

Article

Data Driven Detection of Different Dissolved Oxygen Sensor Faults for Improving Operation of the WWTP Control System

Alexandra-Veronica Luca, Melinda Simon-Várhelyi, Norbert-Botond Mihály and Vasile-Mircea Cristea *

Department of Chemical Engineering, Faculty of Chemistry and Chemical Engineering, Babeş-Bolyai University of Cluj-Napoca, 11 Arany János Street, 400028 Cluj-Napoca, Romania; alexandra.luca@ubbcluj.ro (A.-V.L.); melinda.varhelyi@ubbcluj.ro (M.S.-V.); norbert.mihaly@ubbcluj.ro (N.-B.M.)
* Correspondence: mircea.cristea@ubbcluj.ro; Tel.: +40-264-593833

Abstract: Sensor faults frequently occur in wastewater treatment plant (WWTP) operation, leading to incomplete monitoring or poor control of the plant. Reliable operation of the WWTP considerably depends on the aeration control system, which is essentially assisted by the dissolved oxygen (DO) sensor. Results on the detection of different DO sensor faults, such as bias, drift, wrong gain, loss of accuracy, fixed value, or complete failure, were investigated based on Principal Components Analysis (PCA). The PCA was considered together with two statistical approaches, i.e., the Hotelling's T^2 and the Squared Prediction Error (SPE). Data used in the study were generated using the previously calibrated first-principle Activated Sludge Model no.1 for the Anaerobic-Anoxic-Oxic (A^2O) reactors configuration. The equation-based model was complemented with control loops for DO concentration control in the aerobic reactor and nitrates concentration control in the anoxic reactor. The PCA data-driven model was successfully used for the detection of the six investigated DO sensor faults. The statistical detection approaches were compared in terms of promptness, effectiveness, and accuracy. The obtained results revealed the way faults originating from DO sensor malfunction can be detected and the efficiency of the detection approaches for the automatically controlled WWTP.

Keywords: fault detection; principal component analysis; DO concentration sensors; automatic controlled wastewater treatment plant



Citation: Luca, A.-V.; Simon-Várhelyi, M.; Mihály, 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*, 1633.
<https://doi.org/10.3390/pr9091633>

Academic Editor: Szabolcs Fogarasi

Received: 17 August 2021
Accepted: 8 September 2021
Published: 10 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Clean water availability has become one of the ten major problems of humanity, as water demand due to population growth, urbanization, industrialization, and environmental concerns associated with wastewater-treatment technologies have been considerably amplified during the last decade. Domestic, industrial, agricultural, mining, and commercial sources of wastewater require efficient removal of organic or inorganic pollutants, present in solid or suspended forms.

Taking into account the increasing demands of European and international regulations regarding the standards of wastewater treatment and required water quality characteristics, the wastewater treatment plant (WWTP) becomes the most important actor, but its operation turns is difficult and challenging. Detailed descriptions of wastewater treatment processes and their associated models are very complex, show non-linear behavior, and are characterized by many continuously changing variables. Besides setting the structure and finding the accurate parameter values of the model, the most demanding difficulties for WWTP modeling and control of operation are daily, weekly, or seasonally influent composition changes. Furthermore, comparison of control strategies operating in different plants is complex and requires building simulation benchmarks [1].

The International Water Association (IWA), as a professional organization, is known for its extensive contributions regarding solutions to worldwide water challenges. Following the assumed mission to advance standards for sustainable water management,

IWA's specialized researchers developed state-of-the-art first-principle mathematical models for promoting the design and operation of biological wastewater treatment processes [2]. The most prevalent model is the Activated Sludge Model No. 1 (ASM1). It was developed by IWA (the former IAWQ) in order to set the nomenclature and milestones for modeling WWTPs in a standardized way [3]. Over the years, ASM2, ASM2d, and ASM3 followed with the aim of improving the description of the complex biochemical processes that take place during activated sludge wastewater treatment [4] and are now extensively used for WWTP design, operation, calibration, and optimization [5]. Benchmark Simulation Model No. 1 (BSM1) was developed to complete the ASM1 concept with a defined plant configuration and it became a benchmark procedure for performance evaluation criteria. BSM1 was upgraded to Benchmark Simulation Model No. 2 (BSM2), as an extended plant model including the primary clarifier and anaerobic digestion sludge treatment subunits [6]. BSM1 and BSM2 were devised to assist the application of different control strategies and the evaluation of their performance [7].

The benefits of using mathematical models, either mechanistic or statistical, consist in their ability to estimate the future evolution of the main process variables, under various conditions, and their use for process design or operation improvement. The general issue of reducing greenhouse gas emissions is a growing research topic [8,9]. WWTP greenhouse gas emission reduction may also be investigated by specially developed models [10]. In addition, the mathematical model can be directly involved in automatic-control-system design, performance testing, and tuning, to achieve setpoint tracking and efficient disturbance rejection. Models have become indispensable prerequisites for model-based control solutions by their explicit incorporation in the advanced controllers.

Monitoring and control of WWTP process variables are achieved with various degrees of completeness and accuracy, depending on the wastewater influent, particular bioreactors' configuration, and actual instrumentation available in each plant. Accurate measurements of the composition, flow rate, and temperature, both for influent and process streams, are required for appropriate monitoring and successful control of the process. Besides, effective process control is dependent on the ability to detect sensor faults early [11]. The monitoring and control of the WWTP target its specific equipment, such as biological reactors, settlers, pumps, air blowers, control valves, diffusers, and the particular pipeline system [12]. The composition measurements involve sensors used to monitor the dissolved oxygen, nitrates and nitrites, ammonia, phosphorous, suspended solids, and organic matter [2]. Reliable information provided by these sensors is essential for the safe and efficient management of the whole plant.

Fault detection of sensors is complex and difficult to achieve in wastewater treatment processes because of the high interaction between variables and the continuous change of the plant influent. Notably, the online monitoring system of WWTPs collects a large amount of useful dynamic data [13]. For the municipal WWTP considered in the present study, the daily data set consists of hundreds of thousands individual measurements. Despite the large data set, only a few of the measurements represent meaningful information for the fault identification objective. Different methods have been developed in order to data mine and select the significant variables for fault detection. Multivariate statistical process control (MSPC) methods such as control charts [14,15], partial least squares (PLS) [16,17], independent component analysis (ICA) [18,19], and principal component analysis (PCA) [11,20] have been used for in-deep monitoring, feature extraction, and fault detection. Due to their intrinsic detection capacity, multivariate statistical methods show potential for efficiently finding the faults occurring in time-varying, ill-defined, and non-linear systems [21–23]. Multivariate statistical methodologies are preferred over standard Statistical Process Control approaches because they directly address and work with latent variables, leading to efficient fault detection [24].

Overall economic and environmentally safe WWTP operation largely relies on the nitrification process and its associated aeration control system. The latter is essentially assisted by dissolved oxygen (DO) sensor measurements. Aeration control system effi-

ciency can be dramatically affected due to erroneous information provided by the DO sensor, leading to the deterioration of expected control functioning and with negative consequences on degrading the effluent quality, increasing the energy consumed, decaying the environmental sustainability performance, or even temporary plant shutdown. Few studies have addressed the DO sensors faults issue. One study developed a PCA method for detecting three different fault types [25], one approached a failure type of the DO sensor fault and of the level sensor [14], another paper investigated the wrong output signal originating from the sensor [26], and a study used PCA to detect deviations due to sensor clogging [27]. Other studies focused on implementing several detection methods using set membership identification and radial-basis neural network [28], binary classifiers [29], or impulse response-based methods [30].

The purpose of the present study was to investigate and propose new detection solutions for specific faults that occur due to defective DO sensors, such as: bias, drift, wrong gain, loss of accuracy, fixed value, or complete failure. The novelty and innovation of this work consists in comparatively revealing their occurrence and effects, associated with highlighting reliable improvements brought to the WWTP control system by the PCA-based models aiming to detect different types of DO sensor faults.

The paper is organized to present (i) the first-principle wastewater treatment plant calibrated model used for carrying out dynamic simulations, followed by (ii) the presentation of the core theoretical aspects of the PCA fault-identification approach for (iii) the six different types of investigated DO sensors faults, in association to their software implementation, and (iv) results and discussions of the performance obtained by the proposed PCA fault-identification methodology, and ending with (v) the conclusions of the investigations for improving the operation of the municipal wastewater treatment plant used as a case study.

2. Materials and Methods

2.1. Process Model

The municipal WWTP investigated in this study is characterized by an anaerobic-anoxic-aerobic (A²O) configuration. After wastewater enters the WWTP, the influent is the subject of mechanical filtration, sand and fats separation, followed by primary sedimentation. The wastewater leaves the primary clarifier and reaches the biodegradation tanks, where the activated sludge technology performs the carbon, nitrogen, and phosphorous removal. Biodegradation basins were divided into three different zones: anaerobic, anoxic, and aerobic. The growth of phosphorus-accumulating microorganisms takes place in the anaerobic zone, while the reduction of the nitrates and nitrites (NO) formed in the aerobic reactors takes place in the anoxic zone [31,32]. Physical separation is performed in the secondary settler where the sludge is separated from the semiliquid mixture. The treated water is discharged to the emissary river. Part of the sludge from the secondary clarifier is sent to the anaerobic digestion unit, while another part is returned to the anaerobic biodegradation tank. The latter stream forms the external recycle. An additional recycle stream, denoted as the nitrate or internal recycle, is used to return nitrates from the aerobic reactor into the anoxic zone in order to perform the denitrification step [33]. Tables 1 and 2 present the principal design and operating parameters of the municipal plant.

Table 1. Average flow rates of the municipal plant.

Flow Name	Mean Value	Measurement Unit
Influent flow rate	116,300	
Air flow (total)	297,300	
Nitrate recycling flow	107,400	m ³ /day
Return activated sludge flow	112,500	
Excess waste flow	890	

Table 2. Main dimensions of equipment units at the municipal plant.

Variable		Mean Value	Measurement Unit
Primary clarifier	Area	2125	m ²
	Height	3.5	m
Anaerobic bioreactor	Volume	9015	m ³
Anoxic bioreactor	Volume	12,678	m ³
Aerobic bioreactor 1	Volume	11,022	m ³
Aerobic bioreactor 2	Volume	11,022	m ³
Aerobic bioreactor 3	Volume	11,022	m ³
Secondary settler	Area	67,824	m ²
	Height	3	m

BSM1 and ASM1 represent the foundation of the WWTP model developed and studied in the present paper. The necessary modifications were applied to the original BSM1 model, in order to comply with the A²O configuration of the investigated municipal WWTP. The model of the WWTP consists of a set of differential and algebraic equations that describe each of the main structural units: primary settler [34], anaerobic bioreactor, anoxic bioreactor, three aerobic bioreactors [3], and secondary settler [35]. Data for dry weather were collected from municipal WWTP measurements and underwent a process of reconciliation. The modified BSM1 model was updated with the plant collected data and calibrated accordingly. For this study, a previously developed and calibrated MATLAB model of the municipal WWTP, based on ASM1, BSM1 and MATLAB & SimulinkTM software, was used to conduct dynamic-state simulations of both normal and abnormal DO sensor operation [36]. The calibration of the model parameters was performed based on optimization. The mathematical model was implemented in C++ programming language and compiled as MATLAB executable files to gain simulation speed of the Simulink s-functions and to spare the computation resources.

The importance of automatic control is largely recognized for both the WWTP stable and safe operation, and for the counteraction of pollution propagation in the emissary river [37,38]. The synergy created by automatic control and optimization is highly valuable [39,40]. Two automatic control structures were used in this study: a dissolved oxygen control loop and a nitrates and nitrites concentration control loop. The first loop controlled DO concentration in the third aerated bioreactor, while the second loop controlled the nitrates and nitrites concentration in the anoxic bioreactor. The DO setpoint value of 2 mg O₂/L in the last aerated reactor was imposed for the DO concentration control loop [41]. In order to change the air flow rate entering the aerated reactors, the DO controller computed the control signal to manipulate air control valves. The air flow rate computed by the DO controller is unequally distributed in the sequence of the three aerobic reactors. It used the following weighting factors: 0.9 for aerobic reactor three, 0.5 for aerobic reactor four, and 0.3 for aerobic reactor five. This distribution scheme was established on the basis of oxygen requirements for the nitrification bioreactors. For the second control loop, the nitrates and nitrites (NO) concentration in the second (anoxic) reactor was controlled by manipulating the nitrate recycling flow rate. A setpoint value of 0.01 mg N/L was used for this control loop. Proportional-Integral (PI) controllers were considered for the two feedback control loops. The control system design is presented in Figure 1.

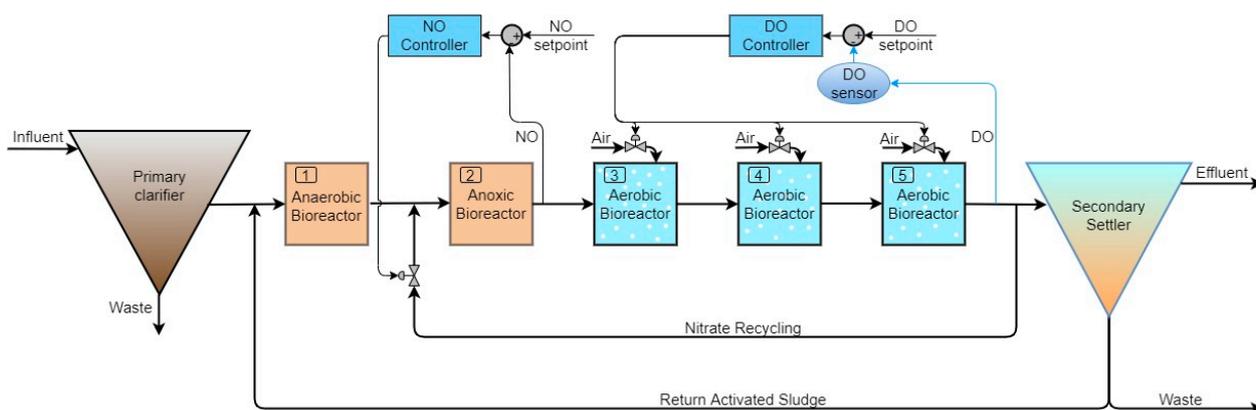


Figure 1. Schematic of the WWTP layout and proposed control system.

Dynamic state simulations for both normal and abnormal operation (i.e., caused by faulty DO sensor operation) were carried out for each type of investigated faults and data were subsequently processed for fault detection.

2.2. Principal Component Analysis

PCA is a data-mining methodology relying on a process model built with ordinary process data and used in statistical process monitoring (SPM). PCA models are principally used to identify from raw data the correlation between process variables. The method reduces the large model training data set by linearly transforming it into a smaller one that still contains the key information. The resulting data set consists of a score matrix and a loadings matrix that include information revealed by the reduced set of variables [42].

Consider the data matrix $X \in \mathbb{R}^{m \times n}$ where m represents the number of samples, and n the number of process variables, collected during normal operation. The matrix X is normalized to zero-mean and unit-variance matrix \hat{X} . To reduce the original space dimension, matrix \hat{X} is decomposed in a score matrix called T and a loading matrix named P . The expression of this decomposition is:

$$\hat{X} = t_1 \cdot p_1^T + t_2 \cdot p_2^T + \dots + t_k \cdot p_k^T + E \tag{1}$$

where t_i is a score vector, p_i is a loading vector with $i = 1, 2, 3, \dots, k$, and E is the residual matrix. The number of selected principal components, k , is calculated using the cumulative percent variance (CPV). It measures the percent of variance encapsulated by the first k principal components [43]. It can be computed using the equation:

$$CPV_k = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^n \lambda_j} \cdot 100 \tag{2}$$

where λ_j is the j^{th} eigenvalue of the covariance matrix, C :

$$C = \frac{1}{m-1} \cdot \hat{X}^T \cdot \hat{X} \tag{3}$$

and

$$C = V \cdot S \cdot V^T \tag{4}$$

where the columns of matrix V are the eigenvectors of the covariance matrix C and S is the diagonal matrix that contains the eigenvalues of C , arranged in decreasing order of magnitude.

The original and normalized data space can be reconstructed according to the following relationship:

$$\dot{\mathbf{X}} = \mathbf{T} \cdot \mathbf{P}^T + \mathbf{E} \quad (5)$$

2.3. Hotelling's T^2 Statistic

In order to measure variations in the principal component subspace (PCS), a multivariate version of a Student's T-test, the Hotelling's T^2 value [44], can be calculated as the sum of the squares of a sample vector \mathbf{x} :

$$T^2 = \mathbf{x}^T \cdot \mathbf{P} \cdot \mathbf{S}_k^{-1} \mathbf{P}^T \cdot \mathbf{x} \quad (6)$$

where \mathbf{S}_k is a matrix formed by the first k lines and columns of the matrix \mathbf{S} . T^2 is a measure of the square distance from the sample (observation) point to the origin point of the principal-components model space.

For a process to be considered normal, it is necessary that the calculated T^2 be less than or equal to the threshold value T_{α}^2 :

$$T_{\alpha}^2 = \frac{(m-1) \cdot k}{(m-k)} \cdot F_{\alpha}(k, m-k) \quad (7)$$

where $F_{\alpha}(k, m-k)$ is the critical value of the F-distribution with α being the confidence level that takes values between 90% and 95% and $k, m-k$ being the degrees of freedom.

2.4. Squared Prediction Error Statistics

Variations in the residual subspace (RS) are measured using the Square Prediction Error, also known as the Q-Statistic [44], by the sum of the squares of residuals:

where \mathbf{r} is the residual vector:

$$\text{SPE} = \mathbf{r}^T \cdot \mathbf{r} \quad (8)$$

$$\mathbf{r} = (\mathbf{I} - \mathbf{P} \cdot \mathbf{P}^T) \cdot \mathbf{x} \quad (9)$$

with \mathbf{I} as the identity matrix, for a sample vector \mathbf{x} . SPE is a measure of the distance from the sample point to the model plane of the principal components.

The threshold for SPE can be determined by:

$$\text{SPE}_{\alpha} = \theta_1 \cdot \left[\frac{h_0 \cdot C_{\alpha} \cdot \sqrt{2 \cdot \theta_2}}{\theta_1} + 1 + \frac{\theta_2 \cdot h_0 \cdot (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \quad (10)$$

with C_{α} the standard normal deviation value for $1-\alpha$ percentile and where:

$$h_0 = 1 - \frac{2 \cdot \theta_1 \cdot \theta_3}{3 \cdot \theta_2^2} \quad (11)$$

$$\theta_i = \sum_{j=k+1}^n \lambda_j^i \text{ for } i = 1, 2, 3 \quad (12)$$

2.5. Sensors Faults

Maintaining the required standards for water quality and efficiency of the wastewater treatment process may be ensured when abnormal operating circumstances are promptly identified and corrected. Many of the problems that occur in WWTPs originate from faulty sensors. Sensor faults are caused by several issues such as questionable-quality signal [45], poor connection, or sensor failure [2].

Considering the major role of the Dissolved Oxygen sensor for the overall WWTP operation, the identification of several DO sensor types of faults was investigated in

the present study. The considered six types of sensors faults are defined and described as follows:

- Bias, also known as a shift or an off-set of the sensor's generated signal values, occurs when the sensor is miscalibrated and the delivered value is shifted in contrast with the true one [46];
- Drift, appears as a deviation of the real value that fluctuates in time [46];
- Wrong gain, is also known as a calibration error; occurs when the slope of the sensor is inaccurately established in the calibration step [47];
- Loss of accuracy is when the sensor exhibits a value that is characterized by imprecision around the true value; is often erroneously qualified as the true value of the measurement [46];
- Fixed value is when the sensor always displays a constant value [47];
- Complete failure (minimum or maximum) is when the measured value is either the minimum or the maximum calibration value of the sensor [47].
- The previously presented faults are graphically exemplified in Figure 2.

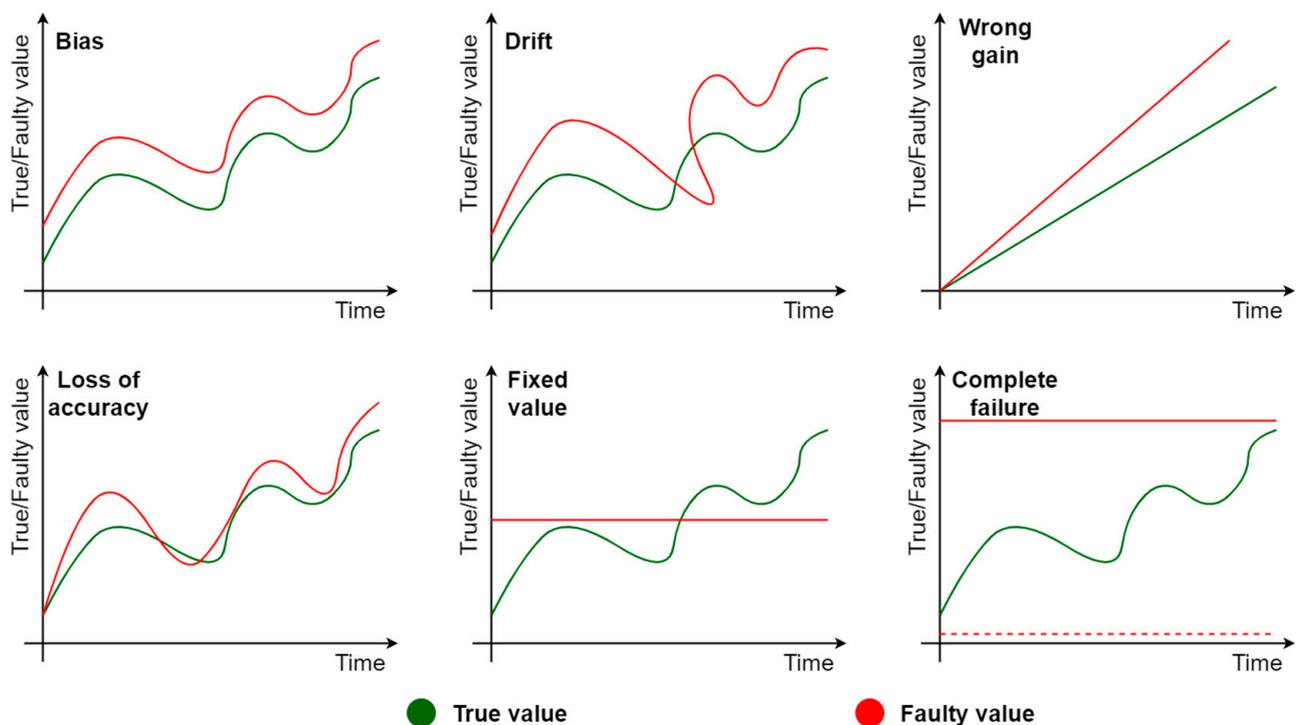


Figure 2. Graphical representation of the considered sensors faults.

Figure 2 shows the comparison between the normal (true) vs. the faulty sensor signals, as they evolve in time.

2.6. Methodology of Faults Implementation

Dedicated software applications were built in order to reproduce the real behavior of the six fault types considered for the DO concentration sensor. They were implemented in the WWTP dynamic model. For each type of faults, the generated faulty value was sent as a measured process variable to the Proportional-Integral DO controller. The DO control loop manipulated the air flow rate accordingly, such as to bring with zero offset the measured DO to the desired setpoint of 2 mg O₂/L.

The bias fault was characterized by a fixed difference between the true value and the faulty output value of the sensor. For this fault type, the DO measured concentration was considered to show a bias of +1.5 mg O₂/L.

The drift was simulated as an increasing bias, being described by a ramp varying signal in time, that was added to the true DO value. The constant value of 0.05 mg O₂/L (i.e., 2.5% of the 2 mg O₂/L nominal setpoint value of the DO control loop) was integrated in time and added to the DO true value, in order to mimic the drift.

The wrong gain fault was simulated by the loss of effectiveness of the sensor due to a gain factor different to 1. It consisted of a wrong calibration gain of 1.4, that led to an inadequate correlation factor between the input and output of the sensor. The wrong gain was smoothly introduced in time, according to a first-order filter with the time constant of 0.3 days.

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

The fixed-value fault was characterized by a constant value of the sensor output signal, irrespective of the changes of the true DO process variable value. For this study, a constant value of 2.2 mg O₂/L was considered (i.e., 10% higher than the 2 mg O₂/L nominal setpoint value of the DO control loop).

The complete failure of a sensor was again characterized by constant values of the sensor output signal, but they were the minimum or the maximum values of the sensor calibration interval. The two extreme values of 0.1 mg O₂/L and of 6 mg O₂/L were chosen for the complete failure minimum and maximum alternatives.

Statistical theory is applied to process monitoring and relies on the assumption that characteristics of the data variations change only when an abnormal condition (fault) occurs in the process [48]. The failure of a single sensor at a time was also assumed.

3. Results and Discussion

3.1. Normal and Abnormal Data Sets

In order to generate the necessary data sets, seven different simulations were built, one for the normal operation and six for faulty operation. The plant model was simulated for 168 days of operation, in the WWTP configuration with the DO and NO control loops. The faults were implemented starting on the 140th day of the simulation scenarios. The first 100 days of normal operation were used to obtain the regular state of the plant. Data obtained by simulation during a period of the next 40 days (i.e., the days ranging from 100 to 139) of normal WWTP operation were collected with a sampling time of 15 min and used to construct data matrix **X** of the PCA model. Data generated in the time period starting with the 140th and 160th days of abnormal (DO sensor fault) operation were used to test the fault detection capability of the PCA model.

The set of twenty WWTP process variables considered for building the PCA model consisted of secondary settler concentration and flow rate variables of the clean water effluent (6 variables), and sludge concentration and flow rate variables of the bottom effluent (14 variables), including temperature. The 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 clean effluent and the last fourteen for its bottom. 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 [49].

3.2. PCA Model Construction

A set of 3840 observations was used from the normal operating period of 40 days in order to set up the PCA model. The obtained data matrix \mathbf{X} of 3840×20 dimensions represented the training matrix. Scores matrix \mathbf{T} and loadings matrix \mathbf{P} were obtained after scaling the data matrix. A threshold value of 98% was chosen for CPV_k . As a result, the appropriate number of principal components k that led to a good representation of the whole data set was considered to be 8. This selection is based on the CPV_k values presented in Table 3. The number of chosen principal components was confirmed by a Scree test [50,51]. The scree plot presented in Figure 3 shows that the selected number of 8 eigenvalues was comprehensive, and they properly captured the variability of the process. Based on this selection the covariance matrix \mathbf{C} and the diagonal matrix \mathbf{S} were determined.

Table 3. Cumulative percent variance and threshold values for different numbers of selected principal components.

Number of Selected Principal Components (k)	$CPV_k(\%)$	T_k^2	SPE_α
7	97.25	18.54	1.92
8	98.24	20.16	1.26
9	98.92	21.75	0.75

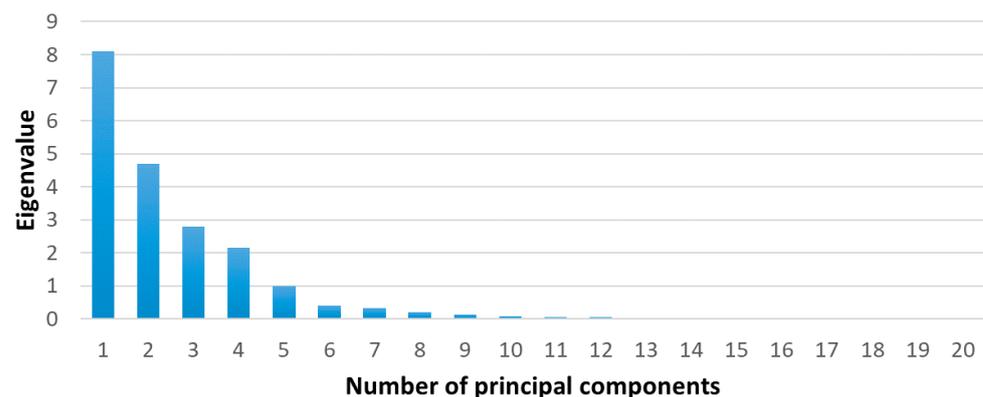


Figure 3. Scree plot representing the model eigenvalues.

Using a 99% confidence level, a threshold value of 20.16 was determined for T_α^2 and a threshold of 1.26 for SPE_α . Any vector \mathbf{x} of measured variables showing a value of either T^2 or of SPE that was higher than the two corresponding thresholds indicated abnormal sensor operation.

3.3. Fault Detection

Seven testing data matrices were constructed with the vectors of abnormal operation measured variables. One was created for each of the six types of faults, excepting the complete-failure fault type where two matrices were built for, respectively, the minimum and the maximum faulty DO sensor. A number of 1920 samples, from simulation day 140 and up to day 160 of the sensor, were used for each testing data matrix. After determining the values of T^2 and SPE for each of the sample vectors of the testing matrices, the values obtained were compared with the previously computed thresholds in order to determine faulty sensor operation. Graphic representations of T^2 and SPE values for each of the normal training and the faulty testing samples, in association to the T_α^2 and SPE_α thresholds, clearly highlight the presence of the DO sensor fault.

Firstly, the training data set was tested with the proposed PCA model. As may be observed in Figure 4, all of the T^2 and SPE values from this data set were confirmed to belong to the normal operation (faulty-free values). Additionally, the T^2 plot reveals that

the difference between the normal operation values and their associated threshold was larger than the difference shown by the SPE plot. This means that in the normal operation mode, the most relevant data are provided by the SPE statistics.

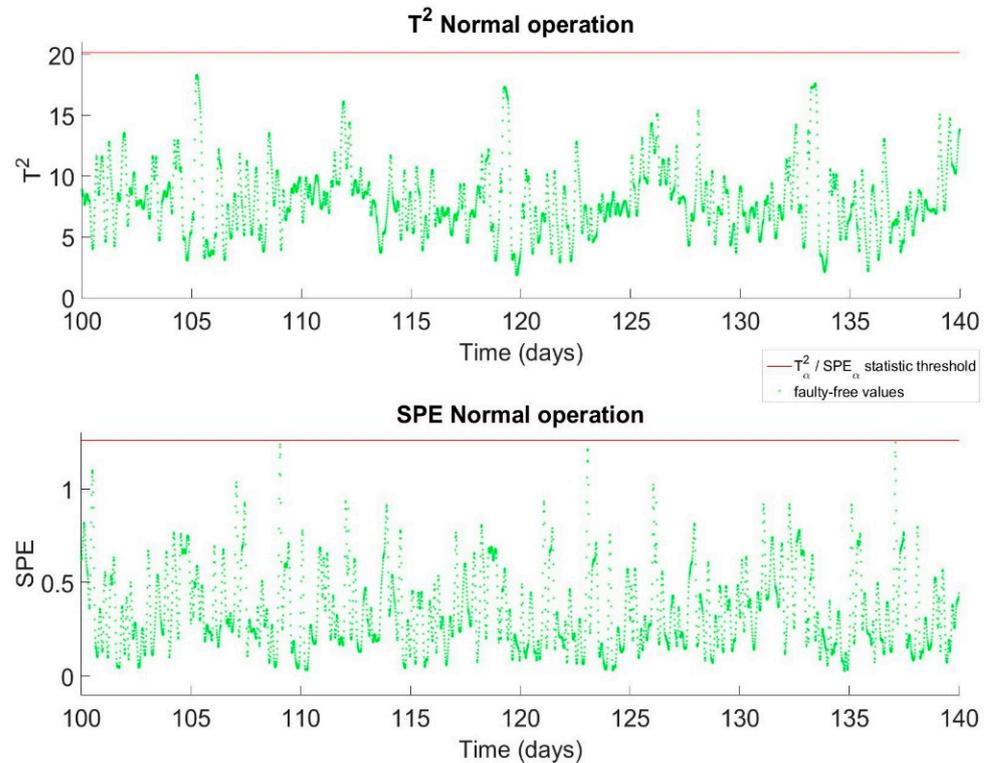


Figure 4. T^2 and SPE plots of PCA monitoring the normal operation data set.

The fault detection results are presented in Figures 5–11, for each type of investigated faults.

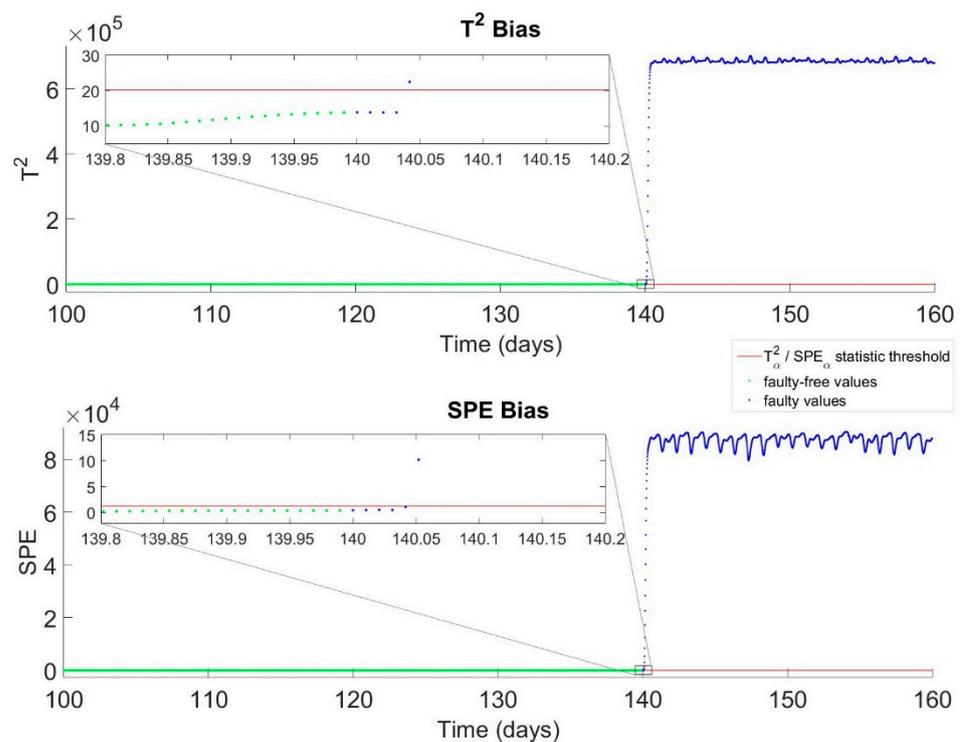


Figure 5. T^2 and SPE plots of PCA detecting the bias-fault data set.

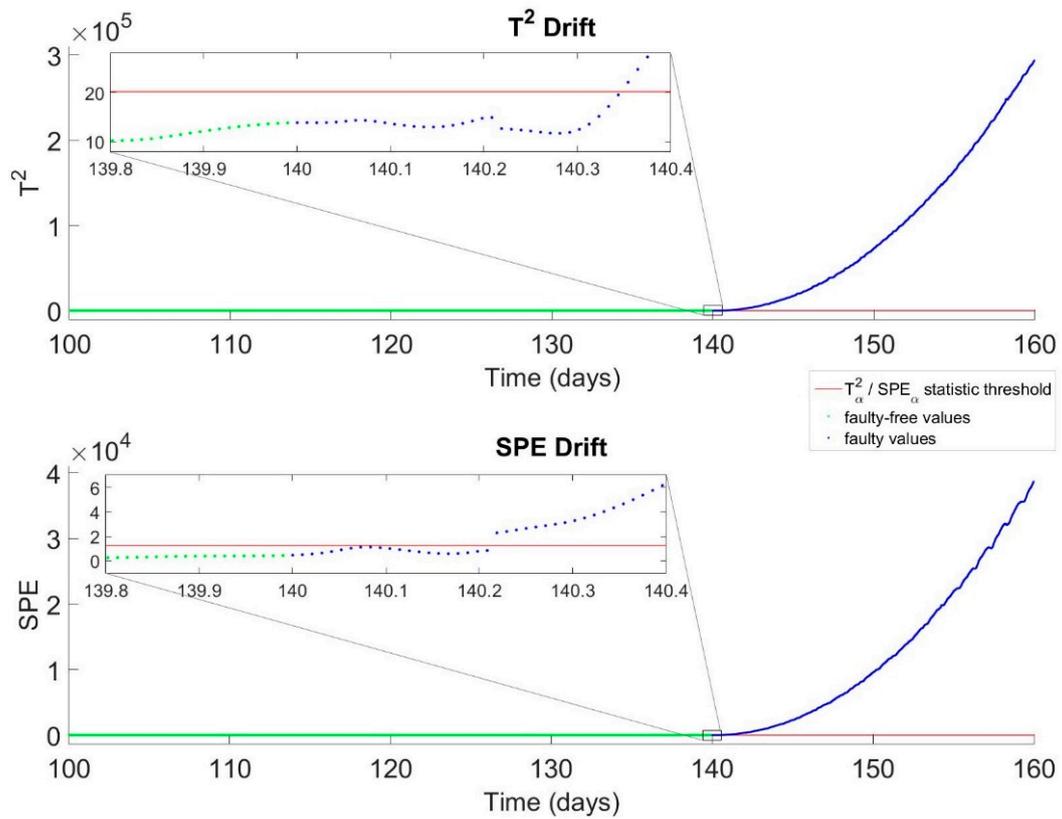


Figure 6. T^2 and SPE plots of PCA detecting the drift fault data set.

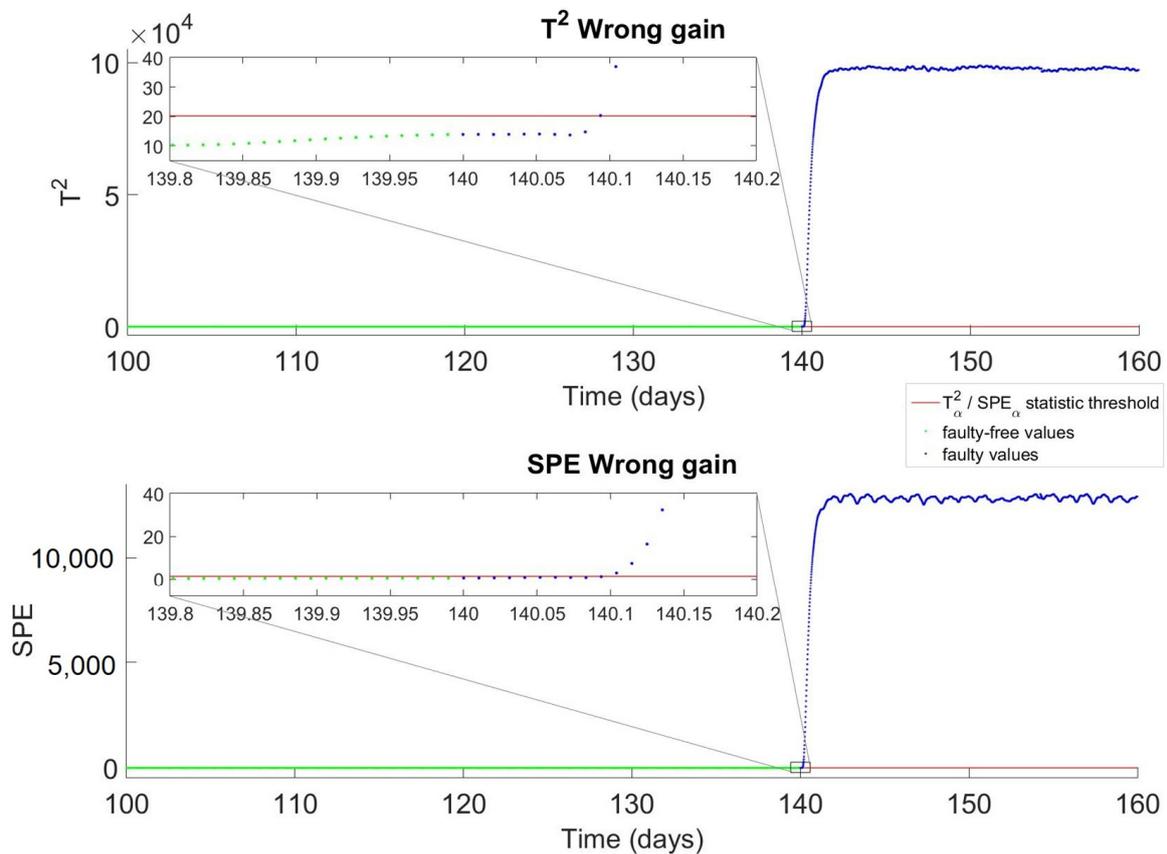


Figure 7. T^2 and SPE plots of PCA detecting the wrong-gain data set.

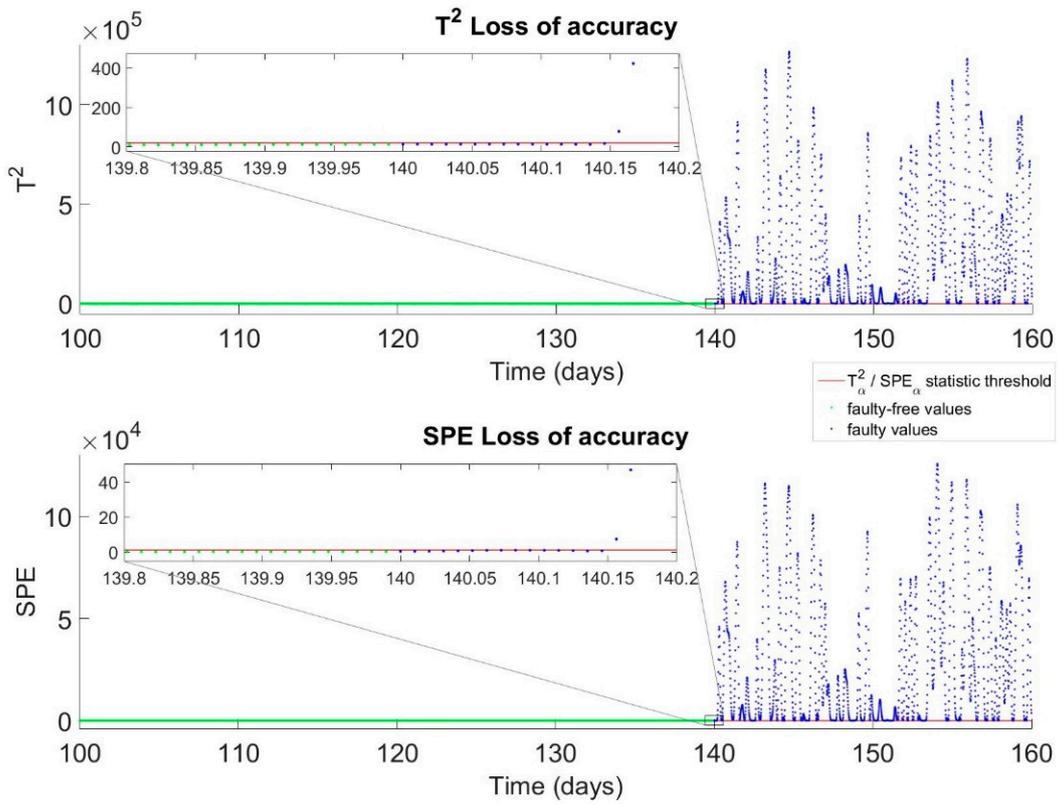


Figure 8. T^2 and SPE plots of PCA detecting the loss-of-accuracy data set.

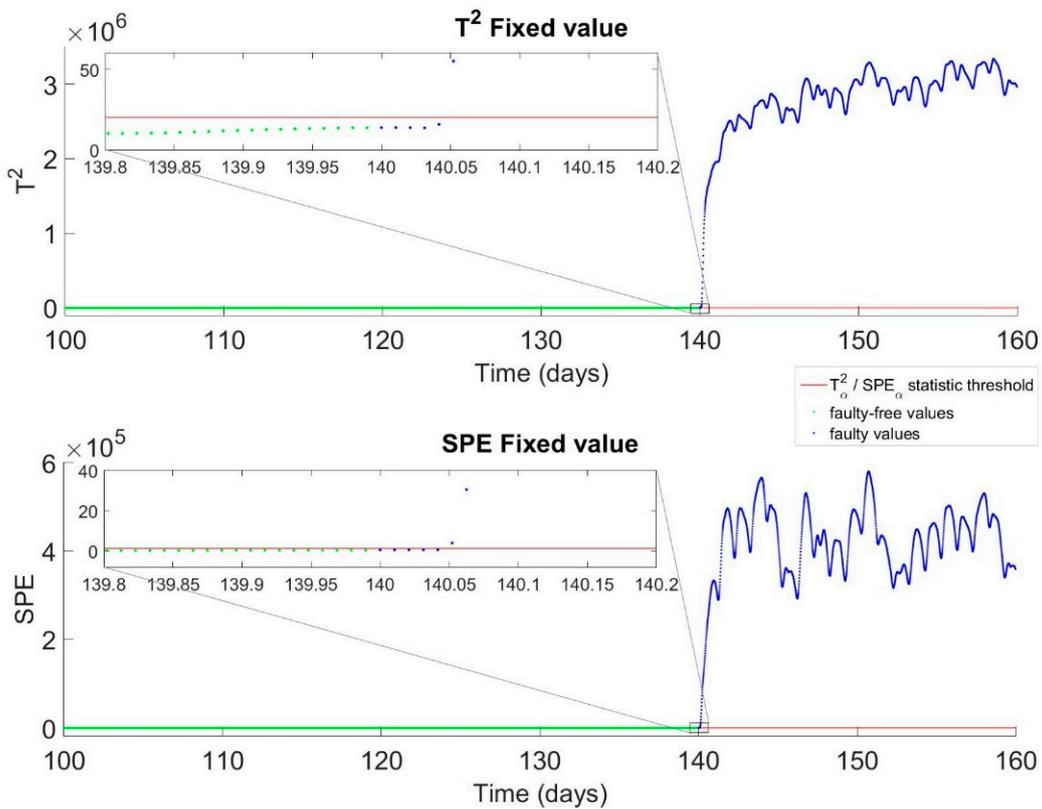


Figure 9. T^2 and SPE plots of PCA detecting the fixed-value data set.

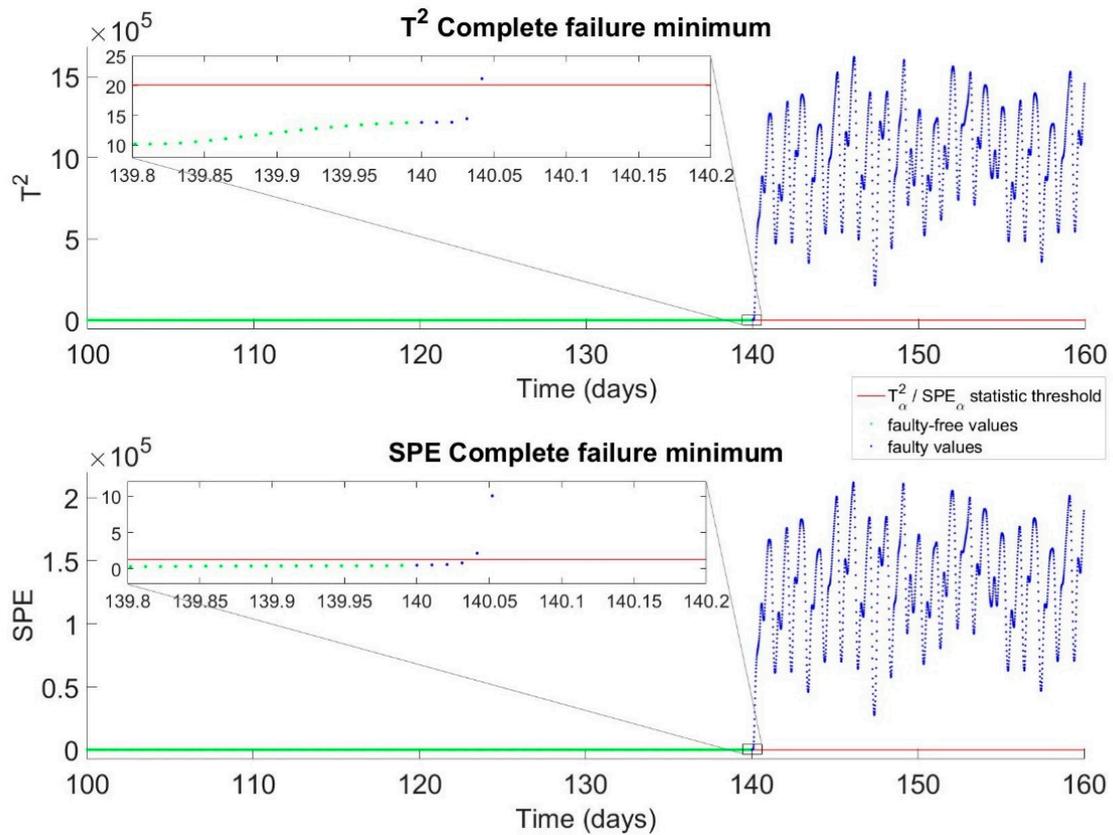


Figure 10. T² and SPE plots of PCA detecting the complete failure-minimum data set.

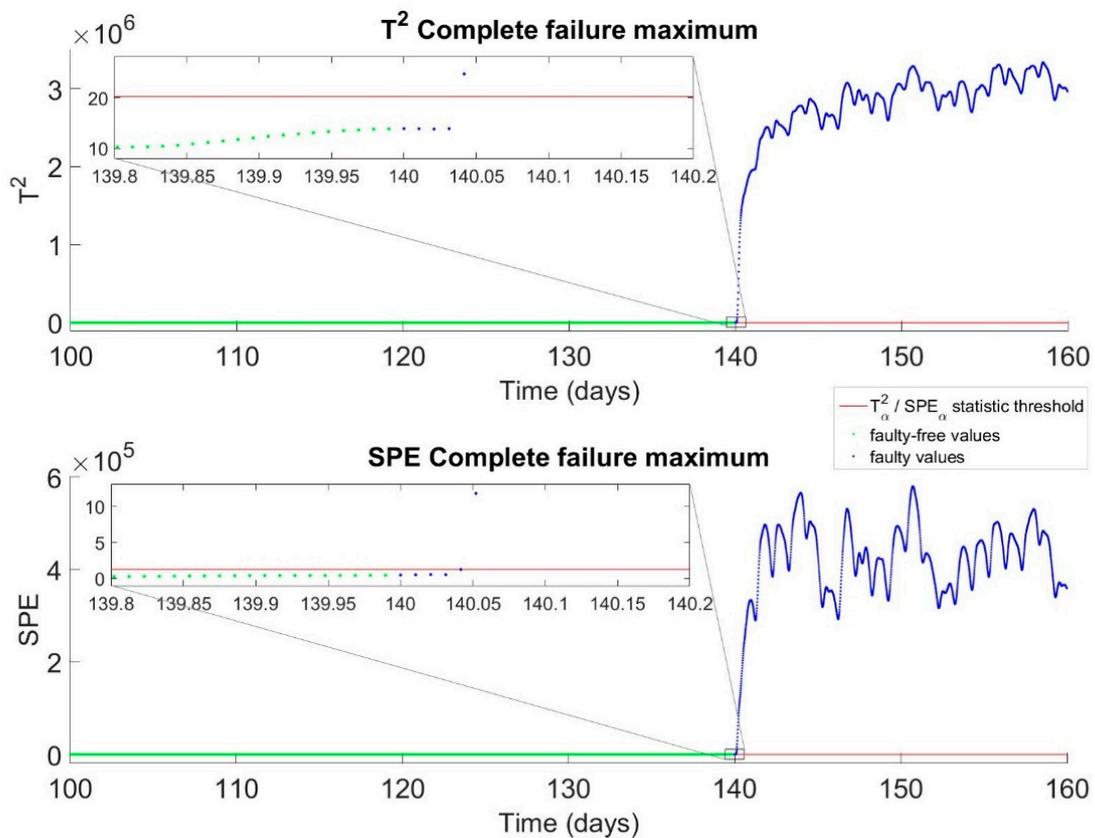


Figure 11. T² and SPE plots of PCA detecting the complete failure-maximum data set.

Results show that WWTP DO sensor faults were successfully detected by the proposed PCA approach. For the majority of the DO sensor fault types, the statistic thresholds were overridden soon after the DO sensor fault acted. This is presented in the detailed parts of Figures 5–11. Moreover, the T^2 and SPE plots revealed high overshooting values of the statistic thresholds during the fault sensor operation.

Figures 5, 10 and 11 show that the detection time of bias and complete failure, either maximum or minimum, were the types of sensor faults which are the most promptly discovered. The fixed-value sensor fault type, presented in Figure 9, closely followed in this ranking. It revealed a detection time of about 1.5 h. The wrong-gain fault detection time was roughly double, when compared to the previous ones, as may be seen in Figure 7. Figure 8 depicts the loss-of-accuracy fault, which was detected in a slightly longer time of 4 h, but this depends on the random amplitude of the fault component value. The fault type requiring the longest time to be uncovered was the sensor drift type of fault. This may be observed from Figure 6 and it revealed a detection time of about 5.5 or 8.75 h. A longer detection time for the drift sensor fault was expected, as the faulty signal grows slowly in time and it passes over the statistical threshold only when its amplitude becomes significant.

Compared to previously reported results, that typically investigated only one specific type of DO sensor fault, the present study had a more comprehensive investigation of six sensor fault types. By exception, one study developed the PCA-based detection for three fault types, i.e., for drift, bias, and precision degradation (loss of accuracy) [25], while another investigated the failure of the DO sensor and the level sensor [11].

In the present study, for the majority of the investigated DO sensor fault types, the Hotelling's T^2 statistic and the Squared prediction error statistic performed the detection with the same efficiency. However, the results of the present work showed that the SPE detection method proved to be more effective than the T^2 statistic for the drift type of fault. Previous results showed that faults could be better detected with SPE statistics because fault-generated changes were mainly identified in RS, rather than in PCS [11]. The results of the present work are also in agreement with another previous study that investigated the detection time for four different faults, out of which only one was related to the DO sensor located in the aerobic reactor [26]. However, the present study revealed that for this specific fault type, the PCA methodology detected the fault in a period of a few hours, while in the prior study the drift fault of a clogging DO sensor was only detected by different PCA algorithms in a period of more than one day [27].

The PCA investigation revealed the most important process variables that contribute, in a decreasing order, to the first four Principal Components that explain 93.18 % of the total variance. They are for Principal Component 1: N_{total} , S_{NO_bottom} ; for Principal Component 2: X_{ND} , X_S ; for Principal Component 3: T, TKN, and; for Principal Component 4: COD, S_O .

4. Conclusions

In this study, a PCA-based sensor fault detection method was proposed and its performance was tested, for the DO sensor of the WWTP used as a case study. The DO sensor is part of the control loop designed to provide efficiency to the nitrification process. Keeping the DO value in the aerated reactors at the desired setpoint was achieved by manipulating the air flow rate, and aeration was of primary importance for WWTP energy consumption.

The detection algorithm and its applicability to six different DO sensor fault types were individually investigated. The considered fault sensor types were: bias, drift, wrong gain, loss of accuracy, fixed value, and complete failure. A calibrated WWTP model was used to generate the data sets corresponding to normal and faulty sensor operation. Both Hotelling's T^2 and SPE statistical tests performed successfully for identifying the presence of the fault, yet in some cases the latter had a faster detection ability. Bias, fixed-value, and complete-failure faults were detected promptly. They showed less than 2 h detection time, following the DO sensor fault occurrence. The loss-of-accuracy detection time depends on the random component of the fault, but it implied a longer time, of about

3–4 h. The detection time of the drift and wrong-gain faults took longer than the detection of the other faults, up to 8–9 h, but strongly depended on the additive or multiplicative magnitude of these fault types.

The PCA fault detection algorithm can be implemented within the software of the Supervisory Control and Data Acquisition system, used for monitoring and control of the WWTP. The available sensors or soft sensors based on the available measurements have to provide the information on the most relevant variables. The operating staff must be trained to interpret the fault detection information, relating to the fault diagnosis analysis, and to take appropriate normal operation restoration actions. Auto-correction algorithms and adaptive PCA may be involved in the more advanced stages of the implementation.

This study has shown the potential of the methodologies and demonstrated the efficiency of the presented PCA-based fault detection approach in detecting the DO sensor faults of the controlled WWTP. Their performance is of major practical importance for complex, non-linear, time-varying and controlled wastewater treatment processes, in support of their safe and efficient management.

Author Contributions: Conceptualization, A.-V.L. and V.-M.C.; methodology, A.-V.L. and V.-M.C.; software, A.-V.L., V.-M.C. and M.S.-V.; investigation, A.-V.L., M.S.-V., and N.-B.M.; data curation, M.S.-V. and N.-B.M.; writing—original draft preparation, A.-V.L. and V.-M.C.; writing—review and editing, A.-V.L., M.S.-V., N.-B.M., and V.-M.C.; supervision, V.-M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the doctoral grant of Babes-Bolyai University.

Acknowledgments: The authors would like to acknowledge the support offered by Compania de Apa Someș S.A., World Federation of Scientists.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

- Alex, J.; Benedetti, L.; Copp, J.; Gernaey, K.V.; Jeppsson, U.; Nopens, I.; Pons, M.N.; Rieger, L.; Rosen, C.; Steyer, J.P.; et al. *Benchmark Simulation Model No. 1 (BSM1)*; Lund University: Lund, Sweden, 2008; pp. 2–29.
- Mamandipoor, B.; Majd, M.; Sheikhalishahi, S.; Modena, C.; Osmani, V. Monitoring and detecting faults in wastewater treatment plants using deep learning. *Environ. Monit. Assess.* **2020**, *192*, 148. [[CrossRef](#)] [[PubMed](#)]
- Rieger, L.; Gillot, S.; Langergraber, G.; Ohtsuki, T.; Shaw, A.; Takacs, I.; Winkler, S. *Guidelines for Using Activated Sludge Models*; IWA Publishing: London, UK, 2012; pp. 11–147, ISBN 9781780401164.
- Henze, M.; Gujer, W.; Mino, T.; van Loosedrecht, M. *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3*; IWA Publishing: London, UK, 2002; pp. 1–120, ISBN 9781780402369.
- 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 Sci. Technol.* **2011**, *63*, 2164–2182. [[CrossRef](#)] [[PubMed](#)]
- Jeppsson, U.; Rosen, C.; Alex, J.; Copp, J.; Gernaey, K.V.; Pons, M.N.; Vanrolleghem, P.A. Towards a benchmark simulation model for plant-wide control strategy performance evaluation of WWTPs. *Water Sci. Technol.* **2006**, *53*, 287–295. [[CrossRef](#)] [[PubMed](#)]
- Jeppsson, U. The benchmark simulation modelling platform—Areas of recent development and extension. *Lect. Notes Civ. Eng.* **2017**, *4*, 81–91. [[CrossRef](#)]
- Miranda, T.; Nogales, S.; Román, S.; Montero, I.; Arranz, J.L.; Sepúlveda, F.J. Control of several emissions during olive pomace thermal degradation. *Int. J. Mol. Sci.* **2014**, *15*, 18349–18361. [[CrossRef](#)] [[PubMed](#)]
- Zheng, X.; Streimikiene, D.; Balezentis, T.; Mardani, A.; Cavallaro, F.; Liao, H. A review of greenhouse gas emission profiles, dynamics, and climate change mitigation efforts across the key climate change players. *J. Clean. Prod.* **2019**, *234*, 1113–1133. [[CrossRef](#)]
- Parravicini, V.; Svardal, K.; Krampe, J. Greenhouse gas emissions from wastewater treatment plants. *Energy Procedia* **2016**, *97*, 246–253. [[CrossRef](#)]
- Tao, E.P.; Shen, W.H.; Liu, T.L.; Chen, X.Q. Fault diagnosis based on PCA for sensors of laboratorial wastewater treatment process. *Chemom. Intell. Lab. Syst.* **2013**, *128*, 49–55. [[CrossRef](#)]
- Morera, S.; Corominas, L.; Rigola, M.; Poch, M.; Comas, J. Using a detailed inventory of a large wastewater treatment plant to estimate the relative importance of construction to the overall environmental impacts. *Water Res.* **2017**, *122*, 614–623. [[CrossRef](#)]

13. Olsson, G.; Carlsson, B.; Comas, J.; Copp, J.; Gernaey, K.V.; Ingildsen, P.; Jeppsson, U.; Kim, C.; Rieger, L.; Rodriguez-Roda, I.; et al. Instrumentation, control and automation in wastewater systems—From London 1973 to Narbonne 2013. *Water Sci. Technol.* **2014**, *69*, 1373–1385. [[CrossRef](#)]
14. Schraa, O.; Tole, B.; Copp, J.B. Fault detection for control of wastewater treatment plants. *Water Sci. Technol.* **2006**, *53*, 375–382. [[CrossRef](#)]
15. Kazemi, P.; Giralt, J.; Bengoa, C.; Steyer, J.P. Data-driven fault detection methods for detecting small-magnitude faults in anaerobic digestion process. *Water Sci. Technol.* **2020**, *81*, 1740–1748. [[CrossRef](#)] [[PubMed](#)]
16. Chen, A.; Zhou, H.; An, Y.; Sun, W. PCA and PLS monitoring approaches for fault detection of wastewater treatment process. In Proceedings of the 25th International Symposium on Industrial Electronics (ISIE), Santa Clara, CA, USA, 8–10 June 2016; pp. 1022–1027. [[CrossRef](#)]
17. Liu, H.; Yang, J.; Zhang, Y.; Yang, C. Monitoring of wastewater treatment processes using dynamic concurrent kernel partial least squares. *Process Saf. Environ. Prot.* **2021**, *147*, 274–282. [[CrossRef](#)]
18. 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), Marrakech, Morocco, 23–25 June 2010; pp. 874–879. [[CrossRef](#)]
19. Yoo, C.K.; Lee, J.M.; Lee, I.B.; Vanrolleghem, P.A. Dynamic monitoring system for full-scale wastewater treatment plants. *Water Sci. Technol.* **2004**, *50*, 163–171. [[CrossRef](#)] [[PubMed](#)]
20. 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), Barcelona, Spain, 30 June–3 July 2009; Volume 42, pp. 1115–1125. [[CrossRef](#)]
21. Dong, Y.; Qin, S.J. A novel dynamic PCA algorithm for dynamic data modeling and process monitoring. *J. Process Control* **2018**, *67*, 1–11. [[CrossRef](#)]
22. Li, Z.; Yan, X. Ensemble model of wastewater treatment plant based on rich diversity of principal component determining by genetic algorithm for status monitoring. *Control Eng. Pract.* **2019**, *88*, 38–51. [[CrossRef](#)]
23. Corominas, L.; Villez, K.; Aguado, D.; Rieger, L.; Rosén, C.; Vanrolleghem, P.D. Performance evaluation of fault detection methods for wastewater treatment processes. *Biotechnol. Bioeng.* **2011**, *108*, 333–344. [[CrossRef](#)]
24. 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, Online Conference, 15 January 2009; pp. 55–60.
25. Lee, C.; Choi, S.W.; Lee, I.B. Sensor fault diagnosis in a wastewater treatment process. *Water Sci. Technol.* **2006**, *53*, 251–257. [[CrossRef](#)]
26. 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), Luxembourg, Luxembourg, 8–11 September 2015; pp. 1–7. [[CrossRef](#)]
27. Andersson, S.; Hallgren, F. Sensor fault detection methods applied on dissolved oxygen sensors at a full scale WWTP. In Proceedings of the 9th IWA Symposium on Systems Analysis and Integrated Assessment (Watermatex 2015), Gold Coast, Australia, 14–17 June 2015.
28. Li, X.; Chai, W.; Liu, T.; Qiao, J. Fault detection of dissolved oxygen sensor in wastewater treatment plants. In Proceedings of the 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 18–21 October 2020; pp. 225–230. [[CrossRef](#)]
29. Chistiakova, T.; Zambrano, J.; Samuelsson, O.; Carlsson, B. Binary classifiers applied to detect DO sensors faults during washing events. In Proceedings of the 2nd New Developments in IT & Water, Rotterdam, The Netherlands, 8–10 February 2015; pp. 1–8. [[CrossRef](#)]
30. Samuelsson, O.; Bjork, A.; Zambrano, J.; Carlsson, B. Fault signatures and bias progression in dissolved oxygen sensors. *Water Sci. Technol.* **2018**, *78*, 1034–1044. [[CrossRef](#)]
31. Ostace, G.S.; Cristea, V.M.; Agachi, P.S. Extension of activated sludge model no 1 with two-step nitrification and denitrification processes for operation improvement. *Environ. Eng. Manag. J.* **2011**, *10*, 1529–1544. [[CrossRef](#)]
32. Focht, D.D.; Chang, A.C. Nitrification and denitrification processes related to waste water treatment. *Adv. Appl. Microbiol.* **1975**, *19*, 153–186. [[CrossRef](#)]
33. Simon-Várhelyi, M.; Cristea, V.M.; Luca, A.V. Reducing energy costs of the wastewater treatment plant by improved scheduling of the periodic influent load. *J. Environ. Manag.* **2020**, *262*, 110294. [[CrossRef](#)]
34. Otterpohl, R.; Freund, M. Dynamic models for clarifiers of activated sludge plants with dry and wet weather flows. *Water Sci. Technol.* **1992**, *26*, 1391–1400. [[CrossRef](#)]
35. Takács, I.; Patry, G.G.; Nolasco, D. A dynamic model of the clarification-thickening process. *Water Res.* **1991**, *25*, 1263–1271. [[CrossRef](#)]
36. Várhelyi, M.; Cristea, V.M.; Brehar, M.; Nemeş, E.D.; Nair, A. WWTP model calibration based on different optimization approaches. *Environ. Eng. Manag. J.* **2019**, *18*, 1657–1670.
37. Cristea, V.M. Counteracting the accidental pollutant propagation in a section of the river Someş by automatic control. *J. Environ. Manag.* **2013**, *128*, 828–836. [[CrossRef](#)]
38. Ani, E.C.; Cristea, V.M.; Agachi, P.S.; Kraslawskp, A. Dynamic simulation of Someş, river pollution using Matlab and Comsol models. *Rev. Chim.* **2010**, *61*, 1108–1112.

39. Revollar, S.; Vilanova, R.; Vega, P.; Francisco, M.; Meneses, M. Wastewater treatment plant operation: Simple control schemes with a holistic perspective. *Sustainability* **2020**, *12*, 768. [[CrossRef](#)]
40. Simon-Várhelyi, M.; Cristea, V.M.; Luca, A.V.; Brehar, M.A. Optimization and control of aeration distribution in the wwtp nitrification reactor. *Rev. Roum. Chim.* **2020**, *65*, 601–609. [[CrossRef](#)]
41. Ostace, G.S.; Gal, A.; Cristea, V.M.; Agachi, P.S. Operational costs reduction for the WWTP by means of substrate to dissolved oxygen correlation—A simulation study. *Lect. Notes Eng. Comput. Sci.* **2011**, *2194*, 939–944.
42. 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 Qual.* **1990**, *1*, 41–51.
43. Valle, S.; Li, W.; Qin, S.J. Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods. *Ind. Eng. Chem. Res.* **1999**, *38*, 4389–4401. [[CrossRef](#)]
44. Jackson, J.E. *A User's Guide to Principal Components*; John Wiley & Sons: New York, NY, USA, 1991; ISBN 9780471725336.
45. Schneider, M.Y.; Carbajal, J.P.; Furrer, V.; Sterkele, B.; Maurer, M.; Villez, K. Beyond signal quality: The value of unmaintained pH, dissolved oxygen, and oxidation-reduction potential sensors for remote performance monitoring of on-site sequencing batch reactors. *Water Res.* **2019**, *161*, 639–651. [[CrossRef](#)]
46. Teh, H.Y.; Kempa-Liehr, A.W.; Wang, K.I. Sensor data quality: A systematic review. *J. Big Data* **2020**, *7*, 11. [[CrossRef](#)]
47. Rosen, C.; Rieger, L.; Jeppsson, U.; Vanrolleghem, P.A. Adding realism to simulated sensors and actuators. *Water Sci. Technol.* **2008**, *57*, 337–344. [[CrossRef](#)] [[PubMed](#)]
48. Chiang, L.H.; Russell, E.L.; Braatz, R.D. *Fault Detection and Diagnosis in Industrial Systems*; Springer: London, UK, 2001; pp. 3–16, ISBN 9781447103479.
49. Tomita, R.K.; Park, S.W.; Sotomayor, O.A. Analysis of activated sludge process using multivariate statistical tools—A PCA approach. *Chem. Eng. J.* **2002**, *90*, 283–290. [[CrossRef](#)]
50. Kellow, J.T. Using principal components analysis in program evaluation: Some practical considerations. *J. Multidiscip. Eval.* **2006**, *3*, 89–107.
51. Ledesma, R.D.; Valero-Mora, P.; Macbeth, G. The Scree test and the number of factors: A dynamic graphics approach. *Span. J. Psychol.* **2015**, *18*, E11. [[CrossRef](#)] [[PubMed](#)]