



UNIVERSIDADE D  
COIMBRA

Miguel Afonso Proença

**INTERPRETABLE DIGITAL TWIN FOR  
INDUSTRIAL PROCESS**

**Dissertation under Master of Science Degree in Electrical and  
Computer Engineering supervised by Professor Doctor Jérôme  
Amaro Pires Mendes, and Professor Doctor Rui Alexandre de  
Matos Araújo, and presented to the Department of Electrical and  
Computer Engineering, Faculty of Science and Technology,  
University of Coimbra.**

February 2024





University of Coimbra  
Faculty of Sciences and Technology  
Department of Electrical and Computer Engineering

# INTERPRETABLE DIGITAL TWIN FOR INDUSTRIAL PROCESS

Miguel Afonso Proença

Dissertation under Master of Science Degree in Electrical and Computer Engineering supervised by Professor Doctor Jérôme Amaro Pires Mendes, and Professor Doctor Rui Alexandre de Matos Araújo, and presented to Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Coimbra.

February 2024



UNIVERSIDADE DE  
COIMBRA





# Agradecimentos

Em primeiro lugar, um agradecimento ao meu orientador, Doutor Jérôme Mendes, por me ter dado a oportunidade de realizar esta dissertação. O seu constante acompanhamento, apoio e os valiosos ensinamentos transmitidos foram fundamentais para a conclusão deste trabalho. Ao orientador, Doutor Rui Araújo, o meu agradecimento pelas críticas e sugestões construtivas.

Esta dissertação é o culminar do meu percurso académico, e para chegar até aqui todo o apoio foi importante. Aos meus amigos de sempre e aqueles que fiz ao longo da minha jornada universitária o meu enorme obrigado por todo o apoio e pelos momentos bem passados.

Aos meus pais, à minha irmã, ao meu cunhado e ao meu sobrinho o meu agradecimento por todo o apoio, motivação e encorajamento ao longo de todo o meu percurso, que agora termina com esta dissertação. Sem eles nada disto tinha sido possível e estou eternamente grato por tudo. Um agradecimento à minha restante família por estarem sempre ao meu lado durante esta caminhada.

Por fim um agradecimento aos meus colegas de laboratório pelo suporte e pela colaboração para superar as dificuldades que foram surgindo.

Esta investigação foi co-financiada pelo Fundo Europeu de Desenvolvimento Regional, através do Programa Operacional Regional do Centro 2014/2020 (Centro2020), do Portugal 2020 no âmbito do Projeto InGestAlgae (CENTRO-01-0247-FEDER-046983).





# Abstract

The digital transformation of industry has gained emphasis in recent years in academia and industry. Organizations need to be more competitive and efficient, improving their processes and performance to cope with changes in environmental legislation, efficient management of resources and energy, and the tendency to zero wastelands. These factors have led to the emergence of a new concept, the Smart Industry, characterized by efficient, flexible, and reliable control of production processes.

This new concept has led to the emergence of new concepts, such as digitalization and Digital Twin (DT), technologies that have undergone a great evolution and that have asserted themselves as technologies capable of improving operational performance, avoiding work accidents, and preventing possible failures that affect the normal performance of the system.

A DT enables cyber-physical integration by creating virtual models for physical objects in a way to simulate their behaviours. With this integration, virtual models can know physical entities' state to predict, estimate and analyze system changes. On the other hand, physical entities can respond to changes according to the simulation. With this cyber-physical integration, the DT can help to improve all manufacturing processes. In this sense, creating DTs to simulate industrial processes is an added value for companies. Moreover, having interpretable DTs allows companies to understand better and use all the data generated by these models, facilitating the decision-making process and detecting anomalies, as well as being fundamental for optimizing processes.

This dissertation presents an approach to perform a DT for industrial processes, which in this dissertation is focused on the Wastewater Treatment Plant (WWTP) waterline. The proposed approach aims to create interpretable DTs, whose models must be easily interpretable by the expert process operators. For this purpose, to create interpretable DT models, different fuzzy-based models were studied, namely a Model based on Fuzzy C-Means and Least Squares Method (FCM-LSM), the Generalized additive models using zero-order T-S fuzzy systems (GAM-ZOTS) model, and the Iterative Learning of multiple univariate zero-order T-S fuzzy systems (iMU-ZOTS) model. For validation of the proposed approach, the Benchmark Simulation Model 2 (BSM2) simulator is used, which simulates the entire wastewater treatment process based on real data. The DT was performed to predict the key variables throughout the waterline of the wastewater treatment process at different points in the plant. Exhaustive validations were performed. The 3 studied models were used to model 16 key variables at 14 different points in the treatment plant. Each model for each key variable for each point was learned/tested with different sets of input variables and using different learning model parameters. These tests were performed to choose the best parameters for the model to obtain the best results for each variable to be estimated. The results showed that the model with the best accuracy was complex and not very interpretable. After analyzing the balance between accuracy and interpretability, a less complex and more interpretable model was found, while maintaining a level of

accuracy comparable to that of the model that obtained the best result.

**Keywords:** Digital Twin, Digitalization, Interpretable Models, Fuzzy Systems, WWTP.

# Resumo

A transformação digital da indústria tem ganho destaque nos últimos anos no meio acadêmico e industrial. As organizações necessitam de ser mais competitivas e eficientes, melhorando os seus processos e o seu desempenho para fazer face às alterações na legislação ambiental, à gestão eficiente dos recursos e de energia, e à tendência para zero desperdícios. Estes fatores levaram ao surgimento de um novo conceito, a *Smart Industry*, caracterizada pelo controlo eficiente, flexível e fiável dos processos de produção.

Este novo conceito levou ao aparecimento de novos conceitos, como a digitalização e o *Digital Twin (DT)*, tecnologias que têm sofrido uma grande evolução e que se têm afirmado como tecnologias capazes de melhorar o desempenho operacional, evitar acidentes de trabalho e prevenir possíveis falhas que afetem o normal desempenho do sistema.

O DT permite a integração ciber-física através da criação de modelos virtuais para objetos físicos, de forma a simular os seus comportamentos. Com esta integração, os modelos virtuais podem conhecer o estado das entidades físicas para prever, estimar e analisar as alterações do sistema. Por outro lado, as entidades físicas podem responder às alterações de acordo com a simulação. Com esta integração ciber-física, o DT pode ajudar a melhorar todos os processos de produção. Neste sentido, a criação de DTs para simular os processos industriais é uma mais-valia para as empresas. Além disso, ter DTs interpretáveis permite às empresas compreender e utilizar melhor todos os dados gerados por estes modelos, facilitando o processo de tomada de decisão e a deteção de anomalias, além de ser fundamental para a otimização dos processos.

Esta dissertação apresenta uma abordagem para realizar um DT para processos industriais, que nesta dissertação é focada na linha de água de uma Estação de Tratamento de Águas Residuais (ETAR). A abordagem proposta tem como objetivo a criação de DTs interpretáveis, cujos modelos devem ser facilmente interpretáveis pelos operadores especializados no processo. Para este efeito, para a criar modelos de DTs interpretáveis, foram estudados diferentes modelos baseados em fuzzy, nomeadamente o *Model based on Fuzzy C-Means and Least Squares Method (FCM-LSM)*, o modelo *Generalized additive models using zero-order T-S fuzzy systems (GAM-ZOTS)* e o modelo *Iterative Learning of multiple univariate zero-order T-S fuzzy systems (iMU-ZOTS)*. Para a validação da abordagem proposta é utilizado o simulador *Benchmark Simulation Model 2 (BSM2)*, que simula todo o processo de tratamento de águas residuais com base em dados reais. O DT foi utilizado para prever as variáveis-chave ao longo da linha de água do processo de tratamento de águas residuais em diferentes pontos da estação. Foram efetuadas validações exaustivas. Os 3 modelos estudados foram utilizados para modelar 16 variáveis-chave em 14 pontos diferentes da estação de tratamento. Cada modelo para cada variável-chave para cada ponto foi aprendido/testado com diferentes conjuntos de variáveis de entrada e utilizando diferentes parâmetros do modelo de aprendizagem. Estes testes foram efetuados para escolher os melhores parâmetros do modelo para obter os melhores resultados para cada variável a estimar. Os resultados mostraram que o modelo com a

melhor exatidão era complexo e pouco interpretável. Após analisar o equilíbrio entre a exatidão e a interpretabilidade, constatou-se a existência de um modelo menos complexo e mais interpretável, mantendo um nível de exatidão comparável ao do modelo que obteve o melhor resultado.

**Palavras-chave:** Gémeo Digital, Digitalização, Modelos Interpretáveis, Sistemas Fuzzy, ETAR.

# Contents

<b>List of Acronyms</b>	<b>ix</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation and Scope . . . . .	1
1.1.1 Wastewater Treatment Plants . . . . .	2
1.2 Objectives and Contributions . . . . .	3
1.3 Structure . . . . .	4
<b>2 Background and Literature Review</b>	<b>5</b>
2.1 Digital Twin . . . . .	5
2.1.1 Definition of Digital Twin . . . . .	5
2.1.2 History of Digital Twin . . . . .	6
2.1.3 Potential of Digital Twin . . . . .	7
2.2 Five-Dimension Digital Twin Model . . . . .	8
2.3 Digital Twin Applications . . . . .	10
2.3.1 Implementation Areas of Digital Twin . . . . .	10
2.3.2 Industrial Applications of Digital Twin . . . . .	11
2.4 Conclusion . . . . .	12
<b>3 Wastewater Treatment Plant Framework</b>	<b>15</b>
3.1 Wastewater Treatment Plants . . . . .	15
3.2 Benchmark Simulation Model 2 . . . . .	16
3.2.1 Activate Sludge Model N°1 - (ASM1) . . . . .	17
3.2.2 Anaerobic Digester Model N°1 - (ADM1) . . . . .	18
3.2.3 BSM2 Variables . . . . .	18
3.3 Digital Twin Framework on BSM2 . . . . .	19
3.4 Conclusion . . . . .	20
<b>4 Digital Twin Methodologies</b>	<b>23</b>
4.1 Fuzzy Logic Systems . . . . .	23
4.1.1 Takagi-Sugeno Fuzzy Model . . . . .	24
4.2 Model based on Fuzzy C-Means and Least Squares Method . . . . .	25
4.2.1 Fuzzy C-Means Method . . . . .	25
4.2.2 Least Squares Method . . . . .	26
4.2.3 Model Design Algorithm . . . . .	27

---

4.3	GAM-ZOTS - Generalized Additive Models Using Zero-Order T-S Fuzzy Systems . . . . .	28
4.3.1	GAM-ZOTS Model . . . . .	28
4.4	iMU-ZOTS - Iterative Learning of Multiple Univariate Zero-Order T-S Fuzzy Systems . . . . .	29
4.4.1	iMU-ZOTS Model . . . . .	31
<b>5</b>	<b>Results and Discussion</b>	<b>33</b>
5.1	Experimental Tests . . . . .	33
5.2	Results . . . . .	36
<b>6</b>	<b>Conclusions and Future Work</b>	<b>49</b>
	<b>Appendix A Submitted Paper</b>	<b>59</b>
	<b>Appendix B Fuzzy rules from iMU-ZOTS for <math>S_{NH_3}</math></b>	<b>75</b>
	<b>Appendix C Fuzzy rules from iMU-ZOTS for <math>S_{NO_3}</math></b>	<b>79</b>
	<b>Appendix D Licenses</b>	<b>83</b>

# List of Acronyms

**ADM1** Anaerobic Digester Model 1.

**AI** Artificial Intelligence.

**ASM1** Activate Sludge Model 1.

**BSM1** Benchmark Simulation Model 1.

**BSM2** Benchmark Simulation Model 2.

**DT** Digital Twin.

**EDT** Experimental Digital Twins.

**FCM** Fuzzy C-Means.

**FCM-LSM** Model based on Fuzzy C-Means and Least Squares Method.

**FLS** Fuzzy Logic Systems.

**GAM-ZOTS** Generalized additive models using zero-order T-S fuzzy systems.

**iMU-ZOTS** Iterative Learning of multiple univariate zero-order T-S fuzzy systems.

**IoT** Internet of Things.

**IWA** International Water Association.

**LSM** Least Squares Method.

**MF** Membership Function.

**MSE** Mean Squared Error.

**NASA** National Aeronautics and Space Administration.

**PHM** Prognostic Health Management.

**T-S** Takagi-Sugeno.

**WWTP** Wastewater Treatment Plant.





# List of Figures

2.1	DT framework. Taken from [Mohr, 2018]. . . . .	6
2.2	The landmarks of DT development. Taken from [Qi et al., 2021]. . . . .	7
2.3	Five-dimension DT model. Taken from [Qi et al., 2021]. . . . .	9
2.4	Different applications fields of DT. Taken from [Qi et al., 2021]. . . . .	10
3.1	Schematic of BSM2. Taken from [Salles et al., 2022b]. . . . .	17
3.2	Schematic of Benchmark Simulation Model 1 (BSM1). Taken from [Salles et al., 2022b]. . . . .	17
4.1	Basic configuration of Fuzzy Logic Systems (FLS). Taken from [Júnior et al., 2023]. . . . .	24
4.2	Multiple zero-order T-S fuzzy systems structure. Taken from [Mendes et al., 2019a]. . . . .	29
4.3	Membership functions for $x_j$ . Taken from [Mendes et al., 2019a]. . . . .	30
5.1	BSM2 data collection points. . . . .	34
5.2	Prediction results for the variable $S_{NH_3}$ . . . . .	39
5.3	Membership functions of the Model based on Fuzzy C-Means and Least Squares Method (FCM-LSM) for the variable $S_{NH_3}$ . . . . .	40
5.4	Membership functions of the Generalized additive models using zero-order T-S fuzzy systems (GAM-ZOTS) method for the variable $S_{NH_3}$ . . . . .	40
5.5	Membership functions of the Iterative Learning of multiple univariate zero-order T-S fuzzy systems (iMU-ZOTS) method for the variable $S_{NH_3}$ . . . . .	42
5.6	Membership functions of the iMU-ZOTS method for the variable $S_{NH_3}$ . . . . .	43
5.7	Prediction results for the variable $S_{NO_3}$ . . . . .	44
5.8	Membership functions of the FCM-LSM for the variable $S_{NO_3}$ . . . . .	44
5.9	Membership functions of the GAM-ZOTS method for the variable $S_{NO_3}$ . . . . .	45
5.10	Membership functions of the iMU-ZOTS method for the variable $S_{NO_3}$ . . . . .	46
5.11	Membership functions of the iMU-ZOTS method for the variable $S_{NO_3}$ . . . . .	47



# List of Tables

5.1	Results of the tests carried out by the 3 models. <b>Var.</b> represents the target variable, $n$ is the number of input variables selected and $thr$ is the threshold value. The best result for each variable is shown in bold and the best result of the other models is shown in underlined. . . . .	37
5.2	Fuzzy rule values for the 3 models. <b>Var.</b> represents the target variable, $thr$ the threshold value, $N$ the number of fuzzy rules, Mean rules the average value of the fuzzy rules and $M_{th}$ the threshold value of the iMU-ZOTS model. The value of the rules for each variable for which the best result was obtained is shown in bold and the best value of the other models is shown in underlined. . . . .	38

# Chapter 1

## Introduction

### 1.1 Motivation and Scope

The digitization process is understood as improving the industrial processes by transforming data into a digital format and introducing new digital technologies, which makes possible a paradigm shift in the industry [Parviainen et al., 2017]. With the advancement of new technologies and the digitization of all processes, in particular, the emergence of a new generation of information technologies, like the Internet of Things (IoT), cloud computing, big data analytics, and Artificial Intelligence (AI), the process of digitization is growing very fast [Botkina et al., 2018]. Digitization is becoming a key driver of innovation in many sectors by merging the physical and virtual worlds.

With the industry's evolution, industrial processes increasingly require efficient, flexible, and reliable control [Misra et al., 2022]. This new trend has led to the emergence of the concept of Industry 4.0, popularized by Klaus Schwab [Schwab, 2016]. This new paradigm has led to the appearance of new concepts, such as the Digital Twin (DT), a virtual representation of a physical object or process that can simulate its real-world behaviour and predict its performance.

With the evolution of technology and industrial processes, a DT can be a competitive advantage for any company. The capability of visualization in real-time of the state of the components of a factory combined with the power to simulate various scenarios makes the DT a powerful tool. The capability to simulate hypothetic scenarios allows tests in different environments and simulations of different system changes, allowing a better understanding of the status of physical entities. The constant monitoring of the performance of the process can allow for an increase in product quality and the reduction of faults and accidents, thus reducing maintenance and repair costs. A DT is a major player in the industrial revolution and the digitization of industrial processes, and it is viewed with great interest by manufacturers.

Industrial processes have undergone major changes in recent decades due to increasing worldwide competition (market globalization) and environmental legislation that has been severely tightened concerning the consumption and degradation of natural resources [Mendes et al., 2018]. The changes in the market have led to more and more restrictions to reduce the consumption of energy and materials, and environmental restrictions have made the processes more complex and more costly [Gonçalves et al., 2022]. Many industrial systems have nonlinear behaviour and often have complex characteristics such as unknown and time-varying dynamics, constraints and disturbances. Thus, it is important to develop efficient systems to deal with nonlinear problems with a high degree of uncertainty. It is in

this sense that fuzzy systems arise to help solve these problems. Fuzzy systems are powerful mathematical tools that can be used to model nonlinear problems with uncertainty [Júnior et al., 2023]. They use fuzzy logic to model uncertainty and imprecision, allowing decision-making based on a subjective representation of reality and are easier to understand and explain since, although they are based on a mathematical representation and formulation, they are described by semantic rules based on natural language.

### 1.1.1 Wastewater Treatment Plants

Water is a vital resource for human life, and has a key role in many human activities, like industry, agriculture, and services. But, this important resource is limited [Salles et al., 2022a]. Planet Earth is covered by about 70% of water, however, it is estimated that about 97% is salt water, which is found in seas and oceans and is unsuitable for human consumption, and about 2% forms inaccessible glaciers. Therefore, only 1% is used for human and animal consumption [Júnior et al., 2021]. With the increase in world population, the supply of quality water for human activities is increasingly challenging. With the increase in the world's population, the supply of quality water for human activities is becoming increasingly difficult as there is greater water consumption and greater wastewater production. The causes of water pollution are many: residential and industrial sewage, fertilizers and pesticides used in agriculture, mining activities, etc [Boretti and Rosa, 2019]. Domestic, agricultural or industrial activities produce effluents containing undesirable pollutants that can be toxic.

Considering the importance of water for human beings and the worrying scarcity of this resource, it is essential to use water resources efficiently and to treat wastewater so that it becomes available again and does not pose a threat to the environment and, consequently, to human health. Thus, Wastewater Treatment Plant (WWTP) emerge as a very important structure to face the new water and wastewater challenges. The wastewater treatment is carried out in WWTPs whose main objective is to speed up the treatment process that takes place in nature and thus remove pollutants from the wastewater.

WWTPs are important structures and can be found in large numbers around the world. In the United States, there are more than 16,000 WWTPs in operation [CARD, 2021]. In the European Union, there are more than 18,000 WWTPs, covering more than 450 million people in all member states [EurEau, 2017]. WWTPs are quite complex and large structures, formed by several interconnected processes. The treatment carried out at WWTPs consists of several phases, with physical, chemical and biological processes, and is subject to climatic events such as periods of drought, rain and storms [de Matos et al., 2023]. Moreover, WWTPs are energy-intensive structures, with 7% of all energy consumed worldwide being consumed by WWTPs [Liu et al., 2012]. In Portugal, around 4% of energy is spent on water distribution and treatment systems, and of this total, 25% is consumed by WWTPs [Oliveira et al., 2021]. Nowadays, the need for efficiency and environmental protection is becoming greater and greater, and the demands placed on WWTPs are increasing. WWTPs have to achieve at least three sustainability goals: social acceptance (human health protection), environmental protection (low pollution load) and economic development (viable operating costs) [Garrido-Baserba et al., 2014, 2015].

Over the years, it was found that the existing models for control and monitoring of WWTPs were not sufficient to keep up with advances in technology and environmental constraints. It is therefore necessary to create systems capable of anticipating problems

that help in the decision-making process. Due to the nonlinearity of the WWTPs, it is complicated to create models that meet the new demands of these structures. With the new technological advances, it is possible to develop new models that can respond to these new challenges. DTs are a promising option in this context, allowing for a dynamic and comprehensive representation of the WWTPs.

Motivated by these problems, this dissertation focuses on creating an interpretable DT framework for wastewater treatment process monitoring. With the help of data-driven fuzzy-based models, the aim is to predict the most important variables in the treatment process.

## 1.2 Objectives and Contributions

Creating an interpretable DT of a WWTP allows the treatment process to be monitored, controlled and optimized. By modelling the key variables, it is possible to understand the behaviour of the system, avoid failures and help with decision-making. The main objective of this dissertation is the creation of an interpretable DT of a WWTP waterline.

To achieve the objective, this dissertation consisted of the following sub-objectives and contributions:

- Literature review on DTs and the different applications they can have in various industrial processes.
- Study of different fuzzy-based models with an interpretable structure to model the key variables at the different points of the waterline of a WWTP. The following models were studied: the Model based on Fuzzy C-Means and Least Squares Method (FCM-LSM) [Rastegar et al., 2017], the Generalized additive models using zero-order T-S fuzzy systems (GAM-ZOTS) model [Mendes et al., 2019a], and the Iterative Learning of multiple univariate zero-order T-S fuzzy systems (iMU-ZOTS) model [Mendes et al., 2019b].
- Validation of the proposed approach using the Benchmark Simulation Model 2 (BSM2), which is a well-known simulator based on real data that simulates the entire wastewater treatment process.
- The key variables at different stages of the waterline of the treatment plant were defined as case studies and an exhaustive validation was performed. The following factors were analyzed on the studied models: the impact of the selected variables and the learning parameters on the model accuracy, and the accuracy versus interpretability tradeoff.
- Exhaustive validation of the models studied. The 3 models were used to model 16 key variables at 14 different points in the treatment plant. Each model for each key variable for each point was learned/tested with 4 different sets of input variables (from the variable selection method) and using different learning parameters. Thus, 520 tests were performed to choose the best parameters for the model to obtain the best model parameters to obtain the best results for each variable to be estimated.
- The development of this work resulted in the submission of a journal article. See Appendix A.

## 1.3 Structure

The structure of the document is divided into 6 chapters. The organization of the information in the remaining chapters will be structured as follows:

- Chapter 2 presents the literature review about DTs focused on industrial processes.
- Chapter 3 is dedicated to describing the WWTPs and the BSM2.
- Chapter 4 describes the fuzzy-based models used in this dissertation.
- Chapter 5 presents the achieved results and their analysis.
- Finally, the main conclusions are highlighted in Chapter 6.

# Chapter 2

## Background and Literature Review

This chapter presents the concepts for the work developed and the literature review. Section 2.1 describes the concept and the history of the Digital Twin (DT). Section 2.2 describes the Five-dimension DT Model. Finally, section 2.3 explains the different applications of the DT.

### 2.1 Digital Twin

#### 2.1.1 Definition of Digital Twin

In recent years, the DT has been used more and more, but no consensus has yet been reached on what its definition is so that it can be defined in various ways [Singh et al., 2021]. It can be described as a virtual model of a product whose purpose is to reveal any flaws of the product in its early development phase and to continuously update it. It is also often defined as a digital model designed to represent a physical object by replicating it virtually. The literature clearly shows that DT differs from computer modelling, such as computer-aided design (CAD) and computer-aided engineering (CAE) [Wanasinghe et al., 2020]. Although many organizations use the term “Digital Twin” as a synonym for 3D model, a 3D model is only one part of DT [LaGrange, 2019].

Some of the definitions of DT found in the literature are as follows:

- “digital copy of the physical systems and act as a connection between physics and digital world” [Nadhan et al., 2018];
- “virtual and simulated model or a true replica of a physical asset” [Poddar, 2018];
- “an immersive data analytic technology that provides insights on human-infrastructure-machine interactions to enable executives to make contextual decisions” [Mohr, 2018];
- “refers to the digital footprint of the physical systems in the various assets which act like a bridge between physics and digital world” [Mayani et al., 2018];
- “a virtual physics and data-based model of a system or an asset that models all the various subsystems, their properties, the interaction among them and the interactions of the system with the environment” [Saini et al., 2018];
- “a virtual model of a physical asset” [Sharma et al., 2018];



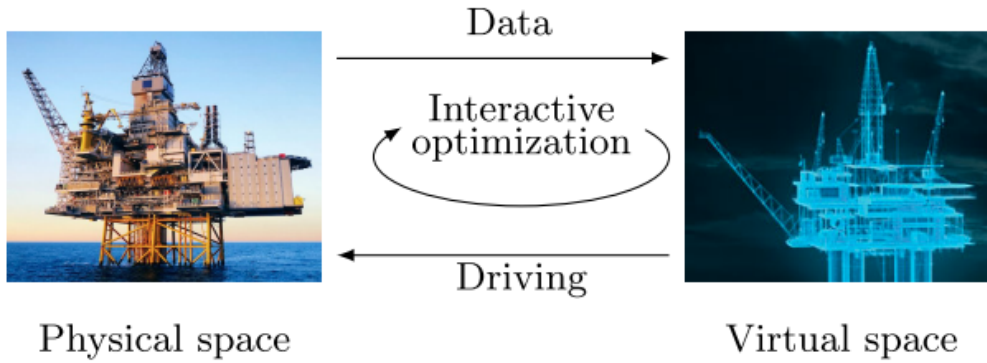


Figure 2.1: DT framework. Taken from [Mohr, 2018].

- “multiphysics, multiscale, probabilistic, ultra fidelity simulation that reflects the state of a corresponding twin based on the historical data, real-time sensor data, and physical model” [Glaessgen and Stargel, 2012].

Although these definitions are somewhat different from each other, they all have in common that they refer to a physical asset, a virtual model, an exchange of data between a physical asset and a virtual model, and data analysis and visualization.

Various frameworks have been developed for the DT, but the one that is most widely accepted and used is the one that includes three main components, the physical space, the virtual space and the links between these spaces, as can be seen in Figure 2.1. The physical space is made up of the physical assets, sensors and actuators. In contrast, the virtual space is made up of simulation models that gather and analyze the data to carry out the simulations necessary to help in optimal control of the physical asset. The links between the physical and virtual spaces allow for the continuous exchange of information between these two spaces.

### 2.1.2 History of Digital Twin

The growing complexity of production in an increasingly demanding market is driving the integration of the physical and digital worlds. At the same time, the growing practical needs of human beings about industrial products are challenging the ability of the digital model to interact with the physical object. The DT was conceived in this context and has triggered a far-reaching industrial revolution [Tao et al., 2018a, Tao and Qi, 2019].

Although DT has gained popularity in recent years, the concept is not new. The concept of DT was introduced by Grieves in 2003, in his course “Product Lifecycle Management” at the University of Michigan, and is considered the origin of DTs [Grieves, 2014]. At that time, this concept was not very specific, and a preliminary form of the DT was proposed to include three parts: physical product, virtual product, and their connections. Initially, the DT was used mainly in the aerospace and military sectors. Nowadays, the DT is undergoing a period of rapid development with its use in various areas of knowledge [Tao et al., 2018b]. Between 2003 and 2011, the development of communication technology, of Internet of Things (IoT) sensor technology, big data analytics, and simulation technologies led to a huge increase in DTs. In 2011, the first journal article was published, explaining how DTs were useful for predicting the structural life of aircraft [Tuegel et al., 2011]. In 2012, the idea of DTs was revisited by the National Aeronautics and Space Administration (NASA)

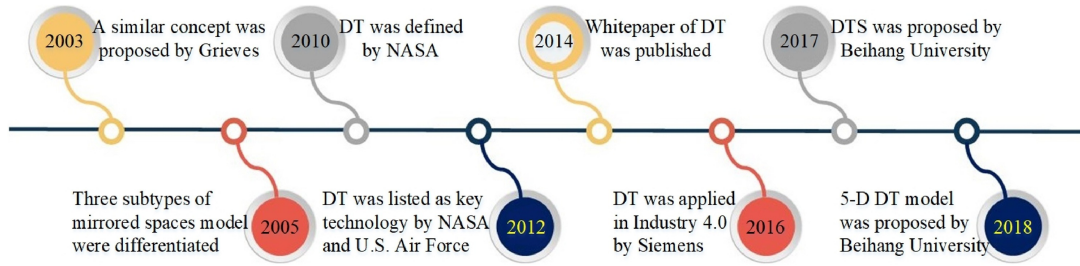


Figure 2.2: The landmarks of DT development. Taken from [Qi et al., 2021].

which defined DT as multiphysics, multiscale, probabilistic, ultra fidelity simulation that reflects the state of a corresponding twin based on the historical data, real-time sensor data, and physical model [Glaessgen and Stargel, 2012]. In 2014, Grieves published a white paper about DT, and according to it, the basic DT model consists of three main parts: (a) physical products in Real Space, (b) virtual products in Virtual Space, as well (c) the connections of data and information that tie the virtual and real products together [Grieves, 2014]. More and more efforts have been devoted to DT research since then. Gartner, in 2017 and 2018, classified DTs as one of the top ten most promising technological trends in the next decade [Rasheed et al., 2019]. Some milestones in the development of DT are presented in Figure 2.2.

The three-dimensional model of DT proposed by Grieves was the one that received the most attention from the entire community, but due to the growing expansion and evolution of application needs, the development and applications of the DT present more and more challenges. In recent years, DT applications have gradually expanded from the military and aerospace sectors to the civilian sector. With the expansion into more sectors, the DT has an increased demand for services from different areas, users and businesses.

Thus, based on the model initially proposed by Grieves, Tao et al. [2019a] proposed the five-dimension model to promote the continuation of DT applications in more areas. This model consists of 5 parts: physical entities, virtual models, DT data, services in the DT, and connections in the DT.

### 2.1.3 Potential of Digital Twin

According to Rasheed et al. [2019], eight assets of the DT can be defined:

1. **Real-time remote monitoring and control:** It is usually difficult to get a real-time overview of a large system. However, a DT offers the advantage of being accessible anywhere, allowing the system to be continuously observed and managed remotely.
2. **Greater efficiency and safety:** The merging of real life and the virtual world allows for greater autonomy, with human intervention where necessary. This will mean that dangerous, boring, repetitive, and dirty jobs can be done by robots, controlled by humans from a distance. This will allow humans to concentrate on more creative and innovative work.
3. **Predictive maintenance and scheduling:** Comprehensive digital twinning will ensure that multiple sensors monitoring physical assets generate large volumes of data in real-time. Intelligent data analysis allows system failures to be detected well in advance. This allows for better maintenance scheduling, avoiding failures.

4. **Scenario and risk assessment:** A DT makes it possible to carry out hypothetical analyses that allow for a better assessment of risks. This makes it possible to introduce changes or disturbances into the system to simulate unexpected scenarios and observe the system's reaction. This type of analysis avoids any risk to the actual physical asset and can only be carried out using a DT.
5. **Better intra and inter-team synergy and collaborations:** Greater autonomy and easy access to information allow teams to optimize their time, promoting better cooperation and synergy, resulting in increased productivity.
6. **More efficient and informed decision support system:** The presence of quantitative data and advanced analytical capabilities in real-time will contribute to more informed and faster decision-making processes.
7. **Personalization of products and services:** Given the comprehensive historical data, knowledge about stakeholder preferences, and the ever-changing dynamics of market trends and competition, demand for products and services customized is expected to increase. So in the context of factories of the future, a DT will facilitate faster and more continuous adjustments to meet evolving requirements.
8. **Better documentation and communication:** The information available on the spot and in real-time, together with automated reporting, allows stakeholders to be well informed at all times, thus improving transparency.

## 2.2 Five-Dimension Digital Twin Model

The formulation of the five-dimensional DT model can be expressed by [Qi et al., 2021]:

$$M_{DT} = F(PE, VM, S_s, DD, CN), \quad (2.1)$$

which it is a function  $F(\cdot)$  of physical entities  $PE$ , virtual models  $VM$ , services  $S_s$ , DT data  $DD$ , and connections  $CN$ . According to this formulation, the DT model in five dimensions is illustrated in Figure 2.3.

The five parts of the five-dimensional model are as follows:

**Physical entities in Digital Twin (PE):** DT consists of creating virtual models of physical entities in a way to simulate all their behaviours [Tao et al., 2018b]. The physical world is the basis of DT and can be a device, a product, a process, or even an organization. Physical entities governed by physical laws and dealing with uncertain environments can be divided into three levels according to their structure and function: the system level, the system of system level (SoS), and the unit level [Tao et al., 2019b].

**Virtual Models in Digital Twin (VM):** Virtual models must represent with maximum accuracy physical entities by replicating their geometries, properties, behaviours, and corresponding physical rules [Tao and Zhang, 2017]. Three-dimensional geometric models describe the appearance, size, tolerance, and structural relationship of a physical entity. Taking into account the physical properties such as velocity, wear, and force, the physical model reflects the physical phenomena occurring in the entities such as deformation, delamination, fracture, and corrosion. The behaviour model describes how entities behave and respond to changes in the external environment, including state transitions, performance, and degradation. Rule models empower DT with logical skills

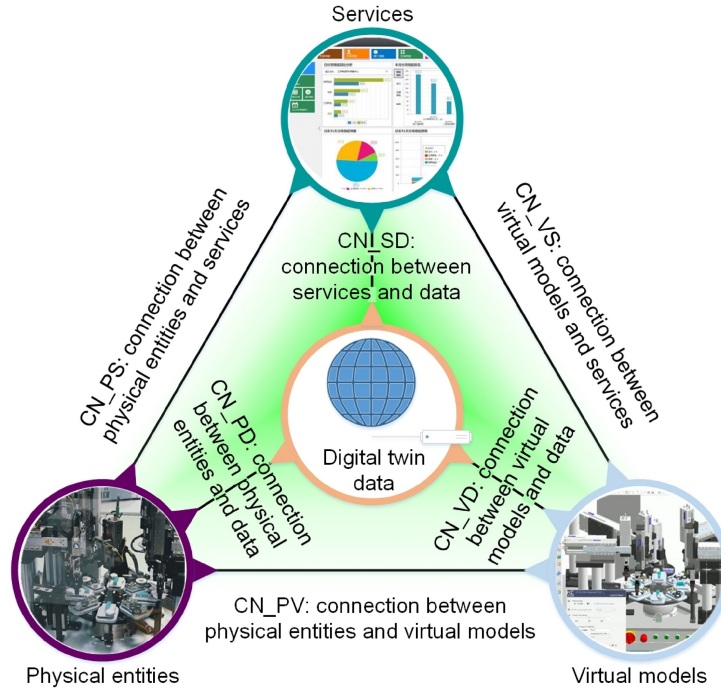


Figure 2.3: Five-dimension DT model. Taken from [Qi et al., 2021].

such as reasoning, judgment, evaluation, and autonomous decision-making by following rules derived from historical data or expert knowledge from specialists in the field.

**Digital Twin data (DD):** Data is the main driver of the DT [Qi et al., 2018]. DT deals with information from various sources at various time scales and dimensions, presenting great heterogeneity. These data can be obtained from physical entities and include both static attributes and dynamic conditions. In addition, data are generated through virtual models, which represent the result of simulations. Data from services, which describe the invocation and execution of these services, are also utilized. Additionally, knowledge provided by domain experts or extracted from existing data is incorporated. Finally, all these data types are merged, resulting in the fusion data.

**Services in Digital Twin ( $S_s$ ):** Given the growing product-service integration in today's society, companies are increasingly recognizing the importance of service [Tao and Qi, 2017]. Service plays a key role in the context of DTs, aligning with the Everything-as-a-Service (XaaS) paradigm. Firstly, services offer users a wide range of application services, including simulation, verification, monitoring, optimization, diagnosis, prognosis, and Prognostic Health Management (PHM), among others. Next, during the process of building a functional DT, various third-party services are required, which can include data services, knowledge services, algorithm services, etc. Finally, the ongoing operation of the DT depends on the support of various platform services, which can be used to develop customized software, create models, and provide services.

**Connections in Digital Twin (CN):** Digital representations act as dynamic links to their real-world counterparts, facilitating simulation, operation, and advanced analyses. These links between physical entities, virtual models, services and data play a fundamental role in enabling the exchange of information and data. In DT, there are six essential types of connections, which are: linking physical entities to virtual models (CN\_PV), the link between physical entities and data (CN\_PD), the link between physical entities and services (CN\_PS), the link between virtual models and data (CN\_VD), the link between

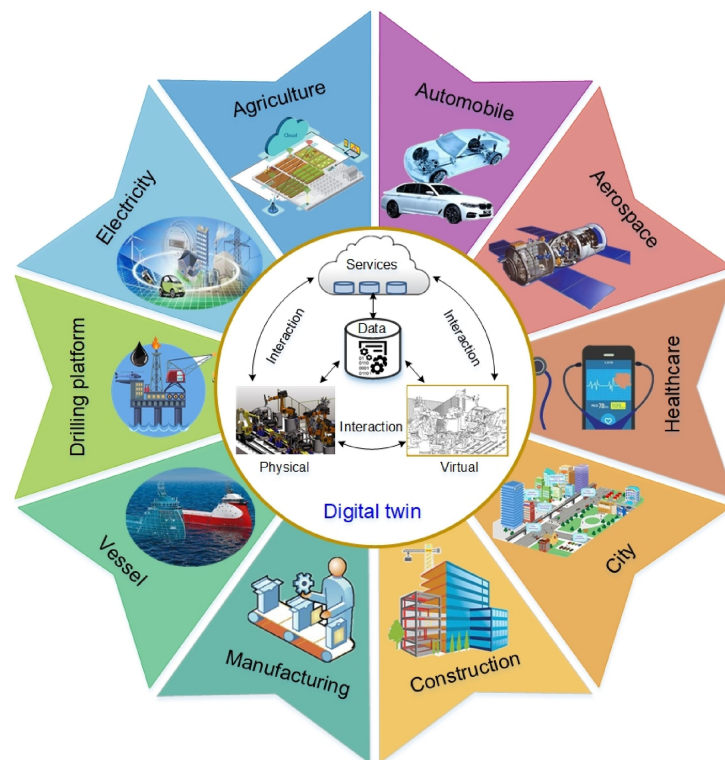


Figure 2.4: Different applications fields of DT. Taken from [Qi et al., 2021].

virtual models and services (CN\_VS), and the link between services and data (CN\_SD) [Tao et al., 2018b]. These connections enable collaboration between the four parties involved.

## 2.3 Digital Twin Applications

This section will summarise the applications of DT in different areas (Subsection 2.3.1) and the applications of DT in the industry (Subsection 2.3.2).

### 2.3.1 Implementation Areas of Digital Twin

The integration of technologies such as cloud computing, mobile Internet, and the analysis of large volumes of data has extended the applicability of DTs to various domains, where physical and virtual domains intersect and evolve together. As illustrated in Figure 2.4, DTs have applications in various sectors, such as construction, healthcare, agriculture, transport, manufacturing, smart cities, electricity, automobile, and aerospace, among many others [Tao et al., 2019c]. Since this technology is relatively new, its application was initially made by large market-leading companies such as General Electric (GE), PTC, Siemens, ANSYS, and Dassault [Schleich et al., 2017, Zheng et al., 2019].

In the field of civil engineering, Dassault used a 3D experience platform to create a “Digital Twin of Singapore”. This DT is a very valuable resource, as it helps with urban planning, construction projects, and the development of new buildings [Systèmes, 2018]. Intellectsoft is exploring the potential applications of DTs on construction sites. The aim is to use DTs to identify potential problems and prevent dangerous operations, thereby increasing the safety and efficiency of construction processes.

In healthcare, Sim&Cure has developed a patient-based DT for the treatment of aneurysms [Marcal, 2018]. Dassault realized a “Living Heart Project (LHP)” for the creation of a human heart [Erol et al., 2020].

As described in Microsoft’s white paper, DTs have the power to significantly speed up farming operations and promote sustainability in the agricultural sector [Microsoft, 2019]. DNV GL has created a “virtual sister ship”, a DT of a ship. This innovation aims to increase reliability, reduce operating costs, and improve safety throughout the ship’s lifecycle [GL, 2019]. In addition, DT was developed for a drilling platform associated with the Blue Whale #1 project in China. This facilitates visualization, operational monitoring, and training in platform design.

In the automotive sector, Tesla company has been working on the implementation of DTs for its electric vehicles, allowing a continuous exchange of data between the vehicles and the factories [Schleich et al., 2017]. LlamaZOO used DT to allow mine supervisors to monitor their operator’s vehicles [McCannel, 2018].

In the aviation sector, Airbus, Boeing, the Air Force Research Laboratory (AFRL), and NASA have adopted DTs to replicate real-world conditions. They use DT to detect defects, predict potential problems, and help meet the challenges of maintaining aircraft structures [Zheng et al., 2019].

General Electric (GE), with the help of the Predix platform, has developed a digital wind farm, creating a DT for each wind turbine. This innovation aims to optimize maintenance strategies, improve reliability, and increase energy production [Energy, 2016].

### 2.3.2 Industrial Applications of Digital Twin

Industrial applications of DTs are mainly concentrated in design, production, PHM, among others, where DTs show a clear advantage over traditional solutions [Tao et al., 2018b].

DTs can be applied to product design to create new products more quickly, efficiently and based on more complete and up-to-date information. Zhuang et al. [2017], investigated the application of DTs in product design and proposed some theories and techniques for implementing this technology in product design. Canedo [2016], considered DTs as a new approach to managing the IoT, using as arguments the fact that product design can be improved by incorporating data feedback from DTs. Yu et al. [2017], proposed a new DT model to manage the three-dimensional configuration of products. With this new model, they believe that the application of DTs in product design can improve the link between the areas of design and production. Tao et al. [2018b], proposed a DT-driven design framework because most design decisions are made without adequate interactions between the expected, interpreted and external spaces. They pointed out some applications of DTs in the different stages of design, such as product planning, conceptual design and detailed design. A case study on bicycle design was carried out to demonstrate their proposal. Schleich et al. [2017], proposed a new DT model to manage geometric variations. They argued that DT allowed designers to assess the quality of a product at the early stage of design. Zhang et al. [2017a], proposed a DT-based technique to design production lines. To study this technique, a glass production line was used to validate the effectiveness of this approach.

DTs can also be used in production, making processes more reliable, flexible and predictable. They can also make it possible to visualize and update the status of products in real-time, making it easy to monitor the entire production process. Rosen et al. [2015],



explored the application of DTs in production operations. They emphasized that DTs can incorporate different types of data, such as environmental conditions, operational information and process data. This makes it possible for autonomous systems to easily adapt to changes in their environment or operating conditions, even during the production process. Brenner and Hummel [2017], examined the hardware and software requirements needed to implement DT in the Logistics Learning Factory at the European School of Business (ESB) to create fluid interactions between humans, machines and products. Tao and Zhang [2017], developed the DT of a shop floor, which included the physical shop floor, the virtual shop floor, the shop floor service system and production data. In addition, they imagined how DTs could serve smart manufacturing.

DTs can also facilitate control. Uhlemann et al. [2017], proposed a data acquisition approach for applying DTs in production systems, thus enabling real-time production control. Schluse et al. [2018], introduced Experimental Digital Twins (EDT) to achieve a solid integration between virtual and physical spaces and to improve simulation. They also considered EDT as an enabler of simulation-based system engineering, optimization and control.

Currently, most DT applications are related to PHM. The first application of DT was done in aircraft PHM. Tuegel et al. [2011], applied DT to predict aircraft structural life through multi-physics modelling, multi-scale damage modelling, integration of finite element structural model (FEM) and damage models, uncertainty quantification and high-resolution structural analysis. Li et al. [2017] built a DT model based on a dynamic Bayesian network to monitor the operational status of aircraft wings. A probabilistic model was built to replace the deterministic physical model. The DT model led to a more accurate diagnosis and prognosis based on a case study of an aircraft wing leading edge. Zakrajsek and Mall [2017], constructed a DT model to predict tyre contact wear and failure probability. The DT model demonstrated many advantages over the traditional model in predicting the probability of failure for varying sink rate, yaw angle and speed.

The application of DT in PHM has other applications besides aircraft. Gabor et al. [2016], developed a simulation-based DT model designed to predict the behaviour of cyber-physical systems, structured into four levels: physical need, machine-environment interface, immediate reaction and planned reaction. In addition, Knapp et al. [2017], applied DT to the additive manufacturing process to anticipate parameters such as cooling rate, temperature gradients, microhardness, density distribution, etc. This allowed for more accurate predictions for cooling and melting rates compared to conventional methods.

Several works have also explored the application of DTs in the context of Wastewater Treatment Plant (WWTP). In Hallaji et al. [2021], DT was applied to predictive maintenance (PdM) to forecast equipment performance based on status monitoring data and to estimate when maintenance is required. In Liu et al. [2023], a DT framework for fault detection in WWTPs is proposed.

## 2.4 Conclusion

The DT has revolutionized the entire industry paradigm, making it possible to improve process efficiency, asset lifecycle optimization and the decision-making process. DT is an extremely powerful tool that makes it possible to create digital replicas of real-world assets, processes and systems. However, DT is not just a digital replica of physical entities, but an extremely complex system that requires analysis, interpretation and integration

with other data.

DTs ability to predict failures, optimize processes and improve maintenance can only be achieved with a proper understanding and interpretation of its capabilities and limitations. Although the scientific community has already carried out several studies on this technology, the problem of the interpretability of these models is still an unresolved gap. It is therefore necessary to continue exploring and studying DTs to better understand them and ensure that their full potential is being used.





# Chapter 3

## Wastewater Treatment Plant Framework

The growing global challenge related to wastewater treatment has led to the emergence of innovative approaches in search of more efficient and sustainable solutions for Wastewater Treatment Plant (WWTP). In this context, Digital Twin (DT) has great potential to improve the understanding and management of WWTPs. In this chapter, Section 3.1, describes the WWTPs, their importance, and the different treatment stages. Section 3.2, explains the Benchmark Simulation Model 2 (BSM2), which simulates a WWTP, detailing its environment and characteristics. Then, Section 3.3 presents the proposed framework and the variables that were considered for this study.

### 3.1 Wastewater Treatment Plants

The treatment process in a WWTP takes place over several stages [de Matos et al., 2023]: preliminary treatment, primary treatment, secondary treatment, tertiary treatment, and sludge treatment. The stages of wastewater treatment are as follows [Spellman, 2008]:

1. **Preliminary treatment:** In this stage, large objects are removed, such as branches, plastics, stones, and other objects that could damage or reduce the treatment capacity due to their size. Sand is also removed through decantation.
2. **Primary treatment:** During this phase, settleable and floatable solids are removed. Floating materials such as oil and grease are also removed.
3. **Secondary treatment:** During this stage biochemical oxygen demand (BOD) and dissolved and suspended organic matter are removed employing a biological and oxidative treatment that takes place in the activated sludge reactor (biological reactor). During this process, the bacteria consume the organic matter, requiring oxygen to do so, hence a portion of the tank is maintained at adequate oxygen levels. At the end of this process, the wastewater goes to a secondary settler, while the organic matter decants. Part of the decanted matter, composed of bacteria and organic matter, is redirected to the activated sludge reactor. Thus, the concentration of bacteria in this reactor is higher than would be natural in wastewater, thus accelerating the decomposition of organic matter. This process can be seen as a way to accelerate a natural process.

4. **Tertiary treatment:** The lack of water in certain regions makes the use of water from wastewater treatment plants an important alternative [Licciardello et al., 2018]. However, before being considered suitable for use, this water must go through an adequate treatment. At this stage, wastewater is subjected to disinfection and nutrient removal processes, where bacteria, suspended solids, excess nutrients, and specific toxic compounds are removed. After this treatment, the water can be used in activities such as agriculture, irrigation of green spaces, and washing of pavements and streets, among many other activities.
5. **Sludge treatment:** The sludge resulting from the treatment process, as well as the final water, must be treated so that it can be disposed of or reused (agriculture, land-fill). These treatments may involve mechanical, chemical, thermal, thermochemical, and biological processes [Zhang et al., 2017b].

In general, these are the processes used in a WWTP, which may vary according to the capacity of the WWTPs and the type of treatment. Each stage can also be divided into sub-processes and can be carried out in different ways. A generic structure of a WWTP is represented by the simulator BSM2, which will be described in Section 3.2.

## 3.2 Benchmark Simulation Model 2

As mentioned before, WWTPs are structures with nonlinear characteristics that are subject to many variations and need to operate within the limits imposed by environmental legislation [Preisner et al., 2020]. To make the wastewater treatment process more efficient, various strategies are being tested, but due to the complexity of the system, it is difficult to compare these methods.

Thus, the simulation environment BSM2 was developed. From 1998 to 2004, the development of benchmark tools for simulation-based evaluation of control strategies for activated sludge plants was issued in Europe by Working Groups of COST Action 682 and 624. This work continued later under the coordination of the International Water Association (IWA) Task Group on Benchmarking of Control Strategies for WWTPs [Alex et al., 1999].

The use of simulators to test and evaluate WWTPs-related strategies is essential. For the development of this dissertation, it is used the BSM2, simulates all the steps in the wastewater treatment process that occurs in a WWTP. The use of BSM2 is widely accepted by the scientific community and has already been used in several studies, such as the following [Morales-Rodelo et al., 2020, Moliner-Heredia et al., 2019, Chen et al., 2014, Rastegar et al., 2014, Salles et al., 2022a].

BSM2 is a simulation environment modelled on real data, where the plant layout, simulation model, influence loads, test procedures and evaluation criteria are defined [Salles et al., 2022a]. The layout of BSM2 is presented in Figure 3.1. BSM2 describes a structure responsible for the treatment of wastewater, consisting of biological and sludge treatment. The simulator is composed of primary settling, an activated sludge system (including anoxic and aerobic reactors and a secondary clarifier), an anaerobic digester, a sludge thickener and dewatering, and a sludge storage unit.

The two models used in BSM2 are the Activate Sludge Model 1 (ASM1) and the Anaerobic Digester Model 1 (ADM1). ASM1 models the biological treatment and ADM1 is responsible for the anaerobic digestion and biogas production.

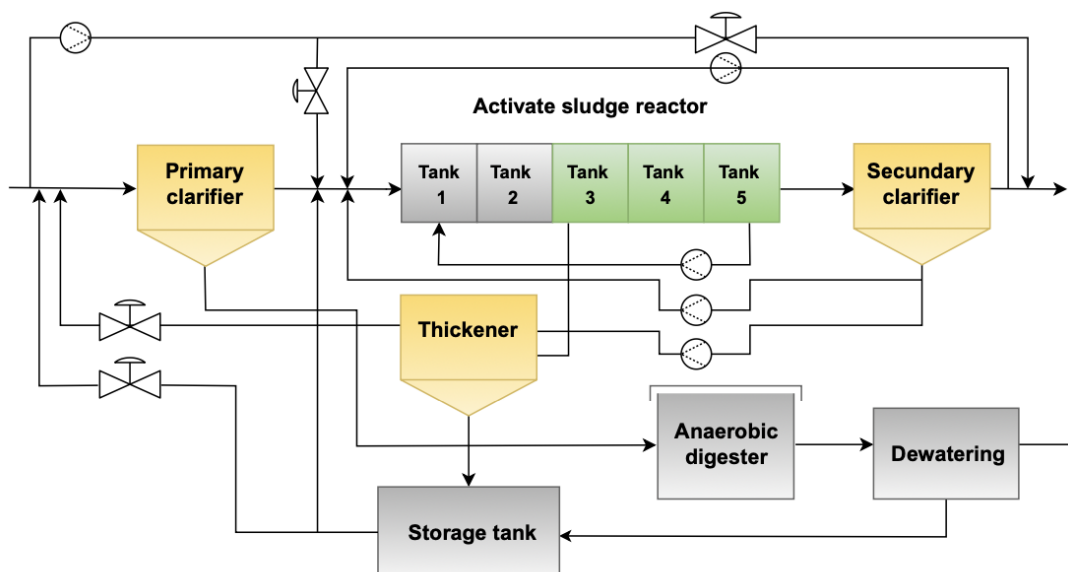


Figure 3.1: Schematic of BSM2. Taken from [Salles et al., 2022b].

### 3.2.1 Activate Sludge Model N°1 - (ASM1)

To promote the use and the proving of models of biological systems for the treatment of effluents in the activated sludge process, a working group was created in 1983 under the responsibility of IWA [Hauduc et al., 2011]. The work developed by this group led to the creation of a family of activated sludge biological treatment models, and the first model launched was ASM1.

ASM1 configures the model of the biologic reactor or activated sludge reactor with five compartments, with an anoxic zone, consisting of two tanks, and an aerated zone, consisting of three tanks. In these two zones, the denitrification and nitrification processes take place. The ASM1 model is part of BSM1 (see Figure 3.2) and describes an activated

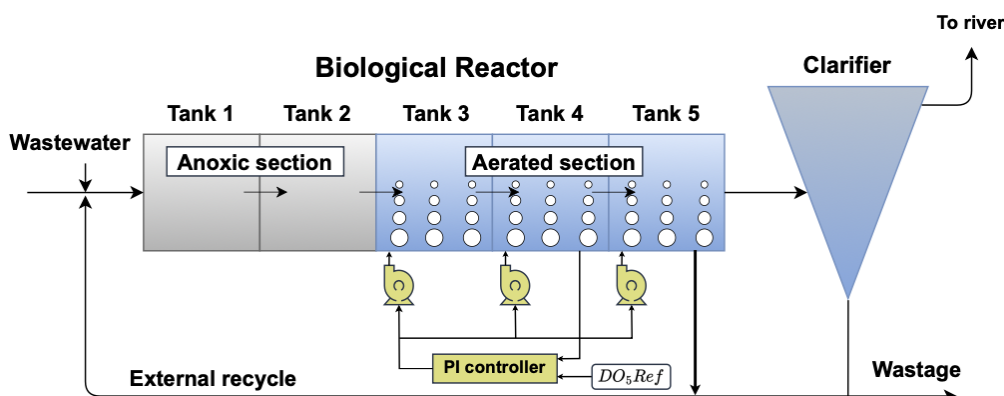


Figure 3.2: Schematic of BSM1. Taken from [Salles et al., 2022b].

sludge system that combines the ASM1 model with a secondary clarifier.[Meneses et al., 2016]. BSM1 served as the basis for the creation of BSM2, adding the primary decanter and sludge treatment tanks to the original design.

To preserve the bacterial population responsible for breaking down organic matter, the sludge from the secondary clarifier is recirculated to the anoxic tanks. As a result

of this recirculation, the sludge concentration is kept constant [Haimi et al., 2009]. The process that takes place in the biological reactor can be described in eight steps [COSTA et al., 2016]:

- Aerobic growth of heterotrophs;
- Anoxic growth of heterotrophs;
- Aerobic growth of autotrophs;
- Decay of heterotrophs;
- Decay of autotrophs;
- Ammonification of soluble organic nitrogen;
- Hydrolysis of organic compounds;
- Hydrolysis of inorganic compounds.

These methods aim to mitigate the risks associated with chemical compounds produced by human activities and to speed up the processes that occur in nature.

### 3.2.2 Anaerobic Digester Model N°1 - (ADM1)

Sludge is one of the products of wastewater treatment, and these compounds need to be treated before being used or disposed of. Thus, the anaerobic digester, described by the ADM1 model [Batstone et al., 2002], performs the anaerobic digestion and the production of biogas. Several factors influence the process of anaerobic digestion, the main ones being temperature and pH, which influence bacterial growth. For optimal bacterial growth, the pH should be between 6.7 and 7.5 [do Amaral et al., 2019]. The production of biogas is one of the results of anaerobic activity. The biogas from digested sludge is composed of methane (CH<sub>4</sub>, 50-70%) and carbon dioxide (CO<sub>2</sub>, 30-50%) [Shen et al., 2015].

Thus the ADM1 and ASM1 models that make up BSM2 are fundamental elements for accurately representing the processes that take place in a WWTP.

### 3.2.3 BSM2 Variables

Throughout the different stages of the treatment process, various variables can be monitored, the most important of which are:

- Time [days];
- $S_I$  (inert soluble material, [g COD/ $m^3$ ]);
- $S_S$  (readily biodegradable substrate, [g COD/ $m^3$ ]);
- $X_I$  (inert particulate material, [g COD/ $m^3$ ]);
- $X_S$  (slowly biodegradable substrate, [g COD/ $m^3$ ]);
- $X_{BH}$  (heterotrophic biomass, [g COD/ $m^3$ ]);
- $X_{BA}$  (autotrophic biomass, [g COD/ $m^3$ ]);

- $X_p$  (inert particulate material from biomass decay, [g COD/ $m^3$ ]);
- $S_O$  (dissolved oxygen, [g COD/ $m^3$ ]);
- $S_{NO}$  (nitrate and nitrite, [g N/ $m^3$ ]);
- $S_{NH}$  (ammonia and ammonium, [g N/ $m^3$ ]);
- $S_{ND}$  (soluble organic nitrogen associated with  $S_S$ , [g N/ $m^3$ ]);
- $X_{ND}$  (particulate organic nitrogen associated with  $X_S$ , [g N/ $m^3$ ]);
- $S_{ALK}$  (alkalinity);
- TSS (total suspended solids, [g  $S_S/m^3$ ]);
- flow rate, [ $m^3$ /day];
- temperature, [ $^{\circ}$ C].

### 3.3 Digital Twin Framework on BSM2

In BSM2, during the various stages of the treatment process, many variables can be assessed. According to experts from WWTP who have collaborated in this work, along the waterline of the WWTP, the main ones and the most important to the prediction case study investigated in this study are:

- total suspended solids,  $TSS$  [g  $S_S/m^3$ ];
- ammonia and ammonium,  $S_{NH}$  [g N/ $m^3$ ];
- dissolved oxygen,  $S_O$  [g COD/ $m^3$ ];
- nitrate and nitrite,  $S_{NO}$  [g N/ $m^3$ ];
- biological oxygen demand,  $BOD$  [g / $m^3$ ];
- chemical oxygen demand,  $COD$  [g / $m^3$ ].

Total Suspended Solids ( $TSS$ ) are one of the main pollutants of the water, contributing to its deterioration and the costs associated with its treatment to be very high. They also affect fauna and flora, as well as the general appearance of the water [Spellman, 2008]. This is an essential parameter because an excess of  $TSS$  causes the dissolved oxygen ( $S_O$ ) to be depleted.

Nitrogen compounds, particularly ammonia ( $S_{NH}$ ), are the result of various human activities, such as agriculture, industry and domestic activities. When discharged excessively into the water, ammonia causes serious environmental problems, such as a decrease in the concentration of dissolved oxygen, accelerated eutrophication of lakes and rivers (excessive algae growth) and toxicity to fish and other aquatic animals [Spellman, 2008]. To avoid these problems, nitrogen compounds are removed in the biological reactor.

Dissolved Oxygen ( $S_O$ ) is an important parameter monitored in WWTPs to assess the level of oxygen available in the water. It represents the concentration of oxygen molecules dissolved in the wastewater. During wastewater treatment the level of  $S_O$  is

measured at various stages of the treatment process. The  $S_O$  concentration provides information on the biological activity and the efficiency of the treatment processes. By monitoring  $S_O$  levels throughout the WWTP, operators can optimize aeration systems, adjust treatment processes, and ensure that sufficient oxygen is available for the biological treatment processes. This helps maintain proper treatment efficiency and minimize the potential impact of wastewater discharge on the receiving environment [Spellman, 2008].

The presence of nutrients in water is beneficial for plant growth, but an excess of nutrients can lead to eutrophication of the water. High levels of nitrates and phosphorus can cause this process to occur more quickly. High levels of nitrogenous substances in wastewater affect fauna and flora, which is why it is essential to reduce nitrogen pollution loads in wastewater. For nitrogenous substances to be eliminated from wastewater, nitrification and denitrification processes must be carried out. During these processes, heterotrophic bacteria use nitrite and nitrate ( $S_{NO}$ ) as oxidizers of organic matter, causing nitrogen to be removed as a gas [Spellman, 2008].

The amount of oxygen consumed by bacteria and other microorganisms to degrade the water contents aerobically is given by the Biological oxygen demand (BOD), for five days at a temperature of 20 °C. This is not a direct measure and depends on the sum of all the biodegradable organic substances in the water. *BOD* indicates the amount of dissolved oxygen required in a given period for the biological degradation of the organic constituents of wastewater and is given by the following expression [Spellman, 2008]:

$$BOD = 0.25(S_S + X_S + (1 - f_p)(X_{BH} + X_{BA})), \quad (3.1)$$

where  $f_p=0.08$  is a stoichiometric parameter,  $S_S$  is the readily biodegradable substrate,  $X_S$  is the slowly biodegradable substrate,  $X_{BH}$  is the heterotrophic biomass, and  $X_{BA}$  is the autotrophic biomass.

Chemical oxygen demand, COD, is the oxygen equivalent of organic matter that can be oxidized by a strong chemical oxidizer in an acidic medium. COD measures both biodegradable and non-biodegradable organic matter. The results can be obtained in a few hours. COD is an essential parameter for WWTPs as it helps assess the organic pollution level and the treatment efficiency of the plant. By monitoring the COD concentrations, it is possible to assess the effectiveness of the treatments and ensure compliance with regulatory standards. COD is given by the following expression [Spellman, 2008]:

$$COD = (S_S + S_I + X_S + X_I + X_{BH} + X_{BA} + X_P), \quad (3.2)$$

where  $S_S$  is the readily biodegradable substrate,  $S_I$  is the inert soluble material,  $X_I$  is the inert particulate material,  $X_{BH}$  is the heterotrophic biomass,  $X_{BA}$  is the autotrophic biomass, and  $X_P$  is the inert particulate material from biomass decay.

### 3.4 Conclusion

By using BSM2 it is possible to implement and test strategies for control, optimization, anomaly and fault detection. It also allows a comparison to be made between the results obtained based on predefined metrics, which makes it possible to achieve innovations and contributions to mitigate the challenges related to wastewater treatment.

Over the last few years, the use of benchmark models has proved important for developing and evaluating control strategies and monitoring systems in WWTPs. The popularity of BSM1 and later BSM2 within the scientific community has led to the

appearance of several publications [Tejaswini et al., 2020]. More than 300 peer-reviewed articles, conference presentations, and dissertations have been published on work related to benchmark systems, which shows that this is a very robust simulator that is viewed favourably by researchers [Flores-Alsina et al., 2012], being a good option for use in this dissertation.





# Chapter 4

## Digital Twin Methodologies

Fuzzy Logic Systems (FLS) has been used in various industrial systems and consumer products, attracting the attention of many researchers. Fuzzy systems are powerful mathematical tools used to model nonlinear problems, which are usually difficult to model and control. They use fuzzy logic to model uncertainty and imprecision, allowing decisions to be made based on a subjective representation of reality, and they are easier to understand and explain since, although they are based on a mathematical representation and formulation, they are described by semantic rules based on natural language rather than complex mathematics.

Using fuzzy-based models to make a Digital Twin (DT) of Wastewater Treatment Plant (WWTP) has numerous advantages for improving the whole wastewater treatment process. Through FLS, it is possible to create models that are easy to understand and interpret by any operator, thus helping in the decision-making process and the optimization of the whole process. Furthermore, FLS can deal with the uncertainty, complexity and variability associated with WWTPs.

This chapter presents the three models useful for this dissertation. Section 4.1 presents the concept of FLS and a generic Takagi-Sugeno (T-S) fuzzy model. Section 4.2 presents the first model, the Model based on Fuzzy C-Means and Least Squares Method (FCM-LSM). Section 4.3 presents the second model, the Generalized additive models using zero-order T-S fuzzy systems (GAM-ZOTS) model. Finally, Section 4.4 presents the last model, the Iterative Learning of multiple univariate zero-order T-S fuzzy systems (iMU-ZOTS) model.

### 4.1 Fuzzy Logic Systems

FLS are used to model nonlinear systems and can be used to approximate unknown nonlinear functions that make plant models [Wang et al., 1992]. Using FLS, it is possible to develop relationships between a set of input and output data, using the theory of fuzzy sets introduced by Zadeh in 1965 [Zadeh, 1965]. In these models, each fuzzy set is represented by a Membership Function (MF), whose values vary between 0 and 1. These MFs can have different shapes, such as triangular, trapezoidal, Gaussian, etc.

In general terms, FLS are made up of 3 stages: fuzzification, modelling the inference engine, which uses the fuzzy rule base, and defuzzification (see Figure 4.1). In the first stage (fuzzification) the input variables are transformed into fuzzy sets. In the second stage (inference engine model) these fuzzy sets are interpreted by a fuzzy inference model to make a correspondence between the inputs and outputs, based on the fuzzy rules. In

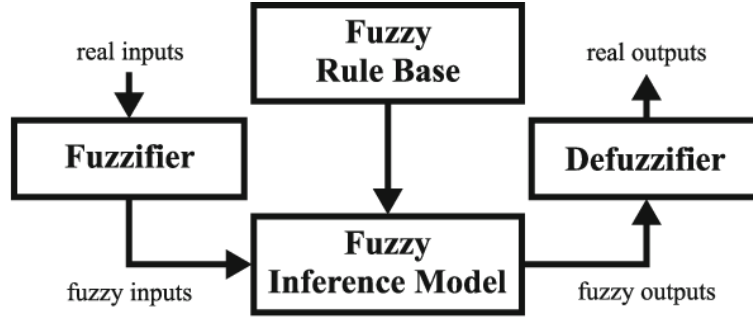


Figure 4.1: Basic configuration of FLS. Taken from [Júnior et al., 2023].

the final step (defuzzification) the mapped fuzzy outputs are transformed into real output variables.

### 4.1.1 Takagi-Sugeno Fuzzy Model

One important type of fuzzy system, and widely used in the scientific community, is the Takagi-Sugeno (T-S) fuzzy system [Takagi and Sugeno, 1985]. This type of fuzzy system has gained great emphasis due to its consequential rule structure. The major difference between this fuzzy system and Mamdani fuzzy systems is the consequent parts of the fuzzy IF-THEN rules, which in a T-S fuzzy system is a real-valued function instead of a fuzzy set as in Mamdani fuzzy systems.

One way of describing a nonlinear system is using a T-S fuzzy model, which is defined by a set of rules as follows:

$$\begin{aligned}
 R_i : & \text{ IF } x_1(k) \text{ is } A_1^i \text{ and } \dots \text{ and } x_n(k) \text{ is } A_n^i \\
 & \text{ THEN } y_i(k) = \theta_{i0} + \theta_{i1}x_1(k) + \dots + \theta_{in}x_n(k), \\
 & i = 1, 2, \dots, N,
 \end{aligned} \tag{4.1}$$

where  $R_i$  represents the  $i$ -th fuzzy rule, and  $N$  is the total number of rules. The input variables of the T-S fuzzy system are  $x_1(k), \dots, x_n(k)$ . The linguistic terms  $A_j^i$  ( $j = 1, 2, \dots, n$ ), are defined by fuzzy MFs, labelled  $\mu_{A_j^i}$ . These MFs describe the plant's operating region.  $\theta_{i1}, \dots, \theta_{in}$  are model (consequent) parameters of  $y_i(k)$ . From (4.1), the T-S fuzzy model  $y(k)$  can be rewritten as:

$$y[\mathbf{x}(k)] = \sum_{i=1}^N \bar{\omega}^i[\mathbf{x}(k)] y_i(\boldsymbol{\theta}_i, \mathbf{x}(k)), \tag{4.2}$$

where for  $i = 1, \dots, N$ ,

$$\mathbf{x}(k) = [x_1(k), \dots, x_n(k)], \tag{4.3}$$

$$\mu_{A_j^i}(x_j) = \exp\left(-\frac{(x_j - v_{ij})^2}{\sigma_{ij}}\right), \tag{4.4}$$

$$\bar{\omega}^i[\mathbf{x}(k)] = \frac{\prod_{j=1}^n \mu_{A_j^i}[\mathbf{x}(k)]}{\sum_{i=1}^N \prod_{j=1}^n \mu_{A_j^i}[\mathbf{x}(k)]}, \tag{4.5}$$

$$\boldsymbol{\theta}_i = [\theta_{i0}, \theta_{i1}, \dots, \theta_{in}]. \tag{4.6}$$

where  $v_{ij}$ ,  $\sigma_{ij}$  and  $\theta_i$  represent the center, width of the MF and model parameters, respectively, which need to be defined/learned.

## 4.2 Model based on Fuzzy C-Means and Least Squares Method

To construct a T-S fuzzy system of the form (4.1) it is necessary to obtain the antecedent parameters ( $\mathbf{v}_i$  and  $\sigma_i$ ), and the consequent parameters ( $\theta_i$ ). The antecedent parameters are given by the Fuzzy C-Means (FCM) algorithm (Subsection 4.2.1), and the consequent parameters are given by the Least Squares Method (LSM) (Subsection 4.2.2). The FCM-LSM is presented in Subsection 4.2.3.

### 4.2.1 Fuzzy C-Means Method

In the FCM algorithm, the main objective is to partition a given dataset, denoted as  $\mathbf{X}$ , into a given number of clusters,  $N$ . In this method, the objects can belong to multiple clusters, with different degrees of membership [Rastegar et al., 2017].

Let's consider a scenario with a set of  $n$  samples that make up an input observation, denoted as  $k$ . Each sample corresponds to a specific input variable, and they are grouped as an  $n$ -dimensional vector  $\mathbf{x}(k) = [x_1(k), \dots, x_n(k)]^T$ , where  $\mathbf{x}(k) \in \mathbb{R}^n$ . A set of  $K$  observations is then denoted as:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{K1} & x_{K2} & \dots & x_{Kn} \end{bmatrix}. \quad (4.7)$$

The fuzzy partitioning of the dataset  $\mathbf{X}$  into a predetermined number of clusters,  $N$ , comprises a family of fuzzy subsets  $\{A^i | 1 \leq i \leq N\}$ . These subsets are characterized by MFs,  $\mu_i(k)$ , which are defined as  $\mu_i(k) = \mu_{A^i}(\mathbf{x}_k)$ . The MFs form the fuzzy partition matrix  $\mathbf{U} = [u_{ik}] = [\mu_i(k)] \in \mathbb{R}^{N \times K}$ . Each row of the  $\mathbf{U}$  matrix, indexed by  $i$ , represents the membership values of the  $i$ -th fuzzy subset  $A^i$  for all observations in the data matrix  $\mathbf{X}$ . Equally, to Dovžan and Škrjanc [2011], the partition matrix must meet the following criteria: the membership degrees are real numbers in interval  $\mu_i(k) \in [0,1]$ , for  $k = 1, \dots, K$ ; the total membership of each sample in all the clusters must be equal to one  $\sum_{i=1}^N \mu_i(k) = 1$ ; and none of the fuzzy clusters is empty nor does any contain all the data  $0 < \sum_{k=1}^K \mu_i(k) < K$ ,  $1 \leq i \leq N$ .

The FCM method aims to minimize an objective function, defined by (4.8). This function has a predefined number of clusters,  $N$ , and also includes a fuzziness parameter,  $\eta$ .

$$J(\mathbf{X}, \mathbf{U}, \mathbf{V}) = \sum_{i=1}^N \sum_{k=1}^K \mu_i^\eta(k) d_{ik}^2(\mathbf{x}(k), \mathbf{v}_i), \quad (4.8)$$

where  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_N]^T \in \mathbb{R}^{N \times n}$  is a matrix of cluster centroid vectors  $\mathbf{v}_i = [v_{i1}, \dots, v_{in}]^T$ ,  $d_{ik}(\mathbf{x}(k), \mathbf{v}_i)$  is the Euclidean distance ( $l^2$ -norm) between the observation  $\mathbf{x}(k)$  and the cluster centroid  $\mathbf{v}_i$  given by (4.10), and  $\eta$  is the fuzziness parameter.

When the derivative of the objective function (4.8) is calculated with respect to the cluster centers  $\mathbf{V}$  and the membership values  $\mathbf{U}$ , the membership values can be calculated

---

**Algorithm 1** Algorithm to learn the antecedent part, based on FCM algorithm.

---

1. Start by obtaining a dataset  $\mathbf{X}$  (4.7) and specify the number of clusters,  $N$ , the degree of fuzziness,  $\eta$ , the termination condition,  $imp$ , and the maximum of iterations,  $Max$ . Initialize the partition matrix  $\mathbf{U}$ , randomly;
  2. Calculate initial cluster centers using the formula (4.11) with the membership values of the initial partition matrix  $\mathbf{U}$ ;
  3. For iteration  $t = 1, \dots, Max$ :
    - (a) Calculate the membership values using (4.9), at iteration  $t$ ,  $\mu_i^{(t)}(k)$ , of each input data object  $\mathbf{x}_k$  in each cluster  $i$ , using the cluster center vector  $\mathbf{v}_i^{t-1}(k)$  from iteration  $(t - 1)$ ;
    - (b) Update the cluster center for each cluster  $i$  at iteration  $t$ ,  $\mathbf{v}_i^t(k)$ , by (4.11), using the membership values (4.9) at iteration  $t$ ,  $\mu_i^{(t)}(k)$ ;
    - (c) If a termination condition is satisfied, e.g  $|\mathbf{v}_i^t - \mathbf{v}_i^{t-1}| \leq imp$ , exit the iterative loop, and save the last iteration of the matrices  $\mathbf{U}$  and  $\mathbf{V}$  and go to Step 4;
  4. Compute the parameters  $\sigma$  using (4.12).
- 

as follows:

$$\mu_i(k) = d_{ik}^2(\mathbf{x}(k), \mathbf{v}_i) \sum_{q=1}^N \left( d_{qk}^{2/(\eta-1)}(\mathbf{x}(k), \mathbf{v}_q) \right)^{-1}, \quad (4.9)$$

where:

$$d_{ik}^2(\mathbf{x}(k), \mathbf{v}_i) = (\mathbf{x}(k) - \mathbf{v}_i)^T (\mathbf{x}(k) - \mathbf{v}_i), \quad (4.10)$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^K \mu_i^\eta(k) \mathbf{x}(k)}{\sum_{k=1}^K \mu_i^\eta(k)}. \quad (4.11)$$

The centers  $v_{ij}$  of the antecedent membership functions are obtained by (4.11) where  $\mathbf{v}_i = [v_{i1}, \dots, v_{in}]^T$ , and the corresponding widths,  $\boldsymbol{\sigma}_i = [\sigma_{i1}, \dots, \sigma_{in}]^T$ ,  $i = 1, \dots, N$ , are obtained using  $\mathbf{U} = [\mu_i(k)]$ , as follows:

$$\sigma_{ij} = \sqrt{\frac{2 \sum_{k=1}^K \mu_i(k) (x_j(k) - v_{ij})^2}{\sum_{k=1}^K \mu_i(k)}}, \quad j = 1, \dots, n. \quad (4.12)$$

The algorithm, based on the FCM algorithm, to learn the antecedent parameters is presented in Algorithm 1.

## 4.2.2 Least Squares Method

After learning of the antecedent parameters (Subsection 4.2.1), the consequent parameters are given by the LSM.

Consider a desired model  $y_d(\mathbf{x}(k))$  for the output of a system and consider a set of  $K$  input patterns  $\mathbf{x}(k)$ , together with the corresponding desired output patterns  $y_d(\mathbf{x}(k))$  for  $k = 1, 2, \dots, K$  and

$$\mathbf{y}_d = [y_d(\mathbf{x}(1)), \dots, y_d(\mathbf{x}(K))]^T, \quad (4.13)$$

$$\mathbf{\Phi} = [\boldsymbol{\phi}(1), \dots, \boldsymbol{\phi}(K)]^T, \quad (4.14)$$

$$\boldsymbol{\phi}(k) = [\bar{\omega}^1[\mathbf{x}(k)]\mathbf{x}_e^T(k), \dots, \bar{\omega}^N[\mathbf{x}(k)]\mathbf{x}_e^T(k)]^T, \quad (4.15)$$

$$\mathbf{x}_e(k) = [1, x_1(k), x_2(k), \dots, x_n(k)]^T, \quad (4.16)$$

then

$$\mathbf{y}_d = \mathbf{\Phi}\boldsymbol{\Theta}. \quad (4.17)$$

Suppose that  $\boldsymbol{\Theta}^*$  represents the optimal values of all consequent parameters  $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_N^T]^T$ , a solution of  $\boldsymbol{\Theta}^*$  can be calculated using the pseudo-inverse least squares method as follows [Mendes et al., 2012]:

$$\boldsymbol{\Theta}^* = \mathbf{\Phi}^+\mathbf{y}_d, \quad (4.18)$$

where, considering  $K \geq N(n+1)$ , the pseudo inverse,  $\mathbf{\Phi}^+$ , can be computed in closed form as follows:

$$\mathbf{\Phi}^+ = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T. \quad (4.19)$$

### 4.2.3 Model Design Algorithm

To build a T-S fuzzy system of the form (4.1), it is needed to obtain the antecedent parameters ( $\mathbf{v}_i$  and  $\boldsymbol{\sigma}_i$ ) and the consequent parameters ( $\boldsymbol{\theta}_i$ ). The antecedent parameters are given by the FCM algorithm (Subsection 4.2.1) and the consequent parameters are provided by the LSM (Subsection 4.2.2). The design of the FCM-LSM is presented in Algorithm 2.

---

**Algorithm 2** Algorithm to design the FCM-LSM.

---

1. Define the inputs,  $K$  data samples  $\{(\mathbf{x}(k), y(k))\}_{k=1}^K$ , with  $\mathbf{x}(k) = [x_1(k), \dots, x_n(k)]^T$ ; Define the parameters to be used on the FCM.
  2. Obtain the antecedent parameters ( $\mathbf{v}_i$  and  $\boldsymbol{\sigma}_i$ ) through the FCM algorithm described in Algorithm 1;
  3. Obtain the consequent parameters using LSM:
    - (a) Construct the vector  $\mathbf{\Phi}$  through (4.14);
    - (b) Compute the consequent values  $\boldsymbol{\Theta}^*$  using (4.18).
-

### 4.3 GAM-ZOTS - Generalized Additive Models Using Zero-Order T-S Fuzzy Systems

Additive modelling was introduced in the 1980s to improve the accuracy and interpretability of learned models. Models obtained by additive modelling are characterized by being composed of a sum of univariate smooth functions representing each input variable [Sánchez and Otero, 2004]. In Mendes et al. [2019a], an approach for learning a system of neo-fuzzy neurons called GAM-ZOTS is proposed. GAM-ZOTS uses the backfitting algorithm to learn multiple univariate zero-order T-S additive fuzzy systems (neo-fuzzy neuron systems). This approach aims to improve the accuracy and interpretability, where each model shows the dependence of the output on its input variable [Sánchez and Otero, 2004].

Additive models are made up of univariate smooth functions  $s_j(x_j)$  ( $j = 1, \dots, n$ ), in form of (4.20), to estimate multivariate regression functions. These functions can be learned using the “backfitting” algorithm [Mendes et al., 2019a].

$$\hat{y} = f(x_1, \dots, x_n) = s_0 + \sum_{j=1}^n s_j(x_j), \quad (4.20)$$

where  $x_j$  ( $j = 1, \dots, n$ ) are the input variables and  $s_j(\cdot)$  can be of any parametric or nonparametric, smooth functions.

#### 4.3.1 GAM-ZOTS Model

The model structure used is based on the concept of additive models (4.20), in which each univariate model is a zero-order T-S fuzzy system representing a relationship between the input variable  $x_j$  (for  $j = 1, 2, \dots, n$ ) and the output. Therefore, the model presented consists of the sum of multiple zero-order T-S fuzzy systems expressed as follows:

$$\begin{aligned} y(\mathbf{x}) &= y_0 + y_1(x_1) + \dots + y_n(x_n), \\ &= y_0 + \sum_{j=1}^n y_j(x_j), \end{aligned} \quad (4.21)$$

where  $y_0$  is a constant, and each individual model  $y_j(x_j)$ , is represented by (4.23), being composed of  $N_j$  zero-order T-S fuzzy rules in the format of (4.22).

$$\begin{aligned} R_j^1 &: \text{IF } x_j(k) \text{ is } A_j^1 \text{ THEN } y_j^1(k) = \theta_j^1, \\ &\vdots \\ R_j^{N_j} &: \text{IF } x_j(k) \text{ is } A_j^{N_j} \text{ THEN } y_j^{N_j} = \theta_j^{N_j}. \end{aligned} \quad (4.22)$$

$$y_j(x_j(k)) = \sum_{i=1}^{N_j} \omega_j^i[x_j(k)]\theta_j^i = \phi_j(x_j(k))\theta_j, \quad (4.23)$$

where for  $i = 1, \dots, N_j$ ,

$$\omega_j^i[x_j(k)] = \frac{\mu_{A_j^i}(x_j(k))}{\sum_{i=1}^{N_j} \mu_{A_j^i}(x_j(k))}, \quad (4.24)$$

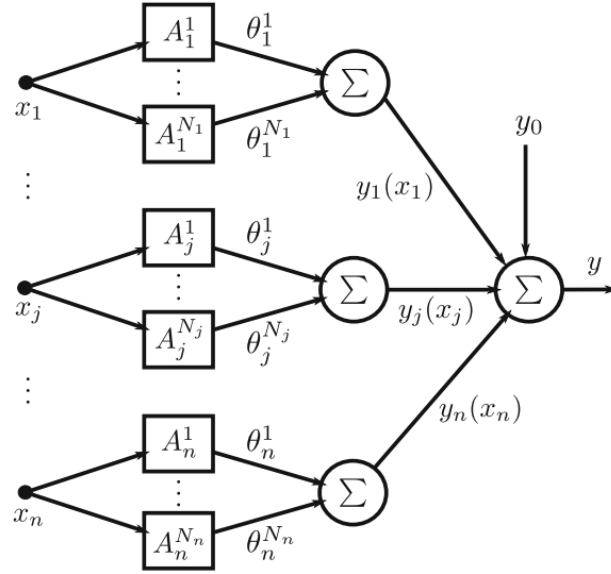


Figure 4.2: Multiple zero-order T-S fuzzy systems structure. Taken from [Mendes et al., 2019a].

$$\boldsymbol{\theta}_j = [\theta_j^1, \dots, \theta_j^{N_j}]^T, \quad (4.25)$$

$$\boldsymbol{\phi}_j(x_j(k)) = [\omega_j^1[x_j(k)], \dots, \omega_j^{N_j}[x_j(k)]] . \quad (4.26)$$

The structure of the GAM-ZOTS model is shown in Figure 4.2. As mentioned before, this model is composed of multiple zero-order T-S fuzzy systems.

The membership functions are complementary triangular membership functions [Silva et al., 2014]. Figure 4.3 shows an example of complementary triangular membership functions, in which the minimum and maximum value of the input variable  $x_j$  are represented by  $x_j^-$  and  $x_j^+$ , respectively. The lower limit, center value, and upper limit of  $A_j^i$  are represented by  $a_{j,i}$ ,  $b_{j,i}$  and  $c_{j,i}$ , respectively.

### GAM-ZOTS Learning

The learning methodology of the GAM-ZOTS is presented in the Algorithm 3. More details are presented in Mendes et al. [2019a]. The first step is to define the MFs for each input which are complementary triangular MFs, as present in Figure 4.3. These MFs are obtained in Steps 2a – 2b of Algorithm 3. The next step is to obtain the consequent parts of the fuzzy rules corresponding to each input variable  $x_j$ , using the Backfitting Algorithm (Step 3 of Algorithm 3).

## 4.4 iMU-ZOTS - Iterative Learning of Multiple Univariate Zero-Order T-S Fuzzy Systems

In the GAM-ZOTS method, the antecedent membership functions are created manually, as presented before, distributing them evenly across the universe of discourse, which means that there are limitations and inefficiencies in the design of the antecedent parts since they



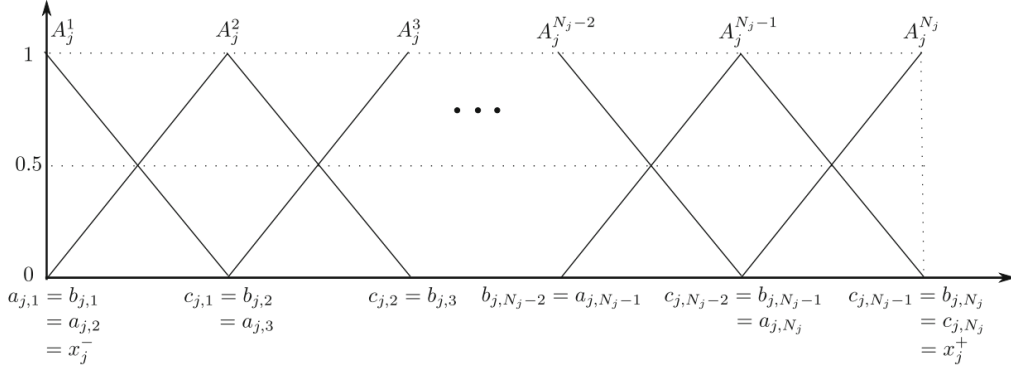


Figure 4.3: Membership functions for  $x_j$ . Taken from [Mendes et al., 2019a].

---

**Algorithm 3** Description of GAM-ZOTS learning method [Mendes et al., 2019a].

---

1. Define the inputs,  $K$  data samples  $\{(\mathbf{x}(k), y(k))\}_{k=1}^K$ , with  $\mathbf{x}(k) = [x_1(k), \dots, x_n(k)]^T$ ;
  2. Define the MFs. For all input variables  $x_j$ :
    - (a) Define the minimum and the maximum values,  $x_j^-$  and  $x_j^+$ ;
    - (b) Set the number of membership functions,  $N_j$ .
    - (c) For each complementary triangular MF,  $A_j^i$ , the parameters are:
      - i. For  $A_j^1$ :  $a_{j,1} = b_{j,1} = x_j^-$  and  $c_{j,1} = b_{j,2}$ ;
      - ii. For  $A_j^i$ :  $a_{j,i} = b_{j,i-1}$  and  $b_{j,i} = c_{j,i-1}$ ,  $c_{j,i} = b_{j,i+1}$ ;
      - iii. For  $A_j^{N_j}$ :  $a_{j,N_j} = b_{j,N_j-1}$  and  $b_{j,N_j} = c_{j,N_j} = x_j^+$ .
  3. Using the Backfitting Algorithm, calculate the consequent parameters ( $\theta_j^i$ ). More details in Mendes et al. [2019a].
- 

are not distributed according to the data. In Mendes et al. [2019b] an iterative learning approach is proposed for a model composed of the sum of multiple univariate T-S fuzzy systems of zero order, called iMU-ZOTS. In this approach, the learning of the consequent part is based on the GAM-ZOTS algorithm described in Section 4.3.

The major difference between iMU-ZOTS and GAM-ZOTS, is in the iterative learning of the antecedent part, where instead of MFs being manually designed, iMU-ZOTS iteratively designs the MFs and selects candidate input variables where new MFs can be added. This selection is based on a novelty detection criterion, similar to what is proposed in the Gath-Geva recursive clustering method [Soleimani-B et al., 2010].

The iMU-ZOTS are divided into the following steps:

- Initialization: In this step, the antecedent and consequent of the fuzzy rules are designed.
- Iterative Learning: In this step, an iterative process is carried out in which new fuzzy rules are added to the selected input variable(s) and other update operations

are performed.

### 4.4.1 iMU-ZOTS Model

#### Initialization

The initialization of the iMU-ZOTS involves applying the GAM-ZOTS method (Algorithm 3), with  $N_j = 2$  fuzzy rules for each input variable,  $x_j$ . In this way, two ( $N_j = 2$ ) complementary triangular MFs are defined,  $A_j^{i_j}$  for all input variables  $x_j$  ( $j = 1, \dots, n$ ):

1. parameters for  $A_j^1$ :  $a_{j,1} = b_{j,1} = x_j^-$  and  $c_{j,1} = b_{j,2}$ ;
2. parameters for  $A_j^2$ :  $a_{j,2} = b_{j,1}$  and  $b_{j,2} = c_{j,2} = x_j^+$ ;

$x_j^-$  and  $x_j^+$  represent the limits of the universe of discourse (minimum and maximum values, respectively) of the input variable  $x_j$ .

The consequent parameters ( $\theta_j$ ) are obtained through the GAM-ZOTS Algorithm 3.

#### Iterative Learning

In the iterative learning stage, a new MF is added iteratively to the selected input variable.

1) *Variable Selection*: The iterative learning approach starts by selecting the candidate input variable, denoted as  $x_j^*$ , where a new fuzzy rule can be added. This selection is done using a novelty detection criterion, which is based on the Gath-Geva recursive clustering method [Soleimani-B et al., 2010]. Thus, a novelty detection criterion is defined to select the input variable where a new fuzzy rule should be added, iteratively measuring, through a value between 0 and 1. This facilitates the definition of the threshold for adding new fuzzy rules to the model [Mendes et al., 2018].

The novelty measurement on the  $i_j$ -th fuzzy rule ( $i_j = 1, \dots, N_j$ ) related to  $x_j(k)$  is obtained by:

$$M_{i_j}(k) = \exp(-|x_j(k) - \mu_{i_j}(k)|), \quad (4.27)$$

where

$$\mu_{i_j}(k) = \mu_{i_j}(k-1) + \frac{x_j(k) - \mu_{i_j}(k-1)}{N_{i_j}(k)} v_{i_j}^m(k), \quad (4.28)$$

$v_{i_j}(k) = \mu_{A_j^{i_j}}(x_j(k))$ ,  $m > 1$  is an integer representing the degree of fuzzification of the clusters,

$$N_{i_j}(k) = N_{i_j}(k-1) + v_{i_j}^m(k) \quad (4.29)$$

indicates the total MF degrees associated to the  $i_j$ -th MF of  $x_j$ , and  $N_{i_j}(0) = 0$ .

**Criterion 1:** the novelty detection criterion for  $x_j$  is defined as follows:

$$M_{j,\max}(k) = \begin{cases} 1, & \text{if } \max_{i_j=1}^{N_j} M_{i_j}(k) \leq M_{\text{th}}^j \\ 0, & \text{otherwise} \end{cases} \quad (4.30)$$

Where  $M_{\text{th}}^j$  is a threshold for  $x_j$ , between zero and one. When  $M_{j,\max}(k) = 0$ , means that the variable  $x_j$  is not well represented by the current model. When  $M_{j,\max}(k) = 1$ , the variable  $x_j$  is perfectly represented by the current model.

---

**Algorithm 4** Description of iMU-ZOTS learning method [Mendes et al., 2019b].

---

1. Define the inputs,  $K$  data samples  $\{(\mathbf{x}(k), y(k))\}_{k=1}^K$  where  $\mathbf{x}(k) = [x_1(k), \dots, x_n(k)]^T$ ; Define their parameters;
  2. Initialization: for each input variable  $x_j$ : define two ( $N_j = 2$ ) MFs and obtain the initial model using the GAM-ZOTS (Algorithm 3);
  3. Iterative learning:
    - (a) For all points in the training dataset  $k = 1, \dots, K$ :
      - i. Using (4.30), the novelty  $M_{j,\max}(k)$  is measured, for input all variables;
      - ii. If Criteria (4.30) and (4.31) are met:
        - A. Define the new MF,  $A_j^{\text{new}}$ , and update  $A_{j^*}^{\text{left}}$  and  $A_{j^*}^{\text{right}}$ ;
        - B. Obtain all consequent parameters using the GAM-ZOTS (Algorithm 3 Step 3).
    - (b) If a fuzzy rule was added in Step 3a, it returns to Step 3a. If not, the learning process ends.
- 

**Criterion 2:** To avoid overfitting and learning of a complex model structure, the criterion defining a minimum distance between the center of the two nearest MFs is as follows:

$$\left| b_{j^*,i_j^*} - b_{j^*,i_j^-} \right| > \eta_j, \quad (4.31)$$

where  $b_{j^*,i_j^*}$  and  $b_{j^*,i_j^-}$  are the centers of MF to be added and of its nearest MF, respectively, and  $\eta_j$  is a threshold, which, is defined as  $\eta_j = |x_j^+ - x_j^-| / MF_{\max}^j$ , where  $MF_{\max}^j$  can be considered as the maximum number of admissible of MFs for  $x_j$ .

**2) New Membership Function:** Once Criteria 1 (4.30) and 2 (4.31) are met, a new fuzzy rule  $R_j^{\text{new}}$ , and a new MF  $A_{j^*}^{\text{new}}$ , are added in  $x_j^*$ .

Thus, the new MF,  $A_{j^*}^{\text{new}}$ , is added and the nearest left and right MFs,  $A_{j^*}^{\text{left}}$  and  $A_{j^*}^{\text{right}}$ , respectively, are updated by:

- $A_{j^*}^{\text{left}}$ : 1)  $c_{j^*,\text{left}} = b_{j^*,\text{new}}$ ;
- $A_{j^*}^{\text{new}}$ : 1)  $a_{j^*,\text{new}} = b_{j^*,\text{left}}$ , 2)  $b_{j^*,\text{new}} = x_j(k)$ , and 3)  $c_{j^*,\text{new}} = b_{j^*,\text{right}}$ ;
- $A_{j^*}^{\text{right}}$ : 1)  $a_{j^*,\text{right}} = b_{j^*,\text{new}}$ .

Algorithm 4 presents the steps of the iMU-ZOTS model learning approach.

# Chapter 5

## Results and Discussion

This chapter presents the results of the Digital Twin (DT) methodologies described in Chapter 4, which were tested to model the key variables on the Wastewater Treatment Plant (WWTP) described in Section 3.3. Section 5.1 presents the datasets used to validate the methodologies as well as the experimental tests. Section 5.2 presents the results and the respective analyses regarding the accuracy and interpretability of the models.

### 5.1 Experimental Tests

As mentioned before, this study consists of creating a DT with the help of the Benchmark Simulation Model 2 (BSM2), which, through fuzzy-based models, makes it possible to predict the essential variables in the WWTP. These models are based on fuzzy logic, which can deal with complex and imprecise data, making them suitable for dealing with the variable and uncertain nature of WWTPs.

The models presented in Chapter 4, Model based on Fuzzy C-Means and Least Squares Method (FCM-LSM), Generalized additive models using zero-order T-S fuzzy systems (GAM-ZOTS) model, and Iterative Learning of multiple univariate zero-order T-S fuzzy systems (iMU-ZOTS) model, were used to predict the variables  $TSS$ ,  $S_{NH}$ ,  $S_O$ ,  $S_{NO}$ ,  $BOD$ , and  $COD$  along the waterline of the WWTP. To collect the data, 14 points were placed along the waterline, as shown in Figure 5.1:

- $TSS$  is predicted at the exit of the primary clarifier, at point P3, using the data from points P1 and P2 as input.
- $S_{NH}$  is predicted throughout the waterline, at point P3, with P1 and P2 as input, at point P10, with P4 and P6 as input, and at point P13, with P11 and P12 as input.
- $S_O$  is predicted at point P10, with P4 and P6 as input.
- $S_{NO}$  is predicted at point P3, with P1 and P2 as inputs, at the outlet of tank 5, at P10, with P4 and P6 as inputs and at the outlet of the secondary clarifier, at P13 using P11 and P12 as inputs.
- $BOD$  and  $COD$  are predicted at point P13, using points P11 and P12 as input.

The datasets of the variables to be predicted are built based on the input and output points, and in the end, there are 10 variables to be predicted distributed among the points on the WWTP waterline. These variables are:  $TSS_3$ ,  $S_{NH_3}$ ,  $S_{NH_{10}}$ ,  $S_{NH_{13}}$ ,  $S_{NO_3}$ ,  $S_{NO_{10}}$ ,

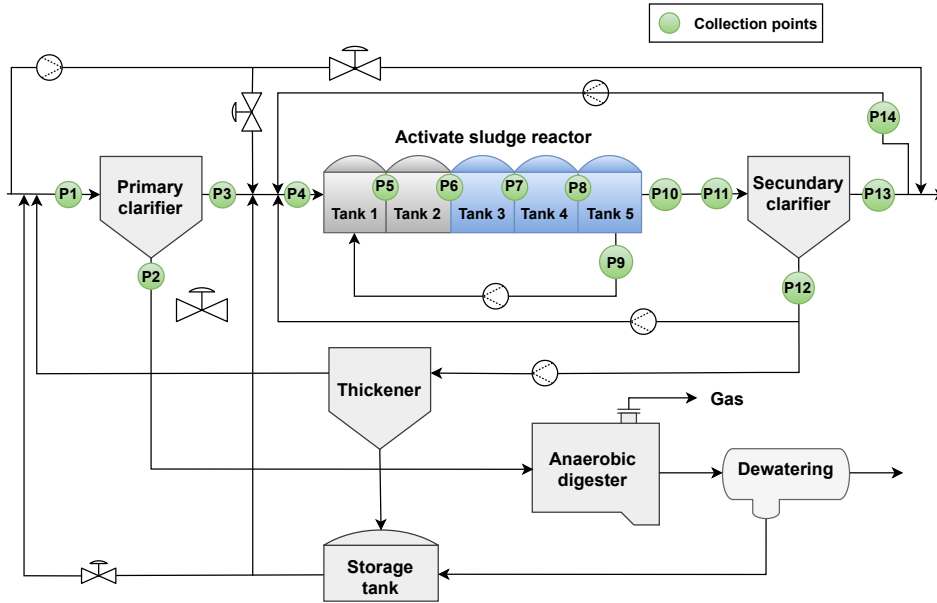


Figure 5.1: BSM2 data collection points.

$S_{NO_{13}}$ ,  $S_{O_{10}}$ ,  $BOD_{13}$  and  $COD_{13}$ . The subscript value of each variable corresponds to the point where that variable is being predicted.

For the evaluation of the proposed framework, to predict each desired variable, 609 days are considered, where the first 245 days are for the stabilization of the plant. The sampling frequency is 15 minutes, totalling 58464 samples. During the 609 days of operation, various factors can make the process even more complex, including climatic phenomena such as droughts, rains, and storms. It is, therefore, necessary to consider all these factors that can influence the results obtained.

The variables used as input were selected according to their correlation with the variables to be predicted. The value of the correlation between the input variables and the variable to be predicted is then compared with a previously defined threshold. Input variables with a higher correlation than the defined threshold are selected. The following thresholds were defined:  $thr = [thr_1, \dots, thr_{max}] = [0.3, 0.5, 0.7, 0.85]$ . In this way, different input variables are selected according to the comparison of the correlation value and the threshold value. Thus, 4 datasets were built with different input variables for each variable to be predicted, except for the variable  $S_{NH_{10}}$  and  $BOD_{13}$ , where it was only possible to build 2 datasets since there were no input variables with a correlation above the thresholds of 0.7 and 0.85. In all, 36 datasets were built and tested for the different models. To carry out training and testing, all the datasets were divided into 70% for training and 30% for testing.

To determine the best configuration of each model, in terms of accuracy and interpretability, different combinations of model parameters were set:

- **FCM-LSM:** number of clusters  $N = [3, 5, 7, 10]$ , degree of fuzziness  $\eta = 2$ , number of iterations  $Max = 100$  and termination condition  $imp = 0.00001$ .
- **GAM-ZOTS:** number of fuzzy rules  $N = [2, 3, 5, 7, 10]$ , maximum number of iterations  $lim_{it} = 100$ , and termination condition  $\epsilon = 0.00001$ .
- **iMU-ZOTS:**  $M_{th}^j = [0.5, 0.7, 0.8, 0.9]$ ,  $\eta_j = |x_j^+ - x_j^-|/10$  (for  $j = 1, \dots, n$ ), maximum number of iterations  $lim_{it} = 100$ , and termination condition  $\epsilon = 0.00001$ .

---

**Algorithm 5** Setup of the proposed framework.
 

---

1. Define the set of variables to be predicted and their collecting points. In this work 10 ( $p_{max}$ ) variables were defined:  $TSS_3$ ,  $S_{NH_3}$ ,  $S_{NH_{10}}$ ,  $S_{NH_{13}}$ ,  $S_{NO_3}$ ,  $S_{NO_{10}}$ ,  $S_{NO_{13}}$ ,  $S_{O_{10}}$ ,  $BOD_{13}$  and  $COD_{13}$ , numbered from  $p = 1$  to  $p = 10$ , respectively;
2. Define the set of thresholds for the variables selection. In this work,  $max = 4$  thresholds  $thr_1, \dots, thr_{max}$  were defined;
3. Define the parameters of the models to be learned;
4. For each variable to be predicted,  $p = 1, \dots, p_{max}$ :
  - (a) Read data from the collected points related to variable  $p$ ;
  - (b) Obtain the correlation between input variables and variable  $p$ ;
  - (c) For each threshold,  $thr_i = thr_1, \dots, thr_{max}$ :
    - i. Select the input variables whose correlation is higher than the threshold  $thr_i$ ;
    - ii. Divide the dataset into training and test;
    - iii. For each model,  $model = 1, \dots, 3$  (models FCM-LSM, GAM-ZOTS, and iMU-ZOTS, respectively)
      - A. For each combination of model parameters,  $m_p = 1, \dots, H$ :
        - Learn the models, FCM-LSM, GAM-ZOTS and iMU-ZOTS;
        - Calculate the MSE of each model.
      - B. Save the best setup for each model of variable  $p$ .
  - (d) Save the best model for variable  $p$ .

---

To obtain the best predictions for each variable, several tests were performed for the different parameter combinations in each model. The best results were evaluated through the Mean Squared Error (MSE) metric described by the following expression:

$$MSE = \frac{1}{K} \sum_{k=1}^K (y_k - \hat{y}_k)^2, \quad (5.1)$$

where  $y_k$  and  $\hat{y}_k$  are the real and estimated target at the instant of time,  $k$ , and  $K$  is the total number of samples.

In Algorithm 5, it is presented the setup of the proposed framework. Firstly, the set of variables to be predicted, the set of thresholds for the input variable selection, and the parameters of the models to be learned are defined (Steps 1 to 3). Then, a cycle is made (Step 4) to go through each variable to be predicted and read its data, the data from the input and output points. Then, the correlation between the variable to be

estimated (variable  $p$ ) and the input variables is calculated (Step 4b). The correlation coefficient is calculated using Pearson's correlation method. This method measures the linear relationship between two variables, varying between -1 and 1. A value of 1 represents a perfect positive correlation, a value of -1 a perfect negative correlation and a value of 0 means no correlation. This is followed by a cycle to run through the different threshold values and select the input variables whose correlation is higher than the threshold  $thr_i$  (Step 4ci). The dataset is divided into training and testing (Step 4cii). The dataset obtained is then run for each model and each combination of parameters. The best result for each model is then saved. In the end, the best result for each variable is saved.

## 5.2 Results

Table 5.1 shows the results of the predictions made by the various models presented, where it is possible to check the information from the tests carried out, namely the best error value for each dataset ("error" columns), the average error ("mean" columns), and the standard deviation ("std. dev" columns). The best result for each variable is shown in bold.

Taking into account the MSE error values in the Table 5.1, it can be seen that the FCM-LSM obtained the best error values for 9 variables, while iMU-ZOTS obtained the best error value for one variable,  $S_{NH_{10}}$ . It can therefore be concluded that FCM-LSM is the method that produces the lowest MSE error value for almost all variables. The GAM-ZOTS and iMU-ZOTS methods have higher error values when compared to FCM-LSM.

The correlation threshold value is an important parameter in this study, as it is through it that the input variables are selected to construct the model. For this study, 4 threshold values were used, and it was found that as this value increased, the number of variables selected decreased. When the results obtained are analyzed, it can be seen that the best results are generally obtained when the threshold value is minimal, where more input variables are selected.

Table 5.2 presents the number of rules corresponding to the MSE results presented in Table 5.1. In addition to the number of rules, the average number of rules for each method is shown and, in the case of iMU-ZOTS, the threshold value,  $M_{th}$  for which the best results were obtained, is also shown. In bold are the values of the rules for which the best result of MSE (in Table 5.1) was obtained. And, the best results obtained by the other two methods are underlined.

The analysis of the number of fuzzy rules used by each method is extremely important because it is through these that the interpretability of a model can be gauged. In the context of WWTPs, precise prediction of variables stands as a linchpin for ensuring their efficient operation. Inaccuracies in prediction can lead to squandered resources and compromised water quality. However, while accuracy remains a critical metric, it doesn't stand alone in significance. The crux lies in striking a balance between accuracy and interpretability. Complex models, while potentially accurate, pose challenges in comprehension, particularly for WWTP operators who require clear insights into the decision-making process.

The analysis of Table 5.1 and Table 5.2 shows that the FCM-LSM obtained the best error value MSE for most of the variables, but it has a large number of rules, which makes it a complex model and difficult to interpret. Finding a balance between accuracy and





Table 5.2: Fuzzy rule values for the 3 models. **Var.** represents the target variable, *thr* the threshold value, *N* the number of fuzzy rules, Mean rules the average value of the fuzzy rules and  $M_{th}$  the threshold value of the iMU-ZOTS model. The value of the rules for each variable for which the best result was obtained is shown in bold and the best value of the other models is shown in underlined.

Var.	<i>thr</i>	FCM-LSM		GAM-ZOTS		iMU-ZOTS		
		<i>N</i>	Mean rules	<i>N</i>	Mean rules	<i>N</i>	$M_{th}$	Mean Rules
$TSS_3$	0.3	<b>10</b>		<u>7</u>		<u>7.11</u>	0.5	
	0.5	10	10	10	9.25	6.5	0.5	6.43
	0.7	10		10		6.11	0.5	
	0.85	10		10		6	0.5	
$SNH_3$	0.3	5		2		6.66	0.9	
	0.5	3	3.5	2	2	<u>6.5</u>	0.9	7.16
	0.7	3		2		7.82	0.5	
	0.85	<b>3</b>		<u>2</u>		7.66	0.9	
$SNH_{10}$	0.3	10		<u>10</u>		<b>7.93</b>	0.5	
	0.5	<u>10</u>	10	10	10	8	0.8	7.97
	0.7	-		-		-	-	
	0.85	-		-		-	-	
$SNH_{13}$	0.3	<b>10</b>		7		<u>8.6</u>	0.5	
	0.5	7	5.75	<u>3</u>	7.5	8.83	0.5	8.86
	0.7	3		10		9	0.5	
	0.85	3		10		9	0.5	
$SN_{O_3}$	0.3	3		3		<u>7.92</u>	0.8	
	0.5	3	4.5	2	4.25	8.1	0.8	8.51
	0.7	5		<u>2</u>		9	0.9	
	0.85	<b>7</b>		10		9	0.5	
$SN_{O_{10}}$	0.3	<b>10</b>		10		<u>8.33</u>	0.9	
	0.5	3	4.75	<u>2</u>	7.5	8	0.8	8.08
	0.7	3		2		8	0.8	
	0.85	3		2		8	0.8	
$SN_{O_{13}}$	0.3	<b>10</b>		<u>7</u>		<u>8.75</u>	0.9	
	0.5	10	8.25	7	7	8.75	0.9	8.88
	0.7	10		7		9	0.7	
	0.85	3		7		9	0.7	
$SO_{10}$	0.3	<b>10</b>		10		<u>8.75</u>	0.8	
	0.5	5	8.75	<u>7</u>	7.75	9	0.5	8.94
	0.7	10		7		9	0.5	
	0.85	10		7		9	0.5	
$BOD_{13}$	0.3	<b>10</b>		<u>10</u>		7.35	0.9	
	0.5	10	8.25	10	8	6.85	0.8	5.86
	0.7	10		10		<u>6.25</u>	0.8	
	0.85	3		2		3	0.8	
$COD_{13}$	0.3	<b>10</b>		<u>10</u>		<u>7.76</u>	0.5	
	0.5	3	6.5	10	10	7	0.7	7.38
	0.7	-		-		-	-	
	0.85	-		-		-	-	

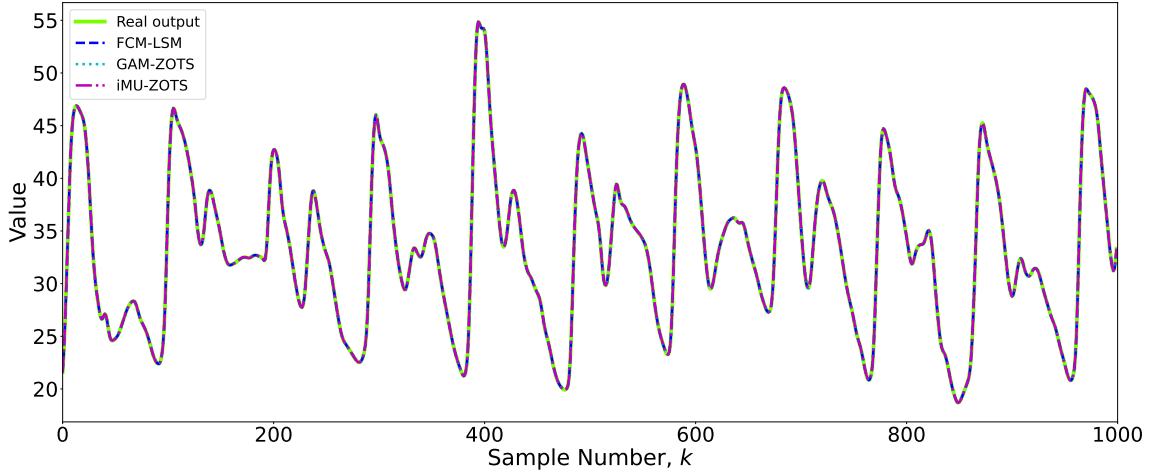


Figure 5.2: Prediction results for the variable  $S_{NH_3}$ .

interpretability is necessary to have efficient and less complex models.

To assess the accuracy and interpretability of the different models, the best results obtained by the 3 methods were analyzed. The results obtained by the methods that did not obtain the best error value for each variable are also chosen taking into account the balance between interpretability and accuracy. A less complex model with fewer rules is chosen if, at the same time, it has a low error value that is relatively close to the best value obtained for that variable. These MSE error values and the rules are shown underlined in Tables 5.1 and 5.2.

To make a more detailed analysis of the different methods and the balance between accuracy and interpretability, two variables were selected,  $S_{NH_3}$  and  $S_{NO_3}$ , as case studied. For the variable  $S_{NH_3}$ , a threshold value of 0.85 was used for the FCM-LSM and GAM-ZOTS methods and 0.5 for the iMU-ZOTS method. For the variable  $S_{NO_3}$ , a threshold value of 0.85 was used for FCM, 0.7 for GAM-ZOTS and 0.3 for iMU-ZOTS.

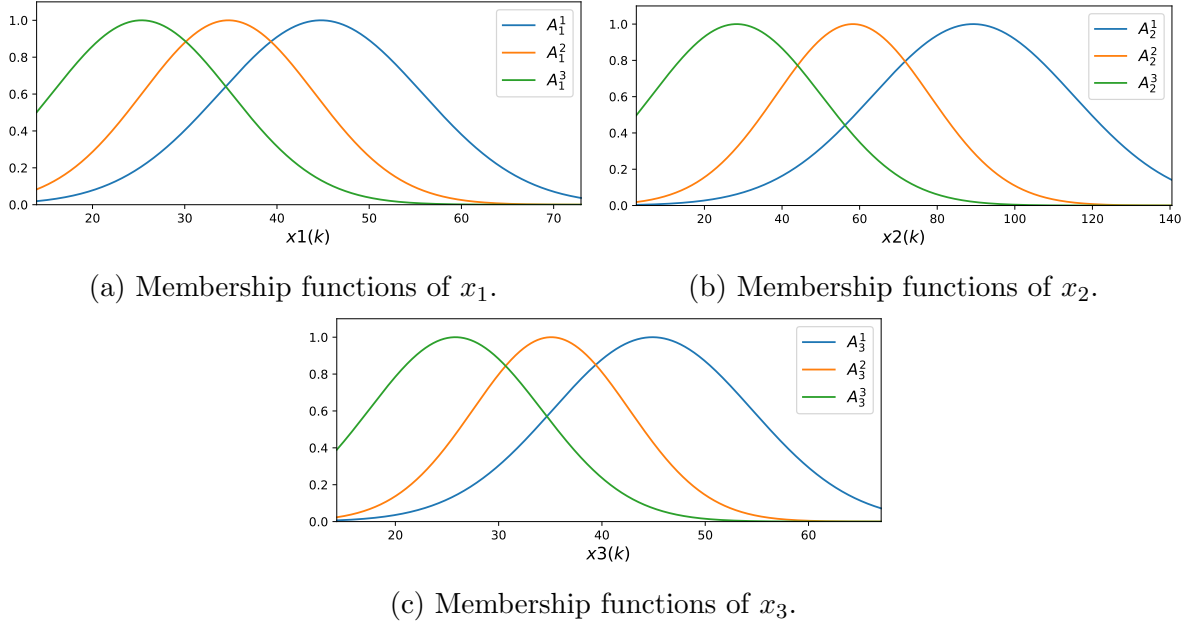
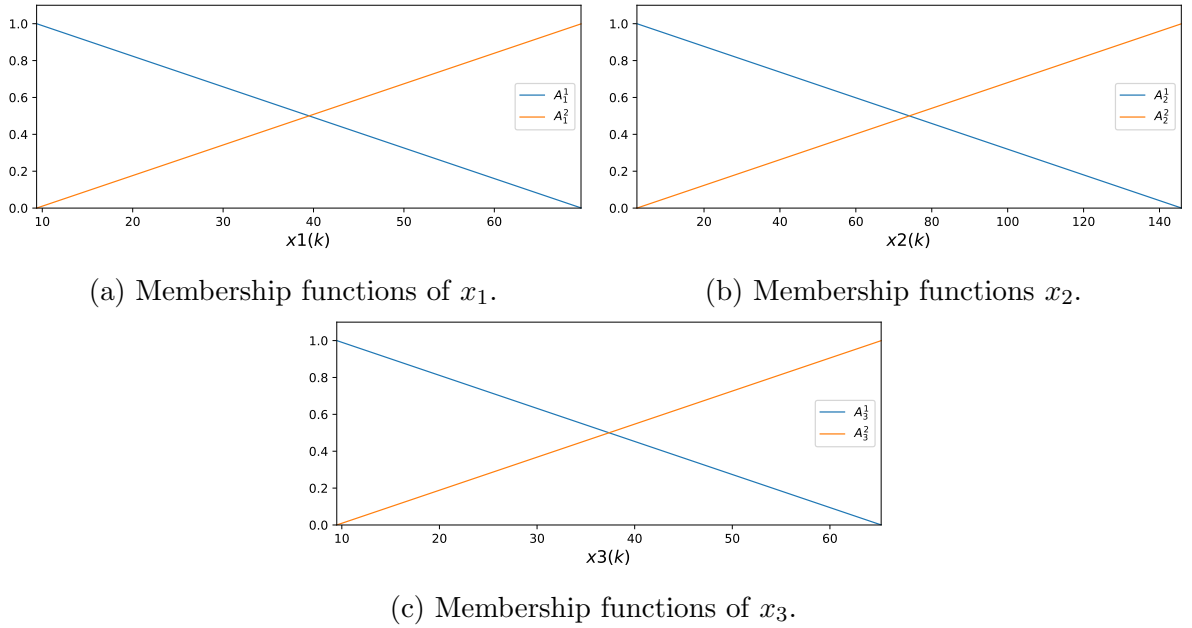
The results of the predictions made by the different methods for the variable  $S_{NH_3}$  are shown in Figure 5.2. By analyzing the figure, it is possible to see that the differences in forecasting between the different methods are minimal, with practically no significant difference. Since all the methods show good accuracy for this variable, is necessary to analyze the fuzzy rules used by each method.

The learned FCM-LSM for  $S_{NH_3}$  describes each of the 3 input variables by 3 fuzzy rules. Figure 5.3 shows the Membership Function (MF) for the 3 input variables. The fuzzy rules are described by:

$$\begin{aligned}
 R_1 : & \quad \text{IF } x_1(k) \text{ is } A_1^1 \text{ and } x_2(k) \text{ is } A_2^1 \text{ and } x_3(k) \text{ is } A_3^1 \\
 & \quad \text{THEN } y_1(k) = 1 + 1.37 \times 10^{-11}x_1(k) + 1.21 \times 10^{-13}x_2(k) + 4.26 \times 10^{-14}x_3(k), \\
 R_2 : & \quad \text{IF } x_1(k) \text{ is } A_1^2 \text{ and } x_2(k) \text{ is } A_2^2 \text{ and } x_3(k) \text{ is } A_3^2 \\
 & \quad \text{THEN } y_2(k) = 1 + 7.66 \times 10^{-12}x_1(k) - 1.97 \times 10^{-13}x_2(k) - 8.75 \times 10^{-14}x_3(k), \\
 R_3 : & \quad \text{IF } x_1(k) \text{ is } A_1^3 \text{ and } x_2(k) \text{ is } A_2^3 \text{ and } x_3(k) \text{ is } A_3^3 \\
 & \quad \text{THEN } y_3(k) = 1 + 5.69 \times 10^{-13}x_1(k) - 1.56 \times 10^{-13}x_2(k) - 1.88 \times 10^{-14}x_3(k).
 \end{aligned}$$

The GAM-ZOTS method describes each of the 3 selected input variables by 2 fuzzy rules. Figure 5.4 shows the MFs for the 3 input variables. The fuzzy rules for variable  $x_1$  are described by:

$$\begin{aligned}
 R_1^1 : & \quad \text{IF } x_1(k) \text{ is } A_1^1 \text{ THEN } y_1^1(k) = -2.85 \times 10^{-6}, \\
 R_1^2 : & \quad \text{IF } x_1(k) \text{ is } A_1^2 \text{ THEN } y_1^2(k) = 4.10 \times 10^{-6},
 \end{aligned}$$

Figure 5.3: Membership functions of the FCM-LSM for the variable  $S_{NH_3}$ .Figure 5.4: Membership functions of the GAM-ZOTS method for the variable  $S_{NH_3}$ .

for the variable  $x_2$  by:

$$\begin{aligned} R_2^1: & \text{ IF } x_2(k) \text{ is } A_2^1 \text{ THEN } y_2^1(k) = -1.32 \times 10^{-5}, \\ R_2^2: & \text{ IF } x_2(k) \text{ is } A_2^2 \text{ THEN } y_2^2(k) = 2.23 \times 10^{-5}, \end{aligned}$$

and for the variable  $x_3$  by:

$$\begin{aligned} R_3^1: & \text{ IF } x_3(k) \text{ is } A_3^1 \text{ THEN } y_3^1(k) = -24.81, \\ R_3^2: & \text{ IF } x_3(k) \text{ is } A_3^2 \text{ THEN } y_3^2(k) = 30.97. \end{aligned}$$

The best model from iMU-ZOTS method is described by 16 selected input variables by on average 6.5 rules by input variable. This number of average fuzzy rules is the sum of

the rules for all the variables divided by the number of variables. This model is composed of a total of  $N = 104$  fuzzy rules, where for each the 16 variables, the numbers of rules is  $N_1 = 6$ ,  $N_2 = 7$ ,  $N_3 = 6$ ,  $N_4 = 9$ ,  $N_5 = 8$ ,  $N_6 = 8$ ,  $N_7 = 9$ ,  $N_8 = 6$ ,  $N_9 = 7$ ,  $N_{10} = 5$ ,  $N_{11} = 3$ ,  $N_{12} = 4$ ,  $N_{13} = 8$ ,  $N_{14} = 8$ ,  $N_{15} = 6$ , and  $N_{16} = 4$ . Figures 5.5 and 5.6 show the MFs for the 16 input variables. The fuzzy rules for each input variable are shown in Appendix B.

The other variable analyzed is  $S_{NO_3}$ . The results of the predictions of  $S_{NO_3}$  made by the 3 methods are shown in Figure 5.7. It can be seen that all the methods obtain good results, with no major differences between them, which leads to the conclusion that accuracy can be achieved by any of them. Therefore, since accuracy can be achieved by any of the methods, it is necessary to analyze interpretability by analyzing the number of fuzzy rules used by each method.

The learned FCM-LSM is described by 4 input variables and 7 fuzzy rules. Figure 5.8 shows the MFs for the 4 input variables.

The fuzzy rules are described by:

$$\begin{aligned}
 R_1 : \quad & \text{IF } x_1(k) \text{ is } A_1^1 \text{ and } x_2(k) \text{ is } A_2^1 \text{ and } x_3(k) \text{ is } A_3^1 \text{ and } x_4(k) \text{ is } A_4^1 \\
 & \text{THEN } y_1(k) = 1 - 4.28 \times 10^{-13}x_1(k) + 2.26 \times 10^{-13}x_2(k) + 4.92 \times 10^{-15}x_3(k) \\
 & \quad + 1.67 \times 10^{-12}x_4(k), \\
 R_2 : \quad & \text{IF } x_1(k) \text{ is } A_1^2 \text{ and } x_2(k) \text{ is } A_2^2 \text{ and } x_3(k) \text{ is } A_3^2 \text{ and } x_4(k) \text{ is } A_4^2 \\
 & \text{THEN } y_2(k) = 1 + 7.88 \times 10^{-15}x_1(k) + 3.03 \times 10^{-14}x_2(k) + 3.29 \times 10^{-17}x_3(k) \\
 & \quad - 1.50 \times 10^{-13}x_4(k), \\
 R_3 : \quad & \text{IF } x_1(k) \text{ is } A_1^3 \text{ and } x_2(k) \text{ is } A_2^3 \text{ and } x_3(k) \text{ is } A_3^3 \text{ and } x_4(k) \text{ is } A_4^3 \\
 & \text{THEN } y_3(k) = 1 - 1.75 \times 10^{-13}x_1(k) + 4.85 \times 10^{-14}x_2(k) + 1.79 \times 10^{-15}x_3(k) \\
 & \quad + 4.23 \times 10^{-13}x_4(k), \\
 R_4 : \quad & \text{IF } x_1(k) \text{ is } A_1^4 \text{ and } x_2(k) \text{ is } A_2^4 \text{ and } x_3(k) \text{ is } A_3^4 \text{ and } x_4(k) \text{ is } A_4^4 \\
 & \text{THEN } y_4(k) = 1 + 1.33 \times 10^{-15}x_1(k) - 3.34 \times 10^{-14}x_2(k) - 1.02 \times 10^{-16}x_3(k) \\
 & \quad + 3.10 \times 10^{-13}x_4(k), \\
 R_5 : \quad & \text{IF } x_1(k) \text{ is } A_1^5 \text{ and } x_2(k) \text{ is } A_2^5 \text{ and } x_3(k) \text{ is } A_3^5 \text{ and } x_4(k) \text{ is } A_4^5 \\
 & \text{THEN } y_5(k) = 1 - 4.29 \times 10^{-14}x_1(k) + 1.31 \times 10^{-13}x_2(k) + 1.68 \times 10^{-15}x_3(k) \\
 & \quad + 9.22 \times 10^{-13}x_4(k), \\
 R_6 : \quad & \text{IF } x_1(k) \text{ is } A_1^6 \text{ and } x_2(k) \text{ is } A_2^6 \text{ and } x_3(k) \text{ is } A_3^6 \text{ and } x_4(k) \text{ is } A_4^6 \\
 & \text{THEN } y_6(k) = 1 - 2.97 \times 10^{-13}x_1(k) + 3.37 \times 10^{-13}x_2(k) + 5.68 \times 10^{-15}x_3(k) \\
 & \quad + 2.03 \times 10^{-12}x_4(k), \\
 R_7 : \quad & \text{IF } x_1(k) \text{ is } A_1^7 \text{ and } x_2(k) \text{ is } A_2^7 \text{ and } x_3(k) \text{ is } A_3^7 \text{ and } x_4(k) \text{ is } A_4^7 \\
 & \text{THEN } y_7(k) = 1 - 5.02 \times 10^{-13}x_1(k) + 3.79 \times 10^{-13}x_2(k) + 7.09 \times 10^{-15}x_3(k) \\
 & \quad + 3.01 \times 10^{-12}x_4(k).
 \end{aligned}$$

The GAM-ZOTS method describes each of the 7 selected input variables by 2 fuzzy rules. Figure 5.9 shows the MFs for the 7 input variables.

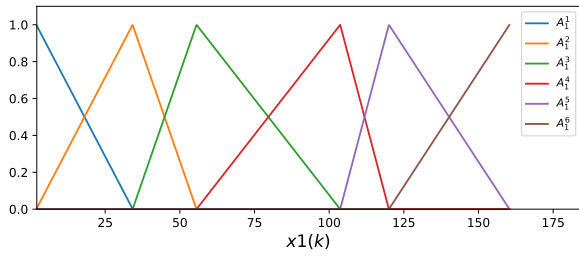
The fuzzy rules of GAM-ZOTS for the variable  $x_1$  are described by:

$$\begin{aligned}
 R_1^1 : \quad & \text{IF } x_1(k) \text{ is } A_1^1 \text{ THEN } y_1^1(k) = 0.02, \\
 R_2^1 : \quad & \text{IF } x_1(k) \text{ is } A_1^2 \text{ THEN } y_1^2(k) = -0.06,
 \end{aligned}$$

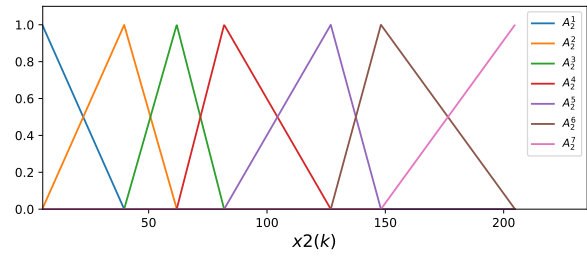
for the variable  $x_2$  by:

$$\begin{aligned}
 R_1^2 : \quad & \text{IF } x_2(k) \text{ is } A_2^1 \text{ THEN } y_2^1(k) = 0.03, \\
 R_2^2 : \quad & \text{IF } x_2(k) \text{ is } A_2^2 \text{ THEN } y_2^2(k) = -0.13,
 \end{aligned}$$

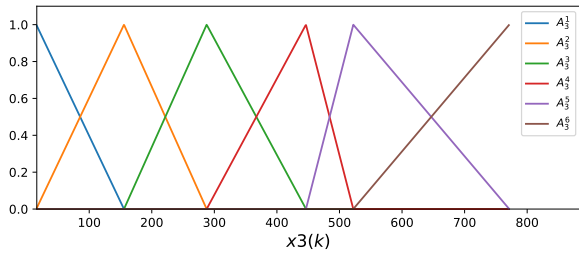
for the variable  $x_3$  by:



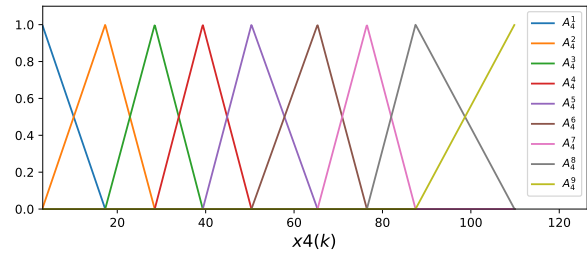
(a) Membership functions of  $x_1$ .



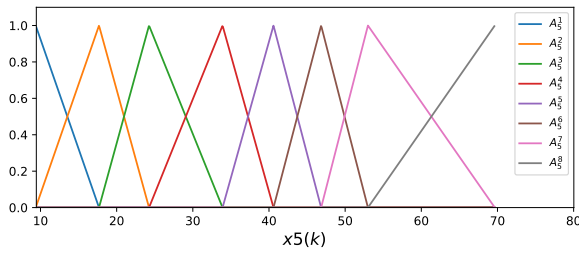
(b) Membership functions of  $x_2$ .



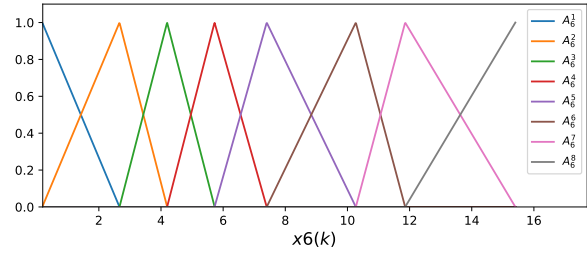
(c) Membership functions of  $x_3$ .



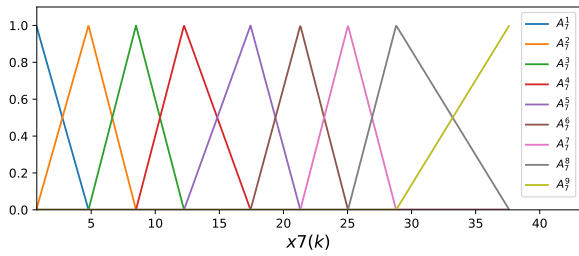
(d) Membership functions of  $x_4$ .



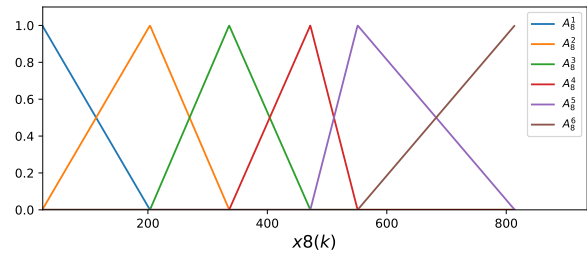
(e) Membership functions of  $x_5$ .



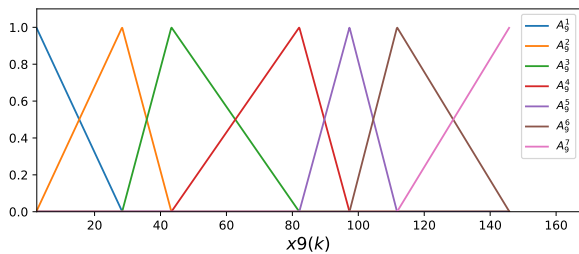
(f) Membership functions of  $x_6$ .



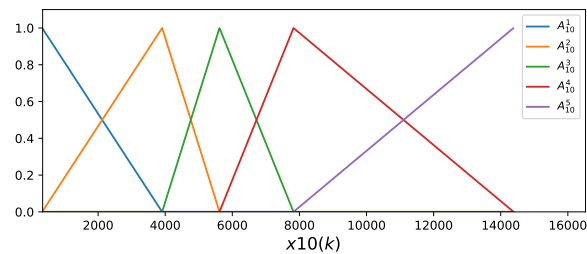
(g) Membership functions of  $x_7$ .



(h) Membership functions of  $x_8$ .



(i) Membership functions of  $x_9$ .



(j) Membership functions of  $x_{10}$ .

Figure 5.5: Membership functions of the iMU-ZOTS method for the variable  $S_{NH_3}$ .

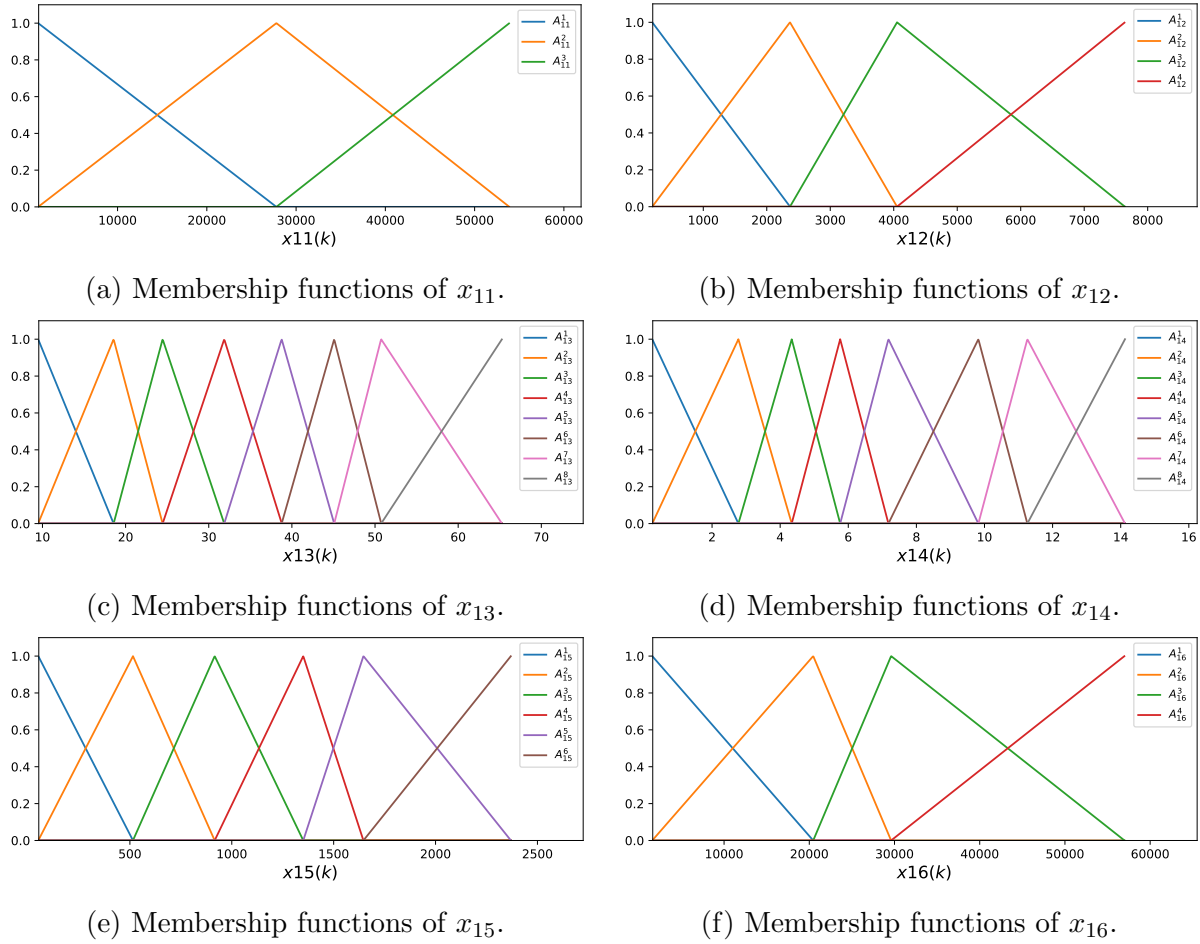


Figure 5.6: Membership functions of the iMU-ZOTS method for the variable  $S_{NH_3}$ .

$$\begin{aligned} R_3^1: & \text{ IF } x_3(k) \text{ is } A_3^1 \text{ THEN } y_3^1(k) = -0.05, \\ R_3^2: & \text{ IF } x_3(k) \text{ is } A_3^2 \text{ THEN } y_3^2(k) = 0.18, \end{aligned}$$

for the variable  $x_4$  by:

$$\begin{aligned} R_4^1: & \text{ IF } x_4(k) \text{ is } A_4^1 \text{ THEN } y_4^1(k) = 0.04, \\ R_4^2: & \text{ IF } x_4(k) \text{ is } A_4^2 \text{ THEN } y_4^2(k) = -0.01, \end{aligned}$$

for the variable  $x_5$  by:

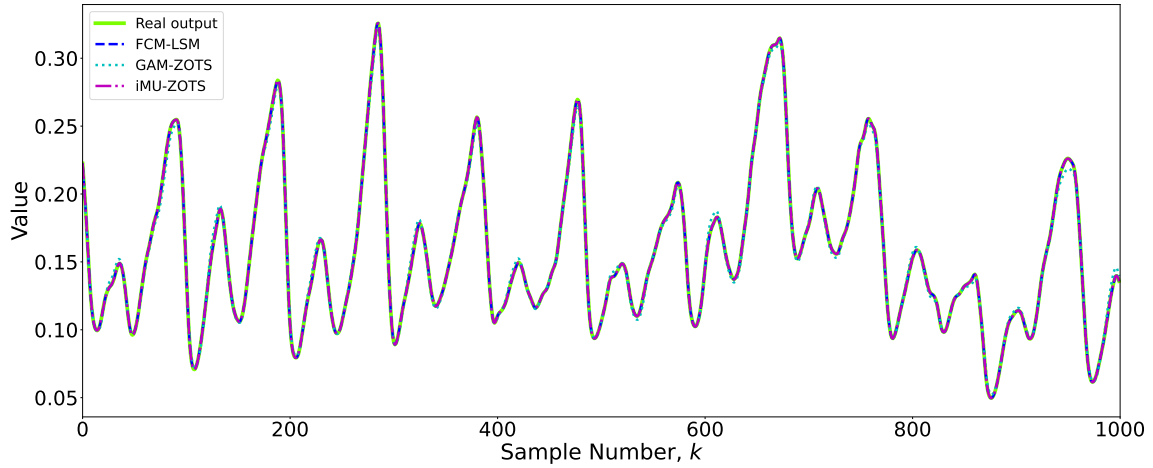
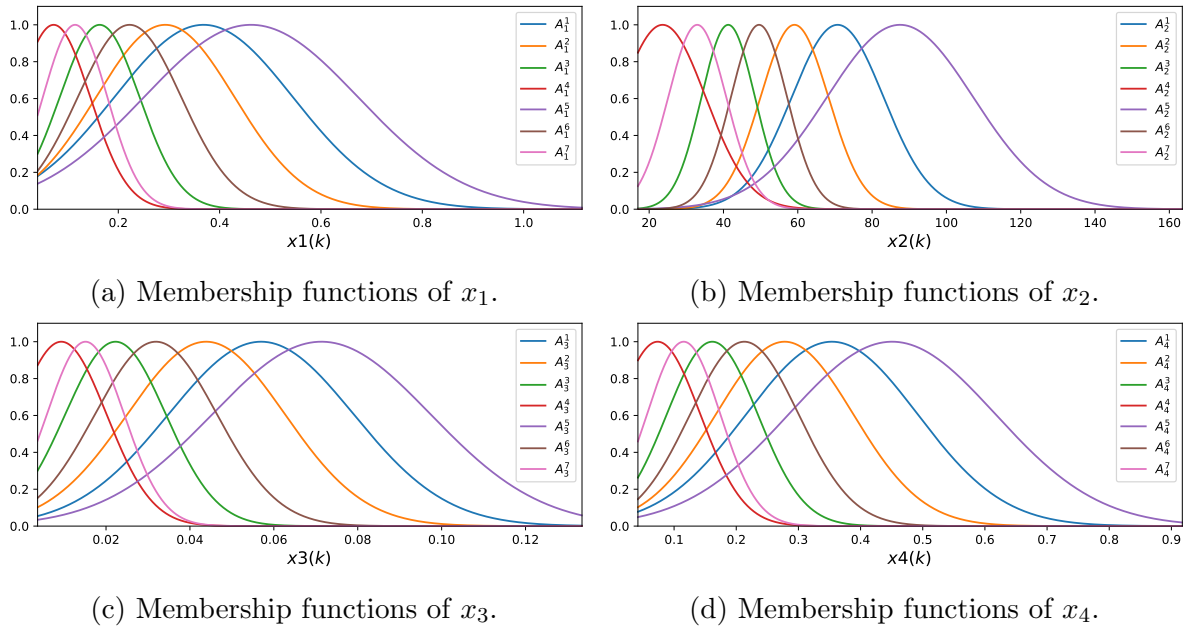
$$\begin{aligned} R_5^1: & \text{ IF } x_5(k) \text{ is } A_5^1 \text{ THEN } y_5^1(k) = -0.03, \\ R_5^2: & \text{ IF } x_5(k) \text{ is } A_5^2 \text{ THEN } y_5^2(k) = 0.07, \end{aligned}$$

for the variable  $x_6$  by:

$$\begin{aligned} R_6^1: & \text{ IF } x_6(k) \text{ is } A_6^1 \text{ THEN } y_6^1(k) = -0.03, \\ R_6^2: & \text{ IF } x_6(k) \text{ is } A_6^2 \text{ THEN } y_6^2(k) = 0.11, \end{aligned}$$

and for the variable  $x_7$  by:

$$\begin{aligned} R_7^1: & \text{ IF } x_7(k) \text{ is } A_7^1 \text{ THEN } y_7^1(k) = -0.12, \\ R_7^2: & \text{ IF } x_7(k) \text{ is } A_7^2 \text{ THEN } y_7^2(k) = 0.33. \end{aligned}$$

Figure 5.7: Prediction results for the variable  $S_{NO_3}$ .Figure 5.8: Membership functions of the FCM-LSM for the variable  $S_{NO_3}$ .

Finally, the iMU-ZOTS model is described by 13 selected input variables and by an average of 7.92 fuzzy rules per input variable. This average number of fuzzy rules is the sum of the number of rules for all the variables divided by the number of variables. This model is made up of a total of  $N = 103$  fuzzy rules, namely  $N_1 = 9$ ,  $N_2 = 9$ ,  $N_3 = 9$ ,  $N_4 = 9$ ,  $N_5 = 5$ ,  $N_6 = 4$ ,  $N_7 = 10$ ,  $N_8 = 7$ ,  $N_9 = 8$ ,  $N_{10} = 10$ ,  $N_{11} = 10$ ,  $N_{12} = 7$  and  $N_{13} = 6$ .

Figures 5.10 and 5.11 shows the MFs for the 13 input variables. The fuzzy rules for each input variable are shown in Appendix C.

After analyzing the results obtained for the above two target variables, it can be seen that although the FCM-LSM has a lower error value than the other methods, the prediction difference is insignificant. In addition, it can be seen that FCM-LSM obtains the best results, but with more complex rules compared to the other methods. A model with a high number of rules or with several variables on the antecedent part becomes complex and difficult to interpret by the operators, who in these cases of WWTPs need

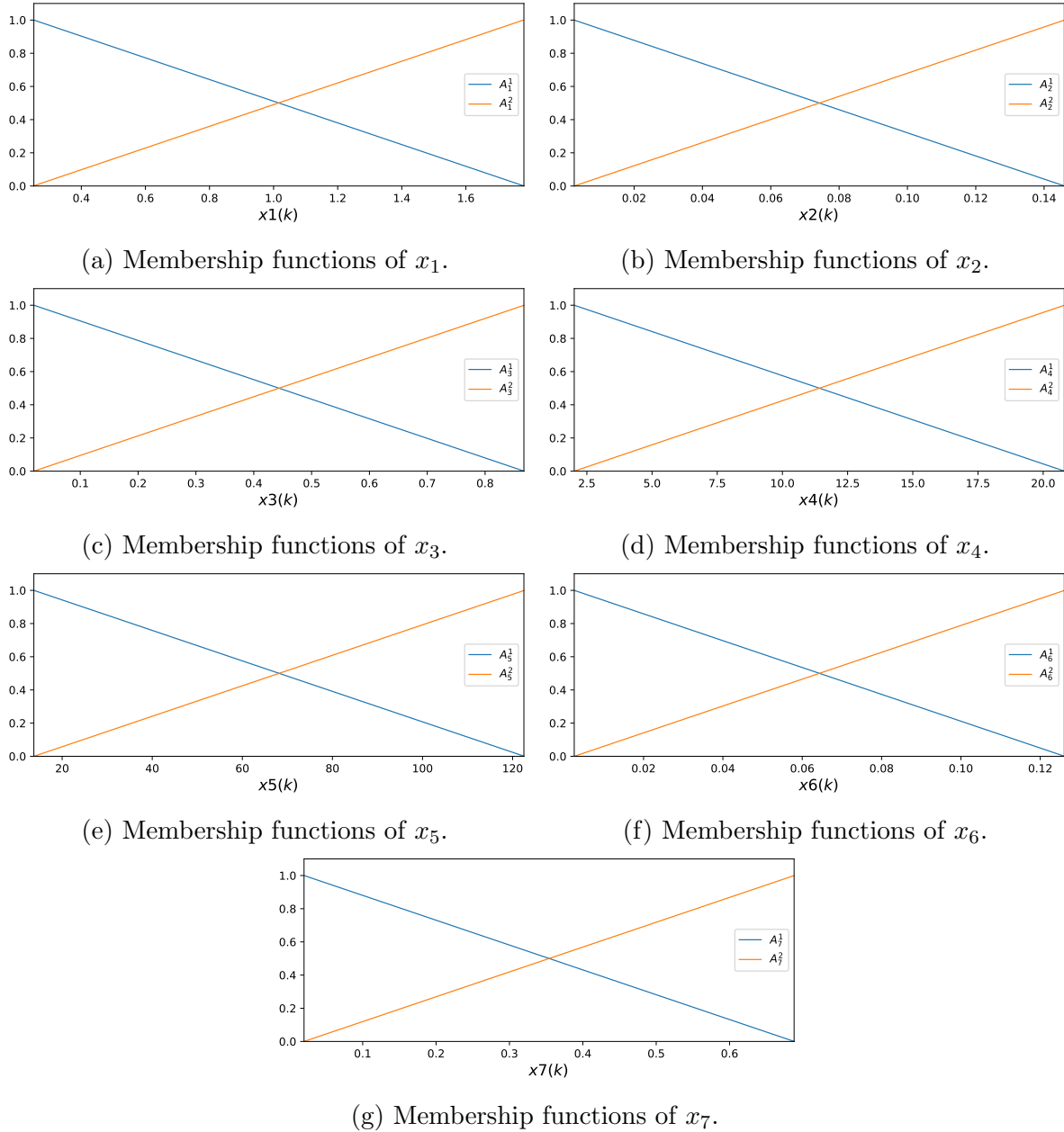
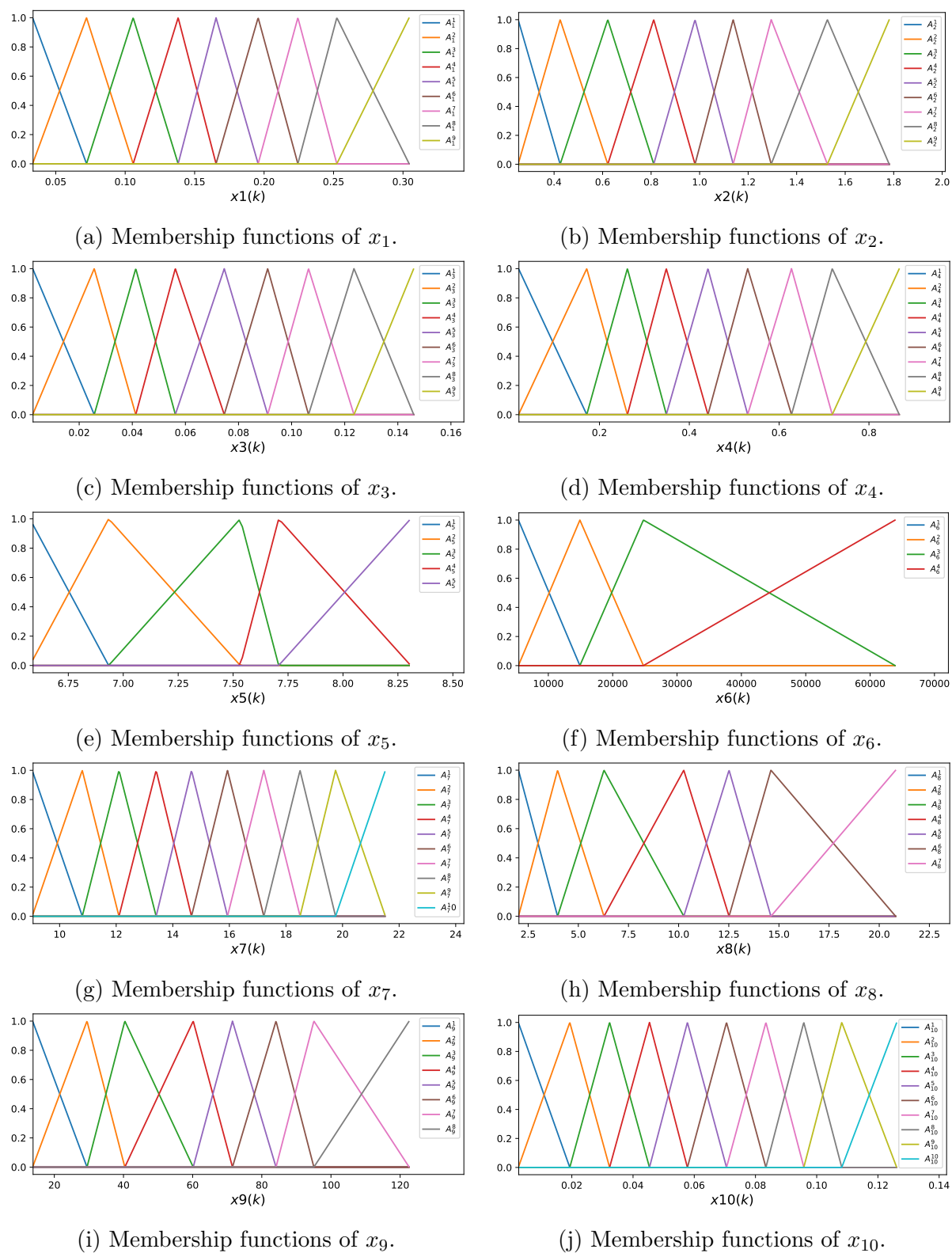


Figure 5.9: Membership functions of the GAM-ZOTS method for the variable  $S_{NO_3}$ .

to understand the models easily to make the right decisions throughout the wastewater treatment process. An interpretable system allows the actions taken in a WWTP to be explained and easily adjusted to requirements.

For the variable  $S_{NH_3}$ , the FCM-LSM obtains an error value of  $MSE=1.06 \times 10^{-23}$  with 3 associated fuzzy rules. GAM-ZOTS obtains an error value of  $MSE=7.90 \times 10^{-12}$ ,  $7.46 \times 10^{13}\%$  higher than the FCM-LSM error, with 2 univariate fuzzy rules per variable. iMU-ZOTS has an error value of  $MSE=1.28 \times 10^{-4}$ ,  $1.21 \times 10^{21}\%$  higher than the FCM-LSM error and 6.5 fuzzy rules per variable. Since the forecast made by the 3 methods is identical (see Figure 5.2), the choice of the best method involves the one with the least complexity, which for this variable is GAM-ZOTS. For this variable in particular, FCM-LSM would also be a possible right choice since it has the lowest error and a low number of fuzzy rules.



Figure 5.10: Membership functions of the iMU-ZOTS method for the variable  $S_{NO_3}$ .

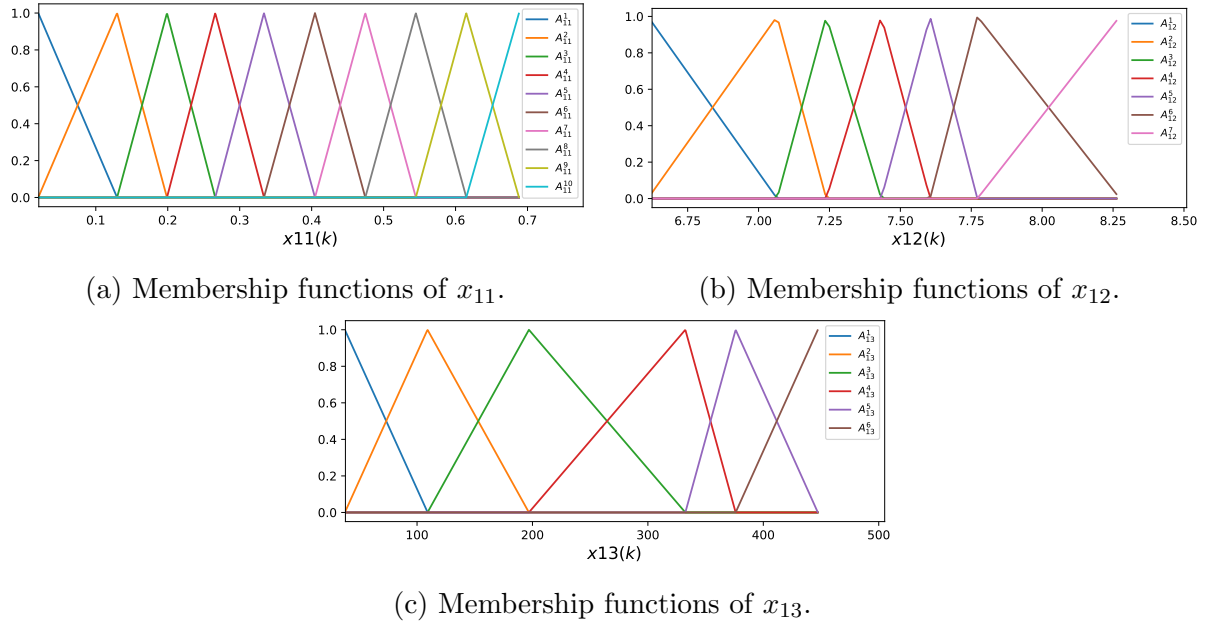


Figure 5.11: Membership functions of the iMU-ZOTS method for the variable  $S_{NO_3}$ .

In the case of the  $S_{NO_3}$  variable, FCM-LSM obtains an error value of  $MSE=2.92 \times 10^{-28}$ , and 7 fuzzy rules. GAM-ZOTS obtains an error value of  $MSE=1.16 \times 10^{-5}$ ,  $3.97 \times 10^{24}\%$  higher than the FCM-LSM error and 2 fuzzy rules per variable. iMU-ZOTS has an error value of  $MSE=2.49 \times 10^{-9}$ ,  $8.53 \times 10^{20}\%$  higher than the FCM-LSM error and 7.92 fuzzy rules per variable. An analysis of Figure 5.7 shows that the differences in the prediction of this variable by the different methods are insignificant, which means that the most appropriate method for this variable should be chosen based on the least complex model. In the case of this variable, GAM-ZOTS is the one with the best balance between accuracy and interpretability, allowing for accurate predictions that are easy for operators to understand.



# Chapter 6

## Conclusions and Future Work

The main aim of this work is to create an interpretable Digital Twin (DT) of the most relevant variables on the waterline of a Wastewater Treatment Plant (WWTP). Through DTs, it is possible to create a replica of these treatment plants, allowing variables to be monitored throughout the process. This makes it possible to understand the system's behaviour better and help operators make the right decisions.

Firstly, Chapter 2 surveys the literature on DTs, explaining the concepts of this innovative technology, its history and the main applications in various areas. The evolution of this technology and its potential for improving industrial processes is described.

Chapter 3 explains the WWTPs and how they work, describing the main stages that water has to go through to be properly treated. To evaluate all the processes in the WWTPs, the Benchmark Simulation Model 2 (BSM2) was used, simulating all the stages of the water treatment process. Various variables can be monitored throughout the treatment process. For this dissertation, 10 variables were predicted, the data for which was collected at various points along the waterline.

Three fuzzy-based models were used to predict the variables, namely Model based on Fuzzy C-Means and Least Squares Method (FCM-LSM), Generalized additive models using zero-order T-S fuzzy systems (GAM-ZOTS) model and Iterative Learning of multiple univariate zero-order T-S fuzzy systems (iMU-ZOTS) model. The Fuzzy Logic Systems (FLS) are used in various industrial systems and are potent tools capable of dealing with nonlinear problems with a high degree of uncertainty and imprecision, making it possible to obtain simpler and easier-to-understand models. Since WWTPs are very complex structures, FLS make it possible to create easy models for operators to understand and interpret, enabling easier and safer decision-making and guaranteeing the optimization of the entire process.

To predict each of the variables, several datasets were constructed for each target variable, using different sets of input variables. These datasets were used to learn each one of the three models, which use different combinations of model parameters, and in the end, the best results from each algorithm were selected for each variable. The results were evaluated using the Mean Squared Error (MSE) metric, and it was found that FCM-LSM was the model that obtained the best accuracy results for most of the variables. Although this model obtained the best results, it is necessary to evaluate the complexity of the models, since a complex model becomes difficult for operators to understand, who in the case of WWTPs need to evaluate the results to make decisions. The number of fuzzy rules can assess the complexity of a model. Analyzing the fuzzy rules of the FCM-LSM model shows that it has a more complex structure than the other models, despite being

the model with the best error result. To make a more detailed analysis of the predictions made by each model, 2 variables, as case studies, were selected, where the prediction errors and the fuzzy rules obtained by the 3 models were analyzed. After this analysis, it was found that the differences in the error obtained by the 3 models were minimal and not very significant. Analyzing the fuzzy rules of each model showed that there was a less complex and more interpretable model than the one that obtained the best error value. It can, therefore, be concluded that the choice of the best model for predicting each variable must take into account not only the accuracy but also the interpretability and complexity of each model. As the complexity of the model increases, it becomes more difficult to understand and interpret. For the variables analyzed, the GAM-ZOTS method showed the best balance between accuracy and interpretability.

Another conclusion that can be drawn from this study is that the choice of the best method that presents the best balance between accuracy and interpretability varies from variable to variable, a model can adapt well to some variables and not to others. It is necessary to analyze which method is best, variable by variable.

To improve this work, some improvements can be made in future studies. One of these improvements could be improving the algorithms used to handle the data from the WWTPs better. Another improvement that could be made is the creation of a generic algorithm that adapts well to all variables and guarantees a balance between accuracy and interpretability, thus making the prediction of each variable easier to understand for treatment plant operators, without compromising the reliability of the results.

# Bibliography

- J. Alex, J.-F. Beteau, J. Copp, C. Hellings, U. Jeppsson, S. Marsili-Libelli, M. Pons, H. Spanjers, and H. Vanhooren. Benchmark for evaluating control strategies in wastewater treatment plants. In *1999 European Control Conference (ecc)*, pages 3746–3751. IEEE, 1999.
- D. J. Batstone, J. Keller, I. Angelidaki, S. Kalyuzhnyi, S. Pavlostathis, A. Rozzi, W. Sanders, H. a. Siegrist, and V. Vavilin. The iwa anaerobic digestion model no 1 (adm1). *Water Science and technology*, 45(10):65–73, 2002.
- A. Boretti and L. Rosa. Reassessing the projections of the world water development report. *NPJ Clean Water*, 2(1):15, 2019.
- D. Botkina, M. Hedlind, B. Olsson, J. Henser, and T. Lundholm. Digital twin of a cutting tool. *Procedia CIRP*, 72:215–218, 2018. ISSN 2212-8271. doi: 10.1016/j.procir.2018.03.178. 51st CIRP Conference on Manufacturing Systems.
- B. Brenner and V. Hummel. Digital twin as enabler for an innovative digital shopfloor management system in the esb logistics learning factory at reutlingen-university. *Procedia Manufacturing*, 9:198–205, 2017.
- A. Canedo. Industrial iot lifecycle via digital twins. In *Proceedings of the Eleventh IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis*, pages 1–1, 2016.
- A. I. R. CARD. Wastewater, asce, usa, tech. rep., 2021. url<https://infrastructurereportcard.org/cat-item/wastewater/>, 2021.
- W. Chen, C. Yao, and X. Lu. Optimal design activated sludge process by means of multi-objective optimization: case study in benchmark simulation model 1 (bsm1). *Water science and technology*, 69(10):2052–2058, 2014.
- A. B. d. COSTA et al. Desenvolvimento de metamodelos kriging e otimização de uma planta de tratamento de efluentes (bsm2). Universidade Federal de Campina Grande, Tese de Doutorado, 2016.
- B. de Matos, R. Salles, J. Mendes, J. R. Gouveia, A. J. Baptista, and P. Moura. A review of energy and sustainability kpi-based monitoring and control methodologies on wwtps. *Mathematics*, 11(1), 2023. doi: 10.3390/math11010173.
- A. C. do Amaral, R. L. R. Steinmetz, A. Kunz, A. C. DO AMARAL, K. AIRTON, et al. O processo de biodigestão. *Embrapa Suínos e Aves-Capítulo em livro científico (ALICE)*, page 26, 2019.

- D. Dovžan and I. Škrjanc. Recursive fuzzy c-means clustering for recursive fuzzy identification of time-varying processes. *ISA transactions*, 50(2):159–169, 2011.
- G. R. Energy. Digital wind farm: the next evolution of wind energy, 2016.
- T. Erol, A. F. Mendi, and D. Doğan. The digital twin revolution in healthcare. In *2020 4th international symposium on multidisciplinary studies and innovative technologies (ISMSIT)*, pages 1–7. IEEE, 2020.
- A. EurEau. Europe’s water in figures—an overview of the european drinking water and waste water sectors, 2017.
- X. Flores-Alsina, D. Ikumi, D. Batstone, K. V. Gernaey, C. Brouckaert, G. A. Ekama, and U. Jeppsson. Towards a plant-wide benchmark simulation model with simultaneous nitrogen and phosphorus removal wastewater treatment processes. In *Proceedings of the IWA 8th World Water Congress and Exhibition, Busan, South Korea*, pages 16–21, 2012.
- T. Gabor, L. Belzner, M. Kiermeier, M. T. Beck, and A. Neitz. A simulation-based architecture for smart cyber-physical systems. In *2016 IEEE international conference on autonomic computing (ICAC)*, pages 374–379. IEEE, 2016.
- M. Garrido-Baserba, A. Hospido, R. Reif, M. Molinos-Senante, J. Comas, and M. Poch. Including the environmental criteria when selecting a wastewater treatment plant. *Environmental modelling & software*, 56:74–82, 2014.
- M. Garrido-Baserba, M. Molinos-Senante, J. Abelleira-Pereira, L. Fdez-Güelfo, M. Poch, and F. Hernández-Sancho. Selecting sewage sludge treatment alternatives in modern wastewater treatment plants using environmental decision support systems. *Journal of Cleaner Production*, 107:410–419, 2015.
- D. GL. Digital twins at work. Available: <https://www.dnvgl.com/feature/digital-twins.html>, 2019.
- E. Glaessgen and D. Stargel. The digital twin paradigm for future nasa and us air force vehicles. In *53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA*, page 1818, 2012.
- M. Gonçalves, P. Sousa, J. Mendes, M. Danishvar, and A. Mousavi. Real-time event-driven learning in highly volatile systems: A case for embedded machine learning for scada systems. *IEEE Access*, 10:50794–50806, 2022. doi: 10.1109/ACCESS.2022.3173376.
- M. Grieves. Digital twin: manufacturing excellence through virtual factory replication. *White paper*, 1(2014):1–7, 2014.
- H. Haimi, M. Mulas, K. Sahlstedt, and R. Vahala. Advanced operation and control methods of municipal wastewater treatment processes in finland. *Water and Wastewater Engineering, Helsinki University of Technology. Helsinki, Finland*, page 25, 2009.

- S. M. Hallaji, Y. Fang, and B. K. Winfrey. A digital twin framework for enhancing predictive maintenance of pumps in wastewater treatment plants. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 38, pages 88–93. IAARC Publications, 2021.
- H. Hauduc, L. Rieger, T. Ohtsuki, A. Shaw, I. Takács, S. Winkler, A. Héduit, P. Vanrolleghem, and S. Gillot. Activated sludge modelling: development and potential use of a practical applications database. *Water Science and Technology*, 63(10):2164–2182, 2011.
- O. S. M. Júnior, C. S. Nunes, J. V. S. Verçosa, M. G. de Souza Silveira, F. I. J. Cunha, L. A. d. F. A. Pereira, and M. J. R. Soares. O porquê da temática água na conferência e sua utilização no mundo: Brasil. In *Trajetória da Conferência Nacional Infantojuvenil pelo Meio Ambiente na Fronteira Oeste-RS está*, page 27, 2021.
- J. S. S. Júnior, J. Mendes, F. Souza, and C. Premebida. Survey on deep fuzzy systems in regression applications: A view on interpretability. *International Journal of Fuzzy Systems*, 25(7):2568–2589, 2023. doi: 10.1007/s40815-023-01544-8.
- G. Knapp, T. Mukherjee, J. Zuback, H. Wei, T. Palmer, A. De, and T. DebRoy. Building blocks for a digital twin of additive manufacturing. *Acta Materialia*, 135:390–399, 2017.
- E. LaGrange. Developing a digital twin: The roadmap for oil and gas optimization. In *SPE Offshore Europe Conference and Exhibition*. OnePetro, 2019.
- C. Li, S. Mahadevan, Y. Ling, S. Choze, and L. Wang. Dynamic bayesian network for aircraft wing health monitoring digital twin. *Aiaa Journal*, 55(3):930–941, 2017.
- F. Licciardello, M. Milani, S. Consoli, N. Pappalardo, S. Barbagallo, and G. Cirelli. Wastewater tertiary treatment options to match reuse standards in agriculture. *Agricultural Water Management*, 210:232–242, 2018.
- F. Liu, A. Ouedraogo, S. Manghee, and A. Danilenko. A primer on energy efficiency for municipal water and wastewater utilities. Technical report, World Bank, Washington, DC, 2012.
- W. Liu, S. He, J. Mou, T. Xue, H. Chen, and W. Xiong. Digital twins-based process monitoring for wastewater treatment processes. *Reliability Engineering & System Safety*, page 109416, 2023.
- T. Marcal. Will digital twins improve healthcare for their real-world counterparts? Available: <https://www.ansys.com/blog/will-digital-twins-improve-healthcare-real-world-counterparts/>, 2018.
- M. G. Mayani, M. Svendsen, and S. Oedegaard. Drilling digital twin success stories the last 10 years. In *SPE Norway One Day Seminar*. OnePetro, 2018.
- D. McCannel. What is a digital twin?(plus 3 industries which already benefit), 2018.
- J. Mendes, F. Souza, R. Araújo, and N. Gonçalves. Genetic fuzzy system for data-driven soft sensors design. *Applied Soft Computing*, 12(10):3237–3245, 2012. doi: 10.1016/j.asoc.2012.05.009.



- J. Mendes, R. Maia, R. Araújo, and G. Gouveia. Intelligent controller for industrial processes applied to a distributed two-tank system. In *Proc. 1st IEEE International Conference on Artificial Intelligence for Industries (AI4I 2018)*, pages 39–43, Laguna Hills, California, USA, September 26–28 2018. IEEE. doi: 10.1109/AI4I.2018.8665701.
- J. Mendes, F. Souza, R. Araújo, and S. Rastegar. Neo-fuzzy neuron learning using backfitting algorithm. *Neural Computing and Applications*, 31:3609–3618, 2019a. doi: 10.1007/s00521-017-3301-4.
- J. Mendes, F. A. Souza, R. Maia, and R. Araújo. Iterative learning of multiple univariate zero-order ts fuzzy systems. In *IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society*, volume 1, pages 3803–3808. IEEE, 2019b. doi: 10.1109/IECON.2019.8927224.
- M. Meneses, H. Concepción, and R. Vilanova. Joint environmental and economical analysis of wastewater treatment plants control strategies: a benchmark scenario analysis. *Sustainability*, 8(4):360, 2016.
- Microsoft. The promise of a digital twin strategy. Available: <https://info.microsoft.com/rs/157-GQE-382/images/Microsoft%27s%20Digital%20Twin%20%27How-To%27%20Whitepaper.pdf>, 2019.
- S. Misra, C. Roy, T. Sauter, A. Mukherjee, and J. Maiti. Industrial internet of things for safety management applications: A survey. *IEEE Access*, 10:83415–83439, 2022. doi: 10.1109/ACCESS.2022.3194166.
- J. Mohr. Digital twins for the oil and gas industry. *Hashplay, San Francisco, CA, USA, Tech. Rep*, 2018.
- R. Moliner-Heredia, I. Peñarrocha-Alós, and R. Sanchis-Llopis. Economic model predictive control of wastewater treatment plants based on bsm1 using linear prediction models. In *2019 IEEE 15th International Conference on Control and Automation (ICCA)*, pages 73–78. IEEE, 2019.
- K. Morales-Rodelo, M. Francisco, H. Alvarez, P. Vega, and S. Revollar. Collaborative control applied to bsm1 for wastewater treatment plants. *Processes*, 8(11):1465, 2020.
- D. Nadhan, M. Mayani, and R. Rommetveit. Drilling with digital twins/iadc/spe asia pacific drilling technology conference and exhibition. *Bangkok, Thailand: Society of Petroleum Engineers*, page 18, 2018.
- P. Oliveira, B. Fernandes, C. Analide, and P. Novais. Forecasting energy consumption of wastewater treatment plants with a transfer learning approach for sustainable cities. *Electronics*, 10(10):1149, 2021.
- P. Parviainen, M. Tihinen, J. Kääriäinen, and S. Teppola. Tackling the digitalization challenge: how to benefit from digitalization in practice. *International journal of information systems and project management*, 5(1):63–77, 2017.
- T. Poddar. Digital twin bridging intelligence among man, machine and environment. In *Offshore Technology Conference Asia*, page D022S001R003. OTC, 2018.

- M. Preisner, E. Neverova-Dziopak, and Z. Kowalewski. An analytical review of different approaches to wastewater discharge standards with particular emphasis on nutrients. *Environmental Management*, 66:694–708, 2020.
- Q. Qi, F. Tao, Y. Zuo, and D. Zhao. Digital twin service towards smart manufacturing. *Procedia Cirp*, 72:237–242, 2018.
- Q. Qi, F. Tao, T. Hu, N. Anwer, A. Liu, Y. Wei, L. Wang, and A. Nee. Enabling technologies and tools for digital twin. *Journal of Manufacturing Systems*, 58:3–21, 2021.
- A. Rasheed, O. San, and T. Kvamsdal. Digital twin: Values, challenges and enablers. *arXiv preprint arXiv:1910.01719*, 2019.
- S. Rastegar, R. Araújo, J. Mendes, T. Matias, and A. Emami. Self-adaptive takagi-sugeno model identification methodology for industrial control processes. In *IECON 2014-40th Annual Conference of the IEEE Industrial Electronics Society*, pages 281–287. IEEE, 2014. doi: 10.1109/IECON.2014.7048512.
- S. Rastegar, R. Araujo, and J. Mendes. Online identification of takagi–sugeno fuzzy models based on self-adaptive hierarchical particle swarm optimization algorithm. *Applied mathematical modelling*, 45:606–620, 2017. doi: 10.1016/j.apm.2017.01.019.
- R. Rosen, G. Von Wichert, G. Lo, and K. D. Bettenhausen. About the importance of autonomy and digital twins for the future of manufacturing. *Ifac-papersonline*, 48(3): 567–572, 2015.
- G. Saini, P. Ashok, E. van Oort, and M. R. Isbell. Accelerating well construction using a digital twin demonstrated on unconventional well data in north america. In *Unconventional Resources Technology Conference, Houston, Texas, 23-25 July 2018*, pages 3264–3276. Society of Exploration Geophysicists, American Association of Petroleum . . . , 2018.
- R. Salles, J. Mendes, R. Araújo, C. Melo, and P. Moura. Prediction of key variables in wastewater treatment plants using machine learning models. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE, 2022a. doi: 10.1109/IJCNN55064.2022.9892661.
- R. Salles, J. Mendes, R. P. Ribeiro, and J. Gama. Fault detection in wastewater treatment plants: Application of autoencoders models with streaming data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 55–70. Springer, 2022b. doi: 10.1007/978-3-031-23618-1\_4.
- L. Sánchez and J. Otero. A fast genetic method for inducing descriptive fuzzy models. *Fuzzy Sets and Systems*, 141(1):33–46, 2004.
- B. Schleich, N. Anwer, L. Mathieu, and S. Wartzack. Shaping the digital twin for design and production engineering. *CIRP annals*, 66(1):141–144, 2017.
- M. Schluse, M. Priggemeyer, L. Atorf, and J. Rossmann. Experimentable digital twins—streamlining simulation-based systems engineering for industry 4.0. *IEEE Transactions on industrial informatics*, 14(4):1722–1731, 2018.

- K. Schwab. *The Fourth Industrial Revolution*. World Economic Forum, Geneva, 2016.
- P. Sharma, D. Knezevic, P. Huynh, and G. Malinowski. Rb-fea based digital twin for structural integrity assessment of offshore structures. In *Offshore Technology Conference*, page D031S037R003. OTC, 2018.
- Y. Shen, J. L. Linville, M. Urgun-Demirtas, M. M. Mintz, and S. W. Snyder. An overview of biogas production and utilization at full-scale wastewater treatment plants (wwtps) in the united states: challenges and opportunities towards energy-neutral wwtps. *Renewable and Sustainable Energy Reviews*, 50:346–362, 2015.
- A. M. Silva, W. Caminhas, A. Lemos, and F. Gomide. A fast learning algorithm for evolving neo-fuzzy neuron. *Applied Soft Computing*, 14:194–209, 2014.
- M. Singh, E. Fuenmayor, E. P. Hinchy, Y. Qiao, N. Murray, and D. Devine. Digital twin: Origin to future. *Applied System Innovation*, 4(2):36, 2021.
- H. Soleimani-B, C. Lucas, and B. N. Araabi. Recursive gath–geva clustering as a basis for evolving neuro-fuzzy modeling. *Evolving Systems*, 1:59–71, 2010.
- F. R. Spellman. *Handbook of water and wastewater treatment plant operations*. CRC press, 2008.
- D. Systèmes. Meet virtual singapore, the city’s 3d digital twin. Available: <https://govinsider.asia/intl-en/article/meet-virtual-singapore-citys-3d-digital-twin>, 2018.
- T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE transactions on systems, man, and cybernetics*, 1:116–132, 1985.
- F. Tao and Q. Qi. New it driven service-oriented smart manufacturing: framework and characteristics. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(1): 81–91, 2017.
- F. Tao and Q. Qi. Make more digital twins. *Nature*, 573(7775):490–491, 2019.
- F. Tao and M. Zhang. Digital twin shop-floor: a new shop-floor paradigm towards smart manufacturing. *IEEE access*, 5:20418–20427, 2017.
- F. Tao, Q. Qi, A. Liu, and A. Kusiak. Data-driven smart manufacturing. *Journal of Manufacturing Systems*, 48:157–169, 2018a.
- F. Tao, H. Zhang, A. Liu, and A. Y. Nee. Digital twin in industry: State-of-the-art. *IEEE Transactions on industrial informatics*, 15(4):2405–2415, 2018b.
- F. Tao, W. Liu, M. Zhang, T.-l. Hu, Q. Qi, H. Zhang, F. Sui, T. Wang, H. Xu, Z. Huang, et al. Five-dimension digital twin model and its ten applications. *Computer integrated manufacturing systems*, 25(1):1–18, 2019a.
- F. Tao, Q. Qi, L. Wang, and A. Nee. Digital twins and cyber–physical systems toward smart manufacturing and industry 4.0: Correlation and comparison. *Engineering*, 5(4): 653–661, 2019b.

- F. Tao, M. Zhang, and A. Y. C. Nee. *Digital twin driven smart manufacturing*. Academic press, 2019c.
- E. Tejaswini, S. Panjwani, G. U. B. Babu, and A. S. Rao. Model based control of a full-scale biological wastewater treatment plant. *IFAC-PapersOnLine*, 53(1):208–213, 2020.
- E. J. Tuegel, A. R. Ingraffea, T. G. Eason, S. M. Spottswood, et al. Reengineering aircraft structural life prediction using a digital twin. *International Journal of Aerospace Engineering*, 2011, 2011.
- T. H.-J. Uhlemann, C. Schock, C. Lehmann, S. Freiburger, and R. Steinhilper. The digital twin: demonstrating the potential of real time data acquisition in production systems. *Procedia Manufacturing*, 9:113–120, 2017.
- T. R. Wanasinghe, L. Wroblewski, B. K. Petersen, R. G. Gosine, L. A. James, O. De Silva, G. K. Mann, and P. J. Warrian. Digital twin for the oil and gas industry: Overview, research trends, opportunities, and challenges. *IEEE access*, 8:104175–104197, 2020.
- L.-X. Wang, J. M. Mendel, et al. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE transactions on Neural Networks*, 3(5):807–814, 1992.
- Y. Yu, S. Fan, G. Peng, S. Dai, and G. Zhao. Study on application of digital twin model in product configuration management. *Aeronautical manufacturing technology*, 526(07): 41–45, 2017.
- L. A. Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.
- A. J. Zakrajsek and S. Mall. The development and use of a digital twin model for tire touchdown health monitoring. In *58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 0863, 2017.
- H. Zhang, Q. Liu, X. Chen, D. Zhang, and J. Leng. A digital twin-based approach for designing and multi-objective optimization of hollow glass production line. *Ieee Access*, 5:26901–26911, 2017a.
- Q. Zhang, J. Hu, D.-J. Lee, Y. Chang, and Y.-J. Lee. Sludge treatment: Current research trends. *Bioresource technology*, 243:1159–1172, 2017b.
- Y. Zheng, S. Yang, and H. Cheng. An application framework of digital twin and its case study. *Journal of Ambient Intelligence and Humanized Computing*, 10:1141–1153, 2019.
- C. Zhuang, J. Liu, H. Xiong, X. Ding, S. Liu, and G. Weng. Connotation, architecture and trends of product digital twin. *Computer Integrated Manufacturing Systems*, 23(4): 753–768, 2017.



# Appendix A

## Submitted Paper

# An Interpretable Digital Twin Framework Using Data-Driven Fuzzy-Based Models for Wastewater Treatment Plant Monitoring

Miguel Proença<sup>a</sup>, Jérôme Mendes<sup>b,\*</sup>, Rui Araújo<sup>a</sup> and Rodrigo Salles<sup>a</sup>

<sup>a</sup>University of Coimbra, Institute of Systems and Robotics, ARISE, Department of Electrical and Computer Engineering, Pólo II, Coimbra, Portugal

<sup>b</sup>University of Coimbra, CEMMPRE, ARISE, Department of Mechanical Engineering, Coimbra, Portugal

## ARTICLE INFO

### Keywords:

Digital Twin

WWTP

Data-driven models

Interpretable models

Digitalization

Fuzzy systems

## ABSTRACT

This paper presents an approach to perform a digital twin (DT) focused on the Wastewater Treatment Plant (WWTP) waterline. The proposed approach aims to create interpretable DTs, whose models must be easily interpretable by the expert process operators. For this purpose, to create interpretable DT models, different fuzzy-based models were studied, namely a model designed using the Fuzzy C-Means and Least Squares Method, the Generalized Additive Models Using Zero-Order T-S Fuzzy Systems (GAM-ZOTS) model, and the Iterative Learning of Multiple Univariate Zero-order T-S Fuzzy Systems (iMU-ZOTS) model. For validation of the proposed approach, the Benchmark Simulation Model 2 (BSM2) simulator is used, which simulates the entire wastewater treatment process based on real data. The DT was performed to predict the key variables throughout the waterline of the wastewater treatment process at different points in the plant. Exhaustive validations were performed. The 3 studied models were used to model 16 key variables at 14 different points in the treatment plant. Each model for each key variable for each point was learned/tested with different sets of input variables and using different learning model parameters. These tests were performed to choose the best parameters for the model to obtain the best results for each variable to be estimated.

## 1. Introduction

The digitization process, encompassing the transformation of data into digital formats and the integration of advanced technologies such as the Internet of Things (IoT), cloud computing, big data analytics, and Artificial Intelligence (AI), is rapidly advancing in the industrial sector [1]. With the industry's evolution, industrial processes increasingly require efficient, flexible, and reliable control [2]. This new trend has led to the emergence of the concept of Industry 4.0, popularized by Klaus Schwab [3]. This new paradigm has led to the appearance of new concepts, such as the Digital Twin (DT), a virtual representation of a physical object or process that can simulate its real-world behavior and predict its performance. The ability to visualize the state of a factory's components in real-time, combined with the power to simulate different scenarios, makes DT a powerful tool. The capability to simulate hypothetical scenarios allows tests in different environments and simulations of different system changes, allowing a better understanding of the status of physical entities. The constant monitoring of the performance of the process can allow for an increase in product quality and the reduction of faults and accidents, thus reducing maintenance and repair costs. DT is a major player in the industrial revolution and the digitization of industrial processes, and it is viewed with great interest by manufacturers.

Considering all its characteristics, DT can represent an important tool in tackling current and urgent problems related to access to water and the treatment of wastewater

resulting from human activities. Water is a vital resource for human life, and has a key role in many human activities, like industry, agriculture, and services. But, this important resource is limited [4]. Planet Earth is covered by about 70% of water, however, it is estimated that about 97% is salt water, which is found in seas and oceans and is unsuitable for human consumption, and about 2% forms inaccessible glaciers. Therefore, only 1% is used for human and animal consumption. With the increase in the world's population, the supply of quality water for human activities is becoming increasingly difficult as there is greater water consumption and greater wastewater production. Considering the importance of water for human beings and the worrying scarcity of this resource, it is essential to use water resources efficiently and to treat wastewater so that it becomes available again and does not pose a threat to the environment and, consequently, to human health. Thus, the Wastewater Treatment Plant (WWTP) emerges as a very important structure to face the new water and wastewater challenges. The wastewater treatment is carried out in WWTPs whose main objective is to remove pollutants from the wastewater. WWTPs are important structures and can be found in large numbers around the world. In the United States, there are more than 16,000 WWTPs in operation [5]. In the European Union, there are more than 18,000 WWTPs, covering more than 450 million people in all member states [6]. The treatment carried out at WWTPs consists of several phases and is subject to climatic events such as periods of drought, rain and storms. These characteristics give WWTPs a highly non-linear demeanor and make it difficult to create reliable models that represent the real behavior of these structures. DTs are a promising option in this context, allowing for a dynamic and comprehensive representation of the WWTPs.

\*Corresponding author

✉ jerome.mendes@uc.pt (J. Mendes); rui@isr.uc.pt (R. Araújo); rodrigo.salles@isr.uc.pt (R. Salles)

ORCID(s):

Although DT has gained popularity in recent years, the concept is not new. The concept of DT was introduced by Grieves in 2003, in his course "product lifecycle management" at the University of Michigan, and is considered the origin of DTs [7]. Initially, the DT was used mainly in the aerospace and military sectors. Nowadays, the DT is undergoing a period of rapid development with its use in various areas of knowledge [8]. Between 2003 and 2011, the development of communication technology, IoT sensor technology, big data analytics, and simulation technologies led to a huge increase in DTs. In 2011, the first journal article was published, explaining how DTs were useful for predicting the structural life of aircraft [9]. In 2012, the idea of DTs was revisited by the National Aeronautics and Space Administration (NASA), which defined DT as multiphysics, multiscale, probabilistic, ultra fidelity simulation that reflects the state of a corresponding twin based on the historical data, real-time sensor data, and physical model [10]. In 2014, Grieves published a white paper about DT, and according to it, the basic DT model consists of three main parts: (a) physical products in Real Space, (b) virtual products in Virtual Space, as well (c) the connections of data and information that tie the virtual and real products together [7]. More and more effort has been devoted to DT research since then. Gartner, in 2017 and 2018, classified DTs as one of the top ten most promising technological trends in the next decade [11]. According to [11], eight assets of the DT can be defined: real-time remote monitoring and control, greater efficiency and safety, predictive maintenance and scheduling, scenario and risk assessment, better intra and inter team synergy and collaborations, more efficient and informed decision support system, personalization of products and services, and better documentation and communication.

Some works have addressed DTs on WWTPs. In [12], DT was applied to predictive maintenance (PdM) to forecast equipment performance based on status monitoring data, to estimate when maintenance is required. In [13], a DT framework for fault detection in WWTPs is proposed. The DT has revolutionized the entire industry paradigm, improving process efficiency, asset lifecycle optimization, and decision-making processes. Despite all its qualities, DT's ability to predict failures, optimize processes, and improve maintenance can only be achieved with a proper understanding and interpretation of its capabilities and limitations. Although the scientific community has already carried out several studies on this technology, the problem of the interpretability of these models is still an unresolved gap, and it is a requirement for most industrial processes.

This paper presents an interpretable DT framework using data-driven fuzzy-based models for wastewater treatment process monitoring. The main contributions of this work are the study of different fuzzy-based models with an interpretable structure to model the key variables at the different points of the waterline of a WWTP. The following models were studied: 1) a model designed using Fuzzy C-Means and Least Squares Method, 2) the Generalized Additive Models Using Zero-Order T-S Fuzzy Systems (GAM-ZOTS)

model [14], and the Iterative Learning of Multiple Univariate Zero-Order T-S Fuzzy Systems (iMU-ZOTS) [15]. The validation of the proposed approach was performed using the Benchmark Simulator Model 2 (BSM2), which is a well-known simulator based on real data that simulates the entire wastewater treatment process. The key variables at different stages of the waterline of the treatment plant were defined as case studies and an exhaustive validation was performed. The following factors were analyzed on the studied models: the impact of the selected variables and the learning parameters on the model accuracy, and the accuracy versus interpretability tradeoff. The 3 models were used to model 16 key variables at 14 different points in the treatment plant. Each model for each key variable for each point was learned/tested with 4 different sets of input variables (from the variable selection method) and using different learning parameters. Thus, 520 tests were performed to choose the best model parameters to obtain the best results for each variable to be estimated.

This paper is organized as follows. Section 2 describes the fuzzy-based models and the models used in this paper. Section 3 describes the WWTPs and the framework used. Section 4 presents the achieved results and their analysis. Section 5 highlights the main conclusions.

## 2. Fuzzy-Based Models

Fuzzy Logic Systems (FLS) are rule-based systems characterized by rules composed of an IF statement (the antecedent part) and by a THEN statement (consequent part), allowing high interpretability of the fuzzy model and the transformation of a human knowledge base into mathematical formulations [16]. Using fuzzy-based models to make a digital representation (DT) of the WWTPs has numerous advantages for improving the whole wastewater treatment process. Through FLS, it is possible to create models that are easy to understand and interpret by any process operator, thus helping in the decision-making process and the optimization of the whole process. Furthermore, FLS can deal with the uncertainty, complexity and variability associated with WWTPs.

This section starts to present a generic Takagi-Sugeno (T-S) fuzzy model (Section 2.1), and the studied models, the model designed using Fuzzy C-Means and Least Squares Method, named here as FCM+LSM (Section 2.2), the GAM-ZOTS model (Section 2.3), and the iMU-ZOTS model (Section 2.4).

### 2.1. Takagi-Sugeno Fuzzy Model

One important type of fuzzy system, which is widely used in the scientific community, is the Takagi-Sugeno (T-S) fuzzy system [17]. One way of describing a non-linear system is using a T-S fuzzy model, which is defined by a set of rules as follows:

$$R_i : \text{IF } x_1(k) \text{ is } A_1^i \text{ and } \dots \text{ and } x_n(k) \text{ is } A_n^i \\ \text{THEN } y_i(k) = \theta_{i0} + \theta_{i1}x_1(k) + \dots + \theta_{in}x_n(k), \quad (1) \\ i = 1, \dots, N,$$



where  $R_i$  represents the  $i$ -th fuzzy rule, and  $N$  is the total number of rules. The input variables of the T-S fuzzy system are  $x_1(k), \dots, x_n(k)$ . The linguistic terms  $A_j^i$  ( $j = 1, \dots, n$ ), are defined by fuzzy membership functions (MFs), labeled  $\mu_{A_j^i}$ . These MFs describe the plant's operating region.  $\theta_{i1}, \dots, \theta_{in}$  are model (consequent) parameters of  $y_i(k)$ . From (1), the T-S fuzzy model  $y(k)$  can be rewritten as:

$$y[\mathbf{x}(k)] = \sum_{i=1}^N \bar{\omega}^i[\mathbf{x}(k)] y_i(\theta_i, \mathbf{x}(k)), \quad (2)$$

where for  $i = 1, \dots, N$ ,

$$\mathbf{x}(k) = [x_1(k), \dots, x_n(k)], \quad (3)$$

$$\mu_{A_j^i}(x_j) = \exp\left(-\frac{(x_j - v_{ij})^2}{2\sigma_{ij}^2}\right), \quad (4)$$

$$\bar{\omega}^i[\mathbf{x}(k)] = \frac{\prod_{j=1}^n \mu_{A_j^i}[\mathbf{x}(k)]}{\sum_{i=1}^N \prod_{j=1}^n \mu_{A_j^i}[\mathbf{x}(k)]}, \quad (5)$$

$$\theta_i = [\theta_{i0}, \theta_{i1}, \dots, \theta_{in}], \quad (6)$$

where  $v_{ij}$ ,  $\sigma_{ij}$  and  $\theta_i$  represent the center, width of the membership function and model parameters, respectively, which need to be defined/learned.

## 2.2. FCM+LSM

This model used the T-S fuzzy system in the form of (1), being necessary to obtain the antecedent parameters ( $\mathbf{v}_i$  and  $\sigma_i$ ), and the consequent parameters ( $\theta_i$ ). The antecedent parameters are given by the Fuzzy C-Means (FCM) algorithm, and the consequent parameters are given by the Least Squares Method (LSM).

### 2.2.1. Fuzzy C-Means

In the FCM algorithm, the main objective is to partition a given dataset into a given number of clusters,  $N$ . These clusters will be used to form the MFs and the antecedent part of the fuzzy rules. The fuzzy partitioning of a given dataset into a  $N$  clusters, comprises a family of fuzzy subsets  $\{A^i | 1 \leq i \leq N\}$ . These subsets are characterized by membership functions,  $\mu_i(k)$ , which are defined as  $\mu_i(k) = \mu_{A^i}(\mathbf{x}(k))$ . The FCM method aims to minimize an objective function, defined by (7). See [18] for more details. This function has a predefined number of clusters,  $N$ , and also includes a fuzziness parameter,  $\eta$ .

$$J = \sum_{i=1}^N \sum_{k=1}^K \mu_i^\eta(k) d_{ik}^2(\mathbf{x}(k), \mathbf{v}_i), \quad (7)$$

where  $\mathbf{v}_i = [v_{i1}, \dots, v_{in}]^T$  is the centroid vector of cluster  $i$ ,  $d_{ik}(\mathbf{x}(k), \mathbf{v}_i)$  is the Euclidean distance ( $l^2$ -norm) between the observation  $\mathbf{x}(k)$  and the cluster centroid  $\mathbf{v}_i$  given by (9), and  $\eta$  is the fuzziness parameter.

When the derivative of the objective function (7) is calculated with respect to the cluster centers and the membership values, the membership values can be calculated as

follows [18]:

$$\mu_i(k) = d_{ik}^2(\mathbf{x}(k), \mathbf{v}_i) \sum_{q=1}^N \left( d_{qk}^{2/(\eta-1)}(\mathbf{x}(k), \mathbf{v}_q) \right)^{-1}, \quad (8)$$

where:

$$d_{ik}^2(\mathbf{x}(k), \mathbf{v}_i) = (\mathbf{x}(k) - \mathbf{v}_i)^T (\mathbf{x}(k) - \mathbf{v}_i), \quad (9)$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^K \mu_i^\eta(k) \mathbf{x}(k)}{\sum_{k=1}^K \mu_i^\eta(k)}. \quad (10)$$

The centers  $v_{ij}$  of the antecedent membership functions are obtained by (10), and the corresponding widths,  $\sigma_i = [\sigma_{i1}, \dots, \sigma_{in}]^T$ , are obtained as follows:

$$\sigma_{ij} = \sqrt{\frac{2 \sum_{k=1}^K \mu_i(k) (x_j(k) - v_{ij})^2}{\sum_{k=1}^K \mu_i(k)}}. \quad (11)$$

### 2.2.2. Least Squares Method

After learning of the antecedent parameters, the consequent parameters are given by the LSM. Consider a desired model  $y_d(\mathbf{x}(k))$  for the output of a system for the sample  $k$  ( $k = 1, \dots, K$ ) and

$$\mathbf{y}_d = [y_d(\mathbf{x}(1)), \dots, y_d(\mathbf{x}(K))]^T, \quad (12)$$

$$\Phi = [\phi(1), \dots, \phi(K)]^T, \quad (13)$$

$$\phi(k) = [\bar{\omega}^1[\mathbf{x}(k)] \mathbf{x}_e^T(k), \dots, \bar{\omega}^N[\mathbf{x}(k)] \mathbf{x}_e^T(k)]^T, \quad (14)$$

$$\mathbf{x}_e(k) = [1, x_1(k), x_2(k), \dots, x_n(k)]^T, \quad (15)$$

then

$$\mathbf{y}_d = \Phi \Theta. \quad (16)$$

Suppose that  $\Theta^*$  represents the optimal values of all consequent parameters  $\Theta = [\theta_1^T, \dots, \theta_N^T]^T$ , a solution of  $\Theta^*$  can be calculated using the pseudo-inverse least squares method as follows:

$$\Theta^* = \Phi^+ \mathbf{y}_d, \quad (17)$$

where, considering  $K \geq N(n+1)$ , the pseudo inverse,  $\Phi^+$ , can be computed in closed form as follows:

$$\Phi^+ = (\Phi^T \Phi)^{-1} \Phi^T. \quad (18)$$

The design of the T-S fuzzy model is presented in Algorithm 1.

## 2.3. GAM-ZOTS model

In [14], an approach for learning a system of neo-fuzzy neurons, called Generalized Additive Models Using Zero-Order T-S Fuzzy Systems (GAM-ZOTS) is proposed. GAM-ZOTS uses the backfitting algorithm to learn multiple univariate zero-order T-S additive fuzzy systems. This approach aims to improve interpretability. The model structure used is

---

**Algorithm 1** Algorithm to design the FCM+LSM model.

1. Define the inputs,  $K$  data samples  $\{\mathbf{x}(k), y(k)\}_{k=1}^K$ , with  $\mathbf{x}(k) = [x_1(k), \dots, x_n(k)]^T$ ; Define the parameters to be used on the FCM.
  2. Obtain the antecedent parameters ( $\mathbf{v}_i$  and  $\sigma_i$ ) through the FCM algorithm (Section 2.2.1);
  3. Obtain the consequent parameters using LSM:
    - (a) Construct the vector  $\Phi$  through (13);
    - (b) Compute the consequent values  $\Theta^*$  using (17).
- 

based on the concept of additive models, in which each univariate model is a zero-order T-S fuzzy system representing a relationship between the input variable  $x_j$  (for  $j = 1, \dots, n$ ) and the output. Therefore, the model presented consists of the sum of multiple zero-order T-S fuzzy systems expressed as follows:

$$\begin{aligned} y(\mathbf{x}) &= y_0 + y_1(x_1) + \dots + y_n(x_n), \\ &= y_0 + \sum_{j=1}^n y_j(x_j), \end{aligned} \quad (19)$$

where  $y_0$  is a constant, and each individual model  $y_j(x_j)$  is represented by (21) being composed of  $N_j$  zero-order T-S fuzzy rules in the format of (20).

$$\begin{aligned} R_j^1 &: \text{IF } x_j(k) \text{ is } A_j^1 \text{ THEN } y_j^1(k) = \theta_j^1, \\ &\vdots \\ R_j^{N_j} &: \text{IF } x_j(k) \text{ is } A_j^{N_j} \text{ THEN } y_j^{N_j} = \theta_j^{N_j}. \end{aligned} \quad (20)$$

$$y_j(x_j(k)) = \sum_{i=1}^{N_j} \omega_j^i [x_j(k)] \theta_j^i = \phi_j(x_j(k)) \theta_j, \quad (21)$$

where for  $i = 1, \dots, N_j$ ,

$$\omega_j^i [x_j(k)] = \frac{\mu_{A_j^i}(x_j(k))}{\sum_{i=1}^{N_j} \mu_{A_j^i}(x_j(k))}, \quad (22)$$

$$\theta_j = [\theta_j^1, \dots, \theta_j^{N_j}]^T, \quad (23)$$

$$\phi_j(x_j(k)) = [\omega_j^1 [x_j(k)], \dots, \omega_j^{N_j} [x_j(k)]]. \quad (24)$$

The MFs are complementary triangular membership functions. Figure 1 shows an example of complementary triangular membership functions, in which the minimum and maximum value of the input variable  $x_j$  are represented by  $x_j^-$  and  $x_j^+$ , respectively. The lower limit, center value, and upper limit of  $A_j^i$  are represented by  $a_{j,i}$ ,  $b_{j,i}$  and  $c_{j,i}$ , respectively.

The learning methodology of the GAM-ZOTS is presented in the Algorithm 2. More details are presented in [14]. The first step is to define the MFs for each input which are complementary triangular MFs. These MFs are obtained in

---

**Algorithm 2** Description of GAM-ZOTS learning method [14].

1. Define the inputs,  $K$  data samples  $\{\mathbf{x}(k), y(k)\}_{k=1}^K$ , with  $\mathbf{x}(k) = [x_1(k), \dots, x_n(k)]^T$ ;
  2. Define the MFs. For all input variables  $x_j$ :
    - (a) Define the range values,  $x_j^-$  and  $x_j^+$ ;
    - (b) Set the number of membership functions,  $N_j$ .
    - (c) For each complementary triangular MF,  $A_j^i$ , the parameters are:
      - i. For  $A_j^1$ :  $a_{j,1} = b_{j,1} = x_j^-$  and  $c_{j,1} = b_{j,2}$ ;
      - ii. For  $A_j^i$ :  $a_{j,i} = b_{j,i-1}$  and  $b_{j,i} = c_{j,i-1}$ ,  $c_{j,i} = b_{j,i+1}$ ;
      - iii. For  $A_j^{N_j}$ :  $a_{j,N_j} = b_{j,N_j-1}$  and  $b_{j,N_j} = c_{j,N_j} = x_j^+$ .
  3. Using the Backfitting Algorithm, calculate the consequent parameters ( $\theta_j^i$ ). More details in [14].
- 

Steps 2a – 2b of Algorithm 2. The next step is to obtain the consequent parts of the fuzzy rules corresponding to each input variable  $x_j$ , using the Backfitting Algorithm (Step 3 of Algorithm 2).

## 2.4. iMU-ZOTS model

In the GAM-ZOTS method, the antecedent membership functions are created manually, as presented before, distributing them evenly across the universe of discourse, which means that there are limitations and inefficiencies in the design of the antecedent parts since they are not distributed according to the data. According to that, in [15] an iterative learning approach, called Iterative Learning of Multiple Univariate Zero-Order T-S Fuzzy Systems (iMU-ZOTS), is proposed. The major difference between iMU-ZOTS and GAM-ZOTS is in the iterative learning of the antecedent part, where instead of MFs being manually designed, iMU-ZOTS iteratively designs the MFs and selects candidate input variables where new MFs can be added.

iMU-ZOTS is divided into 2 stages, initialization and iterative learning. The initialization of the iMU-ZOTS involves applying the GAM-ZOTS method (Algorithm 2), with  $N_j = 2$  fuzzy rules for each input variable,  $x_j$ . In this way, two complementary triangular MFs are defined, and the consequent parameters are obtained through the GAM-ZOTS (Algorithm 2).

The iterative learning approach starts by selecting the candidate input variable, denoted as  $x_j^*$ , where a new fuzzy rule can be added. This selection is done using a novelty detection criterion, Criterion 1 (28), through a value between 0 and 1, which facilitates the definition of the threshold for adding a new fuzzy rule. The novelty measurement on the  $i_j$ -th fuzzy rule ( $i_j = 1, \dots, N_j$ ) related to  $x_j(k)$  is obtained by:

$$M_{i_j}(k) = \exp(-|x_j(k) - \mu_{i_j}(k)|), \quad (25)$$

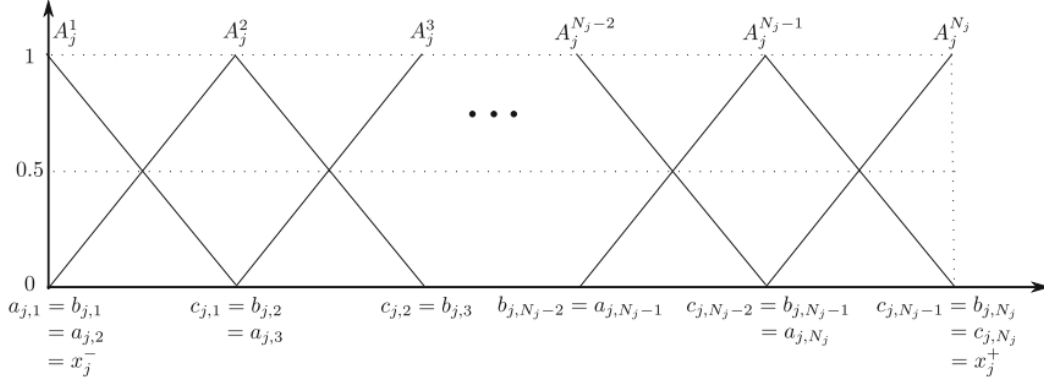


Figure 1: Membership functions for  $x_j$ . Taken from [14].

where

$$\mu_{i_j}(k) = \mu_{i_j}(k-1) + \frac{x_j(k) - \mu_{i_j}(k-1)}{N_{i_j}(k)} v_{i_j}^m(k), \quad (26)$$

$v_{i_j}(k) = \mu_{A_j^m}(x_j(k))$ ,  $m > 1$  is an integer representing the degree of fuzzification of the clusters.

$$N_{i_j}(k) = N_{i_j}(k-1) + v_{i_j}^m(k) \quad (27)$$

indicates the total membership function degrees associated to the  $i_j$ -th MF of  $x_j$ , and  $N_{i_j}(0) = 0$ .

**Criterion 1:** the novelty detection criterion for  $x_j$  is defined as follows:

$$M_{j,\max}(k) = \begin{cases} 1, & \text{if } \max_{i_j=1}^{N_j} M_{i_j}(k) \leq M_{th}^j \\ 0, & \text{otherwise} \end{cases} \quad (28)$$

Where  $M_{th}^j$  is a threshold for  $x_j$ , between zero and one.  $M_{j,\max}(k) = 0$  means that the variable  $x_j$  is not well represented by the current model, and, on the other hand,  $M_{j,\max}(k) = 1$  is perfectly represented.

To avoid overfitting and reach a complex model structure, Criterion 2 (29) was defined.

**Criterion 2:** a minimum distance between the center of the two nearest MFs must be guaranteed as follows:

$$\left| b_{j^*,i_j^+} - b_{j^*,i_j^-} \right| > \eta_j, \quad (29)$$

where  $b_{j^*,i_j^+}$  and  $b_{j^*,i_j^-}$  are the centers of MF to be added and of its nearest MF, respectively, and  $\eta_j$  is a threshold.

Once Criteria 1 (28) and 2 (29) are met, a new fuzzy rule  $R_j^{\text{new}}$ , and a new MF  $A_{j^*}^{\text{new}}$ , are added in  $x_j^*$ . Thus, the new MF,  $A_{j^*}^{\text{new}}$ , is added and the nearest left and right MFs,  $A_{j^*}^{\text{left}}$  and  $A_{j^*}^{\text{right}}$ , respectively, are updated by:

- $A_{j^*}^{\text{left}}$ : 1)  $c_{j^*,\text{left}} = b_{j^*,\text{new}}$ ;

---

### Algorithm 3 Description of iMU-ZOTS learning [15].

---

1. Define the inputs,  $K$  data samples  $\{\mathbf{x}(k), y(k)\}_{k=1}^K$  where  $\mathbf{x}(k) = [x_1(k), \dots, x_n(k)]^T$ ; Define their parameters;
2. Initialization: for each input variable  $x_j$ : define two ( $N_j = 2$ ) MFs and obtain the initial model using the GAM-ZOTS (Algorithm 2);
3. Iterative learning:
  - (a) For all points in the training dataset  $k = 1, \dots, K$ :
    - i. Using (28), the novelty  $M_{j,\max}(k)$  is measured, for input all variables;
    - ii. If Criteria (28) and (29) are met:
      - A. Define the new MF,  $A_{j^*}^{\text{new}}$ , and update  $A_{j^*}^{\text{left}}$  and  $A_{j^*}^{\text{right}}$ ;
      - B. Obtain all consequent parameters using the GAM-ZOTS (Algorithm 2 Step 3).
  - (b) If a fuzzy rule was added in Step 3a, it returns to Step 3a. If not, the learning process ends.

- 
- $A_{j^*}^{\text{new}}$ : 1)  $a_{j^*,\text{new}} = b_{j^*,\text{left}}$ , 2)  $b_{j^*,\text{new}} = x_j(k)$ , and 3)  $c_{j^*,\text{new}} = b_{j^*,\text{right}}$ ;
  - $A_{j^*}^{\text{right}}$ : 1)  $a_{j^*,\text{right}} = b_{j^*,\text{new}}$ .

Algorithm 3 presents the steps of the iMU-ZOTS model learning approach.

### 3. Proposed interpretable digital twin framework for WWTP monitoring

Wastewater is treated in WWTPs, where the main objective is to remove the different types of pollutants present in water used in human activities. Through this treatment, these structures speed up the process that occurs in nature, allowing the water to be returned to the environment without posing any danger. WWTPs are large, highly complex structures where the water treatment process is carried out

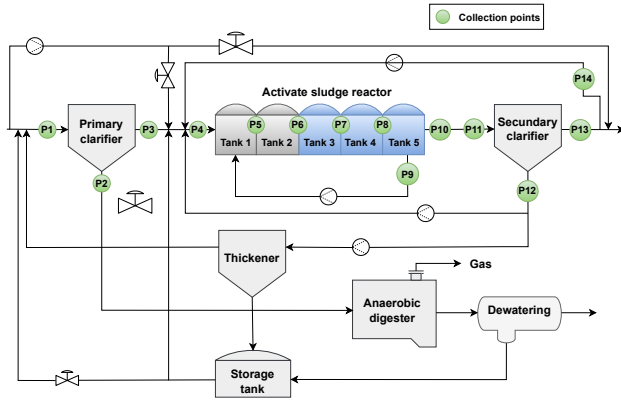


Figure 2: BSM2 data collection points.

in several stages through physical, chemical and biological processes, and need to operate within the limits imposed by environmental legislation [19]. The main stages are: preliminary treatment, primary treatment, secondary treatment, tertiary treatment, and sludge treatment [20]. Some external factors make the treatment process even more complex, such as climatic phenomena (rains, storms, droughts), seasonality, and periods of day and night.

For the development of this work, it is used the Benchmark Simulator Model 2 (BSM2), simulates all the steps in the wastewater treatment process that occurs in a WWTP. The use of BSM2 is widely accepted by the scientific community and has already been used in several studies, such as the following [21, 22, 23, 4]. On BSM2, throughout the different stages of the treatment process, various variables can be monitored.

### 3.1. Framework

According to experts from WWTP who have collaborated in this work, along the waterline of the WWTP, the main and the most important variables to the prediction case study investigated in this work are: total suspended solids,  $TSS$  [ $g\ S_S/m^3$ ], ammonia and ammonium,  $S_{NH}$  [ $g\ N/m^3$ ], dissolved oxygen,  $S_O$  [ $g\ COD/m^3$ ], nitrate and nitrite,  $S_{NO}$  [ $g\ N/m^3$ ], biological oxygen demand,  $BOD$  [ $g\ /m^3$ ], and chemical oxygen demand,  $COD$  [ $g\ /m^3$ ]. To collect the data, 14 points were placed along the water line, as shown in Figure 2.

$TSS$  is predicted at the exit of the primary clarifier, at point P3, using the data from points P1 and P2 as input.  $S_{NH}$  is predicted throughout the water line, at point P3, with P1 and P2 as input, at point P10, with P4 and P6 as input, and at point P13, with P11 and P12 as input.  $S_O$  is predicted at point P10, with P4 and P6 as input.  $S_{NO}$  is predicted at point P3, with P1 and P2 as inputs, at the outlet of tank 5, at P10, with P4 and P6 as inputs and at the outlet of the secondary clarifier, at P13 using P11 and P12 as inputs.  $BOD$  and  $COD$  are predicted at point P13, using points P11 and P12 as input. The datasets of the variables to be predicted are built based on the input and output points, and in the end, there are 10 variables to be predicted distributed among the points on

the WWTP waterline. These variables are:  $TSS_3$ ,  $S_{NH_3}$ ,  $S_{NH_{10}}$ ,  $S_{NH_{13}}$ ,  $S_{NO_3}$ ,  $S_{NO_{10}}$ ,  $S_{NO_{13}}$ ,  $S_{O_{10}}$ ,  $BOD_{13}$  and  $COD_{13}$ . The subscript value of each variable corresponds to the point where that variable is being predicted.

For the evaluation of the proposed framework, to predict each desired variable, 609 days are considered, where the first 245 days are for the stabilization of the plant. The sampling frequency is 15 minutes, totalling 58464 samples.

The variables used as input were selected according to their correlation with the variables to be predicted. The value of the correlation between the input variables and the variable to be predicted is then compared with a previously defined threshold. Input variables with a higher correlation than the defined threshold are selected. The following thresholds were defined:  $thr = [thr_1, \dots, thr_{max}] = [0.3, 0.5, 0.7, 0.85]$ . In this way, different input variables are selected according to the comparison of the correlation value and the threshold value. Thus, 4 datasets were built with different input variables for each variable to be predicted, except for the variables  $S_{NH_{10}}$  and  $BOD_{13}$ , where it was only possible to build 2 datasets since there were no input variables with a correlation above the thresholds of 0.7 and 0.85. In all, 36 datasets were built and tested for the different models. To carry out training and testing, all the datasets were divided into 70% for training and 30% for testing. In Algorithm 4, it is presented the setup of the proposed framework.

To determine the best configuration of each model, in terms of accuracy and interpretability, different combinations of model parameters were set:

- **FCM-LSM:** number of clusters  $N = [3, 5, 7, 10]$ , degree of fuzziness  $\eta = 2$ , number of iterations  $Max = 100$  and termination condition  $imp = 0.00001$ .
- **GAM-ZOTS:** number of fuzzy rules  $N = [2, 3, 5, 7, 10]$ , maximum number of iterations  $lim_{it} = 100$ , and termination condition  $\epsilon = 0.00001$ .
- **iMU-ZOTS:**  $M_{th}^j = [0.5, 0.7, 0.8, 0.9]$ ,  $\eta_j = |x_j^+ - x_j^-|/10$  (for  $j = 1, \dots, n$ ), maximum number of iterations  $lim_{it} = 100$ , and termination condition  $\epsilon = 0.00001$ .

To obtain the best predictions for each variable, several tests were performed for the different parameter combinations in each model. The best results were evaluated through the Mean Squared Error (MSE) metric described by the following expression:

$$MSE = \frac{1}{K} \sum_{k=1}^K (y_k - \hat{y}_k)^2, \quad (30)$$

where  $y_k$  and  $\hat{y}_k$  are the real and estimated target at the instant of time,  $k$ , and  $K$  is the total number of samples.

## 4. Results and Discussion

This section presents the results and the respective analyses regarding the accuracy and interpretability of the models.



---

**Algorithm 4** Setup of the proposed framework.
 

---

1. Define the set of variables to be predicted and their collecting points. In this work 10 ( $p_{max}$ ) variables were defined:  $TSS_3$ ,  $S_{NH_3}$ ,  $S_{NH_{10}}$ ,  $S_{NH_{13}}$ ,  $S_{NO_3}$ ,  $S_{NO_{10}}$ ,  $S_{NO_{13}}$ ,  $SO_{10}$ ,  $BOD_{13}$ , and  $COD_{13}$ , numbered from  $p = 1$  to  $p = 10$ , respectively;
  2. Define the set of thresholds for the variables selection. In this work,  $max = 4$  thresholds  $thr_1, \dots, thr_{max}$  were defined;
  3. Define the parameters of the models to be learned;
  4. For each variable to be predicted,  $p = 1, \dots, p_{max}$ :
    - (a) Read data from the collected points related to variable  $p$ ;
    - (b) Obtain the correlation between input variables and variable  $p$ ;
    - (c) For each threshold,  $thr_i = thr_1, \dots, thr_{max}$ :
      - i. Select the input variables whose correlation is higher than the threshold  $thr_i$ ;
      - ii. Divide the dataset into training and test;
      - iii. For each model,  $model = 1, \dots, 3$  (models FCM-LSM, GAM-ZOTS, and iMU-ZOTS, respectively)
        - A. For each combination of model parameters,  $m_p = 1, \dots, H$ :
          - Learn the models, FCM-LSM, GAM-ZOTS and iMU-ZOTS;
          - Calculate the MSE of each model.
        - B. Save the best setup for each model of variable  $p$ .
    - (d) Save the best model for variable  $p$ .
- 

Table 1 shows the results of the predictions made by the various models presented, where it is possible to check the information from the tests carried out, namely the best error value for each dataset (“error” columns), the average error (“mean” columns), and the standard deviation (“std. dev” columns). The best result for each variable is shown in bold. Taking into account the MSE error values in Table 1, it can be seen that the FCM-LSM obtained the best error values for 9 variables, while iMU-ZOTS obtained the best error value for one variable,  $S_{NH_{10}}$ . It can therefore be concluded that FCM-LSM is the method that produces the lowest MSE error value for almost all variables. The GAM-ZOTS and iMU-ZOTS methods have higher error values when compared to FCM-LSM.

The correlation threshold value is an important parameter in this study, as it is through it that the input variables are selected to construct the model. For this study, 4 threshold values were used, and it was found that as this value increased, the number of variables selected decreased. When the results obtained are analyzed, it can be seen that the best results are generally obtained when the threshold value is minimal, where more input variables are selected.

Table 2 presents the number of rules corresponding to

the MSE results presented in Table 1. In addition to the number of rules, the average number of rules for each method is shown and, in the case of iMU-ZOTS, the threshold value,  $M_{th}$  for which the best results were obtained, is also shown. In bold are the values of the rules for which the best result of MSE (in Table 1) was obtained. And, the best results obtained by the other two methods are underlined.

The analysis of the number of fuzzy rules used by each method is extremely important because it is through these that the interpretability of a model can be gauged. In the context of WWTPs, precise prediction of variables stands as a linchpin for ensuring their efficient operation. Inaccuracies in prediction can lead to squandered resources and compromised water quality. However, while accuracy remains a critical metric, it doesn’t stand alone in significance. The crux lies in striking a balance between accuracy and interpretability. Complex models, while potentially accurate, pose challenges in comprehension, particularly for WWTP operators who require clear insights into the decision-making process.

The analysis of Table 1 and Table 2 shows that the FCM-LSM obtained the best error value MSE for most of the variables, but it has a large number of rules, which makes it a complex, difficult to interpret, model. Finding a balance between accuracy and interpretability is necessary to have efficient and less complex models.

To assess the accuracy and interpretability of the different models, the best results obtained by the 3 methods were analyzed. The results obtained by the methods that did not obtain the best error value for each variable are also chosen taking into account the balance between interpretability and accuracy. A less complex model with fewer rules is chosen if, at the same time, it has a low error value that is relatively close to the best value obtained for that variable. These MSE error values and the rules are shown underlined in Tables 1 and 2.

To make a more detailed analysis of the different methods and the balance between accuracy and interpretability, two variables were selected,  $S_{NH_3}$  and  $S_{NO_3}$ , as case studied. For the variable  $S_{NH_3}$ , a threshold value of 0.85 was used for the FCM-LSM and GAM-ZOTS methods and 0.5 for the iMU-ZOTS method. For the variable  $S_{NO_3}$ , a threshold value of 0.85 was used for FCM, 0.7 for GAM-ZOTS and 0.3 for iMU-ZOTS.

The results of the predictions made by the different methods for the variable  $S_{NH_3}$  are shown in Figure 3. By analyzing the figure, it is possible to see that the differences in forecasting between the different methods are minimal, with practically no significant difference.

The learned FCM-LSM for  $S_{NH_3}$  describes each of the 3 input variables by 3 fuzzy rules. Figure 4 shows the MFs for the 3 input variables. The fuzzy rules are described by:

$$\begin{aligned}
 R_1 : & \text{ IF } x_1(k) \text{ is } A_1^1 \text{ and } x_2(k) \text{ is } A_2^1 \text{ and } x_3(k) \text{ is } A_3^1 \\
 & \text{ THEN } y_1(k) = 1 + 1.37 \times 10^{-11} x_1(k) + 1.21 \times 10^{-13} x_2(k) \\
 & \quad + 4.26 \times 10^{-14} x_3(k),
 \end{aligned}$$

**Table 1**

Results of the tests carried out by the 3 models. **Var.** represents the target variable,  $n$  is the number of input variables selected and  $thr$  is the threshold value. The best result for each variable is shown in bold and the best result of the other models is shown in underlined.

Var.	thr	n	FCM-LSM			GAM-ZOTS			iMU-ZOTS		
			error	mean	std. dev	error	mean	std. dev	error	mean	std. dev
$TSS_3$	0.3	18	<b>64.46</b>	65.61	1.24	<u>84.53</u>	91.06	10.65	<u>82.39</u>	82.99	0.41
	0.5	12	77.88	78.63	0.56	135.66	139.13	5.47	134.11	135.07	0.66
	0.7	9	87.44	87.96	0.55	166.97	171.35	6.35	165.31	167.08	1.19
	0.85	8	90.97	91.29	0.47	210.53	215.13	5.32	210.07	210.68	0.41
$S_{NH_3}$	0.3	18	$1.82 \times 10^{-18}$	$1.31 \times 10^{-17}$	$1.30 \times 10^{-17}$	0.14	0.15	0.0037	$1.13 \times 10^{-4}$	$3.57 \times 10^{-4}$	$2.662 \times 10^{-4}$
	0.5	16	$1.36 \times 10^{-18}$	$1.21 \times 10^{-17}$	$1.29 \times 10^{-17}$	0.13	0.14	0.0026	<u><math>1.28 \times 10^{-4}</math></u>	$5.031 \times 10^{-4}$	$4.007 \times 10^{-4}$
	0.7	11	$6.99 \times 10^{-20}$	$3.73 \times 10^{-19}$	$5.45 \times 10^{-19}$	0.066	0.07	0.0032	$1.87 \times 10^{-8}$	$5.08 \times 10^{-5}$	$1.005 \times 10^{-4}$
	0.85	3	<b><math>1.06 \times 10^{-23}</math></b>	$5.99 \times 10^{-22}$	$1.04 \times 10^{-21}$	<u><math>7.90 \times 10^{-12}</math></u>	$3.61 \times 10^{-10}$	$6.87 \times 10^{-10}$	$1.29 \times 10^{-12}$	$1.71 \times 10^{-12}$	$4.56 \times 10^{-13}$
$S_{NH_{10}}$	0.3	15	0.065	0.069	0.0059	<u>0.043</u>	0.071	0.038	<b>0.038</b>	0.038	$4.08 \times 10^{-4}$
	0.5	5	<u>0.053</u>	0.073	0.021	0.059	0.088	0.041	0.058	0.058	$1.67 \times 10^{-4}$
	0.7	-	-	-	-	-	-	-	-	-	-
	0.85	-	-	-	-	-	-	-	-	-	-
$S_{NH_{13}}$	0.3	15	<b>0.025</b>	0.028	0.0029	0.028	0.029	0.0024	<u>0.027</u>	0.027	$2.68 \times 10^{-4}$
	0.5	6	0.028	0.029	0.0013	<u>0.03</u>	0.032	0.0037	0.031	0.031	$1.30 \times 10^{-4}$
	0.7	1	0.077	-	-	0.076	0.077	$2.49 \times 10^{-4}$	0.076	0.076	$1.33 \times 10^{-5}$
	0.85	1	0.077	-	-	0.076	0.077	$2.49 \times 10^{-4}$	0.076	0.076	$1.33 \times 10^{-5}$
$S_{NO_3}$	0.3	13	$6.44 \times 10^{-26}$	$1.64 \times 10^{-19}$	$3.27 \times 10^{-19}$	$1.21 \times 10^{-5}$	$1.57 \times 10^{-5}$	$3.65 \times 10^{-6}$	<u><math>2.49 \times 10^{-9}</math></u>	$4.57 \times 10^{-9}$	$2.96 \times 10^{-9}$
	0.5	10	$1.45 \times 10^{-26}$	$1.97 \times 10^{-19}$	$3.90 \times 10^{-19}$	$1.24 \times 10^{-5}$	$1.85 \times 10^{-5}$	$4.47 \times 10^{-6}$	$1.30 \times 10^{-9}$	$1.63 \times 10^{-9}$	$2.71 \times 10^{-10}$
	0.7	7	$1.93 \times 10^{-28}$	$8.19 \times 10^{-28}$	$6.09 \times 10^{-28}$	<u><math>1.16 \times 10^{-5}</math></u>	$1.33 \times 10^{-5}$	$1.31 \times 10^{-6}$	$2.29 \times 10^{-10}$	$2.44 \times 10^{-10}$	$1.49 \times 10^{-11}$
	0.85	4	<b><math>2.92 \times 10^{-28}</math></b>	$4.62 \times 10^{-28}$	$2.80 \times 10^{-28}$	$8.88 \times 10^{-12}$	$1.65 \times 10^{-12}$	$2.11 \times 10^{-8}$	$1.46 \times 10^{-12}$	$1.49 \times 10^{-12}$	$2.28 \times 10^{-14}$
$S_{NO_{10}}$	0.3	6	<b>0.14</b>	0.14	0.0043	0.18	0.18	0.0046	<u>0.18</u>	0.18	$6.94 \times 10^{-4}$
	0.5	1	0.97	-	-	<u>0.97</u>	0.97	$8.49 \times 10^{-4}$	0.97	0.97	$2.28 \times 10^{-4}$
	0.7	1	0.97	-	-	0.97	0.97	$8.49 \times 10^{-4}$	0.97	0.97	$2.28 \times 10^{-4}$
	0.85	1	0.97	-	-	0.97	0.97	$8.49 \times 10^{-4}$	0.97	0.97	$2.28 \times 10^{-4}$
$S_{NO_{13}}$	0.3	4	<b>0.12</b>	0.13	0.01	<u>0.14</u>	0.14	0.0062	<u>0.14</u>	0.14	$1.99 \times 10^{-4}$
	0.5	4	<b>0.12</b>	0.13	0.01	0.14	0.14	0.0062	0.14	0.14	$1.99 \times 10^{-4}$
	0.7	1	0.19	-	-	0.19	0.19	$1.20 \times 10^{-4}$	0.19	0.19	$3.64 \times 10^{-5}$
	0.85	1	0.19	-	-	0.19	0.19	$1.20 \times 10^{-4}$	0.19	0.19	$3.64 \times 10^{-5}$
$S_{O_{10}}$	0.3	4	<b>0.016</b>	0.018	0.0018	0.017	0.019	0.0018	<u>0.017</u>	0.017	$1.28 \times 10^{-4}$
	0.5	1	0.032	-	-	<u>0.031</u>	0.032	$2.24 \times 10^{-4}$	0.031	0.031	$2.28 \times 10^{-5}$
	0.7	1	0.032	-	-	0.031	0.032	$2.24 \times 10^{-4}$	0.031	0.031	$2.28 \times 10^{-5}$
	0.85	1	0.032	-	-	0.031	0.032	$2.24 \times 10^{-4}$	0.031	0.031	$2.28 \times 10^{-5}$
$BOD_{13}$	0.3	23	<b>0.003</b>	0.004	$5.82 \times 10^{-4}$	<u>0.005</u>	0.006	0.001	0.005	0.005	$1.62 \times 10^{-4}$
	0.5	13	0.004	0.004	$1.90 \times 10^{-4}$	0.006	0.008	0.001	0.006	0.006	$1.24 \times 10^{-4}$
	0.7	8	0.004	0.005	$7.17 \times 10^{-4}$	0.008	0.010	0.001	<u>0.007</u>	0.008	$4.07 \times 10^{-4}$
	0.85	1	0.097	-	-	0.097	0.098	$5.70 \times 10^{-4}$	0.097	0.097	$3.95 \times 10^{-4}$
$COD_{13}$	0.3	17	<b>0.27</b>	0.29	0.01	<u>0.39</u>	0.49	0.10	<u>0.41</u>	0.42	0.006
	0.5	9	6.75	7.05	0.39	7.49	7.76	0.20	7.28	7.34	0.05
	0.7	-	-	-	-	-	-	-	-	-	-
	0.85	-	-	-	-	-	-	-	-	-	-

$R_2$  : IF  $x_1(k)$  is  $A_1^2$  and  $x_2(k)$  is  $A_2^2$  and  $x_3(k)$  is  $A_3^2$   
 THEN  $y_2(k) = 1 + 7.66 \times 10^{-12} x_1(k) - 1.97 \times 10^{-13} x_2(k) - 8.75 \times 10^{-14} x_3(k)$ ,  
 $R_3$  : IF  $x_1(k)$  is  $A_1^3$  and  $x_2(k)$  is  $A_2^3$  and  $x_3(k)$  is  $A_3^3$   
 THEN  $y_3(k) = 1 + 5.69 \times 10^{-13} x_1(k) - 1.56 \times 10^{-13} x_2(k) - 1.88 \times 10^{-14} x_3(k)$ .

The GAM-ZOTS method describes each of the 3 selected input variables by 2 fuzzy rules. Figure 5 shows the

MFs for the 3 input variables. The fuzzy rules for variable  $x_1$  are described by:

$$\begin{aligned}
 R_1^1: & \text{ IF } x_1(k) \text{ is } A_1^1 \text{ THEN } y_1^1(k) = -2.85 \times 10^{-6}, \\
 R_2^1: & \text{ IF } x_1(k) \text{ is } A_2^1 \text{ THEN } y_1^2(k) = 4.10 \times 10^{-6},
 \end{aligned}$$

for the variable  $x_2$  by:

$$\begin{aligned}
 R_1^2: & \text{ IF } x_2(k) \text{ is } A_2^1 \text{ THEN } y_2^1(k) = -1.32 \times 10^{-5}, \\
 R_2^2: & \text{ IF } x_2(k) \text{ is } A_2^2 \text{ THEN } y_2^2(k) = 2.23 \times 10^{-5},
 \end{aligned}$$

and for the variable  $x_3$  by:

$$\begin{aligned}
 R_1^3: & \text{ IF } x_3(k) \text{ is } A_3^1 \text{ THEN } y_3^1(k) = -24.81, \\
 R_2^3: & \text{ IF } x_3(k) \text{ is } A_3^2 \text{ THEN } y_3^2(k) = 30.97.
 \end{aligned}$$

**Table 2**

Fuzzy rule values for the 3 models. **Var.** represents the target variable, *thr* the threshold value, *N* the number of fuzzy rules, Mean rules the average value of the fuzzy rules and  $M_{th}$  the threshold value of the iMU-ZOTS model. The value of the rules for each variable for which the best result was obtained is shown in bold and the best value of the other models is shown in underlined.

Var.	thr	FCM-LSM		GAM-ZOTS		iMU-ZOTS		
		<i>N</i>	Mean rules	<i>N</i>	Mean rules	<i>N</i>	$M_{th}$	Mean Rules
$TSS_3$	0.3	<b>10</b>		<u>7</u>		<u>7.11</u>	0.5	6.43
	0.5	10		10		6.5	0.5	
	0.7	10	10	10	9.25	6.11	0.5	
	0.85	10		10		6	0.5	
$S_{NH_3}$	0.3	5		2		6.66	0.9	7.16
	0.5	3		2		<u>6.5</u>	0.9	
	0.7	3	3.5	2	2	7.82	0.5	
	0.85	<b>3</b>		<u>2</u>		7.66	0.9	
$S_{NH_{10}}$	0.3	10		<u>10</u>		<b>7.93</b>	0.5	7.97
	0.5	<u>10</u>		10		8	0.8	
	0.7	-	10	-	10	-	-	
	0.85	-		-		-	-	
$S_{NH_{13}}$	0.3	<b>10</b>		7		<u>8.6</u>	0.5	8.86
	0.5	7		<u>3</u>		8.83	0.5	
	0.7	3	5.75	10	7.5	9	0.5	
	0.85	3		10		9	0.5	
$S_{NO_3}$	0.3	3		3		<u>7.92</u>	0.8	8.51
	0.5	3		2		8.1	0.8	
	0.7	5	4.5	<u>2</u>	4.25	9	0.9	
	0.85	<b>7</b>		10		9	0.5	
$S_{NO_{10}}$	0.3	<b>10</b>		10		<u>8.33</u>	0.9	8.08
	0.5	3		<u>2</u>		8	0.8	
	0.7	3	4.75	2	7.5	8	0.8	
	0.85	3		2		8	0.8	
$S_{NO_{13}}$	0.3	<b>10</b>		<u>7</u>		<u>8.75</u>	0.9	8.88
	0.5	10		7		8.75	0.9	
	0.7	10	8.25	7	7	9	0.7	
	0.85	3		7		9	0.7	
$S_{O_{10}}$	0.3	<b>10</b>		10		<u>8.75</u>	0.8	8.94
	0.5	5		<u>7</u>		9	0.5	
	0.7	10	8.75	7	7.75	9	0.5	
	0.85	10		7		9	0.5	
$BOD_{13}$	0.3	<b>10</b>		<u>10</u>		7.35	0.9	5.86
	0.5	10		10		6.85	0.8	
	0.7	10	8.25	10	8	<u>6.25</u>	0.8	
	0.85	3		2		3	0.8	
$COD_{13}$	0.3	<b>10</b>		<u>10</u>		<u>7.76</u>	0.5	7.38
	0.5	3		10		7	0.7	
	0.7	-	6.5	-	10	-	-	
	0.85	-		-		-	-	

The best model from the iMU-ZOTS method is described by 16 selected input variables by on average 6.5 rules by input variable. This number of average fuzzy rules is the sum of the rules for all the variables divided by the number of variables. This model is composed of a total of  $N = 104$  fuzzy rules, where for each the 16 variables, the numbers of rules is  $N_1 = 6$ ,  $N_2 = 7$ ,  $N_3 = 6$ ,  $N_4 = 9$ ,  $N_5 = 8$ ,  $N_6 = 8$ ,  $N_7 = 9$ ,  $N_8 = 6$ ,  $N_9 = 7$ ,  $N_{10} = 5$ ,  $N_{11} = 3$ ,  $N_{12} = 4$ ,  $N_{13} = 8$ ,  $N_{14} = 8$ ,  $N_{15} = 6$ , and  $N_{16} = 4$ . Figure 6 shows the MFs for the 16 input variables.

The other variable analyzed is  $S_{NO_3}$ . The results of the predictions of  $S_{NO_3}$  made by the 3 methods are shown in Figure 7. It can be seen that all the methods obtain good results, with no major differences between them, which leads to the conclusion that accuracy can be achieved by any of them. Therefore, since accuracy can be achieved by any of the methods, it is necessary to analyze interpretability by analyzing the number of fuzzy rules used by each method.

The learned FCM-LSM is described by 4 input variables and 7 fuzzy rules. Figure 8 shows the MFs for the 4 input variables. The fuzzy rules are described by:

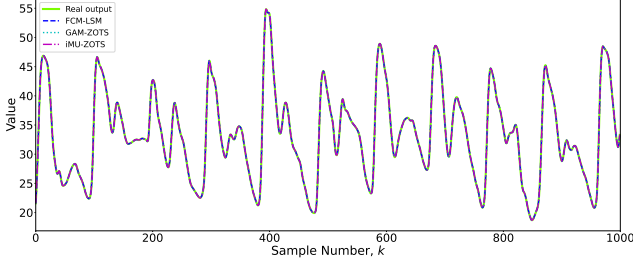


Figure 3: Prediction results for the variable  $S_{NH_3}$ .

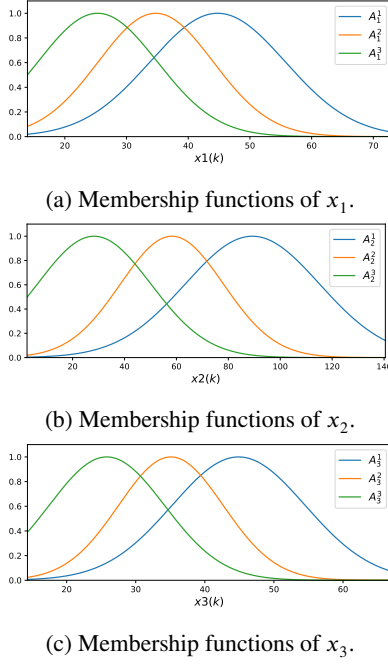
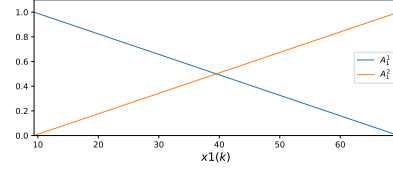
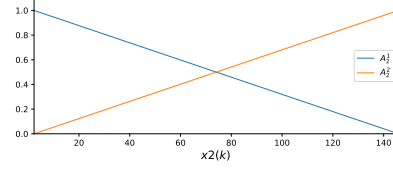


Figure 4: Membership functions of the FCM-LSM for the variable  $S_{NH_3}$ .

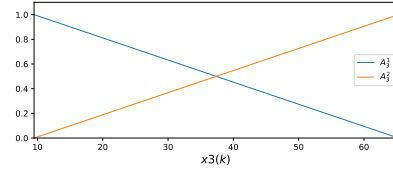
$R_1$  : IF  $x_1(k)$  is  $A_1^1$  and  $x_2(k)$  is  $A_2^1$  and  $x_3(k)$  is  $A_3^1$  and  $x_4(k)$  is  $A_4^1$   
 THEN  $y_1(k) = 1 - 4.28 \times 10^{-13} x_1(k) + 2.26 \times 10^{-13} x_2(k) + 4.92 \times 10^{-15} x_3(k) + 1.67 \times 10^{-12} x_4(k)$ ,  
 $R_2$  : IF  $x_1(k)$  is  $A_1^2$  and  $x_2(k)$  is  $A_2^2$  and  $x_3(k)$  is  $A_3^2$  and  $x_4(k)$  is  $A_4^2$   
 THEN  $y_2(k) = 1 + 7.88 \times 10^{-15} x_1(k) + 3.03 \times 10^{-14} x_2(k) + 3.29 \times 10^{-17} x_3(k) - 1.50 \times 10^{-13} x_4(k)$ ,  
 $R_3$  : IF  $x_1(k)$  is  $A_1^3$  and  $x_2(k)$  is  $A_2^3$  and  $x_3(k)$  is  $A_3^3$  and  $x_4(k)$  is  $A_4^3$   
 THEN  $y_3(k) = 1 - 1.75 \times 10^{-13} x_1(k) + 4.85 \times 10^{-14} x_2(k) + 1.79 \times 10^{-15} x_3(k) + 4.23 \times 10^{-13} x_4(k)$ ,  
 $R_4$  : IF  $x_1(k)$  is  $A_1^4$  and  $x_2(k)$  is  $A_2^4$  and  $x_3(k)$  is  $A_3^4$  and  $x_4(k)$  is  $A_4^4$



(a) Membership functions of  $x_1$ .



(b) Membership functions of  $x_2$ .



(c) Membership functions of  $x_3$ .

Figure 5: Membership functions of the GAM-ZOTS method for the variable  $S_{NH_3}$ .

THEN  $y_4(k) = 1 + 1.33 \times 10^{-15} x_1(k) - 3.34 \times 10^{-14} x_2(k) - 1.02 \times 10^{-16} x_3(k) + 3.10 \times 10^{-13} x_4(k)$ ,  
 $R_5$  : IF  $x_1(k)$  is  $A_1^5$  and  $x_2(k)$  is  $A_2^5$  and  $x_3(k)$  is  $A_3^5$  and  $x_4(k)$  is  $A_4^5$   
 THEN  $y_5(k) = 1 - 4.29 \times 10^{-14} x_1(k) + 1.31 \times 10^{-13} x_2(k) + 1.68 \times 10^{-15} x_3(k) + 9.22 \times 10^{-13} x_4(k)$ ,  
 $R_6$  : IF  $x_1(k)$  is  $A_1^6$  and  $x_2(k)$  is  $A_2^6$  and  $x_3(k)$  is  $A_3^6$  and  $x_4(k)$  is  $A_4^6$   
 THEN  $y_6(k) = 1 - 2.97 \times 10^{-13} x_1(k) + 3.37 \times 10^{-13} x_2(k) + 5.68 \times 10^{-15} x_3(k) + 2.03 \times 10^{-12} x_4(k)$ ,  
 $R_7$  : IF  $x_1(k)$  is  $A_1^7$  and  $x_2(k)$  is  $A_2^7$  and  $x_3(k)$  is  $A_3^7$  and  $x_4(k)$  is  $A_4^7$   
 THEN  $y_7(k) = 1 - 5.02 \times 10^{-13} x_1(k) + 3.79 \times 10^{-13} x_2(k) + 7.09 \times 10^{-15} x_3(k) + 3.01 \times 10^{-12} x_4(k)$ .

The GAM-ZOTS method describes each of the 7 selected input variables by 2 fuzzy rules. Figure 9 shows the MFs for the 7 input variables. The fuzzy rules of GAM-ZOTS for the variable  $x_1$  are described by:

$R_1^1$ : IF  $x_1(k)$  is  $A_1^1$  THEN  $y_1^1(k) = 0.02$ ,  
 $R_2^1$ : IF  $x_1(k)$  is  $A_2^1$  THEN  $y_1^1(k) = -0.06$ ,

for the variable  $x_2$  by:

$R_1^2$ : IF  $x_2(k)$  is  $A_1^2$  THEN  $y_2^1(k) = 0.03$ ,  
 $R_2^2$ : IF  $x_2(k)$  is  $A_2^2$  THEN  $y_2^1(k) = -0.13$ ,

for the variable  $x_3$  by:



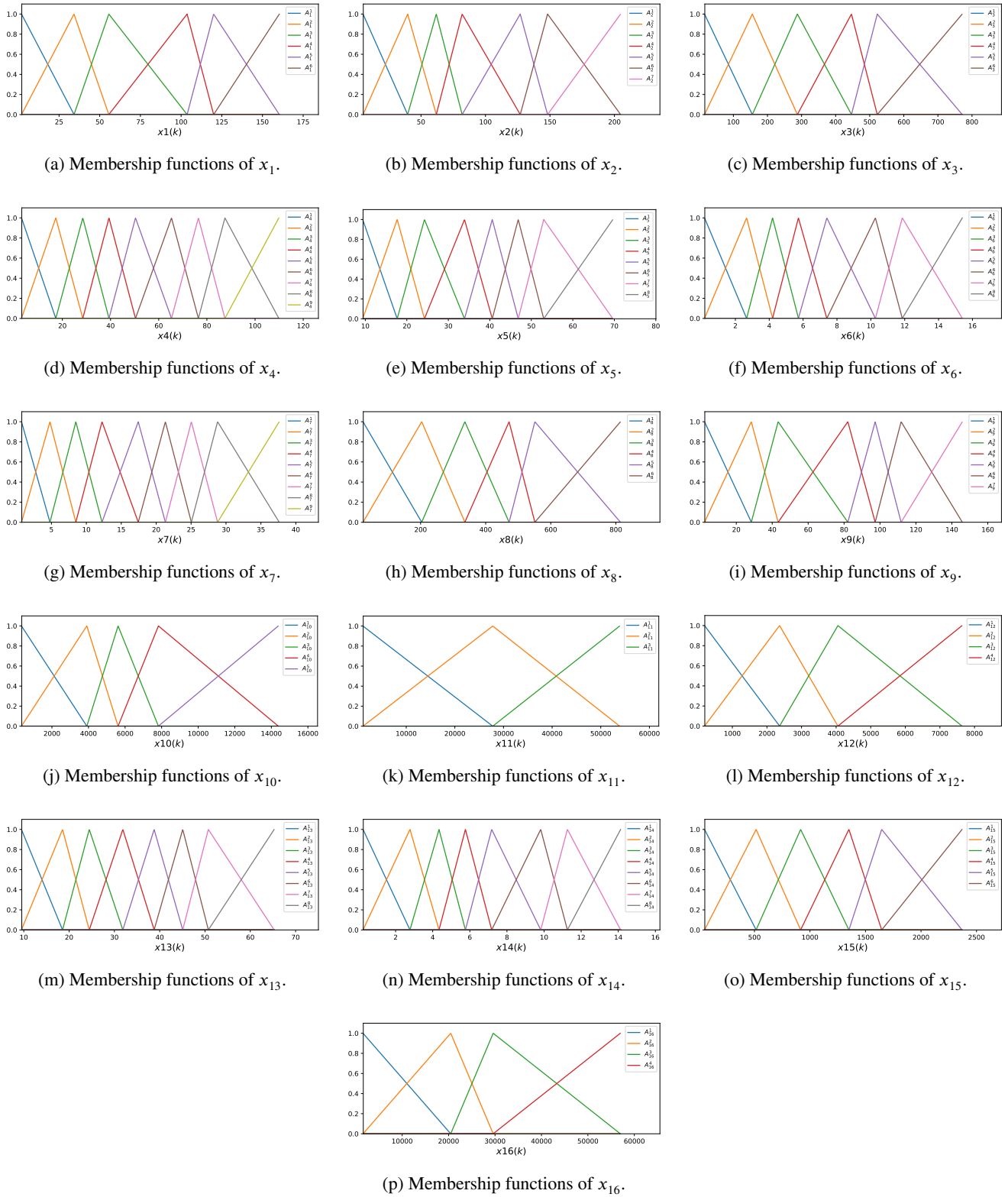


Figure 6: Membership functions of the iMU-ZOTS method for the variable  $S_{NH_3}$ .

$$\begin{aligned} R_3^1: & \text{ IF } x_3(k) \text{ is } A_3^1 \text{ THEN } y_3^1(k) = -0.05, \\ R_3^2: & \text{ IF } x_3(k) \text{ is } A_3^2 \text{ THEN } y_3^2(k) = 0.18, \end{aligned}$$

$$\begin{aligned} R_4^1: & \text{ IF } x_4(k) \text{ is } A_4^1 \text{ THEN } y_4^1(k) = 0.04, \\ R_4^2: & \text{ IF } x_4(k) \text{ is } A_4^2 \text{ THEN } y_4^2(k) = -0.01, \end{aligned}$$

for the variable  $x_4$  by:

for the variable  $x_5$  by:

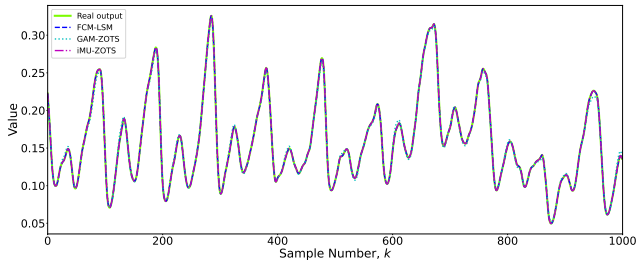
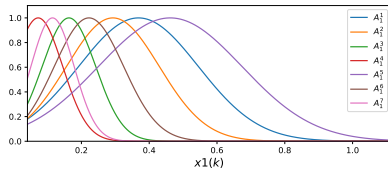
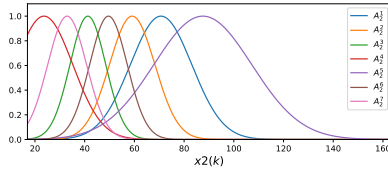


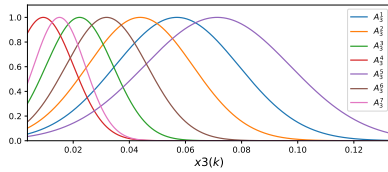
Figure 7: Prediction results for the variable  $S_{NO_3}$ .



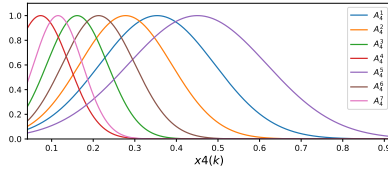
(a) Membership functions of  $x_1$ .



(b) Membership functions of  $x_2$ .



(c) Membership functions of  $x_3$ .



(d) Membership functions of  $x_4$ .

Figure 8: Membership functions of the FCM-LSM for the variable  $S_{NO_3}$ .

$$\begin{aligned} R_5^1: & \text{ IF } x_5(k) \text{ is } A_5^1 \text{ THEN } y_5^1(k) = -0.03, \\ R_5^2: & \text{ IF } x_5(k) \text{ is } A_5^2 \text{ THEN } y_5^2(k) = 0.07, \end{aligned}$$

for the variable  $x_6$  by:

$$\begin{aligned} R_6^1: & \text{ IF } x_6(k) \text{ is } A_6^1 \text{ THEN } y_6^1(k) = -0.03, \\ R_6^2: & \text{ IF } x_6(k) \text{ is } A_6^2 \text{ THEN } y_6^2(k) = 0.11, \end{aligned}$$

and for the variable  $x_7$  by:

$$\begin{aligned} R_7^1: & \text{ IF } x_7(k) \text{ is } A_7^1 \text{ THEN } y_7^1(k) = -0.12, \\ R_7^2: & \text{ IF } x_7(k) \text{ is } A_7^2 \text{ THEN } y_7^2(k) = 0.33. \end{aligned}$$

Finally, the iMU-ZOTS model is described by 13 selected input variables and by an average of 7.92 fuzzy rules

per input variable. This average number of fuzzy rules is the sum of the number of rules for all the variables divided by the number of variables. This model is made up of a total of  $N = 103$  fuzzy rules, namely  $N_1 = 9$ ,  $N_2 = 9$ ,  $N_3 = 9$ ,  $N_4 = 9$ ,  $N_5 = 5$ ,  $N_6 = 4$ ,  $N_7 = 10$ ,  $N_8 = 7$ ,  $N_9 = 8$ ,  $N_{10} = 10$ ,  $N_{11} = 10$ ,  $N_{12} = 7$  and  $N_{13} = 6$ . Figure 10 shows the membership functions for the 13 input variables.

After analyzing the results obtained for the above two target variables, it can be seen that although the FCM-LSM has a lower error value than the other methods, the difference in predictions is insignificant. In addition, it can be seen that FCM-LSM obtains the best results, but with more complex rules compared to the other methods. A model with a high number of rules or with several variables on the antecedent part becomes complex and difficult to interpret by the operators, who in these cases of WWTPs need to understand the models easily to make the right decisions throughout the wastewater treatment process. An interpretable system allows the actions taken in a WWTP to be explained and easily adjusted to requirements.

For the variable  $S_{NH_3}$ , the FCM-LSM obtains an error value of  $MSE=1.06 \times 10^{-23}$  with 3 associated fuzzy rules. GAM-ZOTS obtains an error value of  $MSE=7.90 \times 10^{-12}$ ,  $7.46 \times 10^{13}\%$  higher than the FCM-LSM error, with two univariate fuzzy rules per variable. iMU-ZOTS has an MSE value of  $1.28 \times 10^{-4}$ ,  $1.21 \times 10^{21}\%$  higher than the FCM-LSM error and 6.5 fuzzy rules per variable. Since the forecast made by the 3 methods is identical (see Figure 3), the choice of the best method involves the one with the least complexity, which for this variable is GAM-ZOTS. For this variable in particular, FCM-LSM would also be a possible right choice since it has the lowest error and a low number of fuzzy rules.

In the case of the  $S_{NO_3}$  variable, FCM-LSM obtains an error value of  $MSE=2.92 \times 10^{-28}$ , and 7 fuzzy rules. GAM-ZOTS obtains an error value of  $MSE=1.16 \times 10^{-5}$ ,  $3.97 \times 10^{24}\%$  higher than the FCM-LSM error and 2 fuzzy rules per variable. iMU-ZOTS has an error value of  $MSE=2.49 \times 10^{-9}$ ,  $8.53 \times 10^{20}\%$  higher than the FCM-LSM error and 7.92 fuzzy rules per variable. An analysis of Figure 7 shows that the differences in the prediction of this variable by the different methods are insignificant, which means that the most appropriate method for this variable should be chosen based on the least complex model. In the case of this variable, GAM-ZOTS is the one with the best balance between accuracy and interpretability, allowing for accurate predictions that are easy for operators to understand.

## 5. Conclusions

The main aim of this work is to create an interpretable digital twin of the most relevant variables on the waterline of a WWTP. Through digital twins, it is possible to create a replica of these treatment plants, allowing variables to be monitored throughout the process. This makes it possible to understand the system's behavior better and helps operators make the right decisions. The paper studied three fuzzy-based models for the digital twin, focusing on the accuracy

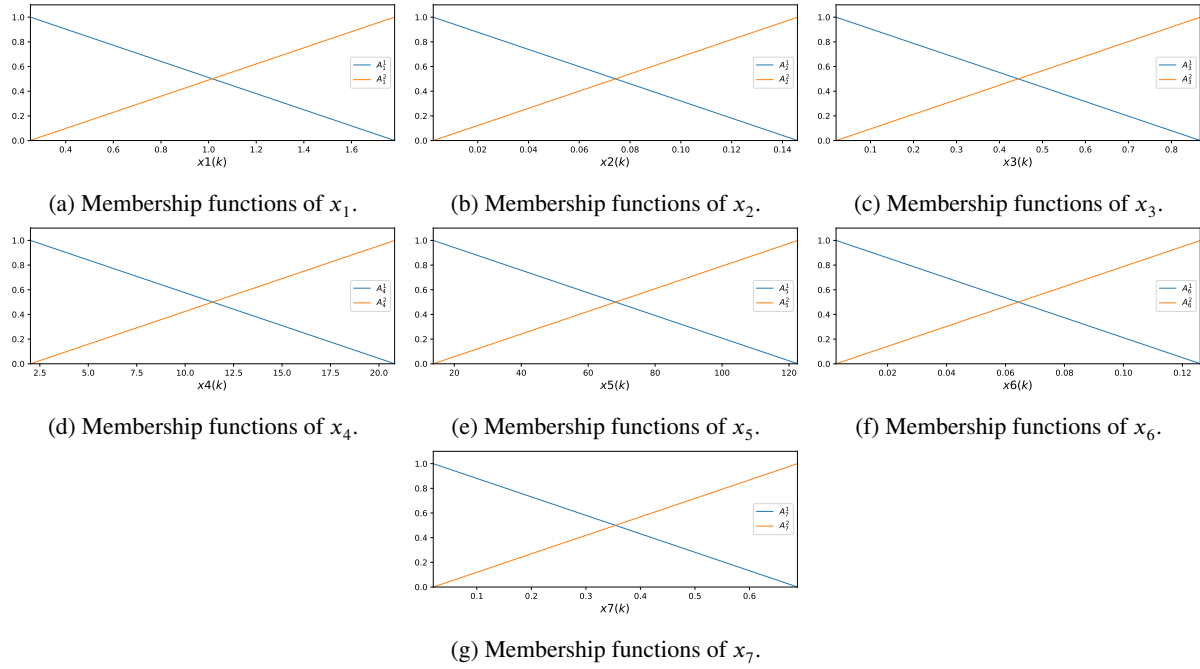


Figure 9: Membership functions of the GAM-ZOTS method for the variable  $S_{NO_3}$ .

and interpretability of these models. Several datasets were constructed for each target variable using different sets of input variables. These datasets were used to learn each one of the three models, which use different combinations of model parameters, and in the end, the best results from each algorithm were selected for each variable. From the results, it was found that the model that obtained the best accuracy results for most of the variables has a more complex structure than the other models, despite being the model with the best error result. To make a more detailed analysis of the predictions made by each model, 2 variables, as case studies, were selected. It was found that the differences in the error obtained by the three models were minimal and not very significant. Analyzing the fuzzy rules of each model showed that there was a less complex and more interpretable model than the one that obtained the best error value. It can, therefore, be concluded that the choice of the best model for predicting each variable must take into account not only the accuracy but also the interpretability and complexity of each model. As the complexity of the model increases, it becomes more difficult to understand and interpret. For the variables analyzed, the GAM-ZOTS method showed the best balance between accuracy and interpretability.

## Acknowledgment

Rodrigo Salles is supported by Fundação para a Ciência e a Tecnologia (FCT) under the grant ref. 2023.01009.BD. This research is partially sponsored by national funds through FCT under projects UIDB/00285/2020 and LA/P/0112/2020.

## References

- [1] Darya Botkina, Mikael Hedlind, Bengt Olsson, Jannik Henser, and Thomas Lundholm. Digital twin of a cutting tool. *Procedia CIRP*, 72:215–218, 2018. 51st CIRP Conference on Manufacturing Systems.
- [2] Sudip Misra, Chandana Roy, Thilo Sauter, Anandarup Mukherjee, and Jhareswar Maiti. Industrial internet of things for safety management applications: A survey. *IEEE Access*, 10:83415–83439, 2022.
- [3] Klaus Schwab. *The Fourth Industrial Revolution*. World Economic Forum, Geneva, 2016.
- [4] Rodrigo Salles, Jérôme Mendes, Rui Araújo, Carlos Melo, and Pedro Moura. Prediction of key variables in wastewater treatment plants using machine learning models. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE, 2022.
- [5] AMERICA'S INFRASTRUCTURE REPORT CARD. Wastewater, asce, usa, tech. rep., 2021. [urlhttps://infrastructurereportcard.org/cat-item/wastewater/](https://infrastructurereportcard.org/cat-item/wastewater/), 2021.
- [6] A EurEau. Europe's water in figures—an overview of the european drinking water and waste water sectors, 2017.
- [7] Michael Grieves. Digital twin: manufacturing excellence through virtual factory replication. *White paper*, 1(2014):1–7, 2014.
- [8] Fei Tao, He Zhang, Ang Liu, and Andrew YC Nee. Digital twin in industry: State-of-the-art. *IEEE Transactions on industrial informatics*, 15(4):2405–2415, 2018.
- [9] Eric J Tuegel, Anthony R Ingraffea, Thomas G Eason, S Michael Spottswood, et al. Reengineering aircraft structural life prediction using a digital twin. *International Journal of Aerospace Engineering*, 2011, 2011.
- [10] Edward Glaessgen and David Stargel. The digital twin paradigm for future nasa and us air force vehicles. In *53rd AIAA/ASME/ASCE/AH-S/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA*, page 1818, 2012.
- [11] Adil Rasheed, Omer San, and Trond Kvamsdal. Digital twin: Values, challenges and enablers. *arXiv preprint arXiv:1910.01719*, 2019.
- [12] Seyed Mostafa Hallaji, Yihai Fang, and Brandon K Winfrey. A digital twin framework for enhancing predictive maintenance of pumps in wastewater treatment plants. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*,

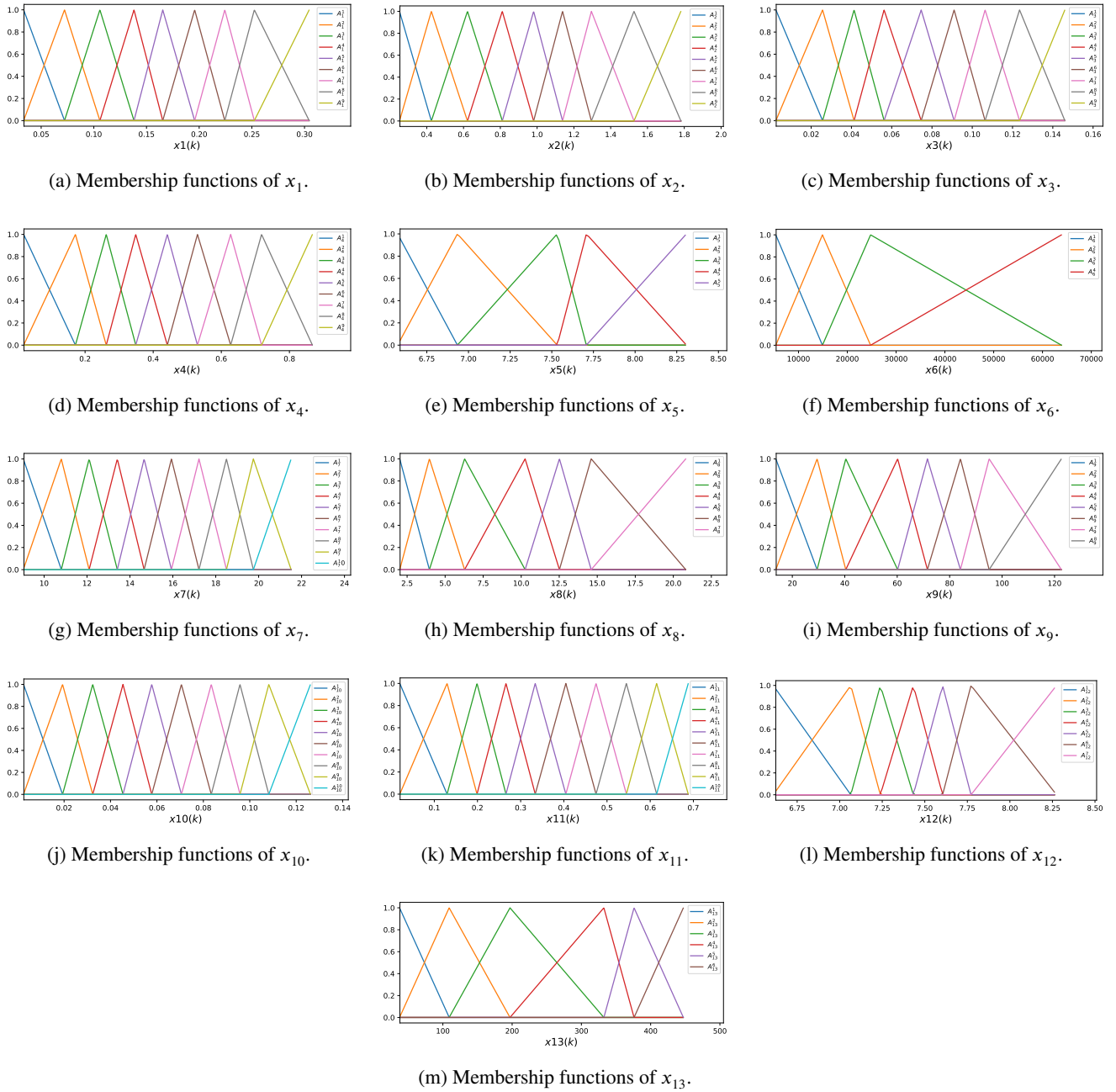


Figure 10: Membership functions of the iMU-ZOTS method for the variable  $S_{NO_3^-}$ .

- volume 38, pages 88–93. IAARC Publications, 2021.
- [13] Wentao Liu, Sudao He, Jianpeng Mou, Ting Xue, Hongtian Chen, and Weili Xiong. Digital twins-based process monitoring for wastewater treatment processes. *Reliability Engineering & System Safety*, page 109416, 2023.
- [14] Jérôme Mendes, Francisco Souza, Rui Araújo, and Saeid Rastegar. Neo-fuzzy neuron learning using backfitting algorithm. *Neural Computing and Applications*, 31:3609–3618, 2019.
- [15] Jerome Mendes, Francisco AA Souza, Ricardo Maia, and Rui Araújo. Iterative learning of multiple univariate zero-order ts fuzzy systems. In *IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society*, volume 1, pages 3803–3808. IEEE, 2019.
- [16] Jorge S. S. Júnior, Jérôme Mendes, Francisco Souza, and Cristiano Premebida. Survey on deep fuzzy systems in regression applications: A view on interpretability. *International Journal of Fuzzy Systems*, 25(7):2568–2589, 2023.
- [17] Tomohiro Takagi and Michio Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE transactions on systems, man, and cybernetics*, 1:116–132, 1985.
- [18] Dejan Dovžan and Igor Škrjanc. Recursive fuzzy c-means clustering for recursive fuzzy identification of time-varying processes. *ISA transactions*, 50(2):159–169, 2011.
- [19] Michal Preisner, Elena Neverova-Dziopak, and Zbigniew Kowalewski. An analytical review of different approaches to wastewater discharge standards with particular emphasis on nutrients. *Environmental Management*, 66:694–708, 2020.
- [20] Bárbara de Matos, Rodrigo Salles, Jérôme Mendes, Joana R. Gouveia, António J. Baptista, and Pedro Moura. A review of energy and sustainability kpi-based monitoring and control methodologies on wwtps. *Mathematics*, 11(1), 2023.

- [21] Keidy Morales-Rodelo, Mario Francisco, Hernan Alvarez, Pastora Vega, and Silvana Revollar. Collaborative control applied to bsm1 for wastewater treatment plants. *Processes*, 8(11):1465, 2020.
- [22] Rubén Moliner-Heredia, Ignacio Peñarrocha-Alós, and Roberto Sanchis-Llopis. Economic model predictive control of wastewater treatment plants based on bsm1 using linear prediction models. In *2019 IEEE 15th International Conference on Control and Automation (ICCA)*, pages 73–78. IEEE, 2019.
- [23] Wenliang Chen, Chonghua Yao, and Xiwu Lu. Optimal design activated sludge process by means of multi-objective optimization: case study in benchmark simulation model 1 (bsm1). *Water science and technology*, 69(10):2052–2058, 2014.

# Appendix B

## Fuzzy rules from iMU-ZOTS for $S_{NH_3}$

The fuzzy rules for the variable  $x_1$  are described by:

$$\begin{aligned} R_1^1: & \text{ IF } x_1(k) \text{ is } A_1^1 \text{ THEN } y_1^1(k) = -0.04, \\ R_1^2: & \text{ IF } x_1(k) \text{ is } A_1^2 \text{ THEN } y_1^2(k) = -0.01, \\ R_1^3: & \text{ IF } x_1(k) \text{ is } A_1^3 \text{ THEN } y_1^3(k) = -0.001, \\ R_1^4: & \text{ IF } x_1(k) \text{ is } A_1^4 \text{ THEN } y_1^4(k) = 0.04, \\ R_1^5: & \text{ IF } x_1(k) \text{ is } A_1^5 \text{ THEN } y_1^5(k) = 0.05, \\ R_1^6: & \text{ IF } x_1(k) \text{ is } A_1^6 \text{ THEN } y_1^6(k) = 0.09. \end{aligned}$$

for the variable  $x_2$  by:

$$\begin{aligned} R_2^1: & \text{ IF } x_2(k) \text{ is } A_2^1 \text{ THEN } y_2^1(k) = -0.14, \\ R_2^2: & \text{ IF } x_2(k) \text{ is } A_2^2 \text{ THEN } y_2^2(k) = -0.08, \\ R_2^3: & \text{ IF } x_2(k) \text{ is } A_2^3 \text{ THEN } y_2^3(k) = -0.05, \\ R_2^4: & \text{ IF } x_2(k) \text{ is } A_2^4 \text{ THEN } y_2^4(k) = -0.02, \\ R_2^5: & \text{ IF } x_2(k) \text{ is } A_2^5 \text{ THEN } y_2^5(k) = 0.06, \\ R_2^6: & \text{ IF } x_2(k) \text{ is } A_2^6 \text{ THEN } y_2^6(k) = 0.09, \\ R_2^7: & \text{ IF } x_2(k) \text{ is } A_2^7 \text{ THEN } y_2^7(k) = 0.14. \end{aligned}$$

for the variable  $x_3$  by:

$$\begin{aligned} R_3^1: & \text{ IF } x_3(k) \text{ is } A_3^1 \text{ THEN } y_3^1(k) = -0.13, \\ R_3^2: & \text{ IF } x_3(k) \text{ is } A_3^2 \text{ THEN } y_3^2(k) = -0.06, \\ R_3^3: & \text{ IF } x_3(k) \text{ is } A_3^3 \text{ THEN } y_3^3(k) = -0.008, \\ R_3^4: & \text{ IF } x_3(k) \text{ is } A_3^4 \text{ THEN } y_3^4(k) = 0.05, \\ R_3^5: & \text{ IF } x_3(k) \text{ is } A_3^5 \text{ THEN } y_3^5(k) = 0.01, \\ R_3^6: & \text{ IF } x_3(k) \text{ is } A_3^6 \text{ THEN } y_3^6(k) = -0.21. \end{aligned}$$

for the variable  $x_4$  by:

$$\begin{aligned} R_4^1: & \text{ IF } x_4(k) \text{ is } A_4^1 \text{ THEN } y_4^1(k) = -0.01, \\ R_4^2: & \text{ IF } x_4(k) \text{ is } A_4^2 \text{ THEN } y_4^2(k) = -0.002, \\ R_4^3: & \text{ IF } x_4(k) \text{ is } A_4^3 \text{ THEN } y_4^3(k) = 0.002, \\ R_4^4: & \text{ IF } x_4(k) \text{ is } A_4^4 \text{ THEN } y_4^4(k) = 0.004, \\ R_4^5: & \text{ IF } x_4(k) \text{ is } A_4^5 \text{ THEN } y_4^5(k) = 0.002, \\ R_4^6: & \text{ IF } x_4(k) \text{ is } A_4^6 \text{ THEN } y_4^6(k) = 9.89 \times 10^{-4}, \\ R_4^7: & \text{ IF } x_4(k) \text{ is } A_4^7 \text{ THEN } y_4^7(k) = -0.008, \\ R_4^8: & \text{ IF } x_4(k) \text{ is } A_4^8 \text{ THEN } y_4^8(k) = -0.026, \\ R_4^9: & \text{ IF } x_4(k) \text{ is } A_4^9 \text{ THEN } y_4^9(k) = -0.05. \end{aligned}$$

for the variable  $x_5$  by:

$$\begin{aligned}
R_5^1: & \text{ IF } x_5(k) \text{ is } A_5^1 \text{ THEN } y_5^1(k) = -0.005, \\
R_5^2: & \text{ IF } x_5(k) \text{ is } A_5^2 \text{ THEN } y_5^2(k) = -0.008, \\
R_5^3: & \text{ IF } x_5(k) \text{ is } A_5^3 \text{ THEN } y_5^3(k) = -7.29 \times 10^{-4}, \\
R_5^4: & \text{ IF } x_5(k) \text{ is } A_5^4 \text{ THEN } y_5^4(k) = 0.001, \\
R_5^5: & \text{ IF } x_5(k) \text{ is } A_5^5 \text{ THEN } y_5^5(k) = -1.02 \times 10^{-4}, \\
R_5^6: & \text{ IF } x_5(k) \text{ is } A_5^6 \text{ THEN } y_5^6(k) = -4.72 \times 10^{-5}, \\
R_5^7: & \text{ IF } x_5(k) \text{ is } A_5^7 \text{ THEN } y_5^7(k) = 0.003, \\
R_5^8: & \text{ IF } x_5(k) \text{ is } A_5^8 \text{ THEN } y_5^8(k) = -0.01.
\end{aligned}$$

for the variable  $x_6$  by:

$$\begin{aligned}
R_6^1: & \text{ IF } x_6(k) \text{ is } A_6^1 \text{ THEN } y_6^1(k) = -0.003, \\
R_6^2: & \text{ IF } x_6(k) \text{ is } A_6^2 \text{ THEN } y_6^2(k) = 0.002, \\
R_6^3: & \text{ IF } x_6(k) \text{ is } A_6^3 \text{ THEN } y_6^3(k) = 0.001, \\
R_6^4: & \text{ IF } x_6(k) \text{ is } A_6^4 \text{ THEN } y_6^4(k) = -3.74 \times 10^{-4}, \\
R_6^5: & \text{ IF } x_6(k) \text{ is } A_6^5 \text{ THEN } y_6^5(k) = -0.001, \\
R_6^6: & \text{ IF } x_6(k) \text{ is } A_6^6 \text{ THEN } y_6^6(k) = -0.003, \\
R_6^7: & \text{ IF } x_6(k) \text{ is } A_6^7 \text{ THEN } y_6^7(k) = 4.31 \times 10^{-4}, \\
R_6^8: & \text{ IF } x_6(k) \text{ is } A_6^8 \text{ THEN } y_6^8(k) = -0.01.
\end{aligned}$$

for the variable  $x_7$  by:

$$\begin{aligned}
R_7^1: & \text{ IF } x_7(k) \text{ is } A_7^1 \text{ THEN } y_7^1(k) = 0.03, \\
R_7^2: & \text{ IF } x_7(k) \text{ is } A_7^2 \text{ THEN } y_7^2(k) = 0.02, \\
R_7^3: & \text{ IF } x_7(k) \text{ is } A_7^3 \text{ THEN } y_7^3(k) = 0.01, \\
R_7^4: & \text{ IF } x_7(k) \text{ is } A_7^4 \text{ THEN } y_7^4(k) = 0.004, \\
R_7^5: & \text{ IF } x_7(k) \text{ is } A_7^5 \text{ THEN } y_7^5(k) = -0.008, \\
R_7^6: & \text{ IF } x_7(k) \text{ is } A_7^6 \text{ THEN } y_7^6(k) = -0.01, \\
R_7^7: & \text{ IF } x_7(k) \text{ is } A_7^7 \text{ THEN } y_7^7(k) = -0.01, \\
R_7^8: & \text{ IF } x_7(k) \text{ is } A_7^8 \text{ THEN } y_7^8(k) = -0.02, \\
R_7^9: & \text{ IF } x_7(k) \text{ is } A_7^9 \text{ THEN } y_7^9(k) = -0.02.
\end{aligned}$$

for the variable  $x_8$  by:

$$\begin{aligned}
R_8^1: & \text{ IF } x_8(k) \text{ is } A_8^1 \text{ THEN } y_8^1(k) = 0.36, \\
R_8^2: & \text{ IF } x_8(k) \text{ is } A_8^2 \text{ THEN } y_8^2(k) = 0.15, \\
R_8^3: & \text{ IF } x_8(k) \text{ is } A_8^3 \text{ THEN } y_8^3(k) = 0.01, \\
R_8^4: & \text{ IF } x_8(k) \text{ is } A_8^4 \text{ THEN } y_8^4(k) = -0.12, \\
R_8^5: & \text{ IF } x_8(k) \text{ is } A_8^5 \text{ THEN } y_8^5(k) = -0.11, \\
R_8^6: & \text{ IF } x_8(k) \text{ is } A_8^6 \text{ THEN } y_8^6(k) = 0.03.
\end{aligned}$$

for the variable  $x_9$  by:

$$\begin{aligned}
R_9^1: & \text{ IF } x_9(k) \text{ is } A_9^1 \text{ THEN } y_9^1(k) = 0.06, \\
R_9^2: & \text{ IF } x_9(k) \text{ is } A_9^2 \text{ THEN } y_9^2(k) = 0.03, \\
R_9^3: & \text{ IF } x_9(k) \text{ is } A_9^3 \text{ THEN } y_9^3(k) = 0.01, \\
R_9^4: & \text{ IF } x_9(k) \text{ is } A_9^4 \text{ THEN } y_9^4(k) = -0.02, \\
R_9^5: & \text{ IF } x_9(k) \text{ is } A_9^5 \text{ THEN } y_9^5(k) = -0.04, \\
R_9^6: & \text{ IF } x_9(k) \text{ is } A_9^6 \text{ THEN } y_9^6(k) = -0.06, \\
R_9^7: & \text{ IF } x_9(k) \text{ is } A_9^7 \text{ THEN } y_9^7(k) = -0.09.
\end{aligned}$$

for the variable  $x_{10}$  by:

$$\begin{aligned} R_{10}^1: & \text{ IF } x_{10}(k) \text{ is } A_{10}^1 \text{ THEN } y_{10}^1(k) = -5.88, \\ R_{10}^2: & \text{ IF } x_{10}(k) \text{ is } A_{10}^2 \text{ THEN } y_{10}^2(k) = -2.49, \\ R_{10}^3: & \text{ IF } x_{10}(k) \text{ is } A_{10}^3 \text{ THEN } y_{10}^3(k) = -0.87, \\ R_{10}^4: & \text{ IF } x_{10}(k) \text{ is } A_{10}^4 \text{ THEN } y_{10}^4(k) = 1.22, \\ R_{10}^5: & \text{ IF } x_{10}(k) \text{ is } A_{10}^5 \text{ THEN } y_{10}^5(k) = 7.70. \end{aligned}$$

for the variable  $x_{11}$  by:

$$\begin{aligned} R_{11}^1: & \text{ IF } x_{11}(k) \text{ is } A_{11}^1 \text{ THEN } y_{11}^1(k) = -22.82, \\ R_{11}^2: & \text{ IF } x_{11}(k) \text{ is } A_{11}^2 \text{ THEN } y_{11}^2(k) = 2.01, \\ R_{11}^3: & \text{ IF } x_{11}(k) \text{ is } A_{11}^3 \text{ THEN } y_{11}^3(k) = 29.42. \end{aligned}$$

for the variable  $x_{12}$  by:

$$\begin{aligned} R_{12}^1: & \text{ IF } x_{12}(k) \text{ is } A_{12}^1 \text{ THEN } y_{12}^1(k) = -3.24, \\ R_{12}^2: & \text{ IF } x_{12}(k) \text{ is } A_{12}^2 \text{ THEN } y_{12}^2(k) = -1.15, \\ R_{12}^3: & \text{ IF } x_{12}(k) \text{ is } A_{12}^3 \text{ THEN } y_{12}^3(k) = 0.49, \\ R_{12}^4: & \text{ IF } x_{12}(k) \text{ is } A_{12}^4 \text{ THEN } y_{12}^4(k) = 4.16. \end{aligned}$$

for the variable  $x_{13}$  by:

$$\begin{aligned} R_{13}^1: & \text{ IF } x_{13}(k) \text{ is } A_{13}^1 \text{ THEN } y_{13}^1(k) = -24.83, \\ R_{13}^2: & \text{ IF } x_{13}(k) \text{ is } A_{13}^2 \text{ THEN } y_{13}^2(k) = -15.74, \\ R_{13}^3: & \text{ IF } x_{13}(k) \text{ is } A_{13}^3 \text{ THEN } y_{13}^3(k) = -9.84, \\ R_{13}^4: & \text{ IF } x_{13}(k) \text{ is } A_{13}^4 \text{ THEN } y_{13}^4(k) = -2.43, \\ R_{13}^5: & \text{ IF } x_{13}(k) \text{ is } A_{13}^5 \text{ THEN } y_{13}^5(k) = 4.47, \\ R_{13}^6: & \text{ IF } x_{13}(k) \text{ is } A_{13}^6 \text{ THEN } y_{13}^6(k) = 10.79, \\ R_{13}^7: & \text{ IF } x_{13}(k) \text{ is } A_{13}^7 \text{ THEN } y_{13}^7(k) = 16.44, \\ R_{13}^8: & \text{ IF } x_{13}(k) \text{ is } A_{13}^8 \text{ THEN } y_{13}^8(k) = 30.97. \end{aligned}$$

for the variable  $x_{14}$  by:

$$\begin{aligned} R_{14}^1: & \text{ IF } x_{14}(k) \text{ is } A_{14}^1 \text{ THEN } y_{14}^1(k) = 0.002, \\ R_{14}^2: & \text{ IF } x_{14}(k) \text{ is } A_{14}^2 \text{ THEN } y_{14}^2(k) = 2.47 \times 10^{-4}, \\ R_{14}^3: & \text{ IF } x_{14}(k) \text{ is } A_{14}^3 \text{ THEN } y_{14}^3(k) = -0.001, \\ R_{14}^4: & \text{ IF } x_{14}(k) \text{ is } A_{14}^4 \text{ THEN } y_{14}^4(k) = 1.48 \times 10^{-4}, \\ R_{14}^5: & \text{ IF } x_{14}(k) \text{ is } A_{14}^5 \text{ THEN } y_{14}^5(k) = 1.83 \times 10^{-4}, \\ R_{14}^6: & \text{ IF } x_{14}(k) \text{ is } A_{14}^6 \text{ THEN } y_{14}^6(k) = 7.01 \times 10^{-4}, \\ R_{14}^7: & \text{ IF } x_{14}(k) \text{ is } A_{14}^7 \text{ THEN } y_{14}^7(k) = 0.005, \\ R_{14}^8: & \text{ IF } x_{14}(k) \text{ is } A_{14}^8 \text{ THEN } y_{14}^8(k) = -0.003. \end{aligned}$$

for the variable  $x_{15}$  by:

$$\begin{aligned} R_{15}^1: & \text{ IF } x_{15}(k) \text{ is } A_{15}^1 \text{ THEN } y_{15}^1(k) = -0.03, \\ R_{15}^2: & \text{ IF } x_{15}(k) \text{ is } A_{15}^2 \text{ THEN } y_{15}^2(k) = -0.01, \\ R_{15}^3: & \text{ IF } x_{15}(k) \text{ is } A_{15}^3 \text{ THEN } y_{15}^3(k) = -0.002, \\ R_{15}^4: & \text{ IF } x_{15}(k) \text{ is } A_{15}^4 \text{ THEN } y_{15}^4(k) = 0.01, \\ R_{15}^5: & \text{ IF } x_{15}(k) \text{ is } A_{15}^5 \text{ THEN } y_{15}^5(k) = 0.01, \\ R_{15}^6: & \text{ IF } x_{15}(k) \text{ is } A_{15}^6 \text{ THEN } y_{15}^6(k) = 0.03. \end{aligned}$$

and for the variable  $x_{16}$  by:

$$\begin{aligned} R_{16}^1: & \text{ IF } x_{16}(k) \text{ is } A_{16}^1 \text{ THEN } y_{16}^1(k) = 31.46, \\ R_{16}^2: & \text{ IF } x_{16}(k) \text{ is } A_{16}^2 \text{ THEN } y_{16}^2(k) = 7.33, \\ R_{16}^3: & \text{ IF } x_{16}(k) \text{ is } A_{16}^3 \text{ THEN } y_{16}^3(k) = -4.38, \\ R_{16}^4: & \text{ IF } x_{16}(k) \text{ is } A_{16}^4 \text{ THEN } y_{16}^4(k) = -41.19. \end{aligned}$$





# Appendix C

## Fuzzy rules from iMU-ZOTS for $S_{NO_3}$

The fuzzy rules for the variable  $x_1$  are described by:

$$\begin{aligned} R_1^1: & \text{ IF } x_1(k) \text{ is } A_1^1 \text{ THEN } y_1^1(k) = 0.001, \\ R_1^2: & \text{ IF } x_1(k) \text{ is } A_1^2 \text{ THEN } y_1^2(k) = 9.25 \times 10^{-4}, \\ R_1^3: & \text{ IF } x_1(k) \text{ is } A_1^3 \text{ THEN } y_1^3(k) = 1.05 \times 10^{-4}, \\ R_1^4: & \text{ IF } x_1(k) \text{ is } A_1^4 \text{ THEN } y_1^4(k) = -7.01 \times 10^{-4}, \\ R_1^5: & \text{ IF } x_1(k) \text{ is } A_1^5 \text{ THEN } y_1^5(k) = -0.001, \\ R_1^6: & \text{ IF } x_1(k) \text{ is } A_1^6 \text{ THEN } y_1^6(k) = -0.002, \\ R_1^7: & \text{ IF } x_1(k) \text{ is } A_1^7 \text{ THEN } y_1^7(k) = -0.02, \\ R_1^8: & \text{ IF } x_1(k) \text{ is } A_1^8 \text{ THEN } y_1^8(k) = -0.003, \\ R_1^9: & \text{ IF } x_1(k) \text{ is } A_1^9 \text{ THEN } y_1^9(k) = -0.004. \end{aligned}$$

for the variable  $x_2$  by:

$$\begin{aligned} R_2^1: & \text{ IF } x_2(k) \text{ is } A_2^1 \text{ THEN } y_2^1(k) = -7.10 \times 10^{-4}, \\ R_2^2: & \text{ IF } x_2(k) \text{ is } A_2^2 \text{ THEN } y_2^2(k) = -3.59 \times 10^{-4}, \\ R_2^3: & \text{ IF } x_2(k) \text{ is } A_2^3 \text{ THEN } y_2^3(k) = -6.45 \times 10^{-5}, \\ R_2^4: & \text{ IF } x_2(k) \text{ is } A_2^4 \text{ THEN } y_2^4(k) = 2.24 \times 10^{-4}, \\ R_2^5: & \text{ IF } x_2(k) \text{ is } A_2^5 \text{ THEN } y_2^5(k) = 4.72 \times 10^{-4}, \\ R_2^6: & \text{ IF } x_2(k) \text{ is } A_2^6 \text{ THEN } y_2^6(k) = 6.66 \times 10^{-4}, \\ R_2^7: & \text{ IF } x_2(k) \text{ is } A_2^7 \text{ THEN } y_2^7(k) = 8.59 \times 10^{-4}, \\ R_2^8: & \text{ IF } x_2(k) \text{ is } A_2^8 \text{ THEN } y_2^8(k) = -0.001, \\ R_2^9: & \text{ IF } x_2(k) \text{ is } A_2^9 \text{ THEN } y_2^9(k) = 6.72 \times 10^{-4}. \end{aligned}$$

for the variable  $x_3$  by:

$$\begin{aligned} R_3^1: & \text{ IF } x_3(k) \text{ is } A_3^1 \text{ THEN } y_3^1(k) = 4.85 \times 10^{-4}, \\ R_3^2: & \text{ IF } x_3(k) \text{ is } A_3^2 \text{ THEN } y_3^2(k) = 4.35 \times 10^{-5}, \\ R_3^3: & \text{ IF } x_3(k) \text{ is } A_3^3 \text{ THEN } y_3^3(k) = -2.29 \times 10^{-4}, \\ R_3^4: & \text{ IF } x_3(k) \text{ is } A_3^4 \text{ THEN } y_3^4(k) = -4.50 \times 10^{-4}, \\ R_3^5: & \text{ IF } x_3(k) \text{ is } A_3^5 \text{ THEN } y_3^5(k) = -6.24 \times 10^{-4}, \\ R_3^6: & \text{ IF } x_3(k) \text{ is } A_3^6 \text{ THEN } y_3^6(k) = -7.36 \times 10^{-4}, \\ R_3^7: & \text{ IF } x_3(k) \text{ is } A_3^7 \text{ THEN } y_3^7(k) = -7.68 \times 10^{-4}, \\ R_3^8: & \text{ IF } x_3(k) \text{ is } A_3^8 \text{ THEN } y_3^8(k) = -7.98 \times 10^{-4}, \\ R_3^9: & \text{ IF } x_3(k) \text{ is } A_3^9 \text{ THEN } y_3^9(k) = -8.46 \times 10^{-4}. \end{aligned}$$

for the variable  $x_4$  by:

$$\begin{aligned}
R_4^1: & \text{ IF } x_4(k) \text{ is } A_4^1 \text{ THEN } y_4^1(k) = -0.001, \\
R_4^2: & \text{ IF } x_4(k) \text{ is } A_4^2 \text{ THEN } y_4^2(k) = -1.31 \times 10^{-4}, \\
R_4^3: & \text{ IF } x_4(k) \text{ is } A_4^3 \text{ THEN } y_4^3(k) = 3.99 \times 10^{-4}, \\
R_4^4: & \text{ IF } x_4(k) \text{ is } A_4^4 \text{ THEN } y_4^4(k) = 8.42 \times 10^{-4}, \\
R_4^5: & \text{ IF } x_4(k) \text{ is } A_4^5 \text{ THEN } y_4^5(k) = 0.001, \\
R_4^6: & \text{ IF } x_4(k) \text{ is } A_4^6 \text{ THEN } y_4^6(k) = 0.001, \\
R_4^7: & \text{ IF } x_4(k) \text{ is } A_4^7 \text{ THEN } y_4^7(k) = 0.002, \\
R_4^8: & \text{ IF } x_4(k) \text{ is } A_4^8 \text{ THEN } y_4^8(k) = 0.002, \\
R_4^9: & \text{ IF } x_4(k) \text{ is } A_4^9 \text{ THEN } y_4^9(k) = 0.002.
\end{aligned}$$

for the variable  $x_5$  by:

$$\begin{aligned}
R_5^1: & \text{ IF } x_5(k) \text{ is } A_5^1 \text{ THEN } y_5^1(k) = 0.001, \\
R_5^2: & \text{ IF } x_5(k) \text{ is } A_5^2 \text{ THEN } y_5^2(k) = -0.001, \\
R_5^3: & \text{ IF } x_5(k) \text{ is } A_5^3 \text{ THEN } y_5^3(k) = -0.003, \\
R_5^4: & \text{ IF } x_5(k) \text{ is } A_5^4 \text{ THEN } y_5^4(k) = 1.02 \times 10^{-4}, \\
R_5^5: & \text{ IF } x_5(k) \text{ is } A_5^5 \text{ THEN } y_5^5(k) = 0.009.
\end{aligned}$$

for the variable  $x_6$  by:

$$\begin{aligned}
R_6^1: & \text{ IF } x_6(k) \text{ is } A_6^1 \text{ THEN } y_6^1(k) = 0.15, \\
R_6^2: & \text{ IF } x_6(k) \text{ is } A_6^2 \text{ THEN } y_6^2(k) = 0.06, \\
R_6^3: & \text{ IF } x_6(k) \text{ is } A_6^3 \text{ THEN } y_6^3(k) = -0.03, \\
R_6^4: & \text{ IF } x_6(k) \text{ is } A_6^4 \text{ THEN } y_6^4(k) = -0.40.
\end{aligned}$$

for the variable  $x_7$  by:

$$\begin{aligned}
R_7^1: & \text{ IF } x_7(k) \text{ is } A_7^1 \text{ THEN } y_7^1(k) = -6.58 \times 10^{-5}, \\
R_7^2: & \text{ IF } x_7(k) \text{ is } A_7^2 \text{ THEN } y_7^2(k) = -2.02 \times 10^{-5}, \\
R_7^3: & \text{ IF } x_7(k) \text{ is } A_7^3 \text{ THEN } y_7^3(k) = -2.41 \times 10^{-5}, \\
R_7^4: & \text{ IF } x_7(k) \text{ is } A_7^4 \text{ THEN } y_7^4(k) = 1.07 \times 10^{-5}, \\
R_7^5: & \text{ IF } x_7(k) \text{ is } A_7^5 \text{ THEN } y_7^5(k) = -7.77 \times 10^{-7}, \\
R_7^6: & \text{ IF } x_7(k) \text{ is } A_7^6 \text{ THEN } y_7^6(k) = 6.01 \times 10^{-6}, \\
R_7^7: & \text{ IF } x_7(k) \text{ is } A_7^7 \text{ THEN } y_7^7(k) = 1.93 \times 10^{-5}, \\
R_7^8: & \text{ IF } x_7(k) \text{ is } A_7^8 \text{ THEN } y_7^8(k) = 1.73 \times 10^{-5}, \\
R_7^9: & \text{ IF } x_7(k) \text{ is } A_7^9 \text{ THEN } y_7^9(k) = 4.33 \times 10^{-5}, \\
R_7^{10}: & \text{ IF } x_7(k) \text{ is } A_7^{10} \text{ THEN } y_7^{10}(k) = -1.82 \times 10^{-5}.
\end{aligned}$$

for the variable  $x_8$  by:

$$\begin{aligned}
R_8^1: & \text{ IF } x_8(k) \text{ is } A_8^1 \text{ THEN } y_8^1(k) = -0.002, \\
R_8^2: & \text{ IF } x_8(k) \text{ is } A_8^2 \text{ THEN } y_8^2(k) = -0.001, \\
R_8^3: & \text{ IF } x_8(k) \text{ is } A_8^3 \text{ THEN } y_8^3(k) = -4.77 \times 10^{-4}, \\
R_8^4: & \text{ IF } x_8(k) \text{ is } A_8^4 \text{ THEN } y_8^4(k) = 0.001, \\
R_8^5: & \text{ IF } x_8(k) \text{ is } A_8^5 \text{ THEN } y_8^5(k) = 0.001, \\
R_8^6: & \text{ IF } x_8(k) \text{ is } A_8^6 \text{ THEN } y_8^6(k) = 0.002, \\
R_8^7: & \text{ IF } x_8(k) \text{ is } A_8^7 \text{ THEN } y_8^7(k) = 0.004.
\end{aligned}$$

for the variable  $x_9$  by:

$$\begin{aligned}
R_9^1: & \text{ IF } x_9(k) \text{ is } A_9^1 \text{ THEN } y_9^1(k) = 0.001, \\
R_9^2: & \text{ IF } x_9(k) \text{ is } A_9^2 \text{ THEN } y_9^2(k) = 4.58 \times 10^{-4}, \\
R_9^3: & \text{ IF } x_9(k) \text{ is } A_9^3 \text{ THEN } y_9^3(k) = 1.36 \times 10^{-4}, \\
R_9^4: & \text{ IF } x_9(k) \text{ is } A_9^4 \text{ THEN } y_9^4(k) = -3.86 \times 10^{-4}, \\
R_9^5: & \text{ IF } x_9(k) \text{ is } A_9^5 \text{ THEN } y_9^5(k) = -6.72 \times 10^{-4}, \\
R_9^6: & \text{ IF } x_9(k) \text{ is } A_9^6 \text{ THEN } y_9^6(k) = -9.54 \times 10^{-4}, \\
R_9^7: & \text{ IF } x_9(k) \text{ is } A_9^7 \text{ THEN } y_9^7(k) = -0.001, \\
R_9^8: & \text{ IF } x_9(k) \text{ is } A_9^8 \text{ THEN } y_9^8(k) = -0.001.
\end{aligned}$$

for the variable  $x_{10}$  by:

$$\begin{aligned}
R_{10}^1: & \text{ IF } x_{10}(k) \text{ is } A_{10}^1 \text{ THEN } y_{10}^1(k) = -4.82 \times 10^{-4}, \\
R_{10}^2: & \text{ IF } x_{10}(k) \text{ is } A_{10}^2 \text{ THEN } y_{10}^2(k) = -1.82 \times 10^{-4}, \\
R_{10}^3: & \text{ IF } x_{10}(k) \text{ is } A_{10}^3 \text{ THEN } y_{10}^3(k) = 8.89 \times 10^{-5}, \\
R_{10}^4: & \text{ IF } x_{10}(k) \text{ is } A_{10}^4 \text{ THEN } y_{10}^4(k) = 3.25 \times 10^{-4}, \\
R_{10}^5: & \text{ IF } x_{10}(k) \text{ is } A_{10}^5 \text{ THEN } y_{10}^5(k) = 4.98 \times 10^{-4}, \\
R_{10}^6: & \text{ IF } x_{10}(k) \text{ is } A_{10}^6 \text{ THEN } y_{10}^6(k) = 6.51 \times 10^{-4}, \\
R_{10}^7: & \text{ IF } x_{10}(k) \text{ is } A_{10}^7 \text{ THEN } y_{10}^7(k) = 7.57 \times 10^{-4}, \\
R_{10}^8: & \text{ IF } x_{10}(k) \text{ is } A_{10}^8 \text{ THEN } y_{10}^8(k) = 8.15 \times 10^{-4}, \\
R_{10}^9: & \text{ IF } x_{10}(k) \text{ is } A_{10}^9 \text{ THEN } y_{10}^9(k) = 8.39 \times 10^{-4}, \\
R_{10}^{10}: & \text{ IF } x_{10}(k) \text{ is } A_{10}^{10} \text{ THEN } y_{10}^{10}(k) = 9.47 \times 10^{-4}.
\end{aligned}$$

for the variable  $x_{11}$  by:

$$\begin{aligned}
R_{11}^1: & \text{ IF } x_{11}(k) \text{ is } A_{11}^1 \text{ THEN } y_{11}^1(k) = -0.17, \\
R_{11}^2: & \text{ IF } x_{11}(k) \text{ is } A_{11}^2 \text{ THEN } y_{11}^2(k) = -0.06, \\
R_{11}^3: & \text{ IF } x_{11}(k) \text{ is } A_{11}^3 \text{ THEN } y_{11}^3(k) = 0.002, \\
R_{11}^4: & \text{ IF } x_{11}(k) \text{ is } A_{11}^4 \text{ THEN } y_{11}^4(k) = 0.06, \\
R_{11}^5: & \text{ IF } x_{11}(k) \text{ is } A_{11}^5 \text{ THEN } y_{11}^5(k) = 0.13, \\
R_{11}^6: & \text{ IF } x_{11}(k) \text{ is } A_{11}^6 \text{ THEN } y_{11}^6(k) = 0.20, \\
R_{11}^7: & \text{ IF } x_{11}(k) \text{ is } A_{11}^7 \text{ THEN } y_{11}^7(k) = 0.27, \\
R_{11}^8: & \text{ IF } x_{11}(k) \text{ is } A_{11}^8 \text{ THEN } y_{11}^8(k) = 0.34, \\
R_{11}^9: & \text{ IF } x_{11}(k) \text{ is } A_{11}^9 \text{ THEN } y_{11}^9(k) = 0.41, \\
R_{11}^{10}: & \text{ IF } x_{11}(k) \text{ is } A_{11}^{10} \text{ THEN } y_{11}^{10}(k) = 0.48.
\end{aligned}$$

for the variable  $x_{12}$  by:

$$\begin{aligned}
R_{12}^1: & \text{ IF } x_{12}(k) \text{ is } A_{12}^1 \text{ THEN } y_{12}^1(k) = -0.001, \\
R_{12}^2: & \text{ IF } x_{12}(k) \text{ is } A_{12}^2 \text{ THEN } y_{12}^2(k) = 0.002, \\
R_{12}^3: & \text{ IF } x_{12}(k) \text{ is } A_{12}^3 \text{ THEN } y_{12}^3(k) = 8.95 \times 10^{-4}, \\
R_{12}^4: & \text{ IF } x_{12}(k) \text{ is } A_{12}^4 \text{ THEN } y_{12}^4(k) = 0.004, \\
R_{12}^5: & \text{ IF } x_{12}(k) \text{ is } A_{12}^5 \text{ THEN } y_{12}^5(k) = 0.001, \\
R_{12}^6: & \text{ IF } x_{12}(k) \text{ is } A_{12}^6 \text{ THEN } y_{12}^6(k) = -9.85 \times 10^{-4}, \\
R_{12}^7: & \text{ IF } x_{12}(k) \text{ is } A_{12}^7 \text{ THEN } y_{12}^7(k) = -0.009.
\end{aligned}$$

and for the variable  $x_{13}$  by:

$$\begin{aligned}
R_{13}^1: & \text{ IF } x_{13}(k) \text{ is } A_{13}^1 \text{ THEN } y_{13}^1(k) = -0.15, \\
R_{13}^2: & \text{ IF } x_{13}(k) \text{ is } A_{13}^2 \text{ THEN } y_{13}^2(k) = -0.05, \\
R_{13}^3: & \text{ IF } x_{13}(k) \text{ is } A_{13}^3 \text{ THEN } y_{13}^3(k) = 0.06, \\
R_{13}^4: & \text{ IF } x_{13}(k) \text{ is } A_{13}^4 \text{ THEN } y_{13}^4(k) = 0.24, \\
R_{13}^5: & \text{ IF } x_{13}(k) \text{ is } A_{13}^5 \text{ THEN } y_{13}^5(k) = 0.30, \\
R_{13}^6: & \text{ IF } x_{13}(k) \text{ is } A_{13}^6 \text{ THEN } y_{13}^6(k) = 0.40.
\end{aligned}$$



# Appendix D

## Licenses

**ELSEVIER LICENSE  
TERMS AND CONDITIONS**

Feb 23, 2024

---

---

This Agreement between Universidade de Coimbra -- Miguel Proença ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5734801181857
License date	Feb 23, 2024
Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Manufacturing Systems
Licensed Content Title	Enabling technologies and tools for digital twin
Licensed Content Author	Qinglin Qi, Fei Tao, Tianliang Hu, Nabil Anwer, Ang Liu, Yongli Wei, Lihui Wang, A.Y.C. Nee
Licensed Content Date	Jan 1, 2021
Licensed Content Volume	58
Licensed Content Issue	n/a
Licensed Content Pages	19
Start Page	3
End Page	21
Type of Use	reuse in a thesis/dissertation

Portion	figures/tables/illustrations
Number of figures/tables/illustrations	3
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	Yes, including English rights
Number of languages	1
Title of new work	Interpretable Digital Twin for Industrial Process
Institution name	Universidade de Coimbra
Expected presentation date	Feb 2024
Portions	Figure 2.2 Figure 2.3 Figure 2.4
Specific Languages	English
Requestor Location	Universidade de Coimbra Paço das Escolas
	Coimbra, 3004-531 Portugal Attn: Universidade de Coimbra
Publisher Tax ID	GB 494 6272 12
Total	0.00 EUR
Terms and Conditions	

## INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and



conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your RightsLink account and that are available at any time at <https://myaccount.copyright.com>).

## GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given. The material may not be reproduced or used in any other way, including use in combination with an artificial intelligence tool (including to train an algorithm, test, process, analyse, generate output and/or develop any form of artificial intelligence tool), or to create any derivative work and/or service (including resulting from the use of artificial intelligence tools).

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier's permissions helpdesk [here](#)). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. **Indemnity:** You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. **No Transfer of License:** This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. **No Amendment Except in Writing:** This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. **Objection to Contrary Terms:** Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. **Revocation:** Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

### LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation:** This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website:** The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> or the Elsevier homepage for books at <http://www.elsevier.com>; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at <http://www.elsevier.com>. All content posted to the web site must maintain the copyright information line on the bottom of each image.

**Posting licensed content on Electronic reserve:** In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

### **Preprints:**

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

**Accepted Author Manuscripts:** An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

**Published journal article (JPA):** A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

**Subscription Articles:** If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

**Gold Open Access Articles:** May be shared according to the author-selected end-user license and should contain a [CrossMark logo](#), the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's [posting policy](#) for further information.

**18. For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

**19. Thesis/Dissertation:** If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

### **Elsevier Open Access Terms and Conditions**

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our [open access license policy](#) for more information.

#### **Terms & Conditions applicable to all Open Access articles published with Elsevier:**

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

#### **Additional Terms & Conditions applicable to each Creative Commons user license:**

**CC BY:** The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by/4.0>.

**CC BY NC SA:** The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-sa/4.0>.

**CC BY NC ND:** The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-nd/4.0>. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

## 20. Other Conditions:

v1.10

Questions? [customercare@copyright.com](mailto:customercare@copyright.com).

---

---

## SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Feb 23, 2024

---

---

This Agreement between Universidade de Coimbra -- Miguel Proença ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5734941287210
License date	Feb 23, 2024
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Neural Computing & Applications
Licensed Content Title	Neo-fuzzy neuron learning using backfitting algorithm
Licensed Content Author	Jérôme Mendes et al
Licensed Content Date	Dec 30, 2017
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Will you be translating?	no
Circulation/distribution	1 - 29

Author of this Springer Nature content	no
Title of new work	Interpretable Digital Twin for Industrial Process
Institution name	Universidade de Coimbra
Expected presentation date	Feb 2024
Portions	Figure 4.2 Figure 4.3
Requestor Location	Universidade de Coimbra Paço das Escolas
	Coimbra, Região 3004-531 Portugal Attn: Universidade de Coimbra
Total	0.00 EUR

## Terms and Conditions

### Springer Nature Customer Service Centre GmbH Terms and Conditions

The following terms and conditions ("Terms and Conditions") together with the terms specified in your [RightsLink] constitute the License ("License") between you as Licensee and Springer Nature Customer Service Centre GmbH as Licensor. By clicking 'accept' and completing the transaction for your use of the material ("Licensed Material"), you confirm your acceptance of and obligation to be bound by these Terms and Conditions.

#### 1. Grant and Scope of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, non-sublicensable, revocable, world-wide License to reproduce, distribute, communicate to the public, make available, broadcast, electronically transmit or create derivative works using the Licensed Material for the purpose(s) specified in your RightsLink Licence Details only. Licenses are granted for the specific use requested in the order and for no other use, subject to these Terms and Conditions. You acknowledge and agree that the rights granted to you under this License do not include the right to modify, edit, translate, include in collective works, or create derivative works of the Licensed Material in whole or in part unless expressly stated in your RightsLink Licence Details. You may use the Licensed Material only as permitted under this Agreement and will not reproduce, distribute, display, perform, or otherwise use or exploit any Licensed Material in any way, in whole or in part, except as expressly permitted by this License.

1. 2. You may only use the Licensed Content in the manner and to the extent permitted by these Terms and Conditions, by your RightsLink Licence Details and by any applicable laws.

1. 3. A separate license may be required for any additional use of the Licensed Material, e.g. where a license has been purchased for print use only, separate permission must be obtained for electronic re-use. Similarly, a License is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the License.

1. 4. Any content within the Licensed Material that is owned by third parties is expressly excluded from the License.

1. 5. Rights for additional reuses such as custom editions, computer/mobile applications, film or TV reuses and/or any other derivative rights requests require additional permission and may be subject to an additional fee. Please apply to [journalpermissions@springernature.com](mailto:journalpermissions@springernature.com) or [bookpermissions@springernature.com](mailto:bookpermissions@springernature.com) for these rights.

## 2. Reservation of Rights

Licensor reserves all rights not expressly granted to you under this License. You acknowledge and agree that nothing in this License limits or restricts Licensor's rights in or use of the Licensed Material in any way. Neither this License, nor any act, omission, or statement by Licensor or you, conveys any ownership right to you in any Licensed Material, or to any element or portion thereof. As between Licensor and you, Licensor owns and retains all right, title, and interest in and to the Licensed Material subject to the license granted in Section 1.1. Your permission to use the Licensed Material is expressly conditioned on you not impairing Licensor's or the applicable copyright owner's rights in the Licensed Material in any way.

## 3. Restrictions on use

3. 1. Minor editing privileges are allowed for adaptations for stylistic purposes or formatting purposes provided such alterations do not alter the original meaning or intention of the Licensed Material and the new figure(s) are still accurate and representative of the Licensed Material. Any other changes including but not limited to, cropping, adapting, and/or omitting material that affect the meaning, intention or moral rights of the author(s) are strictly prohibited.

3. 2. You must not use any Licensed Material as part of any design or trademark.

3. 3. Licensed Material may be used in Open Access Publications (OAP), but any such reuse must include a clear acknowledgment of this permission visible at the same time as the figures/tables/illustration or abstract and which must indicate that the Licensed Material is not part of the governing OA license but has been reproduced with permission. This may be indicated according to any standard referencing system but must include at a minimum 'Book/Journal title, Author, Journal Name (if applicable), Volume (if applicable), Publisher, Year, reproduced with permission from SNCSC'.

## 4. STM Permission Guidelines

4. 1. An alternative scope of license may apply to signatories of the STM Permissions Guidelines ("STM PG") as amended from time to time and made available at <https://www.stm-assoc.org/intellectual-property/permissions/permissions-guidelines/>.



4. 2. For content reuse requests that qualify for permission under the STM PG, and which may be updated from time to time, the STM PG supersede the terms and conditions contained in this License.

4. 3. If a License has been granted under the STM PG, but the STM PG no longer apply at the time of publication, further permission must be sought from the Rightsholder. Contact [journalpermissions@springernature.com](mailto:journalpermissions@springernature.com) or [bookpermissions@springernature.com](mailto:bookpermissions@springernature.com) for these rights.

## 5. Duration of License

5. 1. Unless otherwise indicated on your License, a License is valid from the date of purchase ("License Date") until the end of the relevant period in the below table:

Reuse in a medical communications project	Reuse up to distribution or time period indicated in License
Reuse in a dissertation/thesis	Lifetime of thesis
Reuse in a journal/magazine	Lifetime of journal/magazine
Reuse in a book/textbook	Lifetime of edition
Reuse on a website	1 year unless otherwise specified in the License
Reuse in a presentation/slide kit/poster	Lifetime of presentation/slide kit/poster. Note: publication whether electronic or in print of presentation/slide kit/poster may require further permission.
Reuse in conference proceedings	Lifetime of conference proceedings
Reuse in an annual report	Lifetime of annual report
Reuse in training/CME materials	Reuse up to distribution or time period indicated in License
Reuse in newsmedia	Lifetime of newsmedia
Reuse in coursepack/classroom materials	Reuse up to distribution and/or time period indicated in license

## 6. Acknowledgement

6. 1. The Licensor's permission must be acknowledged next to the Licensed Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract and must be hyperlinked to the journal/book's homepage.

6. 2. Acknowledgement may be provided according to any standard referencing system and at a minimum should include "Author, Article/Book Title, Journal name/Book imprint, volume, page number, year, Springer Nature".

## 7. Reuse in a dissertation or thesis

7. 1. Where 'reuse in a dissertation/thesis' has been selected, the following terms apply: Print rights of the Version of Record are provided for; electronic rights for use only on institutional repository as defined by the Sherpa guideline ([www.sherpa.ac.uk/romeo/](http://www.sherpa.ac.uk/romeo/)) and only up to what is required by the awarding institution.

7. 2. For theses published under an ISBN or ISSN, separate permission is required. Please contact [journalpermissions@springernature.com](mailto:journalpermissions@springernature.com) or [bookpermissions@springernature.com](mailto:bookpermissions@springernature.com) for these rights.

7. 3. Authors must properly cite the published manuscript in their thesis according to current citation standards and include the following acknowledgement: *'Reproduced with permission from Springer Nature'*.

## 8. License Fee

You must pay the fee set forth in the License Agreement (the "License Fees"). All amounts payable by you under this License are exclusive of any sales, use, withholding, value added or similar taxes, government fees or levies or other assessments. Collection and/or remittance of such taxes to the relevant tax authority shall be the responsibility of the party who has the legal obligation to do so.

## 9. Warranty

9. 1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. **You are solely responsible for ensuring that the material you wish to license is original to the Licensor and does not carry the copyright of another entity or third party (as credited in the published version).** If the credit line on any part of the Licensed Material indicates that it was reprinted or adapted with permission from another source, then you should seek additional permission from that source to reuse the material.

9. 2. EXCEPT FOR THE EXPRESS WARRANTY STATED HEREIN AND TO THE EXTENT PERMITTED BY APPLICABLE LAW, LICENSOR PROVIDES THE LICENSED MATERIAL "AS IS" AND MAKES NO OTHER REPRESENTATION OR WARRANTY. LICENSOR EXPRESSLY DISCLAIMS ANY LIABILITY FOR ANY CLAIM ARISING FROM OR OUT OF THE CONTENT, INCLUDING BUT NOT LIMITED TO ANY ERRORS, INACCURACIES, OMISSIONS, OR DEFECTS CONTAINED THEREIN, AND ANY IMPLIED OR EXPRESS WARRANTY AS TO MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, PUNITIVE, OR EXEMPLARY DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE LICENSED MATERIAL REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION APPLIES NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

## 10. Termination and Cancellation

10. 1. The License and all rights granted hereunder will continue until the end of the applicable period shown in Clause 5.1 above. Thereafter, this license will be terminated and all rights granted hereunder will cease.

10. 2. Licensor reserves the right to terminate the License in the event that payment is not received in full or if you breach the terms of this License.

## 11. General

11. 1. The License and the rights and obligations of the parties hereto shall be construed, interpreted and determined in accordance with the laws of the Federal Republic of Germany without reference to the stipulations of the CISG (United Nations Convention on Contracts for the International Sale of Goods) or to Germany's choice-of-law principle.

11. 2. The parties acknowledge and agree that any controversies and disputes arising out of this License shall be decided exclusively by the courts of or having jurisdiction for Heidelberg, Germany, as far as legally permissible.

11. 3. This License is solely for Licensor's and Licensee's benefit. It is not for the benefit of any other person or entity.

**Questions?** For questions on Copyright Clearance Center accounts or website issues please contact [springernaturesupport@copyright.com](mailto:springernaturesupport@copyright.com) or +1-855-239-3415 (toll free in the US) or +1-978-646-2777. For questions on Springer Nature licensing please visit <https://www.springernature.com/gp/partners/rights-permissions-third-party-distribution>

### Other Conditions:

Version 1.4 - Dec 2022

Questions? [customercare@copyright.com](mailto:customercare@copyright.com).

---

---

## SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Feb 23, 2024

---

---

This Agreement between Universidade de Coimbra -- Miguel Proença ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5734950132857
License date	Feb 23, 2024
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Fault Detection in Wastewater Treatment Plants: Application of Autoencoders Models with Streaming Data
Licensed Content Author	Rodrigo Salles, Jérôme Mendes, Rita P. Ribeiro et al
Licensed Content Date	Jan 1, 2023
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Will you be translating?	no

Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title of new work	Interpretable Digital Twin for Industrial Process
Institution name	Universidade de Coimbra
Expected presentation date	Feb 2024
Portions	Figure 3.2 Figure 3.3
Requestor Location	Universidade de Coimbra Paço das Escolas
Total	Coimbra, Região 3004-531 Portugal Attn: Universidade de Coimbra
	0.00 EUR

## Terms and Conditions

### Springer Nature Customer Service Centre GmbH Terms and Conditions

The following terms and conditions ("Terms and Conditions") together with the terms specified in your [RightsLink] constitute the License ("License") between you as Licensee and Springer Nature Customer Service Centre GmbH as Licensor. By clicking 'accept' and completing the transaction for your use of the material ("Licensed Material"), you confirm your acceptance of and obligation to be bound by these Terms and Conditions.

#### 1. Grant and Scope of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, non-sublicensable, revocable, world-wide License to reproduce, distribute, communicate to the public, make available, broadcast, electronically transmit or create derivative works using the Licensed Material for the purpose(s) specified in your RightsLink Licence Details only. Licenses are granted for the specific use requested in the order and for no other use, subject to these Terms and Conditions. You acknowledge and agree that the rights granted to you under this License do not include the right to modify, edit, translate, include in collective works, or create derivative works of the Licensed Material in whole or in part unless expressly stated in your RightsLink Licence Details. You may use the Licensed Material only as permitted under this Agreement and will not reproduce, distribute, display, perform, or otherwise use or exploit any Licensed Material in any way, in whole or in part, except as expressly

permitted by this License.

1. 2. You may only use the Licensed Content in the manner and to the extent permitted by these Terms and Conditions, by your RightsLink Licence Details and by any applicable laws.

1. 3. A separate license may be required for any additional use of the Licensed Material, e.g. where a license has been purchased for print use only, separate permission must be obtained for electronic re-use. Similarly, a License is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the License.

1. 4. Any content within the Licensed Material that is owned by third parties is expressly excluded from the License.

1. 5. Rights for additional reuses such as custom editions, computer/mobile applications, film or TV reuses and/or any other derivative rights requests require additional permission and may be subject to an additional fee. Please apply to [journalpermissions@springernature.com](mailto:journalpermissions@springernature.com) or [bookpermissions@springernature.com](mailto:bookpermissions@springernature.com) for these rights.

## 2. Reservation of Rights

Licensor reserves all rights not expressly granted to you under this License. You acknowledge and agree that nothing in this License limits or restricts Licensor's rights in or use of the Licensed Material in any way. Neither this License, nor any act, omission, or statement by Licensor or you, conveys any ownership right to you in any Licensed Material, or to any element or portion thereof. As between Licensor and you, Licensor owns and retains all right, title, and interest in and to the Licensed Material subject to the license granted in Section 1.1. Your permission to use the Licensed Material is expressly conditioned on you not impairing Licensor's or the applicable copyright owner's rights in the Licensed Material in any way.

## 3. Restrictions on use

3. 1. Minor editing privileges are allowed for adaptations for stylistic purposes or formatting purposes provided such alterations do not alter the original meaning or intention of the Licensed Material and the new figure(s) are still accurate and representative of the Licensed Material. Any other changes including but not limited to, cropping, adapting, and/or omitting material that affect the meaning, intention or moral rights of the author(s) are strictly prohibited.

3. 2. You must not use any Licensed Material as part of any design or trademark.

3. 3. Licensed Material may be used in Open Access Publications (OAP), but any such reuse must include a clear acknowledgment of this permission visible at the same time as the figures/tables/illustration or abstract and which must indicate that the Licensed Material is not part of the governing OA license but has been reproduced with permission. This may be indicated according to any standard referencing system but must include at a minimum 'Book/Journal title, Author, Journal Name (if applicable), Volume (if applicable), Publisher, Year, reproduced with permission from SNCSC'.

## 4. STM Permission Guidelines

4. 1. An alternative scope of license may apply to signatories of the STM Permissions Guidelines ("STM PG") as amended from time to time and made available at <https://www.stm-assoc.org/intellectual->

[property/permissions/permissions-guidelines/](#).

4. 2. For content reuse requests that qualify for permission under the STM PG, and which may be updated from time to time, the STM PG supersede the terms and conditions contained in this License.

4. 3. If a License has been granted under the STM PG, but the STM PG no longer apply at the time of publication, further permission must be sought from the Rightsholder. Contact [journalpermissions@springernature.com](mailto:journalpermissions@springernature.com) or [bookpermissions@springernature.com](mailto:bookpermissions@springernature.com) for these rights.

## 5. Duration of License

5. 1. Unless otherwise indicated on your License, a License is valid from the date of purchase ("License Date") until the end of the relevant period in the below table:

Reuse in a medical communications project	Reuse up to distribution or time period indicated in License
Reuse in a dissertation/thesis	Lifetime of thesis
Reuse in a journal/magazine	Lifetime of journal/magazine
Reuse in a book/textbook	Lifetime of edition
Reuse on a website	1 year unless otherwise specified in the License
Reuse in a presentation/slide kit/poster	Lifetime of presentation/slide kit/poster. Note: publication whether electronic or in print of presentation/slide kit/poster may require further permission.
Reuse in conference proceedings	Lifetime of conference proceedings
Reuse in an annual report	Lifetime of annual report
Reuse in training/CME materials	Reuse up to distribution or time period indicated in License
Reuse in newsmedia	Lifetime of newsmedia
Reuse in coursepack/classroom materials	Reuse up to distribution and/or time period indicated in license

## 6. Acknowledgement

6. 1. The Licensor's permission must be acknowledged next to the Licensed Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract and must be hyperlinked to the journal/book's homepage.

6. 2. Acknowledgement may be provided according to any standard referencing system and at a minimum should include "Author, Article/Book Title, Journal name/Book imprint, volume, page number, year, Springer Nature".

## 7. Reuse in a dissertation or thesis

7. 1. Where 'reuse in a dissertation/thesis' has been selected, the following terms apply: Print rights of the Version of Record are provided for; electronic rights for use only on institutional repository as defined by the Sherpa guideline ([www.sherpa.ac.uk/romeo/](http://www.sherpa.ac.uk/romeo/)) and only up to what is required by the awarding institution.

7. 2. For theses published under an ISBN or ISSN, separate permission is required. Please contact [journalpermissions@springernature.com](mailto:journalpermissions@springernature.com) or [bookpermissions@springernature.com](mailto:bookpermissions@springernature.com) for these rights.

7. 3. Authors must properly cite the published manuscript in their thesis according to current citation standards and include the following acknowledgement: *'Reproduced with permission from Springer Nature'*.

## 8. License Fee

You must pay the fee set forth in the License Agreement (the "License Fees"). All amounts payable by you under this License are exclusive of any sales, use, withholding, value added or similar taxes, government fees or levies or other assessments. Collection and/or remittance of such taxes to the relevant tax authority shall be the responsibility of the party who has the legal obligation to do so.

## 9. Warranty

9. 1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. **You are solely responsible for ensuring that the material you wish to license is original to the Licensor and does not carry the copyright of another entity or third party (as credited in the published version).** If the credit line on any part of the Licensed Material indicates that it was reprinted or adapted with permission from another source, then you should seek additional permission from that source to reuse the material.

9. 2. EXCEPT FOR THE EXPRESS WARRANTY STATED HEREIN AND TO THE EXTENT PERMITTED BY APPLICABLE LAW, LICENSOR PROVIDES THE LICENSED MATERIAL "AS IS" AND MAKES NO OTHER REPRESENTATION OR WARRANTY. LICENSOR EXPRESSLY DISCLAIMS ANY LIABILITY FOR ANY CLAIM ARISING FROM OR OUT OF THE CONTENT, INCLUDING BUT NOT LIMITED TO ANY ERRORS, INACCURACIES, OMISSIONS, OR DEFECTS CONTAINED THEREIN, AND ANY IMPLIED OR EXPRESS WARRANTY AS TO MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, PUNITIVE, OR EXEMPLARY DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE LICENSED MATERIAL REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION APPLIES NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

## 10. Termination and Cancellation

10. 1. The License and all rights granted hereunder will continue until the end of the applicable period shown in Clause 5.1 above. Thereafter, this license will be terminated and all rights granted hereunder will cease.

10. 2. Licensor reserves the right to terminate the License in the event that payment is not received in full or if you breach the terms of this License.



## 11. General

11. 1. The License and the rights and obligations of the parties hereto shall be construed, interpreted and determined in accordance with the laws of the Federal Republic of Germany without reference to the stipulations of the CISG (United Nations Convention on Contracts for the International Sale of Goods) or to Germany's choice-of-law principle.

11. 2. The parties acknowledge and agree that any controversies and disputes arising out of this License shall be decided exclusively by the courts of or having jurisdiction for Heidelberg, Germany, as far as legally permissible.

11. 3. This License is solely for Licensor's and Licensee's benefit. It is not for the benefit of any other person or entity.

**Questions?** For questions on Copyright Clearance Center accounts or website issues please contact [springernaturesupport@copyright.com](mailto:springernaturesupport@copyright.com) or +1-855-239-3415 (toll free in the US) or +1-978-646-2777. For questions on Springer Nature licensing please visit <https://www.springernature.com/gp/partners/rights-permissions-third-party-distribution>

### Other Conditions:

Version 1.4 - Dec 2022

Questions? [customercare@copyright.com](mailto:customercare@copyright.com).

---

---