Machine Learning methods to estimate odour intensity

Author: Ivan Casanovas Rodríguez
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Santiago Marco Colás

Abstract: Odour is a human perception whose relationship with chemical composition is unknown. Contrary to the olfactometric measurement techniques, senso-instrumental methods provide real-time odour monitoring. The study presents a drone equipped with an electronic nose that generates dynamic sensor signals for the classification and quantification of odours in wastewater treatment plants. By calibrating predictive models with Machine Learning algorithms, odour/non-odour samples are classified with 93% accuracy, and odour concentration is predicted 95% limits of agreement within a factor of four, in comparison with dynamic olfactometry measurements.

I. INTRODUCTION

Odour pollution is currently a significant environmental and human health issue. Due to advancements in society, there is a growing number of odour-emitting sources, and wastewater treatment plants (WWTPs) are one of the primary contributors. These emissions have an impact on well-being and air quality in the nearby area, and can even lead to health and psychological problems.

Understanding odour as a human perception, its characterization has become a challenge. This difficulty arises from its subjectivity nature and the complexity of its chemical composition. In recent years, the main purpose has been to estimate odour intensity via standardized methods.

According to the European standard EN13725:2022, its quantification is determined in a laboratory by dynamic olfactometry, a sensorial technique that correlates odour concentration with the human sense of smell. This measurement, carried out by trained human panels, represents the dilution factor required for a sample to reach its odour detection threshold concentration (ou_E/m^3). Since the referenced method is a slow process and requires multiple sampling, real-time odour monitoring is not feasible. These limitations, along with the variability in results, highlight the need for alternative techniques.

Instrumental Odour Monitoring Systems (IOMS) [1], commonly known as electronic noses, are devices trained to classify and quantify odours by analyzing the dynamic electrical signals provided by poorly-selective chemical sensors. Their aim is to establish the relationship between sensors response and odour concentration, which is unknown *a priori*. Using algorithms that learn from samples, we are able to solve the problem and experimentally measure what is actually a human perception.

There is also a necessity to study complex environments such as WWTPS, where there are numerous odour-emitting sources. A novel approach to odour monitoring in these areas is the utilization of drones equipped with e-noses, because they allow for quickly measuring different positions. In this study [2], a similar IOMS was employed to monitor odour in the WWTP of Molina de Segura (Murcia, Spain). The drone hovers above the WWTP following a predefined navigation path and, at each point of interest, the chemical sensors continuously measure for a specific period of time. The drone is also equipped with an odour sampling device to measure the odour concentration using dynamic olfactometry. Moreover, odourless samples (blanks) are characterized through sensor measurements in the surroundings. By covering the entire area, an odour concentration map can be developed to study odour propagation and the interaction of emissions from different sources.

The main goal of this TFG is to estimate the emitting odours of a WWTP by processing the sensor signals using Machine Learning methods. As the odour concentrations of the samples are known, two types of problems are addressed. On one hand, the classification aims to distinguish between odour and non-odour samples. On the other hand, the regression problem involves predicting the odour concentrations for each sample. To approach these issues, the data needs to be initially preprocessed and then analyzed, improving the accuracy of the IOMS.

II. EXPERIMENTAL METHODS

The drone used in Molina de Segura WWTP is the DJI Matrice 600 [2]. It is a rotary-wing small drone with a wide range of applications in aerial monitoring. The structure incorporates an air sampler and a customized IOMS. As pollution gases at the WWTPs environment are composed of a mixture of gases, the e-nose requires a multi-sensor system. It includes low-cost chemical sensors with a fast response time: there are 4 electrochemical sensors (H_2S, NH_3, SO_2, CO), 16 metal oxid sensors (MOX) and one infrared (CO_2). The odour sampling device and IOMS are both connected to vacuum pumps that draw in the samples through a tube. By connecting the two tubes, we ensure that both samples are identical, enabling a reliable comparison between the results of dynamic olfactometry and e-nose signals.

As mentioned, the drone hovers at a point while sensors measure over a 5-minute period. Nine records per minute are logged, so each sensor generates a spectrum signal of 45 data points evenly distributed in time. In the

second minute, the air sample device is opened for one minute to collect the sample. After the 5 minutes, the drone is landed to retrieve and replace the air sampling bag for the next measurement. Once the full process is completed, the samples are transported to Odournet, S.L. (Sant Cugat), a dynamic olfactometry laboratory.

A four-day measurement campaign [2] was carried out at the WWTP during the summer of 2020, when odour intensifies. A total of 72 samples were obtained, including 41 odour measurements and 31 blanks. The odour ones were taken around the four main sources of odour emissions: settler, bioreactor, chimney and desander. Meanwhile, the blank samples were collected in different points from the surrounding area, where odour was not perceived by the expedition members.

III. MACHINE LEARNING METHODS

Machine Learning (ML) methods are used to predict response values. Supervised learning develops models using input and output data.

The sensor signals provide features that characterize the samples. As there are a total of 21 sensors and each sensor generates 45 data points, a vector of $m = 945$ (21×45) values is obtained per sample. Considering all $n = 72$ samples, the input data is a resulting feature matrix $\mathbf{X}$ $(n \times m)$ with 72 samples per 945 features.

Classification and regression problems are addressed using different datasets. For classification, the entire matrix $\mathbf{X}$ is used as input data, and the output is based on a feature $\mathbf{y}$ $(n \times 1)$ with categorical variables to distinguish the odour samples from the blanks. For regression, the blanks and one outlier were excluded, resulting only $n_{\text{reg}} = 40$ samples. The dimensions of matrix $\mathbf{X_{reg}}$ are $n_{\text{reg}} \times m$ and odour concentrations are represented in $\mathbf{y_{reg}}$ $(n_{\text{reg}} \times 1)$ as continuous variables.

This high-dimensional problem, with a few number of samples $(n < m)$ and features strongly correlated, is approached using multiple linear regression models. A linear combination of the variables is built to predict $\mathbf{y}$:

$$\mathbf{y_{pred}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{\beta}$ $(m \times 1)$ is the vector of regression coefficients and $\boldsymbol{\varepsilon}$ $(n \times 1)$ stands for random errors. The aim is to understand the data behavior, predict odour concentration and blanks, and identify important predictors.

A. Dataset preprocessing

Preprocessing data is a crucial step in ML before modelling, in order to identify significant information. The matrix $\mathbf{X}$ and feature $\mathbf{y}$ are logarithmically transformed to map data onto a Gaussian distribution, reducing dispersion, improving normality and linearizing the correlation between $\mathbf{X}$ and $\mathbf{y}$. Once the training and test partition is done, each dataset is mean-centered per feature: this standardization removes any bias or offset and reduces the multicollinearity between predictor variables (features). Furthermore, signal noise is reduced and relative changes in sensor response are amplified.

B. Cross-validation

To avoid overfitting, generalize the models for unseen data and accurately assess its performance, cross-validation (CV) technique splits the data into multiple training and test sets. In this study, a double CV is developed. Firstly, in external validation, four models are built using one of the four days' data as the test set and the remaining three for training. A second data partition (internal validation) is necessary for each model's training set, in order to calculate the optimal hyperparameters of the ML models. Although K-fold would be an interesting data splitting, the different test sets are composed by samples from each source and day. For classification problems, blanks are evenly distributed across all sources.

C. PLS and PLS-DA modelling

Partial Least Squares (PLS) and PLS Discriminant Analysis (PLS-DA) modelling are properly used for the requirements of the problem, as there are more predictors than equations. By maximizing the covariance between $\mathbf{X}$ and $\mathbf{y}$, the algorithm projects data onto a lower dimensional space defined by latent variables (LVs) [3]. $\mathbf{X}$-scores $(n \times h)$ is refered to the resulting projected data, and its dimension corresponds to the number h of LVs, which are orthogonal to each other.

Identify the important variables is essential for variable selection and, consequently, model optimization. The large of the PLS-regression coefficients $\boldsymbol{\beta}$ and PLS-weights $\boldsymbol{\omega}$ $(m \times h)$ can provide an initial idea, but the

TABLE 1: Number of samples collected in WWTP of Molina de Segura and categorized by sources and odour/non-odour.

Day	Date	Settler	Bioreactor	Chimney	Desander	Total (odour)	Blanks	Total
1	24/06/2020	2	3	3	1	9	7	16
2	25/06/2020	2	2	2	2	8	6	14
3	14/07/2020	3	3	3	3	12	11	23
4	15/07/2020	3	3	3	3	12	7	19
All days		10	11	11	9	41	31	72

method used is variable importance in projection (VIP). VIP summarizes the importance of each predictor j by considering the weights and the explained variance in $\mathbf{y}$ of each LV. The VIP score [4] for the jth predictor is calculated as:

$$\nu_j = \sqrt{p \frac{\sum_{a=1}^{h} \mathcal{R}^2(\mathbf{y}, \boldsymbol{t_a}) \left(\frac{\omega_{aj}}{||\boldsymbol{\omega_a}||}\right)^2}{\sum_{a=1}^{h} \mathcal{R}^2(\mathbf{y}, \boldsymbol{t_a})}} \tag{2}$$

where p is the number of predictors (features) and $\mathcal{R}^2$ is the coefficient of determination between $\mathbf{y}$ and the eigenvectors of each component ($\boldsymbol{t_a}$). The criterion used for feature reduction is Recursive Feature Elimination (RFE), an algorithm that iteratively discards a percentage of features with the lowest VIP scores.

PLS modelling predict continuous variables and is used for regression problems. The same algorithm can also be applied for classification but an additional step is required: PLS-DA transforms the $\mathbf{y}$ variables into categorical ones. In this case, a threshold classify the samples: values below it are assigned to one category, while values above it belong to the other one. The chosen threshold is the intermediate value between the two categorical variables that distinguish odour samples from blanks.

D. LASSO and logistic LASSO regression

Regularization models are often used to discard uninformative variables in the data. One of these sparse methods is LASSO (Least Absolute Shrinkage and Selection Operator) regression, which adds a ℓ_1-norm penalization to the cost function. The objective is to minimize:

$$\min_{\boldsymbol{\beta}} \left(\sum_{i=1}^{n} (\mathrm{y}_i - \beta_0 - \sum_{j=1}^{m} \beta_j \mathrm{X}_{ij})^2 + \lambda \sum_{j=1}^{m} |\beta_j| \right) \tag{3}$$

optimizing $\boldsymbol{\beta}$ and where $\lambda > 0$ is the ℓ_1-norm regularization parameter. It sets some model coefficients to zero and thus manages automatic feature selection.

On the other hand, logistic regression is used for odour classification [5] ($y_{\text{odour}} = 1$, $y_{\text{blank}} = 0$). The standard logistic function is referred to the predicted probability:

$$p_i = P(y_i = 1|\mathbf{X_i}) = \frac{1}{1 + e^{-(\beta_0 + \mathbf{X_i}\boldsymbol{\beta})}} \tag{4}$$

This function can be transformed to a linear regression as $\mathrm{y}_{i,\text{pred}} = \beta_0 + \mathbf{X_i}\boldsymbol{\beta} = \log \frac{p_i}{1-p_i}$ and results are found minimizing a cost function with ℓ_1-norm regularization parameter ($\lambda_{\log}$), similar to the previous one. Finally, samples are classified by applying a threshold (0.5) to $\mathbf{y}$ predicted, expecting a categorical target.

IV. RESULTS AND DISCUSSION

The preprocessed matrix $\mathbf{X}$ is illustrated in FIG. 1, where each sample vector is color-coded based on its odour concentration ($\mathbf{y}$). As a preliminary observation, sensors exhibiting greater signal variance demonstrate higher sensitivity to odour, thereby contributing more valuable information to the classification and regression.

The VIP score for each predictor is calculated as the average of the VIP scores for the four models built in external validation. By the results shown at FIG. 2, we confirm the important electrochemical sensors and the few relevant MOX sensors for PLS and PLS-DA modelling. Additionaly, it can be observed that the outstanding measurements over the five-minute period are around the third minute. This pattern holds true for all sensors except for H_2S and CO in PLS-DA.

A. Classification

In classification problems, although it may not be the most suitable metric to evaluate the model due to imbalanced data, the optimal hyperparamaters are selected by maximizing the accuracy score in internal validation.

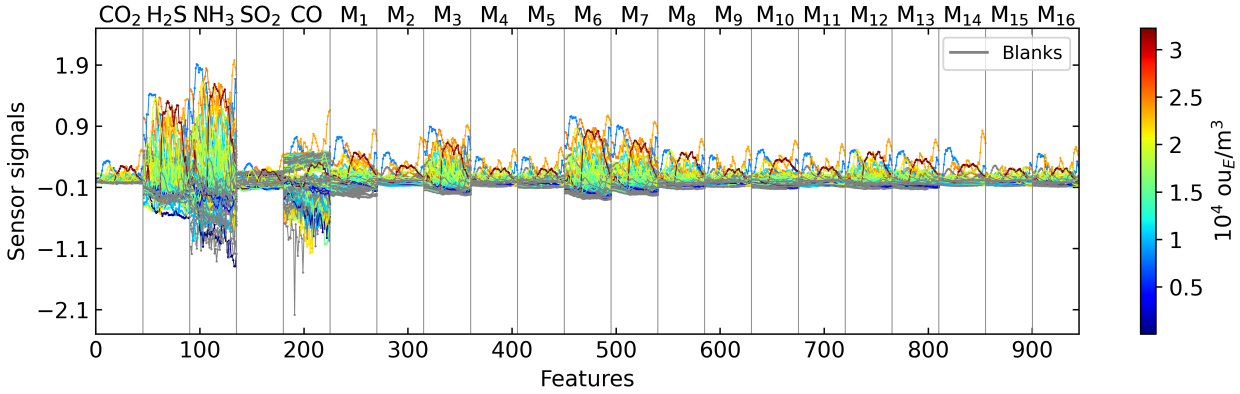

FIG. 1: Mean-centered values of sensor signals for all samples, colored by odour concentration.

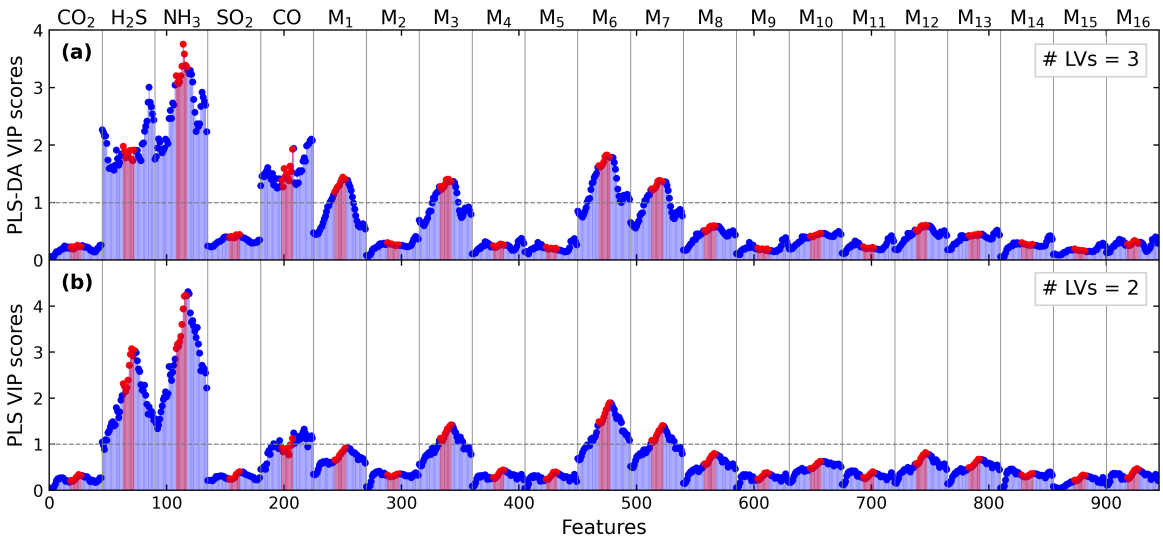

FIG. 2: VIP scores for each feature in PLS-DA and PLS modelling. The air sampling minute per sensor is colored red.

For PLS-DA modelling, ten internal validation models ($i = 1,...,10$) are constructed using a different number of LVs ($LV_i = 1,...,10$). Since the four external models consist of different sub-models, the accuracy and the average accuracy is computed for each. By applying the "elbow rule" [5], the optimal number of LVs determined is 3.

The PLS score plot of FIG. 3 shows a clear separation between odour and non-odour samples in different regions of the reduced-dimensional space. The clustering observed suggests that the model based on all e-nose signals is able to distinguish the blank measurements.

In reference to VIPs presented in FIG. 2(a) and using the the "greater than one rule" [4], H_2S, NH_3, CO, M_1, M_3, M_6 and M_7 (VIP sensors) are selected for model optimization. Then, with the same number of LVs, RFE is applied discarding a 10% of data per iteration. Features are recursively eliminated until a sudden change in accuracy takes place, known as the "knee of the curve". This cutoff point represents where model may reduce its performance due to the removal of important features. It is determined through internal validation, creating sub-models with the different number of features.

Otherwise, in logistic LASSO regression, multiple internal validation models are created considering a wide range of regularization parameter values. After determining the appropriate order of magnitude, the accuracy is maximized again and the most suitable value found is $\lambda_{\log} = 6.5$. The algorithm automatically selects the same sensors as those chosen in PLS-DA previously, setting the coefficients of the remaining sensors to zero.

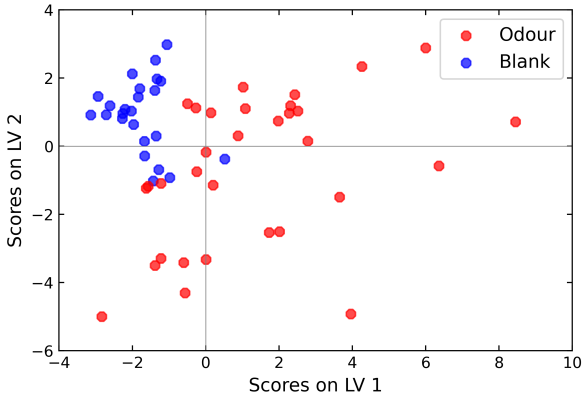

FIG. 3: PLS score plot of model 1 (day 1 as validation day).

By looking at the regression coefficients of both ML algorithms, we ensure that the four external validation models are comparable with each other because their coefficients are highly similar. Moreover, the coefficients for the central period of sensor measurements are the larger ones, as VIPs predicted. It may be a consequence of the response time needed for the sensors and will affect on feature selection. On the contrary, H_2S and CO sensors have a different behaviour. H_2S coefficients appear to be shifted down, possibly indicating the presence of an offset.

TABLE 2: Results of confussion matrix (odours as P; blanks as N) and classification metrics using different ML methods.

Model	# Features	TN	TP	FN	FP	Accuracy	Precision	Sensitivity	Specificity	F1 score
Logistic LASSO regression	945	29	34	2	7	88%	94%	83%	94%	88%
PLS-DA	945	26	38	3	5	89%	88%	93%	84%	90%
PLS-DA-RFE	191	27	40	1	4	93%	91%	98%	87%	94%

CO seems to measure the derivative of the signal observed in the other sensors. This discrepancy could be attributed to sensor sensitivity.

The results shown in TABLE 2 demonstrate the ability of PLS-DA and logistic LASSO regression to classify odour samples from blanks with a 88% and 89% accuracy, respectively, using the entire matrix. PLS-DA-RFE reduces **X** and a 93% accuracy is achieved. Only the central features from MOX VIP sensors are selected, together with all measurements of H_2S, NH_3 and CO.

B. Regression

In regression problems, the root mean squared error (RMSE) is minimized in internal CV using training sets. For PLS modelling and LASSO regression, hyperparameters are selected following the same procedure as for PLS-DA and logistic regression, respectively. The optimal number of LVs is 2 and $\lambda = 3.7 \times 10^{-2}$.

According to FIG. 2(b), 5 sensors take importance in PLS: H_2S, NH_3, M_3, M_6 and M_7. In the case of LASSO regression, just a few features of NH_3 sensor are selected by the 4 models. It suggests that this sensor is the crucial one to predict odour concentration. Even so, the results presented in TABLE 3 show that LASSO predictions are not as favourable as the ones of the optimized PLS model. PLS-RFE improves results by selecting only the features with the highest VIP scores, corresponding to NH_3.

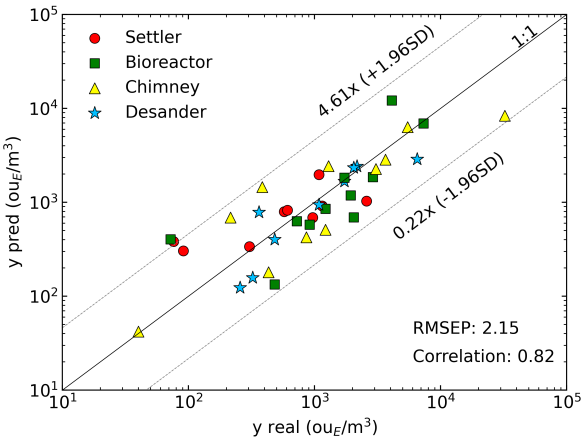

FIG. 4: Predicted and real odour concentration for the optimized PLS-RFE model, labelling the source of the samples.

TABLE 3: Predictive performance of odour concentration for different ML models. Bias is negligible in all results.

Model	# Features	RMSEP	Correlation	LoA
LASSO	945	2.28×	0.79	[0.19×, 5.19×]
PLS	945	2.74×	0.67	[0.13×, 7.52×]
PLS-RFE	8	2.15×	0.82	[0.22×, 4.61×]

Limits of agreement (LoA) represents the 95% confident of the differences between the predicted odour concentration and dynamic olfactometry results. LoA for dynamic olfactometry is already [0.5×, 2×] and a factor of four is obtained with PLS-RFE. RMSEP is reduced to a factor of two and FIG. 4 displays a correlation of 82%.

V. CONCLUSIONS

In this TFG, we have demonstrated that a flying electronic nose predicts with precision the odour in a WWTP, which can have environmental interest. By using the proposed IOMS, we obtain sensor signals and odour concentration for samples from four sources and for blanks. Different ML methods are proposed in order to classify and quantify odour. We present a double CV and a feature selection algorithm to assess their performance and improve their reliability.

On one hand, PLS-DA and logistic LASSO regression models classify accurately odour and blank samples using H_2S, NH_3, CO, M_3, M_1, M_6 and M_7 sensors. On the other hand, PLS-RFE and LASSO identify that NH_3 is the unique relevant sensor for predicting odour concentration. Furthermore, we have discussed that the important sensor response is situated in the middle of measurements, so these features will be selected for optimization.

The satisfactory results found indicate that this methodology could be an alternative to detect odour emissions in WWTP before they impact negatively in nature and society. Even so, due to the few samples used to calibrate the models, we are probably underestimating the complexity of the problem and it is necessary to continue investigating for better conclusions.

Acknowledgments

I would like to thank my advisor for all the guidance.

[1] Oliva, G. et al. (2021). "Optimization of Classification Prediction Performances of an IOMS by Using Temperature Correction Approach." Chemosensors, 9, 147.
[2] Burgués, J., Doñate, S., Esclapez, M., Saúco, L., Marco, S. (2022). "Characterization of odour emissions in a wastewater treatment plant using a drone-based chemical sensor system". Science of The total Environment, 846, 157290.
[3] Wold, S. et al. (2001). "PLS-regression: a basic tool of chemometrics." Chemomet. Intell. Lab. Syst., 58, 109-130.
[4] Chong, I., Jun, C. (2005). "Performance of some variable selection methods when multicollinearity is present". Chemomet. Intell. Lab. Syst., 78, 103-112.
[5] Duchesnay, E. et al. (2021). "Statistics and Machine Learning in Python." Engineering School, France,179-240.