

# Integrating Citizen Science and Machine Learning Algorithms for the Recognition of Odour Classes nearby a Wastewater Treatment Plant

Federico Cangialosi<sup>a\*</sup>, Edoardo Bruno<sup>a</sup>, Antonio Fornaro<sup>b</sup>

<sup>a</sup>Tecnologia e Ambiente (T&A), Putignano (BA), Italia

<sup>b</sup>Labservice Analytica, Anzola Dell'Emilia (BO), Italia  
[federico.cangialosi@icloud.com](mailto:federico.cangialosi@icloud.com)

Odour nuisance is an increasingly topical problem, especially in newly developed urban areas. The use of machine learning algorithms for the classification and quantification of odour sources is becoming more and more widespread using instrumental odour monitoring systems (IOMS) for odour measurements. In this context of odour nuisances, the role of citizens can represent one of the fundamental factors for controlling the environment as several studies have already stressed: citizen science is now considered an additional tool for the smart management of environmental monitoring since it is able to carry out an in-depth analysis of the pollution problem. This paper presents a continuous monitoring study at the fence line of three urban wastewater treatment plants. The study was based on two distinct elements: firstly, continuous monitoring was used at the plant fence line using instrumental odour monitoring systems (IOMS) for odour measurements. Once a database with all IOMS data was obtained, odour classification and quantification algorithms were developed via machine learning techniques, such as Artificial Neural Networks (ANNs) and random forest, which were used to set-up a system capable of automatically recognizing both the odour class and concentration. Then, citizen science was used, by employing the data derived from an app available for the citizens: the app was set up in a way that citizens could enter the type and intensity of the smell they detected so each report would be recorded with GPS location, date, time and weather data allowing a comprehensive data mapping across space and time. We carried out a monitoring campaign over a period of five months, and then we compared the data obtained from the algorithms with the reports of the citizens, then studying the actual causes of the nuisances and verifying whether they were related to the monitored plant. At first, we carried out an analysis of the results provided by the IOMS, so that we could identify the most frequent odour classes and relative odour concentrations: it was decided to investigate different ranges of odour concentration to verify which sources were most influential in the most intense episodes of nuisance. Then, we correlated such information with the weather data and citizens' reports, to find out whether the reports were related to the plant. The description of the odours perceived by the population, alongside the identification of the appropriate wind cone influencing the receptors from the plant, allowed us to identify the events that could be attributed to the known sources. The results obtained from joint analysis of IOMS and citizens data were therefore useful for establishing to what extent the unpleasant odours perceived by the citizens came from the monitored plant.

## 1. Introduction

In the context of atmospheric pollution, odours are classified as a relevant component and an important indicator of the health of urban areas close to urban wastewater treatment plants (WWTP) (Oliva, et al., 2021)

In such a complex and socially sensitive context for the various reports of bad smells in the vicinity of urban wastewater treatment plants (WWTPs), the contribution of citizen science for the identification of odour emissions (Brattoli, et al., 2016, Lotesoriere, et al., 2021, Yen-Cha et al., 2017, Zheng et al., 2017), and the electronic nose or odour monitoring system instrumental (IOMS) (Karakaya et al., 2020) are powerful tools to help competent authorities and / or environmental protection agencies to define appropriate strategies in order to identify, measure and reduce the impact of odours on receptors.

As regards the IOMS, the developments in recent years have been very significant both in hardware and software terms, also thanks to the increasing use of algorithms that provide for suitable signal processing by extracting the most significant features (Zarra et al., 2021) and machine learning techniques (Men et al., 2018, Cangialosi et al., 2021, Yelim et al., 2022). The main goal of this study is to evaluate the potential of a monitoring system that combines the most recent instrumental techniques with the potential of citizen science to assess the odour impact connected to three wastewater treatment plants characterized by multiple emission sources.

## 2. Materials and methods

### 2.1 Plants description

The three urban wastewater treatment plants (WWTPs) considered in the study are located in the industrial area of Monopoli (Bari-Italy), the city of Polignano a Mare (Bari-Italy) and the industrial area of Putignano (Bari-Italy), respectively. Based on the data of various studies carried out for the olfactory characterization of urban wastewater treatment plants (Naddeo et al., 2016), it has been established that the most critical sections from the point of view of odour emissions are the pre-treatments, primary sedimentation, and sludge treatment. Based on these studies and the evidence gathered in the field, the mapping of the odours of the plant was carried out: the first phase involved a complete characterization of the emission sources, and subsequently a sampling program of the most critical sources was defined. The sampling program for the collection of samples for training and testing was designed to also consider environmental variations (temperature, relative humidity) and lasted several months.

### 2.2 IOMS training equipment and procedure

The IOMS (MSEM32® by Sensigent, Baldwin Park, CA, USA and Labservice Analytica, Anzola dell' Emilia, Italy) was positioned close to the fence line of Monopoli WWTP. After duplicate collection, each sample was fed to the IOMS on the same day and the replicate sample was analyzed using dynamic olfactometry (DO) at the T&A Laboratory within 24 hours, using the LEO dynamic olfactometer (ARCO Solutions srl, Trieste, Italy) for the measurement of Odour Concentrations (Cod), expressed as a European olfactory unit (uoE/m<sup>3</sup>). A total of 51 samples were collected, with odour concentrations ranging from 20 to 2435 uoE/m<sup>3</sup>. After a preliminary characterization, the odour classes selected were Class 1 (pretreatments), Class 2 (sludge conditioning), Class 3 (biogas) and Class 0 (unknown source). The dataset used for training the machine learning (ML) algorithms was obtained from the signals acquired by the IOMS which has an array of 32 sensors.

### 2.3 Data pretreatment, algorithms for the classification and quantification of odours and APP description

The response curves, which represent the variation of the sensor signals, were analyzed to extract the characteristics of the signal. The classification of odours and the prediction of the odour unit was carried out using two machine learning algorithms. The first is the Random Forest, and the second is a multilayer neural network (Multi-Layer Perceptron - MLP). All the data collected by the IOMS, representing sensor responses, were then extracted for training. After data pre-treatment, several tests were performed to choose an appropriate subset of the input variables, using the Recursive Feature Elimination with Cross Validation (RFECV) algorithm (Demarchi et al., 2020), to obtain the set of the most significant sensors that they were subsequently used for the construction of both algorithms. The overall dataset with the selected features was then divided into a training set and test set with an 80:20 ratio, thus using 600 data for training and 150 for performance evaluation. For the classification process, a confusion matrix was calculated for both the neural network and the Random Forest and the accuracy, calculated both for each class and overall, and the Cohen's kappa coefficient was used as scoring parameters. For the regression, the absolute differences between the measured odour concentrations and those predicted were calculated, both by the MLP algorithm and by the RF.

As for citizen reports collection, the App "Signal App-Odori", developed by the Municipality of Monopoli for both Android and iOS operating systems, was employed for the Monopoli WWTP. Once the user has logged in, it is possible to report when an odor nuisance is perceived. The user can indicate his level of odor annoyance: weak, easily detectable or very intense, and the type of smell perceived and can also enter a brief description of the perceived odour, to help the classification of the smell. The citizens of Monopoli were informed of the app and the project through a press conference chaired by the mayor who illustrated the objectives and operating methods of the program. For the plants in Polignano a Mare and Putignano, a Telegram bot (Odor-bot by Labservice Analytica) with the same features of signal App was employed. Both the applications allowed the citizen to classify the odour nuisances, among the others, as "wastewater treatment" or "sludges". In this case, the population was informed of the project through social media and online presentations, as public events were prohibited during the COVID period.

### 3. Results and discussion

#### 3.1 On-site training phase and testing for classification and regression

The instrumental signals were processed through a *feature selection* procedure which identifies the most suitable variables to be used in subsequent classification and regression models. Once the models (MLP and RF) were selected, the classification accuracy rates for each class and the overall accuracy rate for the best models were calculated. The results for the training set showed an accuracy for each class of not less than 0.99 for MLP and equal to 1 for RF. After analysing the results of the classification with the data of the test-set, consisting of 150 samples, it was found that only three elements were not correctly classified and both the MLP and the RF scored 0.98 on global accuracy and 0.97 on Cohen's Kappa coefficient. The RMSE mean square deviation for MLP is equal to 130  $\text{uo}/\text{m}^3$ , while for RF the value is equal to 97  $\text{uo}/\text{m}^3$ . The results of the training and testing phase are discussed in detail elsewhere (Cangialosi et al., 2021).

#### 3.2 Joint analysis of class-concentration data

Once the IOMS training was completed, the data were collected on each site in the monitoring periods indicated in Table 1, in which the most representative statistical indices obtained from the univariate analysis of the concentration distribution are also shown. In Figure 1, the cumulative distribution of odour concentrations for all the WWTPs are shown.

Table 1: Univariate analysis of the 3 plants data.

	Monopoli WWTP	Polignano WWTP	Putignano WWTP
Monitoring period	10/02/2021- 11/05/2021	01/07/2021- 05/10/2021	01/10/2021-10/01/2022
Number of data	258,059	258,409	286,231
Median	109 $\text{uo}/\text{m}^3$	7 $\text{uo}/\text{m}^3$	35 $\text{uo}/\text{m}^3$
95° Percentile	382 $\text{uo}/\text{m}^3$	140 $\text{uo}/\text{m}^3$	224 $\text{uo}/\text{m}^3$

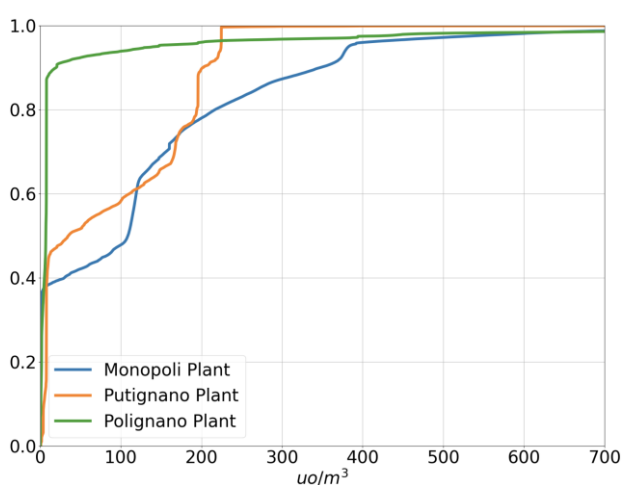


Figure 1: Cumulative distribution of odor concentrations.

Odour concentrations for the Polignano a Mare WWTP are very low, thus odour class analysis was not meaningful. As regards the other two plants, having the data on concentration and odour classes allowed us to jointly examine the data: it was decided to divide all the data from the IOMS into concentration classes with respect to the odour concentration. The lower bound of the first class was set to 100  $\text{uo}/\text{m}^3$  as below 100  $\text{uo}/\text{m}^3$  the results of the classification may not be relevant; the width of each class was chosen to be 100  $\text{uo}/\text{m}^3$ .

For the Monopoli plant, as can be seen from Figure 2(a), the higher the concentration value, the more relevant the contribution of class 0 (other or unknown) is: the lowest contribution of Class 0 (15%) is in the 100-200  $\text{uo}/\text{m}^3$  range and it reaches 69% in the range with values above 1000  $\text{uo}/\text{m}^3$ .

On the other hand, for the Putignano plant (Figure 2b), class 1 (pretreatment) is dominant (98%) throughout the concentration classes, except for the class with odour concentration above 400  $\text{uo}/\text{m}^3$ , where class 2 (sludge conditioning) is detected with a frequency of 20%.

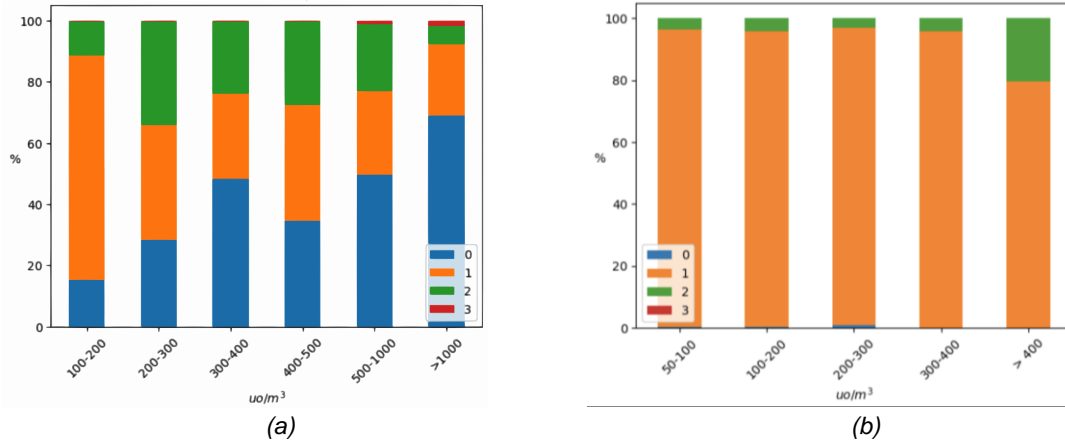


Figure 2: Percentages of odour classes detected among the different odour intensity bands in the Monopoli plant (a) and the Putignano plant (b)

Since the highest concentrations are likely to be more responsible for odor nuisance it is important to analyse the events reported by citizens and verify, using the IOMS data, how many of them may be related to an internal or external (and not known) source, namely Class 0, as discussed in the following paragraph.

### 3.3 Selection of citizen reports and joint analysis with IOMS data

In the monitoring period (February 2021- January 2022), a total of 298 citizens reports were collected, 268 of which from Monopoli and 30 from Putignano. No reports have been received from the municipality of Polignano a Mare, confirming the fact that the plant odour emissions were not of concern.

A two-step selection of citizen reports was adopted: firstly, only the reports for which the wind direction at the time of reporting was aligned to the plant-receptor direction were considered, plus the data with wind calms; then we selected the reports for which the descriptions of the type of odour were also available, selecting those relative to “wastewater treatment” or “sludge”. For Monopoli The citizens’ reports to be ascribed to the plants by wind direction and type of odour are 11 out 268 and 18 out of 30 for Putignano, the remaining reports were not considered because they did not match either the direction or the type of odor reported by the citizens and therefore the relative odor nuisance did not it could in no way have come from the plant. In the municipality of Monopoli there are other several possible sources (another WWTP next to the monitored plant, a waste storage plant and a power plant powered by biofuel). The selected reports are shown in Figure 3.

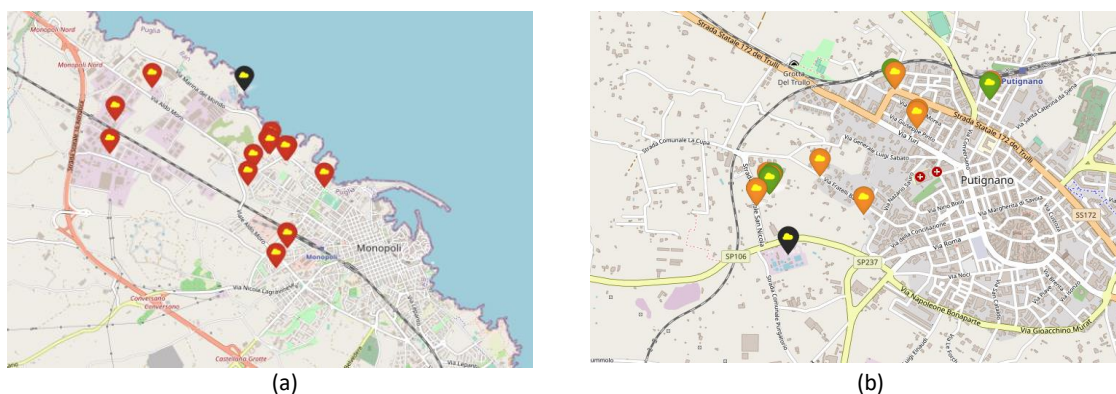


Figure 3: Analysis of the reports in the period of interest, selected by wind direction and type of odour in Monopoli (a) and Putignano (b). In black are indicated the WWTPs.

The selected reports were then correlated with IOMS data. Since the reports refer to a specific time, it was decided to consider, for each report, all the IOMS data within a time window of one hour, centered on the instant of the report, thus having a wider range to analyze, in order to take into account two aspects of human reporting: firstly, there might be delays in reporting with respect to the moment of perception; secondly, there might be reports at the time of initial perception of the odour which could subsequently increase.

Figure 4 shows the daily temporal distribution of the reports, compared with the daily temporal distribution of the events recorded by the IOMS with a concentration greater than  $500 \text{ uo/m}^3$  for the city of Monopoli, and greater than  $300 \text{ uo/m}^3$  for the city of Putignano, both normalized with respect to the maximum number of events. As can be seen, for both WWTP there are some periods of strong correlation between the number of reports and the IOMS data with high concentration.

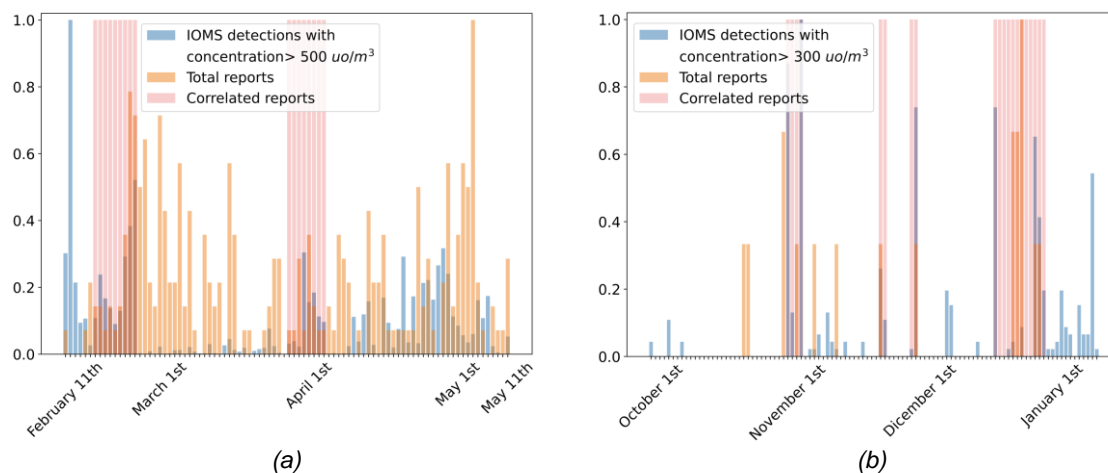


Figure 4 : Comparison between the temporal distribution of all the reports, the selected reports and IOMS data with high concentrations in Monopoli (a) and in Putignano (b)

For the Putignano plant, we wanted to verify if there was a significant difference in odour concentrations between the class of IOMS data aligned with the citizen reports and the other. As shown in Figure 5a, the percentage of data with low concentration ( $<100 \text{ uo/m}^3$ ) is 60% when no reports were recorded; on the other hand, the percentage of data with higher concentration ( $>100 \text{ uo/m}^3$ ) is 70% during the time windows of citizen reports, thus confirming the consistency between IOMS detection at the WWTP fence line and the nuisance reported by citizens. For the Monopoli WWTP, where several odour classes were detected (see Figure 2a), the IOMS data directly correlated with the reports were then analyzed with reference to the odour classes.

As shown in Figure 5b, the percentage of attribution to class 0, i.e. unknown, was found to be 40%, while 60% of the time in which there were reports from citizens is directly attributable to the two main plant odour sources, 27.24% pretreatments (class 1), 32.51% sludge conditioning (class 2).

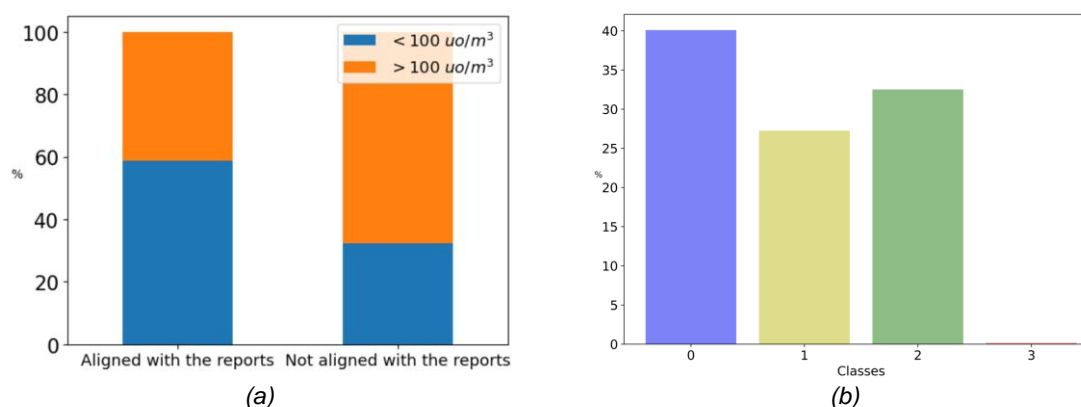


Figure 5: Correlation between odour concentrations and selected citizen reports for Putignano (a) and percentages of odour classes detected during selected reports for Monopoli (b)

Therefore, for the Monopoli plant, the data acquired by the IOMS system and the selected citizens reports, clearly highlighted how a significant part of them (40%) may derive from emission areas not attributable to specific sources within the WWTP, but from other close plants whose emissions were classified by citizens as "wastewater treatment" or "sludges".

#### 4. Conclusions

The present work describes the integration between tools of citizen science and the use of IOMS equipped with artificial intelligence algorithms for monitoring odour emissions from civil wastewater treatment plants.

It was questioned whether the integration between IOMS data and citizens' reports is able to ascertain whether odour nuisance is directly related to a specific WWTP or to other unknown odour sources and to identify the critical sources within a WWTP. In particular, the analysis of the field data, recorded after carefully training of IOMS both for odour concentrations and classes, made it possible to identify the classes of odours that are responsible for the highest concentration values during the various months of monitoring. The reports, an average of more than 20 per month for the 11 months of monitoring, were analyzed to consider the wind direction at the time of the report and the description of the type of odour, to highlight those related to wastewater treatment plants. For the Polignano a Mare WWTP it was not considered to proceed with the analysis of the classes as the values of odour concentrations were very low and citizen reports confirmed that odour emissions were not of concern. For the Putignano WWTP, the analysis of the IOMS data during the hours in which the reports were made, allowed us to quantify the contribution of the emission sources of the plant, thus identifying the pretreatments as the most relevant source for low-medium concentrations, whereas a not negligible contribution (20%) is given at high concentrations by sludge treatment. It was also verified that all the perceived nuisances, classified by citizens as "*wastewater treatment*" or "*sludges*", actually derived from the plant.

On the other hand, for the Monopoli WWTP, the joint analysis of odour concentrations, odour classes and citizen reports allowed to point out that 60% of the odour nuisances are directly related to the plant (27.24% pretreatments and 32.51% sludge conditioning), while 40% of them are not attributable to any source within the plant and may be related to other similar sources, such as a WWTP located nearby.

The combined use of the instrumental approach and data relating to citizens' reports via the App has proven to be useful and effective, especially in the presence of multiple odour emission sources.

#### References

- Brattoli, M.; Mazzone, A.; Giua, R.; Assennato, G.; de Gennaro, G., 2016, Automated Collection of Real-Time Alerts of Citizens as a Useful Tool to Continuously Monitor Malodorous Emissions. *Int. J. Env. Res. Pub. Health*, 13, 263, doi:10.3390/ijerph13030263.
- Cangialosi, F.; Bruno, E.; De Santis, G., 2021 Application of Machine Learning for Fenceline Monitoring of Odor Classes and Concentrations at a Wastewater Treatment Plant. *Sensors*, 21, 4716. <https://doi.org/10.3390/s21144716>
- Demarchi, L.; Kania, A.; Ciężkowski, W.; Piórkowski, H.; Oświecimska-Piasko, Z.; Chormański, J., 2020 Recursive Feature Elimination and Random Forest Classification of Natura 2000 Grasslands in Lowland River Valleys of Poland Based on Airborne Hyperspectral and LiDAR Data Fusion. *Remote. Sens.*, 12, 1842, doi:10.3390/rs12111842.
- Shepherd G. M., 2004. The human sense of smell: Are we better than we think? *PLoS Biol.* 2, 5, e146
- Karakaya, D.; Ulucan, O.; Turkan., 2020, M. Electronic Nose and Its Applications: A Survey. *Int. J. Aut. Comp.*, 17, 179–209, doi:10.1007/s11633-019-1212-9.
- Lotesoriere, B.; Giacomello, A.; Bax, C.; Capelli, L., 2021, The Italian Pilot Study of the D-NOSES Project: An Integrated Approach Involving Citizen Science and Olfactometry to Identify Odour Sources in the Area of Castellanza (VA). *Chem. Eng. Trans.* 85, 145–150.
- Men, H.; Fu, S.; Yang, J.; Cheng, M.; Shi, Y.; Liu, J., 2018, Comparison of SVM, RF and ELM on an Electronic Nose for the Intelligent Evaluation of Paraffin Samples. *Sensors*, 18, 285, doi:10.3390/s18010285.
- Naddeo, V.; Zarra, T.; Oliva, G.; Kubo, A.; Ukida, N.; Higuchi, T., 2016 Odour measurement in wastewater treatment plant by a new prototype of e.Nose: Correlation and comparison study with reference to both European and Japanese approaches. *Chem. Eng. Trans.* 54, 85–90.
- Oliva, G.; Zarra, T.; Massimo, R.; Senatore, V.; Buonerba, A.; Belgiorno, V.; Naddeo, V., 2021, Optimization of Classification Prediction Performances of an Instrumental Odour Monitoring System by Using Temperature Correction Approach. *Chemosensors*, 9, 147, doi: 10.3390/chemosensors9060147.
- Choi, Y.; Kim, K.; Kim, S.; Kim, D., 2022, Identification of odor emission sources in urban areas using machine learning-based classification models, *Atmospheric Environment: X*, Volume 13, 100156, ISSN 2590-1621, <https://doi.org/10.1016/j.aeaoa.2022.100156>.
- Hsu, Y.; Dille, P.; Cross, J.; Dias, B.; Sargent, R.; Nourbakhsh, I., 2017, Community-Empowered Air Quality Monitoring System. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1607–1619. DOI: <https://doi.org/10.1145/3025453.3025853>
- Zarra, T.; Galang, M.G.K.; Ballesteros, F.C. Jr.; Belgiorno, V.; Naddeo, V., 2021, Instrumental Odour Monitoring System Classification Performance Optimization by Analysis of Different Pattern-Recognition and Feature Extraction Techniques. *Sensors*, 21, 114, dx.doi:10.3390/s21010114.
- Zheng, H.; Hong, Y.; Long, D.; and Jing, H., 2017, Monitoring surface water quality using social media in the context of citizen science, *Hydrol. Earth Syst. Sci.*, 21, 949–961, <https://doi.org/10.5194/hess-21-949-2017>.