# Fault detection in wastewater treatment plants - Application and comparison of machine learning techniques with streaming data

Rodrigo Salles
Master in Data Science
DEPARTMENT OF COMPUTER SCIENCE
2022

**Orientador**
Prof. Dr. João Gama, Faculty of Sciences of the University of Porto

**Coorientadora**
Prof. Dra. Rita Ribeiro, Faculty of Sciences of the University of Porto

**Coorientador**
Prof. Dr. Jérôme Mendes , ISR - University of Coimbra

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

# Declaração de Honra

Eu, Rodrigo de Oliveira Salles, inscrito no Mestrado em Ciência de Dados (Data Science) da Faculdade de Ciências da Universidade do Porto declaro, nos termos do disposto na alínea a) do artigo 14.º do Código Ético de Conduta Académica da U.Porto, que o conteúdo da presente dissertação reflete as perspetivas, o trabalho de investigação e as minhas interpretações no momento da sua entrega.

Ao entregar esta dissertação, declaro, ainda, que a mesma é resultado do meu próprio trabalho de investigação e contém contributos que não foram utilizados previamente noutros trabalhos apresentados a esta ou outra instituição.

Mais declaro que todas as referências a outros autores respeitam escrupulosamente as regras da atribuição, encontrando-se devidamente citadas no corpo do texto e identificadas na secção de referências bibliográficas. Não são divulgados na presente dissertação quaisquer conteúdos cuja reprodução esteja vedada por direitos de autor.

Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

Rodrigo de Oliveira Salles

Porto, 04 de Outubro de 2022.

UNIVERSITY OF PORTO

MASTERS THESIS

# Fault detection in wastewater treatment plants - Application and comparison of machine learning techniques with streaming data

*Author:*
Rodrigo SALLES

*Supervisor:*
João GAMA

*Co-supervisor:*
Rita RIBEIRO
Jérôme MENDES

*A thesis submitted in fulfilment of the requirements for the degree of MSc. Data Science*

*at the*

Faculty of Sciences of the University of Porto
Department of Computer Science

November 26, 2022

*" Se eu vi mais longe, foi por estar sobre ombros de gigantes "*

Isaac Newton.

Trecho de uma carta de Newton para Robert Hooke, 5 de Fevereiro de 1676.

# *Acknowledgements*

A presente dissertação de mestrado conclui um período de muito esforço e dedicação. Foi um tempo de muitas incertezas, preocupações, mas também de muitas vitórias. Com certeza não conseguiria concluir esse curso sem a ajuda de pessoas muito especiais.

Gostaria de agradecer aos Professores João Gama e Jérôme Mendes, e à Professora Rita Ribeiro. Muito obrigado pela orientação e toda ajuda na elaboração dessa dissertação.

Agradeço aos meus pais, João Domingos Salles e minha mãe Vera Lúcia, por todo apoio e incentivo. Mesmo distantes, vocês sempre estiveram muito próximos. "O Atlântico é pequeno pra nos separar, porque o sangue é mais forte que a água do mar."

Agradeço aos meus irmãos Fábio e Filipe. O apoio de vocês foi fundamental.

Gostaria de agradecer aos amigos que fiz durante o curso. São pessoas especiais que me ajudaram muito.

Agradeço de forma especial à minha esposa Jacqueline Salles e à minha filha Sophia Salles. Obrigado por toda ajuda e compreensão. Seria impossível concluir esse curso sem a ajuda de vocês. Amo-vos.

Por fim, agradeço a Deus por me sustentar nos momentos de dificuldade e desânimo. Agradeço por colocar pessoas boas em meu caminho. Agradeço por estar sempre presente em minha vida.

UNIVERSITY OF PORTO

# *Abstract*

Faculty of Sciences of the University of Porto
Department of Computer Science

MSc. Data Science

**Fault detection in wastewater treatment plants - Application and comparison of machine learning techniques with streaming data**

by Rodrigo SALLES

Water is an essential element for life, and plays a fundamental role for human beings. Industrial activities, services and agriculture depend on the availability of water resources. And these resources are increasingly scarce. It is estimated that of all the water on Earth, only 1% is available for human consumption. It is a small fraction that must serve an increasing number of people. Approximately 4 billion people face water supply problems at least once a month, and half a billion people face water shortages throughout the year.

In view of the importance of water, the need for its conscious use is clear. Water needs to be used efficiently, and wastewater treated, so that it can be used again, without posing risks to humans and the environment. The objective of the work developed in this dissertation is to implement a fault detection method for sensors used in Wastewater Treatment Plant (WWTP), in order to make them more reliable and efficient. In this dissertation, the use of machine learning techniques, specifically based on autoencoders (AE), were proposed to identify faults in sensors, in WWTPs, using streaming data. The failure detection system was developed, with the help of the Benchmark Simulation Model No 2 (BSM2), to identify faults, in an online way, in the dissolved oxygen sensor, which monitors the oxygen levels in the biological reactor tanks.

Three failure scenarios were created, with the help of BSM2, with variations in the order of appearance, duration and intensity of failures. The purpose of the EA models used is to detect the occurrence of these faults as soon as possible. The AE models, which were used for fault detection, Convolutional and Long Short-Term Memory (LSTM), were efficient, according to the considered metrics. Most of the samples representing the failures

were identified, with few delays. The AEs-based methodologies proved to be a suitable alternative for fault detection in WWTPs. The exploitation of hyperparameters and the combination of AE with other machine learning methods can make the fault detection system even more efficient, and the WWTPs more reliable in performing their important function.

UNIVERSITY OF PORTO

# *Resumo*

Faculty of Sciences of the University of Porto
Department of Computer Science

Mestrado em Ciência de Dados

**Deteção de falhas em estações de tratamento de águas residuais - Aplicação e comparação de técnicas de aprendizado de máquina com dados em streaming**

por Rodrigo SALLES

A água é um elemento essencial para a vida, e desempenha um papel fundamental para os seres humanos. As atividades industriais, serviços e agricultura dependem da disponibilidade dos recursos hídricos. E estes recursos são cada vez mais escassos. Estima-se que de toda a água existente na Terra, apenas 1% está disponível para o consumo humano. Trata-se de uma pequena fração que deve servir a um número crescente de pessoas. Aproximadamente 4 bilhões de pessoas enfrentam problemas de abastecimento de água pelo menos uma vez por mês, e meio bilhão de pessoas enfrentam escassez de água ao longo do ano.

Tendo em conta a importância da água, a necessidade da sua utilização consciente é clara. A água precisa de ser utilizada eficientemente, e as águas residuais tratadas, para que possa ser utilizada novamente, sem representar riscos para os seres humanos e para o meio ambiente. O objetivo do trabalho desenvolvido nesta dissertação é implementar um método de deteção de falhas nos sensores utilizados nas ETARs, de modo a torná-las mais confiáveis e eficientes. Nesta dissertação, foi proposto o uso de técnicas de aprendizagem de máquinas, especificamente baseadas em autoencoders (AE), para identificar falhas em sensores, em ETARs, utilizando dados em streaming. O sistema de deteção de falhas foi desenvolvido, com a ajuda do simulador Benchmark Simulation Model No 2 (BSM2), para identificar falhas, de forma online, no sensor de oxigénio dissolvido, que monitoriza os níveis de oxigénio nos tanques do reator biológico.

Foram criados três cenários de falhas, com a ajuda do BSM2, com variações na ordem de aparecimento, duração e intensidade das falhas. O objetivo do sistema de deteção implementado é detetar a ocorrência destas falhas o mais rapidamente possível. Os modelos

de AE, que foram utilizados para a deteção de falhas, Convolutional e Long Short-Term Memory (LSTM), foram eficientes, de acordo com as métricas consideradas. A maioria das amostras que representavam as falhas foram identificadas, com poucos atrasos. As metodologias baseadas em AEs provaram ser uma alternativa adequada para a deteção de falhas em ETARs. A exploração de hiperparâmetros e a combinação dos AEs com outros métodos de aprendizagem de máquinas podem tornar o sistema de deteção de falhas ainda mais eficiente, e as ETARs mais confiáveis no desempenho da sua importante função.

# Contents

# List of Figures

# List of Tables

# Glossary

| | |
|---|---|
| **TP** | True Positive |
| **FP** | False Positive |
| **TN** | True Negative |
| **FN** | False Negative |
| **ROC** | Receiver Operating Characteristics |
| **WWTP** | Wastewater Treatment Plant |
| **AE** | Aeration Energy |
| **PCA** | Principal Component Analysis |
| **RMSE** | Root Mean Squared Error |
| **GPR** | Gaussian Process Regression |
| **DO** | Dissolved Oxygen |
| **LSTM** | Long Short-Term Memory |
| **BSM2** | Benchmark Simulation Model No 2 |
| **ML** | Machine Learning |
| **ANN** | Artificial Neural Networks |

# Chapter 1

# Introduction

This first chapter aims to present the motivation of this dissertation, namely, the problems related to water and wastewater, the importance and complexity of the Wastewater Treatment Plant (WWTP), and the relevance of a reliable failure detection system. Section 1.1 brings the motivation and context of the work, and in Section 1.1.1 a description of the evolution and challenges faced by WWTPs when performing their tasks is made. Section 1.2 brings a summary of the objectives of the proposed work, and finally in Section 1.3 there is a description of the dissertation organization.

## 1.1 Motivation and Context

Water is a very important resource for human life and represents approximately 60% of body weight in adult men and 50%-55% in adult women [1], which regulates our temperature and is essential to all our organic functions. Furthermore, water plays a fundamental role in many human activities, such as agriculture, industry and services. And this resource so important to our survival is limited. Water covers about 70% of planet Earth [2], but it is estimated that 97% of the water available on the planet is salty, unfit for consumption, found in oceans and seas, and 2% form inaccessible glaciers. Therefore, 1% is used for human and animal consumption [3].

The supply of quality water for human activities is a challenge that increases with population growth. According to United Nations estimates, on 15 November 2022, the Earth will reach the 8 billion people mark [4], and is estimated that 4 billion people already face severe water scarcity at least one month a year. Half a billion people face severe water shortages throughout the year. [5].

With population growth, there is greater use of water and greater amount of wastewater produced, with pollutants of various types. The causes of water pollution are many: residential and industrial sewage, mining activities, pesticides and fertilisers used in agriculture, etc. Domestic, agricultural or industrial activities produce effluents containing undesirable pollutants that can be toxic [6].

In view of the information about the importance of water for human beings and its scarcity, the need for actions that encourage its conscientious use is clear. Water resources must be used efficiently, and the wastewater resulting from this use must be treated, so that it becomes available again and does not pose a threat to the environment and thus to human being. In this context, specialized WWTP, emerge as very important structures to face the new water and wastewater challenges.

### 1.1.1 Wastewater Treatment Plants

Wastewater treatment is carried out in Wastewater Treatment Plants. In general terms, the WWTPs have the function of removing the most different types of pollutants present in wastewater. The stations speed up the treatment that occurs in nature, so that the waters reach the minimum quality requirements and can be returned to the environment without posing threats [7].

WWTPs are very important structures for modern life. They treat wastewater from industrial and domestic environments [8], and they are spread out in large numbers around the world. In the United States of America, there are more than 16,000 public administration WWTPs [9]. The European Union has more than 18.000 treatment units, serving a population of over 450 million people in its member states [10].

WWTPs are large complex plant constituted by several interconnected processes [11]. The sewage treatment comprises several steps, in physical, chemical and biological processes. Decantation of organic matter, addition of coagulants, bacterial action, sludge formation, etc. Each step is made up of its own, and often conflicting, processes and goals. External factors make the wastewater treatment even more complex: climatic events (rains, storms, droughts), seasonality, periods of the day and the year, etc [12, 13]. All these factors must be considered to ensure that the water leaving the treatment station is within the required quality parameters. It is a great challenge for a structure with markedly non-linear characteristics [14].

The efficiency of a WWTP is commonly evaluated by the quality of the water that leaves the station, but many other factors must be considered to confirm its efficiency. The WWTPs perform their functions with a high demand for resources, with special emphasis on electricity. 7% of all energy consumed in the world is consumed by WWTPs [15]. In Portugal, approximately 4% of all energy consumed is spent on water distribution and treatment systems, and of this total, 25% is used by WWTPs [16]. Thus, energy consumption is an important efficiency index. The consumption of chemical reagents involved in the process, the emissions of greenhouse gases, are indices normally considered [17, 18]. These and many other issues must be addressed when designing control, monitoring and optimization models for WWTPs.

In the 21st century the need for efficiency and environmental protection has become even clearer, and the demands placed on the operation of WWTPs have increased [19]. The WWTPs need to achieve at least three sustainability goals: environmental protection (low pollutant load), social acceptance (human sanitary protection) and economic development (feasible operational and construction costs) [20, 21].

Over the years, it was found that the conventional forms of control and monitoring of WWTPs were not sufficient to keep up with the new requirements and restrictions. It is not enough to have reactive systems. It is necessary to have systems that anticipate problems and assist in decision making. And due to the non-linear characteristics of WWTPs, designing models that meet their new demands is not trivial. With technological advances, new techniques can help to overcome new challenges. The massive use of sensors and the consequent data generation, enabled the use of more efficient and promising techniques. The large amount of information generated in a short period of time, in real time, about the various stages of the process, enabled the use of new methods, especially methods involving machine learning algorithms. The availability of data allows the training of algorithms, and obtaining ever more efficient models. The advantages of the new methods are clear, but with the new methods come new challenges.

The amount of data generated by the sensors is very large. A special scheme for handling and storing this information is needed, as well as computing resources. Sensors are exposed to an inhospitable environment, and usually present failures, but a failure must be distinguished from an anomaly in the process. And the detection must be in real time. The results of some measurements of the treatment processes can be obtained instantly or take a few hours or days. Monitoring and control techniques need to deal with different

sampling times. These are just a few issues that have gained relevance with technological advances, and need to be analyzed in the search for efficiency.

## 1.2 Objectives and Contributions

The work developed in this dissertation aims to assess the potential of autoencoders (AEs) in the detection of failures in dissolved oxygen (DO) sensors, in WWTPs, in real time, with streaming data. For that, the simulator Benchmark Simulation Model nº2 (BSM2) [22], which reproduces all phases of the treatment performed in a WWTP, was used to test and validate the fault detection methodologies on the DO sensor in the biological reactor. The contribution of this dissertation is to analyzes the strengths and weaknesses of the Convolution-AE and Long short-term memory (LSTM)-AE models, used to detect failures on DO sensors presented in WWTPs. Moreover, the models are studied and evaluated for the detection of five types of failures: bias, drift, precision degradation, spike and stuck, in three scenarios, with changes in the order of appearance, duration and intensity of the faults. These contributions have resulted on the accepted conference paper (see Appendix A):

- Rodrigo Salles, Jérôme Mendes, Rita P. Ribeiro, and João Gama. Fault Detection in Wastewater Treatment Plants: Application of Autoencoders with Streaming Data. The 7th Workshop on Data Science for Social Good, Grenoble, France, September 23, 2022.

## 1.3 Organization of the Dissertation

The remaining of this dissertation is organized as follows:

- Chapter 2 provides a literature review on the main techniques used for fault detection in WWTPs, and a description of the most common faults found in DO sensors.
- Chapter 3 gives a description of the Benchmark Simulation Model No 2 (BSM2) simulator, used for the development of the work.
- A description of the Autoencoders is given in Chapter 4.
- In Chapter 5, a description of the dataset used, obtained using the BSM2, is made. This chapter also describes the process of injecting the failures in the DO sensor.

- The obtained results and discussions are in Chapter 6.
- In Chapter 7 is the conclusion of the work, with its limitations and possibilities for future work.

# Chapter 2

# Fault Detection in WWTPs

In this chapter, a literature review will be performed on the techniques used to detect failures in WWTPs, as well as the description of the main types of failures found in sensors responsible for monitoring the various stages of the wastewater treatment process. The chapter is organized as follows: Section 2.1 explains the methodology adopted to prepare the literature review, Section 2.2 presents the works found in the literature review that use machine learning techniques to detect failures in WWTPs, in Section 2.3, the works that use statistical methods to detect failures in WWTPs are described, and finally, Section 2.4 describes the the main types of faults found in the DO sensors, drift, bias, Precision degradation, spike and stuck, which will later be used to evaluate the performance of the proposed failure detection system.

## 2.1 Literature Survey Methodology

A fault is an unintentional deviation of a process characteristic that limits the process' ability to achieve its purpose [23]. In this sense, fault identification plays an important role int the WWTP management system. Environmental legislation, increasingly strict, requires the anticipation of the occurrence of failures or anomalies. An unidentified fault in WWTPs can pose serious risks to the environment and severe damage to the station's structures. And, failures in WWTPs can come from a variety of sources: changes in effluent quality (snow melt, industrial discharges, heavy rains), mechanical failures (pumps, aerators, heaters), abnormal growth or death of microorganisms (bacteria and algae),

damage to treatment structures (membranes, valves, pipes), and sensor failures (electrical interference, dirt, sensor degradation). Each failure has its own characteristics and different ways of affecting the treatment process.

Many works have already been proposed with the objective of predicting failures and anomalies in various parts of the wastewater treatment process. In order to better understand these techniques, their strengths and weaknesses, and to identify possible innovations, a search was carried out according to keywords in some repositories.

In order to conduct the literature survey, the title, abstract and field of study were searched in IEEE Xplore*, ScienceDirect†, MDPI‡, Springer§ and IWA¶ repositories. The query was done in publications of the last 20 years, and was made according to a combination of keywords, divided into groups.

TABLE 2.1: Search keywords.

| Group A | Group B | Group C |
|---|---|---|
| sensor | WWTP | fault detection |
| | Wastewater treatment plants | failure detection |
| | | fault identification |
| | | fault isolation |

The words that make up the groups were combined using logical operators: (Group A) AND (Group B) AND (Group C). Only articles related to failures in WWTPs were selected.

Most of the selected works are divided into statistical methods, methods that employ machine learning algorithms or a combination of both approaches. The basic organization of the articles found can be seen in the Figure 2.1. In the next section the works proposed in these categories will be analyzed.

*https://ieeexplore.ieee.org/Xplore/home.jsp
†https://www.sciencedirect.com
‡https://www.mdpi.com/
§https://www.springer.com/gp
¶https://iwa-network.org/

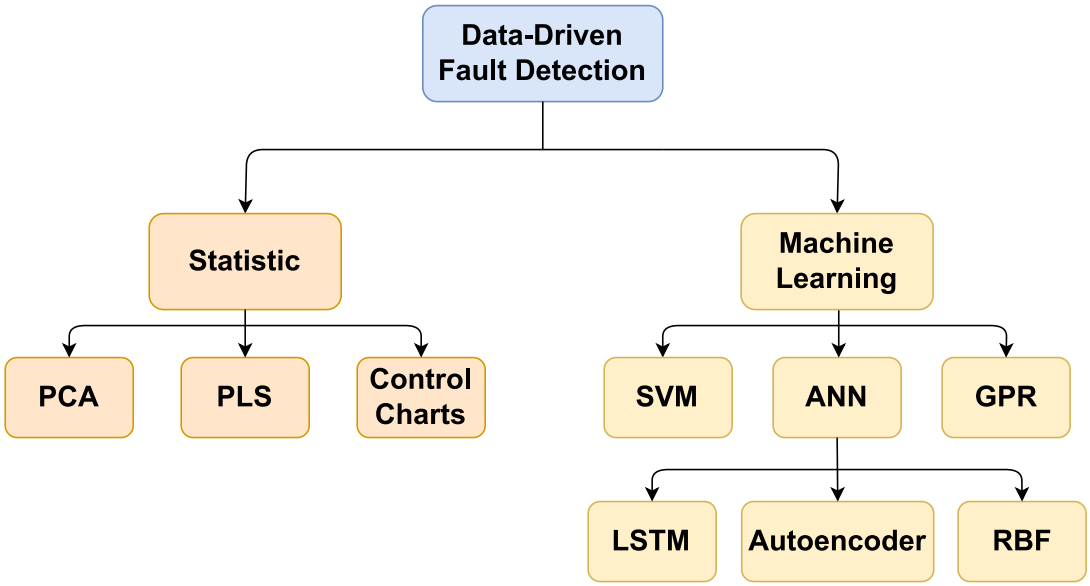



FIGURE 2.1: Data-driven fault detection methods employed in WWTPs.

## 2.2 Machine Learning Techniques

The use of machine learning (ML) techniques makes possible to extract important information from the obtained data from the treatment process carried out in the WWTPs. The ML techniques build a model based on data, modelling the complex relationship in which process inputs are non-linearly linked to outputs, without prior knowledge of an underlying mechanism.

Some machine learning techniques used to identify failures in WWTPs will be described below.

### 2.2.1 Artificial Neural Networks (ANN)

Artificial neural networks (ANN) is a massively parallel combination of simple and interconnected processing units which can acquire knowledge from the environment through a learning process and store the knowledge in its connections [24].

In [25], the use of ANN is proposed to identify six types of failures. The model receives 5 input variables: $D$ (sludge dilution rate), $W$ (aeration rate), $r$ (recirculation rate), $b$ (excess sludge rate), $S_{in}$ (inflow substrate concentration), and can produce seven results: normal operation, failure of the recirculation pumping, failure of the supply pump, failure of the excess of the sludge pump, failure of the biomass concentration sensor, failure of the dissolved oxygen (DO) concentration sensor, and partial failure ( 25%) at the supply

pump. The results of the test proved a good ability of the neural network to recognize the faults, correctly identifying 97.2% of failures in the analyzed cases.

A WWTP produces a large amount of data, most of which come from sensors, and it is a great challenge to identify collective faults in sensors. This is the objective of the work proposed in [26], that uses Long short-term memory (LSTM) networks to identify collective failures in the sensors. The data used were collected from a real WWTP. They were collected every minute, for a year, by 12 sensors, with each sensor responsible for 438.181 observations, totaling 5.1 million observations, and were labeled, by an expert, as normal and faulty. In the wastewater treatment process with the increase in ammonia levels, oxygen is pumped. Then the ammonia levels drop and the pumping of oxygen is stopped. And this cycle is repeated. A fault is detected when the oxygen is pumped, but the ammonia levels does not decrease. The results obtained by the LSTM were compared to the results of the Autoregressive Integrated Moving Average (ARIMA), Principal Component Analysis (PCA) and Support Vector Machines (SVM) models , and achieved a fault detection rate (recall) of over 92%, thus outperforming traditional methods and enabling timely detection of collective faults.

In the biological reactor tanks of the WWTPs, when the DO sensor fails, the system can be wrongly induced to increase or reduce the levels of oxygen pumping. A Radial Basis Function (RBF) neural network is used in [27] to identify faults in DO sensors in the biological reactor tanks. The RBF network is used to construct a time series that corresponds to the variable Oxygen Transfer Coefficient for tank 5 ($K_La_5$). If the actual verified value for the $K_La_5$ variable deviates considerably from the expected value, a failure in the DO sensor is indicated. The method performed well in relation to the delay for fault detection and false alarm rate. In [28], the use of a hybrid system for monitoring and detecting failures in WWTPs is proposed. The model is composed of a multivariate regression, an ANN and a hybrid technique that uses PCA for data pre-processing. The system receives 11 variables as input and aims to predict the value of total Kjeldahl nitrogen present in the biological reactor. The hybrid model performed well in relation to the root mean squared error (RMSE) metric.

There are two forms of training ANNs: supervised and unsupervised training. Supervised training requires data to be labelled in such a way that inputs and outputs are defined. Unsupervised training of a ANN uses data that are unlabeled. In fault detection, unsupervised ANN can be trained to model a process by estimating the values of inputs

and comparing the estimation to the actual values. This unsupervised training ANN is known as Autoencoder (AE). [29] used a deep associative neural network for fault diagnosis and prognosis, with "bottleneck" layer (i.e. the middle, hidden layer contains fewer nodes than the preceding or succeeding layers), as AE. This narrowing in the central layer force the ANN to effectively capture the principal components of the data to detect faults at a WWTP. The proposed model was used for detection of abrupt changes and drift in the sensor signal. The results showed that the proposed methodology is capable of detecting sensor faults and process with good accuracy under different scenarios. In [30], a variational AE is used for fault detection. The proposed model takes into account the temporal evolution of the treatment process. The slow feature variational AE (SFAVAE) model is used to monitor processes, using the Benchmark Simulation Model No 1 (BSM1) simulator, and tries to identify faults such as sludge expansion fault, small magnitude variable step and ramp fault. SFAVAE's performance is compared with the PCA, independent component analysis (ICA), kernel principal component analysis (KPCA) and variational AE (VAE) models. The results indicate that the SFAVAE model has higher effectiveness in process monitoring according to the false alarm rate and the missing alarm rate metrics.

### 2.2.2 Gaussian Process Regression

Gaussian process regression (GPR) is a non-parametric, Bayesian approach to regression which has already been used to identify faults and anomalies in WWTPs. In [31], GPR was used to monitor the Biochemical Oxygen Demand (BOD) value of the WWTP effluent. The value BOD value was used to predict the Sludge Volume Index (SVI), and to predict the occurrence of filamentous sludge bulking. The work develops a GPR model for multi-step-ahead prediction using Recursive, Direct and Direct-Recursive strategies. The data used correspond to 14 days of dry weather, and sampling is done every hour. To assess the system's performance in identifying drifts in the time series, a drift was added in the last three days. The GPR models were able to identify the drifts 6 hours in advance with a 95.5% probability.

The work carried out in [32] proposes two GPR models: GPR with maximum likelihood estimation (GPR-MLE) and GPR using Monte Carlo sequential estimation (GPR-SMC). Two situations are evaluated, the first being the impact of local optimum points in the estimation of missing values in the WWTP flow rate, and the second, the evaluation of

the most efficient method for detecting drifts in the values of sensors that measure ammonia levels. The work demonstrated that the GPR model can be used to monitor WWTPs efficiently. The GPR successfully detected the drift in the ammonium sensor and had few false alarms during the non-faulty test period. In the complete test data set, GPR had 74.5% drift detection percentage, and 15.9% false alarm percentage.

## 2.3 Statistical Techniques

Statistical methods, consolidated in the state-of-the-art, represent important tools for the identification of failures and anomalies in time series. Some of these techniques, used in WWTP, will be analysed below.

### 2.3.1 Control Charts

Control charts are important tools for determining whether a system is in control (IC) or out-of-control (OC). The most popular control chart was developed by Walter Shewhart of Bell Labs. The Shewhart control chart uses upper and lower control limits (UCL and LCL) for a process variable or statistics by adding or subtracting $k$ standard deviations from the variable's mean, with $k = 3$ being the industry standard [33]. If an observation is above the UCL, a statistically significant change has most likely occurred. Shewhart control charts employed at WWTP are typically constructed with a 3 or 5 days arithmetic moving average for variables designed to be stationary. The method proposed by Shewhart has the limitation of being applicable only to normally distributed and independent variables and the stationary time series [34].

To try to overcome the limitations of Shewhart control charts, some techniques were proposed. Exponentially weighted moving average (EWMA) can account for some non-stationarity found in WWTP data [35]. The Shewhart control chart assumes the process is stationary and weighs all past observations equally, ignoring trends [34]. The EWMA gives more weight to the most recent observations, adapting to some process variation [33]. In [36], it is proposed to use EWMA, together with a Kalman filter, to identify failures in WWTPs. The system monitors the variables chemical oxygen demand (COD), DO concentration, active heterotrophic biomass, ammonia and nitrate concentration and active autotrophic biomass. The proposed model is used to calculate the residual between

the actual and estimated values, and then the Control Chart is used to detect the failure. The proposed model performed well in relation to the false alarm rate and missed detection rate metrics.

EWMA is not the best method to distinguish between IC or OC. Like other data-driven methods, its performance is strongly affected by the presence of outliers [37].

### 2.3.2 Principal Component Analysis

A widely used statistical method for monitoring multiple variables simultaneously is principal component analysis (PCA) [38]. PCA is a mathematical method that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PC). PCs account for as much variation as possible and can, therefore, reduce the number of model variables and eliminate noise and redundancy. To use PCA for supervised, data-driven analysis, a training dataset that represents IC conditions are used to calculate the PCs, then testing data are transformed into the model subspace (defined by the PCs). If the overall distance from a new observation to the PCA model is above a desired control limit, then the new observation is considered abnormal.

PCA has many applications in WWTP, from direct fault detection [39] to data reconstruction [40]. For dynamic WWTP data, variations of PCA are often used, with adaptive PCA being the most common [41]. The work developed in [42] aims to detect failures in the DO sensor, of the biological reactor represented by BSM1. The faults are injected into the simulator, and the PCA is used to detect these faults. The detection method obtained good results according to the Hotelling's t-squared Statistic ($T^2$) and Squared Prediction Error (SPE) metrics. The work developed in [43] has the same objective, but here the technique used is the Monte Carlo simulation. The method used obtained good results according to the Fault detection, Fault isolation, and False alarm rate metrics

In [44], the method called Incremental PCA (IPCA) overcomes the difficulties characteristic of the traditional PCA, which has the tendency, due to non-stationary characteristics of WWTPs, to present a high number of false alarms. The method consists of two parts, an offline training and an online monitoring. In offline training the data is normalized, the number of PCs to be retained is determined and the limits, from which a failure is considered, are calculated. In online monitoring new data are collected. The statistical information SPE and $T^2$ are calculated, and it is checked whether these values

are within the limits considered normal. To detect failures 28 variables were monitored. Some failures were implemented to test the proposed system: step in the oxygen level in the aerobic digester, in the sedimentation speed of the secondary clarifier, and a step in the parameters of the bioreactor. In all the proposed failure situations, the IPCA was able to detect the fault, and isolate the variable that originated the failure, with false alarm rate and missed detection rate of 0.07% and 18.53%, respectively.

The method Adaptive PCA updates the model based on a "rolling window" of training observations. The training window is set to $n$ observations. With time evolution the oldest observations are removed, and more recent observations are added to the window, keeping $n$ constant. The rolling training window can thereby account for temporal non-stationarity found in WWTP. However, if the training window is too large, faults could be ignored [41]. If the training window is too small, normal observations could be flagged as faults [45].

A Probabilistic PCA approach in process monitoring and fault diagnosis with application in WWTPs is proposed in [46]. The Probabilistic PCA is compared to PCA, PPCA, GPLVM and Bayesian GPLVM. The PPCA is a probabilistic interpretation of the PCA, the GPLVM is a version of the PPCA for non-linear situations, and the Bayesian GPLVM uses the Beysian theory to train the GPLVM model and thus make it more robust to overfitting. The GPLVM and Bayesian GPLVM models showed better performance in detecting failures. This can be explained by adaptations to non-linearities. The PCA and PPCA methods showed high rates of false positives.

In [47], some methods to identify anaerobic digesters faults in WWTP are compared. A soft sensor implemented with an ANN estimates the value of volatile fatty acids (VFA), and is statistically compared with values considered normal. The techniques used in the comparison are PCA, SPE and Cumulative Sum Control Chart (CUSUM). SPE and CUSUM performed better than PCA in relation to the Recall, Precision and F1 score metrics. CUSUM performed better for errors of lesser magnitude, while SPE performed better for errors of greater magnitude.

The biggest disadvantage for applying the PCA in WWTPs is the assumption that process variables are linearly related to each other [48]. To account for the non-linear components of WWTP, data can first be mapped into a higher-dimensional, non-linear space, where observations are more likely to be linear [49]. PCA can be an important tool

for detecting failures in WWTPs when used in conjunction with other methods that compensate their weaknesses. In [50], it is proposed to use the PCA together with an ANN. The objective is to identify three failures: toxicity shock, Inhabitation and Bulking. The results obtained indicate greater efficiency than the classic PCA, increasing the sensitivity and robustness of the monitoring schemes.

### 2.3.3 Partial Least Squares

Similar to PCA, Partial Least Squares (PLS) identify linear combinations of measured variables, and outliers can be identified with $T^2$ and SPE statistics. In [51], the use of PCA and PLS is proposed, in a system called Multivariate Statistical Process Monitoring. The proposed system is used to monitor collective failures in WWTPs, and showed good performance in relation to the $T^2$ and SPE metrics, but presented considerable delay in the detection of failures. Unlike PCA, PLS differentiates between input variables and output variables and performs dimension reduction on each set of variables separately [52]. PLS only monitors the output variables that are affected by the input variables, while the PCA is used to monitor all variables in the process simultaneously. If an observation is abnormal but does not affect the final water quality, PLS will not signal the process as outside of the region considered normal, but PCA will [53].

### 2.3.4 Summary

A summary with the main information of the analyzed articles can be seen in Table 2.2. It is noticed that the most used techniques for the detection of failures in WWTPs are the ANN, in its multiple forms, and the PCA, with its adaptations.

Several works were analyzed, and many of these dealt with failures in the dissolved oxygen (DO) sensor, which monitors the biological reactor tanks, an important part of the treatment carried out in the WWTPs. The faults commonly found in these types of sensors can be classified into groups [42], which will be further explained in the next section.

TABLE 2.2: Summary of the reviewed works.

| Source | Contribution | Algorithm | Data | Failure Type | Performance Metrics |
|---|---|---|---|---|---|
| [25] | Multiple fault detection | ANN | Simulating scenario in presence of failures | Injection of faults in the analyzed dataset | TP, FP, TN, FN, ROC |
| [26] | Collective failures detection in sensors | LSTM | Real | Data labeled as normal and faulty | Precision, recall, and F1 score |
| [27, 43] | Faults detection in DO sensor | RBF | Simulator (BSM1) | Injection of faults in the analyzed dataset | Detection delay and false alarm |
| [29] | Abrupt changes and drift detection in sensors | AE | Simulator (BSM1) | Injection of faults in the analyzed dataset | False alarm and faulty detection rate |
| [30] | Multiple fault detection | AE | Simulator (BSM1) | Injection of faults in the analyzed dataset | $T^2$, SPE |
| [31] | Fault prognosis of filamentous sludge bulking | GPR | Simulator (BSM1) | Injection of faults in the analyzed dataset | RMSE, Correlation Coefficient (r) |
| [32] | Drift detection and forecast with $NH_4$ sensor data | GPR | Phenomeno-logical influent simulator and real data | Missing data/real drift | Normalized RMSE/Per-centage of detection, false alarm probability |
| [41, 42] | Real-time fault detection and isolation | PCA | Real and Simulator (BSM1) | Faults in nitrogen and DO sensors | Fault detection rate, $T^2$, SPE |
| [44] | Online sensor failure detection | IPCA | Simulator (BSM2) | Injection of faults in the analyzed dataset | False alarm rate, missed detection rate |
| [45] | Multiple fault detection | PCA based models | Real | Data labeled as normal or faulty | Fault detection rate |

TABLE 2.2: Summary of the reviewed work (continuation)

| Source | Contribution | Algorithm | Data | Failure Type | Performance Metrics |
|---|---|---|---|---|---|
| [46] | Multiple fault detection | Probabilistic PCA | Real | Data labeled as normal or faulty | Fault detection rate |
| [47] | Fault detection in anaerobic digestion | ANN, SVM, PCA | Simulator (BSM2) | Injection of faults in the analyzed dataset | Recall, Precision, F1 score |
| [49] | Multiple fault detection | PCA-based and PLS-based faults diagnosis | Simulator (BSM1) | Injection of faults in the analyzed dataset | Fault detection rate |
| [50] | Multiple fault detection | PCA and ANN | Simulator | Injection of faults in the analyzed dataset | Fault detection time |
| [51] | Multiple fault detection | PCA, PLS | Simulator (BSM1) | Injection of faults in the analyzed dataset | $T^2$, SPE |
| [36] | Multiple fault detection | EWMA, Kalman Filter and Control Chart | Simulator | Injection of faults in the analyzed dataset | False alarm and missed detection rate |
| [54] | Multiple fault detection | Lasso | Real data | Real faults present in the dataset | False signal rate, false signal count, detection accuracy and isolation rate |
| [28] | Monitoring of variables and fault detection. | Multivariate regression, ANN, PCA | Real | Deviation from the expected value | RMSE |

## 2.4 Sensor Fault Taxonomy

Deviations from expected behavior in the sensor output are considered faults. They are classified according to the trend in which they deviate from normal behavior. Let $s(t) = h(t) + \eta$ be the expected output of a sensor without the presence of faults, where $h(t)$ is the output of the sensor at time $t$, and $\eta \sim N(0, \delta_n^2)$ is noise [55]. The different types of fault can be defined as follows [56].

### 2.4.1 Drift Fault

When sensor output increases continuously at a constant rate, this type of fault is known as the drift fault. Many conditions can lead to this failure. Degradation of the material that makes up the sensor, extreme conditions where the sensor is installed (common condition in WWTPs), and corrosion, are some causes of drift fault. A drift fault can be mathematically defined as

$$s(t) = h(t) + \eta + b(t),\ b(t) = b(t-1) + v,\ v = constant \tag{2.1}$$

where $b(t)$ is the bias added to the signal at time $t$, and it increments with the passage of time: $b(t) = b(t-1) + v$. In this type of fault, the output value increases linearly from the normal value over time.

### 2.4.2 Bias Fault

In a bias fault, a constant value is added to the sensor output and, as a result, a shift from the normal value is observed. Mathematically, a bias fault is expressed as

$$s(t) \quad = \quad h(t) + \eta + v,\ v = constant \tag{2.2}$$

This fault is mainly due to calibration error.

### 2.4.3 Precision Degradation Fault

This type of fault adds noise with a zero mean and high variance to the output of a sensor. Mathematically, the precision degradation fault is expressed as

$$s(t) \quad = \quad h(t) + \eta + v,\ v \sim N(0, \delta_v^2),\ \delta_v^2 \gg \delta_n^2 \tag{2.3}$$

The reasons for this fault are sparse connections in the circuitry, high-frequency noise in the system, or physical damage in the sensor.

### 2.4.4 Spike Fault

As the name indicates, in spike faults, large-amplitude spikes are observed in sensor output. To obtain spike fault, a constant bias $b_t$ is added to the elements of the normal signal. It is modelled as

$$\begin{aligned} s(t) &= h(t) + \eta + b_t, \\ \forall\, t &= v \times \tau \end{aligned} \tag{2.4}$$

Where $v = \{1, 2, \dots\}$ is a set of natural numbers and $\tau$ is the interval in which the spikes occur in the sensor output, and $\tau \geqslant 2$.

### 2.4.5 Stuck Fault

It is a complete failure, with the sensor output being locked at a fixed value. It could be a temporary or permanent problem. This type of failure is mathematically defined as

$$s(t) = v, v = \text{constant}. \tag{2.5}$$

These are the most common faults found in DO sensors used in WWTPs. Any of these failures can result in heavy fines, damage to the environment, and possible damage to the WWTPs structure [54]. One way to better understand the dynamics of faults, and to evaluate detection methods, is through the use of simulators that represent the treatment process, with its stages and characteristics. For the present work, the BSM2 simulator will be used. The next chapter will provide a brief description of the wastewater treatment process carried out in WWTPs, and its representation by the BSM2 simulator.

# Chapter 3

# Benchmark Simulation Model No 2 - A Wastewater Treatment Plant Simulator

Organic pollution by wastewater discharge from human activities (cities, farming, industry) affects humans and ecosystems worldwide through the global sanitation crisis [57]. Wastewater resulting from various human activities cannot be released directly into nature. They can be a threat to public health and water resources, and so need treatment. This treatment takes place in special places designed for this purpose, the wastewater treatment plant (WWTP), and consists of several stages [58]: preliminary, primary, secondary, tertiary (reuse of treated water), and sludge treatments.

For testing and evaluating strategies related to WWTPs, the use of simulators is essential. For the development of the work presented here, the simulator Benchmark Simulator Model No 2 (BSM2) was used, which simulates all stages of wastewater treatment that occurs in a WWTP. Details about wastewater treatment and the simulator used will be given in this chapter, which is organized as follows. Section 3.1 provides information on the steps in the wastewater treatment process carried out by WWTPs. Section 3.2 describes the BSM2 simulator, with its composition and layout. Section 3.3 describes the Activate Sludge Model No1 module which represents the biological treatment of wastewater. In Section 3.4, the model of the anaerobic digester present in BSM2 is detailed. And, Section 3.5 details general characteristics of the BSM2, such as tank volume and water retention time.

## 3.1 Wastewater Treatment Plant

The treatment of wastewater from various human activities must be treated before returning to the environment, and this treatment takes place in WWTPs. The treatment made in the WWTPs, represented in Figure 3.1, will be briefly described below.

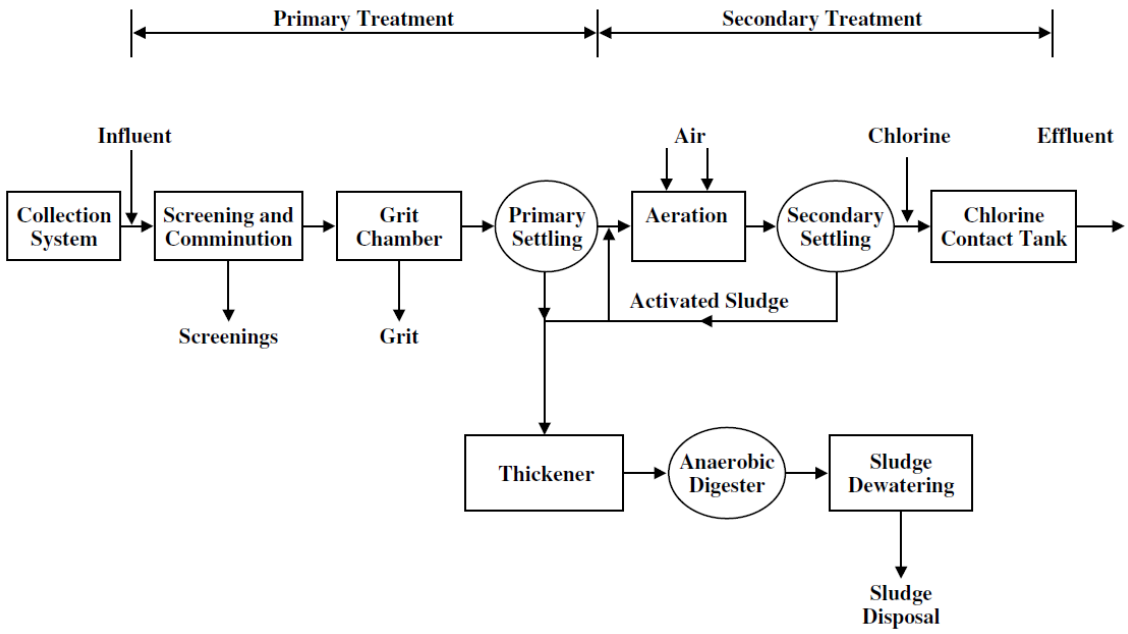



FIGURE 3.1: Schematic of conventional wastewater treatment process [58].

The stages of wastewater treatment are as follows:

**1-Preliminary treatment:** removal, with the help of bars, of larger materials that could damage the treatment plant equipment or reduce the treatment capacity due to its dimensions. At this stage, sand is also removed by decantation.

**2-Primary treatment:** removes settleable and floatable solids (may not be present in all treatment plants [58]). In this step, floating materials such as oils and greases are also removed.

**3-Secondary treatment (activated sludge and decantation):** removes biochemical oxygen demand (BOD) and dissolved and colloidal suspended organic matter by biological action. It is the biological and oxidative treatment, and takes place in the activated sludge reactor (biological reactor). In this process, bacteria consume organic matter, and for that they need oxygen. Thus, this tank has an oxygenated part. At the end of the process, the waste water goes to a secondary settler, and the organic matter decants. Part of this decanting matter, formed by organic matter and bacteria, is redirected to the entrance of the activated sludge reactor. Thus, the concentration of bacteria in this reactor is much

higher than what would be natural in wastewater, and thus, organic matter is consumed more quickly. It's a way to speed up a natural process.

**4-Tertiary treatment (reuse of treated water):** due to water scarcity, in certain regions the use of water from WWTPs represents an important alternative [59]. But this water needs to be treated to reach minimum levels of quality. At this stage of treatment, the wastewater is subjected to disinfection and nutrient removal. Bacteria, suspended solids, excess nutrients and specific toxic compounds are removed. After this treatment, the water can be used in agriculture, watering green spaces, washing pavements and streets, among other possible uses.

**5-Sludge treatment:** the sludge produced in the treatment, as well as the final water, needs to be treated so that it can be discarded, or reused (agriculture, landfills, etc). Mechanical, chemical, thermal, thermochemical, and biological treatments are applied to sludge [60].

The steps described can vary according to the type of treatment and the capabilities of the WWTPs, but in general these are the processes used, and each of these steps can be divided into sub-processes, and done in several different ways. The generic structure of a WWTP is represented by the simulator Benchmark Simulation Model N°2 (BSM2), that will be described in Section 3.2.

## 3.2 Benchmark Simulation Model N°2

WWTPs are structures subject to many variations, and with highly non-linear characteristics [61], and which, nevertheless, need to function within the strict limits imposed by environmental legislation [62].

Many strategies related to attempts to make the wastewater treatment process more efficient have been, and are being implemented, but comparing these methods is difficult due to the complexity of the system. And, to achieve innovations, evaluations must be carried out based on rigorous methods, with well-established performance criteria.

In this sense, the simulation environment Benchmark simulation model no.2 (BSM2) was developed. From 1998 to 2004, the development of benchmark tools for simulation-based evaluation of control strategies for activated sludge plants has been issued in Europe by Working Groups of COST Action 682 and 624. This work has been continued under the umbrella of the International Water Association (IWA) Task Group on Benchmarking of Control Strategies for WWTPs [63]. BSM2 (or part of them, eg. BSM1) has

been widely used by the research community, as for example, in the following works [12, 64–70].

The BSM2 is a simulation environment defining the plant layout, the simulation model, influence loads, test procedures and evaluation criteria. For each of these items, compromises were pursued to combine plainness with realism and accepted standards [22].

BSM2 describes a structure responsible for the treatment of wastewater, consisting of biological and sludge treatment. The simulator is mainly composed by primary settling, activated sludge system (including anoxic and aerobic reactors and a secondary clarifier), anaerobic digester, sludge thickener and dewatering, and a sludge storage unit. The layout of BSM2 can be seen in Figure 3.2.

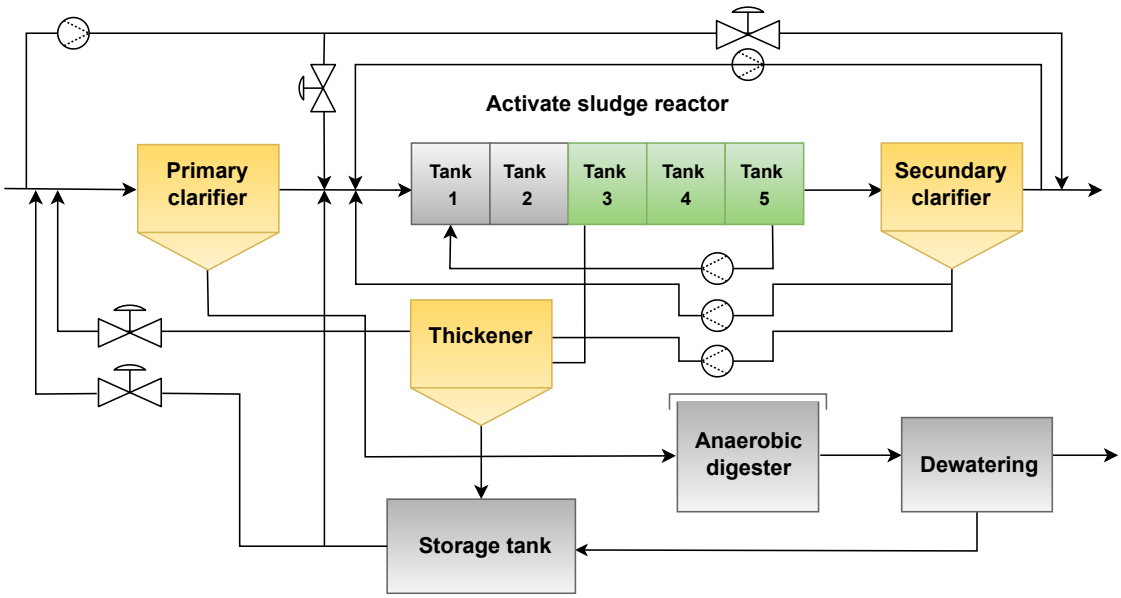



FIGURE 3.2: Layout of BSM2.

Two models used in the BSM2 units are the Activate Sludge Model N°1 (ASM1) and the Anaerobic Digester Model N°1 (ADM1). The ASM1 models the unit where the biological treatment takes place, while ADM1 works with one of the later units, responsible for anaerobic digestion and biogas generation.

## 3.3 Activate Sludge Model N°1 - ASM1

In 1983, a working group was formed, under the responsibility of the IWA, with the objective of providing the application and validation of models of biological systems for the treatment of effluents in activated sludge processes. This work group culminated with the

launch of a family of biological treatment models using activated sludge. The first model released was called Activate Sludge Model N°1, or ASM1.

The ASM1, considered as the reference model for the biological treatment process, configures the biological reactor model or activated sludge reactors, with five compartments, with an anoxic zone, made up of two tanks, and an aerated zone, made up of three tanks. In these two sections the processes of denitrification and nitrification take place. The ASM1 is part of the Benchmark Simulation Model N°1 (BSM1), which can be seen in Figure 3.3, and presents a typical activated sludge system, combining the ASM1 model with a secondary clarifier [71]. The BSM1 simulator was used as a basis for the elaboration of the BSM2, which added the primary decanter and the sludge treatment tanks to the original project.

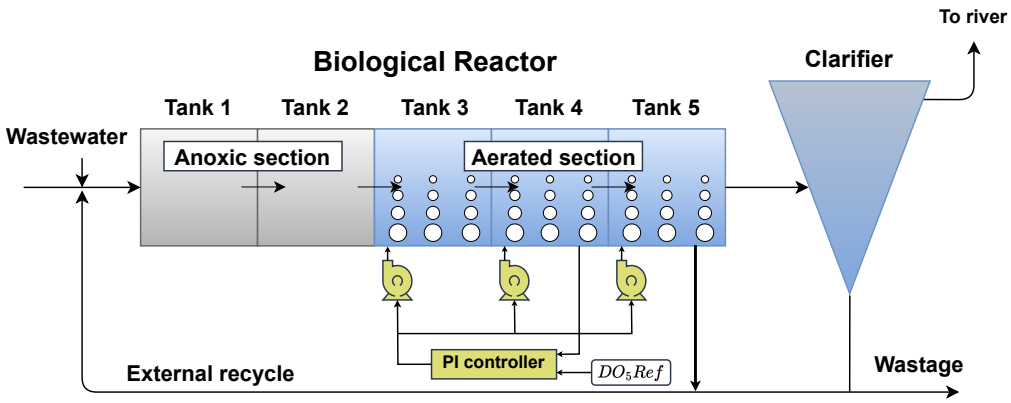



FIGURE 3.3: Layout of BSM1.

To keep the bacterial population responsible for the degradation of organic matter, the sludge from the secondary clarifier is recirculated to the anoxic tanks. The sludge concentration is kept constant through continuous recirculation from the clarifier [72].

The treatment process that takes place in the biological reactor can be summarized in eight processes [61]:

- 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 neutralize threats posed by chemical compounds produced by human activities, and accelerate the process that occurs in nature.

## 3.4 Anaerobic Digester Model N°1 - ADM1

As mentioned earlier, a product of wastewater treatment is sludge. This compound needs to be treated before being used or disposed. In this sense, the anaerobic digester, described by the ADM1 model [73], performs anaerobic digestion and biogas generation.

Many factors influence the anaerobic digestion process. The main ones are temperature and pH, which strongly influence bacterial metabolism. For optimal bacterial growth the pH should be between 6.7 and 7.5 [74]. As a product of the anaerobic bacterial activity, we have the production of biogas. A typical biogas composition of digested sludge is methane (CH4, 50–70%) and carbon dioxide (CO2, 30–50%) [75].

It can be concluded that the ADM1 and ASM1 models, which make up the BSM2 simulator, allow a realistic evaluation of the implemented techniques related to WWTPs.

## 3.5 General Characteristics of BSM2

As general characteristics, the BSM2 simulator is designed for an average influent dry-weather flow rate of $20648,36\,[m^3/day]$ and an average biodegradable Chemical Oxygen Demand (COD) in the influent of $592,53\,[g/m^3]$. Its hydraulic retention time is 22 hours, based on average dry weather flow rate and total tank volume – i.e. primary clarifier ($900\,[m^3]$) + biological reactor ($12000\,[m^3]$) + secondary clarifier ($6000\,[m^3]$). The influent dynamics are defined for 609 days by means of a single file, which takes into account rainfall effect and temperature [76], with the data sampled every 15 minutes.

During the various stages of the treatment process, many variables can be accessed. The main ones 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$^3$]);
- $X_{B,H}$ (heterotrophic biomass, [g COD/m$^3$]);
- $X_{B,A}$ (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, [°C].

At the end of the simulation period, a report is generated with various informations related to indicators, including energy consumption, general expenses in the treatment process, biogas generation and the quality of the resulting water. With the help of BSM2, it is possible to implement and test strategies for control, optimization, prediction of variables, detection of failures and anomalies, etc. It is possible to compare results according to pre-established metrics, and thus achieve innovations and contributions to mitigate the problems related to wastewater.

# Chapter 4

# Fault Detection using Autoencoders

This chapter aims to describe the Autoencoder (AE), their basic structure and usage. It also describes the two models of AEs used in this work. The chapter is organized as follows: Section 4.1 presents some use cases for AEs, and Section 4.2 describes the basic structure of AEs, with their mathematical representation. In this section, the two EA models used in the development of the work are also described.

## 4.1 Autoencoder

AEs are artificial neural networks with a wide application area: classification and regression [77], reconstruction of missing parts in an image [78], hierarchical feature extraction [79], analysis of large neuroimaging datasets [80], online data stream analysis and action recognition [81], 3D face reconstruction [82], natural language processing [83], object detection [84], image classification [85], etc.

The main feature of AEs is the ability to represent high-dimensional data in reduced dimensions, while maintaining the original quality as much as possible. AEs are structures that work like the PCA, which projects high-dimensional data onto smaller-dimensional data. However, PCA applies linear transformation, while AE apply non-linear functions. There are many applications, but there are few works that apply AE to solve problems in WWTPs. In the next section, the basic structure of AEs will be described. Variations and characteristics that allow its application in WWTPs will be explained.

## 4.2 General Architecture of Autoencoders

As stated in the introduction to this chapter, AE is an unsupervised machine learning algorithm that aims to reconstruct its input signal. It was proposed by Geoffrey Everest Hinton in the 1980s to solve the "Backpropagation without a teacher" problem [86]. The generic AE model consists of three parts:

- **Encoder:** responsible for reducing the dimensionality of the data.
- **Code(activation):** latent/compressed representation of the data, contains the reduced representation of the encoder input data. It is a non-linear mapping that transforms the encoded coefficients into the range $[0, 1]$.
- **Decoder:** responsible for expanding the dimensionality represented in the Code, and reconstructing the input signal.

An AE is a typical feed-forward neural network that connect many simple neurons. The output of a neuron can be the input of another, and the parameters can be estimated using a back-propagation algorithm. A forward pass is first run to compute the activations weights (a non-linear mapping) throughout the network. For each middle node, an error term that measures how much that node was responsible for any errors in the output is calculated. For an output node, the difference between the network's output and the true target value can be measured directly and further used to renew the error term. A representation of the AE can be seen in the Figure 4.1, and a generic flowchart of an AE can be seen in the Figure 4.2.

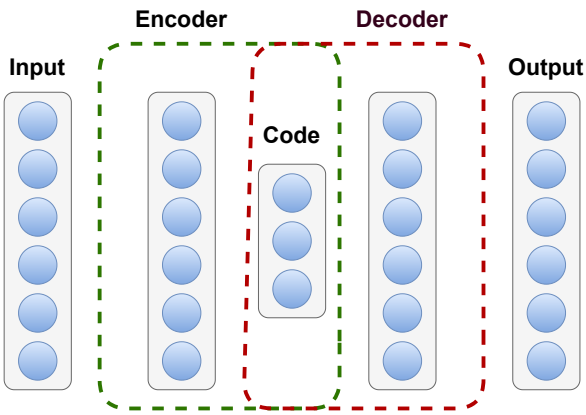



FIGURE 4.1: Generic AE model.

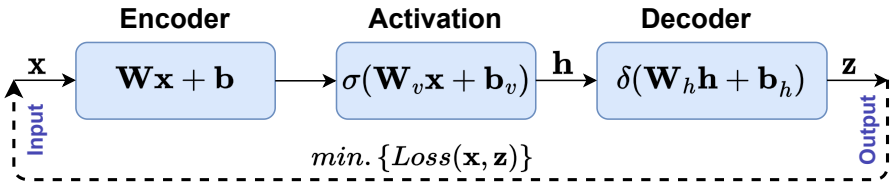



FIGURE 4.2: Generic flowchart of an AE. Adapted from [87].

Given the visible input layer $\mathbf{x}$, the AE first maps it into a hidden layer $\mathbf{h}$ with a weight matrix $\mathbf{W}_v$, bias $\mathbf{b}_v$, and an activation function $\sigma(\cdot) : \mathbb{R} \to [0,1]$:

$$\mathbf{h} = \sigma(\mathbf{W}_v\mathbf{x} + \mathbf{b}_v) \tag{4.1}$$

The hidden units are then listed into the output layer $\mathbf{z}$ inversely with weight matrix $\mathbf{W}_h$, bias $\mathbf{b}_h$, and activation $\delta(\cdot) : \mathbb{R} \to [0,1]$:

$$\mathbf{z} = \delta(\mathbf{W}_h\mathbf{h} + \mathbf{b}_h) \tag{4.2}$$

To simplify the network architecture, the tied weights strategy $\mathbf{W}_v = \mathbf{W}_h = \mathbf{W}$ are used, and the parameters to be determined are $\{\mathbf{W}, \mathbf{b}_v, \mathbf{b}_h\}$. The objective function to train an AE is to minimize the cost function $\mathcal{J}$,

$$\mathcal{J}(\mathbf{W}, \mathbf{b}_v, \mathbf{b}_h). \tag{4.3}$$

Given the training samples, the cost function is defined as

$$\mathcal{J}(\mathbf{W}, \mathbf{b}_v, \mathbf{b}_h) = \mathcal{L}(\mathbf{x}, \mathbf{z}) + g(\mathbf{W}), \tag{4.4}$$

where $\mathcal{L}(\mathbf{x}, \mathbf{z})$ is the loss function and $g(\mathbf{W})$ is a regularization. The weight matrix and bias are obtained by the minibatch stochastic gradient descent (SGD) algorithm

$$\begin{aligned}
\mathbf{W}^{new} &= \mathbf{W}^{old} - \alpha \frac{\partial}{\partial \mathbf{W}} \mathcal{J}(\mathbf{W}, \mathbf{b}_v, \mathbf{b}_h), \\
\mathbf{b}_v^{new} &= \mathbf{b}_v^{old} - \alpha \frac{\partial}{\partial \mathbf{b}_v} \mathcal{J}(\mathbf{W}, \mathbf{b}_v, \mathbf{b}_h), \\
\mathbf{b}_h^{new} &= \mathbf{b}_h^{old} - \alpha \frac{\partial}{\partial \mathbf{b}_h} \mathcal{J}(\mathbf{W}, \mathbf{b}_v, \mathbf{b}_h),
\end{aligned} \tag{4.5}$$

where $\alpha$ is the learning rate, $\mathbf{W}^{old}, \mathbf{b}_v^{old}, \mathbf{b}_h^{old}$ denote the old model parameters, and $\mathbf{W}^{new}, \mathbf{b}_v^{new}, \mathbf{b}_h^{new}$ are the renewed ones.

From the basic structure of AEs, many different models with different applications can be obtained. The next subsection outlines the main variations of AEs.

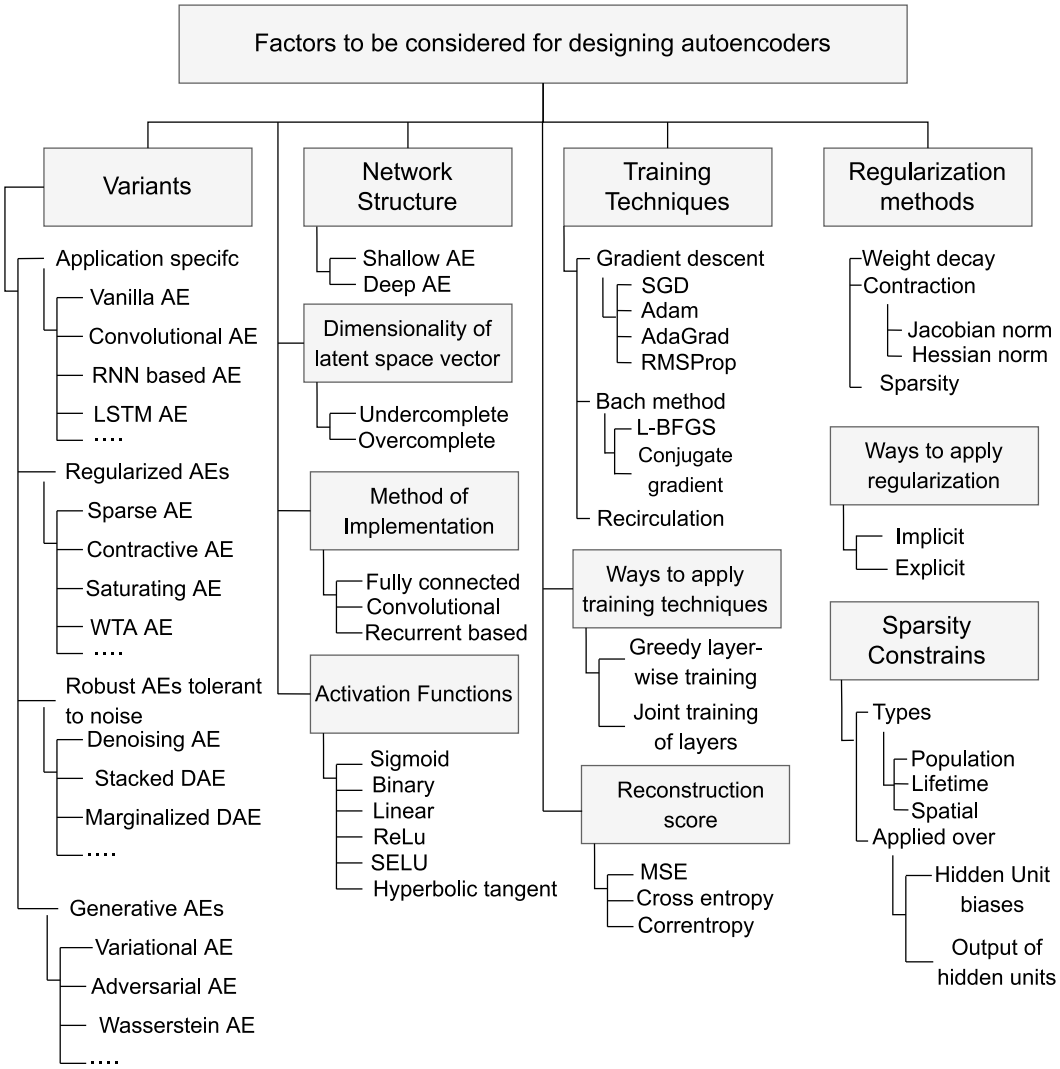



FIGURE 4.3: Factors to consider when implementing AEs. Adapted from [88].

### 4.2.1 Variants of Autoencoders

Many parameters can be modified when implementing AE. It can have a simple structure, with only one hidden layer, or can have several hidden layers, characterizing a deep AE model. The representation of the input data, made by the latent layer $h$ (Figure 4.2), can be classified as under complete or over complete [88]. An under complete representation, where the representation dimension is smaller than the input dimension, forces learning of the most essential features of the training data. The AE can be implemented as fully connected, convolution based or recurrent based units. The number of neurons in each layer can vary according to the objectives to be achieved for the proposed problem. These are just a few of the many factors. The main factors that should be considered when implementing AEs can be seen in the Figure 4.3.

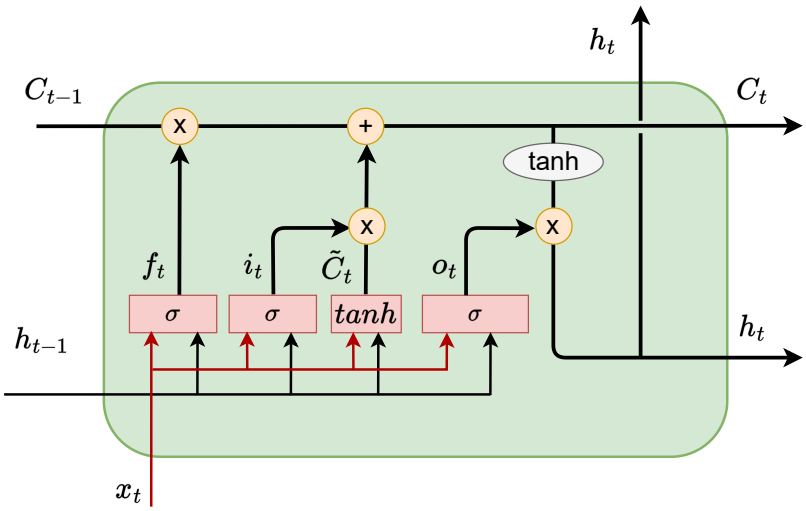



FIGURE 4.4: Generic structure of the LSTM cell. Adapted from [90].

In this work, two forms of AE will be used: Long short-term memory (LSTM) and Convolutional AEs, named here as LSTM-AE and Convolutional-AE, respectively. These models will be described in the next subsections.

#### 4.2.1.1 LSTM Autoencoder

LSTM, proposed in [89], is a recurrent neural network that takes into account the historical context of events to make its predictions, with the help of memory cells. The generic structure of an LSTM cell can be seen in Figure 4.4.

The detailed process of updating the LSTM neural unit is as follows [89, 91]:

- The input $\mathbf{h_{t-1}}$ and $\mathbf{x_t}$ are received as input values of the forget gate at time $t$. The output $\mathbf{f_t}$ of the forget gate is obtained. The formula is as follows:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x_t}] + \mathbf{b}_f), \tag{4.6}$$

  where the value range of $\mathbf{f}_t$ is 0 to 1, $\mathbf{W}_f$ is the weight of the forget gate, and $\mathbf{b}_f$ is the bias of the forget gate.

- The input $\mathbf{h}_{t-1}$ and $\mathbf{x}_t$ are received as input values of the input gate at time $t$. The output $\mathbf{i}_t$ and the candidate cell state $\tilde{\mathbf{C}}_t$ of the input gate are obtained. The formulas are shown in Equations 4.7 and 4.8:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x_t}] + \mathbf{b}_i), \tag{4.7}$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x_t}] + \mathbf{b}_c), \tag{4.8}$$

where the value range of $\mathbf{i}_t$ is 0 to 1, $\mathbf{W}_i$ is the weight of the input gate, $\mathbf{b}_i$ is the bias of the input gate, $\mathbf{W}_c$ is the weight of the candidate input gate, and $\mathbf{b}_c$ is the bias of the candidate input gate.

- Update the cell status $\mathbf{C}_t$ at time $t$. Its equation is as follows:

$$\mathbf{C}_t = \mathbf{f}_t \mathbf{C}_{t-1} + \mathbf{i}_t \tilde{\mathbf{C}}_t, \tag{4.9}$$

  where the value range of $\mathbf{C}_t$ is 0 to 1.

- The input $\mathbf{h}_{t-1}$ and input $\mathbf{x}_t$ are received as input values of the output gate at time $t$, and the output $\mathbf{o}_t$ of the output gate is obtained. The equation is as follows:

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o), \tag{4.10}$$

  where the value range of $\mathbf{o}_t$ is 0 to 1, $\mathbf{W}_o$ is the weight of the output gate, and $\mathbf{b}_o$ is the bias of the output gate.

- The final output value $\mathbf{h}_t$ of the LSTM neural unit is calculated as shown in Equation 4.11.

$$\mathbf{h}_t = \mathbf{o}_t \tanh(\mathbf{C}_t) \tag{4.11}$$

In [92], the LSTM-AE is described as an extension to RNN based AE for learning the representation of time series sequential data. In this model, encoder and decoder are built using LSTM. Encoder LSTM accepts a sequence of vectors in the form of features. Decoder LSTM recreates the target sequence of input vectors in the reverse order. A representation of the LSTM-AE can be seen in Figure 4.5.

#### 4.2.1.2 Convolutional Autoencoder

Fully connected AEs ignore the spatial structure of the input signal, and this spatial structure can represent important information for the final reconstruction. To solve this problem, in [79] is proposed a model known as convolutional AE (convolutional-AE). Instead of using fully connected layers, convolutional-AE use convolutional operators that allow extracting important representations from the input data. Forms of redundancy in parameters are introduced as a way to have a global learning. Because the weights and bias are shared among all locations of input, the spatial locality can be preserved [87].

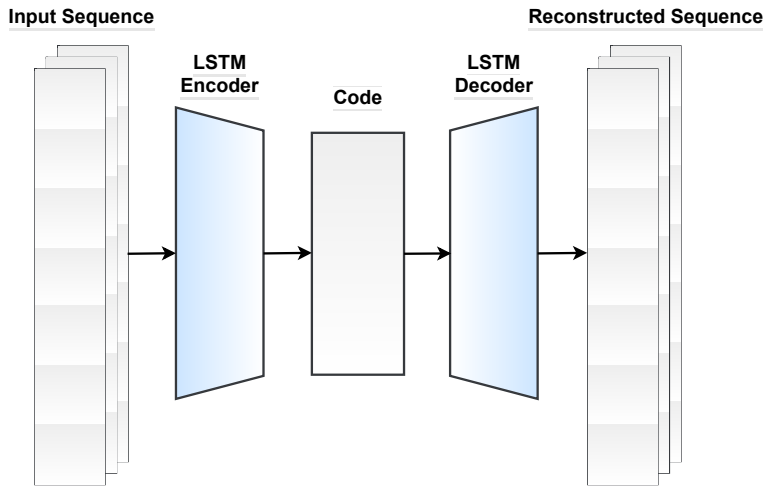



FIGURE 4.5: LSTM-AE. Encoder and decoder implemented with LSTM cells.

Given the input signal $\mathbf{x}$, the latent representation of the $k-th$ feature map is given by

$$\mathbf{h}^k = \sigma(\mathbf{x}\mathbf{W}^k + \mathbf{b}_v) \tag{4.12}$$

Where $\mathbf{h}^k$ is the latent representation of the $k-th$ feature map, $\sigma(\cdot)$ is the activation function, $\mathbf{W}$ is the weight of the units, and the bias of the visible units $\mathbf{b}_v$ is broadcast to the whole map, being one bias per latent map.

As in convolutional neural networks, AE uses max-pooling layers. Max-pooling or Maximum pooling, it is a pooling operation that calculates the maximum, or largest, value in each patch of each feature map. The results are a pooled feature maps that highlight the most present feature in the patch. The max-pooling layer is connected with the convolutional layer, and the reconstruction is obtained using

$$\mathbf{z} = \delta(\sum_k \mathbf{h}^k \tilde{\mathbf{W}}^k + \mathbf{b}_h) \tag{4.13}$$

Where $\tilde{\mathbf{W}}^k$ denotes the flip of $\mathbf{W}^k$ over both dimensions of the weights, and $\mathbf{b}_h$ is the bias of the hidden units. In the reconstruction, the latent coding decreases the average number of filters contributing to the decoding of each input signal, forcing filters to be more general. The cost function $\mathcal{J}_{conv}$ used by convolutional-AE is given by

$$\mathcal{J}_{conv} = \mathcal{L}(\mathbf{x}, \mathbf{z}) = \sum_i \{\|\mathbf{x}_i - \mathbf{z}_i\|_2^2\} \tag{4.14}$$

Where $\mathbf{x}_i$ and $\mathbf{z}_i$ represent the signal value and its reconstruction at time $i$, respectively. A representation of the Convolutional-AE can be seen in Figure 4.6.

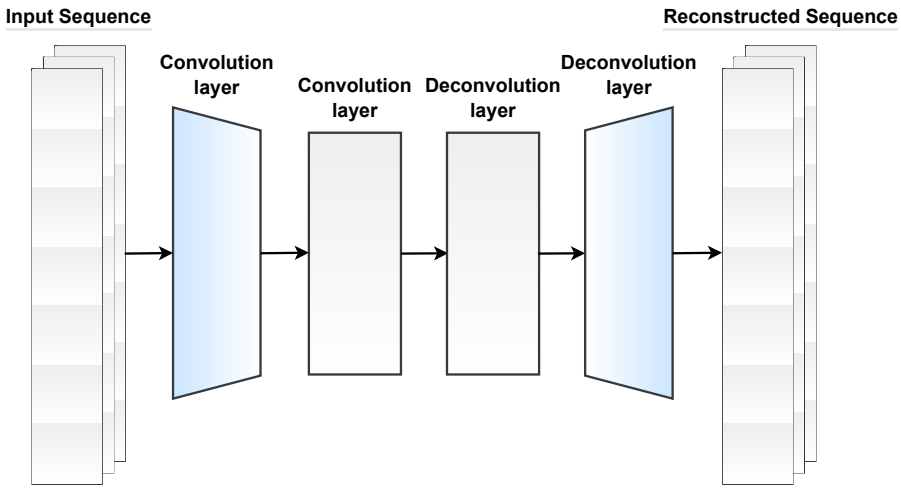



FIGURE 4.6: Convolutional-AE. Encoder and decoder implemented by convolution process.

In the present work the LSTM and Convolutional AEs will be used to detect failures, in real time, in the DO sensor of the biological reactor of a WWTP, represented by the BSM2. By obtaining, as output, the signal representing the expected behavior for the DO sensor, a threshold can be determined from which a failure is indicated.

# Chapter 5

# Dataset and Fault Implementation

This chapter aims to describe the origin of the dataset used for fault detection. The sensors used by the simulator, the faults injected into the system, and the graphical representations of these faults are presented. The chapter is organized as follows: Section 5.1 describes the sensors present in the BSM2, with emphasis on the DO sensor, and describes, in general terms, the functioning of the aeration system of the biological reactor tanks. Section 5.2 details the failure injection process in the dataset used.

## 5.1 BSM2 Dataset

In order to evaluate new methods proposed for WWTPs, the use of simulators is very important. As described in Chapter 3, BSM2 is an important simulator widely used in the development and testing of techniques related to WWTPs. All treatment phases of a conventional treatment plant are represented by the simulator. The BSM2 model implemented in Matlab/Simulink can be seen in Figure 5.1.

Sensor information supports several decisions related to the functioning of WWTP. The information from the sensors helps to determine the amount of chemical reagents that must be added to the treatment process, the time that the aerators must run, the time that purge valves must remain open, etc. Similarly, many of the techniques proposed to optimize WWTPs are based on information obtained by sensors. For the result obtained by the simulator to be realistic, the sensor represented in the BSM2 model must present a behavior close to that found in a real situation. To avoid unrealistic behavior, the BSM2 uses sensor models modeled taking into account information such as saturation, response time and noise. The sensors used by BSM2 can be classified into six classes, according to

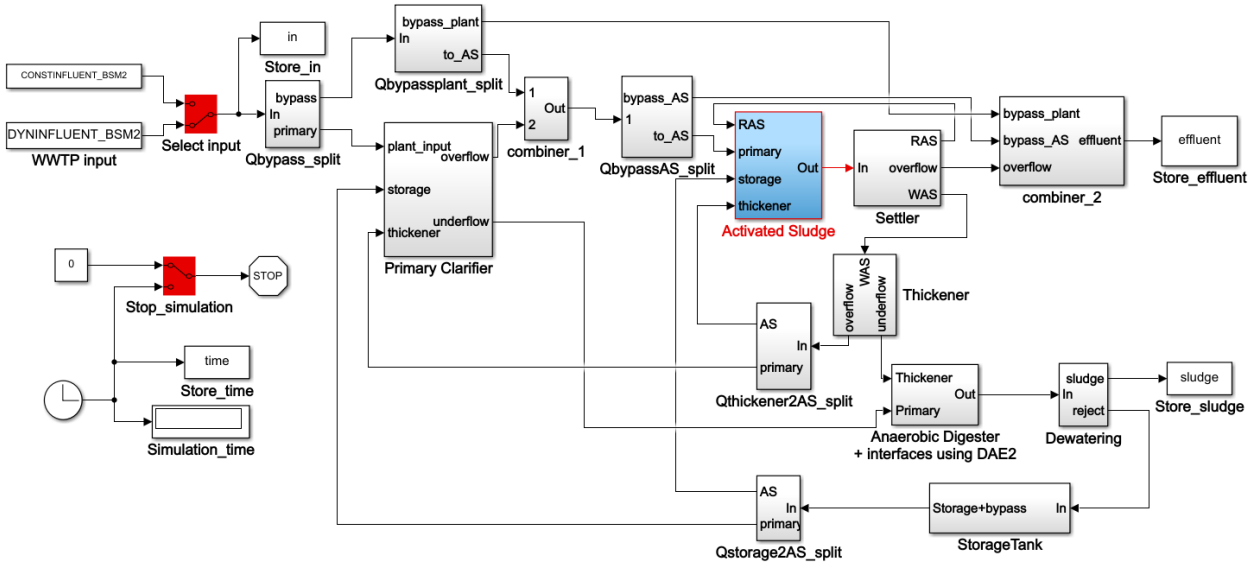



FIGURE 5.1: BSM2 simulator implemented in Matlab/Simulink. Biological reactor and activated sludge system represented by the blue block.

their application. Table 5.1 shows the sensor classes available on the BSM2. The DO sensor, a Class A sensor, can be seen in Figure 5.2, and its representation in the activated sludge treatment system can be seen in Figure 5.3.

TABLE 5.1: Sensor classes. A measuring interval equal to 0 means continuous measurement [93].

| Sensor classes | Response time [min] | Measuring interval [min] | Examples |
|---|---|---|---|
| Class A | 1 | 0 | Ion sensitive, optical without filtration |
| Class B0 | 10 | 0 | Gas-sensitive + fast filtration |
| Class B1 | 10 | 5 | Photometric + fast filtration |
| Class C0 | 20 | 0 | Gas-sensitive + slow filtration |
| Class C1 | 20 | 5 | Photometric + slow filtration or sedimentation |
| Class D | 30 | 30 | Photometric or titrimetric for total components |

The oxygen level in the biological reactor tanks is maintained with the aid of a Proportional-Integral (PI) controller. The PI controller aims to maintain the level of DO in the oxygenated tanks of the biological reactor at 2 [mg/L]. To achieve this objective, the PI controller receives measurements from the DO sensor, compares it with the reference value

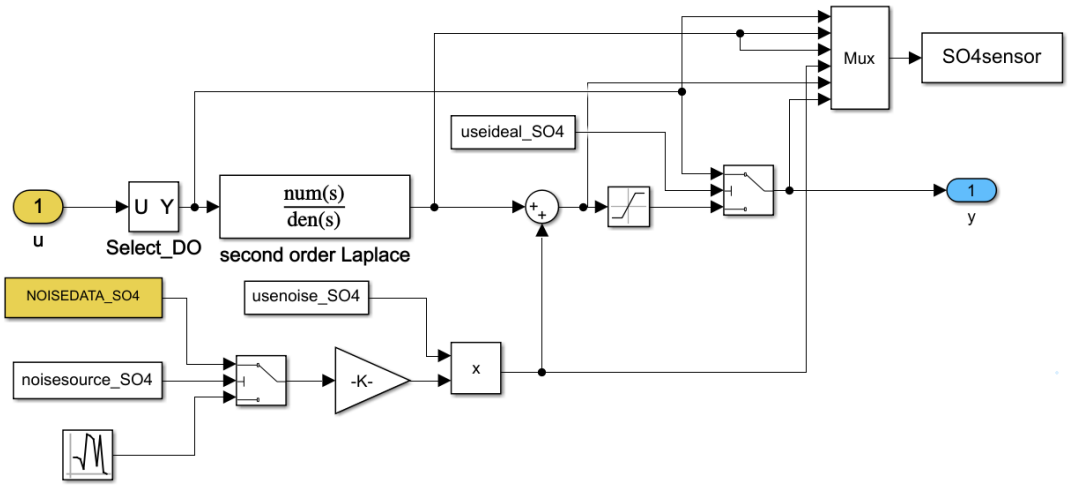



FIGURE 5.2: DO sensor represented in Matlab/Simulink.

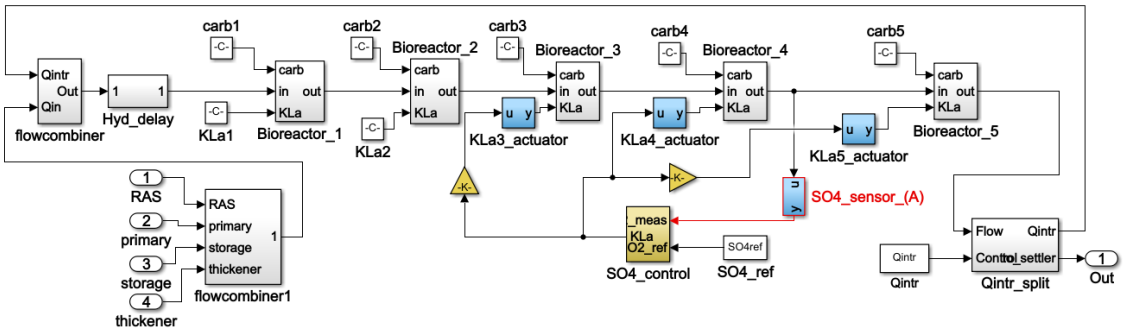



FIGURE 5.3: Activated sludge treatment system with DO sensor.

and provides the operating signal to the aerators of bioreactors tanks 3, 4 and 5.

From the operating process of the treatment system, simulated by the BSM2, the importance of the DO sensor is clear. Failures in the DO sensor can cause unnecessary expenses with excess oxygen pumping, or it can impair the treatment due to lack of oxygen, reducing bacterial activity, and not treating wastewater correctly.

## 5.2 Dissolved Oxygen Sensor Data and Fault Injection

The sensor used to measure the DO level at the exit of tank 4 can be seen in Figures 5.2 and 5.3. The continuous measurement of the sensor provides the DO values for the PI controller. A signal from this sensor under normal operating conditions can be seen in the Figure 5.4.

Figure 5.4 represents only two days of DO sensor signals, but the BSM2 can simulate continuous cycles of 609 days of a WWTP operation. For training and tests of the models

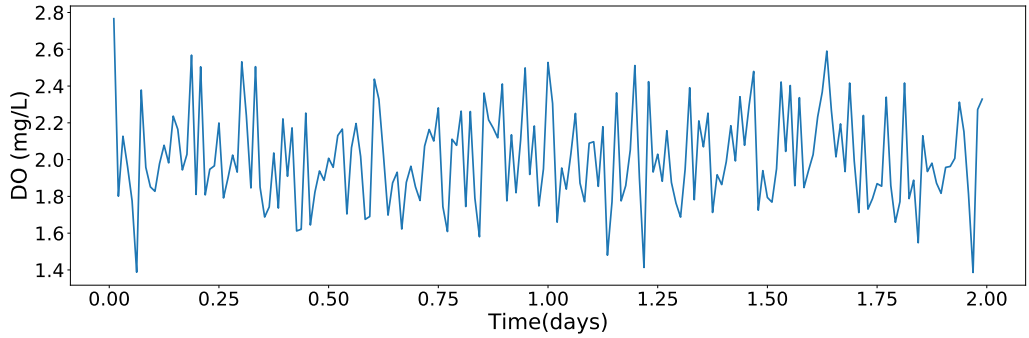



FIGURE 5.4: DO level obtained by the sensor in two days of operation.

used to detect failures in the DO sensor, 200 days of operation were considered, with the last 100 days of a cycle used for training and validation, and the first 100 days of the following cycle used for tests. The data used for training and validation and for the final tests can be seen in Figures 5.5 and 5.6, respectively.

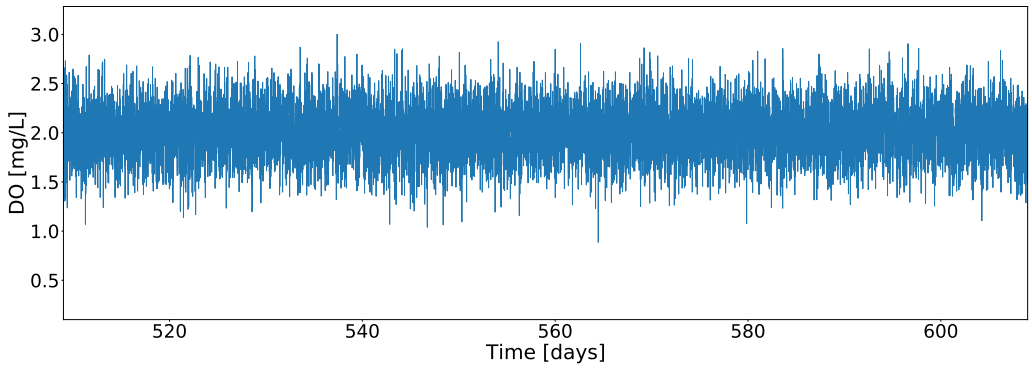



FIGURE 5.5: Time series used for training and validation.

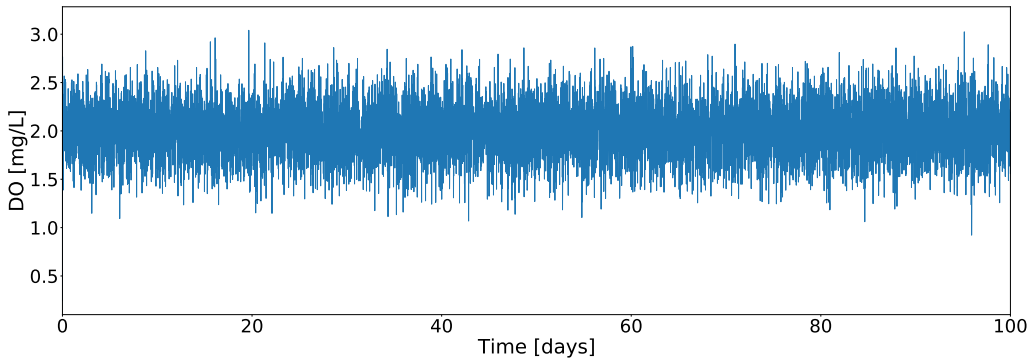



FIGURE 5.6: Time series used for fault injection and tests.

To evaluate the performance of the fault detection techniques analyzed in this work, the faults described in Section 2.4 were injected into the 100 days used for the tests. The fault injection process will be further explained in the following subsection.

## 5.3 Fault Injection

The faults considered were injected directly into the sensor model, and programmed to act on specific days. To give more realism and allow a better evaluation of the detection techniques, three scenarios were evaluated: Scenario 1 - injection of faults on specific days; Scenario 2 - injection of faults with change in the order of appearance and duration of faults; and Scenario 3 - injection of faults with change in order of appearance, intensity and duration of faults. The implementation of sensor failures, in the BSM2 simulator, can be seen in Figure 5.7. Details about the faults injected into the DO sensor, as well as their graphical representation, will be given in the next subsections.

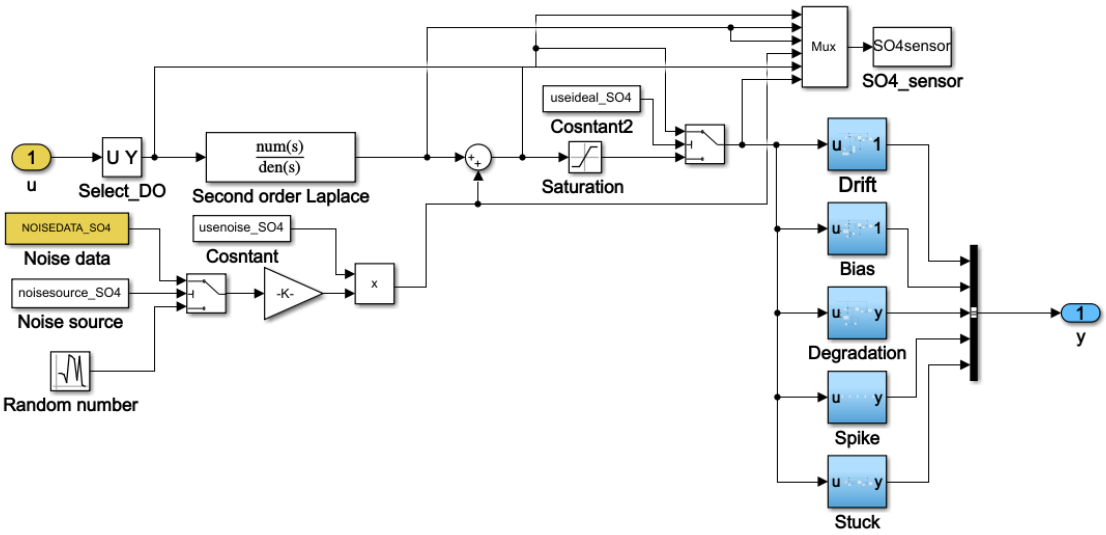



FIGURE 5.7: Faults implemented in the DO sensor: drift, bias, PD, spike and stuck.

### 5.3.1 Fault Injection: First Scenario

In the first scenario, the faults were injected into the DO sensor in the following order: drift, bias, PD, spike, and stuck. The Table 5.2 shows information about the start and duration of these faults. Its graphical representation can be seen in Figure 5.8. To facilitate the visualization, Figures 5.9–5.13 shows the faults individually.

TABLE 5.2: First scenario: faults introduced in the DO sensor signal.

| Fault | Start [day] | Duration [hours] |
|---|---|---|
| Drift | 10 | 120 |
| Bias | 30 | 120 |
| PD | 50 | 120 |
| Spike | 70, 72, 74, 76, 78 | 0.25 |
| Stuck | 90 | 120 |

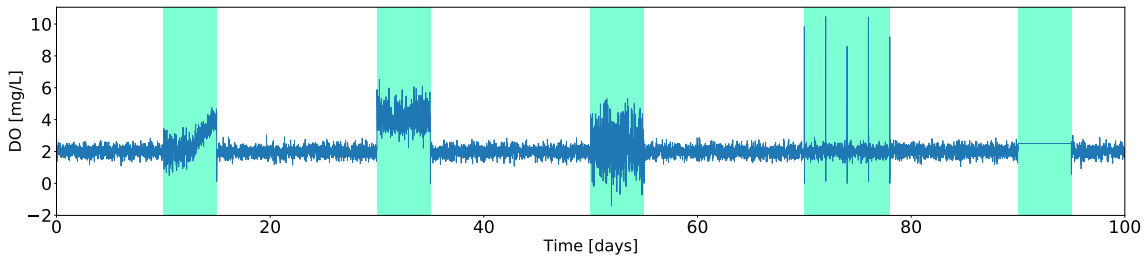



FIGURE 5.8: Faults implemented in the DO sensor signal in the first scenario.

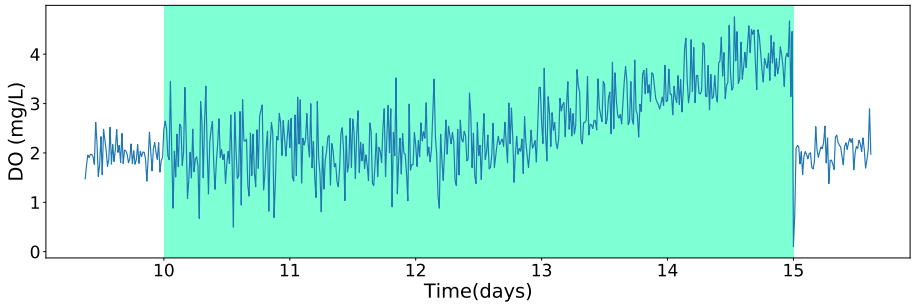



FIGURE 5.9: Faults implemented in the first scenario: drift.

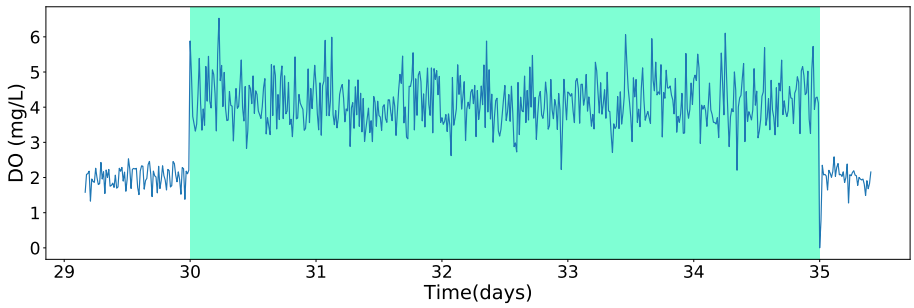



FIGURE 5.10: Faults implemented in the first scenario: bias.

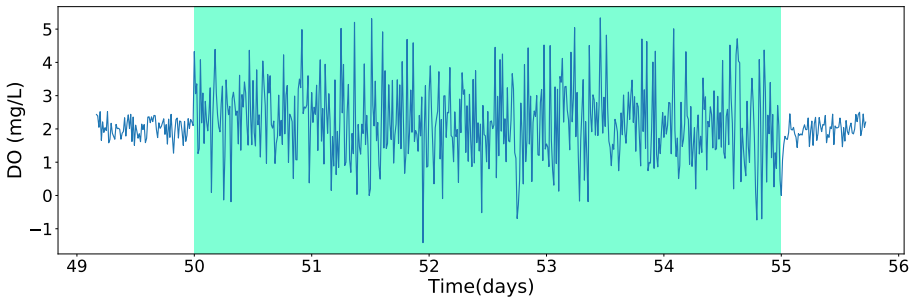



FIGURE 5.11: Faults implemented in the first scenario: PD.

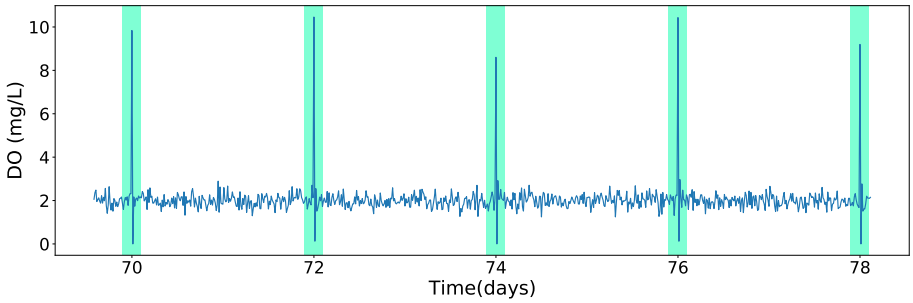



FIGURE 5.12: Faults implemented in the first scenario: spike.

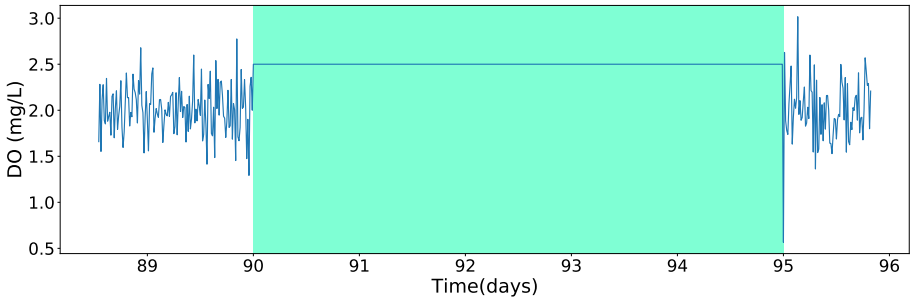



FIGURE 5.13: Faults implemented in the first scenario: stuck.

### 5.3.2 Fault Injection: Second Scenario

In the second scenario, the faults were injected into the DO sensor in the following order: PD, stuck, drift, spike, and bias. In addition to the change in the order of appearance, the duration of the faults was also modified. Table 5.3 shows information regarding the second scenario. Figure 5.14 shows the graphical representation of the faults and Figures 5.15–5.19 show the faults individually.

TABLE 5.3: Second scenario: faults introduced in the DO sensor signal.

| Fault | Start [day] | Duration [hours] |
|---|---|---|
| Drift | 40 | 72 |
| Bias | 92 | 48 |
| PD | 18 | 96 |
| Spike | 60, 62, 64, 66, 68 | 0.25 |
| Stuck | 30 | 72 |

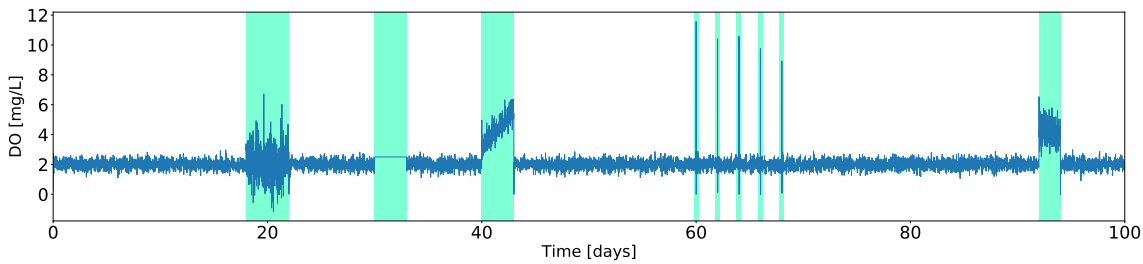



FIGURE 5.14: Second scenario. Faults implemented in the DO sensor signal. Changes implemented in the duration and order of appearance of failures.

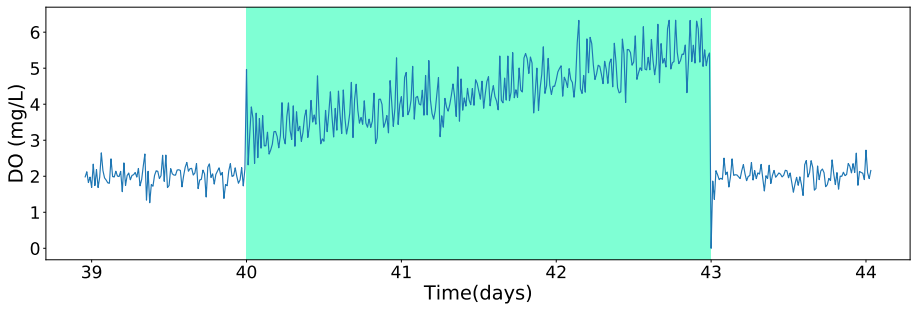



FIGURE 5.15: Faults implemented in the second scenario: drift.

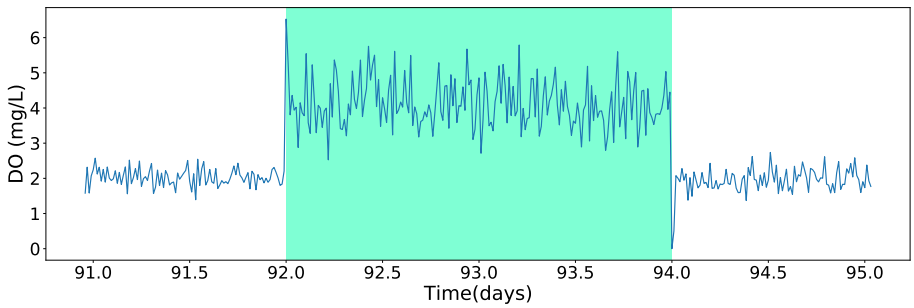



FIGURE 5.16: Faults implemented in the second scenario: bias.

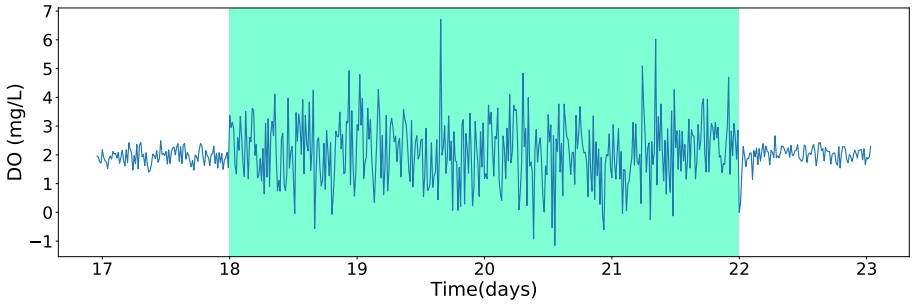



FIGURE 5.17: Faults implemented in the second scenario: PD.

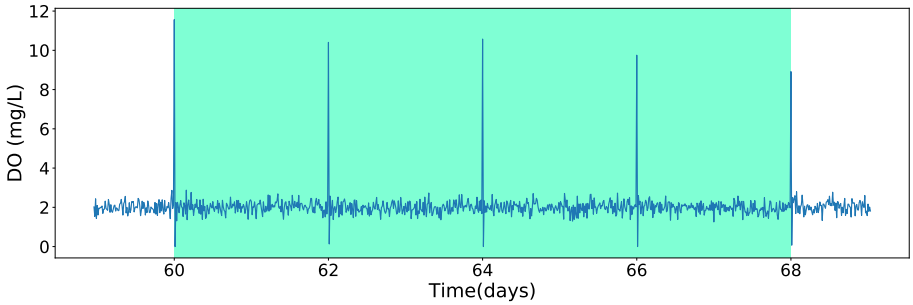



FIGURE 5.18: Faults implemented in the second scenario: spike.

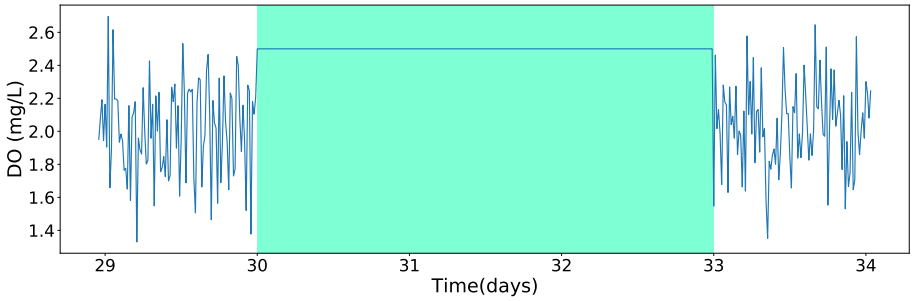



FIGURE 5.19: Faults implemented in the second scenario: stuck.

### 5.3.3 Fault Injection: Third Scenario

In the third scenario, the faults were injected into the DO sensor in the following order: stuck, spike, bias, drift, and PD. In addition to the change in the order of appearance, the duration and intensity of the faults was also modified. The information on the faults implemented in the third scenario can be seen in Table 5.4. The graphical representation of these faults can be seen in Figure 5.20, and their individual representations are represented in Figures 5.21–5.25.

TABLE 5.4: Third scenario: faults introduced in the signal from the DO sensor.

| Fault | Start [day] | Duration [hours] |
|---|---|---|
| Drift | 58 | 96 |
| Bias | 45 | 72 |
| PD | 90 | 72 |
| Spike | 26, 30, 31, 33, 38 | 0.25 |
| Stuck | 15 | 48 |

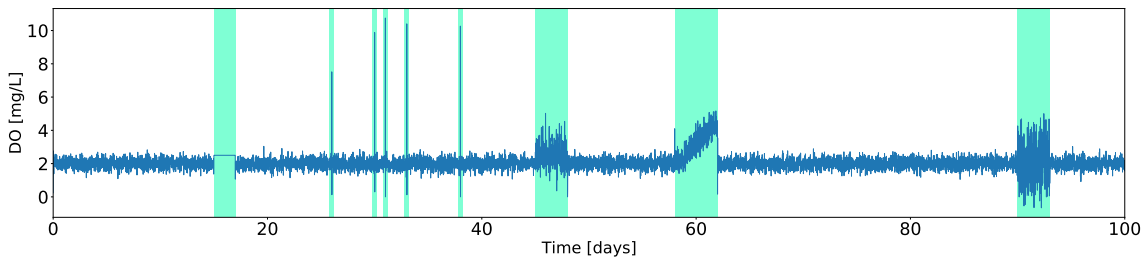



FIGURE 5.20: Third scenario. Modification in the order of appearance, duration and intensity of faults.

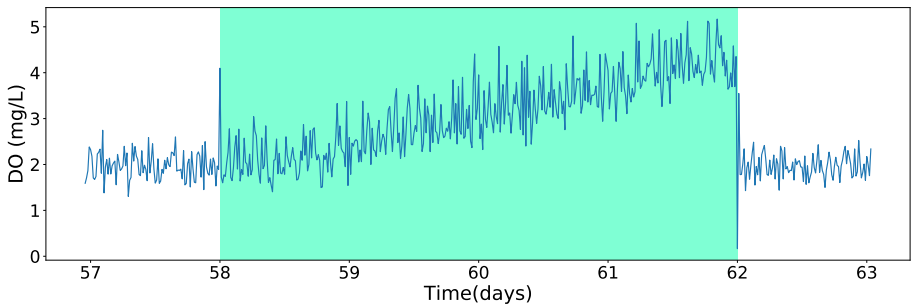



FIGURE 5.21: Faults implemented in the third scenario: drift.

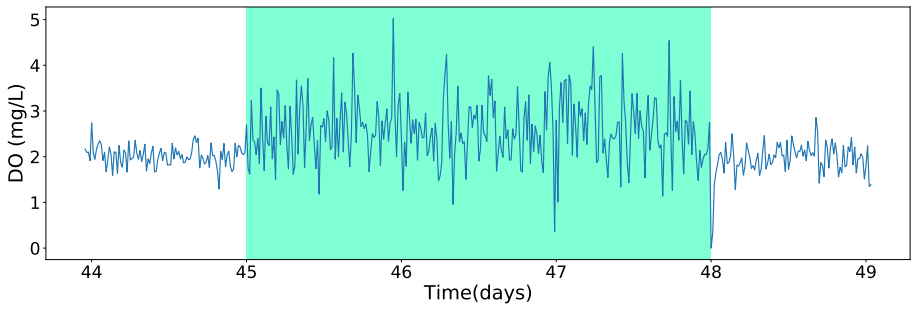



FIGURE 5.22: Faults implemented in the third scenario: bias.

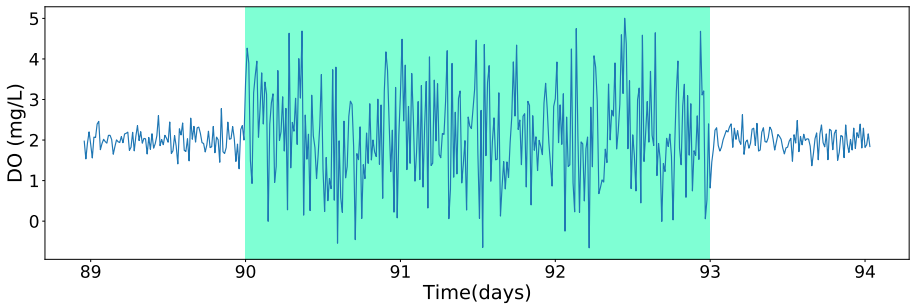



FIGURE 5.23: Faults implemented in the third scenario: pd.

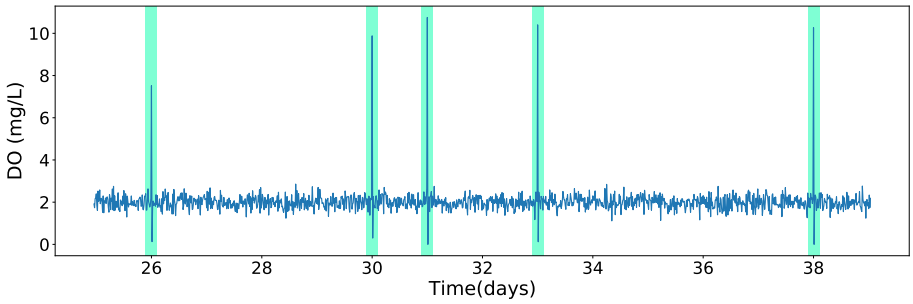



FIGURE 5.24: Faults implemented in the third scenario: spike.

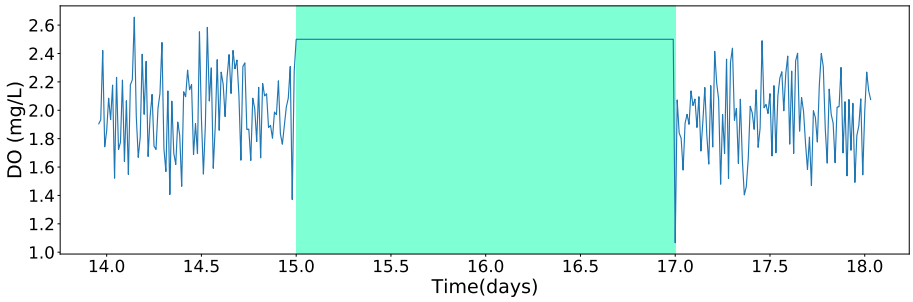



FIGURE 5.25: Faults implemented in the third scenario: stuck.

The faults implemented in the BSM2, and represented in the graphs, are the most common faults found in the sensor responsible for measuring the levels of DO in WWTPs [42]. In order to detect these faults, two autoencoders will be used: convolutional and LSTM autoencoders. The process of training and tuning hyperparameters as well as the results obtained will be detailed in the next section.

# Chapter 6

# Results and Discussion

This chapter brings the training process of the AE-based models (LSTM-AE and Convolutional-AE) presented in Chapter 4 for the detection of the failures presented in the Chapter 5, and the discussion of the obtained results.

The chapter is organized as follows: Section 6.1 and Section 6.2 present the process of training and validating the Convolutional-AE and LSTM-AE, respectively, with the choice of its hyperparameters and determination of its threshold. Section 6.3 presents the algorithms performance evaluation metrics. The results of Convolutional-AE and LSTM-AE models according to each of the three scenarios are described in Section 6.4.

As described in Chapter 5, the dataset, which represents the DO sensor signal, used for training and validating the AEs is equivalent to the last 100 days of a BSM2 operating cycle, that is, the period between days 509 and 609. Graphic representation of the signal of the DO sensor between these days can be seen in Figure 5.5. The sampling frequency is 15 minutes, totaling 9600 samples, which were divided, 70% for training and 30% for validation of Convolutional-AE and LSTM-AE models. All the work described in this section was developed in Python programming language, version 3.7, with the help of the Keras neural network package, version 2.8.0 .

## 6.1 Convolutional Autoencoder

In order to obtain the best Convolutional-AE model, a grid search were performed, with the evaluation of the combination of hyperparameters. The following combinations of hyperparameters were analyzed:

- Epochs $= [10, 20, 30, 40, 50]$;

- Batch size $= [32, 64, 128]$;
- AE layout : $[16, 32, 64, 128]$.

The best model was the one with the lowest mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|s_t - \hat{s}_t|, \tag{6.1}$$

where $s_t$ and $\hat{s}_t$ are the real and estimated DO values at the instant of time $t$, respectively, and $n$ is the number of samples used to validation.

The best models found, according to the MAE (6.1), were:

- Epochs = 20;
- Batch size = 128;
- AE layout = [32, 16, 16, 32].

The best Convolutional-AE, which obtained the lowest MAE, was used to detect the failures implemented in the BSM2 DO sensor signal. The distribution of the input signal reconstruction error, obtained by the Convolutional-AE training, can be seen in Figure 6.1. The maximum reconstruction error obtained, used as threshold, was 0.217238. If the difference between the predicted value and the actual value of a sample is greater than 0.217238, that sample is flagged as a failure.

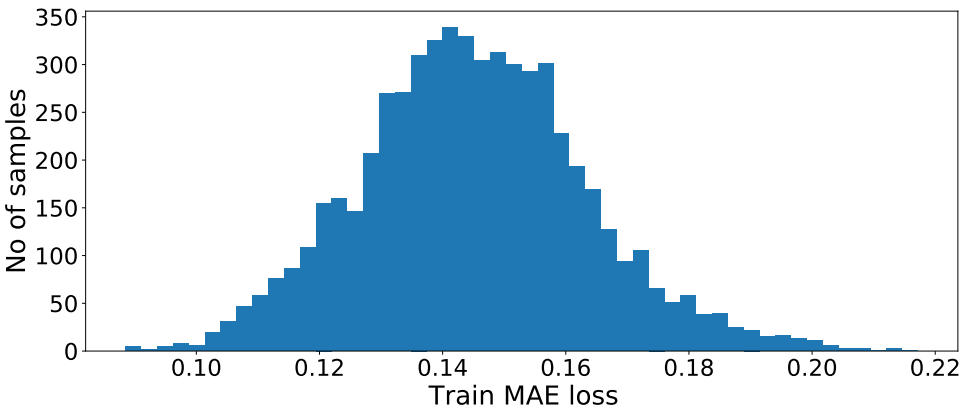



FIGURE 6.1: Signal reconstruction error. If the reconstruction error is greater than the threshold of 0.217238, a sample is considered a failure.

## 6.2 LSTM Autoencoder

To find the best LSTM-AE model, a grid search was performed considering the following hyperparameters:

- Epochs $= [10, 20, 30, 40, 50]$;
- Batch size $= [32, 64, 128]$;
- LSTM cells (AE layout) $= [16, 32, 64, 128]$.

The best model was chosen according to the MAE (6.1). The best model obtained has the following hyperparameters:

- Epochs $= 20$;
- Batch size $= 64$;
- LSTM cells (AE layout) $= [128, 64, 32, 32, 64, 128]$.

The distribution of the input signal reconstruction error, obtained by the best LSTM-AE, can be seen in Figure 6.2. The maximum reconstruction error obtained, adopted as the threshold, was 1.110521. The trained LSTM-AE will be used to detect the failures in the DO sensor signal in the three proposed scenarios. If the difference between the predicted value and the actual value of a sample is greater than 1.110521, that sample is flagged as failure.

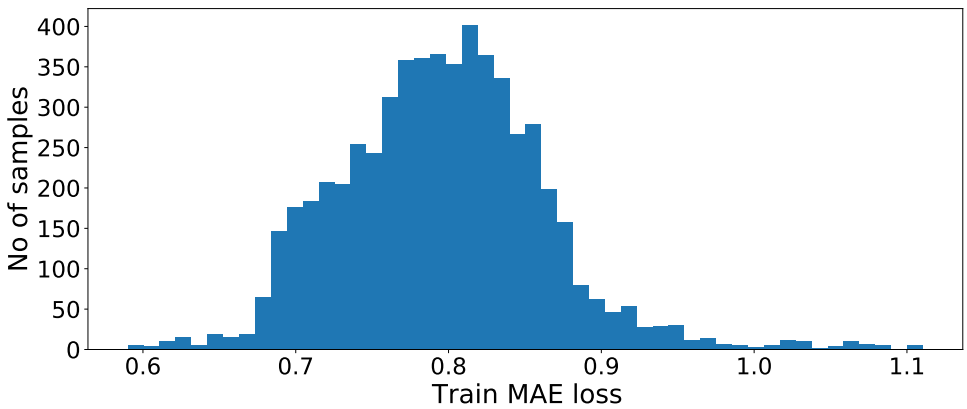



FIGURE 6.2: Signal reconstruction error. If the reconstruction error is greater than the threshold of 1.110521, a sample is considered a failure.

## 6.3 Evaluation Metrics

The purpose of the AE is to reconstruct the input signal. During its training, with the DO sensor data, the maximum value for the reconstruction error is adopted as threshold. In tests, a failure is identified if the difference between the real and estimated DO values is greater than the determined threshold.

The fault identification methods were evaluated as follows:

- If a sample is identified as faulty, within the fault duration period, it is classified as true positive (TP);
- If a sample is identified as a failure, outside the fault duration period, is classified as a false positive (FP);
- If a sample, within the failure duration time, is classified as normal, we have a false negative (FN);
- If a sample outside the fault duration period is identified as normal, it is classified as true negative (TN).

The evaluation metrics used for this study are TP rate ($TP_r$), FP rate ($FP_r$) and FN rate ($FN_r$) given by the Equations (6.2)-(6.4), respectively. The use of these metrics makes it possible to assess the reliability of the implemented error detection failure system.

$$TP_r = TP/(TP + FN) \tag{6.2}$$

$$FP_r = FP/(FP + TN) \tag{6.3}$$

$$FN_r = FN/(FN + TP) \tag{6.4}$$

## 6.4 Results

The faults drift, bias, PD, spike and stuck, injected into the dataset that represents the signal from the DO sensor characterizes three scenarios, explained in the Section 5.3, with variations in the order of appearance, duration and intensity of faults. In the following subsections, the results obtained by Convolutional-AE and LSTM-AE models will be presented.

### 6.4.1 First Scenario

In the first scenario, the faults appear in the following order: drift, bias, PD, spike and stuck. The results achieved by the Convolutional-AE and LSTM-AE can be seen below.

#### 6.4.1.1 Convolutional-AE

The Convolutional-AE has detected most of the proposed faults. Figure 6.3 condenses the results. A good performance can be seen, where the red circles represent the identified abnormal samples. Verifying in a more detailed way the performance of the Convolutional-AE, it can be seen that some samples, which represent faults, were not identified as faults. Figure 6.4 shows the model performance when identifying the fault bias. It is perceived that most of the faults were identified. The first sample representing the fault was not identified, which is equivalent to a delay of 15 minutes in the detection of the bias fault. There was also considerable delay in identifying the drift and stuck faults. The graphs with the individual results can be seen in the Annex B.1.1.

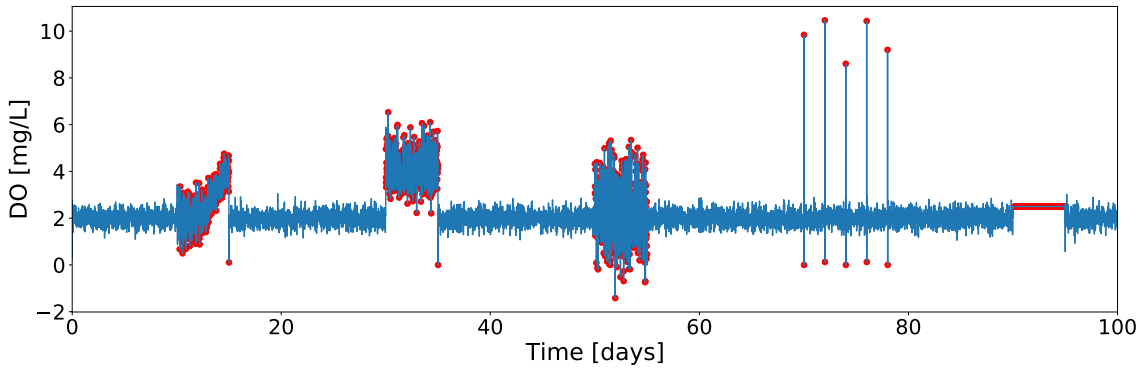



FIGURE 6.3: Result obtained by Convolutional-AE in the first scenario.

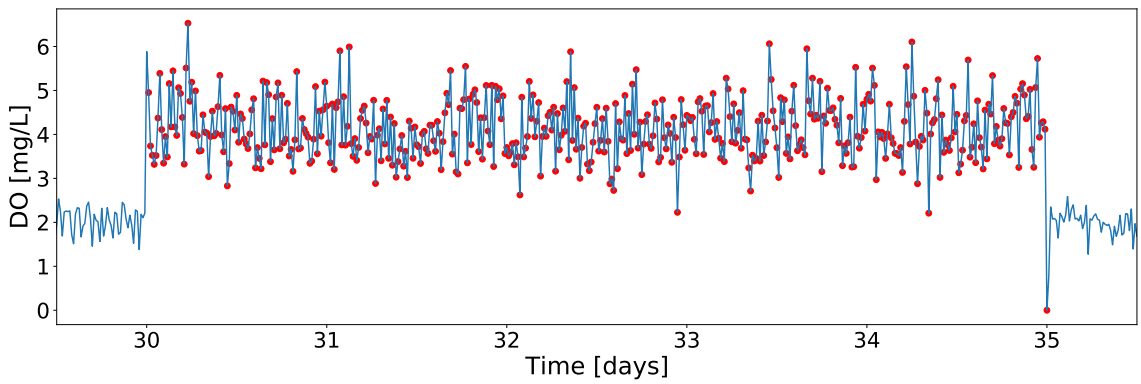



FIGURE 6.4: Convolutional-AE performance in identifying the bias fault. First scenario.

#### 6.4.1.2 LSTM-AE

The graph in Figure 6.5 shows the performance obtained by the LSTM-AE to detect the failures of the first scenario. The performance was very close to that obtained by Convolutional-AE. It is noticed that some samples of the spike failure were not identified as failures. In the graph of figure 6.6 it can be seen that the first sample of the bias fault was not identified, equivalent to a delay of 15 minutes in the detection of this fault. The performance of the LSTM-AE in detecting the failures of the first scenario can be seen in Annex B.2.1.

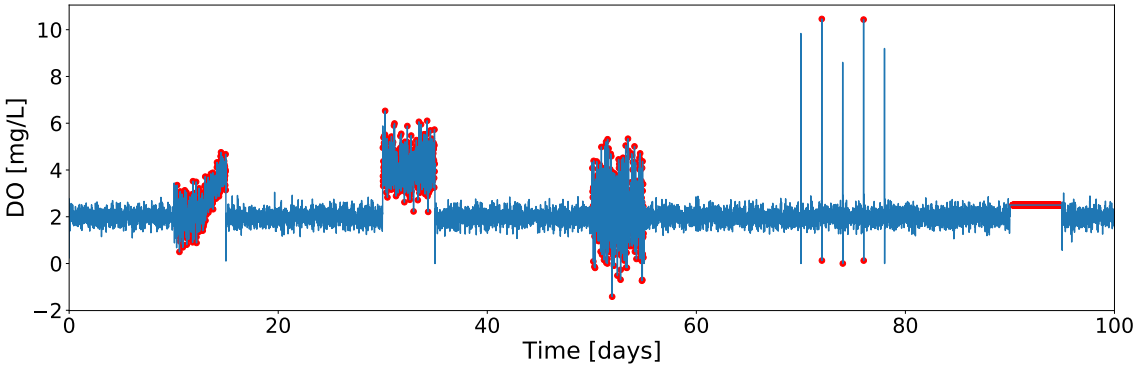



FIGURE 6.5: Result obtained by LSTM-AE in the first scenario.

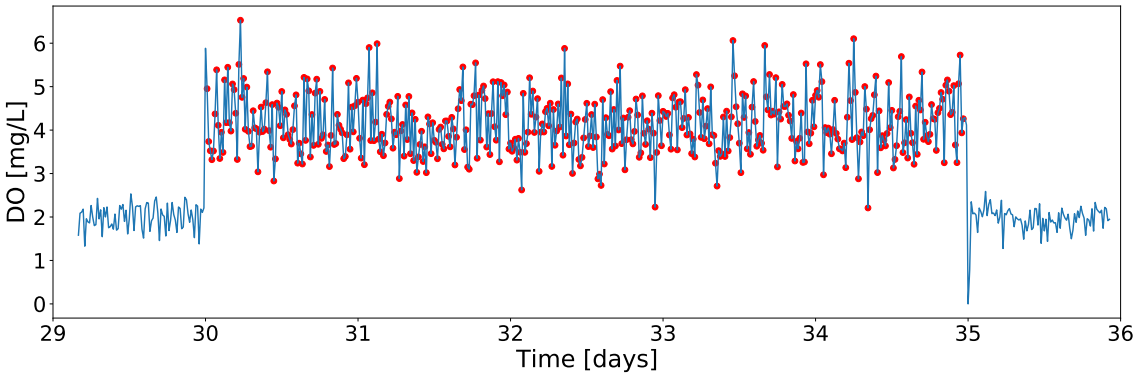



FIGURE 6.6: LSTM-AE performance in identifying the bias fault. First scenario.

### 6.4.2 Second Scenario

The faults appear in the second scenario in the following order: PD, stuck, drift, spike and bias. The results for this scenario can be seen below.

#### 6.4.2.1 Convolutional-AE

The Convolutional-AE was able to detect most of the anomalous samples. Figure 6.7 represents the performance of the model. Figure 6.8 shows a delay in detecting the PD

fault. The first 11 anomalous samples were not identified, which corresponds to a delay of 165 minutes in the detection of the PD fault. The performance of the Convolutional-AE in detecting the the faults implemented for the first scenario can be seen in Annex B.1.2.

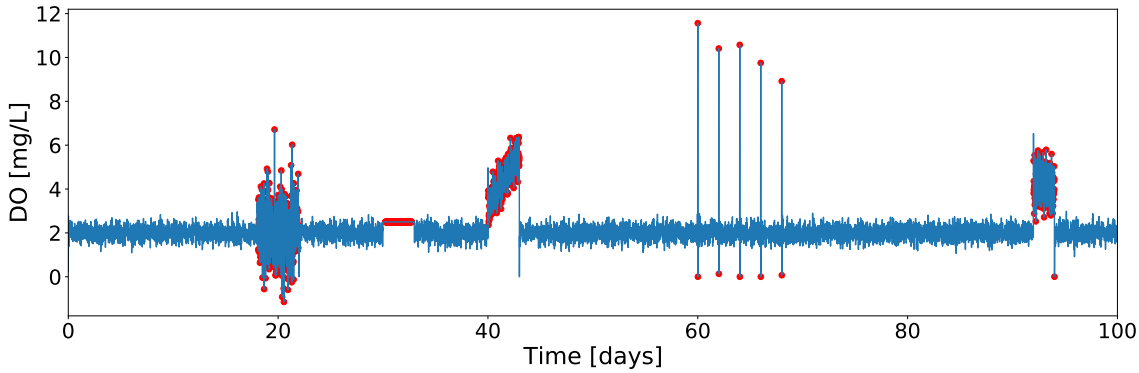



FIGURE 6.7: Result obtained by Convolutional-AE in the second scenario.

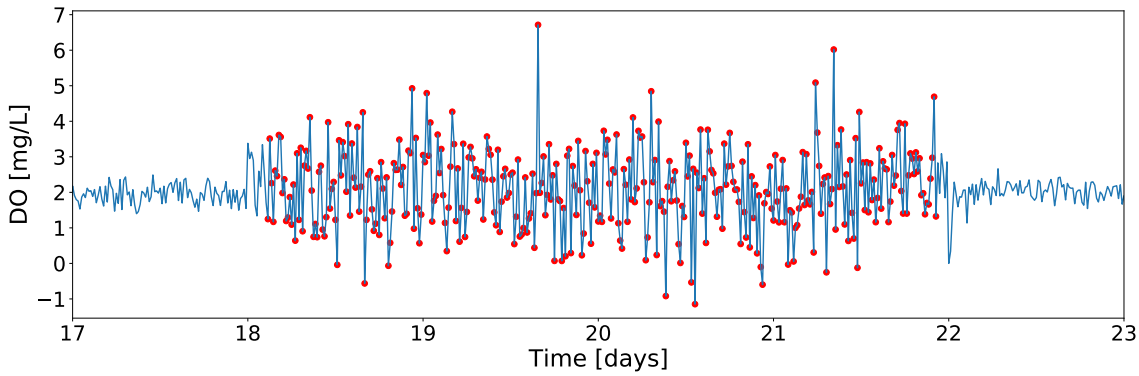



FIGURE 6.8: Convolutional-AE. PD fault identification

#### 6.4.2.2 LSTM-AE

The LSTM-AE was able to identify most of the proposed faults, but it presented difficulties in detecting the spike fault, and similarly to the Convolutional-AE, it presented delays in the identification of some faults. Figure 6.9 shows the general performance of LSTM-AE in the second scenario, and the problems faced in detecting the spike failure. A delay occurs in detecting the PD fault. The first 12 fault samples were not identified, which corresponds to a delay of 180 minutes. The result in the detection of PD failure can be seen in Figure 6.10. The graphs with the individual fault detection results can be seen in the Annex B.2.2.

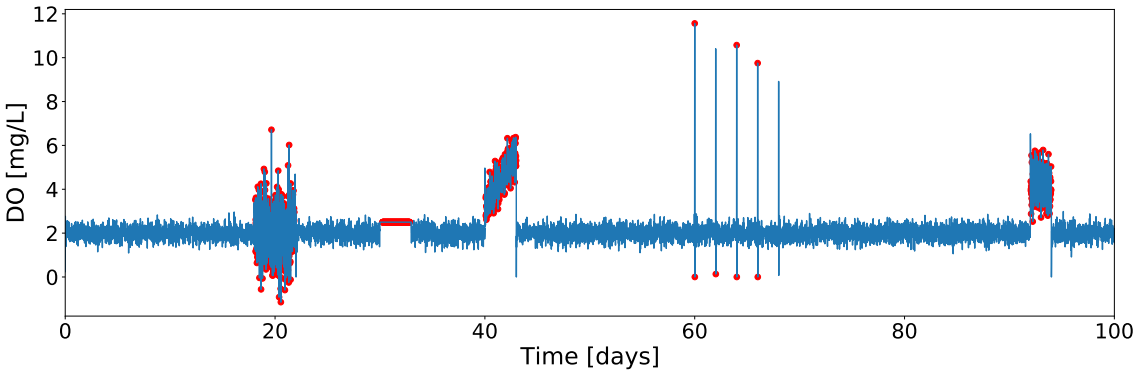



FIGURE 6.9: Result obtained by LSTM-AE in the second scenario.

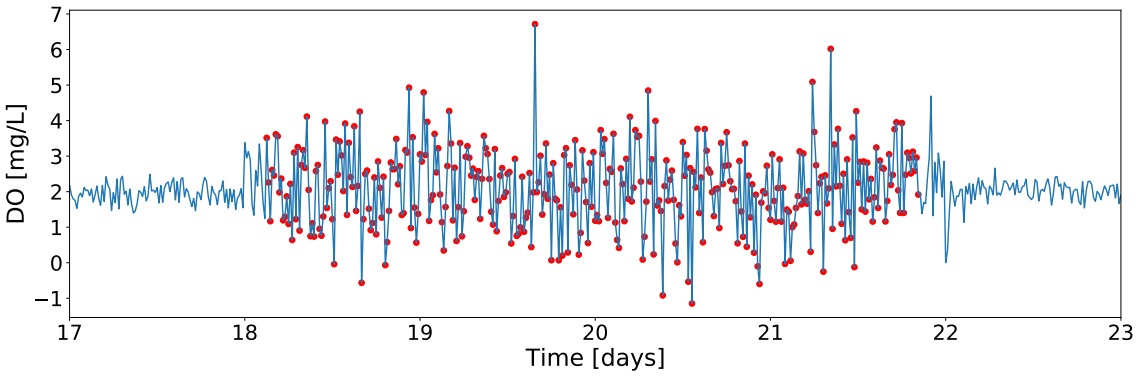



FIGURE 6.10: LSTM-AE. PD fault identification

### 6.4.3 Third Scenario

In the third proposed scenario, the failures appear in the following order: stuck, spike, bias, drift and PD. The results obtained by Convolutional-AE and LSTM-AE can be seen below.

#### 6.4.3.1 Convolutional-AE

The Convolutional-AE was able to identify most of the implemented faults, but it presented considerable delay in its identification, with special emphasis on the stuck fault. Convolutional-AE had problems related to false positives between the peaks of the spike failure, where some normal samples were identified as failure. Figure 6.11 shows the performance of the model, and Figure 6.12 highlights the problems related to false positives. The graphs with details of the failure detection by Convolutional-AE, in the third scenario, can be seen in the Annex B.1.3.

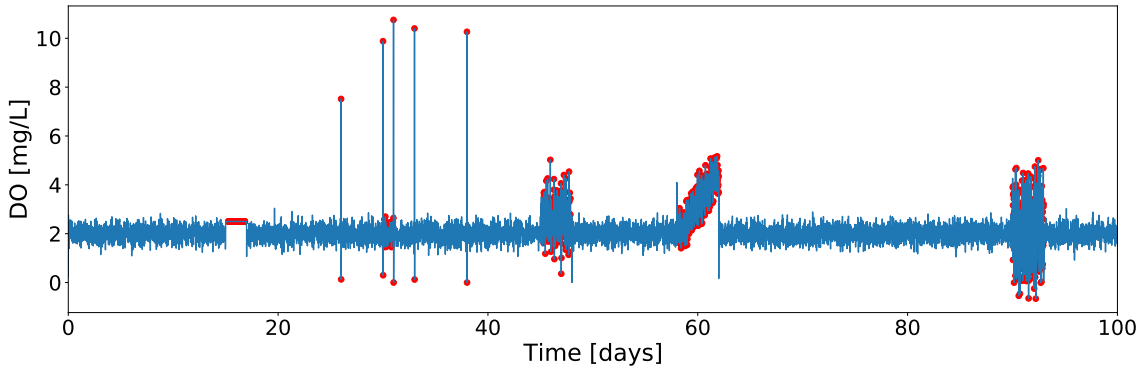



FIGURE 6.11: Result obtained by Convolutional-AE in the third scenario.

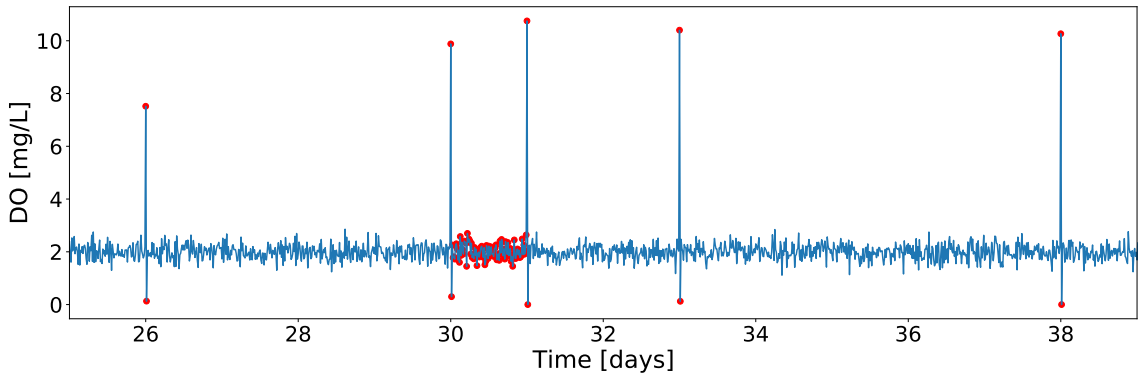



FIGURE 6.12: Identification of the spike fault.

#### 6.4.3.2 LSTM-AE

The LSTM-AE was able to detect most of the samples that correspond to the implemented faults, but it presented the same difficulties as the Convolutional-AE, identifying many normal samples as faults between the second and third peaks of the spike fault. Similar to Convolutional-AE, LSTM-AE showed considerable delay in the detection of implemented faults, with emphasis on the detection of bias and stuck faults. The overall performance can be seen in Figure 6.13, and the detail of the difficulty encountered in the spike failure can be seen in Figure 6.14. The graphs with the individual fault detection results can be seen in the Annex B.2.3.

The results obtained by the autoencoders, graphically represented in Figures 6.3 - 6.14, are summarized in Table 6.1, according to the metrics presented in Subsection 6.3. As the Convolutional-AE and LSTM-AE were used in three different scenarios, the results presented in Table 6.1 represent the arithmetic average of the results obtained in each scenario.

It is noticed that the two models are efficient in identifying the failures according

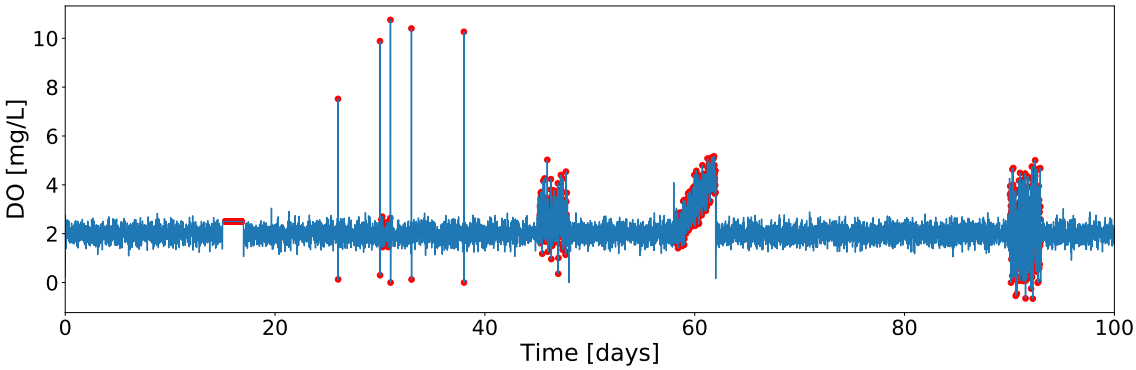



FIGURE 6.13: Result obtained by LSTM-AE in the third scenario.

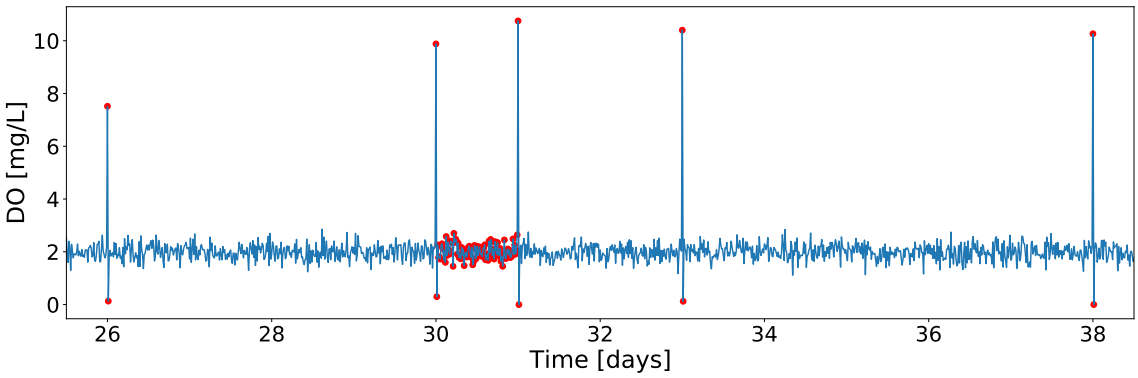



FIGURE 6.14: Identification of the spike fault. There are many false positives.

TABLE 6.1: Performance of LSTM-AE and Convolutional-AE. The values are the arithmetic average of the models' performance in the three scenarios. In bold the best results.

| **Autoencoder** | **Fault** | **Assessment metrics** | | | |
|---|---|---|---|---|---|
| | | **TP$_r$**[%] | **FP$_r$**[%] | **FN$_r$**[%] | **Delay** [h] |
| Convolutional-AE | Bias | **95.6** | 0 | **4.4** | **2.5** |
| | Drift | **95.15** | 0 | **4.85** | **4.41** |
| | PD | **97.8** | 0 | **2.2** | **1.08** |
| | Spike | **88.23** | 12.5 | **11.76** | 0 |
| | Stuck | **88** | 0 | **12** | **5.25** |
| LSTM-AE | Bias | 95 | 0 | 5 | 2.16 |
| | Drift | 94 | 0 | 6 | 5.33 |
| | PD | 94.1 | 0 | 4 | 1.83 |
| | Spike | 84.61 | 12.5 | 15.38 | 0 |
| | Stuck | 86.39 | 0 | 13.61 | 6.03 |

to the evaluation metrics. The bias, drift and PD faults were correctly identified by Convolutional-AE, with $TP_r$ of 95.6%, 95.15% and 97.8%, respectively. The LSTM-AE model presented a little lower performance with $TP_r$ of 95%, 94% and 94.1% for the same failures, respectively. Both algorithms presented greater difficulties with the spike failure. The Convolutional-AE and LSTM-AE showed correct identification of faults in 88.23% and 84.61%, respectively, and considerable $FN_r$, with values of 11.76% and 15.38%, respectively. Both models performed well on the Stuck fault. The Convolutional-AE presented 88%, and the LSTM-AE 86.39% for $TP_r$. But these two models showed considerable value for $FN_r$, with 12% for the Convolutional-AE, and 13.61% for the LSTM-AE model. As evidenced in the analysis of the graphs, there was an identification problem near the spike faults, in the third scenario considered. Outside the evaluation area, between the second and third spikes, the models identified 99 normal samples as faults.

The treatment process carried out by WWTPs presents a characteristic of slow changes. In the BSM2 simulator, the water retention time in the activated sludge tank (where the DO sensor takes its measurements) is 14 hours. Taking water retention time into account, both models had satisfactory results. The delay for fault detection, for each fault was calculated, as in the previous cases, by the arithmetic average of the delay detection times in the three proposed scenarios. The time that each algorithm took to identify the failures can be seen in Table 6.1. Related to the delay, it is observed that the Convolutional-AE obtained better results than the LSTM-AE, with the lowest average delay for fault detection. Only when detecting the bias fault, LSTM-AE obtain a better result, with 13.6% less delay in fault identification, in relation to Convolutional-AE. Both models readily identify the first peak of the spike fault. For the other faults, the Convolutional-AE performed better. It took 20.86%, 69.4% and 14.8% less time to identify drift, PD and stuck faults, respectively, when compared with LSTM-AE.

# Chapter 7

# Conclusion

Water scarcity is already a reality. Much of the world faces problems related to water supply. WWTPs play a fundamental role in facing this problem, and the work developed in this dissertation aims to make WWTPs more efficient. The early detection of failures in the operation of WWTPs is very important to ensure the quality of treated wastewater, protecting the environment, and consequently the human being, in addition to avoiding damage to the station structure, pollution and fines. The most popular methods for fault detection involve statistical analysis, which face many difficulties when applied to WWTP processes due to their non-linear characteristics. Artificial neural networks, in the form of AE, represent a good option because they take into account the non-linearities of the process. To facilitate tests and evaluations, without exposure to real risks, the BSM2 simulator was used. Due to its importance, the dissolved oxygen sensor, which monitors the oxygen levels in tank 4 of the biological reactor, was chosen for the simulation of failures that could occur in its operation. The most common faults in this type of sensor were implemented, and the AE used in their identification. The AE models had their hyperparameters chosen with the help of a grid search process, using the MAE metric to evaluate the reconstruction of the input signal. Convolutional-AE and LSTM-AE were used to detect bias, drift, spike, stuck and precision degradation faults.

For a more realistic evaluation of the AEs, the faults were injected into the dataset, composing three scenarios, with variation in the order of appearance, duration and intensity. The metrics for evaluating the performance of the AE in detecting failures were True Positive, False Positive and False Negative. Convolutional-AE obtained the best results, according to the considered metrics, and a shorter delay in detecting failures.

## 7.1 Future Work

In this work, two methods based on EA were used, Convolutional-AE and LSTM-AE, but as shown in the literature review, there are several models that have not yet been used in problems related to WWTPs. The use of other AE models, the exploration of hyperparameters, and the use of AEs together with other machine learning models can be interesting and help in the development of safer failure detection systems.

# Appendix A

# Paper presented at the 7th Workshop on Data Science for Social Good - SoGood 2022

# Fault Detection in Wastewater Treatment Plants: Application of Autoencoders with Streaming Data

Rodrigo Salles[1,2],
Jérôme Mendes[2], Rita P. Ribeiro[1,3], and João Gama[3,4]

[1] Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal.
[2] Institute of Systems and Robotics, University of Coimbra, 3030-290 Coimbra, Portugal
[3] INESC TEC, 4200-465 Porto, Portugal.
[4] Faculty of Economics, University of Porto, 4200-464 Porto. Portugal

**Abstract.** Water is a fundamental human resource and its scarcity is reflected in social, economic and environmental problems. Water used in human activities must be treated before reusing or returning to nature. This treatment takes place in wastewater treatment plants (WWTPs), which need to perform their functions with high quality, low cost, and reduced environmental impact. This paper aims to identify failures in real-time, using streaming data to provide the necessary preventive actions to minimize damage to WWTPs, heavy fines and, ultimately, environmental hazards. Convolutional and Long short-term memory (LSTM) autoencoders (AEs) were used to identify failures in the functioning of the dissolved oxygen sensor used in WWTPs. Five faults were considered (drift, bias, precision degradation, spike and stuck) in three different scenarios with variations in the appearance order, intensity and duration of the faults. The best performance, considering different model configurations, was achieved by Convolutional-AE.

**Keywords:** Wastewater Treatment Plant · Fault Detection · Autoencoder · BSM2

## 1 Introduction

Water is a strategic and fundamental resource for human beings. Activities carried out in the industry, agriculture, and services depend directly on access to water resources. And access to water is limited. Most of the water on the planet is in the seas and oceans (97%) [1]. There is only 3% of fresh water, but more than two-thirds is frozen in glaciers and polar ice [2]. The small fraction of fresh water remaining needs to serve more and more people. It is estimated that two-thirds of the world's population, 4 billion people, face water scarcity conditions at least one month a year, and approximately 500 million people live with water shortages throughout the year [3]. Water is a strategic resource and must be managed consciously. The used water must be treated so that we

can reused it. Wastewater may contain pollutants that pose risks to the environment and consequently to humans and need to be treated in appropriate places. Wastewater treatment plants (WWTPs) are structures that accelerate the treatment process in nature. The water used in human activities is sent to the WWTPs, which carry out the treatment in several stages, using chemical and physical processes. These structures are present in various parts of the world. In the United States of America, there are more than 16000 public administration WWTPs. Europe has more than 24000 treatment units. Brazil, Mexico, and China have 2820, 2540, and 1486 WWTPs, respectively [4].

WWTPs are important in dealing with water scarcity, but they must carry out their functions sustainably, with high quality and low cost. A monitoring system is needed to provide information from all stages of the treatment process so the necessary actions can be taken at the right time. With technological advances, new techniques were used to improve the functioning of WWTPs. The massive use of sensors in the monitoring of treatment plants generated a large amount of information and enabled the use of new control and optimization techniques. But the use of sensors also poses new problems. The actions taken by the control and optimization methods depend on the quality of the information provided by the sensors, and the quality of this information must be ensured. It is common for sensors to be exposed to extreme conditions at the monitoring site, as for example, temperature, vibrations, dust, chemical reagents, etc. It is of great importance that failures in these sensors are indicated as soon as possible, where undetected failures can represent damage to the structure of WWTPs, heavy fines, and environmental damage.

One of the main wastewater treatment phases occurs in the biological reactor. This reactor is composed of anoxic and oxygenated tanks, and it is the site of action of microorganisms that have the function of removing dangerous pollutants from wastewater. Oxygenated tanks need to maintain minimum oxygen levels. Lack of oxygen can result in the death of microorganisms, and excess oxygen represents a waste of energy spent on pumping. Considering that aeration is the most energy-intensive operation in wastewater treatment, amounting to 45–75% of plant energy costs [5], that of all the energy consumed in the world, approximately 3% is consumed in WWTPs [6], and that the energy spent on pumping depends on the information of the dissolved oxygen (DO) levels provided by the sensor, an efficient fault detection system is essential for the DO sensor. The main objective of this paper is to use autoencoder (AE) models to detect DO sensor failures in WWTPs. Early failure detection is important as it allows the necessary actions to be taken, benefiting higher safety, economy and quality of wastewater treatment. This work aims to assess the potential of AE in detecting failures in DO sensors, in WWTPs, in real-time, with streaming data. The simulator Benchmark Simulation Model nº2 (BSM2), which reproduces all phases of the treatment performed in a WWTP, was used to test the fault detection methodologies on the DO sensor in the biological reactor. This work analyzes the strengths and weaknesses of the Convolution-AE and LSTM-AE models, used to detect failures on DO sensors presented in WWTPs. The mod-

els are evaluated for the detection of five types of failures: bias, drift, precision degradation, spike and stuck. These faults were injected into the dataset, obtained with the help of BSM2, in three scenarios, with changes in the order of appearance, duration and intensity of the faults.

The rest of the paper is organized as follows: Section 2 brings a review of literature related to fault detection in WWTPs. Section 3 describes the simulator used and the case studies. Section 4 presents the structure of AE-based methodologies used to identify faults on DO sensors. Section 5 presents the experimental results, and compares the performance of Convolution-AE and LSTM-AE models in identifying the failures. Finally, in Section 6, the conclusion of the work is elaborated.

## 2 Related Work

Many works have already been proposed with the objective of detecting failures in sensors in WWTPs. The works can be divided into two large groups: failure detection using statistical methods, and failure detection using machine learning techniques. In [7], the use of artificial neural networks (ANNs) is proposed to identify six types of faults, one of which is the DO concentration sensor. The results proved a good ability of the ANN to recognize the faults, identifying 97% of case study failures. In [8], the authors use a Long short-term memory (LSTM) networks to identify collective failures in the sensors. The results obtained by the LSTM were compared to the results of the autoregressive integrated moving average (ARIMA), principal component analysis (PCA) and support-vector machines (SVM) models, and achieved the best performance, with a fault detection rate of over 92%. In [9], a radial basis function (RBF) neural network is used to identify faults in DO sensors by calculating the error limits. The proposed method obtained 0% false alarm, and a delay of 0.22 days in detection. In fault detection, unsupervised ANN can be trained to model a process by estimating the values of inputs and comparing the estimation to the actual values, also known as autoencoder. In [10], the authors used an AE, and the proposed model was used for detection of abrupt changes and drift in the sensor signal. The results showed that AE is capable of detecting sensor faults with good accuracy under different scenarios. In [11], a variational AE is used for fault detection. The proposed model takes into account the temporal evolution of the treatment process. The slow feature variational AE (SFAVAE) model is used to monitor processes and tries to identify faults such as sludge expansion fault and small magnitude variable step. Among the statistical methods the most used is the PCA. The PCA has many applications in WWTP, from direct fault detection [12] to data reconstruction [13]. In [14], the Incremental Principal Component Analysis (IPCA) method was used to identify several types of failures in WWTPs, one of them being failures in the DO sensor. The failures were injected into the dataset, and IPCA proved to be able to detect the failures and isolate the variable that originated the failure, with false alarm rate and missed detection rate of 0.07% and 18.53% respectively. A probabilistic PCA

approach in process monitoring and fault diagnosis with application in WWTP is proposed in [15]. The probabilistic PCA is compared to PCA, PPCA (probabilistic interpretation of the PCA), GPLVM (version of the PPCA for nonlinear situations) and Bayesian GPLVM (uses the Bayesian theory for training). The GPLVM and GPLVM models showed better performance in detecting failures in relation to the other models analyzed in the paper, in relation to the considered metrics. The major drawback of PCA for WWTP, is the assumption that process variables are linearly related to each other [16]. In the present work, models based on AE will be used. The case studies in which the models will be evaluated will be explained in the following section.

## 3 Case Studies

The water resulting from human activities, which carry pollutants, cannot be returned to the environment without undergoing treatment. This treatment occurs in WWTPs and is done in several stages: primary treatment (removes floatable and settleable solids), secondary treatment (secondary decantation and activated sludge), tertiary treatment (reuse of treated water), and sludge treatment (mechanical and biological treatments) [17]. Before implementing and evaluating new techniques in real treatment plants, simulators are commonly used. A widely used simulator of WWTPs is BSM2 [14,18]. Section 3.1 provides a brief description of the BSM2, and Section 3.2 presents the case studies with the description of the injected failures.

### 3.1 Benchmark Simulation Model No 2 - BSM2

The BSM2 is a simulation environment in which the plant layout, the simulation model, influence loads, test procedures and evaluation criteria are defined. For each of these items, compromises were pursued to combine plainness with realism and accepted standards [19]. BSM2 allows the implementation of several techniques and the manipulation of many parameters related to WWTPs [20, 21]. The influent dynamics are defined for 609 days, which takes into account rainfall effect, and temperature [19], with the data sampled every 15 minutes. The structure of BSM2 can be seen in the Fig. 1. The BSM2 was used to test the failure detection methods used in this work.

### 3.2 Faults in Dissolved Oxygen Sensor

BSM2 simulates the various stages of WWTPs. The biological reactor or activated sludge reactor can be seen in Fig. 1. Fig. 2 shows the biological reactor in more detail. It consists of five tanks, the first two being anoxic and the next three oxygenated. In oxygenated tanks, the DO level must be kept at 2 [mg/L], by a proportional-integral (PI) controller, which receives the DO values measured in Tank 4, compares it with the reference value, and drives the pumps responsible for maintaining the DO at the correct levels.

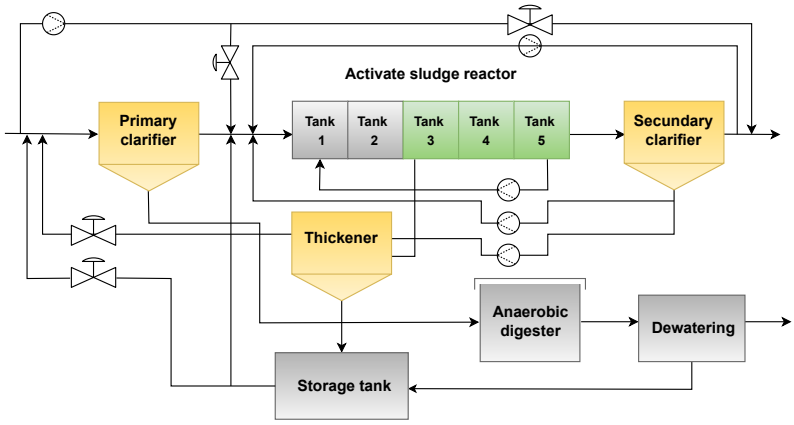



Fig. 1: Layout of Benchmark Simulation Model No 2 - BSM2.

The aeration system depends on the value measured by the DO sensor. A failure in this signal leads the system to malfunction. Through BSM2, we injected several faults. The objective is to identify the faults as early as possible. Also, in Fig. 2, it can be seen that the DO sensor performs its measurement in Tank 4 and it is used by the PI controller, and that the AE, represented in yellow, was positioned between the DO sensor and the PI controller, and has the function of detecting anomalous behavior in the measurement signal coming from the DO sensor.

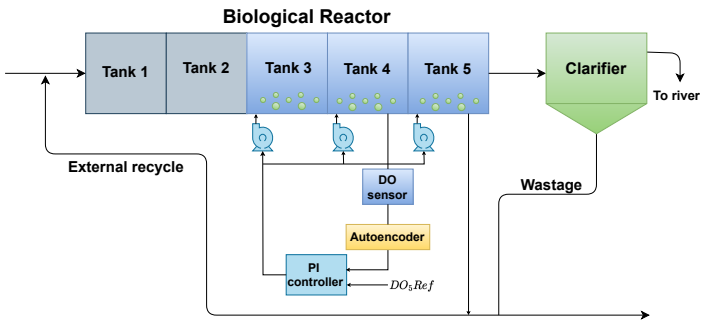



Fig. 2: Biological reactor details and case studies framework.

To evaluate the performance of the fault detection system, some faults were injected into the signal from the sensor. Deviations from expected behavior in the sensor output are considered faults, and they are classified according to the deviation from normal behavior. Let $s(t) = h(t) + \eta$ be the expected output of a sensor without the presence of faults, where $h(t)$ is the output of the sensor at time $t$, and $\eta \sim N(0, \delta_n^2)$ is noise, and $\delta_n^2$ is the noise variance [22]. The failures

considered in the present work are presented below [23, 24]. They are common failures in sensors caused by corrosion, calibration errors, presence of noise and physical damage presented on the WWTPs.

**Drift fault** When the sensor output increases at a constant rate, this type of fault is called as drift fault. A drift fault can be defined as

$$\begin{aligned} s(t) &= h(t) + \eta + b(t), \\ b(t) &= b(t-1) + v, \end{aligned} \tag{1}$$

where $b(t)$ is the bias added to the signal at time $t$, and $v$ is a constant. The $s(t)$ value increases linearly from the normal value over time.

**Bias fault** In a bias fault, a constant value $v$ is added to the sensor output and, as a consequence, a shift from the normal value is observed:

$$s(t) = h(t) + \eta + v. \tag{2}$$

**Precision degradation (PD) fault** This type of fault adds noise with a zero mean and high variance ($\delta_v^2$) to the output of a sensor:

$$s(t) = h(t) + \eta + v,\ v \sim N(0, \delta_v^2),\ \delta_v^2 \gg \delta_n^2. \tag{3}$$

**Spike fault** In spike faults, large amplitude peaks are observed at constant time intervals at the sensor output:

$$\begin{aligned} s(t) &= h(t) + \eta + v(t), \\ \forall t &= v \times \tau,\ h(t) + \eta, \\ \text{otherwise, } v &= \{1, 2, ...\},\ \tau \geq 2, \end{aligned} \tag{4}$$

where $\tau$ is the interval in which the spikes occur in the sensor output.

**Stuck fault** It is a complete failure, with the sensor output being locked at a fixed value $v$:

$$s(t) = v. \tag{5}$$

## 4 Fault Detection using Autoencoders

WWTPs exhibit marked nonlinear characteristics due to biochemical reactions and nitrification processes [25]. Thus, traditional statistical methods present difficulties in correctly identifying changes in the variables involved in the wastewater treatment process. In order to have a good performance in identifying the changes that may occur in WWTPs, the method used must be able

to deal with non-linearities. The AE present, among other characteristics, the ability to deal with non-linear processes.

AE is an unsupervised machine learning algorithm that aims to reconstruct its input signal. The generic AE model consists of three parts: encoder (responsible for reducing the dimensionality of the data), code (reduced representation of the encoder input data), and decoder (responsible for expanding the dimensionality represented in the code, and reconstructing the input signal). A representation of the AE can be seen in the Fig. 3.

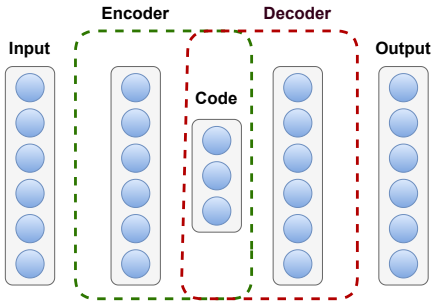



Fig. 3: Illustration of a generic AE model.

AE can have a simple structure, with only the code as the hidden layer, or can have several hidden layers. The representation of the input data, made by the code layer, can be classified as undercomplete or overcomplete. In undercomplete representation, the dimension of the representation of the input data by the code is smaller than the dimension of the input data, which forces the model to learn the most important characteristics of the input data. If the dimension of the code's input data representation is equal to the input dimension, the overcomplete representation, the model will just copy the input signal to the output without learning the most important characteristics of the input signal. The AE can be implemented as fully connected, convolution based or recurrent based units [26]. In this work, two models of AE will be used: LSTM-AE and Convolutional-AE. These models will be described in the next subsections.

### 4.1 LSTM Autoencoder

LSTM is a recurrent neural network that takes into account the historical context of events to make its predictions, with the help of memory cells. In an LSTM cell there are input, forget, memory and output gates.

In [27], the LSTM-AE is described as an extension to RNN based AE for learning the representation of time series sequential data. In this model, encoder and decoder are built using LSTM. The encoder LSTM receives a sequence of vectors that represents the signal from the DO sensor, and the decoder has the function of recreating the target sequence of input vectors in the

reverse order. This is the model used in the present work. The generic structure of a LSTM-AE can be seen in Fig. 4

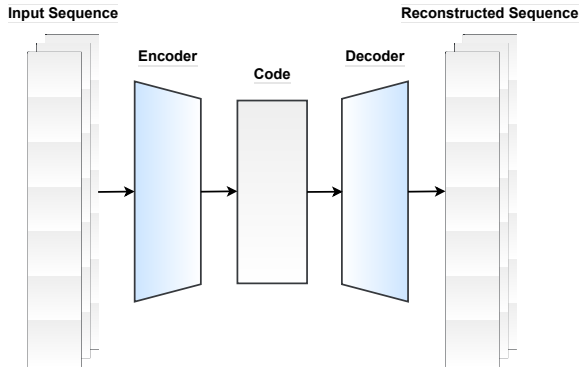



Fig. 4: Generic structure of the LSTM-AE.

### 4.2 Convolutional Autoencoder

Fully connected AEs ignore the spatial structure of the input signal, and this spatial structure can represent important information for the final reconstruction. To solve this problem, in [28] is proposed a model known as Convolutional-AE. Instead of using fully connected layers, Convolutional-AE use convolutional operators.

The Convolutional-AE is trained to reproduce the input signal from the DO sensor to the output layer. The signal passes through the encoder, composed of a convolution layer, which reduces the dimension of the representation of the input signal. In the decoder, composed of deconvolution layers, the compressed signal is reconstructed to obtain the original input signal, the DO sensor signal. The generic structure of a Convolutional-AE can be seen in Fig. 5.

The LSTM-AE and Convolutional-AE will be used to identify failures in these case studies. The experimental results will be described in Section 5.

## 5 Experimental Results

This section presents the results obtained by Convolutional-AE and LSTM-AE presented in Section 4, in detecting the faults described in Section 3.2. This section also describes the characteristics of injected faults and the evaluation metrics.

The dataset from the DO sensor, was separated into two equal parts, each part being equivalent to 100 days (9600 samples), keeping the temporal order. The first part was used to inject the faults described in Table 1 and represented

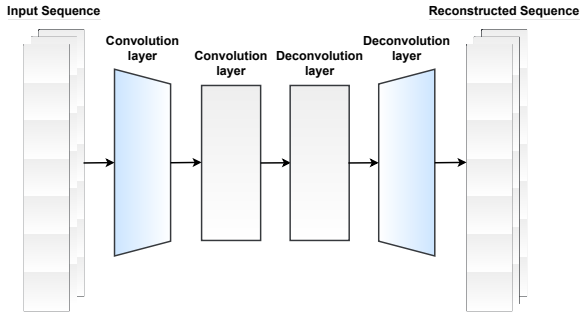



Fig. 5: Generic structure a Convolutional-AE.

in the graphs of Figure 6. From the second part, without failures, 70% of the data was used for training the Convolutional-AE and LSTM-AE, and the remaining 30% was used for evaluating the models, according to the input signal reconstruction error. In order to obtain the best model for each algorithm, a grid search were performed, with the evaluation of the combination of hyperparameters. All the work was developed in Python programming language, version 3.7, with the help of the Keras neural network package, version 2.8.0. The following combinations of hyperparameters were analyzed:

- Convolutional-AE: Epochs $= [10, 20, 30, 40, 50]$; Batch size $= [32, 64, 128]$; AE layout : $[16, 32, 64, 128]$.
- LSTM-AE: Epochs $= [10, 20, 30, 40, 50]$; Batch size $= [32, 64, 128]$; LSTM cells (AE layout) $= [16, 32, 64, 128]$.

The best model was the one with the lowest mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |s_t - \hat{s}_t|, \tag{6}$$

where $s_t$ and $\hat{s}_t$ are the real and estimated DO values at the instant of time $t$, and $n$ is the number of samples used to validation.

The best models found, according to the MAE, were:

- Convolutional-AE: Epochs = 20; Batch size = 128; AE layout = [32, 16, 16, 32]
- LSTM-AE: Epochs = 20; Batch size = 64; LSTM cells (AE layout) = [128, 64, 32, 32, 64, 128]

Table 1 presents the types of faults injected into the dataset, as well as their duration. To better evaluate the performance of the AE three scenarios were considered with variations in the order of appearance, duration and intensity of failures. Figure 6 depicts the signal from the DO sensor, after the faults described in Table 1. The objective of Convolutional-AE and LSTM-AE is to identify the faults that can be seen in Figure 6. The process of training AEs and choosing the best models will be described below.

Table 1: Three scenarios of faults injected in the signal obtained by the DO sensor: drift (Eq. 1), bias (Eq. 2), PD (Eq. 3), spike (Eq. 4) and stuck (Eq. 5). The faults in the scenarios have different order of appearance, duration and intensity.

| Fault | Start [day] | Duration [hours] |
|---|---|---|
| Drift | 10 | 120 |
| Bias | 30 | 120 |
| PD | 50 | 120 |
| Spike | 70, 72, 74, 76, 78 | 0.25 |
| Stuck | 90 | 120 |

(a) Scenario I

| Fault | Start [day] | Duration [hours] |
|---|---|---|
| Drift | 40 | 72 |
| Bias | 92 | 48 |
| PD | 18 | 96 |
| Spike | 60, 62, 64, 66, 68 | 0.25 |
| Stuck | 30 | 72 |

(b) Scenario II

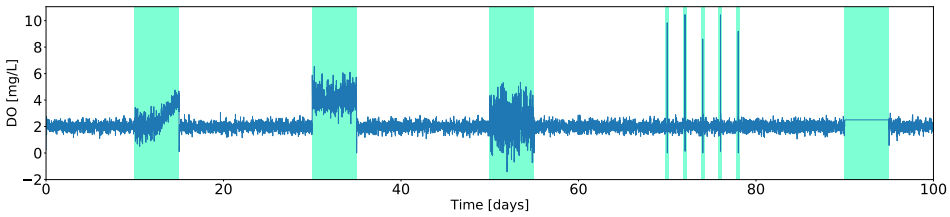
| Fault | Start [day] | Duration [hours] |
|---|---|---|
| Drift | 58 | 96 |
| Bias | 45 | 72 |
| PD | 90 | 72 |
| Spike | 26, 30, 31, 33, 38 | 0.25 |
| Stuck | 15 | 48 |

(c) Scenario III

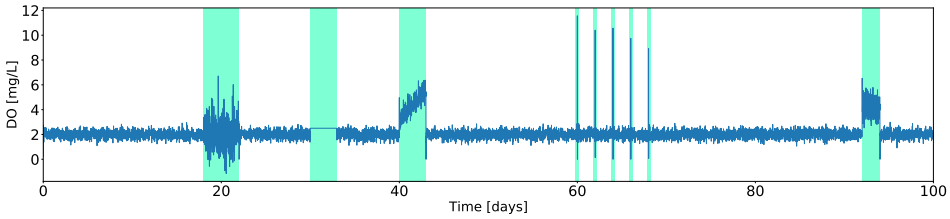
The purpose of the AE is to reconstruct the input signal. During its training, with the DO sensor data, the maximum value for the reconstruction error is adopted as a threshold. In tests, a failure is identified if the difference between the real and estimated DO values is greater than the determined threshold.

The fault identification methods were evaluated as follows:

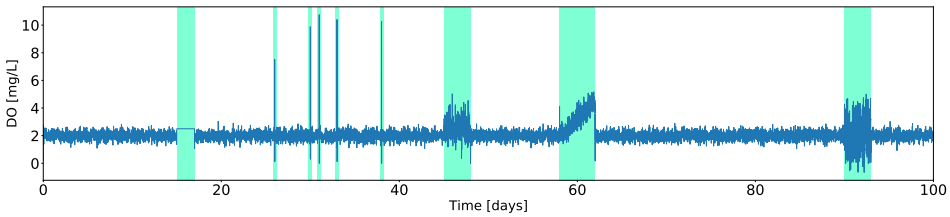
- if a sample is identified as faulty, within the fault duration period, it is classified as true positive (TP);
- if a sample is identified as a failure, outside the fault duration period, is classified as a false positive (FP);
- if a sample, within the failure duration time, is classified as normal, we have a false negative (FN);
- if a sample outside the fault duration period is identified as normal, it is classified as true negative (TN).




(a) Scenario I




(b) Scenario II




(c) Scenario III

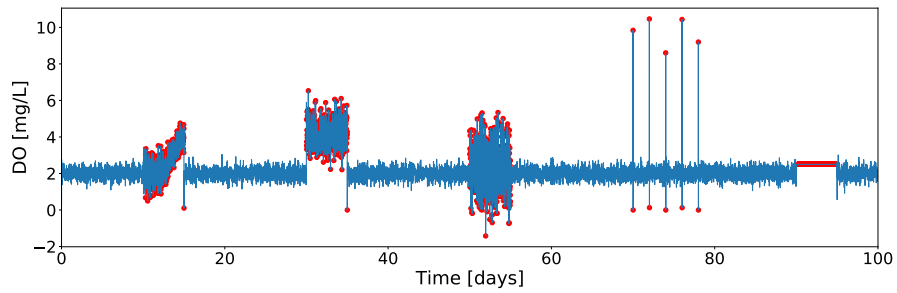
Fig. 6: Faults implemented in the signal obtained from the DO sensor. Variations in the order of appearance, duration and intensity of faults.

The evaluation metrics used was for this study are TP rate ($TP_r$), FP rate ($FP_r$) and FN rate ($FN_r$) - cf. Eq.7-9. The use of these metrics makes it possible to assess the reliability of the implemented error detection system.

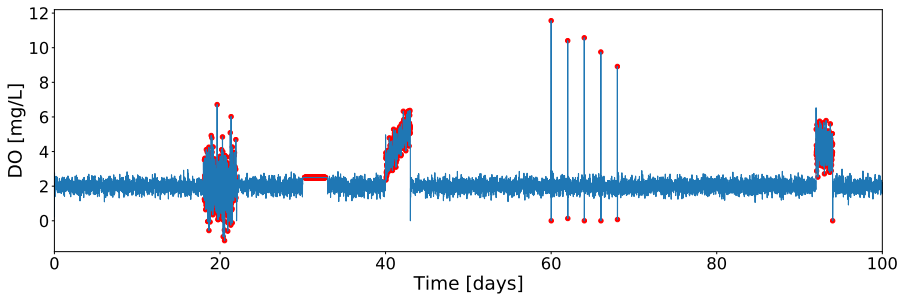
$$TP_r = TP/(TP + FN) \tag{7}$$
$$FP_r = FP/(FP + TN) \tag{8}$$
$$FN_r = FN/(FN + TP) \tag{9}$$

The results obtained by the models are graphically represented in Figures 7–8. Table 2 presents the performance of LSTM-AE and Convolutional-AE in identifying the present faults, where the values represent the arithmetic average of the algorithms' performance, for each fault implemented, in the three scenarios.

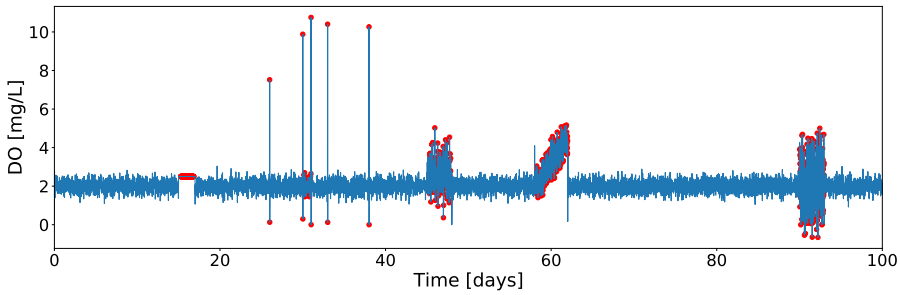
It is noticed that the two models are efficient in identifying the failures according to the evaluation metrics. The bias, drift and PD faults were correctly identified by Convolutional-AE, with $TP_r$ of 95.6%, 95.15% and 97.8%, respectively. The LSTM-AE model presented a little lower performance with $TP_r$ of




(a) Convolutional-AE applied to Scenario I.



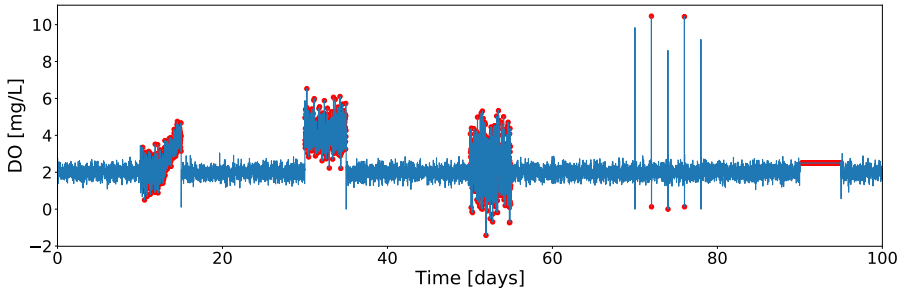

(b) Convolutional-AE applied to Scenario II.



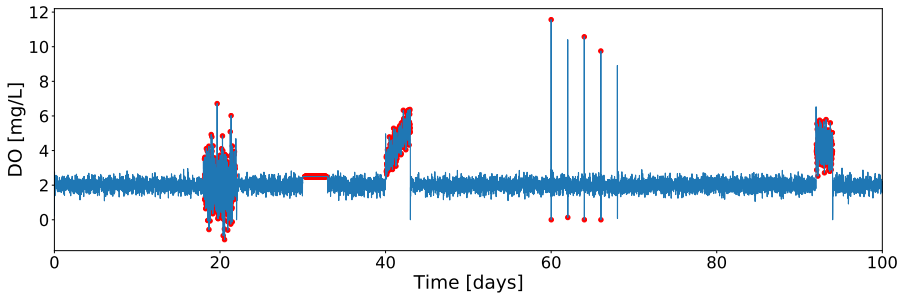

(c) Convolutional-AE applied to Scenario III.

Fig. 7: Faults identified in the three scenarios by the Convolutional-AE.

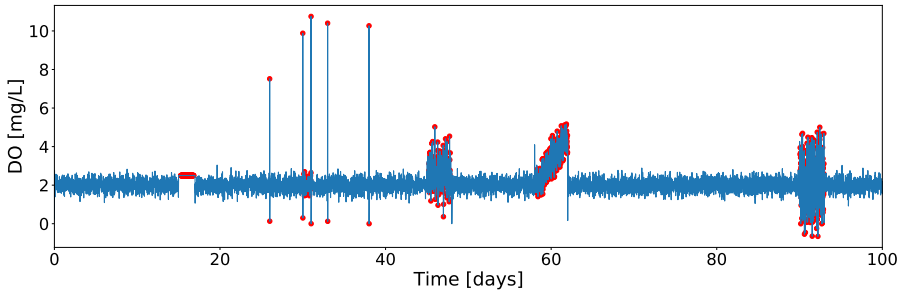
95%, 94% and 94.1% for the same failures, respectively. Both algorithms presented greater difficulties with the spike failure. The Convolutional-AE and LSTM-AE showed correct identification of faults in 88.23% and 84.61%, and considerable $FN_r$, with values of 11.76% and 15.38%, respectively. Both models performed well on the Stuck fault. The Convolutional-AE presented 88%, and the LSTM-AE 86.39% for $TP_r$. But these two models showed considerable value for $FN_r$, with 12% for the Convolutional-AE, and 13.61% for the LSTM-AE model. There was an identification problem near the spike faults, in the third scenario considered. Outside the evaluation area, between the second and third spikes, the models identified 99 normal samples as faults.




(a) LSTM-AE applied to Scenario I.




(b) LSTM-AE applied to Scenario II.




(c) LSTM-AE applied to Scenario III.

Fig. 8: Faults identified in the three scenarios by the LSTM-AE.

The treatment process carried out by WWTPs presents a characteristic of slow changes. In the BSM2 simulator, the water retention time in the activated sludge tank (where the DO sensor takes its measurements) is 14 hours. Taking water retention time into account, both models had satisfactory results. The delay for fault detection detection, for each fault was calculated, as in the previous cases, by the arithmetic average of the delay detection times in the three proposed scenarios. The time that each algorithm took to identify the failures can be seen in Table 2. Related to the delay, it is observed that the Convolutional-AE obtained better results than the LSTM-AE, with the lowest average delay for fault detection. Only when detecting the bias fault, LSTM-AE obtain a better result, with 13.6% less delay in fault identification, in relation to Convolutional-

Table 2: Performance of LSTM-AE and Convolutional-AE. The values are the arithmetic average of the models' performance in the three scenarios. In bold the best results.

| **Autoencoder** | **Fault** | **Assessment metrics** | | | |
|---|---|---|---|---|---|
| | | **TP$_r$**[%] | **FP$_r$**[%] | **FN$_r$**[%] | **Delay** [h] |
| Convolutional-AE | Bias | **95.6** | 0 | **4.4** | **2.5** |
| | Drift | **95.15** | 0 | **4.85** | **4.41** |
| | PD | **97.8** | 0 | **2.2** | **1.08** |
| | Spike | **88.23** | 12.5 | **11.76** | 0 |
| | Stuck | **88** | 0 | **12** | **5.25** |
| LSTM-AE | Bias | 95 | 0 | 5 | 2.16 |
| | Drift | 94 | 0 | 6 | 5.33 |
| | PD | 94.1 | 0 | 4 | 1.83 |
| | Spike | 84.61 | 12.5 | 15.38 | 0 |
| | Stuck | 86.39 | 0 | 13.61 | 6.03 |

AE. Both models readily identify the first peak of the skipe fault. For the other faults, the Convolutional-AE performed better. It took 20.86%, 69.4% and 14.8% less time to identify drift, PD and stuck faults, respectively, when compared with LSTM-AE.

## 6 Conclusions

WWTPs play a key role in dealing with the problem of water scarcity and thus alleviating the resulting economic and social problems. The work proposed in this paper helps to make WWTPs more secure and reliable. This paper proposed the application of AE for fault detection in DO sensors in biological reactors of WWTPs. Convolutional-AE and LSTM-AE were used to detect five types of faults: bias, drift, PD, spike and stuck. The models had their hyperparameters chosen with the help of the grid search process, using the MAE metric to evaluate the input signal reconstruction error. Three scenarios were considered, with variations in the order of appearance, duration and intensity of faults injected into the dataset. The best performance was obtained by the Convolutional-AE, with better detection values, according to the considered metrics, and less delay time when identifying faults. The analysis of other combinations for hyperparameters or the use in conjunction with other methods that allow less delay in fault detection would make the Convolutional-AE a good option to detect faults such as bias and drift in real WWTPs, representing an important contribution to its safety and sustainability.

## Acknowledgments

This research was supported by the ERDF and national funds through the project SYNAPPS (CENTRO-01-0247-FEDER-046978). We also acknowledge the support of the EC project CHIST-ERA-19-XAI-012, and project CHIST-ERA/0004/2019 funded by FCT.

## References

1. W. D. Haglund and M. H. Sorg, "Human remains in water environments," *Advances in forensic taphonomy: method, theory, and archaeological perspectives*, pp. 201–218, 2002.
2. C. Cassardo and J. A. A. Jones, "Managing water in a changing world," *Water*, vol. 3, no. 2, pp. 618–628, 2011.
3. M. M. Mekonnen and A. Y. Hoekstra, "Four billion people facing severe water scarcity," *Science advances*, vol. 2, no. 2, p. e1500323, 2016.
4. H. E. Macedo., B. Lehner, J. Nicell, G. Grill, J. Li, A. Limtong, and R. Shakya, "Distribution and characteristics of wastewater treatment plants within the global river network," *Earth System Science Data*, vol. 14, no. 2, pp. 559–577, 2022.
5. D. Rosso, L. L. E, and M. K. Stenstrom, "Aeration of large-scale municipal wastewater treatment plants: state of the art," *Water Science and Technology*, vol. 57, no. 7, pp. 973–978, 2008.
6. S. Nakkasunchi, N. J. Hewitt, C. Zoppi, and C. Brandoni, "A review of energy optimization modelling tools for the decarbonisation of wastewater treatment plants," *Journal of Cleaner Production*, vol. 279, p. 123811, 2021.
7. M. Miron, L. Frangu, S. Caraman, and L. Luca, "Artificial neural network approach for fault recognition in a wastewater treatment process," in *2018 22nd International Conference on System Theory, Control and Computing (ICSTCC)*, 2018, pp. 634–639.
8. B. Mamandipoor, M. Majd, S. Sheikhalishahi, C. Modena, and V. Osmani, "Monitoring and detecting faults in wastewater treatment plants using deep learning," *Environmental monitoring and assessment*, vol. 192, no. 2, pp. 1–12, 2020.
9. X. Li, W. Chai, T. Liu, and J. Qiao, "Fault detection of dissolved oxygen sensor in wastewater treatment plants," in *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, 2020, pp. 225–230.
10. H. Xiao, D. Huang, Y. Pan, Y. Liu, and K. Song, "Fault diagnosis and prognosis of wastewater processes with incomplete data by the auto-associative neural networks and arma model," *Chemometrics and Intelligent Laboratory Systems*, vol. 161, pp. 96–107, 2017.
11. K. Wang, P. Chang, and F. Meng, "Monitoring of wastewater treatment process based on slow feature analysis variational autoencoder," in *2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS)*. IEEE, 2021, pp. 495–502.
12. K. L. King, Z. Wang, and D. J. Kroll, "Classification of deviations in a process," Feb. 14 2006, uS Patent 6,999,898.
13. O. Schraa, B. Tole, and J. B. Copp, "Fault detection for control of wastewater treatment plants," *Water science and technology*, vol. 53, no. 4-5, pp. 375–382, 2006.
14. P. Kazemi, J. Giralt, C. Bengoa, A. Masoumian, and J.-P. Steyer, "Fault detection and diagnosis in water resource recovery facilities using incremental PCA," *Water Science and Technology*, vol. 82, no. 12, pp. 2711–2724, 08 2020.

15. B. Wang, Z. Li, Z. Dai, N. Lawrence, and X. Yan, "A probabilistic principal component analysis-based approach in process monitoring and fault diagnosis with application in wastewater treatment plant," *Applied Soft Computing*, vol. 82, p. 105527, 2019.
16. K. B. Newhart, R. W. Holloway, A. S. Hering, and T. Y. Cath, "Data-driven performance analyses of wastewater treatment plants: A review," *Water research*, vol. 157, pp. 498–513, 2019.
17. F. R. Spellman, *Handbook of water and wastewater treatment plant operations*. CRC press, 2003.
18. J. Mendes, R. Araújo, T. Matias, R. Seco, and C. Belchior, "Automatic extraction of the fuzzy control system by a hierarchical genetic algorithm," *Engineering Applications of Artificial Intelligence*, vol. 29, pp. 70–78, March 2014.
19. U. Jeppsson, M.-N. Pons, I. Nopens, J. Alex, J. Copp, K. Gernaey, C. Rosén, J.-P. Steyer, and P. Vanrolleghem, "Benchmark simulation model no 2: general protocol and exploratory case studies," *Water Science and Technology*, vol. 56, no. 8, pp. 67–78, 2007.
20. J. Mendes, N. Sousa, and R. Araújo, "Adaptive predictive control with recurrent fuzzy neural network for industrial processes," in *Proc. 16th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2011)*. Toulouse, France: IEEE, September 5-9 2011, pp. 1–8.
21. J. Mendes, R. Araújo, T. Matias, R. Seco, and C. Belchior, "Evolutionary learning of a fuzzy controller for industrial processes," in *Proc. of The 40th Annual Conference of the IEEE Industrial Electronics Society (IECON 2014)*. Dallas, TX, USA: IEEE, October 29 - November 1 2014, pp. 139–145.
22. S. U. Jan, Y. D. Lee, and I. S. Koo, "A distributed sensor-fault detection and diagnosis framework using machine learning," *Information Sciences*, vol. 547, pp. 777–796, 2021.
23. S. U. Jan and I. Koo, "A novel feature selection scheme and a diversified-input svm-based classifier for sensor fault classification," *Journal of Sensors*, vol. 2018, 2018.
24. S. Zidi, T. Moulahi, and B. Alaya, "Fault detection in wireless sensor networks through svm classifier," *IEEE Sensors Journal*, vol. 18, no. 1, pp. 340–347, 2017.
25. F. Li, Z. Su, and G. Wang, "An effective dynamic immune optimization control for the wastewater treatment process," *Environmental Science and Pollution Research*, pp. 1–16, 2021.
26. W. Pedrycz and S.-M. Chen, *Deep learning: Concepts and architectures*. Springer, 2020.
27. N. Srivastava, E. Mansimov, and R. Salakhudinov, "Unsupervised learning of video representations using lstm," in *International conference on machine learning*. PMLR, 2015, pp. 843–852.
28. J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International conference on artificial neural networks*. Springer, 2011, pp. 52–59.

# Appendix B

# Individual Representation of Detected Faults.

This appendix presents the detailed results of Convolutional-AE and LSTM-AE methodologies for the three scenarios.

## B.1 Convolutional-AE

This section presents the Convolutional-AE performance in the three proposed scenarios

### B.1.1 First Scenario

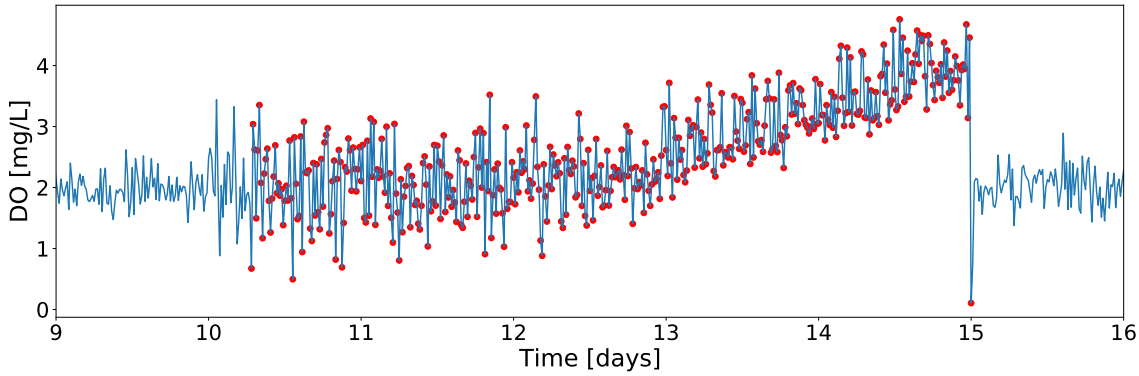



FIGURE B.1: Result obtained by Convolutional-AE - first scenario: drift.

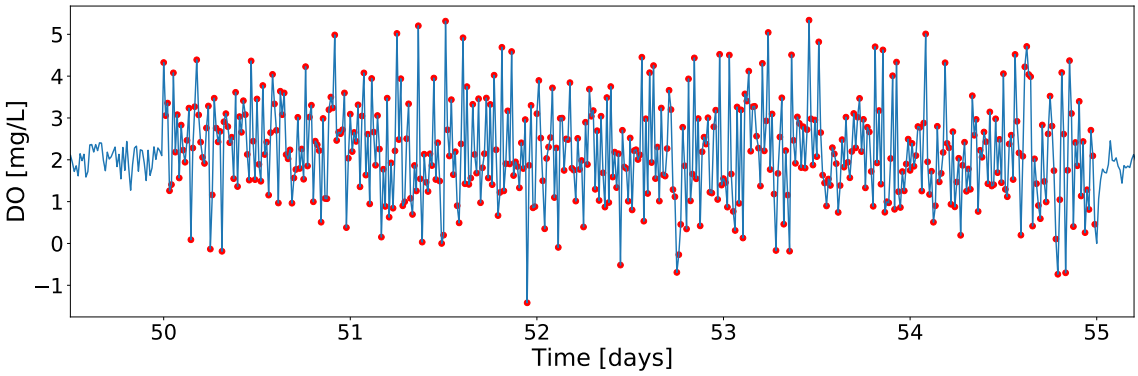



FIGURE B.2: Result obtained by Convolutional-AE - first scenario: PD.

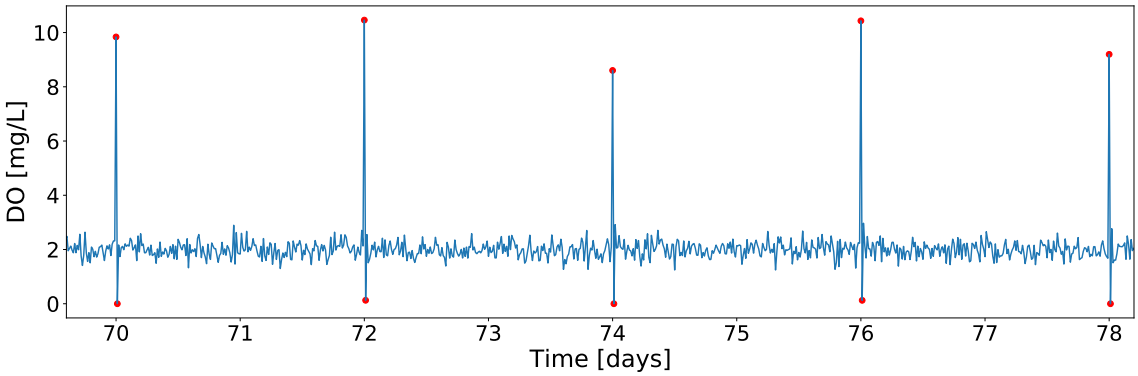



FIGURE B.3: Result obtained by Convolutional-AE - first scenario: spike.

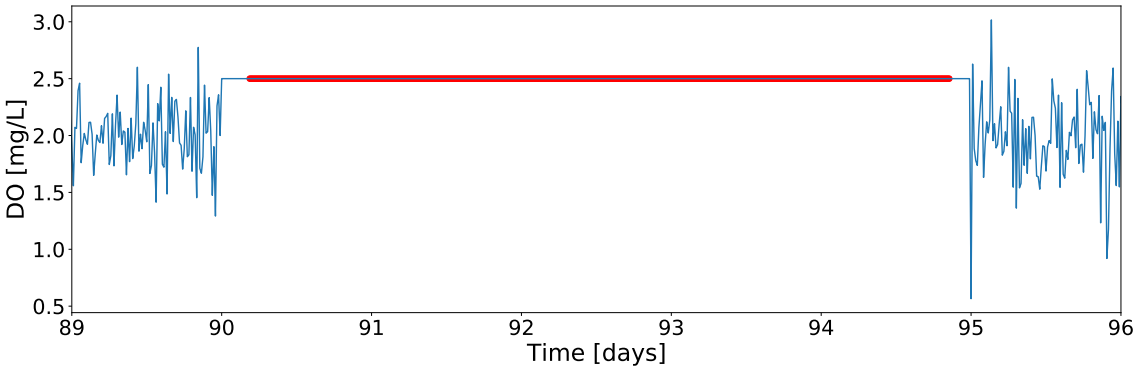



FIGURE B.4: Result obtained by Convolutional-AE - first scenario: stuck.

### B.1.2 Second Scenario

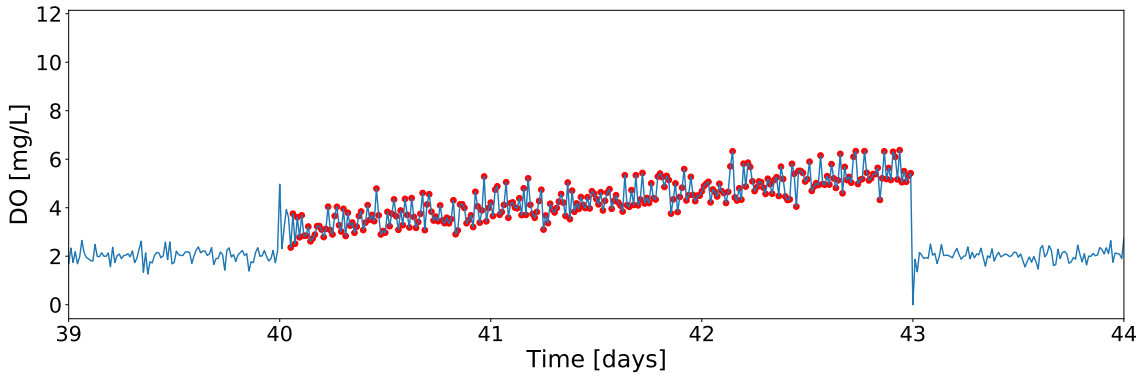



FIGURE B.5: Result obtained by Convolutional-AE - Second scenario: drift.

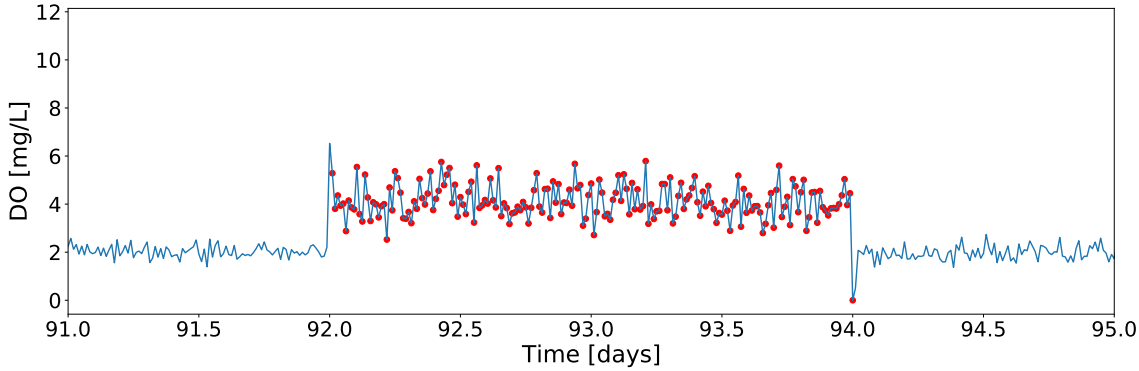



FIGURE B.6: Result obtained by Convolutional-AE - Second scenario: bias.

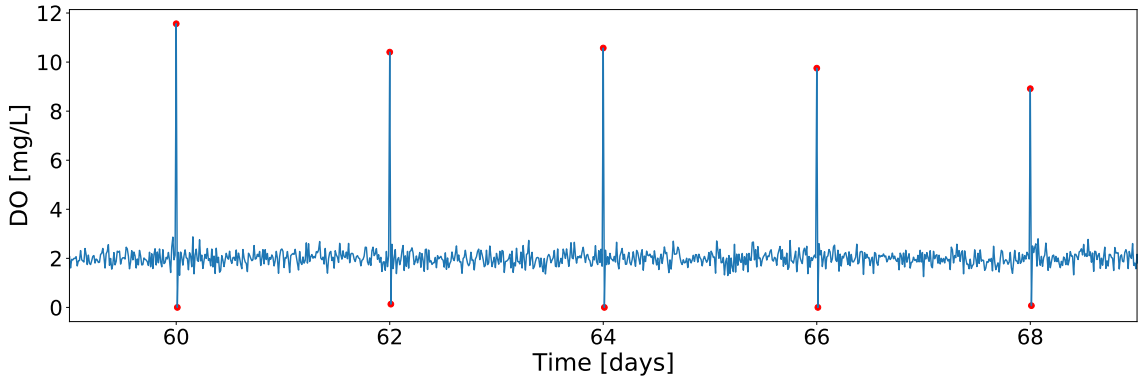



FIGURE B.7: Result obtained by Convolutional-AE - Second scenario: spike.

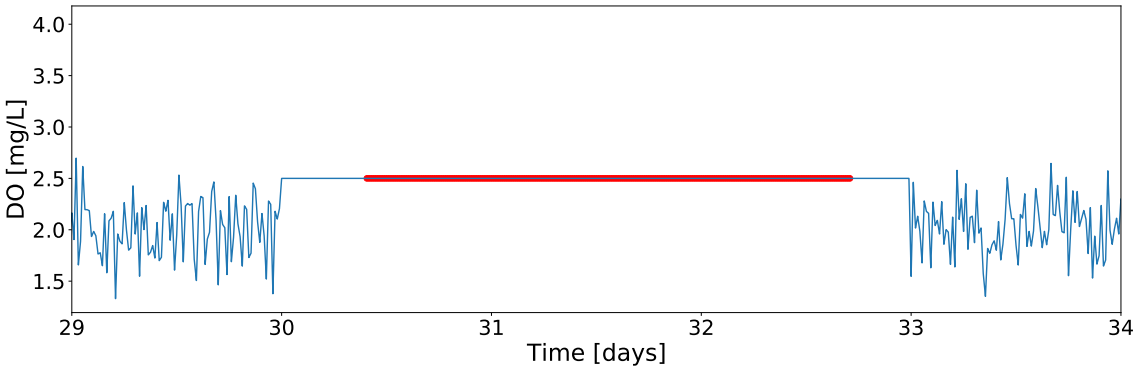



FIGURE B.8: Result obtained by Convolutional-AE - Second scenario: stuck.

### B.1.3 Third Scenario

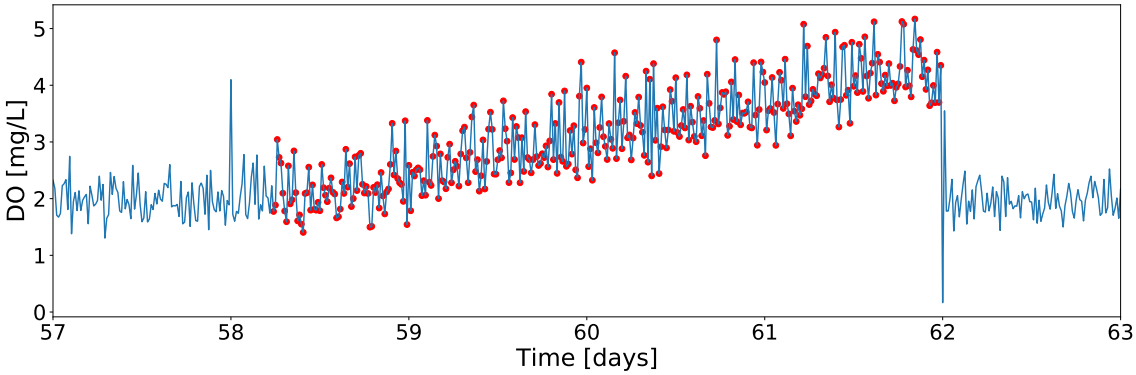



FIGURE B.9: Result obtained by Convolutional-AE - Third scenario: drift.

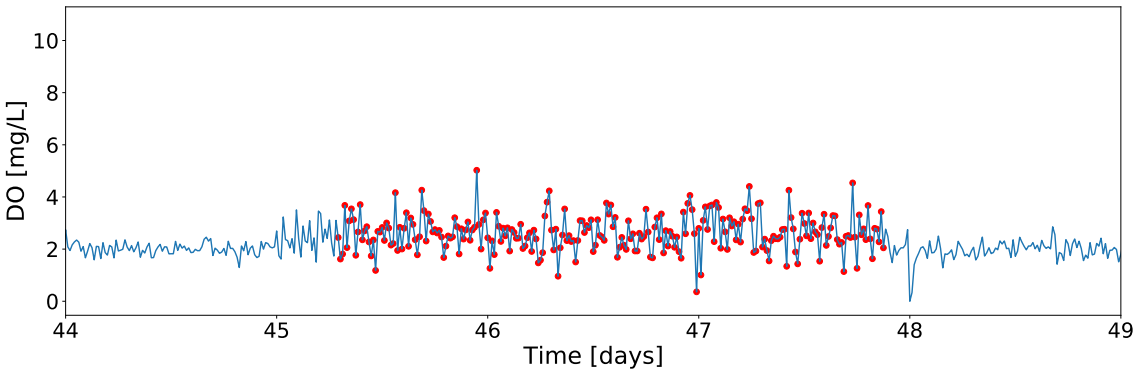



FIGURE B.10: Result obtained by Convolutional-AE - Third scenario: bias.

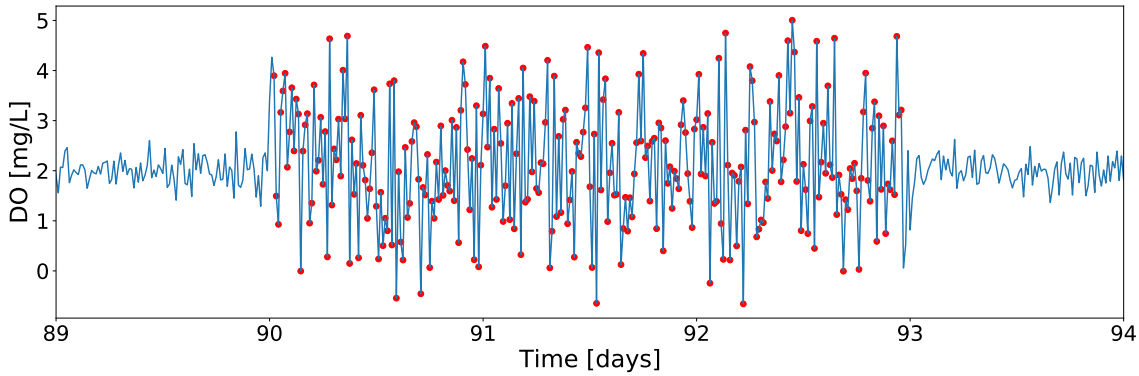



FIGURE B.11: Result obtained by Convolutional-AE - Third scenario: PD.

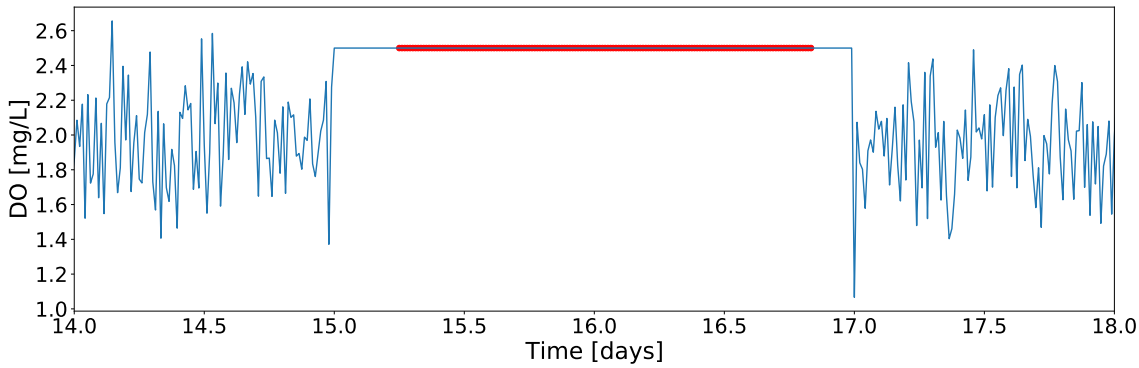



FIGURE B.12: Result obtained by Convolutional-AE - Third scenario: stuck.

## B.2 LSTM-AE

This section presents the LSTM-AE performance in the three proposed scenarios

### B.2.1 First Scenario

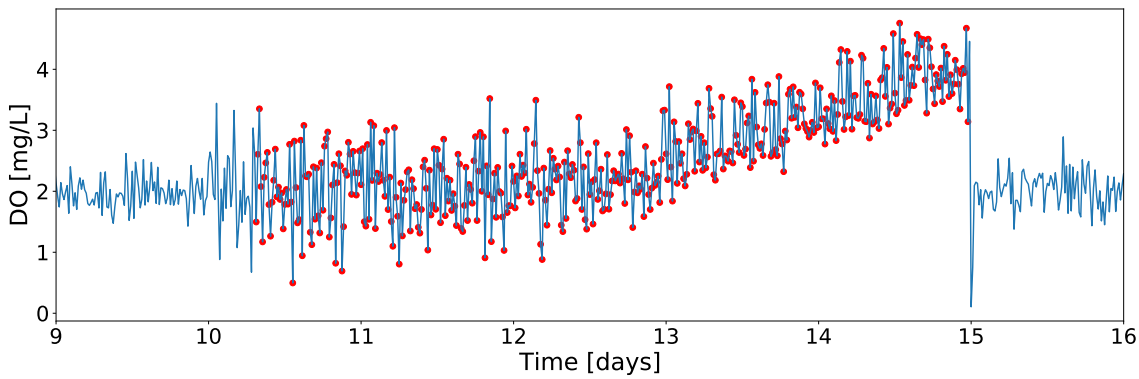



FIGURE B.13: Result obtained by LSTM-AE - first scenario: drift.

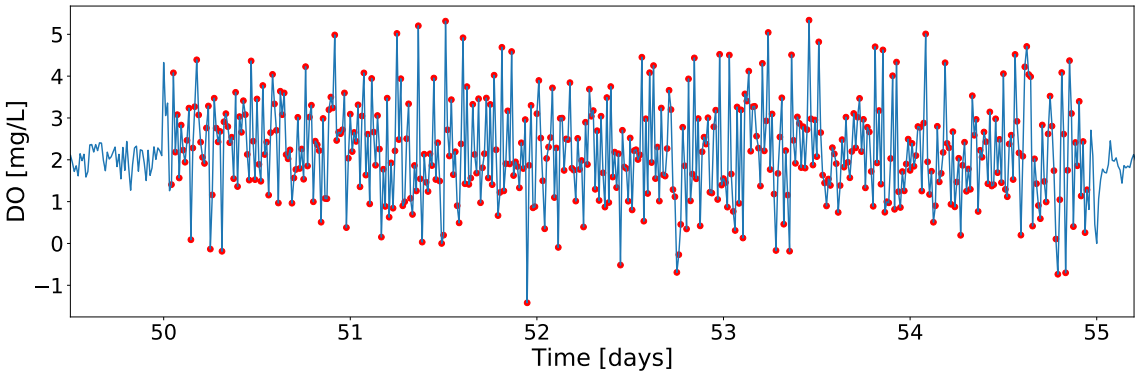



FIGURE B.14: Result obtained by LSTM-AE - first scenario: PD.

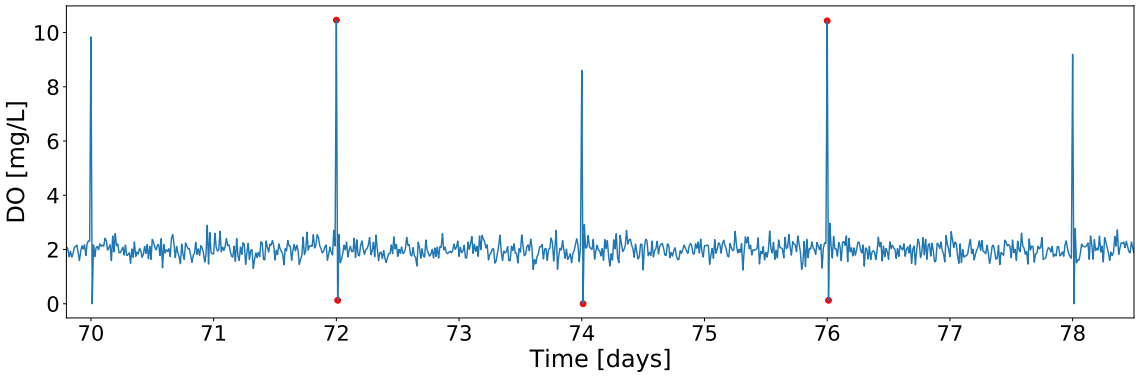



FIGURE B.15: Result obtained by LSTM-AE - first scenario: spike.

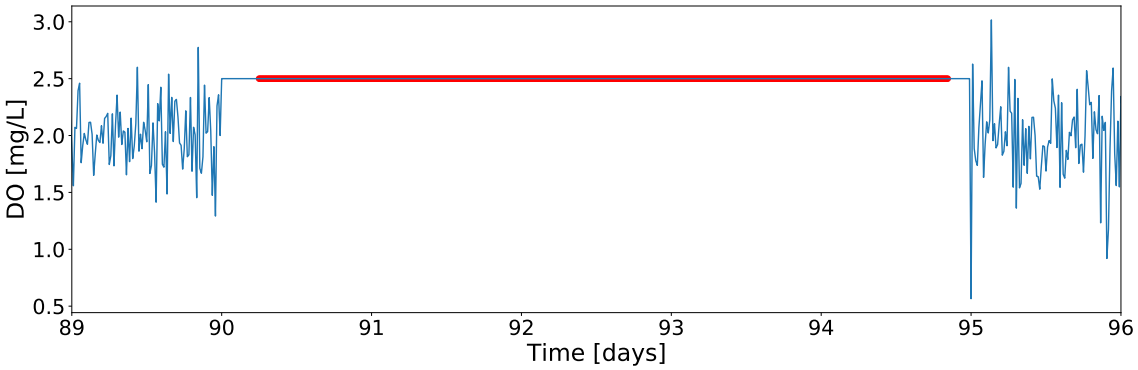



FIGURE B.16: Result obtained by LSTM-AE - first scenario: stuck.

### B.2.2 Second Scenario

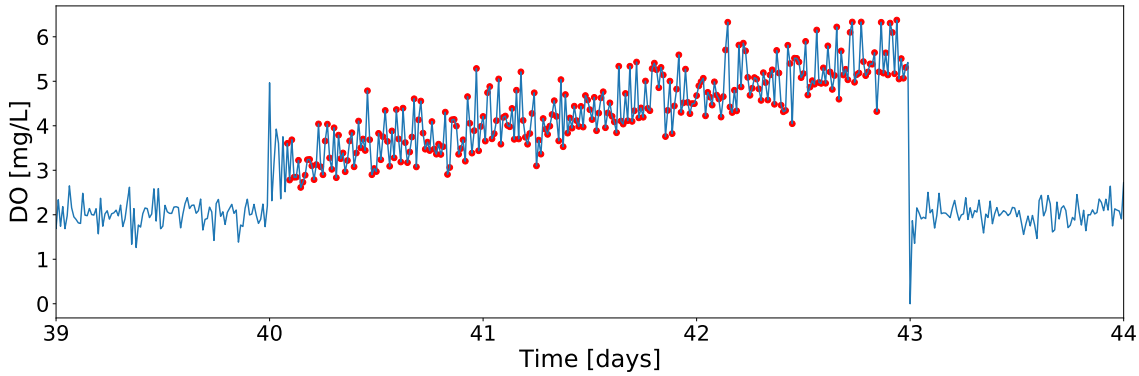



FIGURE B.17: Result obtained by LSTM-AE - Second scenario: drift.

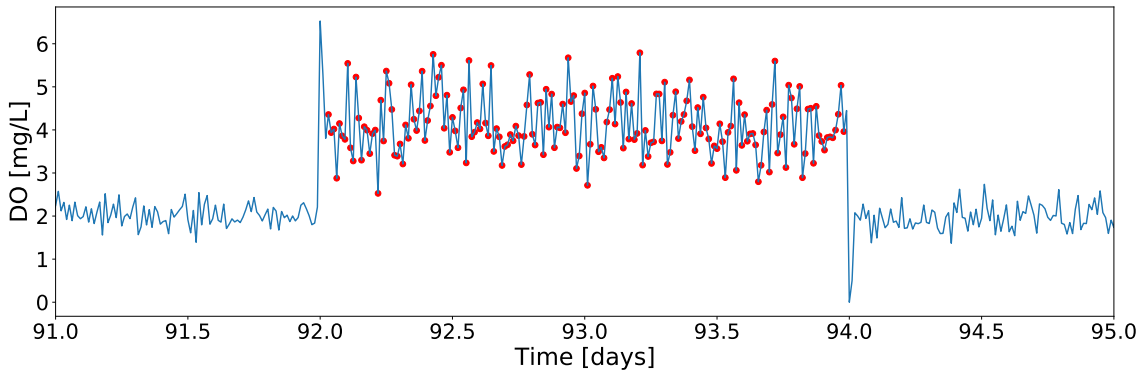



FIGURE B.18: Result obtained by LSTM-AE - Second scenario: bias.

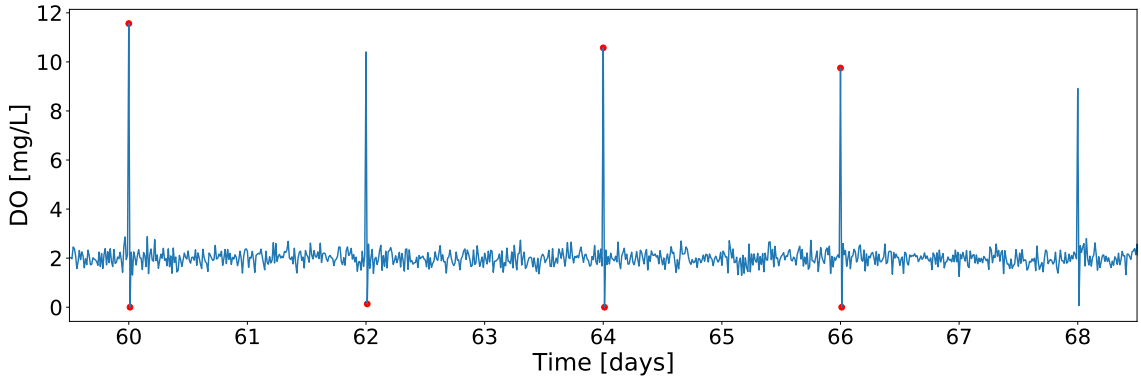



FIGURE B.19: Result obtained by LSTM-AE - Second scenario: spike.

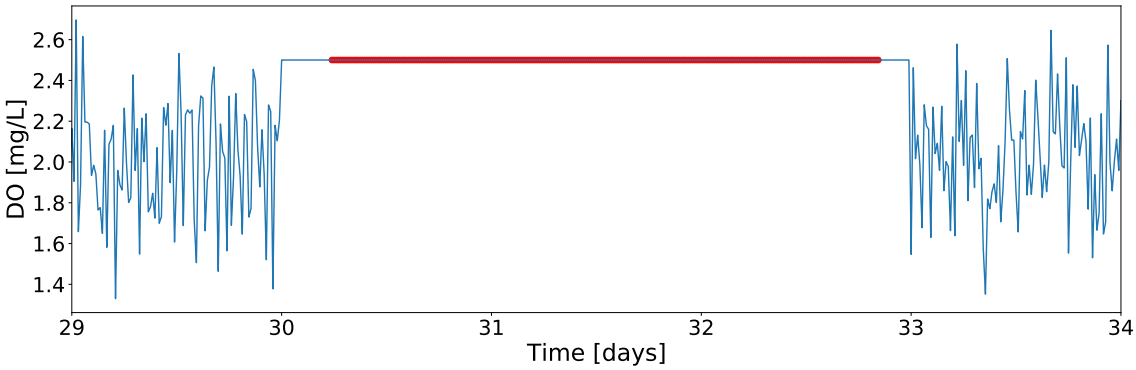



FIGURE B.20: Result obtained by LSTM-AE - Second scenario: stuck.

### B.2.3 Third Scenario

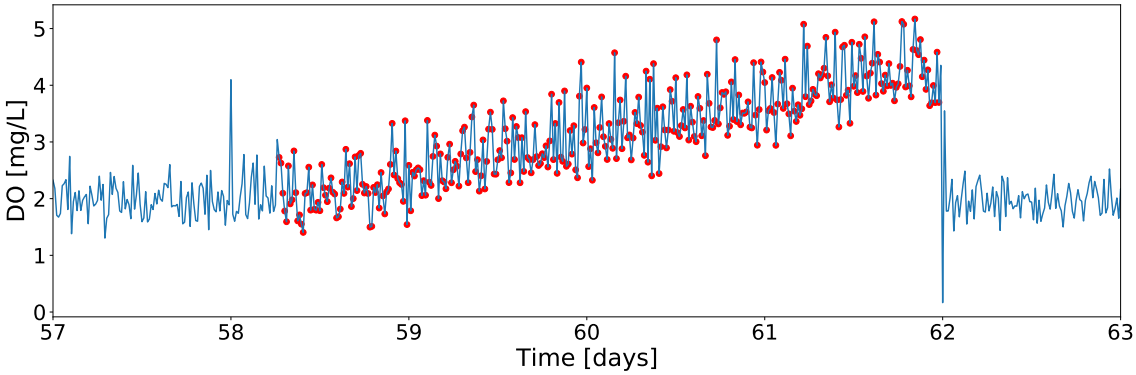



FIGURE B.21: Result obtained by LSTM-AE - Third scenario: drift.

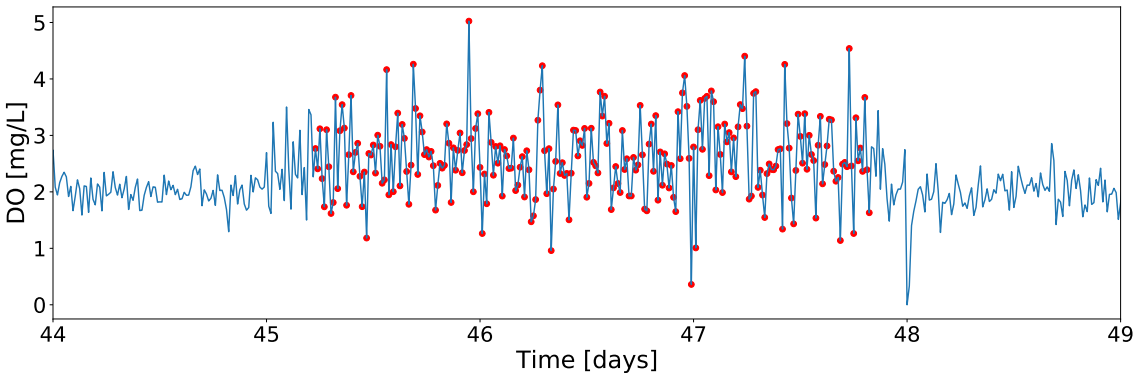



FIGURE B.22: Result obtained by LSTM-AE - Third scenario: bias.

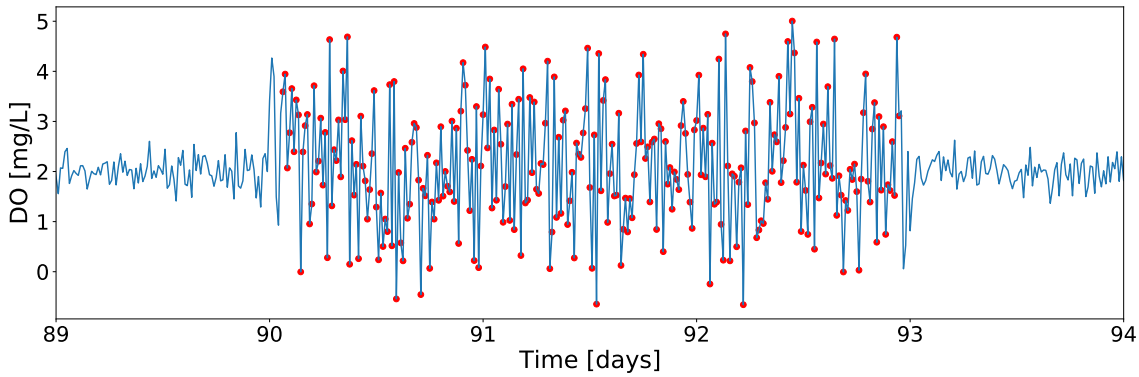



FIGURE B.23: Result obtained by LSTM-AE - Third scenario: PD.

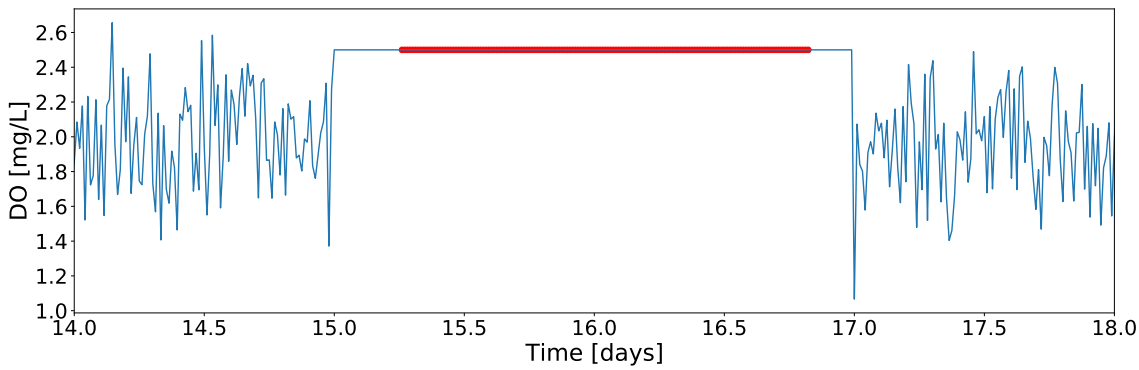



FIGURE B.24: Result obtained by LSTM-AE - Third scenario: stuck.

# Bibliography


[1] M. Serra-Prat, I. Lorenzo, E. Palomera, S. Ramírez, and J. Yébenes, "Total body water and intracellular water relationships with muscle strength, frailty and functional performance in an elderly population. a cross-sectional study," *The journal of nutrition, health & aging*, vol. 23, no. 1, pp. 96–101, 2019. [Cited on page 1.]

[2] F. Kracmar, "Water scarcity and water management. a comparative analysis of different water availability and demand assessments," Ph.D. dissertation, Wien, 2021. [Cited on page 1.]

[3] 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*, 2021, p. 27. [Cited on page 1.]

[4] U. Nations, "The Global Population Will Soon Reach 8 Billion—Then What? | United Nations," [Online; accessed 2022-08-19]. [Cited on page 1.]

[5] M. M. Mekonnen and A. Y. Hoekstra, "Four billion people facing severe water scarcity," *Science advances*, vol. 2, no. 2, p. e1500323, 2016. [Cited on page 1.]

[6] M. N. Khalaf, *Green polymers and environmental pollution control*. CRC Press, 2016. [Cited on page 2.]

[7] C. Kurrer, "Water protection and management," *Articles 191 to 193 of the Treaty on the Functioning of the European Union (TFEU)*, 2021. [Cited on page 2.]

[8] R. Hreiz, A. Latifi, and N. Roche, "Optimal design and operation of activated sludge processes: State-of-the-art," *Chemical Engineering Journal*, vol. 281, pp. 900–920, 12 2015. [Cited on page 2.]

[9] AMERICA'S INFRASTRUCTURE REPORT CARD, "Wastewater," ASCE, USA, Tech. Rep., 2021. [Online]. Available: https://infrastructurereportcard.org/cat-item/wastewater/ [Cited on page 2.]

[10] The European Federation of National Associations of Water Services, "Europe's water in figures - An overview of the European drinking water and waste water sectors," EurEAU, Tech. Rep., 2017. [Online]. Available: https://www.eureau.org/resources/publications/eureau-publications/1460-eureau-data-report-2017-1/file [Cited on page 2.]

[11] R. Salles, J. Mendes, C. H. Antunes, P. Moura, and J. Dias, "Dynamic setpoint optimization using metaheuristic algorithms for wastewater treatment plants," in *48th Annual Conference of the IEEE Industrial Electronics Society (IECON 2022)*. IEEE, October 2022, pp. 1–6. [Cited on page 2.]

[12] J. Mendes, R. Araújo, T. Matias, R. Seco, and C. Belchior, "Evolutionary learning of a fuzzy controller for industrial processes," in *Proc. of The 40th Annual Conference of the IEEE Industrial Electronics Society (IECON 2014)*. Dallas, TX, USA: IEEE, October 29 - November 1 2014, pp. 139–145. [Cited on pages 2 and 24.]

[13] J. A. P. Mendes, "Computational intelligence methodologies for control of industrial processes," Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Coimbra, Coimbra, Portugal, October 2014. [Cited on page 2.]

[14] K. MAYARA ANUSHA, "Optimization of wastewater treatment plant design and operation using simulation and cost analysis," *Journal of Engineering Sciences*, vol. 10, pp. 201–204, 2019. [Cited on page 2.]

[15] F. Liu, A. Ouedraogo, S. Manghee, and A. Danilenko, "A primer on energy efficiency for municipal water and wastewater utilities," Feb 2012. [Online]. Available: http://hdl.handle.net/10986/18060 [Cited on page 3.]

[16] 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*, vol. 10, no. 10, p. 1149, 2021. [Cited on page 3.]

[17] S. Revollar, R. Vilanova, P. Vega, M. Francisco, and M. Meneses, "Wastewater treatment plant operation: Simple control schemes with a holistic perspective," *Sustainability*, vol. 12, no. 3, 2020. [Cited on page 3.]

[18] X. Flores-Alsina, M. Arnell, Y. Amerlinck, L. Corominas, K. V. Gernaey, L. Guo, E. Lindblom, I. Nopens, J. Porro, A. Shaw, L. Snip, P. A. Vanrolleghem, and U. Jeppsson, "Balancing effluent quality, economic cost and greenhouse gas emissions during the evaluation of (plant-wide) control/operational strategies in wwtps," *Science of The Total Environment*, vol. 466-467, pp. 616–624, 2014. [Cited on page 3.]

[19] T. Brinkmann, G. G. Santonja, H. Yükseler, S. Roudier, and L. D. Sancho, "Best available techniques (bat) reference document for common waste water and waste gas treatment/management systems in the chemical sector," *Publications Office of the European Union*, 2016. [Cited on page 3.]

[20] 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*, vol. 56, pp. 74–82, 2014. [Cited on page 3.]

[21] 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*, vol. 107, pp. 410–419, 2015. [Cited on page 3.]

[22] U. Jeppsson, M.-N. Pons, I. Nopens, J. Alex, J. Copp, K. Gernaey, C. Rosén, J.-P. Steyer, and P. Vanrolleghem, "Benchmark simulation model no 2: general protocol and exploratory case studies," *Water Science and Technology*, vol. 56, no. 8, pp. 67–78, 2007. [Cited on pages 4 and 24.]

[23] R. Isermann, "Process fault detection based on modeling and estimation methods—a survey," *Automatica*, vol. 20, no. 4, pp. 387–404, 1984. [Cited on page 7.]

[24] M. Kubat, "Neural networks: a comprehensive foundation by simon haykin, macmillan, 1994, isbn 0-02-352781-7." *The Knowledge Engineering Review*, vol. 13, no. 4, pp. 409–412, 1999. [Cited on page 9.]

[25] M. Miron, L. Frangu, S. Caraman, and L. Luca, "Artificial neural network approach for fault recognition in a wastewater treatment process," in *2018 22nd International Conference on System Theory, Control and Computing (ICSTCC)*, 2018, pp. 634–639. [Cited on pages 9 and 16.]

[26] B. Mamandipoor, M. Majd, S. Sheikhalishahi, C. Modena, and V. Osmani, "Monitoring and detecting faults in wastewater treatment plants using deep learning," *Environmental monitoring and assessment*, vol. 192, no. 2, pp. 1–12, 2020. [Cited on pages 10 and 16.]

[27] X. Li, W. Chai, T. Liu, and J. Qiao, "Fault detection of dissolved oxygen sensor in wastewater treatment plants," in *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, 2020, pp. 225–230. [Cited on pages 10 and 16.]

[28] D.-J. Choi and H. Park, "A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process," *Water research*, vol. 35, no. 16, pp. 3959–3967, 2001. [Cited on pages 10 and 17.]

[29] H. Xiao, D. Huang, Y. Pan, Y. Liu, and K. Song, "Fault diagnosis and prognosis of wastewater processes with incomplete data by the auto-associative neural networks and arma model," *Chemometrics and Intelligent Laboratory Systems*, vol. 161, pp. 96–107, 2017. [Cited on pages 11 and 16.]

[30] K. Wang, P. Chang, and F. Meng, "Monitoring of wastewater treatment process based on slow feature analysis variational autoencoder," in *2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS)*. IEEE, 2021, pp. 495–502. [Cited on pages 11 and 16.]

[31] Y. Liu, H. Xiao, Y. Pan, D. Huang, and Q. Wang, "Development of multiple-step soft-sensors using a gaussian process model with application for fault prognosis," *Chemometrics and Intelligent Laboratory Systems*, vol. 157, pp. 85–95, 2016. [Cited on pages 11 and 16.]

[32] O. Samuelsson, A. Björk, J. Zambrano, and B. Carlsson, "Gaussian process regression for monitoring and fault detection of wastewater treatment processes," *Water Science and Technology*, vol. 75, no. 12, pp. 2952–2963, 03 2017. [Cited on pages 11 and 16.]

[33] C. A. Lowry and D. C. Montgomery, "A review of multivariate control charts," *IIE transactions*, vol. 27, no. 6, pp. 800–810, 1995. [Cited on page 12.]

[34] D. C. Montgomery, *Introduction to statistical quality control*. John Wiley & Sons, 2007. [Cited on page 12.]

[35] S. Wold, "Exponentially weighted moving principal components analysis and projections to latent structures," *Chemometrics and intelligent laboratory systems*, vol. 23, no. 1, pp. 149–161, 1994. [Cited on page 12.]

[36] I. Baklouti, M. Mansouri, H. Nounou, M. B. Slima, and A. B. Hamida, "Fault detection in a wastewater treatment plant," in *2017 international conference on advanced technologies for signal and image processing (ATSIP)*. IEEE, 2017, pp. 1–5. [Cited on pages 12 and 17.]

[37] C. Rosén, J. Röttorp, and U. Jeppsson, "Multivariate on-line monitoring: challenges and solutions for modern wastewater treatment operation," *Water science and technology*, vol. 47, no. 2, pp. 171–179, 2003. [Cited on page 13.]

[38] J. E. Jackson, *A user's guide to principal components*. John Wiley & Sons, 2005. [Cited on page 13.]

[39] K. L. King, Z. Wang, and D. J. Kroll, "Classification of deviations in a process," Feb. 14 2006, uS Patent 6,999,898. [Cited on page 13.]

[40] O. Schraa, B. Tole, and J. B. Copp, "Fault detection for control of wastewater treatment plants," *Water science and technology*, vol. 53, no. 4-5, pp. 375–382, 2006. [Cited on page 13.]

[41] F. Baggiani and S. Marsili-Libelli, "Real-time fault detection and isolation in biological wastewater treatment plants," *Water science and technology*, vol. 60, no. 11, pp. 2949–2961, 2009. [Cited on pages 13, 14, and 16.]

[42] A.-V. Luca, M. Simon-Várhelyi, N.-B. Mihály, and V.-M. Cristea, "Data driven detection of different dissolved oxygen sensor faults for improving operation of the wwtp control system," *Processes*, vol. 9, no. 9, p. 1633, 2021. [Cited on pages 13, 15, 16, and 47.]

[43] B. Carlsson and J. Zambrano, "Fault detection and isolation of sensors in aeration control systems," *Water Science and Technology*, vol. 73, no. 3, pp. 648–653, 2016. [Cited on pages 13 and 16.]

[44] P. Kazemi, J. Giralt, C. Bengoa, A. Masoumian, and J.-P. Steyer, "Fault detection and diagnosis in water resource recovery facilities using incremental pca," *Water Science and Technology*, vol. 82, no. 12, pp. 2711–2724, 2020. [Cited on pages 13 and 16.]

[45] C. Rosén and J. Lennox, "Multivariate and multiscale monitoring of wastewater treatment operation," *Water research*, vol. 35, no. 14, pp. 3402–3410, 2001. [Cited on pages 14 and 16.]

[46] B. Wang, Z. Li, Z. Dai, N. Lawrence, and X. Yan, "A probabilistic principal component analysis-based approach in process monitoring and fault diagnosis with application in wastewater treatment plant," *Applied Soft Computing*, vol. 82, p. 105527, 2019. [Cited on pages 14 and 17.]

[47] P. Kazemi, C. Bengoa, J.-P. Steyer, and J. Giralt, "Data-driven techniques for fault detection in anaerobic digestion process," *Process Safety and Environmental Protection*, vol. 146, pp. 905–915, 2021. [Cited on pages 14 and 17.]

[48] K. B. Newhart, R. W. Holloway, A. S. Hering, and T. Y. Cath, "Data-driven performance analyses of wastewater treatment plants: A review," *Water research*, vol. 157, pp. 498–513, 2019. [Cited on page 14.]

[49] S. Haykin, *Kalman filtering and neural networks*. John Wiley & Sons, 2004, vol. 47. [Cited on pages 14 and 17.]

[50] M. Fuente, D. Garcia-Alvarez, G. Sainz-Palmero, and P. Vega, "Fault detection in a wastewater treatment plant based on neural networks and pca," in *2012 20th mediterranean conference on control & automation (MED)*. IEEE, 2012, pp. 758–763. [Cited on pages 15 and 17.]

[51] A. Chen, H. Zhou, Y. An, and W. Sun, "Pca and pls monitoring approaches for fault detection of wastewater treatment process," in *2016 IEEE 25th International Symposium on Industrial Electronics (ISIE)*. IEEE, 2016, pp. 1022–1027. [Cited on pages 15 and 17.]

[52] A. Höskuldsson, *Prediction Methods in Science and Technology.: Vol 1. Basic theory*. Thor Publishing, 1996. [Cited on page 15.]

[53] P. Nomikos and J. F. MacGregor, "Multi-way partial least squares in monitoring batch processes," *Chemometrics and intelligent laboratory systems*, vol. 30, no. 1, pp. 97–108, 1995. [Cited on page 15.]

[54] M. C. Klanderman, K. B. Newhart, T. Y. Cath, and A. S. Hering, "Fault isolation for a complex decentralized waste water treatment facility," *Journal of the Royal Statistical*

*Society: Series C (Applied Statistics)*, vol. 69, no. 4, pp. 931–951, 2020. [Cited on pages 17 and 19.]

[55] S. U. Jan, Y. D. Lee, and I. S. Koo, "A distributed sensor-fault detection and diagnosis framework using machine learning," *Information Sciences*, vol. 547, pp. 777–796, 2021. [Cited on page 18.]

[56] S. U. Jan and I. Koo, "A novel feature selection scheme and a diversified-input svm-based classifier for sensor fault classification," *Journal of Sensors*, vol. 2018, 2018. [Cited on page 18.]

[57] Y. Wen, G. Schoups, and N. Van De Giesen, "Organic pollution of rivers: Combined threats of urbanization, livestock farming and global climate change," *Scientific reports*, vol. 7, no. 1, pp. 1–9, 2017. [Cited on page 21.]

[58] F. R. Spellman, *Handbook of water and wastewater treatment plant operations*. CRC press, 2003. [Cited on pages 21 and 22.]

[59] 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*, vol. 210, pp. 232–242, 2018. [Cited on page 23.]

[60] Q. Zhang, J. Hu, D.-J. Lee, Y. Chang, and Y.-J. Lee, "Sludge treatment: current research trends," *Bioresource technology*, vol. 243, pp. 1159–1172, 2017. [Cited on page 23.]

[61] A. B. da Costa, "Desenvolvimento de metamodelos kriging e otimização de uma planta de tratamento de efluentes (bsm2)." Masters dissertation, Universidade Federal de Campina Grande, Campina Grande, Brazil, December 2016. [Cited on pages 23 and 25.]

[62] 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*, vol. 66, no. 4, pp. 694–708, 2020. [Cited on page 23.]

[63] J. Alex, J.-F. Beteau, J. Copp, C. Hellinga, 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)*. IEEE, 1999, pp. 3746–3751. [Cited on page 23.]

[64] K. Morales-Rodelo, M. Francisco, H. Alvarez, P. Vega, and S. Revollar, "Collaborative control applied to bsm1 for wastewater treatment plants," *Processes*, vol. 8, no. 11, 2020. [Cited on page 24.]

[65] G. Harja and I. Naşcu, "Control of an activated sludge wastewater treatment process based on a calibrated and modified bsm1 model," in *2019 20th International Carpathian Control Conference (ICCC)*, 2019, pp. 1–6.

[66] 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)*, 2019, pp. 73–78.

[67] S. Rastegar, R. Araújo, and J. Mendes, "A new approach for online t-s fuzzy identification and model predictive control of nonlinear systems," *Journal of Vibration and Control*, vol. 22, no. 7, pp. 1820–1837, April 2016.

[68] S. Rastegar, R. Araújo, J. Mendes, T. Matias, and A. Emami, "Self-adaptive takagi-sugeno model identification methodology for industrial control processes," in *Proc. of The 40th Annual Conference of the IEEE Industrial Electronics Society (IECON 2014)*. Dallas, TX, USA: IEEE, October 29 - November 1 2014, pp. 281–287.

[69] 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*, vol. 69, no. 10, pp. 2052–2058, 2014.

[70] J. Mendes, R. Araújo, and F. Souza, "Adaptive fuzzy identification and predictive control for industrial processes," *Expert Systems with Applications*, vol. 40, no. 17, pp. 6964–6975, December 2013. [Cited on page 24.]

[71] M. Meneses, H. Concepción, and R. Vilanova, "Joint environmental and economical analysis of wastewater treatment plants control strategies: a benchmark scenario analysis," *Sustainability*, vol. 8, no. 4, p. 360, 2016. [Cited on page 25.]

[72] 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*, 2009. [Cited on page 25.]

[73] D. J. Batstone, J. Keller, I. Angelidaki, S. Kalyuzhnyi, S. Pavlostathis, A. Rozzi, W. Sanders, H. Siegrist, and V. Vavilin, "The iwa anaerobic digestion model no 1 (adm1)," *Water Science and technology*, vol. 45, no. 10, pp. 65–73, 2002. [Cited on page 26.]

[74] A. C. do Amaral, R. L. R. Steinmetz, and A. Kunz, "O processo de biodigestão." *Embrapa Suínos e Aves-Capítulo em livro científico (ALICE)*, 2019. [Cited on page 26.]

[75] 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*, vol. 50, pp. 346–362, 2015. [Cited on page 26.]

[76] I. Nopens, L. Benedetti, U. Jeppsson, M.-N. Pons, J. Alex, J. B. Copp, K. V. Gernaey, C. Rosen, J.-P. Steyer, and P. A. Vanrolleghem, "Benchmark simulation model no 2: finalisation of plant layout and default control strategy," *Water Science and Technology*, vol. 62, no. 9, pp. 1967–1974, 2010. [Cited on page 26.]

[77] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006. [Cited on page 29.]

[78] K. Pawar and V. Z. Attar, "Assessment of autoencoder architectures for data representation," in *Deep Learning: Concepts and Architectures*. Springer, 2020, pp. 101–132. [Cited on page 29.]

[79] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International conference on artificial neural networks*. Springer, 2011, pp. 52–59. [Cited on pages 29 and 34.]

[80] M. Makkie, H. Huang, Y. Zhao, A. V. Vasilakos, and T. Liu, "Fast and scalable distributed deep convolutional autoencoder for fmri big data analytics," *Neurocomputing*, vol. 325, pp. 20–30, 2019. [Cited on page 29.]

[81] A. Ullah, K. Muhammad, I. U. Haq, and S. W. Baik, "Action recognition using optimized deep autoencoder and cnn for surveillance data streams of non-stationary environments," *Future Generation Computer Systems*, vol. 96, pp. 386–397, 2019. [Cited on page 29.]

[82] J. Zhang, K. Li, Y. Liang, and N. Li, "Learning 3d faces from 2d images via stacked contractive autoencoder," *Neurocomputing*, vol. 257, pp. 67–78, 2017. [Cited on page 29.]

[83] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, vol. 27, 2014. [Cited on page 29.]

[84] A. R. Pathak, M. Pandey, S. Rautaray, and K. Pawar, "Assessment of object detection using deep convolutional neural networks," in *Intelligent Computing and Information and Communication*. Springer, 2018, pp. 457–466. [Cited on page 29.]

[85] K. Khare, O. Darekar, P. Gupta, and V. Attar, "Short term stock price prediction using deep learning," in *2017 2nd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT)*. IEEE, 2017, pp. 482–486. [Cited on page 29.]

[86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," California Univ San Diego La Jolla Inst for Cognitive Science, Tech. Rep., 1985. [Cited on page 30.]

[87] G. Dong, G. Liao, H. Liu, and G. Kuang, "A review of the autoencoder and its variants: A comparative perspective from target recognition in synthetic-aperture radar images," *IEEE Geoscience and Remote Sensing Magazine*, vol. 6, no. 3, pp. 44–68, 2018. [Cited on pages xv, 31, and 34.]

[88] W. Pedrycz and S.-M. Chen, *Deep learning: Concepts and architectures*. Springer, 2020. [Cited on pages xv and 32.]

[89] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997. [Cited on page 33.]

[90] S. Varsamopoulos, K. Bertels, C. G. Almudever *et al.*, "Designing neural network based decoders for surface codes," *arXiv preprint arXiv:1811.12456*, 2018. [Cited on pages xv and 33.]

[91] R. Salles, J. Mendes, R. Araújo, C. Melo, and P. Moura, "Prediction of key variables in wastewater treatment plants using machine learning models," in *Proc. 2022 IEEE International Joint Conference on Neural Networks (IJCNN 2022), at the 2022 World Congress on Computational Intelligence (WCCI 2022)*. Padova, Italy: IEEE, July 18-23 2022, pp. 1–9. [Cited on page 33.]

[92] N. Srivastava, E. Mansimov, and R. Salakhudinov, "Unsupervised learning of video representations using lstms," in *International conference on machine learning*. PMLR, 2015, pp. 843–852. [Cited on page 34.]

[93] K. V. Gernaey and U. Jeppsson, *Benchmarking of control strategies for wastewater treatment plants*. IWA publishing, 2014. [Cited on pages xvii and 38.]