

Article

Application of Electronic Nose for Evaluation of Wastewater Treatment Process Effects at Full-Scale WWTP

Grzegorz Łagód ^{1,*}, Sylwia M. Duda ², Dariusz Majerek ³, Adriana Szutt ⁴ and Agnieszka Dołhańczuk-Śródka ⁴

¹ Department of Water Supply and Wastewater Disposal, Faculty of Environmental Engineering, Lublin University of Technology, 20-618 Lublin, Poland

² Department of Woman Health, Institute of Rural Health in Lublin, 20-090 Lublin, Poland; sylwia.m.duda@gmail.com

³ Department of Applied Mathematics, Faculty of Fundamentals of Technology, Lublin University of Technology, 20-618 Lublin, Poland; d.majerek@pollub.pl

⁴ Institute of Biotechnology, Faculty of Natural Sciences and Technology, University of Opole, 45-032 Opole, Poland; adriana.szutt@uni.opole.pl (A.S.); agna@uni.opole.pl (A.D.-Ś.)

* Correspondence: g.lagod@pollub.pl; Tel.: +48-81-538-4322

Received: 26 March 2019; Accepted: 24 April 2019; Published: 29 April 2019



Abstract: This paper presents the results of studies aiming at the assessment and classification of wastewater using an electronic nose. During the experiment, an attempt was made to classify the medium based on an analysis of signals from a gas sensor array, the intensity of which depended on the levels of volatile compounds in the headspace gas mixture above the wastewater table. The research involved samples collected from the mechanical and biological treatment devices of a full-scale wastewater treatment plant (WWTP), as well as wastewater analysis. The measurements were carried out with a metal-oxide-semiconductor (MOS) gas sensor array, when coupled with a computing unit (e.g., a computer with suitable software for the analysis of signals and their interpretation), it formed an e-nose—that is, an imitation of the mammalian olfactory sense. While conducting the research it was observed that the intensity of signals sent by sensors changed with drops in the level of wastewater pollution; thus, the samples could be classified in terms of their similarity and the analyzed gas-fingerprint could be related to the pollution level expressed by physical and biochemical indicators. Principal component analysis was employed for dimensionality reduction, and cluster analysis for grouping observation purposes. Supervised learning techniques confirmed that the obtained data were applicable for the classification of wastewater at different stages of the purification process.

Keywords: electronic nose; wastewater treatment processes; wastewater quality; odor nuisances; gas sensor array; multidimensional data analysis

1. Introduction

Objects connected with municipal service management, as well as sewerage systems, pumping stations, sewage sludge processing plants, segregation, recycling, and waste treatment play essential roles in environmental protection. However, they also constitute a potential hazard or nuisance. Nevertheless, if correct operation is ensured and preventive measures are taken if necessary (e.g., when a change in the quality of influent is observed or when hardly-biodegradable or harmful substances (crop protection chemicals, crude-oil derivatives, paints, chemicals from industrial production, disinfectants) appear), the odor nuisance and consequences of malfunctions can be mitigated.

Wastewater treatment plants are technological objects that reduce the pollution load found in wastewater prior to its discharge to a recipient. The processes carried out during wastewater treatment plant operation are connected with the emission of unpleasant odorants into the atmosphere. While discussing the odor nuisance of the considered objects, one should bear in mind that it is connected with the specificity of the medium (i.e., wastewater), as well as the necessity of maintaining certain conditions (e.g., anaerobic or anoxic) that are required to conduct the treatment process correctly [1–5].

Municipal wastewater mainly comprises the spent water discharged from houses, public institutions, industrial wastewater, as well as precipitation, seepage, and thaw water, which percolate to sewerage as a result of pipe leaks. The main groups of pollution found in the considered medium include easily degradable organic substances, other organic compounds, biogenic elements (i.e., nitrogen and phosphorus), microorganisms, heavy metals (i.e., mercury, lead, cadmium, chromium, copper, nickel), refractive and toxic substances, and other inorganic compounds.

However, not all pollutants found in water are determined during the wastewater examination, because there are too many of them and the classification of each of them would be impossible. In practice, groups of the most indicative compounds helpful for the assessment of a negative impact on the environment are determined. Organic compounds found in the considered medium are determined using the amount of spent oxygen (O_2) as BOD (biochemical oxygen demand) and COD (chemical oxygen demand); as well as the amount of generated carbon dioxide (CO_2), as TOC (total organic carbon) [6]. In exploitation practice, total suspension, total nitrogen, or total phosphorus are also determined in the wastewater.

The literature describes the dependencies between a high content of pollution indicators and the odor nuisance of wastewater [5,7–9]. In 2005, Onkal-Engin et al. [10] studied air samples collected from wastewater treatment plants using a multisensor gas array and classified them by means of artificial neural networks (ANNs). This team stated that the obtained results indicated the possibility of using the method for a general classification of samples, and that the e-nose could be used as a tool for BOD monitoring. Research conducted at the Lublin University of Technology in 2014 and 2016 during the operation of a laboratory sequencing batch reactor (SBR) with activated sludge indicated that the signal from an MOS (metal-oxide-semiconductor) sensor changed when the concentration of aromatic substances increased, while the implemented neural network determined the odor nuisance [5,11]. Experiments also indicated that an e-nose could distinguish the stages and states of laboratory-scale bioreactor operation [12] and the possibility of creating a neural network with the parameters enabling to estimate, for example, the values of COD, TSS (total suspended solids), and the concentration of nitrogen compounds in the considered medium [11].

The aforementioned papers confirm the validity of using gas multisensor arrays for monitoring the processes occurring in the course of wastewater treatment both under laboratory conditions as well as in the objects at the technical scale. The majority of papers are focused on the odor nuisance of particular devices, and some of them also related this to wastewater quality. Some papers also focused on detecting the presence of crude-oil derivatives, pesticides, and other chemicals in the wastewater flowing into the wastewater treatment plant (WWTP), which can be harmful for activated sludge and cause malfunction of biological treatment processes. Some of them present the possibility of classifying and evaluating the quality of wastewater treated or prepared in laboratory conditions. However, the analysis and assessment of mechanically treated wastewater subjected to highly efficient treatment in bioreactors with activated sludge at full-scale WWTPs where odor nuisance is virtually absent is underrepresented in the literature.

Identifying and—especially—determining the concentration of aromatic compounds is a difficult task. Nevertheless, studies of air quality are carried out by means of several methods. The odor intensity is frequently analyzed using dynamic olfactometry. This method also has certain disadvantages, because it is based on a subjective assessment of testers, and importantly, can only be employed for the samples considered harmless for humans. In turn, the concentrations of odorants are determined on the basis of chromatographic analyses. Unfortunately, this sensitive method of identifying the

composition cannot provide information on the odor nuisance. At present, gas sensor arrays are becoming increasingly popular [3,7–9,13–21]. Although they are not used for the identification of particular wastewater components and their concentrations (as opposed to chromatographs), they are employed as tools for preliminary sample assessment; additionally, they aid and supplement the traditional methods of determining the pollution indicators of a considered medium.

A system of sensors (i.e., a gas sensor array) coupled with a computing unit (e.g., a computer or microcontroller with an appropriate statistical algorithm) used for the analysis and interpretation of signals constitutes an e-nose (i.e., an imitation of the mammalian olfactory sense) [11,22–24]. Owing to the application of numerous sensors with varied sensitivity and selectivity, it is possible to create a unique combination of signals that is characteristic for a given gas sample—a so-called gas fingerprint [25]. MOS (metal-oxide-semiconductor) sensors are widely used in the considered arrays; they utilize semiconductor metal oxides, usually tin dioxide (SnO_2) with additives such as platinum, gold, and silver (used in order to improve the selectivity of the gas-sensitive layer) [26]. Chemisorption occurs on the surface of the sinter. Gas and semiconductor electrons form a bond, and a change in electrical conductivity occurs, which enables measurements to be taken [27].

Depending on the type of sensors utilized, they can distinguish between particular components or groups of components found in the considered mixture to a certain degree, which allows for attribution of the readouts with various parameters. Specific information on the investigated gas mixture is obtained using appropriately selected methods of analyzing the obtained complex measurement data. The data obtained from e-nose readouts is multidimensional, resulting in relatively complex interpretation algorithms. A number of statistical techniques can be employed for this purpose, such as principal component analysis (PCA), cluster analysis (CA), linear discriminant analysis (LDA), functional discriminant analysis (FDA), partial least squares discriminant analysis (PLS-DA), generalized linear models with regularized path (GLMNETs), or artificial neural networks (ANNs) [21,26,28–34]. Our literature review indicated that PCA, cluster analysis, and methods belonging to the group of supervised learning techniques are the most popular [21,35–37].

The PCA statistical method consists of grouping primary data by means of new low-dimensional space generated by linear combinations of primal variables. By reducing the dimensionality of the original space, they can be subsequently presented on graphs [38]. As a result of transformations, part of the original information is lost, in exchange for a simpler data structure [39].

There are many types of cluster analysis. The main division involves hierarchical and non-hierarchical cluster analysis. Cluster analysis or clustering involves grouping a set of objects in such a way that the objects in the same group (called a cluster) are more similar (in a certain way) to each other than to those in other groups (clusters). Cluster analysis itself is not a single specific algorithm, but rather the general task to be solved. It can be achieved by various algorithms that differ significantly in their understanding of what constitutes a cluster and how to efficiently find them [40]. The main idea of non-hierarchical cluster analysis is to find homogeneous clusters of data based on a distance matrix in multidimensional space [40,41]. Both methods (PCA and CA) are well-known techniques for the unsupervised classification and visualization of observations in multidimensional space [42]. Supervised learning techniques can confirm that the previously mentioned hidden structure (homogeneous clusters of data) is applicable for classification purposes. One of predictive supervised machine learning method is the *classification tree* technique [43]. Any system under test can be described by a set of classifications, holding both input and output parameters. The input parameters can also include environment states, pre-conditions, and other rather uncommon parameters [44]. Each classification can have any number of disjoint classes, describing the occurrence of the parameter. The selection of classes is usually conducted following the principle of equivalence partitioning for abstract test cases and boundary-value analysis for concrete test cases [45]. Together, all classifications form the classification tree. For semantic purposes, classifications can be grouped into compositions.

Bearing the aforementioned information in mind, the aim of this work is to present the possibility of classifying and thus evaluating the quality of wastewater treated at a full-scale municipal plant by

finding a hidden structure in multidimensional space generated on the basis of gas sensors' readout. Unsupervised methods of machine learning were used to find this structure. Principal component analysis was employed for dimensionality reduction, and cluster analysis for grouping observation purposes. Supervised learning techniques confirmed that the previously mentioned hidden structure (homogeneous clusters of data) is applicable for the purpose of classifying wastewater at different stages of the purification process.

2. Materials and Methods

The considered wastewater originated from the “Hajdów” Municipal Wastewater Treatment Plant in Lublin (South-Eastern Poland), with a daily volume of wastewater Q_d averaging $60,000 \text{ m}^3 \text{ d}^{-1}$. This mechanical-biological plant works in a continuous flow mode, and the chambers of the bioreactor utilize a modified Bardenpho [46]. The samples were taken directly from the technological devices in the mechanical and biological parts, at five points: primary settling tank, mixing chamber, bioreactor inlet, bioreactor outlet, and secondary settling tanks (treated wastewater). Each time, the wastewater samples were collected to bottles, filled, and taken directly to the laboratory for analysis. The elapsed time from sample collection to analysis was approximately 30–45 min. The bottles were stored in a travel refrigerator during transport time.

The measurements were performed with a self-constructed gas sensor array comprising 17 MOS sensors by Figaro company (Figaro USA Inc., Arlington Heights, IL, USA): TGS2600-B00 (general air contaminants), TGS2602-B00 (general air contaminants, high sensitivity to VOC), TGS2610-D00 (butane, LP gas; carbon filter), TGS2611-E00 (methane, natural gas; carbon filter), TGS2620-C00 (alcohol, solvent vapors), TGS4161 (carbon dioxide), TGS2444 (ammonia), TGS2442-B02 (carbon monoxide), TGS800 (various gas sensor—carbon monoxide, methane, isobutane, ethanol), TGS825-A00 (hydrogen sulfide), TGS813-A00 (combustible gases), TGS821 (hydrogen gas), TGS823-A00 (ethanol), TGS812 (combustible/flammable/methane gas), TGS830 (chlorofluorocarbons), TGS832-A00 (halocarbon), and TGS2106 (diesel exhaust gas). The utilized sensors were characterized by small size and power consumption [47–49]. The measurements were carried out in a 3–5 arrangement for each port, involving 3 min of flushing the sensors with synthetic air and 5 min of mixture analysis. The measurement set is schematically presented in Figure 1.

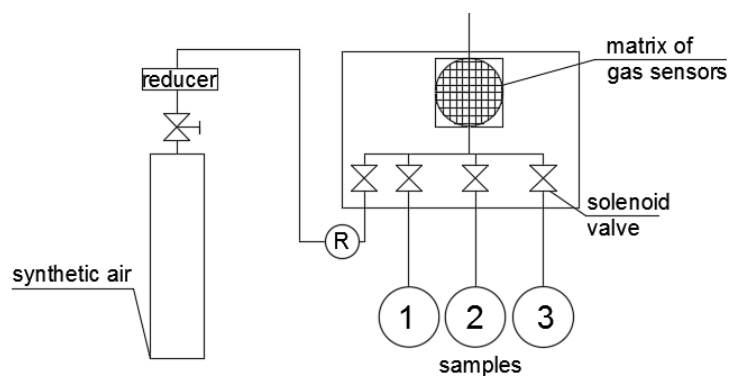


Figure 1. Scheme of the multi-sensor array together with whole laboratory set used in the research.

The experiment was conducted under laboratory conditions. Following intensive stirring until a homogenous composition was obtained, the initial samples of the considered medium (100 mL) were poured in equal amount to three identical conical glass flasks and subsequently subjected to analysis using a multisensor gas array (Figure 2). The procedure was performed in triplicate. The flasks were rinsed with distilled water several times between the consecutive measurements.



Figure 2. Analysis of wastewater samples by means of a gas sensor array.

Additionally, for reference to the array readouts, and in order to determine the pollution level of consecutive samples, one of the basic and most frequently determined parameters (i.e., total organic carbon (TOC)) was determined with the catalytic oxidation method, using a TOC 5050A Total Organic Carbon Analyzer (Shimadzu, Kyoto, Japan). Total suspended solids (TSS) was also determined by means of a HACH DR 3900 spectrophotometer (Hach Lange GmbH, Düsseldorf, Germany), using photometric method 8006 (program 630), according to protocol recommended by producing company.

The employed device recorded the sensor readouts with 5 s intervals; thus, the obtained datasets were appropriately prepared in order to average the presented results and reduce the amount of analyzed points and improve the legibility of graphs. The total size of the initial dataset used for statistical analysis was 185.

Further statistical analyses were carried out in the R-statistical programming environment [50] and packages that expand its functionality, including: factoextra [51], ggplot2 [52], knitr [53], plotly [54], mlr [55], and rpart [56].

As a statistical method for analysis of readouts from the matrix of gas sensors, the following methods were applied: principal components analysis [57] to explain data dispersion, and non-hierarchical cluster analysis—more specifically the *k*-means method described by MacQueen [40]—to find homogeneous clusters of data based on a distance matrix in multidimensional space. As one of the supervised learning techniques, classification tree [43] was used to confirm that the hidden structure of data from the gas sensors matrix was applicable for wastewater classification purposes.

3. Results and Discussion

The results in Table 1 and Figure 3 show that the first five or six principal components were appropriate for the explanation of data dispersion [57]. Another argument for choosing the first six principal components is the Kaiser criterion, which states that only the principal components which have eigenvalues greater than 1 should be chosen [58]. The six principal components explain 78% of the variance; however, in terms of visualizing observations in low-dimensional 2D and 3D space, only 46% and 57% of variance are explained, respectively.

Table 1. Proportion of variance explained by principal component analysis (PCA).

Statistical Parameters	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10	PC 11	PC 12	PC 13	PC 14	PC 15	PC 16	PC 17
Standard deviation	2.19	1.73	1.37	1.22	1.07	1.00	0.96	0.91	0.85	0.76	0.54	0.46	0.32	0.17	0.14	0.08	0.03
Cumulative Proportion	0.28	0.46	0.57	0.66	0.72	0.78	0.84	0.89	0.93	0.96	0.98	0.99	1.00	1.00	1.00	1.00	1.00

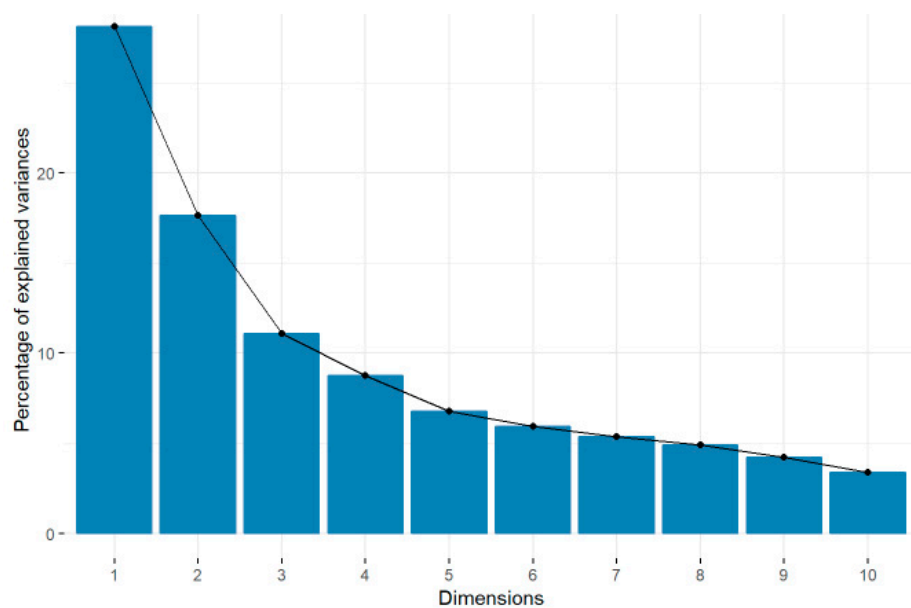


Figure 3. Scree plot of PCA.

Figure 4 presents the final result of the statistical analysis of obtained data, carried out with the PCA method. It aimed at reducing the data dimensionality from the initial 17 variables to only 2 new variables, in order to facilitate the readout and present the dependencies hidden in the data excess. Part of the information is lost during the PCA statistical analysis, but it is possible to reverse the transformation. Figure 4 shows groups of observations of particular stages of the wastewater treatment process. Groups of data from opposite stages were distinct, which means that based on the information from electronic nose sensors, one can group them into relatively homogenous clusters.

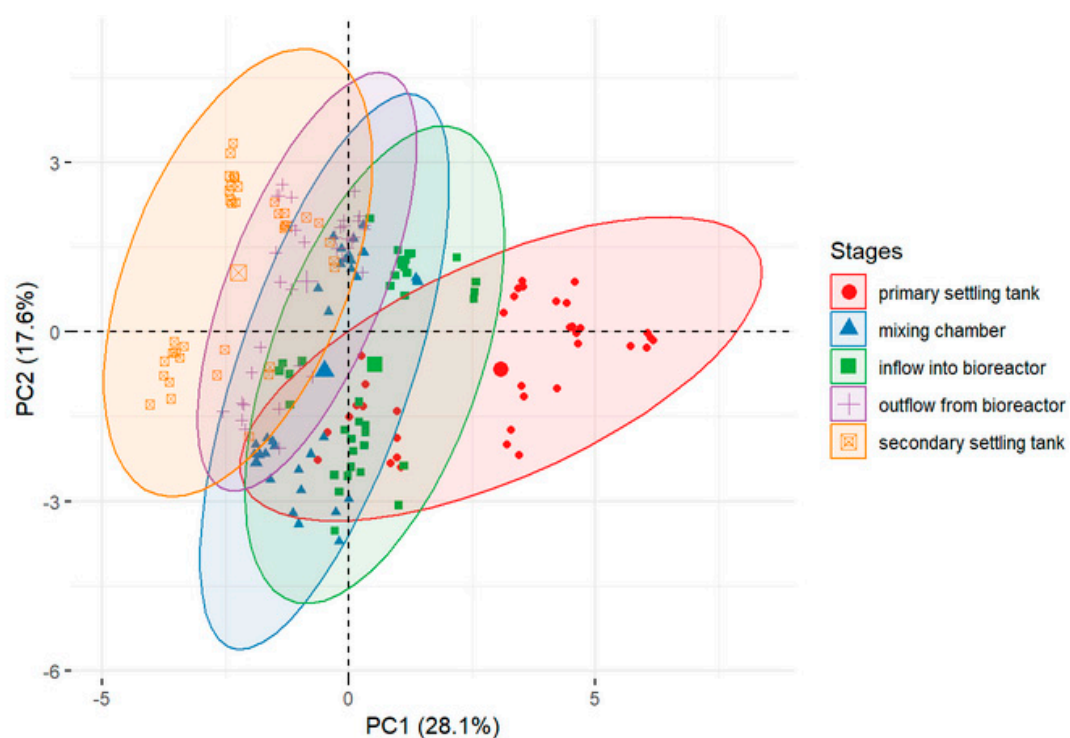


Figure 4. Two-dimensional PCA mapping of data. Colors and ellipses correspond to the stage of the wastewater treatment process.

The image of changes taking place in the wastewater treatment process was clearer in three-dimensional PCA space. Figure 5 shows a 3D representation of the data. The observations from the primary settling tank form a separate group of points, and the data from the secondary settling tank are on the opposite side.

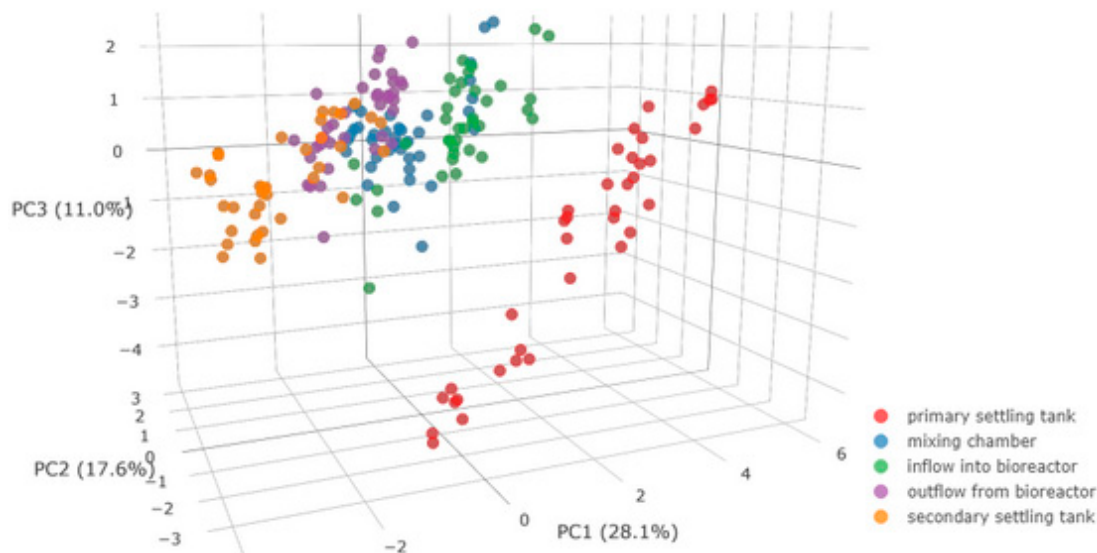


Figure 5. Three dimensional PCA mapping of data. Colors correspond to the wastewater treatment process stages.

Non-hierarchical cluster analysis (i.e., the k -means method) was employed. The main idea behind applying this technique was to find homogeneous clusters of data based on a distance matrix in multidimensional space. Since there are five different stages in the wastewater treatment process, we chose $k = 5$. There was no need to standardize variables, because they were characterized by almost the same variance. Both figures above led us to the conclusion that non-hierarchical cluster analysis operating in 17-dimensional space could group observations according to treatment stages.

Figure 6 illustrates the result of clustering in two-dimensional PC space. There were some misclassifications, but the observations from the primary and secondary settling tanks were in different clusters. First, they were in clusters 1 and 2, whereas later they were in 3 and 4. Therefore, cluster analysis confirmed that there is the potential to discriminate the stages of wastewater treatment process by means of statistically modeling data from an electronic nose.

All unsupervised learning methods showed some structure correlated with the stages of the wastewater treatment process. Supervised learning techniques could confirm that this hidden structure is applicable for classification purposes. As shown below, even a rather simple predictive method [43] was characterized by a very good accuracy.

In order to build and assess the performance of the Classification and Regression Tree model (CART model) [59], the data were split randomly into training and testing samples in a 2:1 proportion. Tuning of the complexity parameter of the model was executed via 5-fold cross-validation. A common method for describing the performance of a classification model is the confusion matrix. This is a simple cross-tabulation of the observed and predicted classes for the data. Diagonal cells denote the cases where the classes were correctly predicted, while the off-diagonals illustrate the number of errors for each possible case. The simplest metric is the overall accuracy rate, which reflects the agreement between the observed and predicted classes.

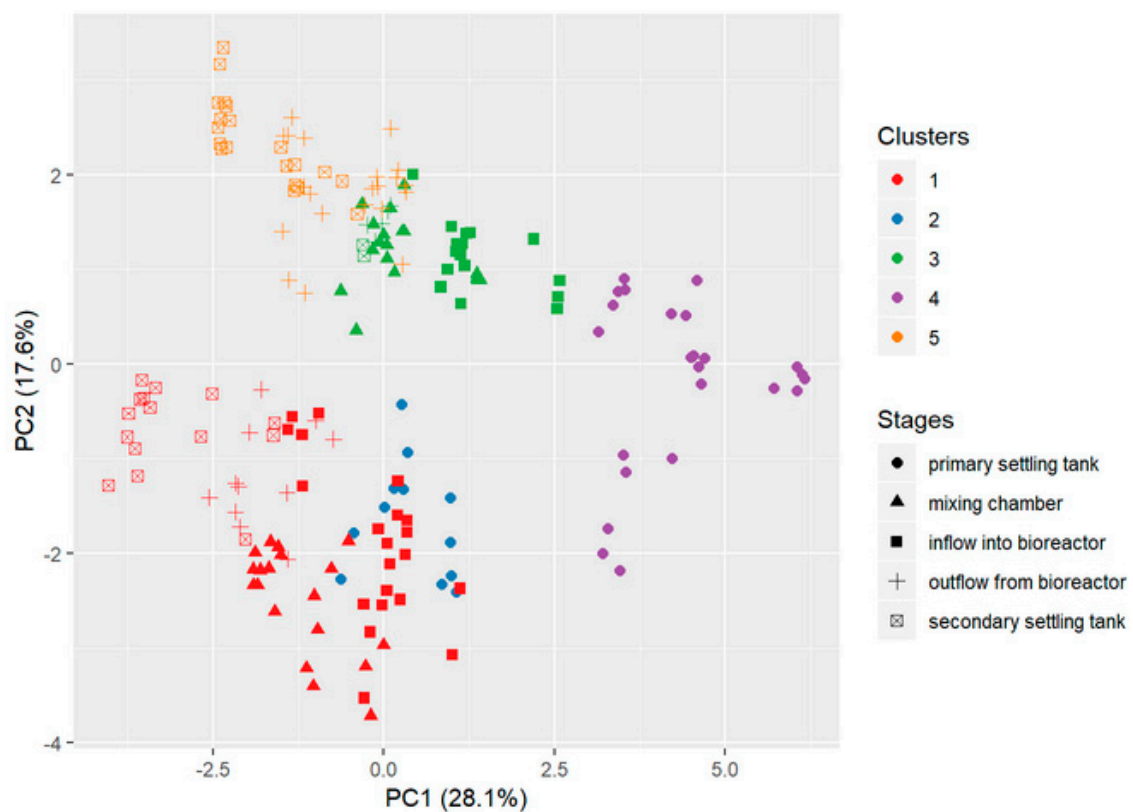


Figure 6. Result of the *k*-means clustering algorithm. Colors denote clusters and shapes correspond to different stages of the wastewater treatment process.

Very good accuracy for training and testing data was obtained by decision tree (i.e., 98% and 97%, respectively). Only two observations of outflow from the bioreactor were mistakenly classified to the secondary settling tank in the test sample (see Table 2), which is understandable because these two stages are adjacent to each other in the wastewater treatment process.

Table 2. Confusion matrix for CART model on the test sample.

Stages	Primary Settling Tank	Mixing Chamber	Inflow into Bioreactor	Outflow from Bioreactor	Secondary Settling Tank	Error
Primary settling tank	12	0	0	0	0	0
Mixing chamber	0	11	0	0	0	0
Inflow into bioreactor	0	0	12	0	0	0
Outflow from bioreactor	0	0	0	11	2	2
Secondary settling tank	0	0	0	0	14	0
Error	0	0	0	0	2	2

The model classified the observations with very good precision as visible in the Figure 7. The purity of leaves (terminal nodes) in the decision tree was very high. Thus, it was possible to obtain a rather simple predictive model with sufficient accuracy.

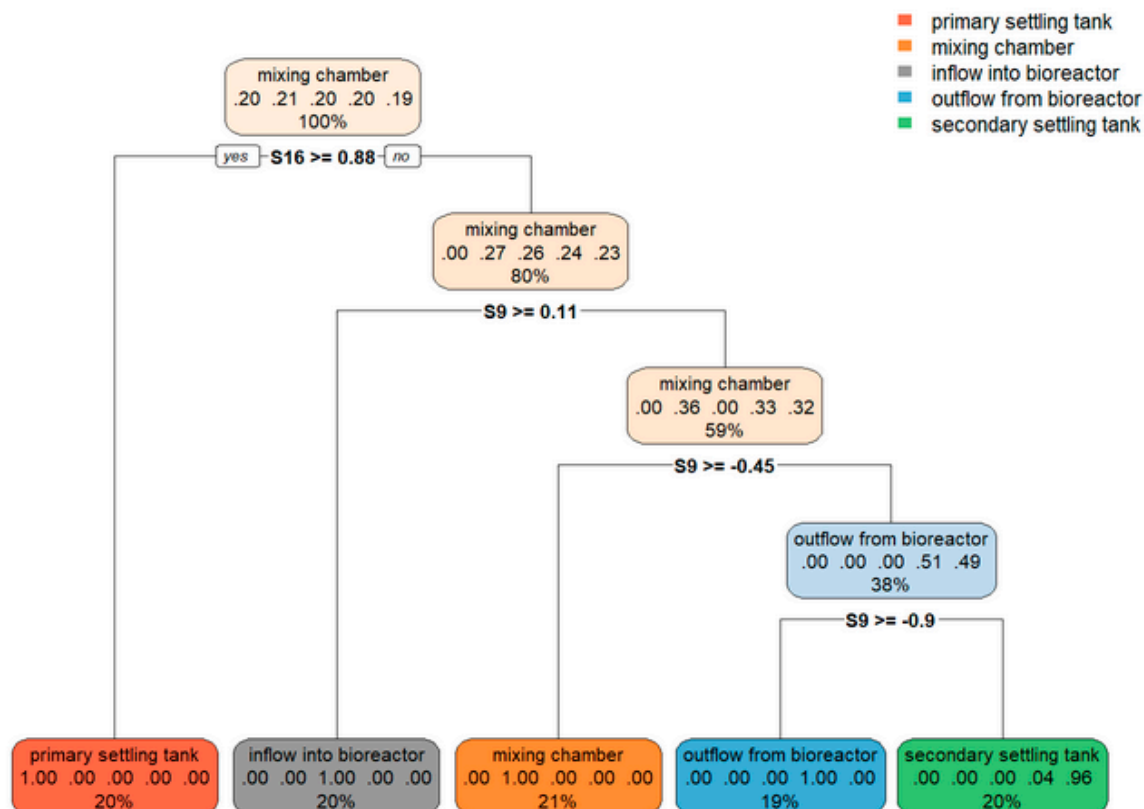


Figure 7. Decision tree for stages of the wastewater treatment process. Numbers in nodes are proportions of observations from particular classes. Percent in each node describes the proportion of all data located in nodes.

All the described results of e-nose readouts can be presented on the background of wastewater quality at the different stages of the treatment process. The treated wastewater was characterized by a varying level of pollution in particular stages of the process. As mentioned before, treatment stage is also connected with the intensity of odor emission [1,3–5]. Figure 4 shows that points (sensor readout data) were grouped into sets corresponding to consecutive devices of the technological line of the wastewater treatment plant and arranged with a certain regularity—from the highest pollution level (on the right-hand side of the graph) to the lowest level (left-hand side). The ellipses overlapped in a certain range, which probably stems from certain similarities of the investigated mixtures, especially apparent in the case of the sample from a mixing chamber and the one collected at the reactor inlet. In turn, the wastewater samples from the primary and secondary settling tanks (treated wastewater) could be distinguished unequivocally. Similar classification of unpurified and purified wastewater is also visible in Figures 5–7. This relation can be confirmed by the bar graph in Figure 8, presenting the results of total organic carbon (TOC) determination in samples.

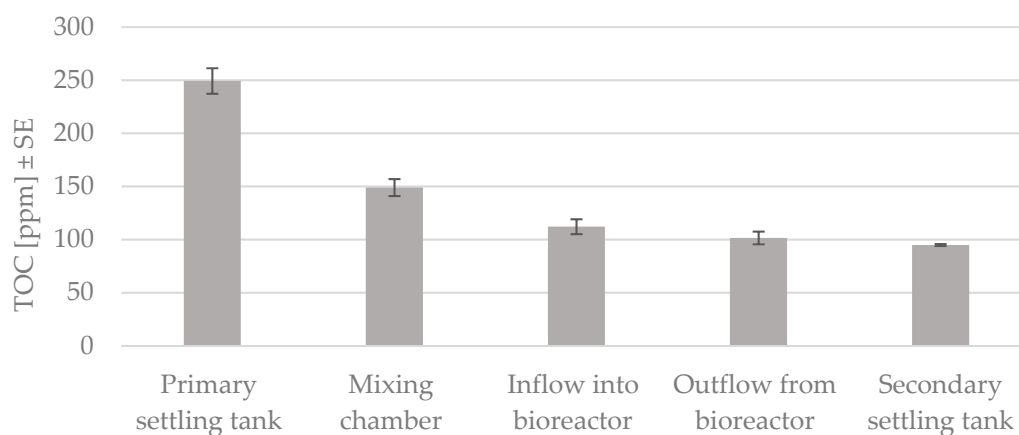


Figure 8. Graph of changes in total organic carbon (TOC) values in the wastewater subjected to consecutive treatment stages.

The only exception to the aforementioned observation stating that the results for samples arranged in line with the level of pollution was the arrangement of the samples described as the bioreactor outlet, beyond the ellipsis describing the secondary settling tank. Some of the points describing a measurement drift of the penultimate sample were located on the edge of the left side of the graph. The treated wastewater should be characterized by a much lower odor nuisance. The explanation for this phenomenon can be found in Figure 9, presenting a graph showing a change in the level of suspended solids. The total TSS values for the samples correspond to the tendency observed in Figure 4.

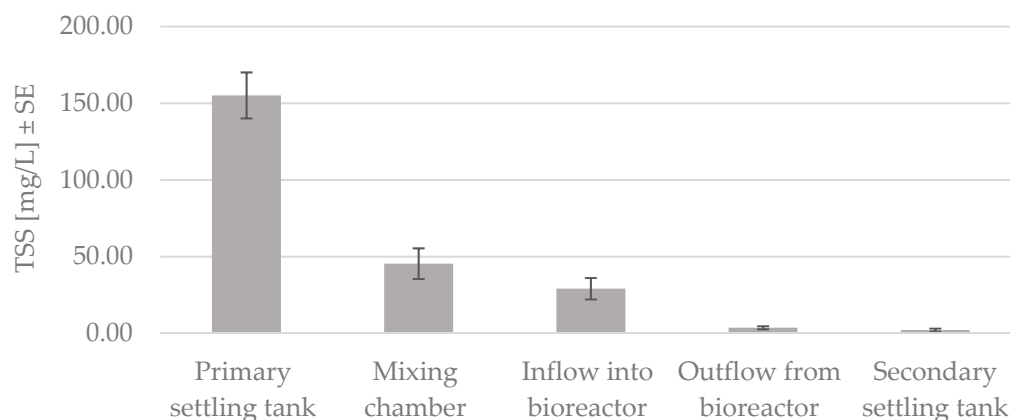


Figure 9. Graph of changes in total suspended solids (TSS) values in the wastewater subjected to consecutive treatment stages.

The observations from primary settling tank and secondary settling tank depicted in Figure 5 were separable, and certain directions of changes which took place in the wastewater treatment process observed in PCA map could be defined. Based on the PCA and the distribution of ellipses on the graph, certain tendencies and correlations between the pollution indicators could be observed, which has already been mentioned in the literature [7,8,11,14–16,60]. Moreover, based on previous research conducted in a laboratory bioreactor [12], a shift in the characteristics of the considered samples towards the positive values on the x -axis was observed, along with an increasing wastewater pollution level [61]. A similar situation occurred in the case of the results discussed in this chapter.

In the course of the analysis of measurement data from the samples of water by means of an e-nose, it was observed that each of the employed statistical methods (Figures 4–7) enabled us to unequivocally distinguish the wastewater treated only in the mechanical part from the biologically treated wastewater. A similar analysis conducted using PCA was presented by Bourgeois et al. [15];

every disturbance in the wastewater quality (caused by, e.g., heavy rain or chemical pollution) was clearly identifiable in the PCA plot.

Other studies also indicated the possibility of employing PCA, for example, for assessing the operation of wastewater treatment plants with activated sludge at technical scale [62]. The possibility of using this method for differentiation between raw sewage and final effluent on the basis of the signal from conducting polymer sensor array was proved [6].

In addition, the literature indicates that it is possible to determine the odor nuisance of treatment devices on the basis of e-nose readouts [9,12,13,18–20]. Blanco-Rodríguez et al. [63] discussed the application of an e-nose in for detecting the odor intensity in a wastewater treatment plant and the analysis of odor in six selected object locations (i.e., BioInlet, BioOutlet, Flotation, Flare, Settling, and Sludge in their work). Generally, the possibility of using the aforementioned devices as an early warning against the odors emitted in the course of the wastewater treatment process and the possibility of utilizing them in situ for odor monitoring were emphasized. However, it was also observed that it is possible to distinguish between particular samples and conduct a basic assessment of a gas mixture, which may help in the preliminary estimation and comparison of compounds emitted from particular locations in a wastewater treatment plant. This may be helpful in detecting changes such as the occurrence of hazardous substances in wastewater, as well as emergencies such as failures in the treatment plant operation.

The possibility of employing a gas sensor array for monitoring the results of wastewater treatment in SBRs operating under laboratory conditions has also been described [11]. The option to utilize an e-nose for the identification of wastewater treatment stages and the occurrence of process failure on the basis of interpreting the changes in sensor resistance, reacting to the change in the air quality of bioreactor headspace using PCA, was confirmed. This analysis enabled the particular states and phases of a laboratory SBR operation to be distinguished and the results of readouts performed in the headspace of untreated unpurified and purified wastewater could be clearly differentiated [12].

Research has shown that the e-nose can be a suitable device for the classification of wastewater as well as odors to their respective location in a WWTP. The samples were collected at different locations: influent, settling tank, activated sludge, and final effluent. A clear classification was obtained with a correlation of 0.99631; 93.06% of the outputs were classified successfully with an error below 10% [10], however, no significant odor problems were recorded within the mentioned plant [64].

Wastewater classification has also been shown in other studies. Dewwetinck et al. [8] showed that processing the gas fingerprints with PCA allowed for the interpretation and differentiation of wastewater samples in terms of origin and quality, relative to their reference (i.e., deionized water). In other WWTPs, the samples collected from the inlet works, settling tank, and final effluent showed that a nonspecific sensor array could distinguish between the different types of sewage samples and from different treatment works [7,65]. The research conducted by Nake et al. [17] showed that conducting-polymer (CP) sensors appeared to be unsuitable for this application while MOS sensors were a better fit. The MOS sensors were able to discriminate between the different odors from the outdoor sludge/bark mixer, the outdoor deodorization tower, outdoor sludge dewatering, and the clarifier.

The fact that each of the applied statistical methods (Figures 4–7) enabled us to unequivocally distinguish the mechanically treated wastewater from the biologically treated wastewater can have a practical application in the future for the management of processes conducted at full-scale WWTPs. The presented methods could be applied in a preliminary assessment and classification of the wastewater treatment result as well as in the rapid on-line detection of a malfunction or treatment process failure in flow bioreactors with activated sludge. In the mentioned case, results can be interpreted as the more the clusters grouping the measurement results of the samples collected in the primary and secondary settling tanks differ from each other, the more efficient the treatment process. The opposite is true as well—the more similar the clusters reflecting features of wastewater before and after biological purification, the less efficient the treatment process. In turn, a sudden decrease in the difference between

the clusters corresponding to the readout of arrays analyzing the two devices (primary and secondary settling tanks) may indicate a failure in the treatment process. Rapid detection of disturbances in the treatment process and a decrease in the quality of treated wastewater would give time for WWTP operators to identify the causes of failure, take preventive actions, and avoid further deterioration which could result in exceeding the permissible level of the quality of treated wastewater discharged to the receiver.

The possibility of using a gas sensor array for the assessment and classification of the wastewater taken up in various treatment plant devices was indicated in the cited works. However, this paper also attempts to familiarize the reader with the mathematical methods for the analysis of multidimensional signals, starting with initial visualization in two- and three-dimensional space, through interpretation of the similarities characterizing the analyzed datasets, confirmation of the observed relations using cluster analysis, to their verification with supervised learning techniques.

4. Summary and Conclusions

The results of the conducted research confirm that the gas sensor arrays could analyze various gases, depending on the configuration of sensors. Quick, relatively inexpensive, and repeatable (provided that the sensors are flushed appropriately prior to use) operation constitute the advantages of the discussed devices. This method is suitable for so-called screening tests (i.e., indicating any deviations from the norm). The on-line measurements enable a constant monitoring of the wastewater table headspace, and thus acquisition of information on the conducted processes. Additionally, the operation of multisensor arrays and the results of conducted studies may constitute a basis for the creation of early warning systems and models of wastewater treatment processes.

The presented research results and cited references indicate that multisensor arrays—especially MOS-type sensors—may find application in objects characterized by high odor nuisance (i.e., wastewater treatment plants). This technology facilitates the classification of samples as well as their preliminary assessment.

Conducting a relatively simple PCA facilitates the interpretation of a set of measurement data, and provides relevant information on the tendencies as well as similarities and differences among the samples.

Cluster analysis which operates in 17-dimensional space could group observations according to their membership to stages of treatment, and confirmed the fact that there is potential to discriminate stages of wastewater treatment by means of statistically modeling data from an electronic nose.

The readings from electronic nose sensors could be applied to build split rules of a decision tree model classifying observations with very good precision. The purity of leaves (terminal nodes) in the decision tree was very high and allowed us to obtain a predictive model with sufficient accuracy describing the classification of samples into stages of wastewater purification at a full-scale treatment plant.

Author Contributions: G.Ł. developed the experimental methodology and the concept of the article; G.Ł. and S.M.D. analyzed the data and wrote the outline of the paper draft. S.M.D. and A.S. performed measurements and improved the paper draft. D.M. prepared statistical analysis of the data as well as visualization, and improved the paper draft. A.D.-S. analyzed the data and improved the paper draft. All authors provided substantive comments.

Funding: This work was partly financially supported by Ministry of Science and Higher Education in Poland, within the statutory research of particular scientific units.

Acknowledgments: Supervisors of WWTP are gratefully acknowledged for access of sampled wastewater.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Frechen, F.B. Odour emissions of wastewater treatment plants—Recent German experiences. *Water Sci. Technol.* **1994**, *30*, 35–46. [[CrossRef](#)]

2. Thomas, O.; Theraulaz, F.; Cerda, V.; Constant, D.; Quevauviller, P. Wastewater quality monitoring. *Trends Anal. Chem.* **1997**, *16*, 419–424. [\[CrossRef\]](#)
3. Gostelow, P.; Persons, S.A.; Stuetz, R.M. Odour measurements for sewage treatment works. *Water Res.* **2001**, *35*, 579–597. [\[CrossRef\]](#)
4. Zarra, T.; Reiser, M.; Naddeo, V.; Belgiorno, V.; Kranert, M. Odour emissions characterization from wastewater treatment plants by different measurement methods. *Chem. Eng. Trans.* **2014**, *40*, 37–42. [\[CrossRef\]](#)
5. Guz, Ł.; Łagód, G.; Jaromin-Gleń, K.; Guz, E.; Sobczuk, H. Assessment of batch bioreactor odour nuisance using an e-nose. *Desalin. Water Treat.* **2016**, *57*, 1327–1335. [\[CrossRef\]](#)
6. Bourgeois, W.; Burgess, J.E.; Stuetz, R.M. On-line monitoring of wastewater quality: A review. *J. Chem. Technol. Biot.* **2001**, *76*, 337–348. [\[CrossRef\]](#)
7. Stuetz, R.M.; Fenner, R.A.; Engin, G. Characterisation of wastewater using an electronic nose. *Water Res.* **1999**, *33*, 442–452. [\[CrossRef\]](#)
8. Dewettinck, T.; Van Hege, K.; Verstraete, W. The electronic nose as a rapid sensor for volatile compounds in treated domestic wastewater. *Water Res.* **2001**, *35*, 75–83. [\[CrossRef\]](#)
9. Gebicki, J.; Byliński, H.; Namieśnik, J. Measurement techniques for assessing the olfactory impact of municipal sewage treatment plants. *Environ. Monit. Assess.* **2016**, *188*, 32. [\[CrossRef\]](#)
10. Onkal-Engina, G.; Demir, I.; Engin, S.N. Determination of the relationship between sewage odour and BOD by neural networks. *Environ. Modell. Softw.* **2005**, *20*, 843–850. [\[CrossRef\]](#)
11. Guz, Ł.; Łagód, G.; Jaromin-Gleń, K.; Suchorab, H.; Sobczuk, H.; Bieganski, A. Application of Gas Sensor Arrays in Assessment of Wastewater Purification Effects. *Sensors* **2015**, *15*, 1. [\[CrossRef\]](#)
12. Łagód, G.; Guz, Ł.; Sabba, F.; Sobczuk, H. Detection of wastewater treatment process disturbances in bioreactors using the e-nose technology. *Ecol. Chem. Eng. S* **2018**, *25*, 405–418. [\[CrossRef\]](#)
13. Stuetz, R.M.; Fenner, R.A.; Engin, G. Assessment of odours from sewage treatment works by an electronic nose, H₂S analysis and olfactometry. *Water Res.* **1999**, *33*, 453–461. [\[CrossRef\]](#)
14. Bourgeois, W.; Stuetz, R.M. Use of a chemical sensor array for detecting pollutants in domestic wastewater. *Water Res.* **2002**, *36*, 4505–4512. [\[CrossRef\]](#)
15. Bourgeois, W.; Hogben, P.; Pike, A.; Stuetz, R.M. Development of a sensor array based measurement system for continuous monitoring of water and wastewater. *Sens. Actuators B-Chem.* **2003**, *88*, 312–319. [\[CrossRef\]](#)
16. Bourgeois, W.; Gardey, G.; Servieres, M.; Stuetz, R.M. A chemical sensor array based system for protecting wastewater treatment plants. *Sens. Actuators B-Chem.* **2003**, *91*, 109–116. [\[CrossRef\]](#)
17. Nake, A.; Dubreuil, B.; Raynaud, C.; Talou, T. Outdoor in situ monitoring of volatile emissions from wastewater treatment plants with two portable technologies of electronic noses. *Sens. Actuators B-Chem.* **2005**, *106*, 36–39. [\[CrossRef\]](#)
18. Littarru, P. Environmental odours assessment from waste treatment plants: Dynamic olfactometry in combination with sensorial analysers “electronic noses”. *Waste Manag.* **2007**, *27*, 302–309. [\[CrossRef\]](#)
19. Capelli, S.; Sironi, P.; Centola, R.; del Rosso, R.; Grande, M. Electronic noses for the continuous monitoring of odours from a wastewater treatment plant at specific receptors: Focus on training methods. *Sens. Actuators B-Chem.* **2008**, *131*, 53–62. [\[CrossRef\]](#)
20. Nicolas, J.; Cerisier, C.; Delva, J. Potential of a Network of Electronic Noses to Assess in Real Time the Odour Annoyance in the Environment of a Compost Facility. *Chem. Eng. Trans.* **2012**, *30*, 133–138. [\[CrossRef\]](#)
21. Szulczynski, B.; Arminski, K.; Namiesnik, J.; Gebicki, J. Determination of Odour Interactions in Gaseous Mixtures Using Electronic Nose Methods with Artificial Neural Networks. *Sensors* **2018**, *18*, 519. [\[CrossRef\]](#)
22. Persaud, K.C.; Dodd, G. Analysis of Discrimination Mechanisms in the Mammalian Olfactory System Using a Model Nose. *Nature* **1982**, *299*, 352–355. [\[CrossRef\]](#)
23. Craven, M.A.; Gardner, J.W.; Bartlett, P.N. Electronic noses—Development and future prospects. *Trends Anal. Chem.* **1996**, *15*, 486–493. [\[CrossRef\]](#)
24. Gardner, J.W.; Bartlett, P.N. Electronic Noses. Principles and Applications. *Meas. Sci. Technol.* **2000**, *11*, 1087.
25. Wilson, A.D.; Baietto, M. Applications and advances in electronic-nose technologies. *Sensors* **2009**, *9*, 5099–5148. [\[CrossRef\]](#)
26. Krivetskiy, V.; Malkov, I.; Garshev, A.; Mordvinova, N.; Lebedev, O.I.; Dolenko, S.; Efitov, A.; Grigoriev, T.; Rumyantseva, M.; Gaskov, A. Chemically modified nanocrystalline SnO₂-based materials for nitrogen-containing gases detection using gas sensor array. *J. Alloy. Compound.* **2017**, *691*, 514–523. [\[CrossRef\]](#)

27. Teterycz, H. *Grubowarstwowe Chemiczne Czujniki Gazów na Bazie Dwutlenku Cyny*; Oficyna Wydawnicza Politechniki Wrocławskiej: Wrocław, Poland, 2005.
28. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*, 1st ed.; Wiley: New York, NY, USA, 1990.
29. Kohonen, T. The self-organizing map. *Neurocomputing* **1998**, *21*, 1–6. [[CrossRef](#)]
30. Fu, J.; Li, G.; Qin, Y.; Freeman, W.J. A pattern recognition method for electronic noses based on an olfactory neural network. *Sens. Actuators B-Chem.* **2007**, *125*, 489–497. [[CrossRef](#)]
31. Friedman, J.H.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [[CrossRef](#)]
32. Smolarz, A.; Kotyra, A.; Wójcik, W.; Ballester, J. Advanced diagnostics of industrial pulverized coal burner using optical methods and artificial intelligence. *Exp. Therm. Fluid Sci.* **2012**, *43*, 82–89. [[CrossRef](#)]
33. Khun, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013.
34. Brereton, R.G.; Lloyd, G.R. Partial least squares discriminant analysis: Taking the magic away. *J. Chemom.* **2014**, *28*, 213–225. [[CrossRef](#)]
35. Gardner, J.W.; Bartlett, P.N. A brief history of electronic noses. *Sens. Actuators B-Chem.* **1994**, *18*, 210–212. [[CrossRef](#)]
36. Llobet, E.; Brezmes, J.; Vilanova, X.; Sueiras, J.E.; Correig, X. Qualitative and quantitative analysis of volatile organic compounds using transient and steady-state response of a thick-film tin oxide gas sensor array. *Sens. Actuators B-Chem.* **1997**, *41*, 13–21. [[CrossRef](#)]
37. Haugen, J.E.; Kvaal, K. Electronic nose and artificial neural network. *Meat Sci.* **1998**, *49*, 273–286. [[CrossRef](#)]
38. Rajagopal, R.; Ranganathan, V. Evaluation of Effect of Unsupervised Dimensionality Reduction Techniques on Automated Arrhythmia Classification. *Biomed. Signal Process.* **2017**, *34*, 1–8. [[CrossRef](#)]
39. Everitt, B.S.; Landau, S.; Leese, M.; Stahl, D. *Cluster Analysis*, Wiley Series in Probability and Statistics; 5th ed.; John Wiley & Sons: Chichester, UK, 2008.
40. MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 21 June–18 July 1965; University of California Press: Berkeley, CA, USA, 1967.
41. Kumari, M.; Tripathi, B.D. Source Apportionment of Wastewater Pollutants Using Multivariate Analyses. *B Environ. Contam. Tox.* **2014**, *93*, 19–24. [[CrossRef](#)]
42. Thomas, M.C.; Romagnoli, J. Extracting Knowledge from Historical Databases for Process Monitoring Using Feature Extraction and Data Clustering. *Comput. Aided Chem. Eng.* **2016**, *38*, 859–864. [[CrossRef](#)]
43. Bergman, L.E.; Wilson, J.M.; Small, M.J.; VanBriesen, J.M. Application of Classification Trees for Predicting Disinfection by-Product Formation Targets from Source Water Characteristics. *Environ. Eng. Sci.* **2016**, *33*, 455–470. [[CrossRef](#)]
44. Mette, A.; Hass, J. *Guide to Advanced Software Testing*; Artech House: Boston, FL, USA, 2008; pp. 179–186.
45. Henry, P. *The Testing Network an Integral Approach to Test Activities in Large Software Projects*; Springer: Berlin, Germany, 2008; p. 87.
46. Jaromin-Gleń, K.; Babko, R.; Łagód, G.; Sobczuk, H. Community composition and abundance of protozoa under different concentration of nitrogen compounds at “Hajdow” wastewater treatment plant. *Ecol. Chem. Eng. S* **2013**, *20*, 127–139. [[CrossRef](#)]
47. TGS—For the Detection of Air Contaminants. Figaro Series Datasheet. 2019. Available online: <http://www.figarosensor.com> (accessed on 25 March 2019).
48. Guz, Ł.; Sobczuk, H.; Wasag, H. Device for determination of odour chemical substances in air. *Przem. Chem.* **2009**, *88*, 446–449.
49. Guz, Ł.; Sobczuk, H.; Suchorab, Z. Odor measurement using portable device with semiconductor gas sensors array. *Przem. Chem.* **2010**, *89*, 378–381.
50. R Core Team. *R: A Language and Environment for Statistical Computing*; Foundation for Statistical Computing: Vienna, Austria, 2018; Available online: <https://www.R-project.org/> (accessed on 25 March 2019).
51. Kassambara, A.; Mundt, F. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. 2017. Available online: <https://CRAN.R-project.org/package=factoextra> (accessed on 25 March 2019).
52. Wickham, H.; Chang, W.; Henry, L.; Pedersen, T.L.; Takahashi, K.; Wilke, C.; Woo, K. Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. 2018. Available online: <https://CRAN.R-project.org/package=ggplot2> (accessed on 25 March 2019).

53. Xie, Y. Knitr: A General-Purpose Package for Dynamic Report Generation in R. 2018. Available online: <https://CRAN.R-project.org/package=knitr> (accessed on 25 March 2019).
54. Sievert, C.; Parmar, C.; Hocking, T.; Chamberlain, S.; Ram, K.; Corvellec, M.; Despouy, P. Plotly: Create Interactive Web Graphics via 'Plotly.js'. 2018. Available online: <https://CRAN.R-project.org/package=plotly> (accessed on 25 March 2019).
55. Bischl, B.; Lang, M.; Kotthoff, L.; Schiffner, J.; Richter, J.; Jones, Z.; Casalicchio, G.; Gallo, M.; Schratz, P. Mlr: Machine Learning in R. 2018. Available online: <https://CRAN.R-project.org/package=mlr> (accessed on 25 March 2019).
56. Therneau, T.; Atkinson, B. Rpart: Recursive Partitioning and Regression Trees. 2018. Available online: <https://CRAN.R-project.org/package=rpart> (accessed on 25 March 2019).
57. Cattell, R.B. The Scree Test for the Number of Factors. *Multivar. Behav. Res.* **1966**, *1*, 245–276. [CrossRef]
58. Kaiser, H.F. A Note on Guttman's Lower Bound for the Number of Common Factors. *Br. J. Stat. Psychol.* **1961**, *14*, 1–2. [CrossRef]
59. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Wadsworth Statistics/Probability Series; Wadsworth Advanced Books and Software: Belmont, CA, USA, 1984.
60. Abbatangelo, M.; Núñez-Carmona, E.; Sberveglieri, V.; Comini, E.; Sberveglieri, G. Array of MOX Nanowire Gas Sensors for Wastewater Management. In Proceedings of the Eurosensors 2018, Graz, Austria, 9–12 September 2018; Volume 2. [CrossRef]
61. Baldwin, E.A.; Bai, J.; Plotto, A.; Dea, S. Electronic noses and tongues: Applications for the food and pharmaceutical industries. *Sensors* **2011**, *11*, 4744–4766. [CrossRef]
62. Tharrault, Y.; Mourot, G.; Ragot, J. WWTP diagnosis based on robust principal component analysis. In Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Barcelona, Spain, 30 June–3 July 2009.
63. Blanco-Rodríguez, A.; Camara, V.F.; Campo, F.; Becherán, L.; Durán, A.; Vieira, V.D.; Melo, H.; Garcia-Ramirez, A.R. Development of an electronic nose to characterize odours emitted from different stages in a wastewater treatment plant. *Water Res.* **2018**, *134*, 92–100. [CrossRef]
64. Onkal-Engin, G.; Demir, I.; Engin, S.N. E-nose response classification of sewage odors by neural networks and fuzzy clustering. In Proceedings of the Advances in Natural Computation: First International Conference (ICNC 2005), Changsha, China, 27–29 August 2005; Proceedings, Part II. pp. 648–651. [CrossRef]
65. Bourgeois, W.; Stuetz, R.M. Measuring wastewater quality using a sensor array: Prospects for real-time monitoring. *Water Sci. Technol.* **2000**, *41*, 107–112. [CrossRef]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).