ODOUR MONITORING WITH PORTABLE ION MOBILITY SPECTROMETRY

Veronica Villa, Luis Fernández Romero, Beatrice Julia Lotesoriere*, Javier Alonso-Valdesueiro, Agustín Gutierrez-Galvez, Lara Terrén, Lidia Sauco, Laura Capelli, Santiago Marco

The monitoring of odorous emissions is becoming more and more important due to the growing attention of citizens to air quality. Specific regulations and methodologies have been developed for odour monitoring from several industrial activities with a particular focus on wastewater treatment plants (WWTPs). Among these methods, Instrumental Odour Monitoring Systems (IOMS), i.e., electronic noses, are widely employed for real-time and continuous analysis.

In this context, in this work we aim to explore the potentiality of monitoring odours from a WWTP with a portable lon Mobility Spectrometry (IMS, from GDA2 by AIRSENSE Analytics), capable of accurately measuring several compounds, such as VOCs. Within the framework of the SNIFFIRDRONE project which aims to characterize odorous emissions from two wastewater treatment plants located in Pinedo (Valencia) named Pinedo 1 (P1) and Pinedo 2 (P2) treating respectively 130.000 m3/day and 220.000 m3/day of civil wastewater.

Three measurement campaigns were carried out in the field (in January, June, and July) to collect data accounting for the emissions variability due to seasonal and wastewater variations of the plant. Data analysis has focused on chemometric pre-processing of spectra and on exploratory analysis using Principal Component Analysis (PCA) individuating two main classes to be distinguished: Water treatments line and Sludge treatments line.

A selection of the most significant features has been performed with Random Forest to mitigate the possible multicollinearity effect, reducing computational effort. Partial Least Squares Discriminant Analysis (PLS-DA) has been employed to develop odour classification model with the January and June dataset from P1. The validation on July dataset of P1 achieved an accuracy of 87%, same results if the validation is performed on January and June of P2. The performances of this approach have been decisive to evaluate the necessity or not to introduce proper calibration transfer techniques. In recent years, specific regulations and methodologies have been developed for monitoring odour emissions from different industrial activities [1], as a response to the increasing attention of citizens toward air quality. In this context, Instrumental Odour Monitoring Systems (IOMS) (e.g., electronic noses) are gaining growing interest because of their ability to continuously monitor odour emissions both at emission sources or at receptors [1]. While some recent studies described off-line explorative analysis with Gas Chromatography Ion Mobility Spectrometry (GC-IMS) in WWTPs [2], [3] no literature has been found about the potential of stand-alone IMS detectors without chromatographic pre-separation. Indeed, IMS is a well-known technology to detect VOCs and other chemical species, characterized by portability, fast response, high sensitivity [4]. It is widely employed by first responders to detect toxic industrial chemicals, chemical warfare agents, explosives and drugs but also in pharmaceutical, food quality and environmental monitoring fields, making it an interesting candidate for odour sources characterisation. In this context, we decided to investigate the potentiality of a portable IMS to characterize malodours in WWTPs, which are considered as a major source of citizens' odour complaints [1]. Three measurement campaigns have been organized and performed at two WWTPs located in Pinedo (Valencia). The investigation mainly focused on the design of suitable protocols for the execution of sampling campaigns at the two plants by IMS. Moreover, considerable efforts were devoted to the implementation of specific data processing procedures for IMS signals aimed at odour detection and classification.

Materials and methods

Instrument and technology

The instrument employed in the field measurements is a gas detector array GDA2 from AIRSENSE Analytics GmbH (Schwerin, Germany). It is a portable and handheld instrument equipped with 4 gas sensors (n° 1 Photo Ionisation Detector (PID), n° 1 Electrochemical Cell (EC), n° 2 Metal Oxide Sensors (MOS)) and an IMS. IMS is an analytical technique detecting VOCs through ionization of the sampled molecules using of reference reactant ions (RIP for positive ions and RIN for negative ions) formed by a radioactive source (Ni- β source in the case of GDA2) [4]. The ionized molecules (e.g., the sampled molecules and the unreacted RIP and RIN) are accelerated by an electric field up to an approximately constant terminal velocity. Depending on the mass, charge, shape and collisional cross-section of the ions their terminal velocity is different causing them to reach the final detector at different drift time moments (in the range of milliseconds [ms]). The resulting positive and negative spectra are composed of peaks, located with respect to the reference ions (i.e., RIP and RIN), providing a unique fingerprint representative of the composition of the sampled gas [4]. This study focuses exclusively on these IMS signals as raw spectra signals representative of odour fingerprints to investigate the possibility to distinguish between different odour sources in a WWTPs.

Measurement campaigns and data collection

The measurements have been performed in the WWTPs of Pinedo (Valencia, Spain) which is composed of two semiindependent sections named Pinedo 1 (P1) and Pinedo 2 (P2) treating different flowrates of wastewater from civil residues. Three measurement campaigns were carried out in the field (in January, June, and July) to collect data (sampled sources reported in Table 1) accounting for the odour-emitting sources variability due to seasonal and wastewater variations of the plant. In the campaign of January, the samples were taken directly in ambient air over the sources with the GDA2. In June and July, a specific sampling system was designed to sample directly over the sources equipping the GDA2 with a PTFE tube and an extensible arm.

Data structure

IMS signals from GDA2 are 3D signals because every second of measurement a scan of both positive (with a totality of *m* drift time points) and negative spectra (with a totality of *m* drift time points) is recorded. For data processing

purposes, the data have been organised in a matrix where the rows are the scans recorded every second of measurement for each sample while the columns contain the intensity of the signals at every drift time point concatenating positive (m drift time points) and negative (m drift time points) spectra in the same row (for a total number of points each row of 2m=1022).

Data pre-processing

The signals recorded with the IMS in the GDA2 have been pre-processed employing standard algorithms for noise reduction (Savitzky-Golay digital filters with a window of width 10 and polynomial of order 2), baseline correction (polynomial approach with fourth order polynomial correction) and peak alignment (using a multiplicative factor with respect to the RIP or RIN position) [4]. Before final data processing the pre-processed signals have been autoscaled.

Data processing

After pre-processing, the preliminary dataset collected in January was investigated through exploratory PCAs to extract useful features to discriminate different odour sources [5]. Random Forest (RF) algorithm was then employed to perform feature selection on the dataset of June, characterized by more repetitions of measurements among the different days of field campaign and randomization, in an attempt to obtain more robust models and reduce multicollinearity effects and overfitting. The most important features have been selected on the basis of the minimal depth distribution among the trees representative of the relevance in classifying the samples. The PLS-DA algorithm has been implemented to build a classification model to distinguish between the *Water* and *Sludge treatment line* units [5]. Calibration and external validation have been performed on different data to optimize the classification models. Finally, the models have been validated on the new dataset of July.

The PLS-DA models were developed to classify the single spectra of each sample. The performance of the models was evaluated on the basis of the accuracy of the classification of the samples whose classification is based on the class assigned by the model to the individual scans belonging to that same sample. To account for the dataset imbalance a balanced accuracy [6] was used which is simply reported from now on as "Classification accuracy". The scheme of the overall data processing workflow is reported in Figure 2.

RESULTS

Preliminary results

The preliminary dataset of January allowed to verify that the pre-processing of the signals was fit-for-purpose allowing the alignment of the spectra and removing noise. PCA scores plots have been used to identify the best features extraction approach to obtain a good data clustering suggesting the potential capability of the IMS of discriminating the different odour classes. Indeed, at first, the entire spectra points (i.e., 1022), as discussed in section II.A and II.C, have been used as variables, obtaining overlapping clusters of the different sources (Figure 1 (A)). With the purpose of reducing dimensionality and extract features informative of the odour fingerprint of the different emissions sources it was decied to focus exclusively on the the peaks' of intensity in the final spectra signal and not on the entire spectra points. Since the RIP and the RIN peaks are only representative of the unreacted reference reactant ions and not of the specific compounds emitted by the sampled sources (i.e. of the odour fingerprint) [4], they have been removed from the extracted features. 16 peaks have been This feature extraction reduced considerably the features dimension from 1022 to 16. PCA score plots show better and clearer clusters (Figure 1 (B)). These scores plots clearly show the difficulty related to separating all the individual sources sampled (Figure 1 (B)). Looking at the distribution of the points belonging to the different samples in the PCs score plots, it was evident the overlapping of the sources belonging to the same process treatment line. This resulted in the definition of two macro-classes: the Water treatment line (Water collector, Grid/Screening, Primary sedimentary tank, Aeration tank) and Sludge treatment line (Thickener, Floater, Mixer) (Figure 1 (C)). These two classes have been considered in the final PCA analysis and in the classification models development.

Odour classification model development

The dataset of June (day 1 & day 2 & day 3) has been employed, for each of the WWTPs respectively, P1 and P2. At first, the same pre-processing and features extraction developed in the preliminary phase were applied to this new data. Concerning P1 June, day 1 & day 2 data were used to calibrate the model and then validating it against day 3 data, achieving 100% of classification rate of the samples (Table 2, first row). A lower classification accuracy of 75% was achieved when the exact same model was tested against the entire P2 June dataset (day 1 & day 2 & day 3) (Table 2, second row). To assess the performance of the classification models within the two WWTPs, the entire dataset of P1 June (day 1 & day 2 & day 3), was then tested against the entire dataset of P2 June (day 1 & day 2 & day3) achieving a classification accuracy of 65% (Table 2, third row). Then, the features selection with RF led to the selection of the 10 most important peaks of intensity within the extracted ones. With these new features selected, the same combinations of calibration and external validation proposed above were repeated with successful results. Indeed, calibrating again on P1 June day 1 & day 2 data and validating against P1 June day 3 data, resulted again in 100% classification accuracy (Table 2, first row) but with a lower computational demand respect before since fewer features are considered. The same model was validated against the entire dataset of P2 June (day 1 & day 2 & day 3) achieving a classification accuracy of 70% (Table 2, second row). The model calibrated on the entire dataset of P1 June (day 1 & day 2 & day3) and validating against the entire P2 June dataset (day 1 & day 2 & day 3), achieved now a classification accuracy of 80% (Table 2, third row). This proves that feature selection helped reducing the potential effect of multi-collinearity improving the classification rates and decreasing the computational demand.

Odour classification model validation

The developed pre-processing and data processing steps of features extraction and selection defined in the previous phases (III.A and III.B) were applied on the dataset in phase of validation. In order to account for seasonal variability and use a larger dataset to achieve higher robustness of the model, the January and June datasets of P1 were used together to calibrate the final model, which is reported as P1 January-June dataset. The PLS-DA classification model calibrated on P1 January-June dataset, including all the data collected in January and June, with the extracted and selected features, was successfully validated against the P1 July dataset achieving a classification accuracy of 87% (Table 3, first row).

Similarly, to the previous case, the same model (calibrated on P1 January-June dataset) was tested against P2 January-June dataset resulting in 87% classification accuracy (Table 3, second row). This rough transfer of the model built for P1 on the dataset of P2 results satisfying since it is achieved the same target accuracy of 87% (Table 3, first row) obtained by carrying out the validation on the same plant (Table 3, second row). This indicates that specific calibration transfer techniques might be not necessary in this case.

CONSIDERATIONS AND CONCLUSIONS

On the basis of these results, the developed strategy and the experimental approach in odour sources classification with a portable IMS instrument in WWTPs can be considered successful, within the analysed scenario. This exploratory analysis highlights the potentiality of IMS for odour emissions monitoring to be used as an IOMS to captured the odour fingerprint of different emissions sources in WWTPs. More measurement campaigns should be performed to have a complete view of the variability of the odour emissions of the plants over at least one year to make the classification models more robust. Moreover, more measurements would be essential to investigate the background odours of the plants, the distinction between air and odour samples, and to investigate the chemical species predominantly present in each class. Including one-year variability, it might be possible to verify the reliability and stability of the model and the necessity of specific calibration transfer techniques. Based on the successful results obtained in odour sources classification, it would be interesting to use the collected dataset also for investigating the potential towards odour quantification. This is made possible by the fact that the odour samples collected were analysed by dynamic olfactometry to determine their odour concentration.

ACKNOWLEDGMENT

The authors gratefully acknowledge Depuración de Aguas del Mediterráneo (DAM) for the collaboration.

References

[1] C. Bax, S. Sironi, and L. Capelli, 'How Can Odors Be Measured? An Overview of Methods and Their Applications', *Atmosphere*, vol. 11, no. 1, Art. no. 1, Jan. 2020, doi: 10.3390/atmos11010092.

[2] L. Vera, E. Companioni, A. Meacham, and H. Gygax, 'Real Time Monitoring of VOC and Odours Based on GC-IMS at Wastewater Treatment Plants', *Chem. Eng. Trans.*, vol. 54, pp. 79–84, Jan. 2016, doi: 10.3303/CET1654014.
[3] W. Gao *et al.*, 'Limitations of GC-QTOF-MS Technique in Identification of Odorous Compounds from Wastewater: The Application of GC-IMS

as Supplement for Odor Profiling', Atmosphere, vol. 12, no. 2, Art. no. 2, Feb. 2021, doi: 10.3390/atmos12020265.

Guamán Novillo, Ana Verónica, 'Multivariate Signal Processing for Quantitative and Qualitative Analysis of Ion Mobility Spectrometry data, [4] applied to Biomedical Applications and Food Related Applications', Ph.D. Thesis, Universitat de Barcelona, 2015. [Online]. Available: http://hdl.handle.net/10803/349210

S. Marco and A. Gutierrez-Galvez, 'Signal and Data Processing for Machine Olfaction and Chemical Sensing: A Review', IEEE Sens. J., vol. 12, [5] no. 11, pp. 3189–3214, Nov. 2012, doi: 10.1109/JSEN.2012.2192920.

A. Tharwat, 'Classification assessment methods', Appl. Comput. Inform., vol. 17, no. 1, pp. 168–192, Jan. 2020, doi: 10.1016/j.aci.2018.08.003. [6]