Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

IMPROVEMENT OF PIXEL CLASSIFICATION BY THE SIMULTANEOUS USE OF SPECTRAL AND SPATIAL INFORMATION IN THE FRAMEWORK OF SPECTROSCOPIC IMAGING

Alessandro Nardecchia

Univ. Lille, CNRS, UMR 8516 – LASIRE – Laboratoire de Spectroscopie pour Les Interactions, La Réactivité et L'Environnement, Lille, F-59000, France.

Raffaele Vitale

Univ. Lille, CNRS, UMR 8516 – LASIRE – Laboratoire de Spectroscopie pour Les Interactions, La Réactivité et L'Environnement, Lille, F-59000, France.

Eric Ziemons

University of Liege (ULiege), CIRM, Vibra-Sante Hub, Laboratory of Pharmaceutical Analytical Chemistry, Avenue Hippocrate 15, 4000, Liege 1, Belgium.

Ludovic Duponchel

Univ. Lille, CNRS, UMR 8516 – LASIRE – Laboratoire de Spectroscopie pour Les Interactions, La R´eactivit´e et L'Environnement, Lille, F-59000, France.

HIGHLIGHTS

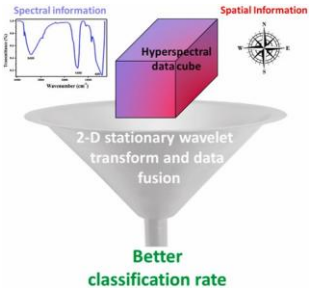
- A way to merge spectral and spectral information of a hyperspectral imaging dataset prior any chemometric exploitation.

- A real opportunity to increase the predictive capabilities of classification models.

- This fusion/data augmentation approach can be used whatever the considered chemometric tool.

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

GRAPHICAL ABSTRACT



Abstract. Hyperspectral imaging technology is developing in a very fast way. We find it today in many analytical developments using different spectroscopies for sample classification purposes. Instrumental developments allow us to acquire more and more data in shorter and shorter periods of time while improving their quality. Therefore, we are going in the right direction as far as the measure is concerned. On the other hand, we can make a more mixed assessment for the hyperspectral imaging data processing. Indeed, the data acquired in spectroscopic imaging have the particularity of encoding both spectral and spatial information. Unfortunately, in chemometrics, almost all classification approaches today only use spectral information from three-dimensional hyperspectral data arrays. To be more precise, an approach encompassing the unfolding/refolding of such arrays is often applied beforehand because the majority of algorithms for analysing these data are not capable of handling them in their original structure. Spatial information is therefore lost during the chemometric exploration. The study of the spectral part of the acquired data array alone is clearly a limitation that we propose to overcome in this work. 2-D Stationary Wavelet Transform will be used in the data preprocessing phase to ensure the joint use of spectral and spatial information. Two spectroscopic datasets will then be used to evaluate the potential of our approach in the context of supervised classification.

Keywords. Hyperspectral imaging, Classification, Partial least squares discriminant analysis (PLS- DA), Spectral and spatial information fusion, 2D stationary wavelet transform (2D-SWT).

1. Introduction

Data analysis is currently the cornerstone of many areas of interest. The development of always more powerful instruments is only one of the reasons of this particular trend. Hyperspectral imaging is one of the most promising analytical approaches in this sense. It is currently the subject of numerous studies due to its important characteristics [1–3]. A hyperspectral image is usually represented by a

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

three-dimensional data array. Its first two dimensions are related to spatial information while the third one is spectral. Differently from bulk analysis, a hyperspectral image can lead to results that are more representative of the heterogeneity of a given sample. At the same time, this kind of analysis is related to the simultaneous acquisition of the spectral, but also and especially, the spatial information inherent to the studied specimen. This aspect is very important nowadays, because a comprehensive analysis of a complex sample must be linked not only to its spectral assessment, but also to the corresponding spatial information linked, for example, to the local distribution of the different components of the observed matrix. Nevertheless, the challenge of fully exploiting the data acquired on a sample is still very complicated. In fact, regardless of the approach used, routine data analysis focuses more on the spectral part of the data. Chemometric approaches, in spite of their interesting features, are not an exception to this limitation. To be more precise, we cannot really explore a hyperspectral imaging data array in its original form by means of most of these approaches. As a consequence, spatial information is most often lost during the chemometric exploration and a spectrum is used independently from all the others acquired in its neighbourhood. Nevertheless, there are a few reports that show the potential for simultaneous use of spectral and spatial information in the data, but not in the context of supervised classification to our knowledge [4,5]. Intuitively, we understand that the simultaneous use of spectral and spatial information should allow us to improve the potential of chemometric approaches, for example in classification. Considering the constant improvement of the various spectroscopic acquisition systems, particularly related to the possibility of obtaining always more high resolution images, this aspect has to be clearly faced, finding new ways to deal with this kind of challenge. In the last decades, chemometrics has been exploited with the purpose of overcoming this limitation [6–8]. Nevertheless, using these approaches some constraints, mainly related to a not very fluent analysis, are unavoidable. One of the most interesting algorithms that could be utilized with such a purpose in mind is the 2D-stationary wavelet transform (2D-SWT) [9–13]. Applying 2D-SWT to a hyperspectral data array permits to extract the spectral information encoded in it, which could be afterwards merged with the original spectral data for a more complete analysis. In contrast to a previous study that our group has conducted in the context of unsupervised exploration of hypespectral data arrays [14], we here apply this information extraction and fusion procedure in order to show how the simultaneous use of spectral and spatial features can be exploited to perform a more accurate classification analysis via partial least squares discriminant analysis (PLS-DA) [15–17]. The potential of this approach will be evaluated using both a simulated and a real-world dataset, this latter resulting from the mid-infrared imaging analysis of micro-plastic samples. This will allow us to span different data complexities and, at the same time, systematically compare the correct classification rates obtained when using only spectral information or when resorting to the spectral/spatial data fusion.

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

2. Materials and methods

2.1. STATIONARY WAVELET TRANSFORM (2D-SWT)

Being this work an extension of a previous research study that our group carried out [14], here the concept of the wavelet transform and the corresponding fusion strategy will only be briefly discussed. Despite the existence of various wavelet algorithms, 2D-stationary wavelet transform (2D-SWT) has shown to be an interesting procedure for the extraction of the spatial information contained in grey-scale, RGB and hyperspectral images [18,19]. Like the well-known Fourier transform, the 2D-SWT is based on the use of low-pass and high-pass filters to decompose the signal under consideration. The application of 2D-SWT to a single image allows generating four other ones, which carry different information. The first one called approximation image (denoted A) reflects the low frequency information of the original image. The other three images called horizontal, vertical and diagonal detail images (denoted H, V and D, respectively) encode the high frequency information retrieved along different directions of the original image. If we go back to the hyperspectral data array, which contains as many images (also called slices in image analysis) as spectral variables in the considered wavelength domain, 2D-SWT can be systematically applied to each one of them as can be seen in Fig. 1. An augmented three-dimensional data structure is, thus, obtained by merging all the images generated by 2D-SWT to the initial hyperspectral array. The subsequent exploration of such a dataset by a chemometric algorithm (here PLS-DA), though, still requires its unfolding, but this time each pixel observation contains simultaneously spectral and spatial information (see Fig. 1). For the sake of clarity, the reverse biorthogonal wavelets family was selected in this work because it guaranteed the best performance [20]. Nonetheless, given the wide range of wavelet families, it is obvious that this choice must be taken so as to properly describe the local structure of the investigated hyperspectral image. Readers interested in this approach may wish to test first the reverse biorthogonal wavelets family and then explore other ones in an attempt of improving their results. For more information regarding the proposed method, the reader can refer to our previous publication [14].

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

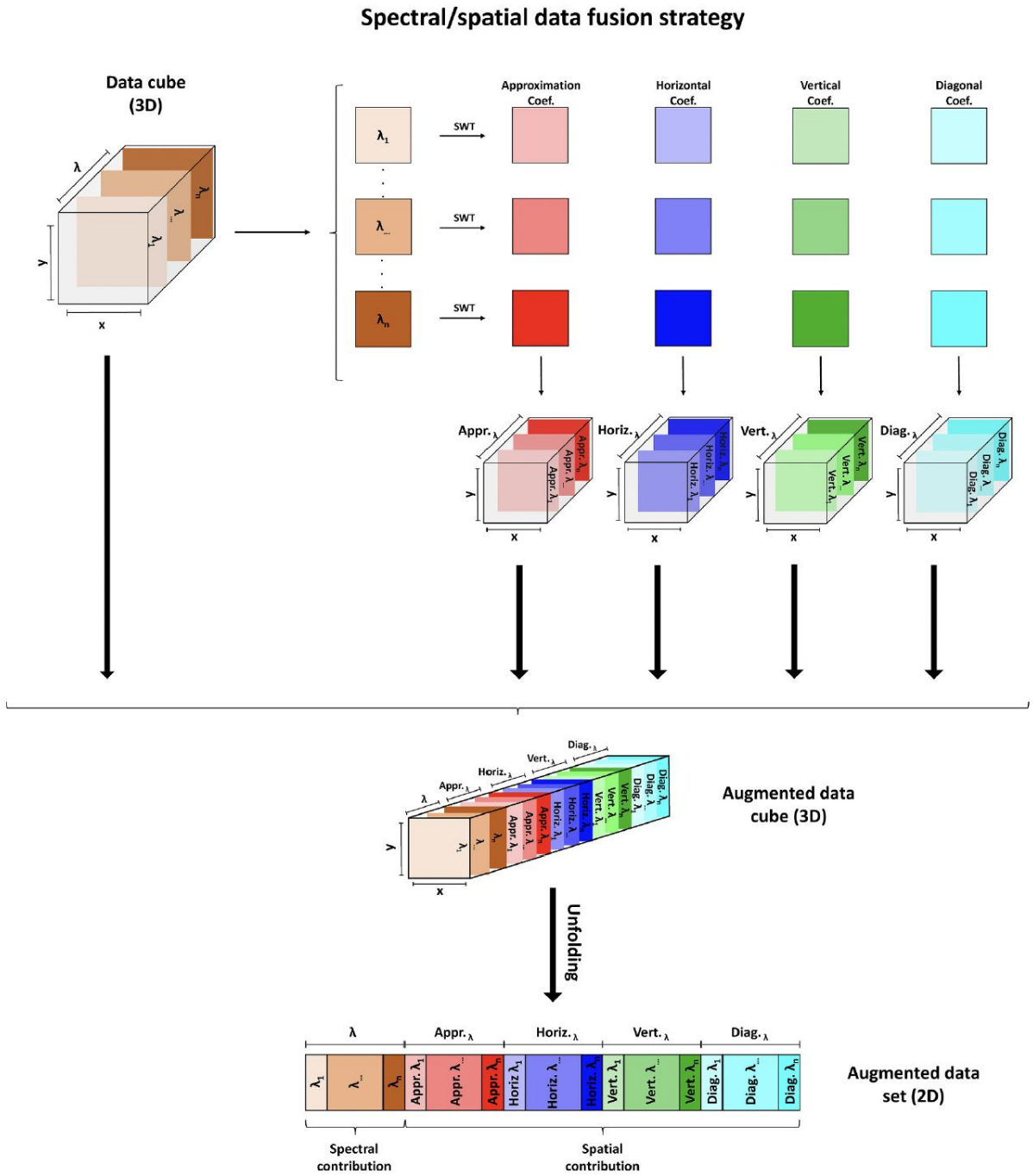


Fig. 1. Schematic representation of the proposed spectral/fusion approach based on 2D-SWT.

2.2. SIMULATED DATASET DESCRIPTION

The first dataset used in this work is a 4-class simulated hyperspectral image which permitted us to directly control the complexity of the analyzed data in terms of spectral interferences and specific spatial features. Fig. 2a shows the spectra of the 4 classes of pixels that we generated from linear combinations of strongly overlapping Gaussian profiles. Noise has also been added to account for the low signal-to-noise conditions often observed in spectroscopic imaging. A closer look at these spectra

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

allows us to see where the main difficulty in this supervised classification task will lie. Classes 1 and 3 seem to be potentially easy to discriminate from classes 2 and 4 because the former have selective contributions between spectral variables 1 and 80 approximately. If we focus only on classes 1 and 3, they could also be potentially discriminated as selective contributions are observed between the spectral variables 150 and 200 approximately. However, the average spectrum of class 3 is a linear combination of the average spectra of classes 1 and 2, which should render the classification more challenging. Things will certainly get even more complicated for classes 2 and 4 because they exhibit exactly the same spectral contributions with the same intensity ratios between bands, only their overall intensity varies. In the context of a classification problem, it can therefore be said that compounds 2 and 4 have roughly the same spectral fingerprint. Fig. 2b shows the location of the pixels belonging to the 4 different classes concerned within the hyperspectral data array (of dimensions 300 pixels × 300 pixels x 300 spectral variables – i.e., a total of 90,000 spectra). It can be seen that each class relates to areas with particular geometric characteristics. More precisely, the first and the second category of pixels are associated to circles of different size, while the third and the fourth ones to rectangles of different size. We thus generated systematic spectral and spatial information in this dataset. Lastly, Fig. 2c displays the two subregions of the dataset which will be used to train the PLS-DA model and evaluate it.

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

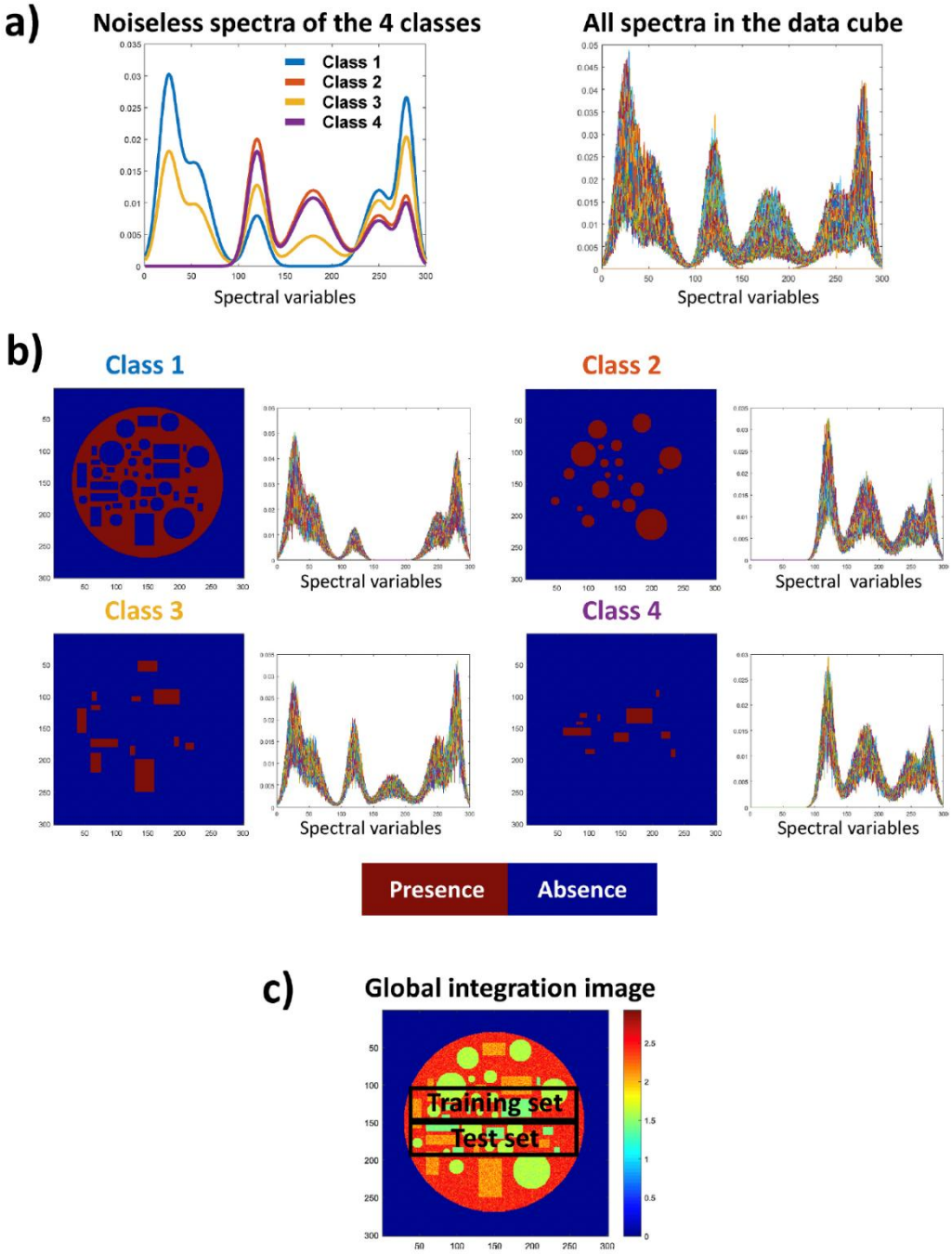


Fig. 2. The simulated dataset: a) average spectra of the 4 classes of pixels and pixel spectral profiles after the addition of noise; b) localisation of the 4 classes of pixels over the sample surface; c) localisation of the areas used for the generation of the training and test sets.

2.3. REAL DATASET DESCRIPTION

The first real dataset we used in this work resulted from a previous study on the characterization of microplastics and relates to a mid- infrared microscopy image (of dimensions 128 pixels × 128 pixels x 312 wavenumber) of polyamide particles deposited on a membrane [21]. This spectroscopic system uses a focal plane array detector (FPA) which allows a direct and simultaneous acquisition of 16,384 spectra (i.e. 128 × 128 pixels) on a sample region of interest. The measurements were conducted in

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

the 1300-3800 cm^{-1} spectral range, with a nominal spectral resolution of 8 cm^{-1} and a pixel size of 5.5 µm. At a first glance, this sample seems trivial since there is only a single polymer located on the membrane. However, this is not necessarily the case, especially with regard to the hypespectral data acquired on such a sample and more particularly the pixels located on the edges of the particles. In these areas, in fact, it is potentially difficult to observe a clear spectral fingerprint for the product due to the limited instrumental sensitivity, a lower signal to noise ratio and potential defocusing. We must not forget, indeed, that a particle is not flat at this scale. The greatest risk in this situation would then be to confound such pixels as background contributions and not to assign them to the polyamide class. This again may seem trivial, since a priori the particle could be detected even if classification errors occur at its edges. Nevertheless, in the context of environmental analysis, the interest of users and practitioners is often to determine the size distribution of plastic microparticles, a procedure that will clearly be influenced by how their contour is determined.

From a practical point of view, a first region of the membrane containing particles has been used to develop the PLS-DA model, while another region containing other particles was used to evaluate the model (i.e. a new set of 16,384 spectra).

The second dataset used in this work corresponds to the analysis of a powdered pharmaceutical sample by Raman imaging. A 1 g blend of powders was made of paracetamol and lactose (50% w/w). Paracetamol (Compap™ PVP3) and lactose (Pharmatose® 200 mesh) were kindly provided by Galephar M/F (Marche-en-Famenne, Belgium). The powders were weighed and placed in a mortar and mixed gently using a pestle. An aliquot was spread on a glass slide prior to Raman imaging analysis. Confocal micro-Raman imaging was performed on a Labram HR Evolution spectroscopic system (Horiba Scientific) equipped with an EMCCD detector (Andor Technology Ltd.), a Leica 50× Fluotar LWD objective and a 785 nm laser (XTRA II single frequency diode laser, Toptica Photonics AG) with a power of 106 mW at sample. The spectrometer was equipped with a 300 gr/mm grating in order to perform a spectral analysis over the 463-1853 cm^{-1} spectral range considering a 100 m s acquisition time per spectrum. The mapping experiment was done considering 10.3 µm and 6.2 µm steps in both directions of the sample plane. The size of the hyperspectral data cube was therefore 100 pixels × 100 pixels × 1600 wavenumbers i.e. 10000 acquired in approximately 16 min. The aim here was to identify the pharmaceutical product that had a very limited raman signal compared to glass, which was in the majority. From a practical point of view, a first region of the slide containing particles has been used to develop the PLS-DA model, while another region containing other particles was used to evaluate the model (i.e. a new set of 5000 spectra).

2.4. CLASSIFICATION MODEL

The aim of this work is above all to provide a first proof of concept on the benefits of using simultaneously spectral and spatial information in the framework of pixel classification in spectroscopic imaging. As outlined in section 2.1, the key aspect of our approach lies in the augmentation of the hyperspectral data array, which is performed so that even the final unfolded dataset to be explored contains both spatial and spectral information. This is an interesting point since we circumvent the development of new classification algorithms capable of taking into account all the dimensions of the original three-dimensional data structure. All classical classification tools could, in

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

principle, be used. Here, we have chosen to apply PLS-DA, one of the most commonly exploited methodologies in chemometrics [15–17]. As mentioned in the previous sections, from every hyperspectral image dealt with, we will extract a training set to build the corresponding classification model and a different test set for its validation. All calculations were performed in a Matlab 2016b environment (The MathWorks, Inc., Natick, United States of America) using PLS_Toolbox (ver. 8.2.1, Eigenvector Research Inc.).

3. Results and discussion

We used a relatively simple strategy to evaluate the potential of our spectral/spatial data fusion approach in the framework of classification. First, a classification model is built using the original raw data, i.e., exploiting only the spectral information encoded in the hyperspectral image at hand, and then another PLS-DA model is calibrated on the image pixel spectral profiles augmented with the spatial information yielded by 2D-SWT, as detailed before. Fig. 3 shows the prediction results on the test set for the simulated hyperspectral image when spectral information only is considered as well as when spatial information is added. Ground truth images are also provided for the 4 classes in order to visualise the relevance of the pixel membership prediction, with red and blue being used to differentiate assigned and unassigned pixels, respectively. This visual assessment already highlights interesting aspects: if we focus on the results yielded by the PLS-DA model constructed on the raw data (middle column of Fig. 3), we can observe that the pixels belonging to class 1 are mostly correctly recognized as such, which was rather expected. This is not really the case for the other 3 categories of pixels, which are characterized by strong spectral overlaps (see also Section 2.2). More specifically, i) classes 2 and 4 were found to be particularly confounded given their high spectral similarity, which translated into a relatively large number of false positive; ii) also for class 3, a significant amount of false positives, mainly proceeding from classes 1 and 2, can be observed (remind that the spectral profiles of the pixels belonging to class 3 were generated as linear combinations of those simulated for class 1 and class 2). If we now look at the outcomes returned by the PLS-DA model trained on the initial data augmented with the corresponding wavelet coefficients (right column of Fig. 3), we can see that all the previous ambiguities/confusions are evidently reduced, even if not completely. At the same time, it is worth noticing that, in this particular scenario, the optimal number of PLS-DA latent variables (LV) increased from 3 when using the raw spectral data to 13 when these are augmented with spatial information. This is due to the fact that the wavelet coefficients are orthogonal to each other, but, most importantly, that they provide complementary information with respect to the spectral data themselves. For a quantitative comparison, Table 1 lists the figures of merit (accuracy, sensitivity and specificity) related to the two different PLS-DA models calibrated in this first case-study. As a reminder, the equations for calculating these criteria are:

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$specificity = \frac{TN}{TN + FP} \quad (3)$$

where TP = True Positive, TN = True Negative, FP = False Positive and FN = False Negative.

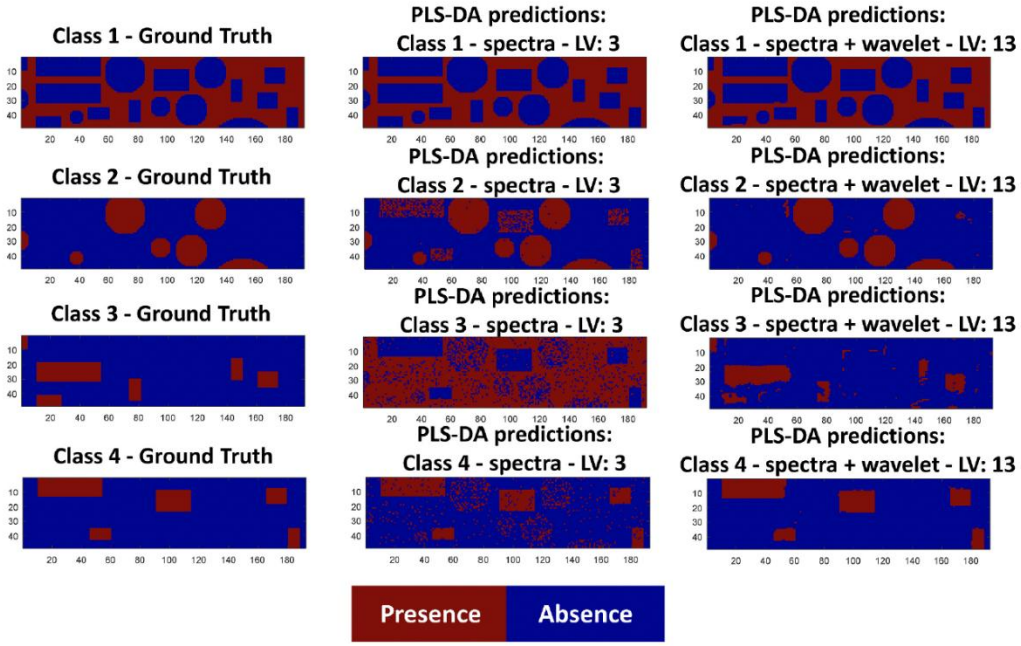


Fig. 3. PLS-DA results obtained for the simulated dataset when only spectral information is taken into account (center column) and when both spectral and spatial information are modelled (right column).

Table 1
PLS-DA figures of merit for the four classes obtained modelling only spectral and both spectral and spatial information, respectively (simulated dataset).

Only spectral information – 3 LV PLS-DA model			
Class	Specificity	Sensitivity	Accuracy
1	1	1	1
2	0.8932	0.9892	0.9400
3	0.4508	0.6599	0.5454
4	0.9140	0.9344	0.9241
Spectral + spatial information (wavelets) – 13 LV PLS-DA model			
Class	**Specificity**	**Sensitivity**	**Accuracy**
1	0.9919	0.9996	0.9957
2	0.9910	0.9989	0.9949
3	0.9754	0.8185	0.8935
4	0.9943	0.9863	0.9903

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

The same approach has then been applied to the microplastics dataset. Fig. 4a provides a visual representation of the class memberships of the test set pixels predicted by the different PLS-DA models. Here we faced a binary classification problem, with classes 1 and 2 being associated to the microplastic and to the image background, i.e., the membrane, respectively. For this new dataset, an optimal number of 6 latent variables was required when only the spectral data were exploited, while 13 were needed when spatial information was additionally taken into account. At a first glance, the differences in the prediction performance seem much smaller compared to the previous example. Nevertheless, it can be said that i) many isolated pixels located in the surroundings of large plastic microparticles and mistakenly classified as polyamide ones by the standard PLS-DA approach were instead properly recognized as belonging to the membrane support when also spatial information was processed (see, e.g., Fig. 4a) and b) microplastic particle edges seem to be better defined when the augmented spectral/spatial data are analyzed. The figures of merit for the two PLS-DA models constructed in this second case-study are given in Table 2 (once again, overall differences might be slight, but all are systematically localized in particular regions of interest of the image, e.g., the borders between microparticles and membrane). We show here that the simultaneous use of spectral and spatial information in hyperspectral imaging allows to better discriminate the pixels at the edges of the objects present in the investigated scene and, by extension, to potentially improve the determination of their size and their size distribution. In order to corroborate these conclusions, we have also recalculated the aforementioned classification figures of merit for each image sub-area represented in Fig. 4b (see Table 3). Such an assessment permitted to emphasize the actual difference between the two distinct classification models.

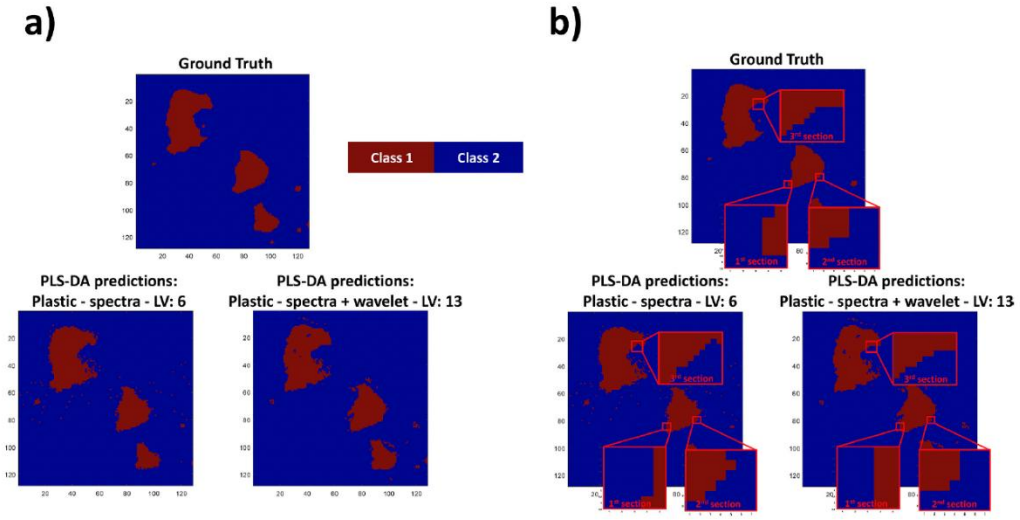


Fig. 4. a) PLS-DA results obtained for the microplastics dataset when only spectral information is taken into account and when both spectral and spatial information are modelled; b) zoomed insets highlighting the classification improvement at the edges of the plastic microparticles.

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

Table 2
PLS-DA figures of merit for the two classes obtained modelling only spectral and both spectral and spatial information, respectively (microplastics dataset).

Only spectral information – 6 LV PLS-DA model			
Class	Specificity	Sensitivity	Accuracy
1	0.9719	0.9010	0.9358
2	0.9010	0.9719	0.9358
Spectral + spatial information (wavelets) – 13 LV PLS-DA model			
Class	**Specificity**	**Sensitivity**	**Accuracy**
1	0.9709	0.9760	0.9735
2	0.9760	0.9709	0.9735

Table 3
PLS-DA figures of merit for the two classes obtained modelling spectral and both spectral and spatial information, respectively, and related to the three different image sub-areas represented in Fig. 4b (microplastics dataset).

Only spectral information – 6 LV PLS-DA model			
Selected subarea	Specificity	Sensitivity	Accuracy
1	0.8947	0.5714	0.7150
2	0.8571	1	0.9258
3	0.8333	0.9216	0.8763
Spectral + spatial information (wavelets) – 13 LV PLS-DA model			
Selected subarea	**Specificity**	**Sensitivity**	**Accuracy**
1	0.8333	1	0.8333
2	0.9643	1	0.9643
3	0.9792	1	0.9895

This last section concerns the classification of the powdered pharmaceutical product. Fig. 5a provides a visual representation of the class memberships of the test set pixels predicted by the two PLS-DA models (i.e. using only spectral information or both spectral and spatial ones) which can be compared with the ground truth. We are still in the framework of a two-class classification, the first being the paracetamol/ lactose mixture and the second being the glass slide. Again, it is quite difficult to see differences at first glance. In fact, the potentially observable differences are still on the edge of the particles such as the previous case. It should be noted once again that the PLS-DA model using spectral and spatial data simultaneously requires more components, i.e. 6 as opposed to 3 when only spectral data are used. In order to better observe these differences we have enlarged two specific areas of the sample surface represented in Fig. 5b. Tables 4 and 5 respectively give the figures of merit for the whole surface of the sample and two sub- sections as for the previous sample. These results still show significant differences. It is again interesting to consider simultaneously the spectral and spatial dimensions of a hyperspectral dataset in the framework of supervised classification.

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

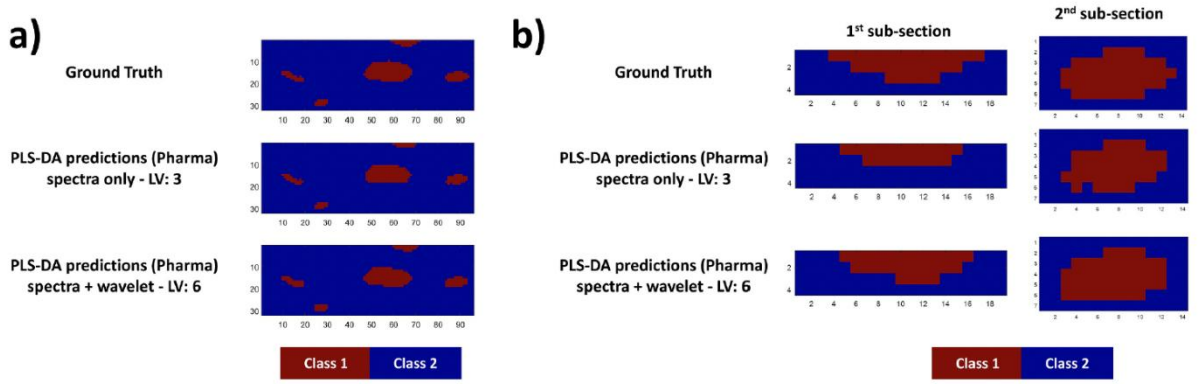


Fig. 5. a) PLS-DA results obtained for the pharmaceutical dataset when only spectral information is taken into account and when both spectral and spatial information are modelled; b) sub-section highlighting the classification improvement at the edges of the particles.

Table 4
PLS-DA figures of merit for the two classes obtained modelling only spectral and both spectral and spatial information, respectively (pharma. dataset).

Only spectral information – 3 LV PLS-DA model			
Class	Specificity	Sensitivity	Accuracy
1	1	0,8681	0.9317
2	0.8681	1	0.9317
Spectral + spatial information (wavelets) – 6 LV PLS-DA model			
Class	**Specificity**	**Sensitivity**	**Accuracy**
1	0.9975	0.9487	0.9728
2	0.9487	0.9975	0.9728

Table 5
PLS-DA figures of merit for the two classes obtained modelling spectral and both spectral and spatial information, respectively, and related to two different image sub-sections (pharma. dataset).

Only spectral information – 3 LV PLS-DA model			
Selected sub-section	Specificity	Sensitivity	Accuracy
1	1	0.6552	0.8684
2	1	0.8780	0.9490
Spectral + spatial information (wavelets) – 6 LV PLS-DA model			
Selected sub-section	**Specificity**	**Sensitivity**	**Accuracy**
1	1	0.8966	0.9605
2	1	0.9756	0.9898

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

4. Conclusions

Spectroscopic imaging data naturally encode spectral and spatial information on the investigated scenes. Nevertheless, it must be noted that this latter is rarely exploited when applying chemometric approaches for their analysis, which may lead to a less accurate characterization of such scenes. To overcome this limitation taking advantage of all the useful information contained in a hyperspectral data array, we have here proposed an approach for spectral/spatial fusion fundamentally based on the principles of 2D-SWT. Using three datasets with different characteristics and complexity, we have shown that this approach could potentially enhance the prediction performance of PLS- DA models constructed only on the raw unfolded hyperspectral measurements. This is an opportunity to recall that this work is primarily a feasibility study of the concept. Consequently, it will be necessary to observe the behaviour of this approach on even more hyperspectral datasets in order to definitively demonstrate its potential. It is worth mentioning that this methodology not only suits PLS-DA data modelling, but can easily be resorted to in combination with other classification/ discrimination techniques that can readily benefit from it without having to be adapted or somehow modified.

CRediT authorship contribution statement

Alessandro Nardecchia: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. Raffaele Vitale: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. Eric Ziemons: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. Ludovic Duponchel: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

Acknowledgement

We would like to thank Dr. Jose Manuel Amigo from the University of Basque Country for providing the mid-infrared imaging dataset and for the constructive discussions about this project. It should be noted that the same dataset is now contained in a free matlab toolbox called HYPERTOOLS (https://www.hypertools.org/) that he developed for chemometric analysis in spectroscopic imaging [2,22].

References

[1] J.M. Amigo, I. Martí, A. Gowen, Hyperspectral imaging and chemometrics, in: Data Handling in Science and Technology, Elsevier, 2013, pp. 343–370, https://doi.org/ 10.1016/B978-0-444-59528-7.00009-0.

[2] J.M. Amigo, H. Babamoradi, S. Elcoroaristizabal, Hyperspectral image analysis. A tutorial, Anal. Chim. Acta 896 (2015) 34–51, https://doi.org/10.1016/j. aca.2015.09.030.

[3] A. Gowen, C. Odonnell, P. Cullen, G. Downey, J. Frias, Hyperspectral imaging – an emerging process analytical tool for food quality and safety control, Trends Food Sci. Technol. 18 (2007) 590–598, https://doi.org/10.1016/j.tifs.2007.06.001.

[4] R. Vitale, S. Hugelier, D. Cevoli, C. Ruckebusch, A spatial constraint to model and extract texture components in Multivariate Curve Resolution of near-infrared hyperspectral images, Anal. Chim. Acta 1095 (2020) 30–37, https://doi.org/ 10.1016/j.aca.2019.10.028.

[5] A. de Juan, M. Maeder, T. Hancewicz, R. Tauler, Use of local rank-based spatial information for resolution of spectroscopic images, J. Chemometr. 22 (2008) 291–298, https://doi.org/10.1002/cem.1099.

[6] M.H. Bharati, J.J. Liu, J.F. MacGregor, Image texture analysis: methods and comparisons, Chemometr. Intell. Lab. Syst. 72 (2004) 57–71, https://doi.org/ 10.1016/j.chemolab.2004.02.005.

[7] J.J. Liu, J.F. MacGregor, On the extraction of spectral and spatial information from images, Chemometr. Intell. Lab. Syst. 85 (2007) 119–130, https://doi.org/ 10.1016/j.chemolab.2006.05.011.

[8] F. Jamme, L. Duponchel, Neighbouring pixel data augmentation: a simple way to fuse spectral and spatial information for hyperspectral imaging data analysis, J. Chemometr. 31 (2017), e2882, https://doi.org/10.1002/cem.2882.

[9] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, IEEE Trans. Inf. Theor. 36 (1990) 961–1005, https://doi.org/10.1109/ 18.57199.

[10] L. Debnath, F.A. Shah, in: Wavelet Transforms and Their Applications, Birkhauser ¨ Boston, Boston, MA, 2015, https://doi.org/10.1007/978-0-8176-8418-1.

[11] M. Li Vigni, J.M. Prats-Montalban, A. Ferrer, M. Cocchi, Coupling 2D-wavelet decomposition and multivariate image analysis (2D WT-MIA): coupling 2D-WT to multivariate image analysis (2D WT-MIA), J. Chemometr. 32 (2018), e2970, https://doi.org/10.1002/cem.2970.

Published in: Analytica Chimica Acta (2023), vol. 1242, 340805
DOI: 10.1016/j.aca.2023.340805
Status : Postprint (Author's version)

[12] M. Li Vigni, M. Cocchi, Multiresolution analysis and chemometrics for pattern enhancement and resolution in spectral signals and images, in: Data Handling in Science and Technology, Elsevier, 2016, pp. 409–451, https://doi.org/10.1016/ B978-0-444-63638-6.00013-9.

[13] M. Ahmad, R. Vitale, C.S. Silva, C. Ruckebusch, M. Cocchi, A novel proposal to investigate the interplay between the spatial and spectral domains in near-infrared spectral imaging data by means of Image Decomposition, Encoding and Localization (IDEL), Anal. Chim. Acta 1191 (2022), 339285, https://doi.org/ 10.1016/j.aca.2021.339285.

[14] A. Nardecchia, R. Vitale, L. Duponchel, Fusing spectral and spatial information with 2-D stationary wavelet transform (SWT 2-D) for a deeper exploration of spectroscopic images, Talanta 224 (2021), 121835, https://doi.org/10.1016/j. talanta.2020.121835.

[15] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemometr. 17 (2003) 166–173, https://doi.org/10.1002/cem.785.

[16] U.G. Indahl, H. Martens, T. Næs, From dummy regression to prior probabilities in PLS-DA, J. Chemometr. 21 (2007) 529–536, https://doi.org/10.1002/cem.1061.

[17] S. Chevallier, D. Bertrand, A. Kohler, P. Courcoux, Application of PLS-DA in multivariate image analysis, J. Chemometr. 20 (2006) 221–229, https://doi.org/ 10.1002/cem.994.

[18] G.P. Nason, B.W. Silverman, The stationary wavelet transform and some statistical applications, in: A. Antoniadis, G. Oppenheim (Eds.), Wavelets and Statistics, Springer New York, New York, NY, 1995, pp. 281–299, https://doi.org/10.1007/ 978-1-4612-2544-7_17.

[19] Cé Vonesch, T. Blu, M. Unser, Generalized daubechies wavelet families, IEEE Trans. Signal Process. 55 (2007) 4415–4429, https://doi.org/10.1109/ TSP.2007.896255.

[20] S. Mallat, in: A Wavelet Tour of Signal Processing, Elsevier, 2009, https://doi.org/ 10.1016/B978-0-12-374370-1.X0001-8.

[21] V.H. da Silva, F. Murphy, J.M. Amigo, C. Stedmon, J. Strand, Classification and quantification of microplastics (<100 µm) using a focal plane array–fourier transform infrared imaging system and machine learning, Anal. Chem. 92 (2020) 13724–13733, https://doi.org/10.1021/acs.analchem.0c01324.

[22] N. Mobaraki, J.M. Amigo, HYPER-Tools. A graphical user-friendly interface for hyperspectral image analysis, Chemometr. Intell. Lab. Syst. 172 (2018) 174–187, https://doi.org/10.1016/j.chemolab.2017.11.003.