

Uncommon multivariate statistical methods for environmental studies: A review



Mauro Mecozzi ^{a,*}, Mabel Beatriz Tudino ^b, Maria Grazia Finoia ^c, Marcelo Enrique Conti ^d

^a Italian National Institute for Environmental Protection and Research, Laboratory of Chemometrics and Environmental Applications, Via di Castel Romano 100, 00128 Rome, Italy

^b INQUIMAE, Departamento de Química Inorgánica, Analítica y Química Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

^c Italian National Institute for Environmental Protection and Research, Via Vitaliano Brancati 60, 00166 Rome, Italy

^d Department of Management, Sapienza University of Rome, Via del Castro Laurenziano 9, 00161 Rome, Italy

ARTICLE INFO

Article history:

Received 24 November 2014

Received in revised form 19 February 2015

Accepted 24 February 2015

Keywords:

Environmental studies

Multivariate methods

Independent component analysis

Two-dimensional correlation spectroscopy

Spectroscopic analysis

Chromatographic analysis

ABSTRACT

In this paper we describe the characteristics and the applications of the multivariate methods for spectroscopic and chromatographic techniques independent component analysis (ICA) and two-dimensional correlation spectroscopy (2DCOS) focused to their use in environmental studies. In our opinion, these methods are important because they allow to characterize environmental samples with different aims and scopes from those generally obtained by means of more common multivariate methods such as principal component analysis (PCA) and partial least squares (PLS). The new insights of these methods in recent environmental studies are reviewed and debated.

© 2015 Elsevier B.V. All rights reserved.

Contents

1. Introduction	31
2. Discussion on the application of ICA and 2DCOS for environmental studies	33
2.1. Pre-processing of analytical signals for the correct interpretation of ICA and 2DCOS results	33
2.2. ICA theory	33
2.2.1. Identification of the correct number of ICs	33
2.2.2. Application of ICA in environmental FTIR spectroscopic studies	34
2.2.3. Application of ICA in environmental UV–vis and fluorescence spectroscopic studies	34
2.2.4. Application of ICA in environmental chromatographic studies	34
2.3. Theory of 2DCOS	36
2.3.1. Wavelength–wavelength 2DCOS	36
2.3.2. Sample to sample 2DCOS	36
2.3.3. Spectroscopic applications of 2DCOS	37
2.3.4. Application of 2DCOS in chromatographic studies	38
3. Conclusions	38
References	38

1. Introduction

In analytical chemistry, principal component analysis (PCA) and partial least squares analysis (PLS) are multivariate statistic techniques widely applied as methods for quantitative analysis

* Corresponding author. Tel.: +39 0650073287.

E-mail addresses: mauro.mecozzi@isprambiente.it, mauromecozzi2004@libero.it (M. Mecozzi).

and for the discrimination and differentiation of multivariate data samples [1–3] by means of specific chemometric software [1,4,5].

As far as environmental studies concern, PCA has been applied to study the distribution of hydrocarbons [6,7] and heavy metals in marine sediments [8], to support the identification of baseline levels of heavy metals in marine and terrestrial organisms [9–14], to study the different driving force involved in the mechanisms of organic matter degradation in East Siberian and Laptev seas [15] and to determine physiological cell states in environmental water quality control by FTIR spectroscopy [16].

With the same aim, PLS has been applied to support vibrational spectroscopy in the assessment of environmental quality of soils and sediments [17,18], to improve the analytical accuracy in the determination of total carbohydrate contents in seawater [19] and to support the simultaneous determination of ten polycyclic aromatic hydrocarbons in natural water by fluorescence spectroscopy [20].

Being well known and widely applied, PCA and PLS applications in environmental studies are practically unlimited and for this reason we do not include the discussion of their uses in this review.

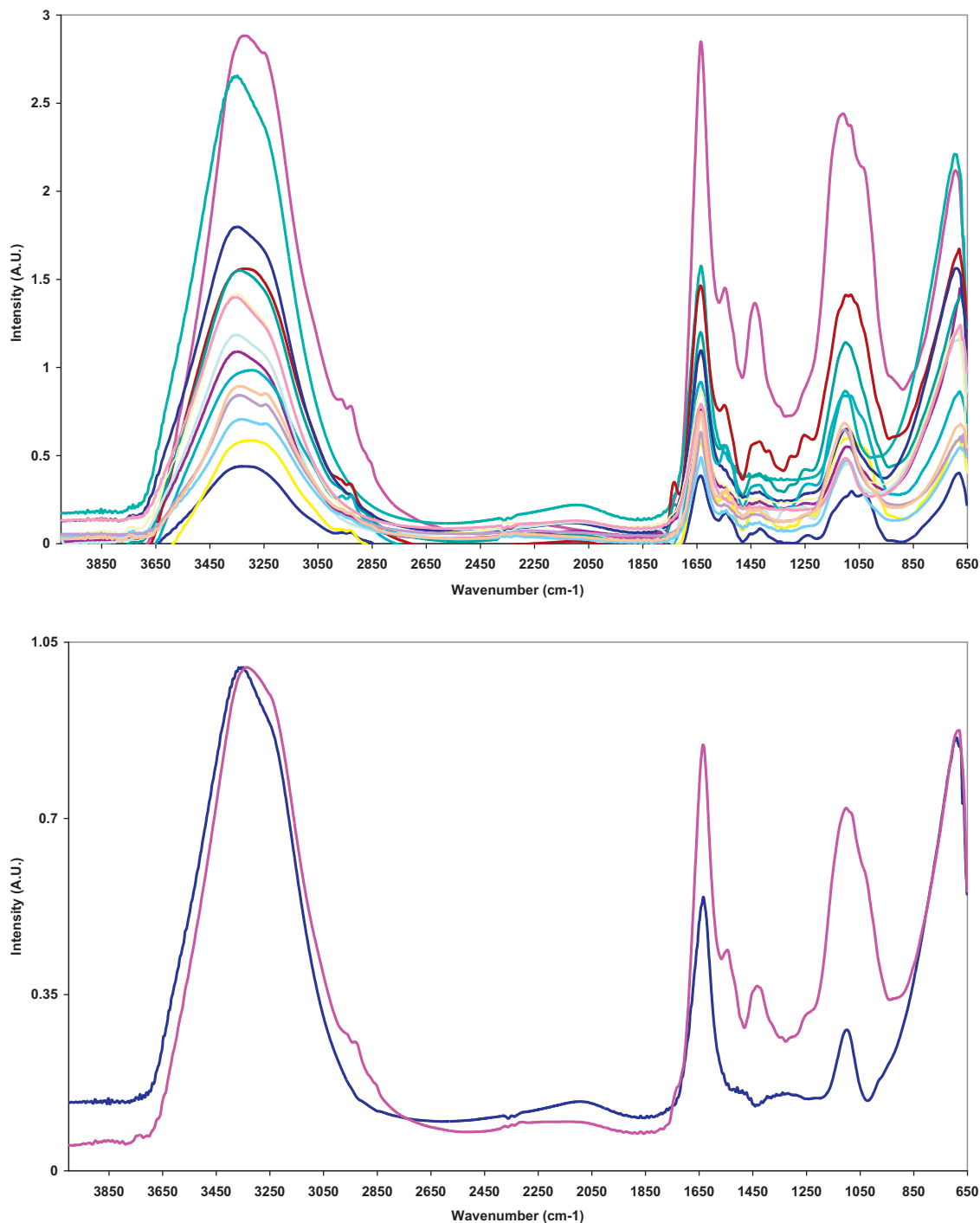


Fig. 1. Example of ICA application by means of the MILCA algorithm to a FTIR spectral data set of samples of marine organic matter during the process of evolution (followed for 21 days). The bottom plot reporting two spectra (i.e., two ICs) only shows that the whole process of organic matter aggregation is described by the formation of amino acids (the blue spectrum) and a complex mixture of compounds (the pink spectrum) where polysaccharides proteins high polymerized are present [31]. (For interpretation of the reference to color in this figure legend, the reader is referred to the web version of this article.)

Here, we describe and review the applications in environmental studies by means of spectroscopic and chromatographic techniques of two less common multivariate techniques such as independent component analysis (ICA) and two-dimensional correlation spectroscopy (2DCOS) because they are scarcely applied in environmental studies. In fact, with respect to PCA and PLS, they are neither applied as quantitative nor as discrimination methods for chemical analysis.

ICA identifies the real sources of the analytical signals corresponding to well defined subsamples present in the data set and do not describe abstract factors like PCA and PLS, which remain basically methods for reducing the data dimension [21]. 2DCOS allows to describe reactions and evolutionary mechanisms of molecular systems submitted to external perturbations which produce structural changes [22].

According to these characteristics, ICA and 2DCOS give new tools and perspectives in environmental studies and in this review these aspects are fully debated by means of the recent literature.

2. Discussion on the application of ICA and 2DCOS for environmental studies

2.1. Pre-processing of analytical signals for the correct interpretation of ICA and 2DCOS results

Like all the multivariate methods, ICA and 2DCOS require appropriate pre-processing treatments for the correct interpretation of spectroscopic and chromatographic signals. In fact, pre-processing (i.e., baseline correction, smoothing and standardization) of analytical signals is necessary to support a correct interpretation of spectra and chromatograms by enhancing the signal to noise ratio and by improving spectral and chromatographic resolution. The discussion of these pre-processing methods is out of the aim of this review, so we recommend some papers as skill guides for the pre-processing of analytical signals [23–25].

2.2. ICA theory

ICA belongs to the group of neural network methods [26], aimed to identify the number of independent components (ICs, the pure and latent sources of signals) present in unknown proportion in the mixture of analytical signals. Basically, given the complex matrix \mathbf{X} consisting of t samples and n signals each, ICA can identify the number of the significant j -ICs which determine

$$\mathbf{X} = a_1\text{IC}_1 + a_2\text{IC}_2 + \dots + a_j\text{IC}_j \quad 1 \leq j \leq t \quad (1)$$

It is important to underline two fundamental characteristics and properties of ICA. The ICs are the latent and independent signals of samples describing the original \mathbf{X} matrix and they are restored

according to a blind process. Blind means that no one preliminary information of these components is required for the decomposition (i.e., deconvolution) of the signals in the \mathbf{X} matrix [21,26]. The blind characteristics of ICA also differentiate it from other multivariate methods such as PCA and PLS. PCA and PLS describe latent and abstract factors which are linear combination of the original variables belonging to the system (i.e., the \mathbf{X} matrix) under investigation. So PCA and PLS latent factors are able to discriminate samples present in the \mathbf{X} matrix but they do not describe any existing physical reality. Conversely, ICA describes and identifies the real source signals corresponding to all the existing subsamples present in the system under investigation. These subsamples are the ICs which produce the final spectral or chromatographic plot of overlapped signals [27]. Common ICA applications include sound and image analysis studies mainly, however, chemists have been also attracted by its powerful ability in the analysis of complex chemical signals. Wang et al. [27] report the first review of ICA applications for analytical chemistry by means of the several existing ICA algorithms. Most of these ICA algorithms (FastICA, JADE, MILCA) are freely available on the web [28–30].

Fig. 1 reports an example of ICA application to a FTIR spectral data set of marine organic matter during the evolution of their aggregation process (Fig. 1, upper plot), while the other plot (Fig. 1, bottom plot) shows the two independent (i.e., chemical) components describing the whole characteristics of the process [31]. The two spectra in the bottom plot are the two ICs (latent subsamples) able to describe all the structural characteristics of the fourteen spectra present in the upper plot of Fig. 1. Table 1 shows the list of ICA applications in environmental studies reviewed in this paper.

2.2.1. Identification of the correct number of ICs

Like PCA and PLS, the decomposition of the \mathbf{X} matrix by ICA needs the correct identification of the significant ICs, so to separate the ICs having well defined physical–chemical meanings from the remaining ICs which consist of signal artefacts (i.e., noise) only. The separation between significant and noisy ICs determines reliability, stability and robustness of the ICA results [21,32]. Both in FastICA and JADE algorithms, the \mathbf{X} matrix is preliminary submitted to a PCA treatment in order to obtain a general noise reduction. This implies that the IC number is the same of the components identified by PCA and consequently this ensures the stability of the algorithm when the size of the samples in the dataset is small compared to the data dimension [32].

Specific methods support the identification of the significant IC number. The minimal value of the sum-square root of residues between the original and the reconstructed data is a common tool for ICs number identification, like in PCA and PLS [27,33]. The Amari index value when lower than 0.05 is the evidence of a good decomposition of the \mathbf{X} matrix whereas values above 0.2 correspond to a poor and unsatisfactory ICA model [34]. According

Table 1
Applications of ICA in environmental studies.

Analytical technique	Sample type	Determined properties	Data treatment	Reference
FTIR	Petroleum	Petroleum sources	JADE	[36]
FTIR	Petroleum	Petroleum refinery sources	JADE	[37]
FTIR	Marine mucilage	Mucilage formation	FastICA	[38]
FTIR	Organic matter	Structural characteristics	FastICA	[39]
FTIR	Organic matter	Structural characteristics	FastICA, JADE, MILCA	[40]
FTIR	<i>Ostreopsis</i> o., alga	Spectral interferences correction	FastICA	[41]
UV–vis	Seawater	Anion determination	MILCA	[42]
UV–vis	Natural water	Metal determination	MILCA	[44]
GC	Sediment	PAHs	FastICA	[46]
GC	Petroleum	PAHs	FastICA	[47]

to the Durbin–Watson criterion, the related index tends to 0 when the identified ICs are noise free whereas it tends toward 2 when one or more ICs consist of noise [35].

2.2.2. Application of ICA in environmental FTIR spectroscopic studies

The identification of the petroleum constituents and fractions in commercial gasoline is important for quality control purposes and for determining the fingerprinting of petroleum sources when oil spills occur in terrestrial or in marine environments. The JADE algorithm application to the FTIR spectral matrix of different commercial gasoline performs the extraction of the spectral components which characterize each gasoline or oil spill sample allowing the recognition of the sources [36]. This study also shows that ICA application for the identification of specific refinery fractions outperforms the results obtained by means of PCA applied to the same FTIR matrix. In fact, the correlation coefficients calculated between the derived ICs and the original spectra of the constituents, are significantly higher than the respective ones obtained from PCA. Furthermore, the same authors improved the performances of this FTIR–ICA approach for the identification of refinery fractions by applying techniques of spectral encoding as alternative to the correlation coefficient as similarity tool [37].

Marine mucilages are anomalous size aggregates of marine organic matter floating on sea surface and along the water column which cause problems for fish and tourist activities; this phenomena has an unpredicted recurrence in the Mediterranean sea. As mucilages can be easily reproduced in laboratory by the degradation of marine algae, the FTIR spectra of these organic matter aggregates were collected and then examined by the FastICA algorithm with the aim to identify the chemical mechanisms leading to their formation [38]. ICA results showed that the process of mucilage formation can be described by two ICs mainly. One IC corresponds to the presence and formation of oligopeptides and oligosaccharides coming from the degradation of higher molecular weight compounds such as proteins and polysaccharides. The second IC corresponds to the compounds arising from the repolymerization of oligopeptides and oligosaccharides, leading to higher polymerized compounds such as proteins and polysaccharides, with in addition the development of supramolecular structures involving their interactions with lipids. Due to these results, it is possible to state that marine mucilage formation depends on specific hydrological factors present in the marine environment which are not strictly related to marine pollution which remains the most diffuse opinion on the causes of this phenomena.

Calace et al. [39] applied the FastICA algorithm to the classification of a two hundred FTIR spectral data base of organic matter samples extracted from lake, river and marine sediments.

ICA determined three ICs in the data set, linked to the proteins, lipids and carbohydrates; these three ICs describe the chemical reactions present in organic matter formation and aggregation. This study also confirmed the importance of the supramolecular interactions for the formation and the stabilization of the types of organic matter aggregates [36,38].

The study of the complex mechanisms of organic matter aggregation is also the aim for comparing the performances of the three FastICA, JADE and MILCA algorithms [40]. Results showed that MILCA outperforms FastICA and JADE, giving more resolved and noiseless ICs, without the spectral ambiguities arising from the presence of both positive and negative spectral peaks. This study also reports the mathematic reasons explaining the best performances of the MILCA algorithm with respect to the FastICA and JADE.

Ostreopsis ovata is a marine algae present in Italian seas which can produce toxins that enter the food chain causing several problems to the marine environment and human health [41]. Cultures of this alga, growth in batch reactor under different

hydrological conditions, were obtained and examined by FTIR spectroscopy to test the physiological effects on this alga by the hydrological conditions. However, if the adhesion of algal biofilms on the wall of the batch reactor is not good, a spectral interference from the bicarbonate present in seawater makes hardly to compare the spectral bands of samples growth in different conditions. The application of FastICA to FTIR spectra allowed to separate the IC bicarbonate from the IC of algal samples and then to perform the correct (i.e., interference free) spectral comparison among the algal samples of *O. ovata* [41].

2.2.3. Application of ICA in environmental UV–vis and fluorescence spectroscopic studies

The UV absorption of seawater depends on the presence of several anions such as bromide, sulphide, bicarbonate and nitrate, producing strongly overlapped peaks between 200 and 250 nm range. In the monitoring activity of the marine environment, the determination of these anions requires a specific analytical method for each anion. PCR, PLS and the MILCA and JADE algorithms were tested and compared to identify the most reliable methods to perform the simultaneous determination of the above ions by UV spectroscopy [42]. MILCA method gave comparable results with PLS for nitrate, bicarbonate and bromide determination and even better in the case of sulphide determination, so that MILCA outperformed PLS. Many reasons can explain these results. Like PLS, MILCA is able to take into account the non linear UV spectral absorption of seawater depending on the high amount of dissolved anions; this peculiar ability is not present in PCR. Moreover, the better performances of MILCA with respect to PLS observed in this study for sulphide determination depend reasonably by the specific characteristics of the MILCA algorithm which works in the first derivative space (derivatives of spectral curves with respect to wavelength). This allows to correct some spectral artefacts depending on strongly overlapped peaks and then to improve spectral resolution [43].

Monakhova et al. [44] apply and compare several ICA algorithms for the simultaneous UV–vis determination of seven metals in different food, pharmaceutical and environmental samples. ICA application gives comparable and in some specific cases even better performances than the classical PCR and PLS methods.

The simultaneous determination of several organic pollutants such as polycyclic aromatic hydrocarbons by fluorescence spectroscopy, supported by multivariate calibration, is actually an interesting alternative approach to HPLC methods due to the possibility of saving time and costs of analysis [20]. In this field of studies, Gao and Ren [45] apply the combination of ICA and Neural Network Regression, obtaining remarkable improvements of the analytical accuracy for pollutant determinations with respect to the determinations performed by using PLS or neural network regression only.

2.2.4. Application of ICA in environmental chromatographic studies

The FastICA algorithm has been applied for the estimation of the biogenic and anthropogenic hydrocarbon sources in marine sediments for a GC data set of seventy samples from four different areas of Italian seas [46]. FastICA identified five ICs and by means of the hydrocarbon markers observed in these ICs, it was possible to estimate the hydrocarbon sources present in the four marine areas. In addition, FastICA allowed to establish a direct compositional link between one of the coastal areas considered in the study and one of the identified IC.

Quantitative analysis of polycyclic aromatic hydrocarbons in petroleum products by HPLC techniques is a topic of several environmental studies. However, the analytical accuracy of these determinations is often affected by the complex hydrocarbon

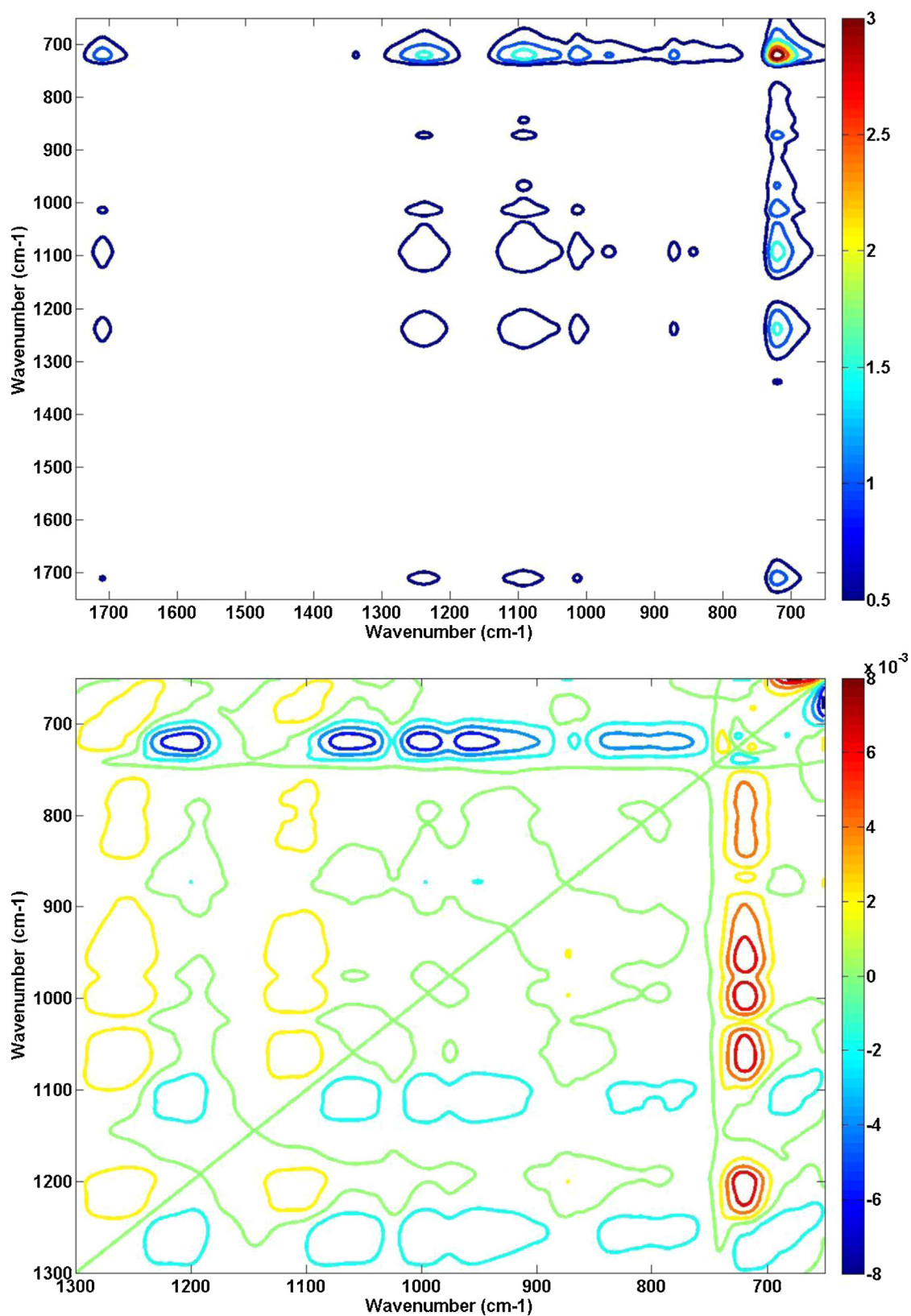


Fig. 2. Example of synchronous (upper plot) and asynchronous (bottom plot) spectra of a sixteen polyethylene terephthalate sample data set coming from marine plastic debris litter, collected along the coasts of Tyrrhenian sea [31]. Asynchronous spectra are very helpful to detect differences in the crystallinity and orientation of the polymeric structures present as shown by the crosspeaks in the 700–1300 cm⁻¹ range [72].

composition of samples showing strongly overlapped peaks. Being ICA a deconvolution technique, its application improves the analytical accuracy of the determinations by enhancing the chromatographic resolution of overlapped peaks [47,48].

2.3. Theory of 2DCOS

Two-dimensional correlation analysis is a family of multivariate methods for investigating the dynamic evolution of a molecular system submitted to an external factor and/or perturbation which determines changes in its structure and composition [22]. This technique has been originally developed for spectroscopic analysis in the field of polymer and material sciences, however, due to its mathematic fundamentals, the applications of 2DCOS to other analytical techniques and to other fields of studies, including environment are of potential interest [22].

2.3.1. Wavelength-wavelength 2DCOS

In 2DCOS, the X data matrix is constructed by sorting samples according to a specific criteria which can be the time of sample acquisition or the concentration of the perturbation agent [22]. Then synchronous (Synww), asynchronous (Asynww) and disrelation (Dis) spectra of the molecular system are determined. Synww spectra describe the structural changes occurring in the system according to linear (i.e., in phase) relationships among the variables. For spectra aligned as column vectors, Synww spectra are determined according to

$$\text{Synww} = \frac{X \times X^T}{(s-1)} \quad (2)$$

where X is the spectral data matrix previously described, X^T is its transposed matrix and s is the number of samples in the matrix.

Asynww spectra describe the nonlinear (out of phase) relationships among the variables of the molecular systems and are determined according to

$$\text{Asynww} = \frac{X \times HX^T}{(s-1)} \quad (3)$$

where X , X^T and s have been previously defined and H denotes the Noda's modification of the Hilbert transform matrix.

Dis spectra are determined according to

$$\text{Dis} = [\text{Var}(X) \times \text{Var}(X^T) - \text{Synww}^2]^{0.5} \quad (4)$$

where Var is the total variance of the X and X^T matrices and Synww^2 is the quadratic Synww matrix. Dis spectra display all the information occurring in the absence of correlations among the variables, so describing the unrelated events occurring in the systems under investigation.

2DCOS is often applied using one single spectroscopic technique for the X , however, when we consider two different spectroscopic techniques for the above equations (for instance FTIR for X and FTNIR for X^T), 2DCOS is termed two-dimensional hetero-correlation spectroscopy (2DHCOS); it has the peculiar advantage to exploit the complementary information present in each spectroscopic technique [22].

Fig. 2 reports an example of the synchronous and asynchronous FTIR 2DCOS spectra of a polyethylene terephthalate data set of samples coming from marine macrolitter [31]. Table 2 shows the list of the 2DCOS applications in environmental studies reviewed in this paper.

Detailed supports for the interpretation of synchronous, asynchronous and disrelation spectra are available [22,49]. In addition, Pazderka and Kopeck [50] report simple MATLAB routines for the determination of synchronous and asynchronous 2DCOS spectra.

2.3.2. Sample to sample 2DCOS

By means of some modifications of the synchronous and asynchronous equations for the wavelength-wavelength 2DCOS, it is also possible to determine the synchronous and asynchronous 2DCOS spectra in the so called sample to sample (ss) mode [22,51].

With respect to the wavelength-wavelength mode which describes relationships among variables, the sample to sample mode describes linear and nonlinear relationships among samples. The joint examination of spectral data by both wavelength-wavelength and sample to sample modes specific insights into the structure of complex molecular systems [51,52]. For analytical signals aligned as column vectors, Synss spectra are determined according to

$$\text{Synss} = \frac{X^T \times X}{(n-1)} \quad (5)$$

Table 2
Applications of 2DCOS in environmental studies. "ww" and "ss" mean wavelength-wavelength and sample-sample mode respectively. "FS" means fluorescence spectroscopy.

Analytical technique	Sample type	Determined properties	Data treatment	Reference
FTIR	<i>Dunaliella t.</i> , alga	Pollutant effects	ww 2DCOS	[53]
FTIR	Humic acids	Structural characteristics	ww 2DCOS	[54]
F	Humic acids	Structural characteristics	2DHCOS	[55]
TIR, NMR				
UV-VIS, FS	Humic acids	Structural characteristics	ww 2DCOS	[56]
FS	Humic acids	Structural characteristics	ww 2DCOS	[57]
FTIR	Derived organic matter	Sequential order of degradation	ww 2DCOS	[58]
FTIR, UV-vis	Marine mucilage	Structural characteristics	2DCOS, 2DHCOS	[59]
FTIR, FTNIR	Organic matter	Browning development	2DHCOS	[60]
FTIR	Organic matter	Organic matter degradation	2DCOS	[61]
FTIR	Organic matter	Recalcitrant organic matter formation	2DCOS	[62]
FTIR	Clay mineral	Xenobiotic absorption	ww 2DCOS	[63]
FTIR	<i>Caulerpa r.</i> , alga	Phytoremediation ability	ww 2DCOS	[64]
UV-vis	Natural water	Calibration and variable selection	ww 2DCOS	[65]
FTIR	<i>Vicia faba</i> roots	Effects of essential oils	ww 2DCOS, D2DCOS	[68]
GC	Hydrocarbons in sediments	Estimation of similarity and source	ss 2DCOS	[45]
GC	Jet fuel	Estimation of similarity and classification	ss 2DCOS	[70]
GC	Hydrocarbons in Antarctic cores	Estimation of similarity	ss 2DCOS	[71]

where X is the spectral data matrix previously described, X^T is its transposed matrix and n is the number of variables (i.e., wavelengths) in the sample spectra.

Asynss spectra describe the nonlinear (out of phase) relationships among samples of the molecular systems and are determined according to

$$\text{Asynss} = \frac{X^T \times HX}{(n-1)} \quad (6)$$

where X , X^T and n have been previously defined and H is the Noda's modification of the Hilbert transform matrix.

2.3.3. Spectroscopic applications of 2DCOS

In the exposure of the marine microalga *Dunaliella tertiolecta* to several organic and inorganic pollutants, 2DCOS applied to FTIR spectra depicts all the structural changes arising from the interaction between the organism and the pollutants [53]. In addition, the FTIR–2DCOS asynchronous spectra describe the non linear relationships existing among the concentration of the tested pollutants and the inhibition growth effect on the microalga.

Spectroscopic applications of 2DCOS are present in the studies concerning the structural characterization of fulvic and humic acids from aquatic and terrestrial environments. 2DCOS FTIR studies on fulvic and humic acids extracted from marine sediments reveal many interesting characteristics involving the roles played by carbohydrates, proteins and lipids in the mechanisms of organic matter aggregation [54].

Hussain et al. [55] apply 2HDCOS by FTIR and ^{13}C NMR spectroscopy to investigate changes in the compositional and structural characteristics of dissolved organic matter along an estuarine transect submitted to different hydrologic conditions. The 2DCOS maps reveal that each of the three main identified components of organic matter consist of dynamic mixtures of compounds that share similar backbone structures but have significant differences in the functional groups present [55].

It is well known that fulvic and humic acids of the aquatic environment are involved into oxidative degradation reactions by microbial activity and solar irradiation and by binding reactions with metals. Hur et al. [56] investigated the chemical transformation of fulvic and humic acids submitted to UV irradiation by means of 2DCOS applied to UV–vis and synchronous fluorescence spectroscopy. Their results showed that three types of bands changed sequentially in the order of 290–400 nm, 200–250 nm and 250–290 nm ranges respectively; this suggested that aromatic chromophores in aminoacids and tannin-like structures were preferentially oxidized generating non UV-absorbing compounds. The same research team studied the changes of algal derived organic matter in a microbial fuel cell by FTIR spectroscopy supported by 2DCOS [57], determining the sequential order of transformation reactions. This topic, obtained by means of the so-called Noda's rules [22], showed the sequential order proteins → acidic functional groups → polysaccharides → amino acids → oligopeptide/proteins.

Nakashima et al. [58] investigated the binding capacity of humic acids by 2DCOS applied to fluorescence spectroscopy with special emphasis to Ca^{2+} and Pb^{2+} . Asynchronous spectra identified two different binding species in humic acids and the sequential order of the quenching perturbation of the two metals on the structure of humic acids.

Conventional 2DCOS (by FTIR spectroscopy) and 2HDCOS (by FTIR and UV–vis spectroscopy) were jointly applied to the spectra of anomalous size aggregates of marine organic matter called mucilage, giving a valuable support to the comprehension of the mechanisms involved in their formation [59].

2DHCOS by FTIR and FTNIR spectroscopy has been applied to elucidate potential mechanisms involved in the browning observed in aggregates of organic matter formed under anoxic conditions [60]. This study describes the importance of hydrogen bond interactions among carbohydrates and proteins in the texture of the aggregates, texture which also suggests that the color characteristics of the aggregates can be related to the presence of both Maillard and enzymatic browning reactions.

Li et al. [61] apply 2DCOS to the FTIR investigation of the organic matter transformations along the section of core sediments from the Lake Superior (USA). Synchronous spectra reveal that carbohydrates and aliphatic esters are degraded significantly with the increasing core depth causing their enrichment in biogenic silica and inactive polysaccharides. Asynchronous spectra point out that carbonyl groups of aliphatic esters and amide in proteins are degraded faster than those present in carbohydrate and aromatic compounds.

2DCOS FTIR spectroscopic studies are present in other fields of environmental studies too. Harvey et al. [62] investigated the mechanism for the development of surface charge and recalcitrant organic matter in plant derived biochars.

Yan et al. [63] applied 2DCOS-FTIR spectroscopy to study the absorption of the polar xenobiotic enrofloxacin on clay minerals commonly present in soils. The electrostatic interaction between dissolved organic matter and enrofloxacin is the predominant force to describe its presence in sediments and in any case the role played by H-donor-acceptor and π – π interactions is always significant.

In a study concerning the phytoremediation ability of the marine macro-alga *Caulerpa racemosa* for removing hydrocarbons in seawater, 2DCOS FTIR disrelation spectra differentiate the structural changes caused by the three hydrocarbons tested for the phytoremediation tests [64].

2DCOS can support multivariate calibration methods for reducing problems arising from data redundancy and high computation time. Zhou et al. apply 2DCOS as variable selection tool for UV spectroscopic calibration [65]. They select the wavelength ranges where autopeaks of synchronous spectra are placed because in the autopeak ranges the correlation is maximized [22], then they use the wavelength ranges of autopeaks only to perform calibration by PLS.

2DCOS can be also applied as tool for explaining the different performances observed for PCR and PLS in multivariate calibration. In the study for the simultaneous determination of bromide, bicarbonate, sulphate and nitrate in seawater samples by UV spectroscopy, asynchronous spectra show the wavelength range where non linear relationships among absorption and concentrations are present [42]. These non linear relationships explain the best calibration performances of PLS which is able to take into account the non linear relationships among absorptions and concentrations, a property absent in PCR [66].

Recently, Noda [67] has proposed a further development of 2DCOS termed double two-dimensional correlation spectroscopy (D2DCOS). By means of an accurate matrix manipulation it is possible to have new maps of synchronous and asynchronous spectra with a marked improvement of spectral resolution with respect to conventional 2DCOS maps. Noda recommends the application of D2DCOS in the case of hetero two-dimensional correlation spectroscopy, however, D2DCOS showed to be helpful even in omo-2DCOS. In fact, vegetal samples produce complex FTIR spectra due to the simultaneous presence of polysaccharides, proteins and lipids. In the investigation of the effects of some essential oil mixtures on the secondary protein structure of *Vicia faba* roots [68], asynchronous D2DCOS spectra are more resolved than those observed by conventional 2DCOS, allowing to evidence better the specific structural changes caused by each type of

essential oil mixture tested in the study. These results are useful to clarify the possible use of essential oils as alternative to pesticides in agricultural practices.

2.3.4. Application of 2DCOS in chromatographic studies

The chromatograms of hydrocarbons from environmental samples often show the presence of several peaks and due to the lack of specific tools and criteria, the qualitative and quantitative comparison of the hydrocarbon distributions can be hardly performed [69]. 2DCOS in sample–sample mode is a valid support for this topic. Wang et al. describe the so called two-dimensional correlation coefficient mapping (2DMAP) as a powerful tool for jet fuel classification in environmental analysis [70]. Basically, 2DMAP consists of a 2DCOS sample–sample mode analysis of chromatographic samples where a threshold value of correlation coefficients, generally 0.95 is selected. Then for 2DMAP results reported by means of a contour map, samples having correlation higher than the threshold value are visualized whereas samples with correlation lower than the threshold value are hidden and not visualized. According to this simple approach of similarity estimation, 2DMAP becomes an efficient screening method for assessing the similarity and/or dissimilarity among samples. The approach of 2DMAP has been further applied for the comparison of hydrocarbon distributions in marine sediments coming from different areas of the Italian coasts [45]. This study showed that several biogenic, anthropogenic and petrogenic sources can produce high qualitative and quantitative dissimilarities even for samples coming from neighbouring sampling sites. At last, in the study of the hydrocarbon distributions in two Antarctic cores, disrelation 2DCOS evidenced the presence of different anthropogenic, petrogenic and biogenic sources [71].

3. Conclusions

In this paper we reviewed the innovative applications of the ICA and 2DCOS in environmental studies underlying the different targets with respect to those obtained with the conventional PCA and PLS techniques. Their use allows to obtain new relevant information present in spectroscopic and chromatographic data sets belonging to environmental studies.

References

- [1] M. Forina, S. Lanteri, M. Casale, *J. Chromatogr. A* 1158 (2007) 61–93.
- [2] R.G. Brereton, *Analyst* 125 (2000) 2125–2154.
- [3] R.G. Brereton, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, 4th ed., Hollis Horwood Series, Chichester, 2003.
- [4] M. Daszykoeski, S. Serneels, K. Kaczmarek, P. Van Espen, C. Croux, B. Walczak, *Chemom. Intell. Lab. Syst.* 85 (2007) 269–277.
- [5] [www.http://cran.r-project.org/web/packages/chemometrics/vignettes/chemometrics-vignette.pdf](http://cran.r-project.org/web/packages/chemometrics/vignettes/chemometrics-vignette.pdf)
- [6] M. Mecozzi, M. Pietroletti, F. Oteri, R. Di Mento, Applications of PCA to the monitoring of hydrocarbon content in marine sediments by means of gas chromatographic measurements, in: S. Parinya (Ed.), *Principal Component Analysis – Engineering Applications*, Intech Edition, 2012. www.intechopen.com.
- [7] Z.-Y. Zhao, Y. Chu, J.-D. Gu, *Ecotoxicology* 21 (2012) 1743–1752.
- [8] A. Ausili, M. Mecozzi, M. Gabellini, G. Ciuffa, F. Mellara, *Water Sci. Technol.* 37 (1998) 131–138.
- [9] M.E. Conti, B. Bocca, M. Iacobucci, M.G. Finoia, M. Mecozzi, A. Pino, A. Alimonti, *Arch. Environ. Contam. Toxicol.* 58 (2010) 79–95.
- [10] M.E. Conti, J. Stripeikis, M.G. Finoia, M.B. Tudino, *Ecotoxicology* 21 (2012) 1112–1125.
- [11] M.E. Conti, M.G. Finoia, B. Bocca, G. Mele, A. Alimonti, A. Pino, *Environ. Monit. Assess.* 184 (2012) 527–538.
- [12] M.E. Conti, J. Stripeikis, M.G. Finoia, M.B. Tudino, *Ecotoxicology* 20 (2011) 1341–1353.
- [13] M.E. Conti, A. Pino, F. Bottrè, B. Bocca, A. Alimonti, *Ecotox. Environ. Saf.* 72 (2009) 1082–1089.
- [14] M.E. Conti, M. Iacobucci, D. Cucina, M. Mecozzi, *Int. J. Environ. Pollut.* 29 (2007) 333–343.
- [15] E.S. Karlsson, V. Brüchert, T. Tesi, A. Charkin, O. Dudarev, I. Semiletov, Ö. Gustafsson, *Mar. Chem.* 170 (2015) 11–22.
- [16] H. Wagner, S. Dunker, Z. Liu, C. Wilhelm, *Curr. Opin. Biotechnol.* 24 (2013) 88–94.
- [17] J. Moros, R.J. Cassella, M.C. Barciela-Alonso, A. Moreda-Piñeiro, P. Herbello-Hermelo, P. Bermejo-Barrera, S. Garrigues, M. de la Guardia, *Vib. Spectrosc.* 53 (2010) 204–213.
- [18] S. Armenta, M. De la Guardia, *Trends Environ. Anal. Chem.* 2 (2014) 43–52.
- [19] M. Mecozzi, *Chemom. Intell. Lab. Syst.* 79 (2005) 84–90.
- [20] J.L. Beltrán, R. Ferrer, J. Guiteras, *Anal. Chim. Acta* 373 (1998) 311–319.
- [21] D.N. Rutledge, J.-R. Bouveresse, *Trends Anal. Chem.* 50 (2013) 22–32.
- [22] I. Noda, Y. Ozaki, *Two-Dimensional Correlation Spectroscopy, Application in Vibrational and Optical Spectroscopy*, John Wiley & Sons, Chichester, UK, 2004.
- [23] J.H. Christensen, G. Tomasi, *J. Chromatogr. A* 1169 (2007) 1–18.
- [24] I. Noda, *J. Mol. Struct.* 883–884 (2008) 216–227.
- [25] J. Engel, J. Gerretzen, E. Szymanska, J.J. Jansen, G. Downey, L. Blanchet, L.M.C. Buydens, *Trends Anal. Chem.* 50 (2013) 96–106.
- [26] A. Hyvärinen, E. Oja, *Neural Network* 13 (2000) 411–430.
- [27] G. Wang, Q. Ding, Z. Hou, *Trends Anal. Chem.* 27 (2008) 368–376.
- [28] <http://research.ics.aalto.fi/jica/fastica/>.
- [29] <http://perso.telecom-paristech.fr/~cardoso/Algo/Jade/jadeR.m>.
- [30] www.ucl.ac.uk/departments/sobell/Research/RLemon/MILCA/MILCA for MILCA.
- [31] M. Mecozzi, submitted for publication.
- [32] F. Szabo de Edeleny, A.W. Simonetti, G. Postma, R. Huo, L.M.C. Buydens, *Anal. Chim. Acta* 544 (2005) 36–46.
- [33] D. Jouan-Rimbaud Bouveresse, A. Moya-González, F. Ammari, D.N. Rutledge, *Chem. Int. Lab. Syst.* 112 (2012) 24–32.
- [34] Y.B. Monakhova, S.A. Astakhov, A. Kraskov, S.P. Mustakhova, *Chemom. Intell. Lab. Syst.* 103 (2010) 108–115.
- [35] F. Ammari, C.B. Cordella, N. Bougham, D.N. Rutledge, *Chemom. Intell. Lab. Syst.* 113 (2012) 32–42.
- [36] N. Pasadakis, A.A. Kardamakis, *Anal. Chim. Acta* 578 (2006) 250–255.
- [37] A.A. Kardamakis, A. Mouchtaris, N. Pasadakis, *Chemom. Intell. Lab. Syst.* 89 (2007) 51–58.
- [38] M. Mecozzi, M. Pietroletti, M. Scarpiniti, R. Acquistucci, M.E. Conti, *Environ. Monit. Assess.* 184 (2012) 6025–6036.
- [39] N. Calace, M. Pietroletti, B.M. Petronio, M. Scarpiniti, M. Mecozzi, *Int. J. Environ. Health* 5 (2011) 246–261.
- [40] Y.B. Monakhova, A.M. Tsikin, S.P. Mushtakova, M. Mecozzi, *Microchem. J.* 118 (2015) 211–222.
- [41] M. Mecozzi, M. Pietroletti, A. Tornambè, *Spectrochim. Acta Part A* 78 (2011) 1572–1580.
- [42] M. Mecozzi, Y.B. Monakhova, *Int. J. Environ. Health* 6 (2013) 235–251.
- [43] N. Pasadakis, V. Gaganis, P. Smaragdis, www.merl.com/publications/docs/TR2003-140.pdf.
- [44] Y.B. Monakhova, S. Kolesnikova, S.P. Mushtakova, *Anal. Methods* 5 (2013) 2761–2772.
- [45] L. Gao, S. Ren, *J. Fluoresc.* 22 (2012) 1595–1602.
- [46] M. Pietroletti, S. Mattiello, F. Moscato, F. Oteri, M. Mecozzi, *Chromatographia* 75 (2012) 961–971.
- [47] Z. Liu, W. Cai, X. Shao, *J. Chromatogr. A* 1190 (2008) 358–364.
- [48] B. Debrus, P. Lebrun, A. Ceccato, G. Caliaro, B. Govaerts, B.A. Olsen, E. Rozet, B. Boulanger, P. Hubert, *Talanta* 79 (2009) 77–85.
- [49] P. de, B. Harrington, A. Urbas, P.J. Tandler, *Chem. Intell. Lab. Syst.* 50 (2000) 149–174.
- [50] T. Pazderka, V. Kopecký, http://dsp.vsch.cz/konference_matlab/MATLAB08/prispevky/081_pazderka.pdf
- [51] S. Šašić, T. Amari, Y. Ozaki, *Anal. Chem.* 73 (2001) 5184–5190.
- [52] S. Šašić, J. Jiang, Y. Ozaki, *Chem. Intell. Lab. Syst.* 65 (2003) 1–15.
- [53] M. Mecozzi, M. Pietroletti, R. Di Mento, *Vib. Spectrosc.* 44 (2007) 228–235.
- [54] M. Mecozzi, E. Pietrantonio, M. Pietroletti, *Spectrochim. Acta Part A* 71 (2009) 1874–1877.
- [55] A. Hussain, N. Abdulla, E.C. Minor, P.G. Hatcher, *Environ. Sci. Technol.* 44 (2010) 8044–8049.
- [56] J. Hur, K. Young Jung, Y. Mee Young, *Water Res.* 45 (2011) 2965–2974.
- [57] J. Hur, B.M. Lee, K.S. Choi, B. Min, *Environ. Sci. Pollut. Res.* 21 (2014) 2230–2239.
- [58] K. Nakashima, S. Xing, Y. Gong, T. Miyajima, *J. Mol. Struct.* 883–884 (2008) 155–159.
- [59] M. Mecozzi, M. Pietroletti, V. Gallo, M.E. Conti, *Mar. Chem.* 116 (2009) 18–35.
- [60] M. Mecozzi, R. Acquistucci, L. Nisini, M.E. Conti, *Infrared Phys. Technol.* 63 (2014) 74–83.
- [61] H. Li, E.C. Minor, P.K. Zigah, *Org. Geochem.* 58 (2013) 125–136.
- [62] O.R. Harvey, B.E. Herbert, L. Kuo, P. Louchouart, *Environ. Sci. Technol.* 46 (2012) 10641–10650.
- [63] W. Yan, J. Zhang, C. Jing, *J. Colloid Interface Sci.* 390 (2013) 196–209.
- [64] M. Pietroletti, A. Capobianco, E. Ragosta, M. Mecozzi, *Spectrochim. Acta Part A* 77 (2010) 673–679.
- [65] X. Zhou, B. Xiang, M. Zhang, *Anal. Lett.* 46 (2013) 340–348.
- [66] I.S. Helland, *Chemom. Intell. Lab. Syst.* 58 (2001) 97–107.
- [67] I. Noda, *J. Mol. Struct.* 974 (2010) 108–115.
- [68] M. Mecozzi, E. Sturchio, *Spectrochim. Acta Part A* 137 (2015) 90–98.
- [69] G.S. Fraser, J. Ellis, L. Hussain, *Mar. Poll. Bull.* 56 (2008) 9–13.
- [70] G. Wang, J. Karnes, C.E. Bunker, M. Lei Geng, *J. Mol. Struct.* 799 (2006) 247–252.
- [71] M. Mecozzi, M. Pietroletti, G. Trifirò, L. Nisini, *Mar. Chem.* 159 (2014) 9–18.
- [72] Z. Chen, J.N. Hay, M.J. Jenkins, *Eur. Polym. J.* 48 (2012) 1586–1610.