solvers returned a low COD in shorter CPU times than those obtained by the DDPG, the latter algorithm accounts for plant-model mismatch, which makes it a suitable candidate for a more realistic operation. When the single-stage and the two-stage AD configurations were compared using the same substrate, the later resulted in a significant reduction in the accumulated COD at the effluent, thus resulting in a more attractive operation. The DDPG algorithm was also able to accommodate disturbances in the inlet feed concentrations of tequila vinasses combined with uncertainty in the kinetic parameters for both configurations. Furthermore, the proposed algorithm was able to return an acceptable performance in the presence of stochastic disturbances in the inlet stream. Although several disturbances and parametric uncertainty were considered, the open-loop control actions can accommodate those effects while maintaining the COD concentration in the methanogenic reactor at low values. Additionally, the robust EMPC for

the single-stage AD system was able to maintain a low COD in the presence of disturbances and parametric uncertainty. The results presented in this study have shown the potential that RL algorithms could offer to address control for chemical systems. Although the DDPG algorithm demonstrated the ability to learn optimal control policies under the different scenarios considered in this work, in particular for cases involving time-varying random disturbances, the CPU time required by this method is still considerable and may not be acceptable for applications involving short closed-loop time constants. Further improvements are thus needed to realize the application of this technique for online industrial-scale applications.

A methodology to address an integrated process design and control using DDPG was presented in Chapter 4. The key contribution of this methodology is the consideration of the interaction of time-dependent (manipulated variables $\boldsymbol{u}$) and time-independent (design variables ***des***) for the DDPG through a customized activation function and the consideration of stochastic disturbances. By maximizing an economic function, the DDPG aimed to identify optimal designs while obtaining open-loop control profiles that can accommodate the transient operation of the system. Three different scenarios using a single-stage and two-stage AD systems were tested. Overall, the single-stage AD system resulted in lower capital costs but also returned lower profits compared to the two-stage AD system. Although the two-stage configuration may be seen as a larger investment on equipment, it exhibited better performance in terms of COD reduction and thus, more methane production was observed for this case study. Regarding the variability cost (VC), this study considered a penalty function in the objective function to enforce this condition as a constraint. For this case study, the constraint's objective was formulated such that it was required to meet a specified (user-defined) set-point in COD in the effluent stream. The results showed that the DDPG

was able to accommodate the process constraint in the two-stage AD system for most of the operation whereas a higher frequency of constraint violation was observed for the single-stage system. This suggests that it might be more attractive for the single-stage AD system to pay a higher environmental cost, i.e., release more organic content through the effluent, at the expense of producing more biogas and therefore increasing the plant profits. Moreover, it was observed that the single-stage system treated a less volume of vinasses while the two-stage tend to deal with higher capacities of vinasses. When comparing the integrated approach with a traditional sequential approach, it was observed a more stable and smooth operation and a higher biogas production for the integrated approach. Overall, the results showed a promising potential for RL in large-scale applications, particularly for slow processes such as AD systems, where daily control actions may be sufficient to operate the process near optimal conditions.

5.2.- Recommendations for Future Work

The study presented in this thesis can be extended in different ways to provide a further insight on RL applications in process control for AD systems, as well as different applications for addressing an integrated process design and control problem. The recommendations that can be pursued as part of the future work in this study are as follows:

- The methodologies presented in this work obtained the hyperparameter's tuning through trial-and-error experiments. Although this methodology resulted in an acceptable performance, the tuning procedure was time consuming. Hence, further improvement can be achieved by performing a hyperparameter optimization such as search grid or Bayesian model-based optimization (Dewancker et al., 2016).
- The proposed methodology of an integrated process design and control assumed that the design variables were continuous variables. Consideration of integer variables in the process design

and control methodology using DDPG is therefore suggested as a method to further refine the present methodology and offer more attractive design and control schemes while taking into account structural decisions.

- This study focused on the case study of AD systems treating Tequila vinasses. A potential future work could also consider different substrates of AD systems to determine the benefits and limitations in terms of AD systems performance.
- An attractive area of opportunity would be to implement the proposed methodology on large chemical systems to explore the feasibility of the DDPG algorithm to handle even higher dimensional problems.
- The case studies considered for the integrated process design and control only included a single constraint in the formulation, which was represented as a penalty function in the DDPG formulation. Process models that include a more realistic representation consider multiple constraints related to economics, sustainability, and process operation. Further improvements of this methodology could include the consideration of more constraints as penalty functions.
- A key assumption made in the integrated design and control framework is full access to the system states. In practice, only a limited number of states can be accessed online using measurements, soft sensors or state estimation methods (Valipour et al., 2021; Valipour & Ricardez-Sandoval, 2021a) Moreover, states and model parameters are subject to constraints that are often ignored in the analysis (Valipour & Ricardez-Sandoval, 2021b, 2022). A future area of research can consider the development of robust integrated design and control formulations that explicitly consider state and parameter estimation schemes that are subject to process constraints.

- The methodologies considered in this work were based on the DDPG algorithm, which is one of the most advanced algorithms. To date, there is one algorithm that is very similar to DDPG, called Twin Delayed DDPG (TD3). Future work in this area could consider the adaption of the proposed methodologies to the TD3 algorithm. Such modifications are expected to improve the quality of the solution, but they may also require higher computational costs.

REFERENCES

Ahring, B. K., Angelidaki, I., de Macario, C. C., Gavala, H. N., Hofman-Bang, J., Elfering, S. O., & Zheng, D. (2003). *Biomethanation I* (Vol. 81). Springer.

Alvarado-Morales, M., Hamid, M. K. A., Sin, G., Gernaey, K. v., Woodley, J. M., & Gani, R. (2010). A model-based methodology for simultaneous design and control of a bioethanol production process. *Computers and Chemical Engineering*, *34*(12), 2043–2061. https://doi.org/10.1016/J.COMPCHEMENG.2010.07.003

Arai, S., Sycara, K., & Payne, T. R. (2000). Multi-agent reinforcement learning for planning and scheduling multiple goals. *Proceedings - 4th International Conference on MultiAgent Systems, ICMAS 2000*, 359–360. https://doi.org/10.1109/ICMAS.2000.858474

Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). *Layer Normalization*. https://arxiv.org/abs/1607.06450v1

Bahakim, S. S., & Ricardez-Sandoval, L. A. (2014). Simultaneous design and MPC-based control for dynamic systems under uncertainty: A stochastic approach. *Computers and Chemical Engineering*, *63*, 66–81. https://doi.org/10.1016/J.COMPCHEMENG.2014.01.002

Bangi, M. S. F., & Kwon, J. S. il. (2021). Deep reinforcement learning control of hydraulic fracturing. *Computers & Chemical Engineering*, *154*, 107489. https://doi.org/10.1016/J.COMPCHEMENG.2021.107489

Bansal, V., Sakizlis, V., Ross, R., Perkins, J. D., & Pistikopoulos, E. N. (2003). New algorithms for mixed-integer dynamic optimization. *Computers & Chemical Engineering*, *27*(5), 647–668. https://doi.org/10.1016/S0098-1354(02)00261-2

Bemporad, A., Morari, M., Dua, V., & Pistikopoulos, E. N. (2002). The explicit linear quadratic regulator for constrained systems. *Automatica*, *38*(1), 3–20. https://doi.org/10.1016/S0005-1098(01)00174-1

Bernal, D. E., Carrillo-Diaz, C., Gómez, J. M., & Ricardez-Sandoval, L. A. (2018). Simultaneous design and control of catalytic distillation columns using comprehensive rigorous dynamic models. *Industrial and Engineering Chemistry Research*, *57*(7), 2587–2608. https://doi.org/10.1021/ACS.IECR.7B04205/SUPPL_FILE/IE7B04205_SI_002.PDF

Boulesnane, A., & Meshoul, S. (2021). Reinforcement Learning for Dynamic Optimization Problems. *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. https://doi.org/10.1145/3449726

Bouwmans, T., Javed, S., Sultana, M., & Jung, S. K. (2019). Deep neural network concepts for background subtraction:A systematic review and comparative evaluation. *Neural Networks*, *117*, 8–66. https://doi.org/10.1016/J.NEUNET.2019.04.024

Calvo, F., Gómez, J. M., Alvarez, O., & Ricardez-Sandoval, L. (2022). Trends and perspectives on emulsified product design. *Current Opinion in Chemical Engineering*, *35*, 100745. https://doi.org/10.1016/J.COCHE.2021.100745

Cavalcanti, F. M., Kozonoe, C. E., Pacheco, K. A., Maria, R., & Alves, B. (2021). Application of Artificial Neural Networks to Chemical and Process Engineering. *Deep Learning Applications*. https://doi.org/10.5772/INTECHOPEN.96641

Chachuat, B., Singer, A. B., & Barton, P. I. (2006). Global Methods for Dynamic Optimization and Mixed-Integer Dynamic Optimization. *Industrial and Engineering Chemistry Research*, *45*(25), 8373–8392. https://doi.org/10.1021/IE0601605

Chen, K., Wang, H., Valverde-Pérez, B., Zhai, S., Vezzaro, L., & Wang, A. (2021). Optimal control towards sustainable wastewater treatment plants based on multi-agent reinforcement learning. *Chemosphere*, *279*, 130498. https://doi.org/10.1016/J.CHEMOSPHERE.2021.130498

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2018). Neural Ordinary Differential Equations. *NIPs*, *109*(NeurIPS), 31–60. https://doi.org/10.48550/arxiv.1806.07366

Cohen, A., Breure, A. M., van Andel, J. G., & van Deursen, A. (1980). Influence of phase separation on the anaerobic digestion of glucose—I maximum COD-turnover rate during continuous operation. *Water Research*, *14*(10), 1439–1448. https://doi.org/10.1016/0043-1354(80)90009-3

Cohen, A., Zoetemeyer, R. J., van Deursen, A., & van Andel, J. G. (1979). Anaerobic digestion of glucose with separated acid production and methane formation. *Water Research*, *13*(7), 571–580. https://doi.org/10.1016/0043-1354(79)90003-4

Dankwa, S., & Zheng, W. (2019). Twin-Delayed DDPG: A Deep Reinforcement Learning Technique to Model a Continuous Movement of an Intelligent Robot Agent. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3387168.3387199

de Carvalho, R. F., & Alvarez, L. A. (2020). *Simultaneous Process Design and Control of the Williams−Otto Reactor Using Infinite Horizon Model Predictive Control*. https://doi.org/10.1021/acs.iecr.0c01953

DeCoursey, W. J. (2004). *Statistics and Probability for Engineering Applications With Microsoft ® Excel*.

Dewancker, I., Mccourt, M., & Clark, S. (2016). *Bayesian Optimization for Machine Learning A Practical Guidebook*.

Diangelakis, N. A., Burnak, B., Katz, J., & Pistikopoulos, E. N. (2017). Process design and control optimization: A simultaneous approach by multi-parametric programming. *AIChE Journal*, *63*(11), 4827–4846. https://doi.org/10.1002/AIC.15825

Dogru, O., Velswamy, K., & Huang, B. (2021). Actor–Critic Reinforcement Learning and Application in Developing Computer-Vision-Based Interface Tracking. *Engineering*, *7*(9), 1248–1261. https://doi.org/10.1016/J.ENG.2021.04.027

Edgar, T., Himmelblau, D., & Lasdon, L. (2001). *Optimization of chemical processes* (Vol. 2). McGraw Hill . https://www.yumpu.com/en/document/view/65250018/mcgraw-hill-chemical-engineering-series-thomas-f-edgar-david-m-himmelblau-optimization-of-chemical-processes-mcgraw-hill-2001

el Naqa, I., Murphy, M. J., el Naqa, I., & Murphy, M. J. (2015). What Is Machine Learning? *Machine Learning in Radiation Oncology*, 3–11. https://doi.org/10.1007/978-3-319-18305-3_1

Ferenci, T. (1999). `Growth of bacterial cultures' 50 years on: towards an uncertainty principle instead of constants in bacterial growth kinetics. *Research in Microbiology*, *150*(7), 431–438. https://doi.org/10.1016/S0923-2508(99)00114-X

Flores-Tlacuahuac, A., & Biegler, L. T. (2007a). Simultaneous mixed-integer dynamic optimization for integrated design and control. *Computers & Chemical Engineering*, *31*(5–6), 588–600. https://doi.org/10.1016/J.COMPCHEMENG.2006.08.010

Flores-Tlacuahuac, A., & Biegler, L. T. (2007b). Simultaneous mixed-integer dynamic optimization for integrated design and control. *Computers & Chemical Engineering*, *31*(5–6), 588–600. https://doi.org/10.1016/J.COMPCHEMENG.2006.08.010

Ghanimeh, S., Al-Sanioura, D., Saikaly, P. E., & El-Fadel, M. (2020). Comparison of Single-Stage and Two-Stage Thermophilic Anaerobic Digestion of SS-OFMSW During the Start-Up Phase. *Waste and Biomass Valorization*, *11*(12), 6709–6716. https://doi.org/10.1007/S12649-019-00891-8/FIGURES/6

Gupta, D., Agrawal, U., Arora, J., & Khanna, A. (2020). Bat-inspired algorithm for feature selection and white blood cell classification. *Nature-Inspired Computation and Swarm Intelligence*, 179–197. https://doi.org/10.1016/B978-0-12-819714-1.00022-1

Hua, H., Qin, Y., Hao, C., & Cao, J. (2019). Optimal energy management strategies for energy Internet via deep reinforcement learning approach. *Applied Energy*, *239*, 598–609. https://doi.org/10.1016/J.APENERGY.2019.01.145

Hubbs, C. D., Li, C., Sahinidis, N. v., Grossmann, I. E., & Wassick, J. M. (2020). A deep reinforcement learning approach for chemical production scheduling. *Computers & Chemical Engineering*, *141*, 106982. https://doi.org/10.1016/J.COMPCHEMENG.2020.106982

Kingma, D. P., & Ba, J. L. (2014). Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. https://arxiv.org/abs/1412.6980v9

Koller, R. W., Ricardez-Sandoval, L. A., & Biegler, L. T. (2018). Stochastic back-off algorithm for simultaneous design, control, and scheduling of multiproduct systems under uncertainty. *AIChE Journal*, *64*(7), 2379–2389. https://doi.org/10.1002/AIC.16092

Kuo, J., & Dow, J. (2017). Biogas production from anaerobic digestion of food waste and relevant air quality implications. *Https://Doi.Org/10.1080/10962247.2017.1316326*, *67*(9), 1000–1011. https://doi.org/10.1080/10962247.2017.1316326

Lan, T., & An, Q. (2021). Discovering Catalytic Reaction Networks Using Deep Reinforcement Learning from First-Principles. *Journal of the American Chemical Society*, *143*(40), 16804–16812. https://doi.org/10.1021/JACS.1C08794/SUPPL_FILE/JA1C08794_SI_002.XLSX

LA-Ricardez-Sandoval. (2008). *Simultaneous Design and Control of Chemical Plants: A Robust Modelling Approach*. University of Waterloo.

Lee, J. M., & Lee, J. H. (2004). Approximate dynamic programming strategies and their applicability for process control: A review and future directions. *International Journal of Control, Automation, and Systems*, *2*(3), 263–278. https://www.koreascience.or.kr/article/JAKO200411922627185.page

Liang, Y., Guo, C., Ding, Z., & Hua, H. (2020). Agent-Based Modeling in Electricity Market Using Deep Deterministic Policy Gradient Algorithm. *IEEE Transactions on Power Systems*, *35*(6), 4180–4192. https://doi.org/10.1109/TPWRS.2020.2999536

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). Continuous control with deep reinforcement learning. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*. https://arxiv.org/abs/1509.02971v6

López Velarde Santos, M., Ventura Ramos Jr., E., Rodríguez Morales, J. A., Oliver, H., López Velarde Santos, M., Ventura Ramos Jr., E., Rodríguez Morales, J. A., & Oliver, H. (2020). EFFECT OF INOCULUM SOURCE ON THE ANAEROBIC DIGESTION OF MEZCAL VINASSES AT DIFFERENT SUBSTRATE-INOCULUM RATIOS. *Revista Internacional de Contaminación Ambiental*, *36*(1), 81–95. https://doi.org/10.20937/RICA.2020.36.53276

López-López, A., Davila-Vazquez, G., León-Becerril, E., Villegas-García, E., & Gallardo-Valdez, J. (2010). *Tequila vinasses: generation and full scale treatment processes*. https://doi.org/10.1007/s11157-010-9204-9

Luo, G., Xie, L., Zhou, Q., & Angelidaki, I. (2011). Enhancement of bioenergy production from organic wastes by two-stage anaerobic hydrogen and methane production process. *Bioresource Technology*, *102*(18), 8700–8706. https://doi.org/10.1016/J.BIORTECH.2011.02.012

MacHalek, D., Quah, T., & Powell, K. M. (2020). Dynamic Economic Optimization of a Continuously Stirred Tank Reactor Using Reinforcement Learning. *Proceedings of the American Control Conference*, *2020-July*, 2955–2960. https://doi.org/10.23919/ACC45564.2020.9147706

Malcolm, A., Polan, J., Zhang, L., Ogunnaike, B. A., & Linninger, A. A. (2007). Integrating systems design and control using dynamic flexibility analysis. *AIChE Journal*, *53*(8), 2048–2061. https://doi.org/10.1002/AIC.11218

Mallon, S., & Weersink, A. (2007). *The financial feasibility of anaerobic digestion for Ontario's livestock industries*. https://ageconsearch.umn.edu/record/7295/

Mansouri, S. S., Huusom, J. K., Gani, R., & Sales-Cruz, M. (2016). Systematic integrated process design and control of binary element reactive distillation processes. *AIChE Journal*, *62*(9), 3137–3154. https://doi.org/10.1002/AIC.15322

Martinez-Orozco, E., Gortares-Moroyoqui, P., Santiago-Olivares, N., Napoles-Armenta, J., Ulloa-Mercado, R. G., de La Mora-Orozco, C., Leyva-Soto, L. A., Alvarez-Valencia, L. H., & Meza-Escalante, E. R. (2020). Tequila Still Distillation Fractioned Residual Streams for Use in Biorefinery. *Energies 2020, Vol. 13, Page 6222*, *13*(23), 6222. https://doi.org/10.3390/EN13236222

Meidanshahi, V., & Adams, T. A. (2016). Integrated design and control of semicontinuous distillation systems utilizing mixed integer dynamic optimization. *Computers & Chemical Engineering*, *89*, 172–183. https://doi.org/10.1016/J.COMPCHEMENG.2016.03.022

Méndez-Acosta, H. O., Campos-Rodríguez, A., González-Álvarez, V., García-Sandoval, J. P., Snell-Castro, R., & Latrille, E. (2016). A hybrid cascade control scheme for the VFA and COD regulation in two-stage anaerobic digestion processes. *Bioresource Technology*, *218*, 1195–1202. https://doi.org/10.1016/J.BIORTECH.2016.07.076

Méndez-Acosta, H. O., Snell-Castro, R., Alcaraz-González, V., González-Álvarez, V., & Pelayo-Ortiz, C. (2010). Anaerobic treatment of Tequila vinasses in a CSTR-type digester. *Biodegradation*, *21*(3), 357–363. https://doi.org/10.1007/S10532-009-9306-7

Mendiola-Rodriguez, T. A., & Ricardez-Sandoval, L. A. (2022). Robust control for anaerobic digestion systems of Tequila vinasses under uncertainty: A Deep Deterministic Policy Gradient Algorithm. *Digital Chemical Engineering*, *3*, 100023. https://doi.org/10.1016/J.DCHE.2022.100023

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T. P., Harley, T., Silver, D., & Kavukcuoglu, K. (2016). Asynchronous Methods for Deep Reinforcement Learning. *33rd International Conference on Machine Learning, ICML 2016*, *4*, 2850–2869. https://arxiv.org/abs/1602.01783v2

Moguel-Castañeda, J. G., Puebla, H., Méndez-Acosta, H. O., & Hernandez-Martinez, E. (2020). Modeling pH and temperature effects on the anaerobic treatment of tequila vinasses. *Journal of Chemical Technology & Biotechnology*, *95*(7), 1953–1961. https://doi.org/10.1002/JCTB.6361

Mohideen, M. J., Perkins, J. D., & Pistikopoulos, E. N. (1996). Optimal synthesis and design of dynamic systems under uncertainty. *Computers & Chemical Engineering*, *20*(SUPPL.2), S895–S900. https://doi.org/10.1016/0098-1354(96)00157-3

Morán, A., Alonso, S., Prada, M. A., Fuertes, J. J., Díaz, I., & Domínguez, M. (2017). Analysis of parallel process in hvac systems using deep autoencoders. *Communications in Computer and Information Science*, *744*, 15–26. https://doi.org/10.1007/978-3-319-65172-9_2/FIGURES/5

Mowbray, M., Zhang, D., Antonio, E., & Rio Chanona, D. (2022). *Distributional Reinforcement Learning for Scheduling of Chemical Production Processes*. https://doi.org/10.48550/arxiv.2203.00636

Neumann, M., & Palkovits, D. S. (2022). Reinforcement Learning Approaches for the Optimization of the Partial Oxidation Reaction of Methane. *Industrial and Engineering Chemistry Research*, *61*(11), 3910–3916. https://doi.org/10.1021/ACS.IECR.1C04622/ASSET/IMAGES/MEDIUM/IE1C04622_0011.GIF

Ngah, S., Abu Bakar, R., Embong, A., & Razali, S. (2016). *TWO-STEPS IMPLEMENTATION OF SIGMOID FUNCTION FOR ARTIFICIAL NEURAL NETWORK IN FIELD PROGRAMMABLE GATE ARRAY*. *11*(7). www.arpnjournals.com

Nguyen, D., Gadhamshetty, V., Nitayavardhana, S., & Khanal, S. K. (2015). Automatic process control in anaerobic digestion technology: A critical review. *Bioresource Technology*, *193*, 513–522. https://doi.org/10.1016/J.BIORTECH.2015.06.080

Oyama, H., & Durand, H. (2020). Interactions between control and process design under economic model predictive control. *Journal of Process Control*, *92*, 1–18. https://doi.org/10.1016/J.JPROCONT.2020.05.009

Palma-Flores, O., & Ricardez-Sandoval, L. A. (2022). Integration of Design and NMPC-based Control for Chemical Processes under Uncertainty: An MPCC-based Framework. *Computers & Chemical Engineering*, 107815. https://doi.org/10.1016/J.COMPCHEMENG.2022.107815

Pan, E., Petsagkourakis, P., Mowbray, M., Zhang, D., & Rio-Chanona, E. A. del. (2021). Constrained model-free reinforcement learning for process optimization. *Computers & Chemical Engineering*, *154*, 107462. https://doi.org/10.1016/J.COMPCHEMENG.2021.107462

Patilas, C. S., & Kookos, I. K. (2021). A novel approach to the simultaneous design & control problem. *Chemical Engineering Science*, *240*, 116637. https://doi.org/10.1016/J.CES.2021.116637

Petsagkourakis, P., Sandoval, I. O., Bradford, E., Zhang, D., & del Rio-Chanona, E. A. (2020a). Constrained Reinforcement Learning for Dynamic Optimization under Uncertainty. *IFAC-PapersOnLine*, *53*(2), 11264–11270. https://doi.org/10.1016/J.IFACOL.2020.12.361

Petsagkourakis, P., Sandoval, I. O., Bradford, E., Zhang, D., & del Rio-Chanona, E. A. (2020b). Reinforcement learning for batch bioprocess optimization. *Computers & Chemical Engineering*, *133*, 106649. https://doi.org/10.1016/J.COMPCHEMENG.2019.106649

Pettigrew, L., & Delgado, A. (2016). *Neural Network Based Reinforcement Learning Control for Increased Methane Production in an Anaerobic Digestion System*. https://www.researchgate.net/publication/304895187

Piceno-Diaz, E. R. (2018). *Optimización y Control Robusto de un Digestor Anaerobio de Dos Etapas para el Procesamiento de Vinazas de Tequila* [Universidad Autonoma

Metropolitana-Unidad Azcapotzalco]. http://zaloamati.azc.uam.mx/bitstream/handle/11191/6107/Optimizacion_y_control_robusto_Piceno_Diaz_E_R_2018.pdf?sequence=1&isAllowed=y

Piceno-Díaz, E. R., Ricardez-Sandoval, L. A., Gutierrez-Limon, M. A., Méndez-Acosta, H. O., & Puebla, H. (2020). Robust Nonlinear Model Predictive Control for Two-Stage Anaerobic Digesters. *Industrial & Engineering Chemistry Research*, *59*(52), 22559–22572. https://doi.org/10.1021/ACS.IECR.0C03809

Poh, P. E., Gouwanda, D., Mohan, Y., Gopalai, A. A., & Tan, H. M. (2016). Optimization of Wastewater Anaerobic Digestion Using Mechanistic and Meta-heuristic Methods: Current Limitations and Future Opportunities. *Water Conservation Science and Engineering 2016 1:1*, *1*(1), 1–20. https://doi.org/10.1007/S41101-016-0001-3

Porru, M., & Özkan, L. (2019). Simultaneous design and control of an industrial two-stage mixed suspension mixed product removal crystallizer. *Journal of Process Control*, *80*, 60–77. https://doi.org/10.1016/J.JPROCONT.2019.04.011

Powell, B. K. M., Machalek, D., & Quah, T. (2020). Real-time optimization using reinforcement learning. *Computers & Chemical Engineering*, *143*, 107077. https://doi.org/10.1016/J.COMPCHEMENG.2020.107077

Quah, T., Machalek, D., & Powell, K. M. (2020). Comparing Reinforcement Learning Methods for Real-Time Optimization of a Chemical Process. *Processes 2020, Vol. 8, Page 1497*, *8*(11), 1497. https://doi.org/10.3390/PR8111497

Rafiei, M., & Ricardez-Sandoval, L. A. (2018). Stochastic Back-Off Approach for Integration of Design and Control under Uncertainty. *Industrial and Engineering Chemistry Research*, *57*(12), 4351–4365. https://doi.org/10.1021/ACS.IECR.7B03935/ASSET/IMAGES/MEDIUM/IE-2017-03935P_0012.GIF

Rafiei, M., & Ricardez-Sandoval, L. A. (2020a). New frontiers, challenges, and opportunities in integration of design and control for enterprise-wide sustainability. *Computers & Chemical Engineering*, *132*, 106610. https://doi.org/10.1016/J.COMPCHEMENG.2019.106610

Rafiei, M., & Ricardez-Sandoval, L. A. (2020b). New frontiers, challenges, and opportunities in integration of design and control for enterprise-wide sustainability. *Computers and Chemical Engineering*, *132*, 106610. https://doi.org/10.1016/J.COMPCHEMENG.2019.106610

Rafiei, M., & Ricardez-Sandoval, L. A. (2020c). Integration of design and control for industrial-scale applications under uncertainty: a trust region approach. *Computers & Chemical Engineering*, *141*, 107006. https://doi.org/10.1016/J.COMPCHEMENG.2020.107006

Rangel-Martinez, D., Nigam, K. D. P., & Ricardez-Sandoval, L. A. (2021). Machine learning on sustainable energy: A review and outlook on renewable energy systems, catalysis, smart grid and energy storage. *Chemical Engineering Research and Design*, *174*, 414–441. https://doi.org/10.1016/J.CHERD.2021.08.013

Ricardez-Sandoval, L. A., Budman, H. M., & Douglas, P. L. (2008). Application of Robust Control Tools to the Simultaneous Design and Control of Dynamic Systems. *Industrial and Engineering Chemistry Research*, *48*(2), 801–813. https://doi.org/10.1021/IE800378Y

Ricardez-Sandoval, L. A., Budman, H. M., & Douglas, P. L. (2009). Simultaneous design and control of chemical processes with application to the Tennessee Eastman process. *Journal of Process Control*, *19*(8), 1377–1391. https://doi.org/10.1016/J.JPROCONT.2009.04.009

Ricardez-Sandoval, L. A., Budman, H. M., & Douglas, P. L. (2010). Simultaneous Design and Control: A New Approach and Comparisons with Existing Methodologies. *Industrial and Engineering Chemistry Research*, *49*(6), 2822–2833. https://doi.org/10.1021/IE9010707

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature 1986 323:6088*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Rummery, G. A., & Niranjan, M. (1994). *ON-LINE Q-LEARNING USING CONNECTIONIST SYSTEMS*.

Sachio, S., del-Rio Chanona, A. E., & Petsagkourakis, P. (2021). Simultaneous Process Design and Control Optimization using Reinforcement Learning. *IFAC-PapersOnLine*, *54*(3), 510–515. https://doi.org/10.1016/J.IFACOL.2021.08.293

Saddoud, A., Hassaïri, I., & Sayadi, S. (2007). Anaerobic membrane reactor with phase separation for the treatment of cheese whey. *Bioresource Technology*, *98*(11), 2102–2108. https://doi.org/10.1016/J.BIORTECH.2006.08.013

Sakizlis, V., Perkins, J. D., & Pistikopoulos, E. N. (2004). Recent advances in optimization-based simultaneous process and control design. *Computers & Chemical Engineering*, *28*(10), 2069–2086. https://doi.org/10.1016/J.COMPCHEMENG.2004.03.018

Sánchez-Sánchez, K., & Ricardez-Sandoval, L. (2013). Simultaneous process synthesis and control design under uncertainty: A worst-case performance approach. *AIChE Journal*, *59*(7), 2497–2514. https://doi.org/10.1002/AIC.14040

Schievano, A., Tenca, A., Lonati, S., Manzini, E., & Adani, F. (2014). Can two-stage instead of one-stage anaerobic digestion really increase energy recovery from biomass? *Applied Energy*, *124*, 335–342. https://doi.org/10.1016/J.APENERGY.2014.03.024

Schulman, J., Levine, S., Moritz, P., Jordan, M., & Abbeel, P. (2015). Trust Region Policy Optimization. *32nd International Conference on Machine Learning, ICML 2015*, *3*, 1889–1897. https://doi.org/10.48550/arxiv.1502.05477

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Openai, O. K. (2017). *Proximal Policy Optimization Algorithms*. https://doi.org/10.48550/arxiv.1707.06347

Schweidtmann, A. M., Esche, E., Fischer, A., Kloft, M., Repke, J. U., Sager, S., & Mitsos, A. (2021). Machine Learning in Chemical Engineering: A Perspective. *Chemie Ingenieur Technik*, *93*(12), 2029–2039. https://doi.org/10.1002/CITE.202100083

Seo, G., Yoon, S., Kim, M., Mun, C., & Hwang, E. (2021). Deep Reinforcement Learning-Based Smart Joint Control Scheme for On/Off Pumping Systems in Wastewater Treatment Plants. *IEEE Access*, *9*, 95360–95371. https://doi.org/10.1109/ACCESS.2021.3094466

Sharifzadeh, M. (2013). Integration of process design and control: A review. *Chemical Engineering Research and Design*, *91*(12), 2515–2549. https://doi.org/10.1016/J.CHERD.2013.05.007

Shi, X., Li, Y., Sun, B., Xu, H., Yang, C., & Zhu, H. (2020). Optimizing zinc electrowinning processes with current switching via Deep Deterministic Policy Gradient learning. *Neurocomputing*, *380*, 190–200. https://doi.org/10.1016/J.NEUCOM.2019.11.022

Shin, J., Badgwell, T. A., Liu, K. H., & Lee, J. H. (2019). Reinforcement Learning – Overview of recent progress and implications for process control. *Computers & Chemical Engineering*, *127*, 282–294. https://doi.org/10.1016/J.COMPCHEMENG.2019.05.029

Sikarwar, V. S., Pohořelý, M., Meers, E., Skoblia, S., Moško, J., & Jeremiáš, M. (2021). Potential of coupling anaerobic digestion with thermochemical technologies for waste valorization. *Fuel*, *294*, 120533. https://doi.org/10.1016/J.FUEL.2021.120533

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature 2016 529:7587*, *529*(7587), 484–489. https://doi.org/10.1038/nature16961

Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., & Riedmiller, M. (2014). *Deterministic Policy Gradient Algorithms* (pp. 387–395). PMLR. https://proceedings.mlr.press/v32/silver14.html

Solera, R., Romero, L. I., & Sales, D. (2001). *The Evolution of Biomass in a Two-phase Anaerobic Treatment Process During Start-up*.

Tang, W., & Daoutidis, P. (2018). Distributed adaptive dynamic programming for data-driven optimal control. *Systems & Control Letters*, *120*, 36–43. https://doi.org/10.1016/J.SYSCONLE.2018.08.002

Tessler, C., Mankowitz, D. J., & Mannor, S. (2018). Reward Constrained Policy Optimization. *7th International Conference on Learning Representations, ICLR 2019*. https://arxiv.org/abs/1805.11074v3

The Chemical Engineering Plant Cost Index - Chemical Engineering. (n.d.). Retrieved May 14, 2022, from https://www.chemengonline.com/pci-home

Tian, Y., Pappas, I., Burnak, B., Katz, J., & Pistikopoulos, E. N. (2021). Simultaneous design & control of a reactive distillation system – A parametric optimization & control approach. *Chemical Engineering Science*, *230*, 116232. https://doi.org/10.1016/J.CES.2020.116232

Toffolo, K., & Ricardez-Sandoval, L. (2021). Optimal Design and Control of a Multiscale Model for a Packed Bed Chemical-Looping Combustion Reactor. *IFAC-PapersOnLine*, *54*(3), 615–620. https://doi.org/10.1016/J.IFACOL.2021.08.310

Trainor, M., Giannakeas, V., Kiss, C., & Ricardez-Sandoval, L. A. (2013). Optimal process and control design under uncertainty: A methodology with robust feasibility and stability analyses. *Chemical Engineering Science*, *104*, 1065–1080. https://doi.org/10.1016/J.CES.2013.10.017

Tu, J. v. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, *49*(11), 1225–1231. https://doi.org/10.1016/S0895-4356(96)00002-9

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the Theory of the Brownian Motion. *Physical Review*, *36*(5), 823. https://doi.org/10.1103/PhysRev.36.823

U.S. Energy Information Administration. (2022). *United States Natural Gas Industrial Price (Dollars per Thousand Cubic Feet)*. https://www.eia.gov/dnav/ng/hist/n3035us3m.htm

Valipour, M., & Ricardez-Sandoval, L. A. (2021a). Assessing the Impact of EKF as the Arrival Cost in the Moving Horizon Estimation under Nonlinear Model Predictive Control. *Industrial and Engineering Chemistry Research*, *60*(7), 2994–3012. https://doi.org/10.1021/ACS.IECR.0C06095/SUPPL_FILE/IE0C06095_SI_001.PDF

Valipour, M., & Ricardez-Sandoval, L. A. (2021b). Constrained Abridged Gaussian Sum Extended Kalman Filter: Constrained Nonlinear Systems with Non-Gaussian Noises and Uncertainties. *Industrial and Engineering Chemistry Research*, *60*(47), 17110–17127. https://doi.org/10.1021/ACS.IECR.1C02804/ASSET/IMAGES/MEDIUM/IE1C02804_0008.GIF

Valipour, M., & Ricardez-Sandoval, L. A. (2022). A robust moving horizon estimation under unknown distributions of process or measurement noises. *Computers & Chemical Engineering*, *157*, 107620. https://doi.org/10.1016/J.COMPCHEMENG.2021.107620

Valipour, M., Toffolo, K. M., & Ricardez-Sandoval, L. A. (2021). State estimation and sensor location for Entrained-Flow Gasification Systems using Kalman Filter. *Control Engineering Practice*, *108*, 104702. https://doi.org/10.1016/J.CONENGPRAC.2020.104702

Vergara-Fernández, A., Vargas, G., Alarcón, N., & Velasco, A. (2008). Evaluation of marine algae as a source of biogas in a two-stage anaerobic reactor system. *Biomass and Bioenergy*, *32*(4), 338–344. https://doi.org/10.1016/J.BIOMBIOE.2007.10.005

Wächter, A., & Biegler, L. T. (2005). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming 2005 106:1*, *106*(1), 25–57. https://doi.org/10.1007/S10107-004-0559-Y

Waschneck, B., Reichstaller, A., Belzner, L., Altenmüller, T., Bauernhansl, T., Knapp, A., & Kyek, A. (2018). Optimization of global production scheduling with deep reinforcement learning. *Procedia CIRP*, *72*, 1264–1269. https://doi.org/10.1016/J.PROCIR.2018.03.212

Wastewater rates :: East Bay Municipal Utility District. (n.d.). Retrieved May 14, 2022, from https://www.ebmud.com/wastewater/rates-and-charges

Watkins, C. J. C. H., & Dayan, P. (1992). *Q-Learning*. *8*, 279–292.

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning 1992 8:3*, *8*(3), 229–256. https://doi.org/10.1007/BF00992696

Yoo, H., Kim, B., Kim, J. W., & Lee, J. H. (2021). Reinforcement learning based optimal control of batch processes using Monte-Carlo deep deterministic policy gradient with phase segmentation. *Computers & Chemical Engineering*, *144*, 107133. https://doi.org/10.1016/J.COMPCHEMENG.2020.107133

Yoo, H., Zavala, V. M., & Lee, J. H. (2021). A Dynamic Penalty Function Approach for Constraint-Handling in Reinforcement Learning. *IFAC-PapersOnLine*, *54*(3), 487–491. https://doi.org/10.1016/J.IFACOL.2021.08.289

Yuan, Z., Chen, B., Sin, G., & Gani, R. (2012). State-of-the-art and progress in the optimization-based simultaneous design and control for chemical processes. *AIChE Journal*, *58*(6), 1640–1659. https://doi.org/10.1002/AIC.13786

Zarate, M. (2013). *Estimación de parámetros y validación del modelo AM2 para un proceso de digestión anaerobia a escala piloto. Master's thesis*.

Zhang, B., Hu, W., Cao, D., Huang, Q., Chen, Z., & Blaabjerg, F. (2019). Deep reinforcement learning–based approach for optimizing energy conversion in integrated electrical and heating system with renewable energy. *Energy Conversion and Management*, *202*, 112199. https://doi.org/10.1016/J.ENCONMAN.2019.112199

Zheng, C., Ji, T., Xie, F., Zhang, X., Zheng, H., & Zheng, Y. (2021). From active learning to deep reinforcement learning: Intelligent active flow control in suppressing vortex-induced vibration. *Physics of Fluids*, *33*(6), 063607. https://doi.org/10.1063/5.0052524

Zhou, Z., Li, X., & Zare, R. N. (2017). Optimizing Chemical Reactions with Deep Reinforcement Learning. *ACS Central Science*, *3*(12), 1337–1344. https://doi.org/10.1021/ACSCENTSCI.7B00492

Zhu, W., Rendall, R., Castillo, I., Wang, Z., Chiang, L. H., Hayot, P., & Romagnoli, J. A. (2021). Control of A Polyol Process Using Reinforcement Learning. *IFAC-PapersOnLine*, *54*(3), 498–503. https://doi.org/10.1016/J.IFACOL.2021.08.291