UNIVERSITAT DE BARCELONA

Multivariate Signal Processing for Quantitative and Qualitative Analysis of Ion Mobility Spectrometry data, applied to Biomedical Applications and Food Related Applications

Ana Verónica Guamán Novillo

FACULTAT DE FÍSICA

Departament d'Electrònica

MEMÒRIA PER OPTAR AL TÍTOL DE DOCTOR PER LA UNIVERSITAT DE BARCELONA

Doctorat en Enginyeria i Tecnologies Avançades (RD 99/2011)

Multivariate Signal Processing for Quantitative and Qualitative Analysis of Ion Mobility Spectrometry data, applied to Biomedical Applications and Food Related Applications

by

Ana Verónica Guamán Novillo

Director:

Dr. Antonio Pardo

Codirector:

Dr. Josep Samitier

Tutor:

Dr. Antonio Pardo

CONCLUSIONS

Conclusions of this thesis

In this thesis, several methodologies and strategies have been presented for improving and tackling open issues in IMS field when biomedical and food samples are analyzed. The results of this thesis have been presented in two major groups. On the one hand, different approaches were established for resolving typical problems of qualitative analysis, covering also pre-processing strategies. The second group was focused on proposals for enhancing quantitative analysis, especially pointing out the importance of using multivariate algorithms instead of univariate ones.

The main conclusions of this thesis are following presented:

i. In chapter two an open discussion about comparison of three different commercial spectrometers were done. To address the comparison study, a set of experiments with three different biogenic amines have been designed and developed. The limit of detection and sensitivity of the three spectrometers were calculated for the three different amines. The main conclusion is the three spectrometers show similar performance in LOD and sensitivity, even though they differ so much from each other in their operating condition. It is important to mention that spectra from the three spectrometers were carefully pre-processed in order to get the best performance of each device. Thus, we do not guarantee similar results without a proper signal processing analysis.

ii. An additional conclusion is preprocessing of the spectra and specifically, the noise filtering and signal enhancement procedures are key elements for the LOD calculations in IMS instruments.

iii. About noise filtering, a multivariate strategy for eliminating low frequency noise coupled to IMS spectra has been developed. As much as I know this is the first time this strategy has been applied in the IMS field.

Principal Component Analysis (PCA) and Independent Component Analysis (ICA) have been tested as multivariate strategy for specific low frequency noise filtering. Both procedures provide good performance but slightly better results were obtained with ICA rather than with PCA.

These multivariate techniques were tested against conventional smoothing techniques. It has been shown that an improvement on the signal to noise ratio is obtained using multivariate strategies. Despite of the fact this methodology was designed for giving a solution of a particular device, it can be extrapolated for other commercial IMS with the same problem.

iv. About preprocessing, this work also discusses about misalignment of IMS peaks. The necessity to correct misalignment within and between samples has been demonstrated. In this sense, solutions based on additive correction to two different scenarios (spectra with or without a peak of reference) have been tested.

For the first scenario (spectra with a reference peak), additive correction provides good results and it is simply to implement. If IMS measurements do not provide a set of data with a peak that can be used as reference, it is

advisable to introduce a substance able to generate a reference peak without disturbing (or disturbing the less) the spectra. For the second scenario (spectra without a reference peak or peaks) a whole spectrum (instead of a single peak) is used as reference for the correction, providing also good results.

v. Usually procedures for qualitative detection substances with IMS use univariate estimations of height or area of a single peak. This work has demonstrated that multivariate strategies perform better results and interpretation. New IMS applications in either biomedical field or food and beverage field cannot be addressed with univariate strategies.

An application about wine discrimination has been solved using a multivariate strategy consisting on a combination of principal component analysis and discriminant analysis called PCDA.

An application about discrimination of rats with sepsis and healthy rats has been also solved using multivariate strategies. In this case, Multivariate Curve Resolution (MCR) in combination with feature selection techniques has been used. MCR allows the extraction of pure components from spectra, and feature selection allows the selection of the most significate components for class separability. Even though the use of MCR and feature selection is not novel in IMS field, the fact of combining both techniques is new. The main advantage of this methodology is allowing the identification of the main compounds. In the hypothetical case of having a library of spectra, this procedure will allow an identification of the compounds.

vi. In this thesis a specific effort on robust validation has been done. Cross-validation methodology using bootstrap has been implemented in order to guarantee the results. Overoptimistic results and overfitting of the models has been avoided with the validation procedures implemented. Validation strategies have received less or null attention in the multivariate processing of IMS spectra. This thesis claim for the use of robust validation procedures in the signal processing of IMS datasets.

vii. This thesis were also motivated by giving solutions to open issues of IMS such as nonlinearities and mixture effect. Again, it has been proved that univariate techniques are not a good alternative for the analysis of IMS complex data.

Synthetic dataset were used for tacking non linearity's and mixture effect in IMS. We demonstrate that the use of multivariate techniques, such as Partial Least Square and Multivarite Curve Resolution, are able to solve these issues in IMS.

In the case of non linearites, it was evident that multivariate techniques offers better results than univariate techniques. In addition, the fact of using MCR-ALS offers the alternative to extract only the pure compounds relevant in the quantification. In terms of nonlinearities both PLS and MCR-ALS gives quite similar results.

The mixture effect was evaluated in terms of limit of detection. The main idea was to observe what happen with the LOD of an analyte when it is mixed with other analyte with a similar proton affinity. When the univariate technique was used, the LOD was twice bigger than LOD of the dataset with the single anlayte. When the multivariate algorithms were used, the results show that LOD was slightly different but closer to LOD of the single analyte. Both algorithms were not enough able to solve the mixture effect, but a good approximation

were obtained. From these results we can conclude that there is a clear effect in presence of a complex matrix, which need to be deeply explored.

viii. Once was demonstrated the usefulness of applying multivariate strategies. The methodologies proposed in this thesis were tested in real samples in clinical and food and beverages applications. The results were promising showing the enormous potential of IMS in food industry and clinical fields.