Instrumental Odour Monitoring System Application in Complex Refinery Plant: Comparison of Different Pattern Recognition Algorithms and Data Features

Mark K.Gino Galang, Giuseppina Oliva, Vincenzo Belgiorno, Vincenzo Naddeo, Tiziano Zarra

Sanitary Environmental Engineering Division (SEED), Department of Civil Engineering, Università degli Studi di Salerno, via Giovanni Paolo II, 132 - 84084 Fisciano (SA), Italy
mgalang@unisa.it

The paper presents and discusses the development and application of different odour monitoring models (OMMs) for the classification and quantification of odour emissions with Instrumental Odour Monitoring Systems (IOMSs). Feed-forward neural network and linear discriminant analysis were considered for the classification of different type of odours, while feed-forward neural network and partial least square were investigated for the odour quantification. The prediction accuracy of the models was examined by analyzing different data extracted from the sensors' response curve (at rise, intermediate and peak period). The application has been carried out in a complex petroleum refinery plant. A total of 44 potential odour sources were monitored and grouped into 7 different classes.

Results highlight that the feed-forward neural network prevails in terms of high prediction capability having an architecture with three layers (input, hidden and output) of respectively 14-8-7 for odour classification, and of 14-8-1 for odour quantification, at ≥0.982 R^2. Meanwhile, the most useful data were found using the peak period. The research contributes to the understanding of IOMS applications, providing data on refinery plant odour emissions and applicable mathematical models to ensure great data reliability. The study highlights the influence of the pattern recognition algorithms in the Odour Montoring Model (OMM) elaboration and suggests the utility of promoting the implementation of flexible and adaptable IOMS.

1. Introduction

Odour has been listed among atmospheric pollutants (Piccardo et al., 2022), due to the fact that they can be an indicator of unhealthy air quality for the people and a source of private or public nuisance resulting in concerning environmental issue (Full et al., 2020). Among all the industries, petroleum industry is one of the contributors of odour pollution through the emission of different gaseous substances in the form of aromatic compounds, specially BTEX (benzene, C_6H_6; toluene, C_7H_8; ethylbenzene, C_8H_{10}; and xylene, C_8H_{10}), and sulphur (SO_X) compounds. Moreover, the emission from these plants may generate off-site impacts, thus, requiring comprehensive management by reliable assessment (Spinazzè et al., 2022).

At present, there are three (3) main strategies employed for odour emissions assessment such as the use of instrumental techniques, design of dispersion models, and citizens active involvement (Full et al., 2020). Firstly, the use of instrumental techniques can be categorized as analytical, sensorial, and combined analytical-sensorial methods (Giuliani et al., 2012). Analytical techniques are usually detectors and gases analysers in the laboratory that provide information about specific gases concentration in the odorant mixture (Blanco-Rodríguez et al., 2018). This technique can reveal the responsible gaseous compounds in the odour emission; however, it is not possible to quantify odour as a single parameter. For sensorial methods, dynamic olfactometry (EN 13725) and field inspection (EN 16841) are the methods specifically developed for odour measurement, which have been standardized at the European level (Bokowa et al., 2021). For combined analytical-sensorial techniques, Instrumental Odour Monitoring Systems (IOMSs) are applied (Zarra et al., 2021a). IOMS can be coupled with

other instruments in the field and implemented using smart technologies (Zarra et al., 2019), thus making it suitable to deploy on-site, giving real-time measurements, but more studies are needed in order to improve its functionality. On the other hand, odour dispersion models are able to predict odour concentration values at ground level in the simulation space-time domain (Botta et al., 2020). Lastly, citizens active involvement in terms of mapping and managing odour pollution problems, represents a proactive strategy to address socio-economic conflicts within the impacted communities (Conti et al., 2020). This approach is not expensive but takes time and lacks scientific stability due to the psychological effects of the people engaged in monitoring the odours.

In this study, the development and comparison of different odour monitoring models (OMMs) for the application of Instrumental Odour Monitoring System in a petroleum refinery plant is discussed. OMM can be parametric or non-parametric depending on the behaviour of the dataset (Zarra et al., 2021b), while data reduction is coupled in the data processing to support the OMM's effectiveness. In this way, robust information from the sensor response can be achieved with less redundancy. The research aims to contribute to the understanding of the methods to improve the accuracy of odour classification monitoring model (OCMM) and odour quantification monitoring model (OQMM) using the IOMS, in complex odorous plants characterized by a significant number of potential sources of odour emissions.

The investigations have been conducted in five phases: i) collecting samples at a real refinery plant and odour classes definition; ii) IOMS data acquisition and dynamic olfactometry analysis; iii) IOMS, extraction of the signals from the complete acquired response curve from the measurement sensors, in correspondence with the rise (first 1 minute), intermediate (mid 1 minute) and peak (last 1 minute) periods, and (b) IOMS, processing of the extracted data for the development of the odour monitoring models (OMMs) for OCMM and OQMM application; v) comparison studies.

2. Material and methods

2.1 Odour sampling

Research studies were carried out by collecting real samples from a big refinery plant. A total of 135 gaseous samples from different points, were carried out over a period of one year. The collected samples were clustered into seven classes (S1 – S7) (Figure 1) based on their position and the compounds emitted, in order to identify the odour classes to be investigated. The "lung technique" was employed for the sampling by withdrawing the inlet odorous air using a vacuum pump into a 7-L Nalophan® bag. Moreover, odourless ambient air samples were also sampled to build the baseline reference.
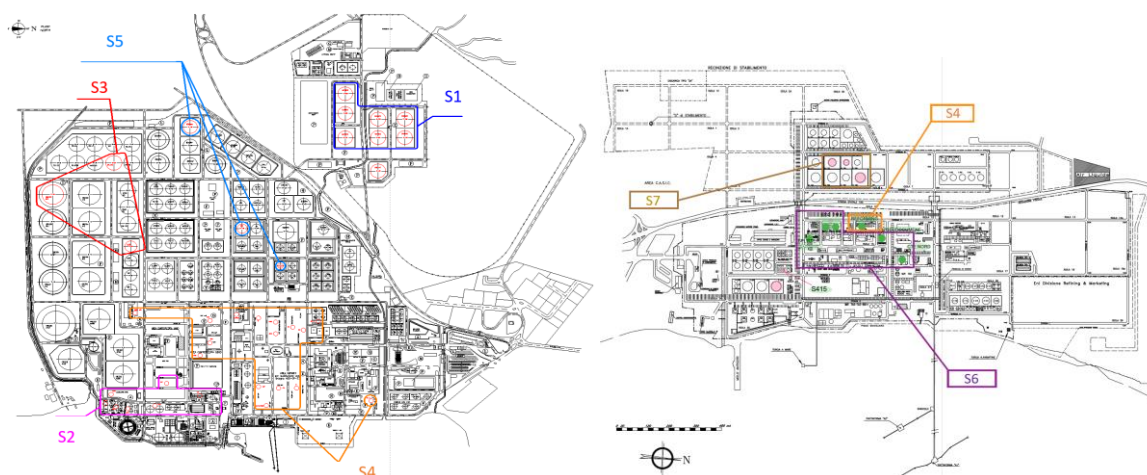

Figure 1: Investigated petroleum refinery plant, odour sources and classes

2.2 IOMS Technology and Dynamic Olfactometry Analysis

The IOMS presented in the study of Oliva et al. (2021) has been employed for the experiments. While TO8 olfactometer, with 4 panel and "yes/no" method was applied in accordance with EN 13725, to determine the odour concentration of the collected air samples, at the Olfactometric Laboratory of the Sanitary Environmental Engineering Division of the University of Salerno.

2.3 Feed-Forward Neural Network

Feed-Forward Neural Network (FFNN) was architected with three layers (input, hidden and output) under supervised learning by using back-propagating algorithm (e.g., Bayesian regularization) and tan-sigmoid function [$f(x) = (1+e^{-\Sigma f})^{-1}$] to obtain the ideal network (Keller et al., 2017). The different electrical resistances values ($x_1 \ldots x_n$) serve as inputs. 7 output classes ($\lambda_1 \ldots \lambda_7$) (Figure 2a) were used for the development of the odour classification monitoring model (OCMM) and one (δ, OU m^{-3}) (Figure 2b) as target output for the odour quantification monitoring model (OQMM). The ideal FFNN is based on the number of nodes in the hidden layer.
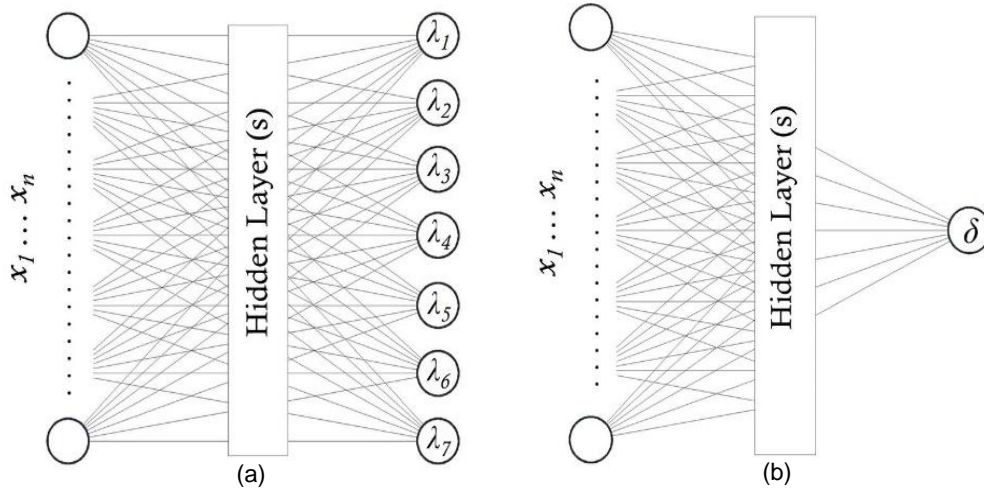

Figure 2: Target FFNN's typology for odour classification (a) and odour quantification (b) monitoring model

2.4 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) was used to reduce the number of features before classification. The consequent OCMM seeks a linear transformation that maximizes the separation of classes in a reduced dimensional space (Zarra et al., 2021a). During the calibration phase in the LDA method, the coefficients are calculated; in our case: $k, a, b\ldots \alpha$ of the different discriminant function equations (γ) (Equation 1-3) defined for each representative group (i.e., λ, β and n_{th}). To do this, the output values of the sensors are substituted for the variables ($x_1, x_2 \ldots x_n$)

$$\gamma_\lambda = k_1 + ax_1 + bx_2 + \ldots \alpha x_n$$

$$\gamma_\beta = k_2 + ax_1 + bx_2 + \ldots \alpha x_n \tag{1}$$

$$\gamma_{nth} = k_3 + ax_1 + bx_2 + \ldots \alpha x_n$$

(2)

2.5 Partial Least Square

(3)

Meanwhile the partial least square method was used for the OQMM. PLS finds the best relationship between dependent variable (Y) and independent variables (X_m) for prediction by considering the ideal number of components. Moreover, during PLS training, cumulative (CUM) R^2 and Q^2 are considered. Q^2 is another R^2 when the model applied to a test set, meaning adding more variables can increase the R^2, but might not make Q^2 increase (Wang et al., 2019). Considering these components, the PLS model can be expressed in the form of multiple linear regression model (Equation 4):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

where:

(4)

– $\beta_i, i = 14$ are the coefficient of the predictive variables (X_n).

2.6 Data Reduction

IOMS registered signals given in kΩ (Equation 5):

$$R_S = (R - R_O) / R_O$$

where:

(5)

– R = resistance value after the reaction with a gaseous compound;
– R_O = default resistance value of the sensor (baseline resistance).

The relationship between resistance and the gas concentration is inversely proportional (Equation 6):

$R (k\Omega) = A(C)^{-\alpha}$

where:

− R = electrical resistance supplied by the sensor;
− A = constant defined by the material;
− C = concentration of analyzed gas;
− α = slope (for instance the experimental quantity of the gas).

Since the acquisition time is 2 minutes, 3 extracted data are taken into account: rise (first 1 min), intermediate (mid 1 min) and peak (last 1 min).

2.7 Statistical analysis and comparison study

Microsoft Excel, Statistica StatSoft 10 and MATLAB R2021a were used as statistical software. While classification accuracy rates (CAR, %), coefficient of determination (R^2) and root-mean square errors (RMSE) served as criteria for the accuracy tests. Comparison study was based on which of the models provided the best criteria.

3. Results

3.1 Odour emissions characterization in terms of odour concentration

Figure 3 highlights the box-whisker diagram of all the determined odour concentrations for the overall monitored period. As shown, the minimum value was found to be 50 OU_E m^{-3}, while the maximum value of 1800 OU_E m^{-3}, covering a confined variability of values. The average measured odour concentration is 366 OU_E m^{-3} and the standard deviation is 270 OU_E m^{-3}. Furthermore, it has been found that most of the measured data belong to the 3-4th quartile, characterized by an odour concentration included between 500 – 800 OU_E m^{-3}.
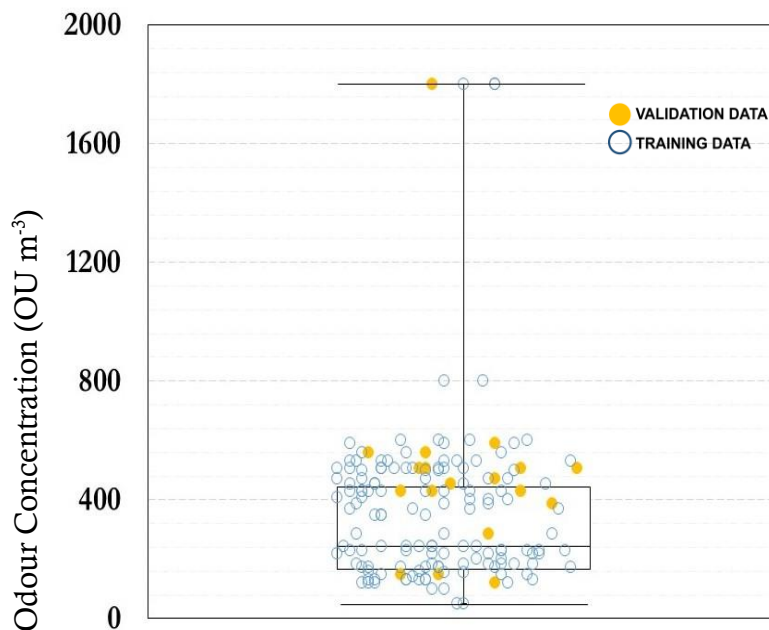

Figure 3: Odour concentration characterization and distribution for the collected samples

3.2 OCMMs analysis

During LDA calibration, a tolerance level of 0.001 and step wise analysis were employed. The obtained calibration (C), validation (V) values were respectively: i) for the whole curve: C,76.96%; V, 70.35%; Wilk's Lambda, Λ = 0.028084; ii) for the rise: C, 74.54%; V, 60.90%; Wilk's Lambda, Λ = 0.0304486; iii) for the intermediate: C, 77.60%; V, 64.87%; Wilk's Lambda, Λ = 0.0261409; iv) for the peak period: C, 78.54%; V, 73.62%; Wilk's Lambda, Λ = 0.0241204.

During the calibration phase, the Wilk's Lambda which signify the LDA's discriminatory power, was found to be highest for the data in the peak period at Λ = 0.0241204, and a 78.54% accuracy rate, while the rise period has the poorest due to the uncertainty in the dataset at this point (i.e., Λ = 0.0304486, % = 74.54%).

Meanwhile, after the validation stage, the peak period still shows the best accuracy rates of 73.62%, promoting the most stable points in this interval.

Meanwhile, in the FFNN environment, during the training, the higher the number of nodes assigned to the hidden layer, the more chances for the network to over-fit, while few numbers of nodes may lead to an under-fit network (Licen et al., 2018; Zarra et al., 2019). In the investigated analysis, training the network under Bayesian regularization, the optimum FFNN typology was found as 14-8-7. After extracting FFNN 14-8-7, the validation dataset was fed to the trained network to assess its generalizing capability.

The obtained training (T), validation (V) values were respectively: i) for the whole curve: T, 0.8784; V, 84.43%; ii) for the rise: T, 0.9067%; V, 99.14%; iii) for the intermediate: T, 0.9327; V, 99.29%; iv) for the peak period: T, 0.9608%; V, 99.71%. Consequently, the results show that the peak period contains the most useful data for FFNN as OCMM such as 0.9608 R^2 (training) and 99.71% (validation).

3.3 OQMMS analysis

PLS results show that by evaluating all the piecemeal signal values, all reduced data points were in good agreement with the 7 components, thus considering the ideal number of components for PLS OQMM. Moreover, during the calibration of PLS, the R^2 (CUM) (0.973) and Q^2 (CUM) (0.722) show only a negligible change, therefore it only highlights that data reduction presumably it does not have a significant impact for the calibration. However, the PLS models were verified to ensure their individual reliability, thus, revealing the ideal PLS OQMM entails the best data in the peak period on the basis of 0.9399 R^2 and 136.65 OU m^{-3} during validation.

In the case of FFNN, the experiment was carried out by manually assigning a specific number of nodes in the hidden layer from 5 – 22. After a complete run, it has been found that the FFNN typology is 14-8-1. A strong correlation (R ≥ 0.987), coefficient of determinations (R^2 ≥ 0.974) and root-mean squared errors (RMSE ≥ 22.25 OU m^{-3}) were obtained during FFNN training and validation, expressing the reliability of FFNN as OQMM. In fact, this phenomenon shows that FFNN can build a strong relationship based on the available data resources, but only varies on the training time (Herrero et al., 2016; Pomponi et al., 2021). The higher the number of nodes in the hidden layer, the more complex the network, thus longer training time is required.

3.4 Comparison Study

During the training/calibration phase, both LDA and PLS demonstrated simplicity in the execution and calibration. Their mechanisms are based on linear patterns and assumptions (e.g., parametric). In LDA, it has been found the linear features and combines them before classification, while PLS initially calculate the principal components and make a regression model with respect to the response variable. Meanwhile, FFNN is complex due to the numerous configurations considered, and must be done manually and carefully. It can establish a strong pattern independent of linearity, thus providing more accurate measurements, but the main drawback to researchers is visualizing the variable interactions.

Furthermore, during the validation of the QCMMs and OQMMs, the ANN revealed the highest classification rates (%) and coefficient of determinations (R^2), respectively. The accuracy of FFNN application is at least 80%. Finally, the comparison analysis highlights that the best combination should be the peak data and ANN as pattern-recognition algorithm for both OCMM and OQMM.

4. Conclusions

In the study, the data processing system of an advanced Instrumental Odour Monitoring System (IOMS), was studied and implemented with reference to odour emissions from a complex refinery plant, characterized of a high number of potential different odour sources.

Different processing approaches were used such as FFNN as non-parametric technique, and PLS and LDA as parametric techniques. Moreover, 7 classes were identified for the OCMM (odour classification), while odour concentration serves as target output for OQMM (odour quantification).

Results show that for the investigated study, FFNN emerged as the ideal OCMM and OQMM using the piecemeal data in the peak period. However, it should be emphasized that the development of FFNN is laborious and time-consuming and has limitations on exploiting variables' interaction. On the other hand, PLS and LDA, show fast execution, are less complex, and consist of features that reveal and independent and dependent variable interaction. But, PLS and LDA have a bounded prediction capability, being dependent to linearity. This issue could only be overcome when supplying more samples in the dataset, unlike ANN, it could adopt a strong pattern based on the available data resources. Furthermore, FFNN OCMM and OQMM typology was found to be 14-8-7 and 14-8-1, respectively. For PLS, 7 components are ideal, and 7 discriminant functions for LDA.

96

References

Blanco-Rodríguez A., Camara V.F., Campo F., Becherán L., Durán A., Vieira V.D., de Melo H., Garcia-Ramirez A.R., 2018. Development of an electronic nose to characterize odours emitted from different stages in a wastewater treatment plant. Water Research 134, 92–100. https://doi.org/10.1016/j.watres.2018.01.067

Bokowa A., Diaz C., Koziel J.A., McGinley M., Barclay J., Schauberger G., Guillot J.M., Sneath R., Capelli L., Zorich V., Izquierdo C., Bilsen I., Romain A.C., Del Carmen Cabeza M., Liu D., Both R., Van Belois H., Higuchi T., Wahe L., 2021. Summary and overview of the odour regulationsworldwide. Atmosphere (Basel). 12. https://doi.org/10.3390/atmos12020206

Botta S., Onofrio M., Spataro R., 2020. A review on the use of air dispersion models for odour assessment. Int. Journal Environmental Pollution 67, 1. https://doi.org/10.1504/ijep.2020.10030406

Conti C., Guarino M., Bacenetti J., 2020. Measurements techniques and models to assess odor annoyance: A review. Environmental International 134, 105261. https://doi.org/10.1016/j.envint.2019.105261

Full J., Delbrück L., Sauer A., Miehe R., 2020. Systematic Derivation of New Fields of Application for Innovative Bio-based Odour Sensors with Transfected Cells and Analysis of Economic Potentials 7029. https://doi.org/10.3390/iecb2020-07029

Giuliani S., Zarra T., Nicolas J., Naddeo V., Belgiorno V., Romain A.C., 2012. An alternative approach of the e-nose training phase in odour impact assessment. Chemical Engineering Transactions 30, 139–144. https://doi.org/10.3303/CET1230024

Herrero J.L., Lozano J., Santos J.P., Suárez J.I., 2016. On-line classification of pollutants in water using wireless portable electronic noses. Chemosphere 152, 107–116. https://doi.org/https://doi.org/10.1016/j.chemosphere.2016.02.106

Keller A., Gerkin R.C., Guan Y., Dhurandhar A., Turu G., Szalai B., Mainland J.D., Ihara Y., Yu C.W., Wolfinger R., Vens C., Schietgat L., De Grave K., Norel R., Stolovitzky G., Cecchi G.A., Vosshall L.B., Meyer P., Bhondekar A.P., Boutros P.C., Chang Y.C., Chen C.Y., Cherng B.W., Dimitriev A., Dolenc A., Falcao A.O., Golińska A.K., Hong M.Y., Hsieh P.H., Huang B.F., Hunyady L., Kaur R., Kazanov M.D., Kumar R., Lesiński W., Lin X., Matteson A., Oyang Y.J., Panwar B., Piliszek R., Polewko-Klim A., Raghava G.P.S., Rudnicki W.R., Saiz L., Sun R.X., Toplak M., Tung Y.A., Us P., Várnai P., Vilar J., Xie M., Yao D., Zitnik M., Zupan B., 2017. Predicting human olfactory perception from chemical features of odor molecules. Science. 355, 820–826. https://doi.org/10.1126/science.aal2014

Licen S., Barbieri G., Fabbris A., Briguglio S.C., Pillon A., Stel F., Barbieri P., 2018. Odor control map: Self organizing map built from electronic nose signals and integrated by different instrumental and sensorial data to obtain an assessment tool for real environmental scenarios. Sensors Actuators, B Chem. 263, 476–485. https://doi.org/10.1016/j.snb.2018.02.144

Oliva G., Zarra T., Pittoni G., Senatore V., Galang M.G., Castellani M., Belgiorno V., Naddeo V., 2021. Next-generation of instrumental odour monitoring system (IOMS) for the gaseous emissions control in complex industrial plants. Chemosphere 271, 129768. https://doi.org/10.1016/j.chemosphere.2021.129768

Piccardo M.T., Geretto M., Pulliero A., Izzotti A., 2022. Odour emissions: A public health concern for health risk perception. Environmental Research 204, 112121. https://doi.org/https://doi.org/10.1016/j.envres.2021.112121

Pomponi J., Scardapane S., Uncini A., 2021. Bayesian Neural Networks with Maximum Mean Discrepancy regularization. Neurocomputing 453, 428–437. https://doi.org/10.1016/j.neucom.2021.01.090

Spinazzè A., Polvara E., Cattaneo A., Invernizzi M., Cavallo D.M., Sironi S., 2022. Dynamic Olfactometry and Oil Refinery Odour Samples: Application of a New Method for Occupational Risk Assessment. Toxics 10. https://doi.org/10.3390/toxics10050202

Wang H., Gu J., Wang S., Saporta G., 2019. Spatial partial least squares autoregression: Algorithm and applications. Chemometrics and Intelligent Laboratory Systems. 184, 123–131. https://doi.org/10.1016/j.chemolab.2018.12.001

Zarra T., Galang M.G., Ballesteros F., Belgiorno V., Naddeo V., 2019. Environmental odour management by artificial neural network – A review. Environ. Int. 133, 105189. https://doi.org/10.1016/j.envint.2019.105189

Zarra T., Galang M.G.K., Ballesteros F.C., Belgiorno V., Naddeo V., 2021a. Instrumental odour monitoring system classification performance optimization by analysis of different pattern-recognition and feature extraction techniques. Sensors (Switzerland) 21, 1–16. https://doi.org/10.3390/s21010114

Zarra T., Galang M.G.K., Belgiorno V., Naddeo V., 2021b. Environmental odour quantification by ioms: Parametric vs. non-parametric prediction techniques. Chemosensors 9. https://doi.org/10.3390/chemosensors9070183