BABEȘ-BOLYAI UNIVERSITY
Faculty of Chemistry and Chemical Engineering
Doctoral School of Chemical Engineering

Detection and Identification of Wastewater Treatment Plant Main Sensors Faults for Improving Operation Performance

Summary of Ph.D. Thesis

PhD Student: Eng. Alexandra-Veronica Luca

Scientific supervisor: Prof. Dr. Eng. Vasile-Mircea Cristea

Table of contents

INTRODUCTION

1. Motivation

Water management and treatment have been crucial to human health, environmental protection, and sustainable development since ancient civilizations. Early systems focused primarily on diverting wastewater, with treatment emerging only in the 19th century during industrialization due to increasing urbanization, pollution, and waterborne diseases. The development of the activated sludge process in 1914 marked a breakthrough in wastewater treatment, enabling more efficient removal of organic pollutants. Over the 20th century, advances in biological treatment, aeration, and chemical processes, combined with stricter regulations have driven the adoption of modern, large-scale, and energy-intensive treatment plants. Today, the focus extends beyond pollution removal to resource recovery, energy efficiency, and sustainability.

As global populations rise and water scarcity intensifies, the demand for high-quality water treatment is increasing. Modern WWTPs rely heavily on accurate sensor measurements for key variables such as dissolved oxygen, nitrate, and nitrite. These sensors are critical for effective process control, energy management, and compliance with environmental regulations. Faulty sensors can lead to inefficient treatment, higher energy consumption, reduced effluent quality, and increased greenhouse gas (GHG) emissions. Given that wastewater treatment contributes up to 5% of global GHG emissions, reducing emissions through efficient operation has become a key environmental priority.

The complexity of WWTP processes, characterized by non-linear dynamics, variable influent composition, and interdependent biological reactions, requires advanced monitoring and control strategies. Mathematical and data-driven models, including Principal Component Analysis (PCA) and multivariate statistical process monitoring, play a crucial role in early fault detection, diagnosis, and process optimization. These tools help minimize environmental impact, reduce energy costs, and enhance operational reliability. Moreover, integrating energy recovery and renewable energy solutions offers the potential for plants to move toward energy neutrality, contributing to both economic and environmental sustainability.

This research is motivated by the urgent need to improve WWTP reliability, sustainability, and efficiency. By investigating sensor fault detection, diagnosis, and the associated environmental and economic impacts, the study addresses a critical knowledge gap. Accurate sensor monitoring not only ensures compliance with evolving regulations but also supports optimized energy use, GHG emission reduction, and the broader goals of circular economy and water resource recovery. In essence, advancing sensor-based monitoring and control in WWTPs is vital for the safe, cost-effective, and environmentally responsible management of modern wastewater treatment systems.

CURRENT STATE OF KNOWLEDGE

2. Current state of the knowledge in modelling activated sludge technology based WWTP

2.1. Introduction

Chemical process monitoring in WWTPs is essential for efficiency, effluent quality, and advanced control, relying on mechanistic or data-driven models to predict process variables and optimize operations (Frank, 1990; Yoon & MacGregor, 2001). Dynamic models aid in nutrient removal plant optimization, crisis management, and testing of operational scenarios (Henze *et al*., 2000). The International Water Association developed standardized Activated Sludge Models (ASM1-ASM3) to represent carbon, nitrogen, and phosphorus removal, supporting design, operation, and optimization of WWTPs (Rieger *et al*., 2012; Hauduc *et al*., 2011; Gernaey *et al*., 2004; Henze *et al*., 2000).

European COST Action groups established Benchmark Simulation Models (BSM1 and BSM2) to evaluate control strategies and plant performance, including influent scenarios, energy use, sludge production, and long-term facility assessment, enabling objective comparison of control laws and optimization of processes such as anaerobic digestion and biogas generation (Alex *et al*., 1999, 2008; Beraud *et al*., 2009).

2.2. Description of the Activated Sludge Model No. 1

The ASM1 model is perhaps the model most commonly employed for describing WWT processes worldwide, despite the fact that it has been expanded since its conception to incorporate more chemical oxygen demand fractions, to accommodate new experimental findings (Sollfrank and Gujer, 1991), to define the growth and population dynamics of floc forming and filamentous bacteria (Gujer and Kappeler, 1992), and to define improved biological phosphorus removal (Henze *et al*., 1995).

The various processes included in the ASM1 model are defined as it follows (Jeppsson, 1996):

Aerobic growth of heterotrophs:

$$process1 = \mu_H \cdot \left(\frac{S_S}{K_S + S_S}\right) \cdot \left(\frac{S_O}{K_{O,H} + S_O}\right) \cdot X_{B,H} \tag{2.1}$$

Anoxic growth of heterotrophs

$$process2 = \mu_H \cdot \left(\frac{S_S}{K_S + S_S}\right) \cdot \left(\frac{K_{O,H}}{K_{O,H} + S_O}\right) \cdot \left(\frac{S_{NO}}{K_{NO} + S_{NO}}\right) \cdot \eta_g \cdot X_{B,H} \tag{2.2}$$

Aerobic growth of autotrophs:

$$process3 = \mu_A \cdot \left(\frac{S_{NH}}{K_{NH} + S_{NH}}\right) \cdot \left(\frac{S_O}{K_{O,A} + S_O}\right) \cdot X_{B,A} \tag{2.3}$$

Decay of heterotrophs:

$$process4 = b_H \cdot X_{B,H} \quad (2.4)$$

Decay of autotrophs:

$$process5 = b_A \cdot X_{B,A} \quad (2.5)$$

Ammonification of soluble organic nitrogen:

$$process6 = k_a \cdot S_{ND} \cdot X_{B,H} \quad (2.6)$$

Hydrolysis of entrapped organics:

$$process7 = k_h \cdot \frac{X_S/X_{B,H}}{K_x + (X_S/X_{B,H})} \cdot \left[\left(\frac{S_O}{K_{O,H} + S_O} + \eta_h \cdot \left(\frac{K_{O,H}}{K_{O,H} + S_O}\right) \cdot \left(\frac{S_{NO}}{K_{NO} + S_{NO}}\right)\right] \cdot X_{B,H} \quad (2.7)$$

Hydrolysis of entrapped organic nitrogen:

$$process8 = k_h \cdot \frac{X_S/X_{B,H}}{K_x + (X_S/X_{B,H})} \cdot \left[\left(\frac{S_O}{K_{O,H} + S_O} + \eta_h \cdot \left(\frac{K_{O,H}}{K_{O,H} + S_O}\right) \cdot \left(\frac{S_{NO}}{K_{NO} + S_{NO}}\right)\right] \cdot X_{B,H} \cdot (X_{ND}/X_S) \quad (2.8)$$

2.3. Description of the Benchmark Simulation Model No. 1

The concept of Benchmark Simulation Model No. 1 appears to be straightforward. It is based on a modified Ludzack-Ettinger configuration that allows the WWT plant to meet stringent COD and nitrogen removal regulations (Tchobanoglous *et al*., 2003).

The benchmark layout consists of a five-compartment activated sludge reactor. Out of the five compartments, two are anoxic tanks and the other three are aerobic tanks.

The general equation for the mass balance of the first reactor (non-aerated) is (Alex *et al*., 2008):

$$\frac{dZ_{as,1}}{dt} = \frac{1}{V_{as,1}}(Q_{int}Z_{int} + Q_rZ_r + Q_{po}Z_{po} + r_{Z,1}V_{as,1} - (Q_{int} + Q_r + Q_{po})Z_{as,1}) \quad (2.9)$$

where Q_{int} is the internal recycle, Q_r represents the external recycle and Q_{po} is the overflow rate of the primary settler.

For the rest of the reactors, the mass balance equation is:

$$\frac{dZ_{as,k}}{dt} = \frac{1}{V_{as,k}}(Q_{k-1}Z_{as,k-1} + r_{Z,k}V_{as,k} - Q_kZ_{as,k}) \quad (2.10)$$

where *k* represents the of the bioreactor number (2-5).

A secondary settler is positioned after the activated sludge reactor. The secondary settler is represented using a double-exponential velocity function (Takacs *et al*., 1991):

$$v_s(X_{sc}) = \max\left[0, \min\left\{v_0', v_0\left(e^{-r_h(X_{sc} - X_{\min})} - e^{-r_p(X_{sc} - X_{\min})}\right)\right\}\right] \quad (2.11)$$

where v_s represents the settling velocity, v_0 is the theoretical maximum settling velocity (m/d), v_0' is the practical maximum Vesiling settling velocity (m/d), r_h is the hindered zone settling parameter (m^3/g), r_p is the flocculant zone settling parameter (m^3/g), X_{sc} is the sludge concentration and X_{min} is the minimum attainable suspended solids concentration (g/m^3).

The plant includes an internal recirculation flow to supply nitrates in the anoxic reactors from the

aerobic reactors, and an external recirculation in order to provide sludge from the secondary settler to the anoxic reactors. The benchmark layout comprises a simple control strategy with two PI control loops.

2.4. Tools used for dynamic modelling

Matlab is a multipurpose programming language and numerical computing environment that includes tools for matrices manipulation, algorithm implementation, and a variety of functions (MathWorks, 2025a). Simulink is a graphical programming environment for block diagrams that works with Matlab (MathWorks, 2025b). It is generally used for modelling and simulating dynamic systems, as well as continuous testing and verification systems. Through the specifically designed S-functions, the Benchmark Simulated Models are applied in this work.

2.5. WWTP operation performance evaluation

In the present work four separate performance indices were used to assess the overall impact of the WWTP dysfunctional operation on plant performance: aeration energy (AE), pumping energy (PE), effluent quality (EQ) and the newly introduced greenhouse gas emissions.

The aeration energy index is calculated using the oxygen mass transfer coefficient of aerobic bioreactors (K_{La}), which is directly related to the injected air flow rate, as shown in equation (2.12).

$$AE = \frac{S_{O_{sat}}}{T \cdot 1.8 \cdot 1000} \cdot \int_0^{Tdays} \sum_{aerated\ reactors} V_{bioreactor} \cdot K_L a_i(t) dt \quad (2.12)$$

where $S_{O_{sat}}$ is the oxygen saturation concentration (mg O_2/L), T_{days} is the time in days of the evaluation, $V_{bioreactor}$ represents the volume of the bioreactor (m^3) and $K_L a_i$ is the mass transfer coefficient in the aerated bioreactor i.

The pumping energy index is calculated using the flow rates of nitrate recirculation, return activated sludge recycling, and waste, as shown in equation (2.13):

$$PE = \frac{1}{T} \cdot \int_0^{Tdays} [0.004 \cdot Q_{NR}(t) + 0.08 \cdot Q_{RAS}(t) + 0.05 \cdot Q_{waste}(t)] dt \quad (2.13)$$

with Q_{NR} the flow rate of the nitrate recirculation (m^3/day), Q_{RAS} the flow rate of the return activated sludge (m^3/day), Q_{waste} the flow rate of waste from the secondary settler (m^3/day). Both aeration and pumping energy are measured in kWh/day.

The effluent quality index is calculated using the total suspended solids (TSS), chemical oxygen demand (COD), biochemical oxygen demand (BOD), total Kjeldahl nitrogen (TKN), and nitrate and nitrite concentrations in the effluent flow stream, as presented in equation (2.14). They are multiplied by the effluent flow rate. This index is expressed in kilograms of pollutant units per day.

$$EQ = \frac{1}{T \cdot 1000} \cdot \int_{0}^{Tdays} [PU_{TSS}(t) + PU_{COD}(t) + PU_{BOD}(t) + PU_{TKN}(t) + PU_{NO}(t)] \cdot \cdot Q_{effluent}(t)dt \quad (2.14)$$

where PU_{TSS} denotes total suspended solids, PU_{COD} refers to the chemical oxygen demand and PU_{BOD} to the biochemical oxygen demand, PU_{TKN} considers the Total Kjeldahl Nitrogen, PU_{NO} accounts for the nitrate and nitrite and $Q_{effluent}$ is the effluent flow rate.

2.6. Evaluation of the detection and diagnosis performance

The confusion matrix or plot is a widely used tool in classification analysis, providing a visual representation of the performance of a model by comparing actual outcomes to predicted outcomes. Confusion matrices were constructed for each of the analysed cases, with each matrix displaying the counts of true positives, false positives, true negatives, and false negatives. This tabular representation serves as the foundation for calculating key performance metrics that assess the system's detection capabilities as they are displayed bellow:

The formula for calculating Positive Predictive Value is:

$$PPV = \frac{Number\ of\ true\ positives}{Number\ of\ true\ positives + Number\ of\ false\ positives} \quad (2.15)$$

Accuracy is a metric used to evaluate the performance of a classification model:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad (2.16)$$

The False Alarm Rate is defined as follows:

$$FAR = \frac{Number\ of\ false\ alarms}{Total\ number\ of\ non-fault\ conditions} \quad (2.17)$$

The Missed Alarm Rate is defined as:

$$MAR = FNR = \frac{Number\ of\ missed\ alarms\ (false\ negatives)}{Total\ number\ of\ actual\ faults\ (true\ positives + false\ negatives)} \quad (2.18)$$

The F1 score is defined as:

$$F1\ score = 2\ x \frac{Precision\ x\ Recall}{Precision + Recall} = 2\ x \frac{PPV\ x\ TPR}{PPV + TPR} \quad (2.19)$$

2.7. GHG assessment

Total GHG emissions from the municipal wastewater treatment plant's water line include both on and off-site CO_2 and N_2O gas emissions. Aside from CO_2, N_2O is considered a significant contributor to GHGs due to its global warming potential (GWP) of approximately 265-298 times that of CO_2 (Vallero, 2019).

Off-site CO_2 emissions (kg CO_2/day) involve indirect CO_2 emissions from the power plant related to the power consumed in the wastewater treatment process. They are defined as follows:

$$P_{CO_2,off-site} = k_{PG} \cdot e_D \quad (2.20)$$

where k_{PG} is the site-specific emission factor per unit of energy generated, considered with a value of

0.19 kg CO_{2e}/kWh, and e_D is the total energy demand, calculated as the sum of the aeration energy and pumping energy (Mannina *et al*., 2016; Listowski *et al*., 2011).

Off-site N_2O emissions involve N_2O produced by biological degradation in the wastewater treatment plant's effluent (downstream) (Mannina *et al*., 2016; Prendez and Lara-Gonzales, 2008):

$$P_{N_2O,off-site} = N_{effluent} \cdot EF_{effluent} \quad (2.21)$$

where $N_{effluent}$ is the nitrogen load in the effluent discharged into aquatic environments and $EF_{effluent}$ = 0.005·44/28 kg N_2O/kg N (IPCC, 2006) is the emission factor for N_2O emissions from the discharged wastewater.

The on-site CO_2 emissions from the water-line of aerobic biological processes are determined using the following expression:

$$P_{CO_2,on-site} = Q_{influent} \cdot 0.99 \cdot (1 - Y_H) \cdot \eta_{ASP} \cdot bCOD + Q_{influent} \cdot 1.03 \cdot Y_H \cdot \eta_{ASP} \cdot bCOD \cdot \frac{k_{d,H} \cdot MCRT}{1 + k_{d,H} \cdot MCRT} \quad (2.22)$$

where: $Q_{influent}$ is the plant influent flow rate (m^3/day), 0.99 kg CO_{2e}/kg COD is the emission factor related to organic compounds, Y_H is the heterotrophic biomass yield (massVSS/massCOD) (Alex *et al*., 2008), η_{ASP} is the biodegradable COD (*bCOD*) removal in the activated sludge reactors, 1.03 kg CO_{2e}/kg COD is the emission factor related to activated sludge biomass, $k_{d,H}$ is the decay rate of heterotrophic biomass and has a value of 0.3 day^{-1} (Alex *et al*., 2008), and MCRT is the mean cell retention time, that is 15 days for this case (Mannina *et al*., 2016; Gori *et al*., 2011).

The following relationship can be used to determine the on-site N_2O emissions from the water-line.

$$P_{N_2O,on-site} = Q_{influent} \cdot (TN_{in} - TN_{out}) \cdot r_{N_2O} \quad (2.23)$$

where TN_{in} represents the total nitrogen from the influent (kg N/m^3), TN_{out} is the total nitrogen in the effluent (kg N/m^3) (Huang and Shen, 2019) and r_{N_2O} is the emission rate of N_2O (kg N_2O/kg N) (Baresel *et al*., 2016).

PERSONAL CONTRIBUTIONS

3. Model of the municipal WWTP water line

3.1. Introduction

This work's case study is the wastewater treatment plant (WWTP) in Someseni, Cluj, which uses an anaerobic-anoxic-aerobic (A^2O) arrangement. The facility treats both sewage and urban runoff and is designed with a water line that integrates physical separation, biochemical treatment, settling, and a sludge line with biogas production. Following the A^2O process, the biological stage involves the sequential passage of wastewater through anaerobic, anoxic, and aerobic tanks, which facilitates the release and uptake of phosphorus, denitrification, nitrification, and carbon degradation. Following this, solid-liquid separation occurs in clarifiers. The majority of the activated sludge is recycled, but any extra is thickened, anaerobically digested, and used to produce biogas for energy recovery.

3.2. WWTP layout and design parameters

Following its entry into the WWTP, the influent wastewater undergoes primary sedimentation. After the primary clarifier, the wastewater enters the biodegradation tanks, where the removal of carbon, nitrogen, and phosphorus is accomplished by processes specific to the activated sludge technology. The three zones of biodegradation basins are anaerobic, anoxic, and aerobic. The anaerobic zone is where phosphorus-accumulating microorganisms grow. In the anoxic zone denitrification occurs with the help of heterotrophic bacteria that use nitrate and nitrite for cellular respiration in the absence of molecular oxygen. As a result, nitrates and nitrites are converted to nitrous oxides and nitrogen gas (Wanner and Grau, 1988). The last three bioreactors are forming the aerated section of the plant. Here, air is introduced in the wastewater to ensure the necessary level of dissolved oxygen with the help of compressors. In this compartment organic matter is converted into aerobic biomass via assimilation and phosphorus is absorbed by phosphorus-accumulating microorganisms. The internal recirculation sends the formed nitrates and nitrites back to the anoxic basins to complete the denitrification step. The remaining liquid-solid mixture is fed into the secondary settlers. 99.5% of the activated sludge, rich in microorganisms, is returned to the anaerobic zone. The sludge from the settler's base is returned to the anaerobic reactor via the return activated sludge stream, while the designated river receives the treated water discharge (Simon-Varhelyi *et al*., 2020a). Additionally, the plant includes a biogas production section.

3.3. Municipal WWTP model

The basis of the WWTP simulator developed and used in this work consists of BSM1 and ASM1. The original BSM1 model was modified as needed to conform to the anaerobic-anoxic-oxic, A^2O, configuration of the municipal WWTP under investigation.

The facility uses two feedback control loops based on PI controllers. The first loop controls the concentration of dissolved oxygen in the third aerated bioreactor. Its goal is to maintain the dissolved oxygen to the 2 mg O_2/L setpoint by adjusting the airflow fed to the three aerated reactors. The second loop regulates the nitrates and nitrites levels in the anoxic bioreactor. It keeps the NO concentration at the desired setpoint of 0.01 mg N/L by adjusting the mixed-liquid recycle flow, to perform appropriate denitrification. As presented in Figure 3.1, the municipal WWTP also comprises a primary clarifier and a secondary one. The model of the plant follows this water line structure.

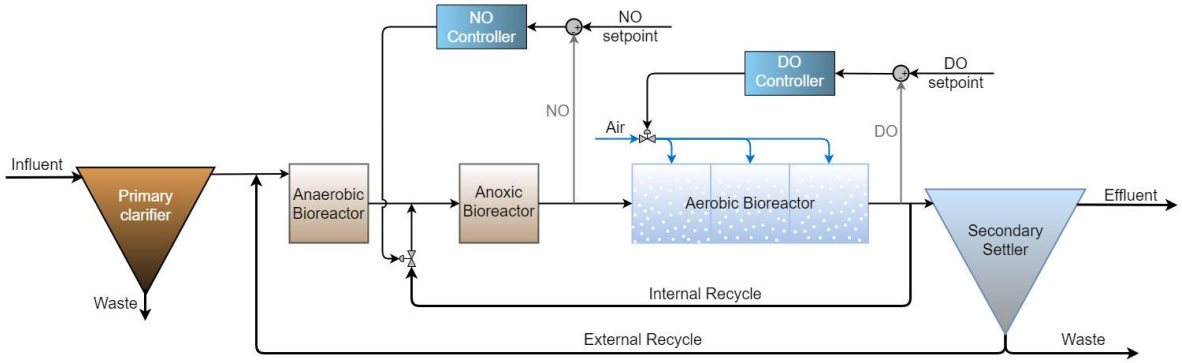

Figure 3.1. Municipal WWTP layout

The dynamic-state simulations of both normal and faulty DO and NO concentration sensors operation were carried out for this study using the previously developed and calibrated Matlab model of the municipal WWTP, based on ASM1, BSM1, and Matlab & Simulink software (Varhelyi *et al.*, 2019).

The advantages of employing mathematical models embedded in dynamic simulators, whether they are mechanistic or statistical, rely on their capacity to predict the future evolution of the main process variables under diverse conditions. They offer a valuable tool for the development of the design and new operation improvement solutions.

3.4. Conclusions

The Someseni municipal WWTP is a state-of-the-art treatment facility whose A^2O configuration, along with sophisticated separation, biochemical, and sludge management processes, enable energy recovery through the production of biogas and guarantee the effective removal of organic matter and nutrients. Understanding the intricate dynamics of wastewater treatment and accurately representing site-specific conditions through meticulous calibration with actual operational data have been made possible by the development of a mathematical simulator. The model provides a dependable tool for forecasting plant performance in both steady-state and dynamic scenarios by combining ongoing

online monitoring with laboratory validation.

4. Methods for detection and diagnosis of the DO and NO sensor faults

4.1. Introduction

Maintaining the mandatory standards for water quality and wastewater treatment process efficiency can be difficult when particular anomalies occur in the system. When abnormal operating conditions are promptly identified and corrected, an efficient and safe operation can be ensured. Various cases of abnormal functioning that appear in WWTPs might be caused by faulty sensors.

Advanced fault detection and diagnosis methods that use data-driven modelling, machine learning, and artificial intelligence have been created to lessen the impact of malfunctioning sensors.

Among the most important monitoring tools in wastewater treatment facilities are dissolved oxygen and nitrate sensors, which are crucial for preserving effective biological treatment processes. Effective aeration and nitrogen removal are essential components of wastewater treatment, and precise real-time data from these sensors are important for maximizing plant efficiency, cutting operating costs, and guaranteeing adherence to environmental standards (Tchobanoglous *et al*., 2014).

4.2. Investigated types of faults

Given the critical importance of the DO and NO sensors in the overall operation of the WWTP, the detection and identification of different DO and NO sensors types of errors were examined in this thesis. The six investigated types of sensor faults are defined and described in the following.

Bias fault type, also known as a shift or an off-set of the sensor's generated signal values, may be frequently caused by a miscalibrated sensor(Teh *et al*., 2020).

Drift type of fault appears as a time-varying deviation from the true value and it is distinguished by a time-varying ramp signal (Teh *et al*., 2020).

Wrong gain type of fault is a loss of sensor effectiveness caused by a calibration error, specifically when the calibration slope of the sensor is affected by an inappropriate gain factor or not precisely defined in the calibration step (Rosen *et al*., 2008).

Loss of accuracy is an irregular degradation of the sensor's performance (Teh *et al*., 2020).

Fixed value fault is characterized by a sensor that generates the same constant value (Rosen *et al*., 2008). Complete failure (with two instances, minimum or maximum) appears when the sensor's output signal equals or exceeds the sensor's minimum or maximum calibration value (Rosen *et al*., 2008). This fault is a particular case of the fixed value fault.

4.3. Methods for sensor fault detection and diagnosis

Several studies have used various Multivariate Statistical Process Control (MSPC) methodologies to examine sensor abnormal operation and strategies for detecting inaccurate information or malfunctioning of the sensors. Control charts (Kazemi *et al*., 2020), independent component analysis

(ICA) (Villegas *et al*., 2010) and principal component analysis (Tao *et al*., 2013; Garcia-Alvarez *et al*., 2009; Qin, 2009) are MSPC methods that have been used for in-depth monitoring and fault detection.

Few studies have looked into the issue of DO sensor failures. One study developed a PCA method for detecting three different fault types (Lee *et al*., 2006) and another paper investigated the wrong output signal originating from the sensor (Sanchez-Fernandez *et al*., 2015).

Only a few studies examined the issues resulting from an out-of-range value of the nitrate and nitrite concentration sensor of WWTPs. One of these studies focused on the drift of an amperometric nitrite sensor (Britschgi *et al*., 2020), another one studied the estimation accuracy of a soft sensor for nitrate concentration (Corona *et al*., 2013).

Fisher Discriminant Analysis (FDA) is a well-known pattern classification technique (Duda *et al*., 2000), and its application in chemical process data analysis has increased steadily over the last two decades (Fuente *et al*., 2008; He *et al*., 2005). The FDA's application to faulty sensors was investigated for air processing units (Du and Jin, 2008).

4.3.1. Fault detection methods (Principal Component Analysis)

PCA is a data-mining methodology that is based on a process model built with ordinary process data. PCA models are primarily used to emphasize the correlation between process variables from raw data. The method reduces the large model training data set by linearly transforming it into a smaller one that retains the critical information, enabling facile data interpretation. The resulting data set includes a score matrix and a loadings matrix, which contain information revealed by the reduced set of variables. When the number of measurements exceeds the number of states, PCA can help to simplify process monitoring and fault detection (Wise *et al*., 1990).

The PCA method, which is frequently used with Hotelling's T^2 and Squared Prediction Error (*SPE*), assumes a series of procedures to detect faults in model's principal or residual subspace.

4.3.2. Fault diagnosis methods (Fisher Discriminant Analysis)

Fisher discriminant analysis is a pattern categorization method with a high classification potential. The FDA's primary goal is to find the Fisher optimal discriminant vector, which maximizes the Fisher criterion function. For the purpose of constructing a lower-dimensional feature space, the higher-dimensional feature space of process measurements can be projected onto the obtained optimal discriminant vectors space.

In order to identify the defects, FDA examines measured data collected under different faults and uses a discriminant function to assess the similarity between both the current data and the data belonging to each class. The observation is assigned to class i when the maximum discriminant function value, g_i, meets the following conditions:

$$g_i(x) > g_j(x), \forall j \neq i \quad (4.1)$$

$g_i(x)$ is the discriminant function for class i given by a measured vector $x \in R^m$, and $g_j(x)$ is the discriminant function for class j given by the measured vector x.

4.4. Conclusions

The reliability of sensor networks in wastewater treatment plants is critical for achieving high water quality standards and maximising operational efficiency. It is impossible to overstate the importance of crucial sensors like DO and NO sensors in biological treatment processes. Inaccurate data from DO sensors can cause poor effluent quality and energy inefficiencies. On the other hand, faulty NO sensors can cause eutrophication, which harms the environment.

For real-time sensor monitoring, methods such as PCA and FDA provide effective tools. These techniques are essential for reducing downtime and guaranteeing ongoing adherence to water quality requirements because they can identify sensor malfunctions, categorize them, and recommend remedial actions.

The reliability and efficiency of WWTPs can be significantly improved by integrating efficient fault detection and diagnosis techniques with routine sensor maintenance and calibration, while enhancing the plant's sustainable and energy-efficient operation.

5. Implementation of the faults in the Simulink model of the municipal WWTP

5.1. Introduction

Fault detection and identification are crucial to the management of wastewater treatment systems because inaccurate sensor readings can result in incorrect chemical dosage, ineffective aeration, or even regulatory violations. By using PCA and FDA, this work seeks to improve the capacity to identify and diagnose sensor problems. The DO and NO sensor errors that are essential to preserving the best possible wastewater treatment procedures are detected and identified by these models.

To improve diagnostic methods and support fault investigation, controlled simulations of sensor errors were performed using Simulink. Six different faults of the sensors were investigated: bias, drift, wrong gain, loss of accuracy, fixed value and complete failure. Moreover, the implementation of sensor faults in Simulink together with the faulty free operation is a novel aspect of this work.

The datasets obtained from the simulations serve as the foundation for training and validating PCA and FDA models. In addition, an assessment of fault detection and diagnosis methods was rigorously performed using the confusion matrix. This metric provided a quantitative evaluation of the classification performance under both normal and faulty conditions.

5.2. Normal and abnormal operation data sets

To generate the necessary data sets for fault detection or diagnosis, thirteen distinct simulation scenarios were built to thoroughly examine these operational circumstances (normal and abnormal operation) of the WWTP, in order to produce the datasets required for the subsequent analysis.

The calibrated WWTP model was simulated for a time period of 168 days, for both normal and faulty scenarios, resulting in a set of fifteen different simulations. To reach the plant's quasi-stationary state, the simulations were carried out under normal operating conditions for the first 139 days. For the normal operation scenario, the simulation was carried further, up to a total of 168 days,

without the introduction of a fault. For the malfunctioning scenarios the sensor faults were deliberately introduced in the 140th day of each simulation. In total, twelve simulations were performed individually, each corresponding to a different type of fault. Simulated data of the WWTP operation were collected with a sampling time of 15 minutes to create comprehensive data sets for further analysis.

The obtained data sets were further used to build, train, validate and test the PCA and FDA models. The objective of these machine learning techniques was to improve the reliability and robustness of WWTP monitoring systems by evaluating the efficacy of fault detection and identification models.

5.3. Data sets for building the PCA fault detection model

Data collected via the simulation process over a period of 40 days of normal operation, specifically from day 100 to day 139, were used to build the data matrix X of the PCA model. This time frame was selected to ensure that the model was trained under consistent, error-free operating conditions, enabling it to capture the inherent variability of the wastewater treatment plant in normal conditions. With the help of this dataset, the PCA model was able to detect normal operating patterns and establish a baseline (threshold) that could be used to detect deviations caused by sensor faults.

In order to simulate various types of sensor malfunctions, artificial faults were introduced into the system at the beginning of day 140.

The fault detection capacity of the PCA model was investigated using data generated and recorded from the 140th day till 160th day (a time frame of 20 days), separately for each of the DO and NO sensor faulty operation. The statistical measures Hotelling's T^2 and Squared Prediction Error, were used to track the system's behaviour by projecting the faulty operational data onto the principal component space. The model was able to detect possible abnormalities by using significant deviations in these statistical indices as indicators of sensor malfunctions.

5.4. Data sets for building the FDA fault identification model

The FDA model was trained using data generated between the 141st and 145th days (five days) of either normal or faulty DO and NO sensor operation. In the case of faulty operation, the data sets used for training included instances for each of the fault types. As a result, the observation training matrix for the DO sensor operation is composed of eight separate classes, with one class representing normal operation and the remaining seven representing each of the various fault types. For the NO sensor operation, the observation training matrix consists of six distinct classes, one for the normal operating conditions and five for the faulty operating conditions.

The trained FDA model's fault identification performance was evaluated using data from the first day of abnormal sensor operation (the 140th day). This testing strategy was developed to investigate the ability of the FDA diagnosis method to identify the type of fault within the early hours after its appearance. This assessment offered significant insight on the efficiently of the diagnosis method and

its application potential for real-time working scenarios.

5.5. Assessment metrics for fault detection and diagnosis performance

This study's performance on fault detection and diagnosis capabilities is based on the assessment metrics and methodology. This includes the construction and interpretation of confusion matrices, as well as the calculation of key performance indicators such as accuracy, PPV, MAR and F1 score. The confusion matrix was computed and systematically applied to each specific scenario to precisely assess the effectiveness of the fault detection and identification techniques.

In the confusion matrix, the rows correspond to the predicted categories (Output Class), whereas the columns denote the actual categories (Target Class). Accurately classified instances are placed along the diagonal cells, while misclassified instances appear in the off-diagonal cells. Each cell displays both the count of instances and the corresponding percentage of the total observations. The rightmost column of the matrix shows the proportion of cases predicted for each category that were classified correctly or incorrectly, aligning with metrics such as precision (or PPV) and the false discovery rate. Likewise, the bottom row displays the proportion of actual instances of each category that were classified correctly or incorrectly, corresponding to recall (or true positive rate) and MAR. The overall classification accuracy is displayed in the bottom-right cell of the matrix.

6. DO sensor fault detection, diagnosis and impact on energy, water quality and GHG emissions

6.1. Introduction

The simulation framework and the methodology designed to detect and diagnose faults of the DO concentration sensor with the support of the dynamic simulator for a municipal WWTPs are presented in this chapter. Particular simulations were developed for both, normal and faulty operation. In the case of faulty operation, seven different faults of the DO concentration sensor were considered: bias, drift, wrong gain, loss of accuracy, fixed value, complete failure minimum and maximum. Each fault was implemented using an individual procedure that accurately replicated actual DO sensor faults. Two statistical process monitoring techniques were considered and PCA-based and FDA-based models were developed to detect and diagnose the considered DO sensor faults.

The overall economic and environmental impact of the DO sensors faults in the municipal plant is also examined. An extensive analysis is performed with special attention to energy usage and costs, effluent quality and GHG emissions. This study emphasises the necessity of preventive maintenance, periodic calibration, and automated fault detection systems by revealing the correlating sensor faults with plant performance indices.

6.2. Methodology of DO sensor faults implementation

To simulate the behaviour of the seven failure types considered for the DO concentration sensor,

customized software modules were created. They were integrated into the WWTP dynamic simulator model. For each type of fault, the generated fault-affected output of the sensor was sent as measured DO process variable to the Proportional-Integral DO controller. The layout of the WWTP and the DO sensor location are presented in Figure 6.1.

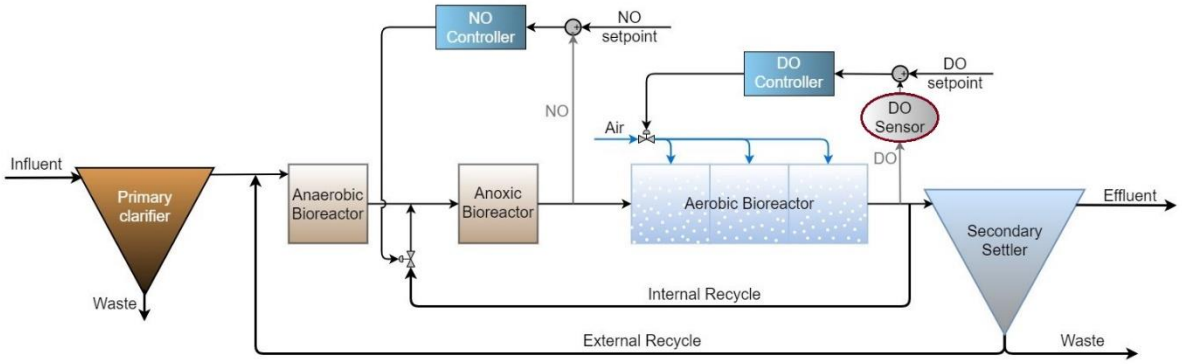

Figure 6.1. Placement of the DO sensor in the municipal WWTP

The DO control loop adjusts the air flowrate to bring the measured DO to the desired setpoint of 2 mg O_2/L, with zero offset. For each of the faults, the investigation scenario was carried out as follows. For a period of 139 days, the simulation was run without any fault. Then, beginning on the 140th day of the simulations, each fault was implemented by an individual fault operating scenario for a total of 28 days. The process variables considered for fault identification were taken from the first day of faulty operation, i.e., the day no. 140. For collecting data that describe normal operation, the same measurement period of 28 days was considered.

The bias fault type was defined as a predetermined constant difference between the sensor's true and faulty output values. The DO measured concentration was considered to have a bias of +1.5 mg O_2/L for this fault type.

The drift was simulated as an increasing bias, which was described by a time-varying ramp signal added to the true DO value. To simulate the drift, a constant value of 0.05 mg O_2/L was integrated in time and added to the DO true value.

The wrong gain fault was simulated by the sensor's loss of effectiveness because of a gain factor other than 1. It consisted of an incorrect calibration gain of 1.4, which resulted in a wrong correlation factor between the sensor's input and output signal. A first-order filter with a time constant of 0.3 days smoothly introduced the incorrect gain in time.

The loss of accuracy fault was simulated using a uniform random-number generator that delivered a value in the interval [-2.5, 2.5] with a sampling time of 0.1 days. This signal was passed through a first-order filter with a time constant of 0.01 days, and then to a saturation block, which limited its output value to the positive interval [0, 6]. This random signal was then added to the DO true value to generate the fault-affected measured process variable.

The fixed value fault was implemented by a constant value of the sensor output signal, regardless of the true DO process variable value. For this study, a constant value of 2.2 mg O_2/L was

used (which is 10% higher than the DO control loop's nominal setpoint value of 2 mg O_2/L).

A sensor's complete failure was once again characterized by unvarying values of the sensor output signal, but these constant values were the minimum or maximum of the sensor's calibration interval. For the complete failure minimum and maximum alternatives, the two extreme values of 0.1 mg O_2/L and 6 mg O_2/L were chosen. They represent 5%, respectively 300% of the nominal setpoint value of 2 mg O_2/L considered for the DO control loop.

The application of statistical theory in process monitoring is based on the assumption that the characteristics of the data change only when an unexpected situation (fault) takes place in the process. Any significant deviation from this normal behaviour suggests that an unexpected event or fault has occurred in the system (Chiang *et al*., 2001). It was also assumed that a single sensor would fail at a time.

The purpose of this chapter is to support the field of wastewater treatment plant operation and management by investigating the DO concentration sensor typical faults, the detection and identification of the faults and revealing how these issues affect the energy use, gas emissions and quality of effluent water. This work was published in two scientific papers (Luca *et al*., 2023; Luca *et al.*, 2021).

6.3. Process variables considered for the detection and identification of the DO sensor faults

The total set of twenty WWTP process variables considered for building the PCA model consisted of secondary settler flow rate and clean water effluent concentration variables (6 variables), and sludge flow rate and settler bottom effluent concentration variables (14 variables), including temperature. The entire set of variables taken into consideration were: total nitrogen (N_{total}), total Kjeldahl nitrogen (TKN), chemical oxygen demand (COD), nitrate and nitrite nitrogen (S_{NO}), free and saline ammonia (S_{NH}), total suspended solids (TSS), slowly biodegradable substrate (X_S), heterotrophic biomass ($X_{B,H}$), autotrophic biomass ($X_{B,A}$), inert particulate products (X_P), particulate biodegradable organic nitrogen (X_{ND}), soluble biodegradable organic nitrogen (S_{ND}), dissolved oxygen concentration (S_O), readily biodegradable substrate (S_S), alkalinity (S_{alk}), waste flow rate (Q_W), and temperature (t). The first six variables of the presented set were considered for the settler clean water effluent and the last fourteen for its bottom stream. Inert suspended organic matter (X_I) and inert soluble organic matter (S_I) concentration variables were excluded, as they would not add significant information to the study (Tomita *et al*., 2002). These variables were chosen based on relative standard deviation (RSD) values obtained during the faulty time period of simulations. The RSD is a measure of the dispersion of data points in a data set relative to its mean. It was used as an indicator to determine which variables provided meaningful variation. The higher the value of RSD, the higher the variability, which meant that the data points were widely spread. Values of S_{NO}, S_O and Q_W were

excluded from the FDA sets as they induced singularity on the S_w matrix (the determinant of the matrix was equal to zero and the matrix had no inverse).

6.4. PCA model construction for DO sensor faults detection

The PCA model was built using 3840 observations from the normal operating period of 40 days. The obtained data matrix *X* of 3840x20 dimensions represented the training matrix. After scaling the data matrix, the scores matrix *T* and the loadings matrix *P* were obtained. For CPV_k, a threshold value of 98.24% was chosen. As a result, the number of principal components *k* that contributed to a consistent representation of the entire data set was determined to be eight. A T^2 threshold of 20.16 and a *SPE* threshold of 1.26 were determined for the 99% confidence level. Any vector *x* of measured variables with a T^2 or *SPE* value greater than the two equivalent thresholds implied the abnormal sensor operation.

Seven testing data matrices were created using the vectors of abnormal-operation measured variables. One was created for each of the first five types of faults, to which was added the complete-failure fault type, which had two matrices built for the minimum and maximum faulty values of the DO sensor. For each testing data matrix (1920x20), a total of 1920 samples were used, originating from simulation day 140 up to day 160 (excluding day no. 160). To determine faulty sensor operation, T^2 and *SPE* values were determined for each of the sample vectors of the testing matrices and compared to previously calculated thresholds. T^2 and *SPE* values were graphically represented for each of the normal (training) and faulty (testing) samples, in association with the T_α^2 and SPE_α thresholds.

6.5. FDA model construction for DO sensor faults identification

Data generated during the period of the 141st and 145th days of normal and faulty DO sensor operation were used to train the FDA model. A class was defined as the set of 480 observations of the 17 variables. These variables created the observation training matrix of 3840 lines (8 classes with 480 measurements) and 17 columns. The fault diagnosis performance of the trained FDA model was evaluated using data from the 140th day, i.e., the first day of abnormal operation. Each case of normal or abnormal operation had 96 measurements in the testing data set. This testing strategy was devised to investigate the ability of the FDA diagnosis method (model) to identify the type of fault within the first few hours following its appearance.

The values obtained for each class's discriminant function, g_i, were compared in order to diagnose the faulty sensor operation. The values of the discriminant functions g_i were computed for each of the 15-minute time-sampled measurements of testing day no. 140, which were affected by the various types of faults (7 classes), as well as measurements corresponding to normal operation (1 class). The discriminant function with the highest value identified the faulty sensor's class and, as a result, diagnosed the fault type.

6.6. Results and discussions

6.6.1. Assessment metrics for fault detection of DO sensor

The effectiveness of the PCA-based and FDA-based methods for detecting and diagnosing DO sensor faults was assessed using two different confusion matrices. This approach allowed for a thorough evaluation of the models' performance by methodically examining their classification accuracy. This comprehensive assessment ensured an accurate and dependable analysis of sensor functioning by offering a quantitative estimate of the models' capabilities.

The DO sensor detection confusion matrix analyses two different classes: abnormal (with defect) and normal (no defect). The normal class is shown in the second column (target), whereas the abnormal class is represented in the first column (target). Figure 6.2 presents the DO sensor confusion matrix for faults detection.

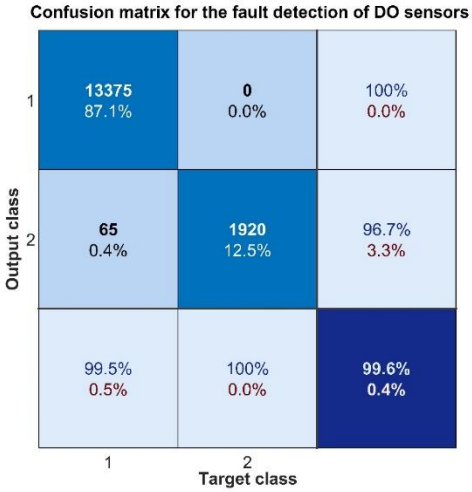

Figure 6.2. Confusion matrix of the fault detection for the DO sensor

With a 100% precision rate for the normal class, the model showed no false positives. For the faulty class the precision was 96.7%. Furthermore, the model's overall accuracy was 99.6%, demonstrating high proficiency in correctly classifying observations. The F1 score for the normal class was 0.983, while the faulty class achieved a remarkable F1 score of 0.997. These scores indicate an exceptional balance between recall and precision, with the faulty class performing particularly well.

6.6.2. Fault detection of DO sensor faults

First, the proposed and developed PCA model was applied to the normal data set. As shown in Figure 6.3, all of the T^2 and *SPE* values from this data set were validated to be faulty-free values, therefore they were classified as normal operation values. Furthermore, the T^2 plot reveals that the difference between the normal operation values and their associated threshold was greater for T^2 than that shown by the *SPE* plot. This means that in normal operation mode, the *SPE* statistics provide the most reliable data.

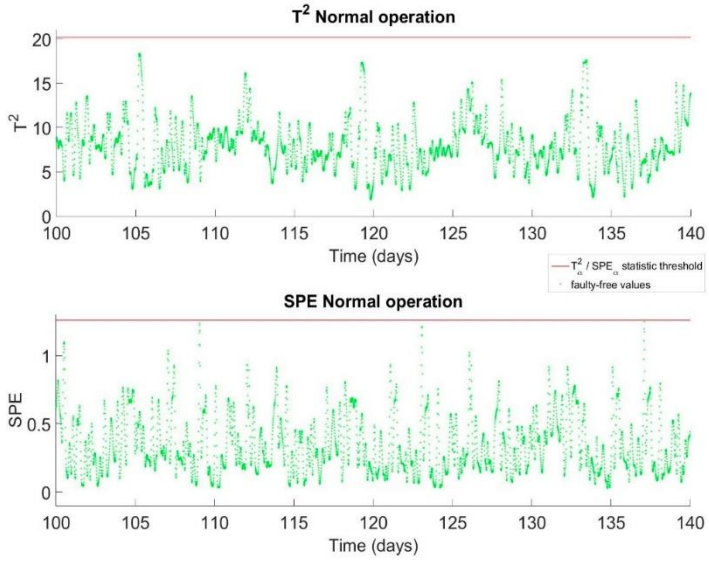

Figure 6.3. T^2 and *SPE* plots of PCA monitoring the normal operation data set for the DO sensor

Figures 6.4-6.10 show the fault detection results for each type of the DO sensor investigated faults.

The bias fault type was detected after 1h with T^2 statistics and in 1 h and 15 minutes with *SPE* statistics, as it can be observed in Figure 6.4.

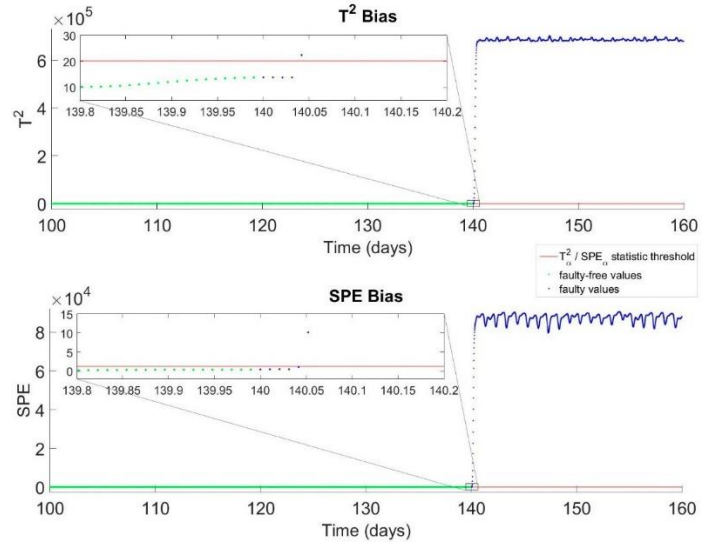

Figure 6.4. T^2 and *SPE* plots of PCA detecting the bias type of fault for the DO sensor, with details on the period next to the fault occurrence moment

Figure 6.5 shows the detection of the drift type of fault. The plots confirm the existence of a fault after 8.5 h with the T^2 method. In this case *SPE* method proved to be more rapid and detected the fault in 5.25 h. The drift type of sensor fault was expected to take longer to be detected because the faulty signal grows slowly over time and crosses the statistical threshold only when its amplitude becomes significant.

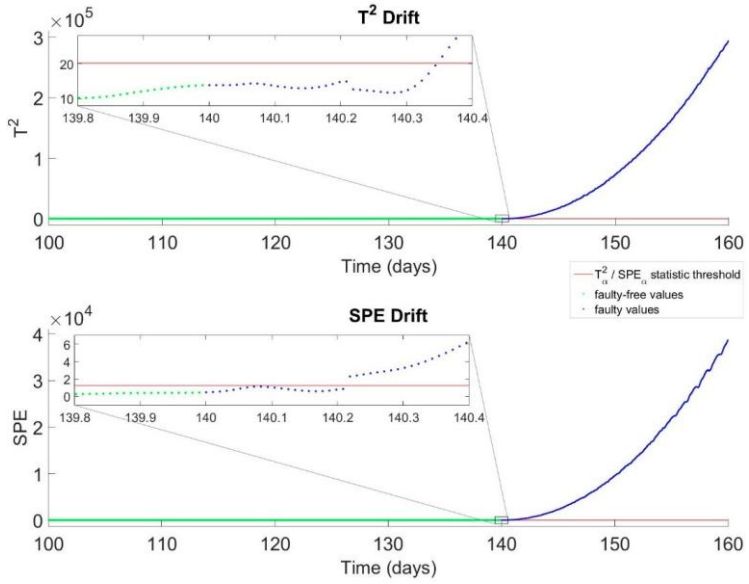

Figure 6.5. T^2 and *SPE* plots of PCA detecting the drift type of fault for the DO sensor, with details on the period next to the fault occurrence moment

For wrong gain fault type, both statistics methods proved to be just as fast, and the fault was detected in 2 h and 30 minutes. The results for the detection of the wrong gain fault of the DO sensor are presented in Figure 6.6.

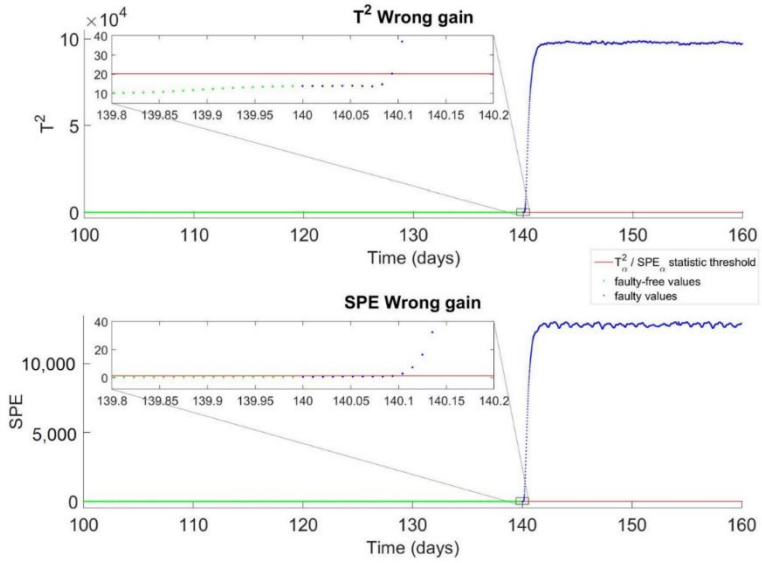

Figure 6.6. T^2 and *SPE* plots of PCA detecting the wrong gain type of fault for the DO sensor, with details on the period next to the fault occurrence moment

Figure 6.7 presents the results for the loss of accuracy DO sensor type of fault. In this case, the fault presence was detected in 3 h and 45 minutes by both, T^2 and *SPE* methods. In the case of this fault, the detection is dependent on the random amplitude of the fault component value.

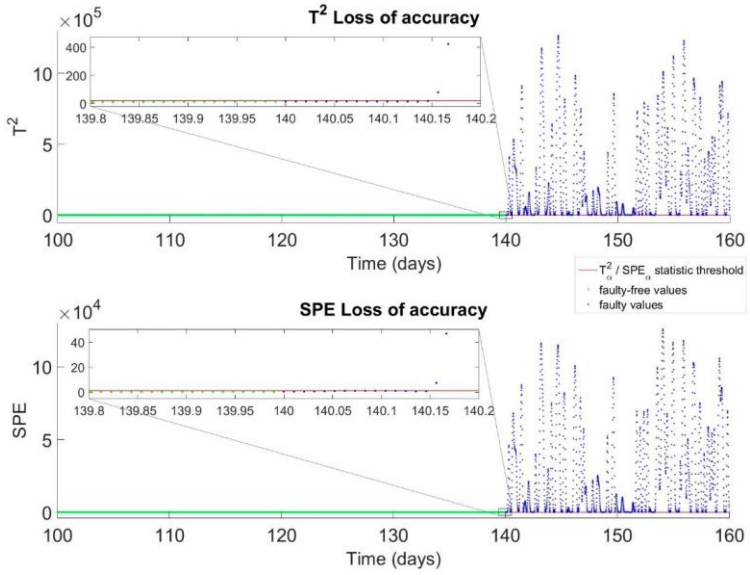

Figure 6.7. T^2 and *SPE* plots of PCA detecting the loss of accuracy type of fault for the DO sensor, with details on the period next to the fault occurrence moment

The detection of the fixed-value type of fault was detected in 1.25 h. For this fault, both methods proved to be just as efficient. Figure 6.8 presents the results for the fixed value DO sensor fault.

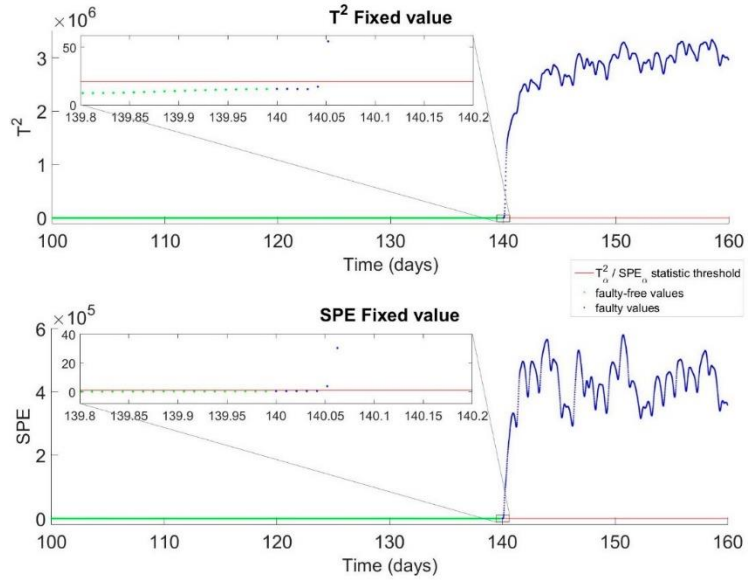

Figure 6.8. T^2 and *SPE* plots of PCA detecting the fixed value type of fault for the DO sensor,with details on the period next to the fault occurrence moment

Again, in the complete failure minimum case, the T^2 statistic proved to be as fast as *SPE* statistic and the detection time was 1 h. The results for the complete failure minimum type of fault for the DO sensor are presented in Figure 6.9.

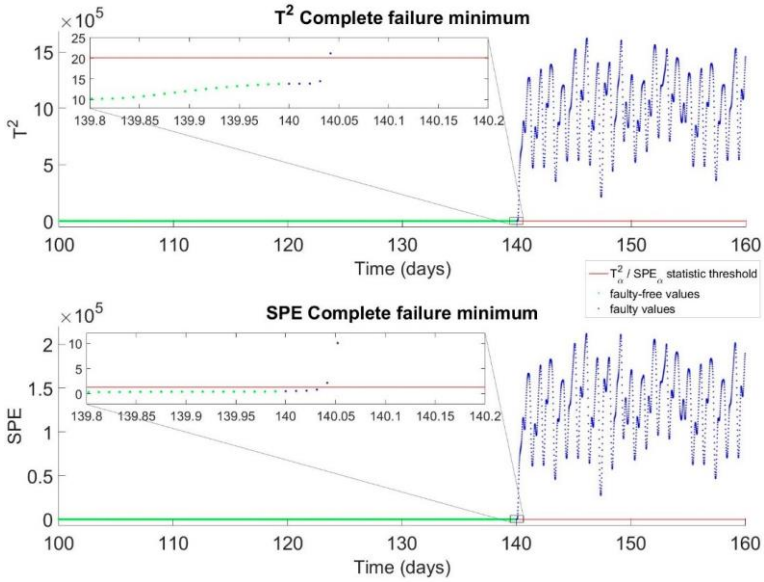

Figure 6.9. T^2 and *SPE* plots of PCA detecting the complete failure minimum type of fault for the DO sensor, with details on the period next to the fault occurrence moment

As shown in Figure 6.10 complete failure maximum was detected after 1 h, similar to the complete failure minimum and the bias type of faults.

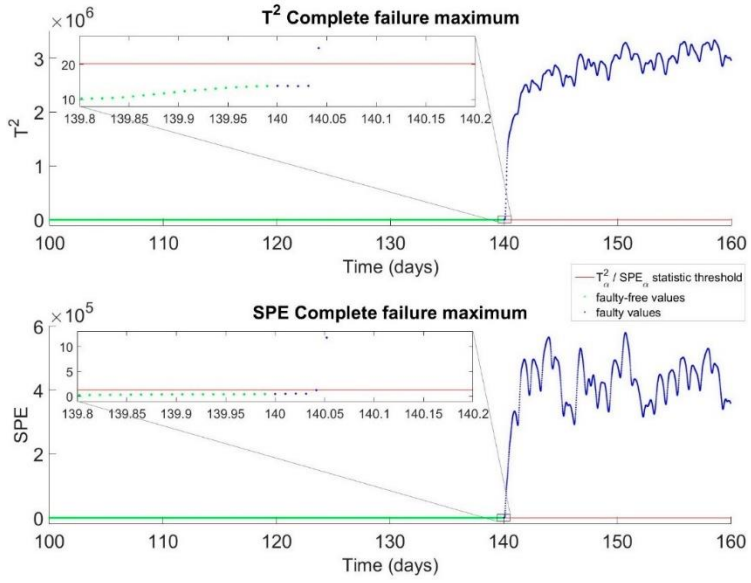

Figure 6.10. T^2 and *SPE* plots of PCA detecting the complete failure maximum type of fault for the DO sensor, with details on the period next to the fault occurrence moment

The proposed PCA model successfully detected the occurrence of different types of DO sensor faults. *SPE* usually provided faster detection than the T^2 method, with a higher promptitude especially for the drift type of fault. The efficiency of T^2 and *SPE* statistics was comparable for bias, wrong gain, loss of accuracy, fixed value, and complete failure type of defects. These results demonstrate the capability of the designed PCA-based method for early detection of DO sensor failure by prompt fault detection and providing reliability to the efficient operation.

6.6.3. Assessment metrics for DO sensor faults diagnosis

For the FDA-based identification method, a complete confusion matrix was computed to evaluate the method's performance in classifying different sensor types of faults. This matrix's normal class (fault-free operation, designated as target 1) is represented by the first column. Listed in order of occurrence, the following columns represent the following seven different sensor faults: bias (target 2), drift (target 3), wrong gain (target 4), loss of accuracy (target 5), fixed value (target 6), complete failure minimum (target 7) and complete failure maximum (target 8).

As shown in Figure 6.11, the FDA technique showed good classification performance, with an overall accuracy of 85.8%. This value demonstrates the ability of the FDA-based method to accurately identify different types of sensor errors.

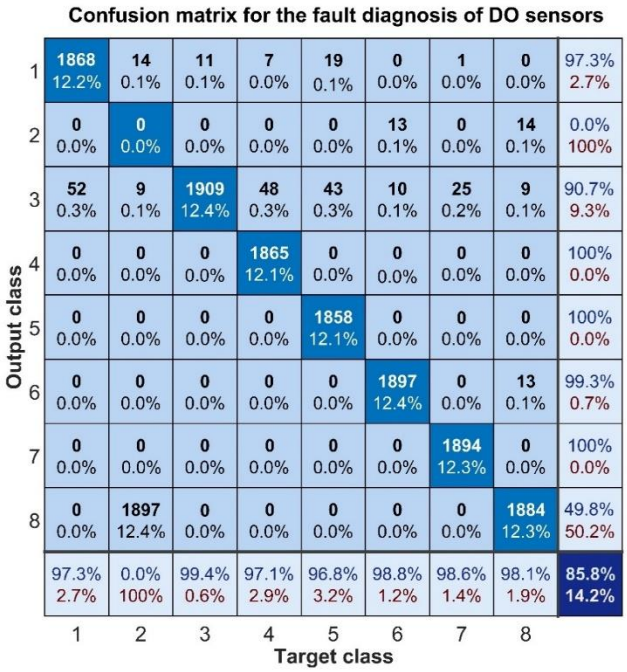

Figure 6.11. Confusion matrix of the fault diagnosis for the DO sensor type of faults

6.6.4. Fault diagnosis of the DO sensor type of faults

As shown in Figure 6.12, the g_i values revealed the consolidated normal operation (fault-free values) diagnosis after 16.5 hours (confident identification moment). In this case, a small set of 52 observations was misclassified as drift, resulting in a MAR of 2.7%. Additionally, the F1 score was calculated to 0.973, reflecting a good balance between precision and recall.

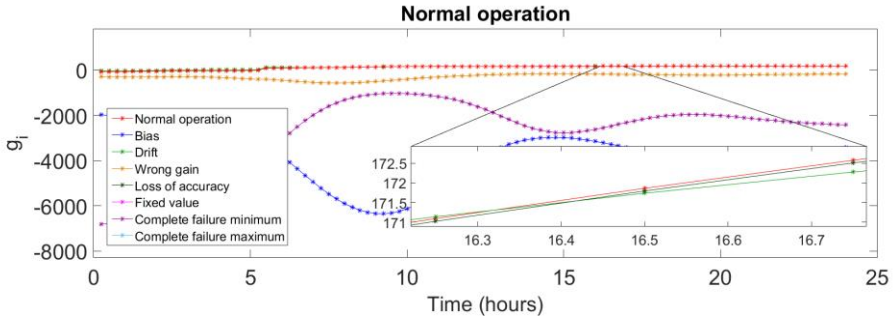

Figure 6.12. Faulty-free operation diagnosis for the DO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and seven types of faults; details on the time period of confident identification

The identification of the sensor types of fault is presented in Figures 6.13-6.19.

As shown in Figure 6.13, the bias fault type diagnosis is confirmed after 2.5 hours, for a total of 5.75 hours. As a result, the FDA discrimination reveals the complete failure maximum class for the remaining period of the day. This resulted in a MAR value of 100% and a F1 score of 0.

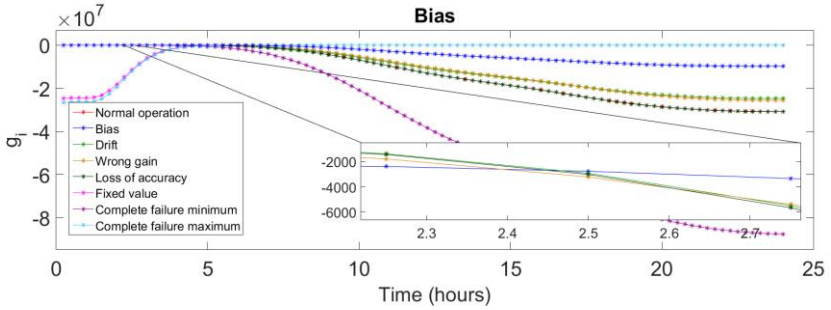

Figure 6.13. Bias type of fault diagnosis for the DO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and seven types of faults; details on the time period of confident identification

As shown in Figure 6.14, the diagnosis of drift type of fault is firmly acknowledged after 13.75 hours (confident identification moment). In this instance, the MAR was only 0.6%. This low percentage of misclassification demonstrates how well the model can discriminate between faulty and fault-free conditions. Additionally, the F1 score demonstrated a well-balanced performance between precision and recall, reaching the value of 0.949.

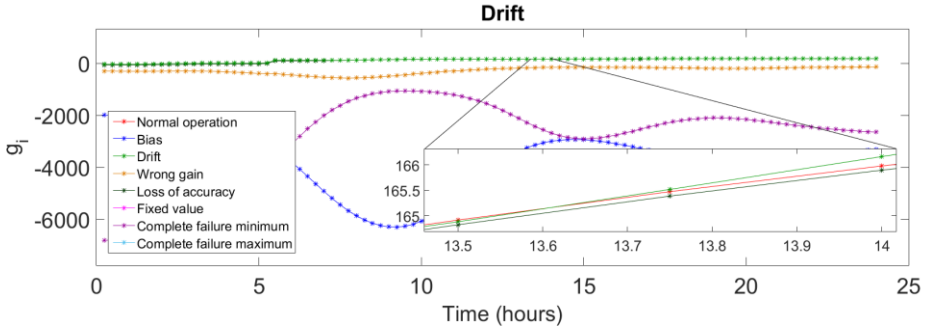

Figure 6.14. Drift type of fault diagnosis for the DO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and seven types of faults; details on the time period of confident identification

After 14 hours of the fault incidence moment, the diagnosis of wrong gain fault was confirmed.

Figure 6.15 depicts the graphical representation for the diagnosis of this fault type. The model's overall efficacy in correctly identifying observations was demonstrated by its F1 score of 0.985.

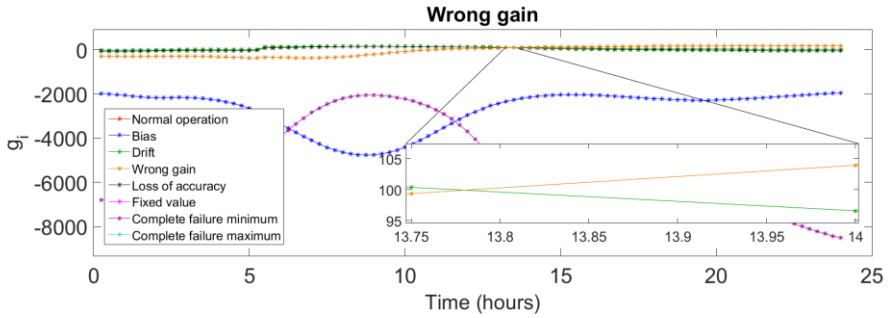

Figure 6.15. Wrong gain type of fault diagnosis for the DO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and seven types of faults; details on the time period of confident identification

Figure 6.16 shows that the loss of accuracy fault is identified after 16.5 hours from the starting moment of the fault action. Nevertheless, 62 observations were misclassified, with 19 being mistakenly classified as normal operation and 43 as drift, resulting in a MAR of 3.2%.

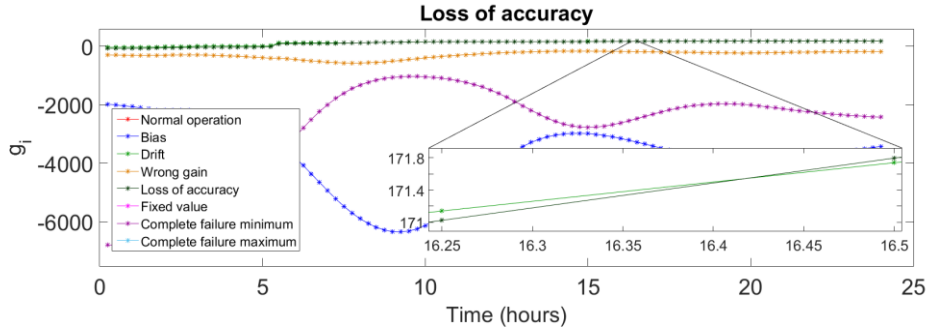

Figure 6.16. Loss of accuracy type of fault diagnosis for the DO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and seven types of faults; details on the time period of confident identification

Figure 6.17 shows the graphical representation of the fixed value fault. This fault type was accurately diagnosed after 6 hours (confident identification moment). With the second-highest F1 score of 0.990, the model showed remarkable performance for identifying the fixed value fault. A good balance between precision and recall is revealed by this high score, which further supports the reliability of the categorisation.

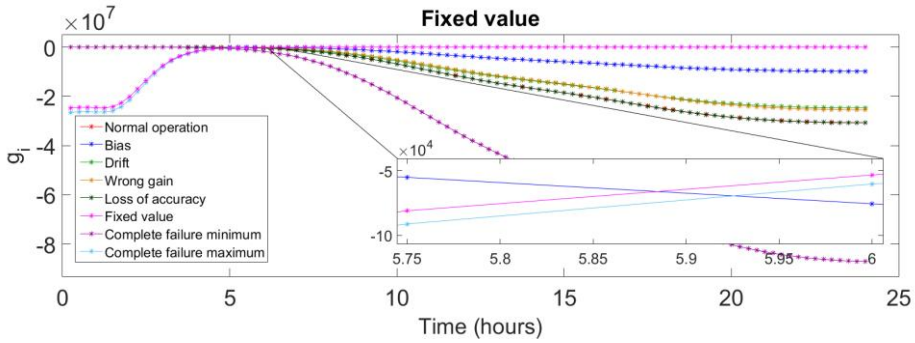

Figure 6.17. Fixed value type of fault diagnosis for the DO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and seven types of faults; details on the time period of confident identification

Figure 6.18 shows that the complete failure minimum is correctly diagnosed after 6.75 hours (confident identification moment) of its appearance. The identification of complete failure was most accurately achieved, as indicated by the highest F1 score of 0.993. The MAR was 1.4%.

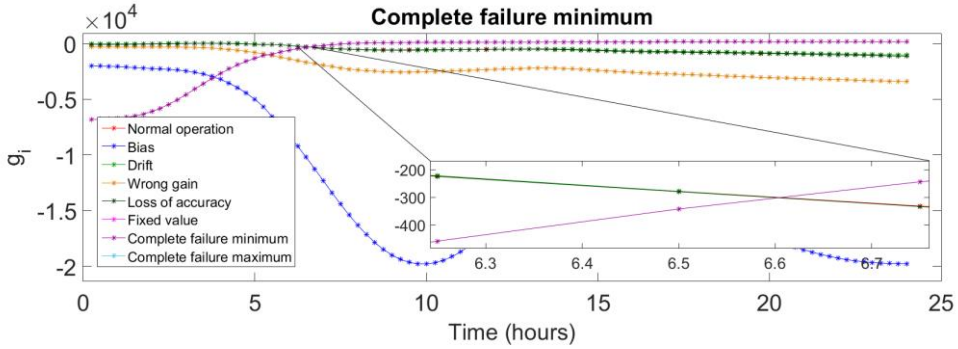

Figure 6.18. Complete failure minimum type of fault diagnosis for the DO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and seven types of faults; details on the time period of confident identification

As shown in Figure 6.19, the complete failure maximum is successfully diagnosed after 9.5 hours (confident identification moment). For complete failure maximum, the identification approach showed less accurate results, with a value of 1.9% for MAR.

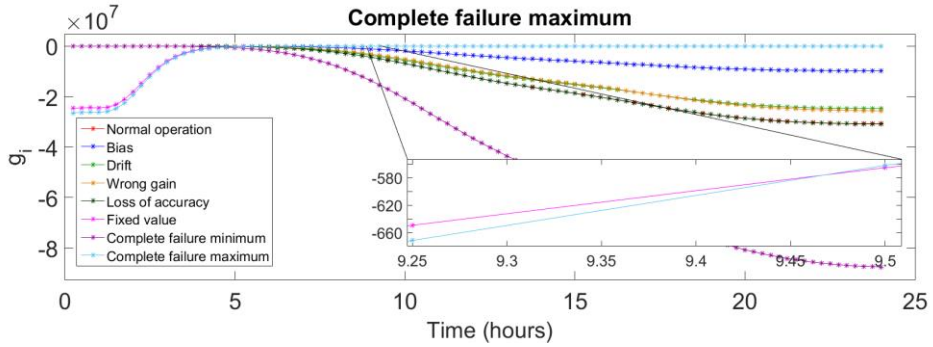

Figure 6.19. Complete failure maximum type of fault diagnosis for the DO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and types of seven faults; details on the time period of confident identification

The FDA-based method for DO sensor faults offers a trustworthy framework for diagnosing sensor faults in a timely manner and efficiently monitoring the development of faults.

6.6.5. WWTP performance evaluation during operation with DO sensor faults

For both normal and faulty operation scenarios of the DO sensor faults, the performance indices AE, PE, and EQ were determined as mean values over 28 days (starting from the 140th, up to the 168th day). The indices value for normal operating conditions and for the six different defects are shown in Table 6.1 as AE, PE, and EQ performance indicators.

Operating regime	AE (kWh/day)	PE (kWh/day)	Total energy demand (kWh/day)	EQ (kg PU/day)
Normal operation	16,992	1,329	18,321	16,852
Bias fault	14,206	2,415	16,621	21,461
Drift fault	15,569	1,746	17,315	17,134
Wrong gain fault	15,866	1,593	17,459	16,706
Loss of accuracy fault	9,150	2,403	11,553	219,189
Fixed value fault	1,968	2,415	4,383	338,737
Complete failure minimum fault	23,537	1,039	24,576	19,804
Complete failure maximum fault	1,968	2,415	4,383	338,750

Table 6.1. Values of the performance indices for the normal and faulty operation cases

It is evident that the PE and EQ indices values for the fault types of bias, loss of accuracy, fixed value, and complete failure maximum are significantly greater than the values for normal operation, although AE is lower. The high DO sensor signal values of the previously four fault categories, which are delivered to the DO controller as feedback, justify the AE values. As a result of the increased by fault signal, the DO controller reduces aeration in an effort to lower the value of the incorrect DO signal. This results in a reduction in aeration energy. After that, nitrification declines, which lowers the concentration of nitrate and nitrite in the aerobic reactor and, ultimately, in the anoxic reactor. The internal recycle flowrate is increased by the NO control loop in response to this shift, which results in a higher pumping energy need. The increased air flow rate controlled by the DO controller in response to the low, but incorrect value of the DO sensor signal is what causes the high AE value in the case of the complete failure minimum type of fault.

6.6.6. Energy costs assessment for operation affected by the DO sensor faults

In general, operating a WWTP involves significant energy costs and, implicitly, considerable economic expenses. Geopolitics, the nation's unique energy mix, distribution network costs, environmental protection fees, extreme weather conditions, and excise and taxation levels are only a few of the supply and demand elements that affect energy prices. The energy source also has an impact on the cost of energy. According to Table 6.2, based on the source, energy costs range from 5.84 to 18.75 eurocents. These are average values calculated based on two different reports on energy

costs in Europe (Kost *et al*., 2024), United States and global markets (Lazard, 2024) and on an online calculator made available by the International Energy Agency (IEA) (IEA, 2024).

Table 6.2. Average costs of different energy sources

Source/Technology	**€cents/kWh**
Lignite	9.89
Coal	9.37
Gas (CCGT[1])	10.78
Nuclear	18.75
Wind onshore	6.75
Wind offshore	8.55
Solar PV[2] commercial	8.07
Solar PV residential	12.70
Solar thermal (CSP[3])	14.8
Hydro reservoir	7.95
Hydro run of river	5.84
Geothermal	8.63
Biomass	11.57

[1]CCGT - combined cycle gas turbines

[2]PV - photovoltaic

[3]CSP - concentrating solar power

For the WWTP taken into consideration in the current study, the daily cost for regular-normal and for each type of faulty operation modes was determined. The daily costs for different energy sources were evaluated based on energy production technologies. The costs of various energy sources were multiplied by the total energy demand values calculated for both normal and faulty operating scenarios to generate the data shown in Table 6.3.

Table 6.3. Source depending sum of AE and PE energy costs computed for normal and fault affected operation of the WWTP water line for the DO sensor case

Source/Technology	Daily operation costs (€)							
	Normal operation	**Bias**	**Drift**	**Wrong gain**	**Loss of accuracy**	**Fixed value**	**Complete failure minimum**	**Complete failure maximum**
Lignite	1,812	1,644	1,712	1,727	1,143	433	2,431	433
Coal	1,717	1,557	1,622	1,636	1,083	411	2,303	411
CCGT	1,975	1,792	1,867	1,882	1,245	472	2,649	472
Nuclear	3,435	3,116	3,247	3,274	2,166	822	4,608	822
Onshore wind	1,237	1,122	1,169	1,178	780	296	1,659	296
Offshore wind	1,566	1,421	1,480	1,493	988	375	2,101	375
Solar PV commercial	1,479	1,341	1,397	1,409	932	354	1,983	354
Solar PV residential	2,327	2,111	2,199	2,217	1,467	557	3,121	557
Solar thermal (CSP)	2,712	2,460	2,563	2,584	1,710	649	3,637	649
Hydro reservoir	1,457	1,321	1,377	1,388	918	348	1,954	348
Hydro run of river	1,070	971	1,011	1,020	675	256	1,435	256
Geothermal	1,581	1,434	1,494	1,507	997	378	2,121	378
Biomass	2,120	1,923	2,003	2,020	1,337	507	2,843	507

Eroare! Utilizați fila Pornire pentru a aplica Heading 2 la textul care doriți să apară aici.

When examining the costs associated with electrical energy used for the various fault cases, it can be seen that the complete failure minimum fault of the DO sensor is associated with the highest energy costs, as the control system is forced to significantly increase the air flowrate due to the reduced faulty value of the DO sensor. On the other hand, the large values of the DO sensor signal associated to the complete failure maximum or the (high) fixed value type of faults, cause the DO controller to limit the air flowrate, resulting in low energy costs. However, in these latter situations, the effluent quality rapidly declines by a factor of more than 20, making these defects the ones with the worst overall effects.

6.6.7. Environmental assessment of CO_2 and N_2O emissions for operation with DO sensor faults

For each case of faulty operation, as well as for normal operation, the on-site and off-site emissions of CO_2 and N_2O, the principal contributors to the Green House Gases produced by the water line of the WWTP, were estimated. Table 6.4 displays their daily mean values.

Table 6.4. GHG Emissions due to the different types of the DO sensor defects

Site of emissions	Process	Type of emitted gas	Type of fault							
			Normal	**Bias**	**Drift**	**Wrong gain**	**Loss of accuracy**	**Fixed value**	**Complete failure minimum**	**Complete failure maximum**
Off-site emissions	Power generation	CO_2, $P_{CO_2,off-site}$ (kg CO_2/day)	3,481	3,158	3,290	3,317	2,195	833	4,669	833
	Biological degradation in the WWT effluent	N_2O, $P_{N_2O,off-site}$, (kg N_2O/day)	3.61	2.47	2.89	2.97	14.39	21.49	6.50	21.49
On-site emissions	Water-line aerobic biological processes	CO_2, $P_{CO_2,on-site}$(kg CO_2/day)	13,689	30,459	17,851	16,178	461,439	921,028	10,604	921,168
		N_2O, $P_{N_2O,on-site}$ (kg N_2O/day)	10.35	10.81	10.64	10.05	6.07	3.27	9.20	3.27
Total emissions		CO_2, $P_{CO_2,total}$(kg CO_2/day)	17,170	33,617	21,141	19,495	463,634	921,861	15,274	922,001
		N_2O, $P_{N_2O,total}$ (kg N_2O/day)	13.96	13.28	13.53	13.02	20.46	24.76	15.70	24.76
Total overall emissions		CO_{2e}, $P_{CO_{2e},overall}$ (kg CO_{2e}/day)	21,330	37,574	25,173	23,375	469,731	929,239	19,953	929,379

According to the GHG assessment results presented in Table 6.4, on-site emissions are the most important, accounting for 75%-80% of the total emissions during normal operation, for both $P_{CO_2,total}$ and $P_{N_2O,total}$. With the exception of the complete failure minimum fault type, all cases of the DO sensor defects resulted in increased values for the computed total CO_2 emissions, that is, the sum of the on-site and off-site values.

Increased values for the N_2O emissions are also found when the total N_2O emissions are assessed for the fixed value, complete failure maximum, and loss of accuracy faults. It should be noted that complete failure minimum fault type led to higher values of the N_2O emissions, which is the opposite of the observed trend for the total CO_2 emissions. The total CO_2 and total N_2O emissions were summed together to calculate the overall CO_2e emission values. All faults exhibit higher CO_2 emission levels than the case of normal operation, according to the results of the overall CO_2 emission. The single exception to this trend is the complete failure minimum type of fault. However, in this specific instance the effluent quality deteriorates by more than 17.50% and the sum of the aeration and pumping energy has the highest values of all analysed situations.

6.7. Conclusions

Using the DO sensor of a municipal A^2O WWTP as a case study, PCA-based fault detection and FDA-based fault diagnosis methods were evaluated, alongside energy and water quality performance. The DO sensor is central to nitrification efficiency, with aeration being the main energy consumer. Six fault types were analysed: bias, drift, wrong gain, loss of accuracy, fixed value, and complete failure. PCA detected faults with 99.6% accuracy, typically within 2-9 hours depending on fault type, while FDA diagnosis reached 85.8% accuracy within 2.5-16.5 hours.

Faults impacted effluent quality and energy differently: bias and drift moderately reduced energy but worsened water quality, while loss-of-accuracy and fixed-value faults caused severe effluent deterioration. Complete failure faults showed extremes, with either high energy demand or drastically reduced effluent quality. CO_2 and N_2O emissions were highest from on-site sources, particularly under complete failure, fixed-value, and loss-of-accuracy faults. Energy costs were also fault-dependent, with low DO readings increasing aeration demand.

The study demonstrates that PCA and FDA approaches effectively detect and diagnose DO sensor faults, supporting safe, efficient, and sustainable WWTP operation. Results inform control system design, energy optimization, GHG mitigation, and intelligent sensor development.

7. NO sensor fault detection, diagnosis and impact on energy, water quality and GHG emissions

7.1. Introduction

Reliable sensor measurements are essential for regulatory compliance and optimal process control in wastewater treatment plants. Among the many sensors employed in WWTP operations,

the NO concentration sensor is essential for monitoring the dynamic concentrations of nitrate and nitrite. Nitrogen compounds present in WWTPs can cause significant environmental issues.

This study focuses on detecting and identifying different NO sensor errors. Specific types of faults of the NO concentration sensor were incorporated into the WWTP dynamic model and several different simulations were developed to reveal the plant behaviour as a consequence of these faults. Five different fault types were considered: bias, drift, wrong gain, loss of accuracy and fixed value. To effectively detect these faults a PCA-based model was developed, validated and tested. Another model, based on FDA was implemented and used to diagnose the NO sensor type of faults.

Furthermore, the confusion matrix was employed to thoroughly evaluate fault diagnosis and detection methods, offering a numerical assessment of the classification performance in both normal and faulty scenarios.

The impact of NO sensor faults on WWTP performance is thoroughly examined in this chapter through the energy cost, effluent quality and environmental assessments. For this, three different performance indices were analysed together with the GHG emissions, all of them quantified under normal and faulty sensor conditions.

7.2. Methodology of NO sensor faults implementation

Dedicated software modules were developed to simulate the specific failure types considered for the NO concentration sensor. They were incorporated into the WWTP dynamic simulation. For each fault type, the sensor's generated fault-affected output was fed into the Proportional-Integral NO controller as a measured NO process variable. This control loop is directly responsible for the denitrification process of the ASM technology. The placement of this sensor is presented in Figure 7.1.

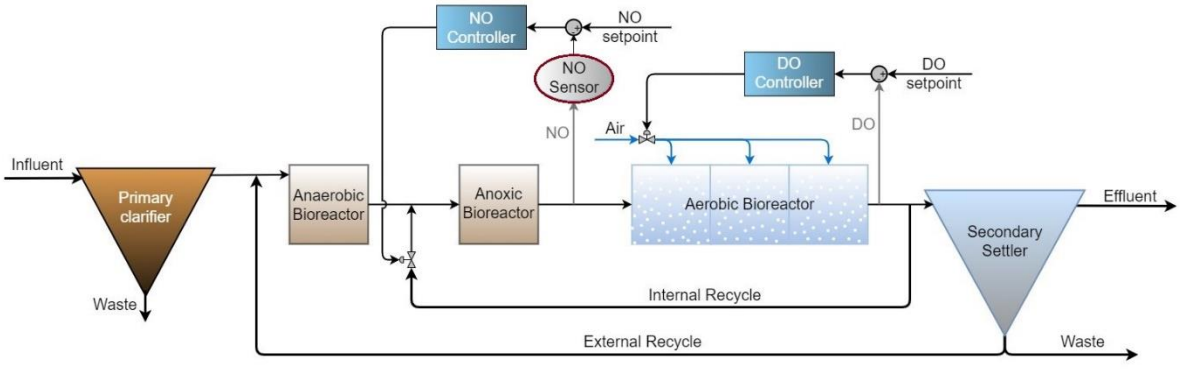

Figure 7.1. Placement of the NO sensor in the municipal WWTP

The NO control loop adjusts the internal recycle flow-rate of nitrate and nitrite to achieve the desired setpoint of 0.01 mg N/L, with zero offset. For each of the faults, the investigation scenario was performed similarly to the DO sensor case in terms of time. First, the simulation was run without any of the sensor's faults for 139 days. Each fault was then implemented for a total of 28 days,

beginning on the 140th day of the simulations. The values of the process variables considered in the fault identification and testing methodology were collected starting from day 140, i.e., the first day of faulty operation. The same duration of 28 days for the measurements collection period was considered for data that describe the normal operation.

The bias fault was characterized by a preset constant-difference between the true and faulty output values of the sensor. For the implementation of the fault in the simulation, a negative constant value of -0.0098 mg N/L was added to the true value of the NO concentration.

The drift was simulated as a time-varying ramp signal subtracted from the true NO value. A negative constant value of -0.0075 mg N/L was integrated in time and added to the NO true value. (i.e. -75% of the NO control loop's nominal setpoint value of 0.01 mg N/L).

The sensor's loss of effectiveness as a result of a negative gain factor was used to simulate the wrong gain fault. It had an incorrect calibration gain of -0.7, resulting in an incomplete correlation between the sensor's input and output signal. The incorrect gain was gradually introduced in time by a first-order filter with a time constant of 0.1 days.

A uniform random-number generator that delivered a value in the interval [-0.06, 0.06] with a sampling time of 0.1 days was used to generate the loss of accuracy faulty signal. This signal was processed by a first-order filter with a time constant of 0.01 days, followed by a saturation block, which limited its output value to the positive interval [0, 3]. This random signal was then subtracted from the NO true value to produce the inaccurately measured NO process variable.

The fixed value fault was characterized by the constant value of the sensor output signal. For this study, a constant value of -0.0095 mg N/L (which is 95% of the nominal setpoint value of the NO control loop of 0.01 mg N/L) was added to the NO control loop's nominal setpoint value of 0.01 mg N/L, resulting in the sensor constant value of 0.0005 mg N/L.

Each of NO concentration sensor types of faults was individually introduced at the specified time in the dynamic simulations and generated the set of data for describing the respective particular abnormal situation.

This chapter proposes supporting the WWTP management by investigating the way NO sensor malfunctions affect energy usage, greenhouse gas emissions, and effluent quality. It evaluates the effectiveness of fault detection and identification using the PCA and FDA methods, and offers recommendations for long-term operations and maintenance of the WWTP. The presented work was published in one scientific paper (Luca *et al*., 2025).

7.3. Process variables considered for the detection and identification of the NO sensor faults

Twenty process variables were considered for the PCA model and twenty-one for the FDA model. They consist of secondary settler concentrations (3 variables), anoxic bioreactor concentrations (8 variables), aerated bioreactors concentrations (7 variables), internal recycle flow

rate (1 variable), aeration flow rate (1 variable) and clean water effluent flow rate (1 variable). The variables considered were as follows: nitrate and nitrite nitrogen (S_{NO}), free and saline ammonia (S_{NH}), soluble biodegradable organic nitrogen (S_{ND}), particulate biodegradable organic nitrogen (X_{ND}), dissolved oxygen concentration (S_O), readily biodegradable substrate (S_S), slowly biodegradable substrate (X_S), alkalinity (S_{alk}), reactors flow rate (Q), nitrate recirculation flow rate (Q_{NR}) and aeration flow rate (Q_{air}). Measurements values of Q_{NR} were excluded from the FDA model.

7.4. PCA model construction for the detection of the NO sensor faults

The PCA model was developed using the 3840 observations from the 40-days of normal operating period. A matrix of 3840 observations of the 21 considered variables defined the training matrix *X*. The matrix *T* of scores and matrix *P* of loadings were obtained after scaling the training data matrix. A threshold value of 98.98% was chosen for *CPVk*. As a result, the number of principal components *k* considered necessary to represent the entire data set was determined to be 8. Using a 99% confidence level, the T_α^2 threshold was 20.16 and the *SPEα* threshold was 0.97. A fault was indicated by any vector *x* of measured variables with a T^2 or *SPE* value greater than the two corresponding thresholds.

Five different testing data matrices were constituted for each type of fault using the vectors of abnormal operation measured variables. Each testing data matrix was built from 1920 observations of 20 variables from the simulated time period of the day 140 up to the day 160 (excluding day no. 160). One value of the index T^2 and one of the index *SPE* were determined for each of the sample vectors of the testing matrices. To determine faulty sensor operation, T^2 and *SPE* values were determined for each of the sample vectors of the testing matrices and compared to previously calculated thresholds.

The T^2 and *SPE* values were graphically represented for each of the normal (training) and faulty (testing) samples, in association with the T_α^2 and SPEα thresholds.

7.5. FDA model construction for the identification of the NO sensor faults

Six distinct data classes were created for each scenario: one for normal operation mode and five for each of the NO sensor operation in the presence of a fault. These classes were created using data from days no.141 up to day no. 145 (five days), and they were used to train the FDA model. Each class was defined by 480 observations. There were 2880 lines and 20 columns in the observation training matrix. Data from day 140, the first day of the abnormal operation, was used to assess the trained FDA model's fault diagnosis performance. The testing data set included 96 measurements for each case, normal or abnormal. The purpose of this testing strategy was to investigate the ability of the FDA diagnosis method to identify the type of fault within the first few hours of its appearance.

In order to diagnose the faulty sensor operation, the values g_i, obtained for each class's

discriminant function, were compared. The discriminant functions g_i were computed for each of the 15-minute time-sampled measurements from the testing day 140, which were affected by the various types of faults (5 classes), as well as measurements corresponding to normal operation (1 class). The highest-valued discriminant function identified the faulty sensor's class and, as a result, diagnosed the type of fault.

7.6. Results and discussions

7.6.1. Assessment metrics for fault detection of NO sensor

Two different confusion matrices were employed to assess the effectiveness of the PCA-based and FDA-based methods, one for detecting and one for diagnosing the NO sensor considered malfunctions. The confusion matrix for the detection of the NO sensor faults considers two classes: normal (faulty-free) and abnormal (faulty). The first column (target) displays the abnormal class, while the second column (target) displays the normal class. The confusion matrix for the faults detection of the NO sensor is presented in Figure 7.2.

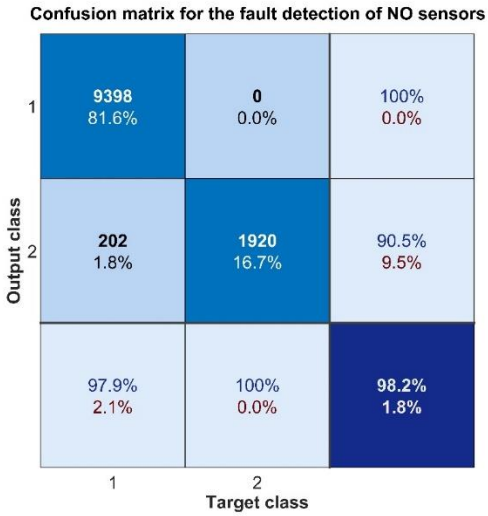

Figure 7.2. Confusion matrix of the fault detection for the NO sensor

For the normal class, the precision is 100%, meaning there are no false positives; for the abnormal class, it is 90.5%. Only 202 observations, representing 1.8% of all observations, were incorrectly classified as belonging to the abnormal class. The model's overall accuracy of 98.2% shows that it performed well in accurately detecting the NO sensor faults.

The F1 score value for the normal class equals 0.950, and for the abnormal class the F1 score is 0.989. Both scores reveal an outstanding balance between recall and precision, with the normal class performing very well.

7.6.2. Fault detection of NO sensor faults

Based on the PCA algorithm the presence of a NO sensor fault is detected when a measured variable vector x exceeds the threshold values for both T^2 and SPE indices.

The PCA-based NO sensor fault detection method demonstrated high efficiency in identifying sensor defects at an early stage. Among the two statistics, the *SPE*-based approach exhibited superior fault detection performance compared to the T^2 method. However, utilizing both indices in combination can enhance reliability by providing a more comprehensive assessment of fault occurrence and persistence.

The proposed PCA model was first tested on the normal operation data set. Graphical representations of the T^2 and *SPE* values for the samples from this data set against their corresponding thresholds are shown in Figure 7.3. They confirm the fault-free scenario, except for some T^2 statistics points. This means that for normal operation mode, the *SPE* statistic provides the most relevant outcomes on fault detection, as it was also reported in literature (Yoo *et al*., 2004).

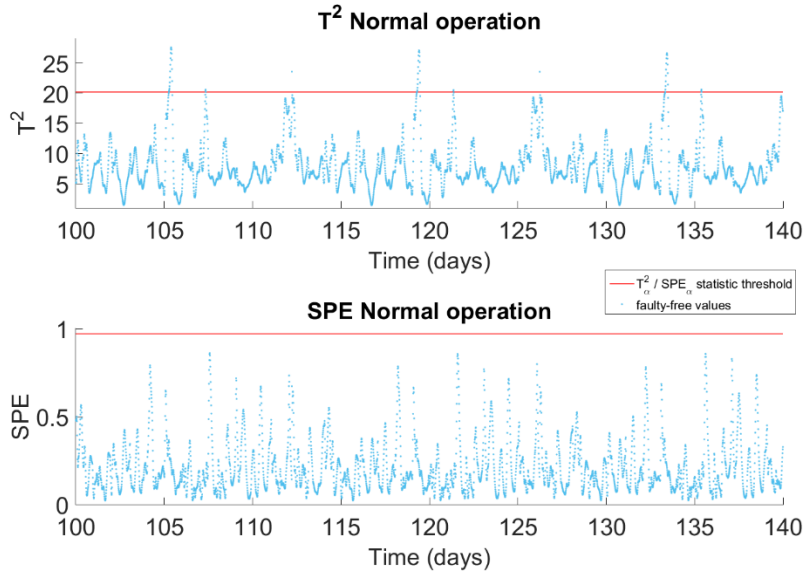

Figure 7.3. T^2 and *SPE* plots of PCA monitoring the normal operation data set for the NO sensor

Figures 7.4 to 7.8 show the fault detection results for each of the investigated NO sensor types of faults.

Figure 7.4 shows that the bias fault was discovered 1.25 hours after its incidence (during the 140th day). With only a few exceptions on the days no. 145, 152, and 159, the *SPE* statistic correctly detects the faulty regime, with better results than the T^2 statistic.

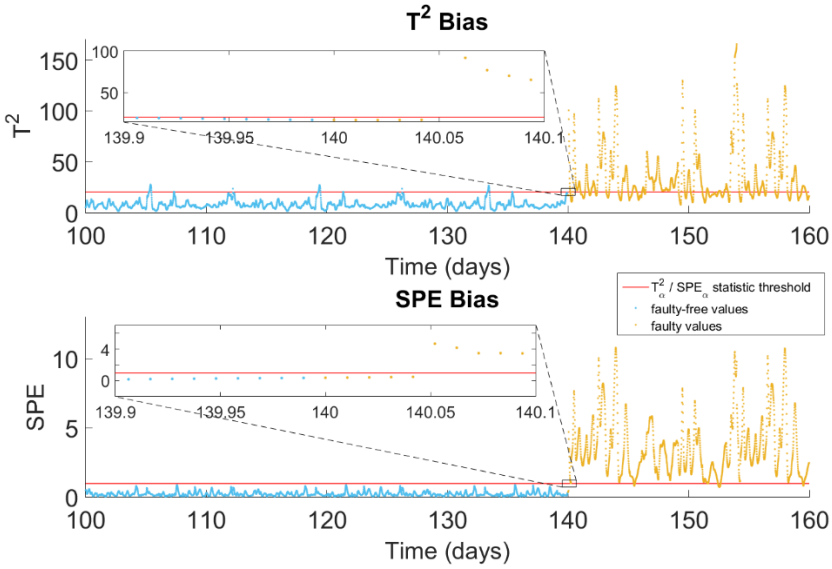

Figure 7.4. T^2 and *SPE* plots of PCA detecting the bias type of fault for the NO sensor, with details on the period next to the fault occurrence moment

The *SPE* method detected the drift fault in 19 hours and 15 minutes, faster than the Hotelling's T^2 method. Figure 7.5 shows that the fault detection with T^2 took 35 h and 15 minutes.

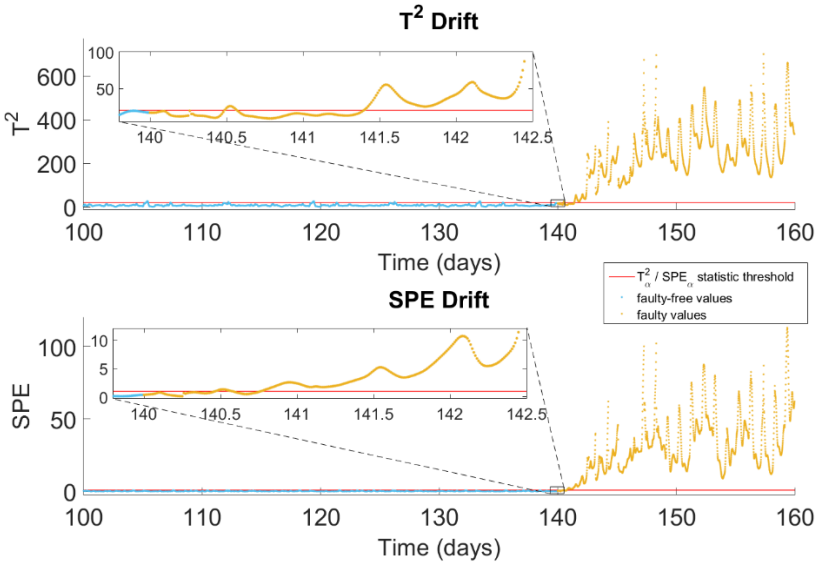

Figure 7.5. T^2 and *SPE* plots of PCA detecting the drift type of fault for the NO sensor, with details on the period next to the fault occurrence moment

Figure 7.6 presents the detection of the wrong gain fault. Moreover, in this fault-type case again the *SPE* method also proved to be more accurate. Even though the T^2 method initially detects the fault in 1.5 hours, it confirms the fault presence for the entire period of time just after 6.5 hours, whereas the *SPE* method firmly recognizes the wrong gain fault for the entire 28-day interval after 2 hours and 45 minutes.

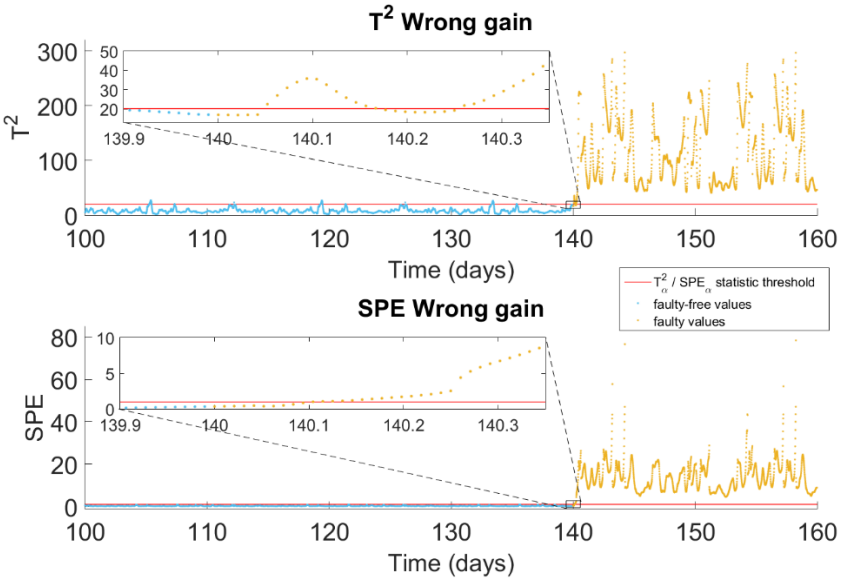

Figure 7.6. T^2 and *SPE* plots of PCA detecting the wrong gain type of fault for the NO sensor, with details on the period next to the fault occurrence moment

Even though the T^2 confirms the loss of accuracy fault presence in 1 hour and 45 minutes while the *SPE* acknowledges its presence after 3 hours and 45 minutes, the latter method detects it better on the next days. i.e., also, for the days no. 141, 143, and 144, as shown in Figure 7.7.

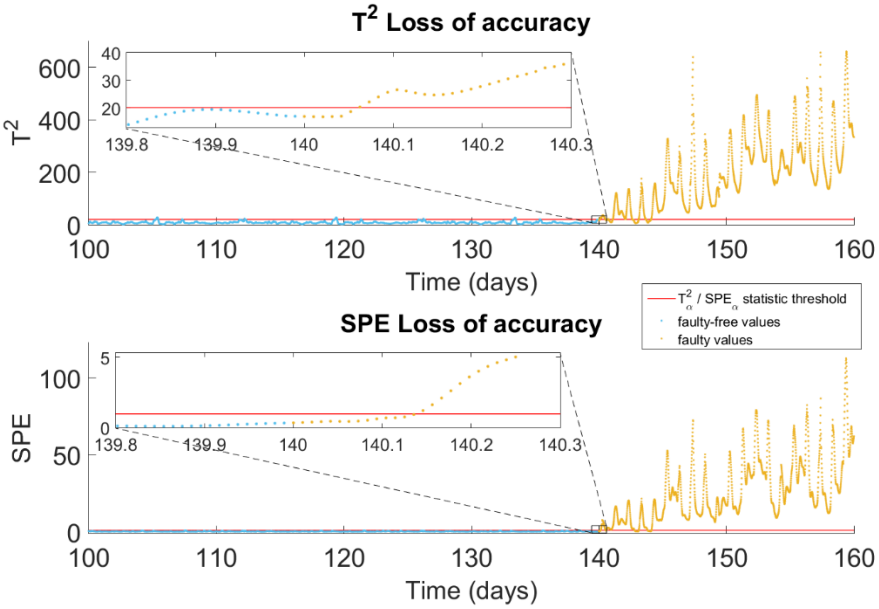

Figure 7.7. T^2 and *SPE* plots of PCA detecting the loss of accuracy type of fault for the NO sensor, with details on the period next to the fault occurrence moment

The PCA detection of the fixed value fault is shown in Figure 7.8. Both methods were equally effective and fast in this case, detecting the fixed value fault in 1 hour and 30 minutes.

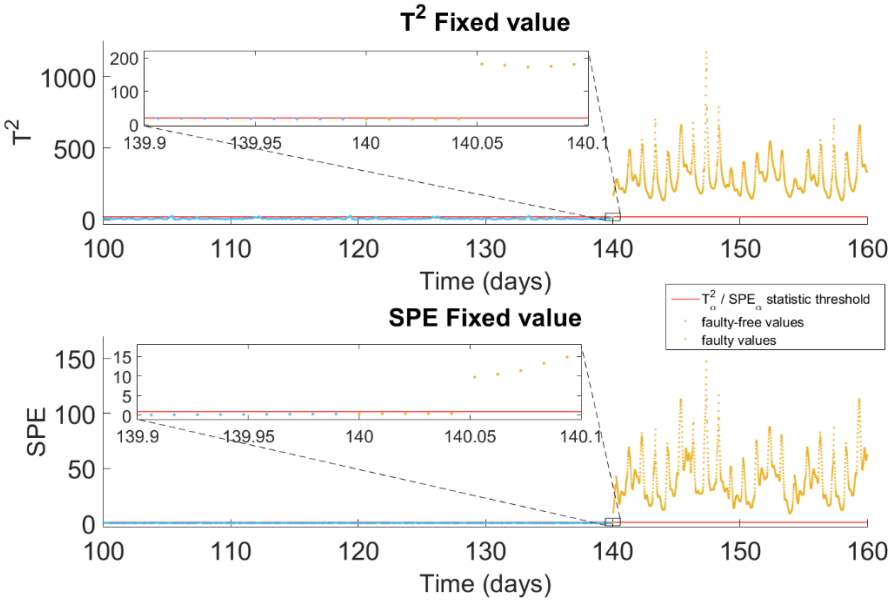

Figure 7.8. T^2 and *SPE* plots of PCA detecting the fixed value type of fault for the NO sensor, with details on the period next to the fault occurrence moment

The PCA-based NO sensor fault detection method proved to be highly effective and early in identifying faults, with the *SPE* statistic demonstrating superior performance over the T^2 method.

7.6.3. Assessment metrics for NO sensor fault diagnosis

For the FDA identification method, a confusion matrix was also developed. The first column in this matrix displays the normal class (fault-free operation, target 1). The following columns list the faulty classes (targets) in the order: bias (target 2), drift (target 3), wrong gain (target 4), loss of accuracy (target 5), and fixed value (target 6). The overall diagnosing accuracy was 97.7%, as illustrated in Figure 7.9.

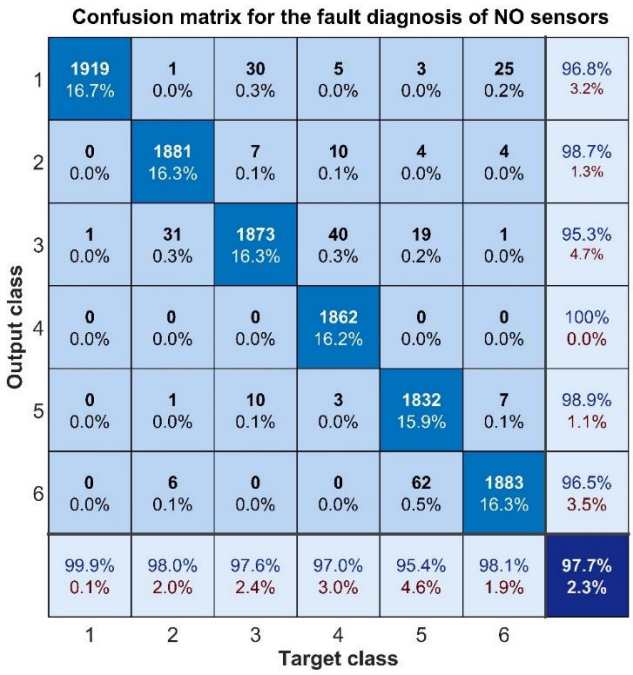

Figure 7.9. Confusion matrix of the fault diagnosis for NO sensors type of faults

7.6.4. Fault diagnosis of the NO sensor type of faults

The g_i values that correspond to the normal operation curve, represented with red colour on the graph, identified the fault-free normal operation. From the starting moment of diagnosing the 24-hour observation period, each of the g_i points of the red curve showed the highest values during the testing day, indicating normal operation with one single exception. This manifested for one value at 5.5 h.

The recall value in this case was 99.9%, and the long-term precision was 96.8%. A balanced assessment of precision and recall was provided by the computed F1 score of 0.983. In this instance, the MAR has a low value of 0.1%. The graphical representation of the diagnosis for the normal operation of the NO sensor is shown in Figure 7.10.

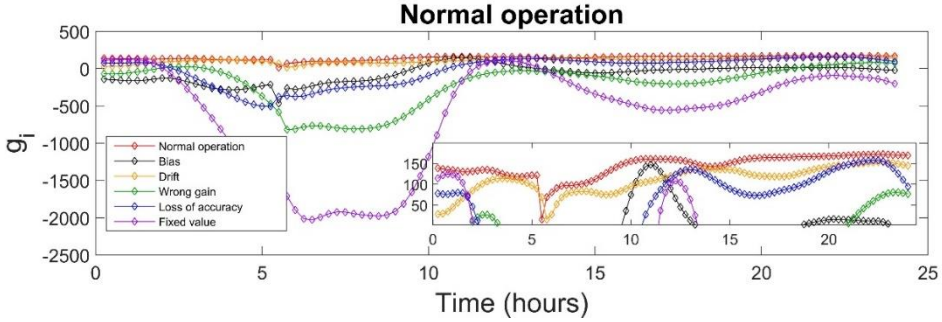

Figure 7.10. Faulty-free operation diagnosis for the NO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and five types of faults; details on the time period of confident identification

As presented in Figure 7.11, the bias fault type was successfully diagnosed for the following

time periods: 1h-1.5h, 5h-5.25h, 5.75h, 8.75h-13.75h, and then from 17 h until the end of the day. The latter value was considered as the confident moment of identification. In this case the MAR was 2%. The F1 score for this classification scenario was 0.983, reflecting a good detection of the fault type.

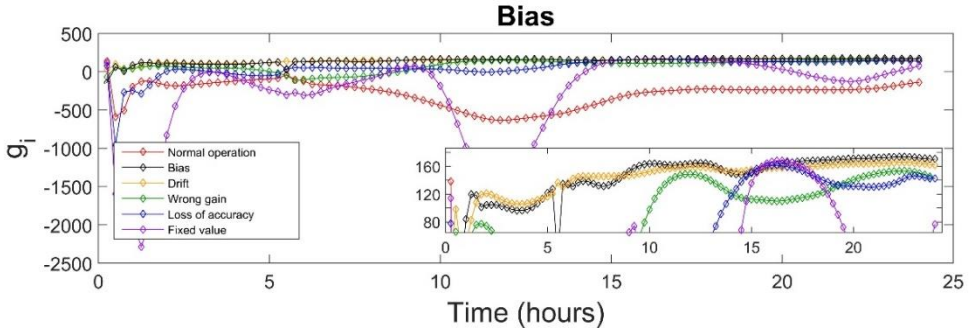

Figure 7.11. Bias diagnosis for the NO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and five types of faults; details on the time period of confident identification

The diagnosis of the drift fault was reported at the 5.5 hours point, then between 7.25h-8.25h, 9h-9.5h, 11.75h-14.75h and finally from 18h until the end of the day. After 18 hours, the yellow curve becomes dominant in comparison with other curves, as seen in Figure 7.12. The F1 score is

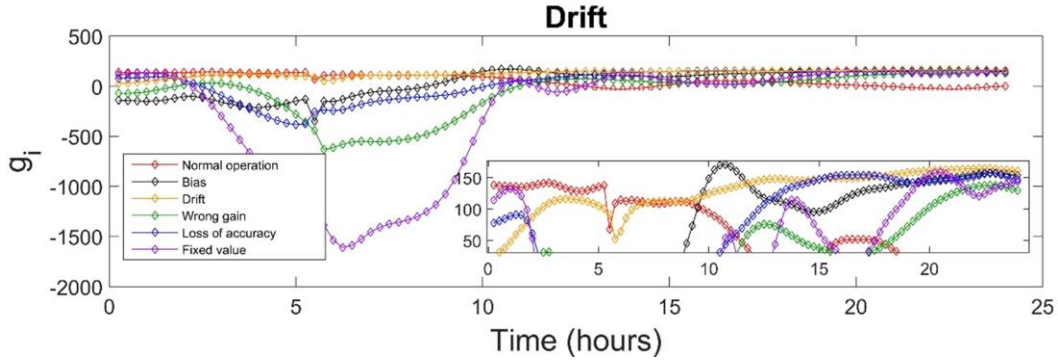

Figure 7.12. Drift diagnosis for the NO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and five types of faults; details on the time period of confident identification

0.964 and MAR has a value of 2.4%.

Figure 7.13 displays the wrong gain fault diagnosis. This kind of defect was definitively identified after 17.75 hours. However, it was also appropriately identified three periods of time, previous to the confident moment identification, i.e., between 9.75h-11h, 12.25h-12.75h time intervals and at the time moment of 14.75h. This results in a long-term precision of 100%, with a value of the recall of 97.0% and a MAR of 3%. The F1 score is 0.985, suggesting a very good capacity for diagnosing the fault.

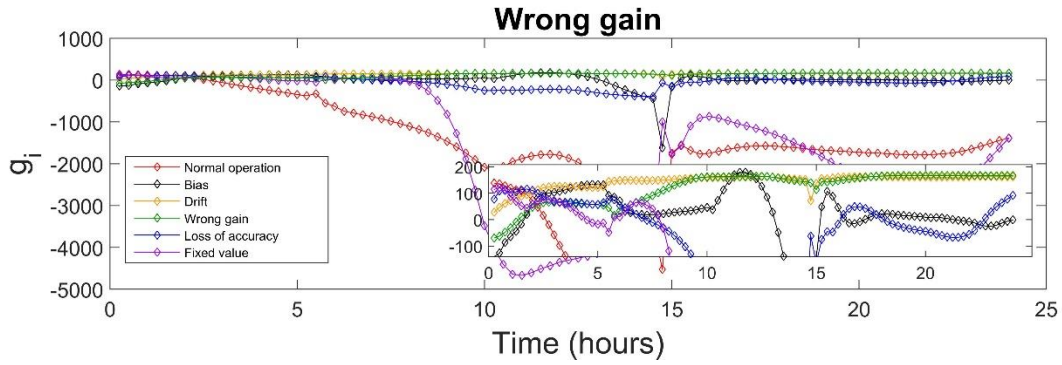

Figure 7.13. Wrong gain diagnosis for the NO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and five types of faults; details on the time period of confident identification

Loss of accuracy fault is correctly diagnosed between 12h-13.75h. During the first day of fault occurrence, the fault was temporarily misidentified either as a fixed value or as a drift fault. Occasionally, this defect is mistakenly diagnosed as normal functioning or drift. Figure 7.14 illustrates the diagnostic of the fault's loss accuracy type. The long-term precision for this fault was 98.9% and the recall value was 95.4%. Both indices are demonstrating a good proportion of correct identifications. The F1 score was 0.971 and MAR had a value of 4.6%, reflecting a good capability to identify the loss of accuracy type of fault.

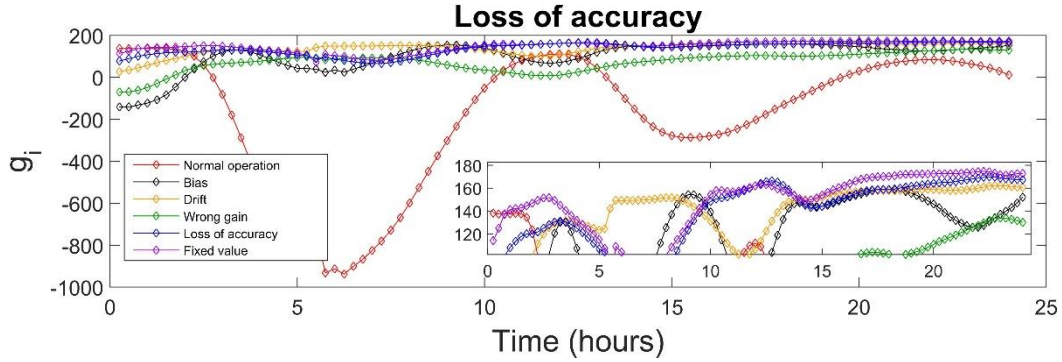

Figure 7.14. Loss of accuracy diagnosis for the NO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and five types of faults; details on the time period of confident identification

As presented in Figure 7.15 the fixed value is correctly diagnosed between 2h-5.25h, 10.75h-12.75h, 14.75h-15h and then, is persistently and rightly diagnosed after the time moment of 15.5h (confident identification moment). For the instances where this fault was misdiagnosed, the MAR is 1.9%. The F1 score of 0.973 and the long-term precision of 96.5% reflect a strong ability to correctly identify the fixed value type of fault.

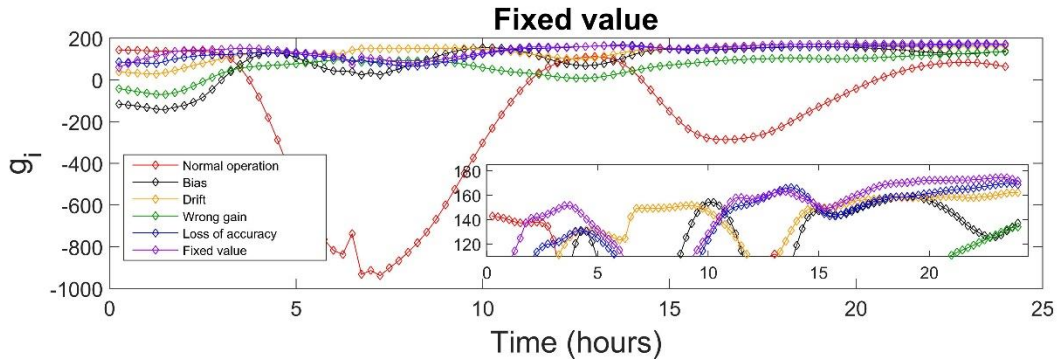

Figure 7.15. Fixed value diagnosis for the NO sensor: graphs of the FDA discriminant functions $g_i(x)$ for each of the normal and five types of faults; details on the time period of confident identification

It may be noted that the comparable NO sensor signal values considered in the faulty scenarios lead to some very close values of the NO process variable. In the end, this resulted in similar g_i values of the discriminant function. Therefore, it was more likely that some faults would be misdiagnosed in the immediate moment of the fault incidence, as well as in the longer or shorter time periods that follow.

7.6.5. Performance evaluation during operation with NO sensor faults

Similar to the DO sensor faults investigations, performance indices AE, PE, and EQ were determined as a daily mean value over the period of 28 days, for both normal and defective operating cases of the NO sensor (140th day up to the 168th day, the last one being excluded). The AE, PE, and EQ values for the normal operation and operation with the five types of fault scenarios are shown in Table 7.1.

Table 7.1. Values of the performance indices for the normal and faulty operation cases of the NO sensor

Operating regime	AE (kWh/day)	PE (kWh/day)	Total energy demand (kWh/day)	EQ (kg PU/day)
Normal operation	16,992	1,329	18,321	16,852
Bias fault	17,253	1,633	18,886	15,050
Drift fault	17,776	2,320	20,096	14,911
Wrong gain fault	17,496	1,956	19,452	17,108
Loss of accuracy fault	17,752	2,232	19,984	17,186
Fixed value fault	17,848	2,415	20,263	17,198

All faults negatively influence AE, PE, and overall energy demand, resulting in higher energy costs for aeration and for pumping the internal recycle flow. Increased pumping occurs as a result of

the lower NO sensor signal value that drives the NO controller's manipulated variable to higher internal recycle flow (nitrate and nitrite flow). Additionally, due to the DO controller's increased manipulated air flowrate variable, the AE is also increased. Bias and drift faults have a favourable effect on EQ values, while wrong gain, loss of accuracy and fixed value faults result in lower values of this performance index.

7.6.6. Energy costs assessment for operation affected by NO sensor faults

The daily cost for each type of faulty operation of the WWTP under consideration in this study was calculated similarly to the ones computed for the DO sensor faults investigation. Each of the possible types of energy source or energy production technology, potentially to be used by the AE and PE of the WWTP, had its daily costs evaluated. Table 7.2 provides a summary of these evaluations.

Table 7.2. Source depending sum of AE and PE energy costs computed for normal and fault

Source/Technology	Daily operation costs (€)					
	Normal operation	**Bias**	**Drift**	**Wrong gain**	**Loss of accuracy**	**Fixed value**
Lignite	1,812	1,868	1,987	1,924	1,976	2,004
Coal	1,717	1,770	1,883	1,823	1,873	1,899
CCGT	1,975	2,036	2,166	2,097	2,154	2,184
Nuclear	3,435	3,541	3,768	3,647	3,747	3,799
Onshore wind	1,237	1,275	1,356	1,313	1,349	1,368
Offshore wind	1,566	1,615	1,718	1,663	1,709	1,732
Solar PV commercial	1,479	1,524	1,622	1,570	1,613	1,635
Solar PV residential	2,327	2,399	2,552	2,471	2,538	2,573
Solar thermal (CSP)	2,712	2,795	2,974	2,879	2,958	2,999
Hydro reservoir	1,457	1,501	1,598	1,547	1,589	1,611
Hydro run of river	1,070	1,103	1,174	1,136	1,167	1,183
Geothermal	1,581	1,630	1,734	1,679	1,725	1,749
Biomass	2,120	2,185	2,325	2,251	2,312	2,344

affected operation of the WWTP water line for the NO sensor case

The fixed value fault of the NO sensor is the one that results in the highest energy costs, according to an analysis of the costs of electrical energy spent by the WWTP during operation in the presence of the various types of faults. On the other hand, the bias fault implies the lowest energy cost.

Prior to replacing the defective sensor, the implied energy costs may be reduced by switching to alternative, less expensive energy sources and implementing a special plan for the use of a mixed energy source in the event of faulty sensor operation, depending on the availability of technical solutions.

7.6.7. Environmental assessment of CO_2 and N_2O emissions for operation with NO sensor faults

Estimates were made of the on-site and off-site CO_2 and N_2O emissions, which are significant sources of the greenhouse gases emitted by the water line of the WWTP. Their daily average values were computed when the operation of the WWTP was carried out in the presence of each of the NO sensor types of faults. The GHG emission results are shown in Table 7.3.

Table 7.3. GHG emissions due to the different types of the NO sensor defects

Site of emissions	Process	Type of emitted gas	Type of fault					
			Normal	Bias	Drift	Wrong gain	Loss of accuracy	Fixed value
Off-site emissions	Power generation	CO_2, $P_{CO_2,off-site}$ (kg CO_2/day)	3,481	3,588	3,818	3,696	3,797	3,850
	Biological degradation in the WWT effluent	N_2O, $P_{N_2O,off-site}$, (kg N_2O/day)	3.61	3.65	3.50	3.63	3.51	3.47
On-site emissions	Water-line aerobic biological processes	CO_2, $P_{CO_2,on-site}$ (kg CO_2/day)	13,689	13,869	14,175	14,037	14,137	14,207
		N_2O, $P_{N_2O,on-site}$ (kg N_2O/day)	10.35	10.34	10.40	10.35	10.40	10.42
Total emissions		CO_2, $P_{CO_2,total}$ (kg CO_2/day)	17,170	17,457	17,993	17,733	17,934	18,057
		N_2O, $P_{N_2O,total}$ (kg N_2O/day)	13.96	13.99	13.90	13.98	13.91	13.89
Total overall emissions		CO_{2e}, $P_{CO_{2e},overall}$ (kg CO_{2e}/day)	21,330	21,626	21,135	22,899	22,079	22,196

According to the GHG emission values presented in Table 7.3, on-site emissions are the most substantial, making up almost 80%, respectively 75% of all emissions under typical operating conditions for $P_{CO_2,total}$ and $P_{N_2O,total}$. For all instances of investigated NO sensor faults, the computed total CO_2 emissions, that is, the total of on-site and off-site values, showed higher values. Bias and wrong gain faults showed higher total N_2O emission levels than the other faults. The latter presented slightly lower values. It is important to note that the lowest N_2O emissions values were produced by fixed value fault type, in contrast to the trend in total CO_2 emissions, where the bias fault produced the highest emissions.

To determine the overall CO_2 emissions, the sum of the total CO_2 and total N_2O emissions was used. According to the findings, the majority of the malfunctions produce more overall CO_2 emissions than in the event of normal operation.

The evaluation of CO_2 and N_2O emissions computed for various NO sensor faults offers useful quantitative data on the scope and ranking of the most problematic sensor types of faults that may have a significant impact on the long-term operation of the WWTP.

7.7. Conclusions

This study evaluated PCA-based fault detection and FDA-based fault identification for the NO sensor controlling nitrite/nitrate in an anoxic WWTP reactor. Faults considered included loss of accuracy, fixed value, drift, bias, and wrong gain, using a calibrated WWTP model with two primary control loops for DO and NO. PCA detection was fast and accurate (98.2%), with SPE generally outperforming T^2, while FDA achieved 97.7% diagnosis accuracy, with initial detection times of 1-12h and consolidated diagnosis in 5.5-18h.

Faults affected energy use, effluent quality (EQ), and GHG emissions. Normal operation consumed 18,321 kWh/day with EQ 16,852 kg PU/day; faults raised energy use by 10.6-20,263 kWh/day. Bias and drift slightly improved EQ but increased energy, while wrong gain, loss of accuracy, and fixed value faults worsened EQ (>17,000 kg PU/day). GHG emissions were mainly on-site (~75%), with fixed value faults causing the largest increases (~5%).

The results demonstrate that PCA and FDA methods are effective for NO sensor fault management, supporting safe, efficient, and sustainable WWTP operation. Findings also provide guidance for control system design, energy management, environmental impact mitigation, maintenance, calibration, and intelligent sensor development.

8. Final conclusions and personal contributions

This research examined DO and NO sensor faults in a municipal WWTP, focusing on early detection, diagnosis, and their effects on effluent quality, energy use, and GHG emissions. PCA-based fault detection and FDA-based fault identification proved effective for various fault types, including bias, drift, wrong gain, loss of accuracy, fixed value, and complete failure. Faults significantly impacted the EQ index and energy consumption, with complete failure, fixed value, and loss of accuracy being the most detrimental. Energy-intensive aeration was particularly affected, and all faults increased total energy use by up to 10.6%. GHG emissions were also strongly influenced, with certain faults notably worsening the plant's environmental footprint. The findings demonstrate the importance of reliable sensors and effective fault management for sustainable, efficient WWTP operation.

The most significant conclusions summarizing both quantitative and qualitative results obtained by the present research are the following:

- Fault detection results:
    - The PCA-based model proved highly effective in detecting faults in DO and NO concentration sensors, achieving an overall accuracy of 99.6% for the DO sensors and 98.2% for the NO sensors;
    - Comparing Hotelling's T^2 and *SPE* statistic indices, the SPE method was faster at detecting faults, typically identifying them in less than two hours;
    - Detection times for DO sensor faults ranged from less than one hour (bias, complete failure) to 5 hours 15 minutes (drift);
    - Detection times for NO sensor faults ranged from less than 1 hour 15 minutes (bias) to 19 hours 15 minutes (drift);
- Fault diagnosis results:
    - The FDA-based model demonstrated strong diagnostic performance, accurately distinguishing between faulty and fault-free operations. It achieved an overall accuracy of 85.8% for DO sensor faults and 97.7% for NO sensor faults;
    - For DO sensor faults, FDA identified faults within 2.5 to 16.5 hours;
    - For NO sensor faults, FDA identified faults in as little as 15 minutes to 1 hour, with values of the confident identification moment ranging between 15 minutes and 18 hours;
- Impact on effluent quality (EQ index):
    - For DO sensor faults, the most detrimental scenarios were complete failure maximum, fixed value, and loss of accuracy, while wrong gain had minimal impact;
    - For NO sensor faults, fixed value faults had the worst effect on the EQ index,

while drift, wrong gain, and loss of accuracy also caused deterioration to varying degrees;

- Impact on GHG emissions:
    - For DO sensor faults, complete failure maximum and fixed value scenarios resulted in 40 times increase in GHG emissions compared to normal operation;
    - For NO sensor faults, on-site sources contributed 75% of total CO_2 and N_2O emissions, with fixed value faults causing the most severe impact, i.e., increasing emissions by approximately 5%;
- Energy cost implications:
    - The complete failure minimum type of fault of the DO sensor led to the highest energy costs due to the induced excessive aeration;
    - For the NO sensor faults, fixed value faults resulted in the highest energy costs among all fault types.

Future work may focus on improving the detection and identification of multiple simultaneous sensor faults in wastewater treatment plants. This could involve developing more advanced diagnostic models to enhance accuracy and reliability. Machine learning techniques, such as deep learning, may help automate fault detection and improve response times.

Real-time monitoring could also be improved with adaptive control strategies and self-correcting algorithms that adjust detection thresholds based on system data. Integrating advanced fault detection into SCADA system of the municipal WWTP could streamline operations and reduce costs.

By exploring these areas, future research can help make wastewater treatment systems more efficient, reliable, and environmentally friendly, ensuring better water quality and lower energy consumption.

REFERENCES

1. Alex J., Benedetti L., Copp J., Gernaey K.V., Jeppsson U., Nopens I., Pons M.N., Steyer J.P., Vanrolleghem P., Benchmark simulation model no. 1 (BSM1), 2008.
2. Alex J., Tschepetzki R., Jumor U., Obenaus F., Rosenwinkel K.H., Analysis and design of suitable model structures for activated sludge tanks with circulating flow, *Water Science and Technology*, 1999, 39(4), 55-60. DOI: https://doi.org/10.1016/S0273-1223(99)00053-0.
3. Beraud B., Lemoine C., Steyer J.P., Computational intelligence techniques for bioprocess modelling, supervision and control, Springer, Berlin, 2009, Chapter Multiobjective Genetic Algorithms for the Optimisation of Wastewater Treatment Processes, pp. 163-195. DOI: https://doi.org/10.1007/978-3-642-01888-6_6.
4. Britschgi L., Villez K., Schrems P., Udert K.M., Electrochemical nitrite sensing for urine nitrification, *Water Research X*, 2020, 9, 100055. DOI: https://doi.org/10.1016/j.wroa.2020.100055.
5. Chiang L.H., Russell E.L., Braatz R.D., Fault detection and diagnosis in industrial systems, Springer Science & Business Media, London, 2001. ISBN: 1-85233-327-8.
6. Corona F., Mulas M., Haimi H., Sundell L., Heinonen M., Vahala R., Monitoring nitrate concentrations in the denitrifying post-filtration unit of a municipal wastewater treatment plant, *Journal of Process Control*, 2013, 23(2), 158-70. DOI: https://doi.org/10.1016/j.jprocont.2012.09.011.
7. Du Z., Jin X., Multiple faults diagnosis for sensors in air handling unit using Fisher discriminant analysis, *Energy Conversion and Management*, 2008, 49(12), 3654-3665. DOI: https://doi.org/10.1016/j.enconman.2008.06.032.
8. Duda R.O., Hart P.E., Stork D.G., Pattern classification, 2nd edition, Wiley, New York, 2000. ISBN: 978-0-471-05669-0.
9. Frank P., Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: a survey and some new results, *Automatica*, 1990, 26(3), 459-474. DOI: https://doi.org/10.1016/0005-1098(90)90018-D.
10. Fuente M., Garcia G., Sainz G., Fault diagnosis in a plant using fisher discriminant analysis, in Proceedings of the 16th Mediterranean Conference on Control and Automation Congress Centre, 25th-27th of June 2008, Ajaccio, France, 53-58. DOI: https://doi.org/10.1109/MED.2008.4602082.
11. Garcia-Alvarez D., Fault detection using principal component analysis (PCA) in a wastewater treatment plant (WWTP), in Proceedings of the International Student's Scientific Conference, 15th of January 2009, Online Conference, 55-60.
12. Gernaey K.V., van Loosdrecht M.C.M., Henze M., Lind M., Jorgensen S.B., Activated sludge wastewater treatment plant modelling and simulation: state of the art, *Environmental Modelling & Software*, 2004, 19, 763-783. DOI: https://doi.org/10.1016/j.envsoft.2003.03.005.
13. Gori R., Jiang L.-M., Sobhani R., Rosso D., Effects of soluble and particulate substrate on the carbon and energy footprint of wastewater treatment processes, *Water Research*, 2011, 45, 5858-5872. DOI: https://doi.org/10.1016/j.watres.2011.08.036.

14. Gujer W., Kappeler J., Modelling population dynamics in activated sludge systems, *Water Science and Technology*, 1992, 25(6), 93-104. DOI: https://doi.org/10.2166/wst.1992.0116.
15. Hauduc H., Rieger L., Ohtsuki T., Shaw A., Takacs I., Winkler S., Heduit A., Vanrolleghem P.A., Gillot S., Activated sludge modelling: development and potential use of a practical applications database, *Water Science and Technology*, 2011, 63, 2164-2182, DOI: https://doi.org/10.2166/wst.2011.368.
16. He Q., Qin S., Wang J., A new fault diagnosis method using fault directions in fisher discriminant analysis, *AIChE Journal*, 2005, 51(2), 555-571. DOI: https://doi.org/10.1002/aic.10325.
17. Henze M., Gujer W., Mino T., Matsuo T., Wentzel M.C., Marais G.V.R., Activated sludge model no. 2, IAWQ Scientific and Technical Report No. 3, IAWQ, London, 1995.
18. Henze M., Gujer W., Mino T., van Loosedrecht M., Activated sludge models ASM1, ASM2, ASM2d and ASM3, IWA Publishing, London, 2000, pp. 1-120. DOI: https://doi.org/10.2166/9781780402369.
19. Huang F., Shen W., Calculation of N_2O emissions in the wastewater treatment process of paper mill, in Proceedings of the 2nd International Conference on Sustainable Energy, Environment and Information Engineering (SEEIE 2019), 24th-25th of March 2019, Beijing, China. DOI: https://doi.org/10.2991/seeie-19.2019.29.
20. IEA - International Energy Agency, Levelised cost of electricity calculator, https://www.iea.org/data-and-statistics/data-tools/levelised-cost-of-electricity-calculator, 2024, accessed on 7th of September 2025.
21. IPCC - Intergovernmental Panel on Climate Change, Guidelines for national greenhouse gas inventories. Intergovernmental panel on climate change, 2006, Table 6.11, 5(6), http://www.ipcc-nggip.iges.or.jp/public/2006gl/index.html, accessed on 7th of September 2025.
22. Jeppsson U., Modelling aspects of Wastewater Treatment Processes, PhD thesis,1996.
23. Kazemi P., Bengoa C., Steyer J.-P., Giralt J., Data-driven techniques for fault detection in anaerobic digestion process, *Process Safety and Environmental Protection*, 2020, 146, 905-915. DOI: https://doi.org/10.1016/j.psep.2020.12.016.
24. Kost C., Muller P., Schweiger J.S., Fluri V., Thomsen J., Levelized cost of electricity -Renewable energy technologies, *Fraunhofer ISE Report*, 2024, 1-48.
25. Lazard, Lazard's levelized cost of energy +, https://www.lazard.com/research-insights/levelized-cost-of-energyplus, 2024, accessed on 7th of September 2025.
26. Lee C., Choi S.W., Lee, I.B., Sensor fault diagnosis in a wastewater treatment process, *Water Science and Technology*, 2006, 53(1), 251-257. DOI: https://doi.org/10.2166/wst.2006.027.
27. Listowski A., Ngo H.H., Guo W.S., Vigneswaran S., Shin H.S., Moon H., Greenhouse gas (GHG) emissions from urban wastewater system: future assessment framework and methodology, *Journal of Water Sustainability*, 2011, 1(1), 113-125.

28. Luca A.-V., Simon-Varhelyi M., Mihaly N.-B., Cristea V.-M., Data driven detection of different dissolved oxygen sensor faults for improving operation of the WWTP control system, *Processes*, 2021, 9(9),1633. DOI: https://doi.org/10.3390/pr9091633.
29. Luca A.-V., Simon-Varhelyi M., Mihaly N.-B., Cristea V.-M., Fault detection and diagnosis of the wastewater nitrate and nitrite sensors using PCA and FDA combined with assessment of the economic and environmental impact of the faults, *Environmental Monitoring and Assessment*, 2025, 197, 121. DOI: https://doi.org/10.1007/s10661-024-13593-z.
30. Luca A.-V., Simon-Varhelyi, M., Mihaly N.-B., Cristea V.-M., Fault type diagnosis of the WWTP dissolved oxygen sensor based on Fisher Discriminant Analysis and assessment of associated environmental and economic impact, *Applied Sciences*, 2023, 13(4), 2554. DOI: https://doi.org/10.3390/app13042554.
31. Mannina G., Ekama G., Caniani D., Cosenza A., Esposito E., Gori R., Garrido-Baserba M., Rosso D., Olsson G., Greenhouse gases from wastewater treatment - a review of modelling tools, *Science of The Total Environment*, 2016, 551-552, 254-270, DOI: https://doi.org/10.1016/j.scitotenv.2016.01.163.
32. MathWorks, Matlab, https://www.mathworks.com/products/Matlab.html, 2025a, accessed on 7th of September 2025.
33. MathWorks, Simulink, https://www.mathworks.com/products/Simulink.html, 2025b, accessed on 7th of September 2025.
34. Prendez M., Lara-Gonzalez S., Application of strategies for sanitation management in wastewater treatment plants in order to control/reduce greenhouse gas emissions, *Journal of Environmental Management*, 2008, 88(4), 658-664. DOI: https://doi.org/10.1016/j.jenvman.2007.03.041.
35. Qin S.J., Data-driven fault detection and diagnosis for complex industrial processes, in Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes (IFAC), 30th of June - 3rd of July 2009, Barcelona, Spain, *IFAC Proceedings Volumes*, 42(8), 1115-1125. DOI: https://doi.org/10.3182/20090630-4-es-2003.00184.
36. Rieger L., Gillot S., Langergraber G., Ohtsuki T., Shaw A., Takacs I., Winkler, S., Guidelines for using activated sludge models, IWA Publishing, London, 2012, pp. 11-147. DOI: https://doi.org/10.2166/9781780401164.
37. Rosen C., Rieger L., Jeppsson U., Vanrolleghem P.A., Adding realism to simulated sensors and actuators, *Water Science and Technology*, 2008, 57(3), 337-44. DOI: https://doi.org/10.2166/wst.2008.130.
38. Sanchez-Fernandez A., Fuente M.J., Sainz-Palermo G.I., Fault detection in wastewater treatment plants using distributed PCA methods, in Proceedings of the 20th Conference on Emerging Technologies & Factory Automation (ETFA), 8th-11th of September 2015, Luxembourg, Luxembourg, 1-7. DOI: https://doi.org/10.1109/ETFA.2015.7301504.
39. Simon-Varhelyi M., Cristea V.-M., Luca A.-V., Reducing energy costs of the wastewater treatment plant by improved scheduling of the periodic influent load, *Journal of Environmental Management*, 2020a, 262, 110294, 1-10. DOI: https://doi.org/10.1016/j.jenvman.2020.110294.

40. Sollfrank U., Gujer W., Characterization of domestic wastewater for mathematical modelling of the activated sludge process, *Water Science and Technology*, 1991, 23, 4-6, 1057-1066.
41. Takacs I., Patryio G.G., Nolasco D., A dynamic model of the clarification-thickening process, *Water Research*, 1991, 25(10), 1263-1271. DOI: https://doi.org/10.1016/0043-1354(91)90066-Y.
42. Tao E.P., Shen W.H., Liu T.L., Chen X.Q., Fault diagnosis based on PCA for sensors of laboratorial wastewater treatment process, *Chemometrics and Intelligent Laboratory Systems*, 2013, 128, 49-55. DOI: https://doi.org/10.1016/j.chemolab.2013.07.012.
43. Tchobanoglous G., Burton F.I., Stensel H.D., Wastewater engineering - treatment and reuse, Metcalf & Eddy Inc., Boston, 2003. ISBN: 7-302-05857-1.
44. Tchobanoglous G., Stensel H.D., Tsuchihashi R., Burton F.I., Wastewater engineering - treatment and reuse, Metcalf & Eddy Inc., Boston, 2014. ISBN: 0-073-40118-8.
45. Teh H.Y., Kempa-Liehr A.W., Wang K.I.K., Sensor data quality: a systematic review, *Journal of Big Data*, 2020, 7(1), 1-49. DOI: https://doi.org/10.1186/s40537-020-0285-1.
46. Tomita R.K., Park S.W., Sotomayor O.A., Analysis of activated sludge process using multivariate statistical tools - a PCA approach, *Chemical Engineering Journal*, 2002, 90(3), 283-290. DOI: https://doi.org/10.1016/S1385-8947(02)00133-X.
47. Vallero D.A., Air pollution calculations, Elsevier, Amsterdam, 2019, Chapter Air pollution biogeochemistry, pp. 175-206, ISBN 978-0-12-814934-8.
48. Varhelyi M., Cristea V.-M., Brehar, M., Nemes E.-D., Nair A., WWTP model calibration based on different optimization approaches, *Environmental Engineering and Management Journal*, 2019, 18(8), 1657-1670.
49. Villegas T., Fuente M.J., Sainz-Palmero G.I., Fault diagnosis in a wastewater treatment plant using dynamic independent component analysis, in Proceedings of the 18th Mediterranean Conference on Control and Automation (MED'10), 23rd-25th of June 2010, Marrakech, Morocco, 874-879. DOI: https://doi.org/10.1109/MED.2010.5547760.
50. Wanner J., Grau P., Filamentous bulking in nutrient removal activated sludge systems, *Water Science and Technology*, 1988, 20(4-5), 1-8. DOI: https://doi.org/10.2166/wst.1988.0149.
51. Wise B.M., Ricker N.L., Veltkamp D.F., Kowalski B.R., Theoretical basis for the use of principal component models for monitoring multivariate processes, *Process Control and Quality*, 1990, 1, 41-51.
52. Yoo C.K., Lee J.M., Lee I.B., Vanrolleghem P.A., Dynamic monitoring system for full-scale wastewater treatment plants, *Water Science and Technology*, 2004, 50, 163-171. DOI: https://doi.org/10.2166/wst.2004.0684.
53. Yoon S., MacGregor J., Fault diagnosis with multivariate statistical models part I: using steady state fault signatures, *Journal of Process Control*, 2001, 11(4), 387-400. DOI: https://doi.org/10.1016/S0959-1524(00)00008-1.