Application of multivariate statistics in assessment of green analytical chemistry parameters of analytical methodologies

ARTICLE *in* GREEN CHEMISTRY · MAY 2013

Impact Factor: 8.02 · DOI: 10.1039/C3GC36976E

4 AUTHORS:

Marek Tobiszewski
Gdansk University of Technology
28 PUBLICATIONS **540** CITATIONS
SEE PROFILE

Stefan Leonidov Tsakovski
Sofia University "St. Kliment Ohridski"
75 PUBLICATIONS **851** CITATIONS
SEE PROFILE

Vasil Simeonov
Sofia University "St. Kliment Ohridski"
180 PUBLICATIONS **2,122** CITATIONS
SEE PROFILE

Jacek Namieśnik
Gdansk University of Technology
618 PUBLICATIONS **8,358** CITATIONS
SEE PROFILE

Green Chemistry

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Cite this: *Green Chem.*, 2013, **15**, 1615

Application of multivariate statistics in assessment of green analytical chemistry parameters of analytical methodologies

Marek Tobiszewski,*[a] Stefan Tsakovski,[b] Vasil Simeonov[b] and Jacek Namieśnik[a]

Received 7th December 2012,
Accepted 2nd April 2013

DOI: 10.1039/c3gc36976e

www.rsc.org/greenchem

The study offers a multivariate statistical analysis of a dataset, including the major metrological, "greenness" and methodological parameters of 43 analytical methodologies applied for aldrin determination (a frequently analyzed organic compound) in water samples. The variables (parameters) chosen were as follows: metrological (LOD, recovery, RSD), describing the "greenness" (amount of the solvent used, amount of waste generated) and general methodological parameters (sample volume, time of analysis, injection volume) and scores of greenness assessment with NEMI and eco-scale. The results of the study show that all analytical methodologies have been grouped into three clusters. The first one consisted of "non-green" LLE and SPE methodologies and the other two consisted of solventless or virtually solventless methodologies. The NEMI and eco-scale scores are well correlated, which indicates the similarity between these two assessment scales. A self-organizing maps technique is not feasible for easy and quick labeling of analytical methodologies in terms of their greenness. However, the multivariate analysis of analytical methodologies can give information about clustering of methodologies to "green" or "non-green" groups and some extra information about relations between objects inside clusters of interest.

1. Introduction

Green analytical chemistry is becoming a more widely recognized and established idea throughout the analytical laboratories worldwide. The increase in analyst's attention to this topic can be seen by the rapidly growing number of publications dealing with green analytical chemistry.[1] The idea of green analytical chemistry can be brought into analytical practice – in research studies but more importantly in routine analyses in several ways: by introduction of sample preparation methods that consume less materials, especially organic solvents,[2,3] *e.g.* by introduction of solventless, microextraction techniques. Sample preparation steps might be omitted by introduction of direct analytical techniques.[4] Some other approaches to green analytical chemistry involve dealing with analytical waste already produced,[5] improving efficiency of analytes chromatographic separation and reducing analysis time.[6]

One of the most important issues with green analytical chemistry is the lack of tools allowing for assessment of the "greenness" of each analytical method. So far a few methods have been proposed for analytical methodologies environmental impact assessment, each with its advantages and disadvantages. One of the oldest and well established methods is using National Environmental Methods Index (NEMI) greenness profile symbols.[7] Greenness profile is divided into four fields, each related to different aspects of methodologies environmental impact. If the condition is fulfilled the appropriate field is colored green. The condition is fulfilled when none of the chemicals used during the procedure is listed as persistent, accumulative and toxic, listed on the hazardous chemicals list when the sample pH during analysis is within the range 2–12 and when the amount of waste generated during analysis is less than 50 g. The main advantage of this method of evaluation is simplicity of both creating and reading methodology greenness profile. The second proposed method for evaluating methodologies greenness is application of an analytical eco-scale,[8] which is based on the eco-scale used for assessing environmental impact of organic syntheses. The result of analysis with an eco-scale is a number lower than 100. Penalty points are given for the amount and type of reagents, energy consumption, analyst occupational hazard, solid waste generated and the way they are treated (or not) and are subtracted from the initial value of 100. This method requires more time and effort to assess the analytical procedure, however, it takes more parameters into consideration.

[a]*Department of Analytical Chemistry, Chemical Faculty, Gdańsk University of Technology (GUT), 11/12 G. Narutowicza St., 80-233 Gdańsk, Poland. E-mail: marektobiszewski@wp.pl*
[b]*Chair of Analytical Chemistry, Faculty of Pharmacy and Chemistry, University of Sofia "St. Kl. Okhridski", J. Bourchier Blvd. 1, 1164 Sofia, Bulgaria*

Table 1 Variables being input to statistical data analysis

Variable	Abbreviation	Units	Remarks
Limit of detection	LOD	ng L^{-1}	—
Relative standard deviation	RSD	%	If RSD was calculated for several concentration levels, the RSD for the lowest one was chosen
Recovery	REC	%	If recovery was calculated for more than one sample matrix, the cleanest was chosen: tap water > groundwater > surface water
Amount of organic solvent	AOS	mL	The total amount of all organic solvents used in the analytical protocol
Amount of organic solvent × hazard	AOSH	—	Amount of each solvent used is multiplied by the number of warning (×1) or danger (×2) pictograms. Similar approach to eco-scale
Volume of sample	VS	mL	—
Injection volume	INJV	μL	The volume of liquid injected to the injector. If analytes were thermally desorbed the injection volume was assumed to be 0.01 μL, because value "0" cannot be handled with most multivariate techniques algorithms. The value of 0.01 is negligible compared to other analytical methodologies, involving injection of certain sample volume
Number of other analytes determined	OANAL	—	Number of other analytes determined in a single analytical run. 'Per analyte' concept assumes that all material and energetic inputs can be divided by the number of determined analytes in a protocol
Time	Time	h	Total time of analysis (estimated sample preparation time + chromatographic run)
Amount of solid waste	SW	g	The total amount of all waste generated during analysis with the analytical protocol
NEMI score	NEMI	Unitless (range 0–4)	NEMI score was calculated for each analytical methodology, unless it was accessible
Eco-scale score	ECOSC	Unitless (range 0–100)	Eco-scale score was calculated for each analytical methodology

A self-organizing map is a multivariate statistical technique allowing for classification of objects or variables, with very intuitive visualization of the dataset. It has been used in many very different fields, *e.g.* environmental analysis,[9] predicting consumers preferences,[10] environmental sustainability,[11] visualization of financial trajectories,[12] exploration of psychological datasets,[13] *etc.* Self-organizing maps are very flexible tools as they allow for visualization of huge datasets with linear and non-linear distributions of data.

The aim of this paper is to investigate, with the aim of multivariate statistics tools, the relationship between factors determining the greenness of an analytical procedure. The second aim is to assess the eco-scale and NEMI applicability as methods for the assessment of analytical methodology "greenness". Finally, the self-organizing maps technique will be discussed as a greenness assessment method itself. To our knowledge, it is the first approach to apply multivariate statistics as a tool for assessment of analytical methodologies in the context of green chemistry.

2. Methods

2.1 Data preparation

To obtain the possibility of comparison of the different analytical methodologies, all the methodologies included in the multivariate analysis should consider a single analyte in one and the same sample matrix. All the data collected describe analytical methodologies for the determination of aldrin in water samples. Aldrin was chosen as an example of a frequently determined organic compound in water samples. The dataset consists of 12 variables that describe 43 analytical methodologies. The data were extracted from the methodologies published in journal articles and from standard procedures. The methodologies were found in NEMI database for standard procedures and in RSC, ACS, Springerlink and ScienceDirect databases for publication years 1999–2012, with most of the papers published during the last five years. The variables are described in detail in Table 1 and the abbreviation for each methodology is presented in Table 2.

2.2 Cluster analysis

Cluster analysis is a well established technique allowing us to group objects or variables according to their similarity. The similarity is expressed as a distance in *n*-dimensional space.[57] The cluster analysis was performed using the Statistica 7.0 software package.

2.3 Self-organizing maps

Self-organizing maps (SOM) technique is a multivariate statistics tool based on neural network theory described by Kohonen.[58] It is usually applied for visualization of large high-dimensional sets of data. SOM algorithm is a non-linear projection of the dataset from multi-dimensional space into a two-dimensional array of neurons, also called nodes. The algorithm involves the projection as each node in the grid is initialized as a random unit vector in multidimensional space. Each object from the data set is considered as an *n*-dimensional input vector. The "winner-takes-all" rule involves the selection of the node whose vector most closely matches the input data. This winning vector modifies its weights to better match the input data. Vectors ("neighbors") that match the nodes surrounding the winning node are also modified. In

Table 2 Analytical methodologies assessed during the study

No.	Analytical methodology	Analytical methodology abbreviation	Reference
1	Liquid–liquid extraction–gas chromatography–electron capture detection, Method 6630	6630 LLE–GC–ECD	14
2	Liquid–liquid extraction–gas chromatography–electron capture detection, Method 505	505 LLE–GC–ECD	15
3	Liquid–liquid extraction–gas chromatography–electron capture detection, Method 508	508 LLE–GC–ECD	16
4	Liquid–liquid extraction–gas chromatography–electron capture detection, Method 1104	O-1104 LLE–GC–ECD	17
5	Liquid–liquid extraction–gas chromatography–mass spectrometry Method 625	625 LLE–GC–MS	18
6	Membrane assisted solvent extraction–large volume injection–gas chromatography–mass spectrometry	MASE–LVI–GC–MS	19
7	Microwave assisted extraction–headspace solid phase microextraction–gas chromatography–electron capture detection	MAE–HS–SPME–GC–ECD	20
8	Headspace solid phase microextraction–gas chromatography–tandem mass spectrometry	HS–SPME–GC–MS–MS	21
9	Solid phase microextraction–high performance liquid chromatography–detection in ultraviolet	SPME–HPLC–UV	22
10	Solid phase microextraction–gas chromatography–tandem mass spectrometry	SPME–GC–MS–MS	23
11	Solid phase microextraction–gas chromatography–electron capture detection	SPME–GC–ECD	24
12	Solid phase microextraction–gas chromatography–mass spectrometry	SPME–GC–MS	25
13	Stir bar sorbent extraction–thermal desorption–two dimensional gas chromatography–mass spectrometry	SBSE–TD–GC–GC–MS	26
14	Stir bar sorbent extraction–gas chromatography–mass spectrometry	SBSE–GC–MS	27
15	Stir bar sorbent extraction–two dimensional gas chromatography–time-of-flight mass spectrometry	SBSE–GC–GC–TOF–MS	28
16	Stir bar sorbent extraction–thermal desorption–gas chromatography–tandem mass spectrometry	SBSE–TD–GC–MS–MS	29
17	Stir bar sorbent extraction–thermal desorption–gas chromatography–mass spectrometry	SBSE–TD–GC–MS	30
18	Dispersive liquid–liquid microextraction–gas chromatography–electron capture detection	DLLME–GC–ECD	31
19	Dispersive liquid–liquid microextraction–gas chromatography–mass spectrometry	DLLME–GC–MS	32
20	Dispersive liquid–liquid microextraction–gas chromatography–mass spectrometry	DLLME–GC–MS	33
21	Single drop microextraction–gas chromatography–mass spectrometry	SDME–GC–MS	34
22	Improved homogeneous liquid–liquid extraction–gas chromatography–electron capture detection	IHLLE–GC–ECD	35
23	Solvent cooling–assisted dynamic hollow-fibre-supported headspace liquid-phase microextraction–gas chromatography–electron capture detection	SC–DHF–HS–LPME–GC–ECD	36
24	Liquid-phase microextraction–gas chromatography–electron capture detection	LPME–GC–ECD	37
25	Hollow fibre–liquid-phase microextraction–gas chromatography–tandem mass spectrometry	HF–LPME–GC–MS–MS	38
26	Liquid-phase microextraction–gas chromatography–electron capture detection	LPME–GC–ECD	39
27	Dynamic headspace time–extended helix liquid-phase microextraction–gas chromatography–tandem mass spectrometry	DHS–TEH–LPME–GC–MS–MS	40
28	Micro-liquid–liquid extraction–large volume on-column injection–gas chromatography–electron capture detection	MLLE–LVOCI–GC–ECD	41
29	Micro-liquid–liquid extraction–large volume injection–negative ion chemical ionization–mass spectrometry	MLLE–LVI–NICI–MS	42
30	Solid phase extraction–liquid chromatography–mass spectrometry	SPE–LC–MS	43
31	Solid phase extraction–liquid chromatography–Mass spectrometry	SPE–LC–MS	44
32	Solid phase extraction–two dimensional gas chromatography–time-of-flight mass spectrometry	SPE–GC–GC–TOF–MS	45
33	Solid phase extraction–gas chromatography–mass spectrometry	SPE–GC–MS	46
34	Solid phase extraction–gas chromatography–mass spectrometry	SPE–GC–MS	47
35	Solid phase extraction–gas chromatography–mass spectrometry, method 525.2	525.2 SPE–GC–MS	48
36	Solid phase extraction–gas chromatography–electron capture detection, Method 508.1	508.1 SPE–GC–ECD	49
37	Solid phase extraction–gas chromatography–mass spectrometry	SPE–GC–MS	50
38	Solid phase extraction–gas chromatography–electron capture detection	SPE–GC–ECD	51
39	Solid phase extraction–gas chromatography–microwave-induced plasma–atomic emission spectroscopy	SPE–GC–MIP–AES	52
40	On-line solid phase extraction–gas chromatography–mass spectrometry	On-line–SPE–GC–MS	53
41	On-line solid phase extraction–gas chromatography–electron capture detection	On-line–SPE–GC–ECD	54
42	On-line solid phase extraction–programmed temperature vaporization–gas chromatography–mass spectrometry	On-line–SPE–PTV–GC–MS	55
43	On-line solid phase extraction–large volume injection–gas chromatography–mass spectrometry	On-line–SPE–LVI–GC–MS	56

such a way the self-organizing maps algorithm recognizes vectors similar to themselves. The final result of the analysis is two dimensional which preserves the topology of initial data set in a non-parametric way without any external information.

The trained map is usually graphically presented on two-dimensional planes for each variable, indicating a variable distribution pattern on the different regions of the map with different colors. On the right side of each variable plane there is a color bar scale that indicates the variable value. The SOM classification gives as a result also a U matrix plane. It visualizes distances between the nodes and helps to identify cluster structure of the map. The high values in the U-matrix plane indicate the places of clusters' borders while the areas with low values indicate clusters themselves. Matlab 6.5 software was used for calculations. All calculations concerning SOM classification were performed by a free toolbox (SOM Toolbox 2.0), which can be downloaded from http://www.cis.hut.fi/projects/somtoolbox/.

3. Results and discussion

The main goal of the chemometric data treatment was:

1. to find groups of similarity between the parameters chosen;

2. to find groups of similarity between the objects of the study (analytical methods used);

3. to clarify the data set structure.

The interpretation of the data could give some answers to the questions stated in the aims of the paper (the major issue being which is the "greenest" method). Before answering definitively to this difficult question some conclusions from the multivariate statistical approach will be presented.

3.1 Cluster analysis

The first step of the analysis was application of a relatively simple chemometric tool, like cluster analysis. The aim of such an approach was to identify the data structure.

First, the clustering of variables was performed, as shown in Fig. 1. In principle, two major clusters are formed (level of significance 66.7% D_{max}). As seen, one of them includes the "theoretical concepts" of greenness ECOSC and NEMI along with the methodological variables injection volume, time and recovery. The second one involves the specific analytical parameters related to the sample preparation. The clustering could be done also with a more strict significance test (33.3% D_{max}) and, thus, the major clusters could be subdivided:

- K1 (eco-scale score, NEMI score, injection volume)
- K2 (time, recovery)
- K3 (amount of solvent, amount of solvent × hazard)
- K4 (volume of the liquid sample, amount of solid waste, other analytes per analysis, RSD, LOD)

The formation of the "*green estimator*" *pattern* (K1; injection volume is not so strongly related to the other two parameters) is an indication that, indeed, these eco-scale and NEMI concepts are reliable because the results of assessments are similar.

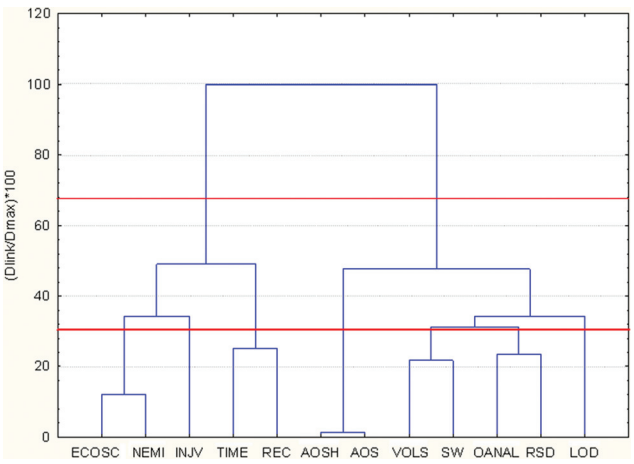

Fig. 1 Hierarchical dendrogram of the variables (parameters). The abbreviations are explained in Table 1.

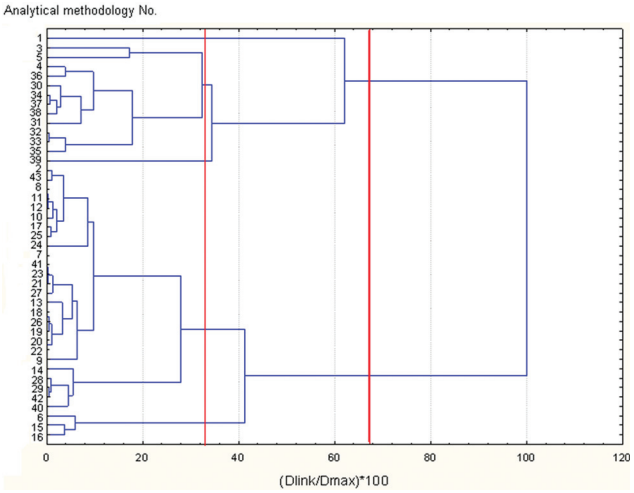

Fig. 2 Hierarchical dendrogram for all 43 objects. Number of methodologies like in Table 2.

K2 is a typical "*method-oriented*" cluster and its close linkage to the green estimators (see clustering at significance level 66.7% D_{max}) indicates probably that the green estimators assess dominantly methods applied.

K3 is related to the use of the amount of solvents and along with K4 form the sample preparation pattern. K4 is formed by correlated "volume of the liquid sample" with the "amount of solid waste". The second subcluster is formed by "precision of method" and "the amount of analytes determined in single run". LOD is less correlated with other parameters in cluster K4.

Next step was to link the methodologies used, which seems to be a more important issue. In Fig. 2 the hierarchical dendrogram of the objects (methods) is presented. Again, two major clusters have been found at significance level 66.7% D_{max} and four clusters are identified at significance level 33.3% D_{max}.

- K1 (15, 16, 6) formed by SBSE extraction methods and MESI with GC separation.
- K2 (40, 42, 29, 28, 14, 9, 22, 20, 19, 26, 18, 13, 27, 23, 21, 41, 7, 24, 25, 17, 10, 12, 11, 8, 43, 2) formed by methodologies based on solventless extraction and SPE in *on-line* mode techniques.
- K3 (39, 35, 33, 32, 31, 38, 37, 34, 30, 36, 4, 5, 3) formed by SPE methodologies together with three liquid–liquid extraction methodologies. SPE methodologies apply organic solvents and require additional material inputs (cartridges). There is a statistically important difference between SPE methodologies and "solventless" ones.
- K4 (1) an outlier in the clustering (LLE methodology).

Solid phase extraction is sometimes considered to be a green sample preparation technique.[3,7,59] However SPE needs organic solvents and generates solid waste. The results of classification show that development of microextraction techniques has put SPE techniques among non-green sample preparation choices. SPE should not be considered or chosen as a green sample pretreatment technique.

The great amount of analyses for determination of aldrin is routinely performed in many laboratories. The laboratories that perform routine analyses are considered to generate high negative environmental impact.[4] All recommended standard procedures (except Method 505 which uses a relatively small volume of solvent), which were included in this study, are categorized as non-green. These results suggest that the list of standard analytical procedures requires revision.

It is of substantial interest to find the discriminating parameters for each one of the clusters formed. It is obvious that the clustering of the objects (methods) is not directly related to the analytical technique used – the same techniques are found in different clusters. That is why specific discriminating parameters for each one of the clusters were determined.

In Fig. 3 the averages for each parameter and for each cluster are presented. It could be concluded that K1 is characterized by the highest values of the NEMI score, time, injection volume and recovery, which is an indication for a relatively high level of "greenness" but prolonged time of preparation, high injection volume and good recovery. The methods included in K1 can be considered "green".

The methodologies grouped into K2 have the highest level of the eco-scale score and lowest levels of the amount of solvent, amount of sample × hazard, solid waste, volume of the sample, other analytes per analysis and time of analysis. It means that the methods in K2 are "green" with very good analytical indicators – low amount of organic solvents used, low amount of generated solid waste and sample amount and quick in performance. K1 and K2 form a large group of "green" methods.

The other two clusters (actually K4 is an outlier, not a cluster) are characterized with lower levels of "greenness". K3 shows a maximal level of LOD which is usually not acceptable as an analytical parameter and lowest levels in NEMI and the injection volume (INJV) is very low. High LOD can be partially explained with low injection volumes. K4 is an unfavorable method since it has no acceptable methodical (analytical) parameters (maximal levels of RSD, AOS, AOSH, SW and VOLS) parallel to minimal levels of NEMI and ECOS although with low limits of detection.

Comparison of variables mean values for "green" clusters (K1 and K2) and "non-green" clusters can give some information. The amounts of organic solvents used, generated solid waste and initial volumes of the sample are considerably higher for methodologies clustered as "non green" compared to "green" ones. This is reflected by both NEMI and eco-scale scores as they are higher for methodologies ranked as "green". The differences in values of RSD, LOD and recoveries are not so well visible but it can be stated that green methodologies have slightly better metrological parameters. It can be stated that the main difference in analytical methodologies analyzed during a cluster analysis study is environmental impact, not the analytical performance.

3.2 Self-organizing maps analysis

Compared to cluster analysis, SOM has some advantages. It allows us to easily see if variables are positively or negatively correlated. In the second case, the color distribution on the planes of such variables is opposite. It also carries semiquantitative information about the distribution of a given variable through the dataset.[60]

The dataset was treated by the SOM method keeping in mind the same goals: classification of the methods (objects) based on their "greenness" and detection of relationships between the variables supposed to determine "greenness" quality. Fig. 4 shows the results of SOM analysis for the whole dataset.

The self-organizing maps readily indicate some patterns of similarity between parameters, *e.g.* ECOS and NEMI, AOS and AOSH, *etc.* A better separation between groups of similar variables is shown in Fig. 5 where the variables are grouped according to their similarity. Some of the results obtained by SOM analysis could seem obvious. Experimental confirmation

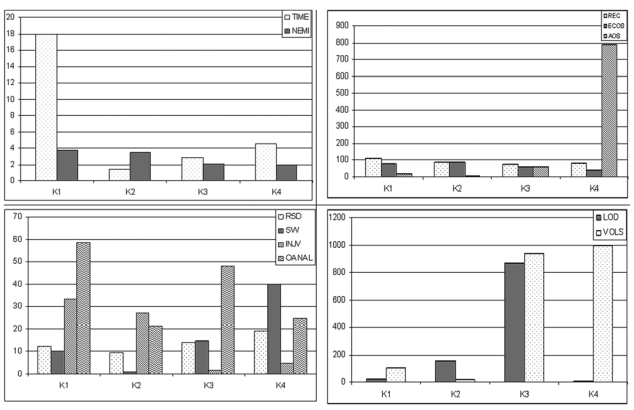

Fig. 3 Average values of each variable for each cluster of objects (methodologies).

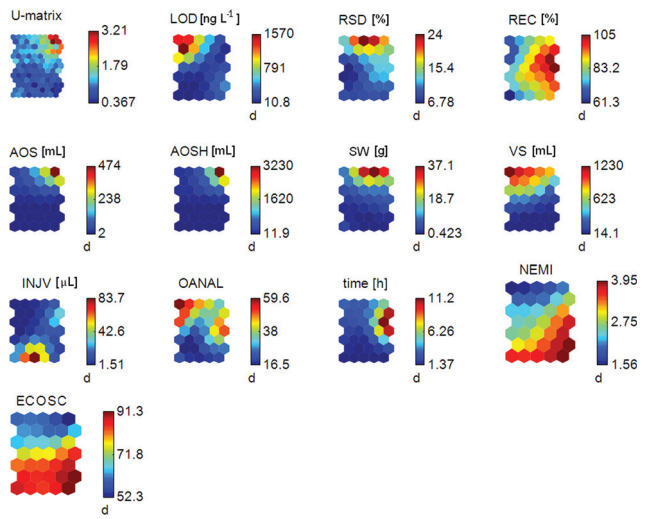

Fig. 4 SOM patterns of the data.

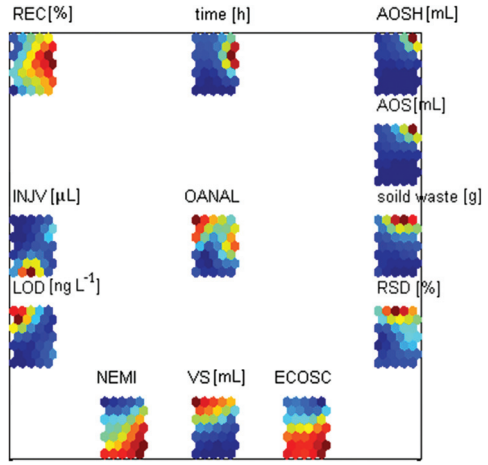

Fig. 5 Grouping of variables according to their similarity.

of such classification results can be treated as validation of the SOM technique.

Four groups of correlated variables are formed. The first group consists of the NEMI score factor, the eco-scale factor and negatively correlated with them sample volume. High correlation of NEMI and eco-scale has been observed during cluster analysis and means that similar results are obtained with both assessment techniques. Negative correlation of initial sample volume means that good eco-scale and NEMI scores are obtained if samples of low volume are analyzed.

The second group was formed by the positively correlated amount of organic solvents, the amount of solvents multiplied by its hazard, the mass of solid waste and RSD. It can be stated that methodologies with a high solvent input have also high solid materials input. The presence of low precision (high RSD) in this group seems to be also logical. The methodologies with more operations and higher material and solvent inputs are characterized by lower precision. Every sample treatment step could be a source of error which makes the precision of methodologies involving many preparation steps low.

The third group was formed by negatively correlated sample injection volume and limit of detection. This relation is very logic as large volume injection techniques can give much lower limits of detection. It is worth noting that the LOD variable is also negatively correlated with NEMI and eco-scale factors. Methodologies ranked as "green" have also low limits of detection.

The fourth group consists of weakly positively correlated recovery of aldrin and time sample of analysis. This weak correlation may suggest that higher recovery can be obtained with prolonged sample treatment. These two variables are correlated neither with variables describing material inputs nor with two "greenness" assessment methods.

In this classification the "number of other analytes determined" variable takes a specific position not being closely linked to any other variable. However, it can be noted that this variable is weakly correlated (similar patterns in color distribution) to time of analysis, LOD and sample volume. Therefore it may be stated that multi-residue methodologies require large initial sample volume with long time for preparation and these methodologies allow for obtaining relatively high limits of detection.

SOM clustering of objects confirms almost completely the clustering obtained by cluster analysis – four major clusters, one of them consisting of one outlier (method 1) and the other three with the same composition as indicated above. Therefore, SOM clustering of objects will not be presented.

It seems reasonable to reduce the number of the input variables and to select less but more informative number of them. This pre-processing step is usually applied in ranking procedures. Next step in the multivariate statistical analysis aiming at fine ranking of the analytical methods for determination of aldrin with respect to their "greenness" was to perform SOM analysis by the use of only pre-selected variables for the group of methods having been ranked "green" from cluster analysis (respective numbers: 15, 16, 6, 40, 42, 29, 28, 14, 9, 22, 20, 19, 26, 18, 13, 27, 23, 21, 41, 7, 24, 25, 17, 10, 12, 11, 8, 43, 2). The pre-selected variables were AOS, SW, VS, INJV, OANAL, NEMI, ECOS (according to the components plane). The aim of this step analysis is identification of the greenest methodologies from the set identified in the data pre-treatment step as "green".

In Fig. 6 and 7 the new SOMs and components planes are shown. As seen in Fig. 7 four types of parameters are responsible for the determination of methodologies ranking – "chemical impact" (OANAL and AOS); "objective assessment" (NEMI and ECOSC); "sample size impact" (SW and VS) and "method impact" (INJV).

As seen in Fig. 8 two major clusters are formed during the classification of objects. From Fig. 6 it can be seen that objects (methodologies) grouped into the upper cluster are characterized by higher amounts of solvents, solid waste and initial volume of the sample, as they allow for determination of larger

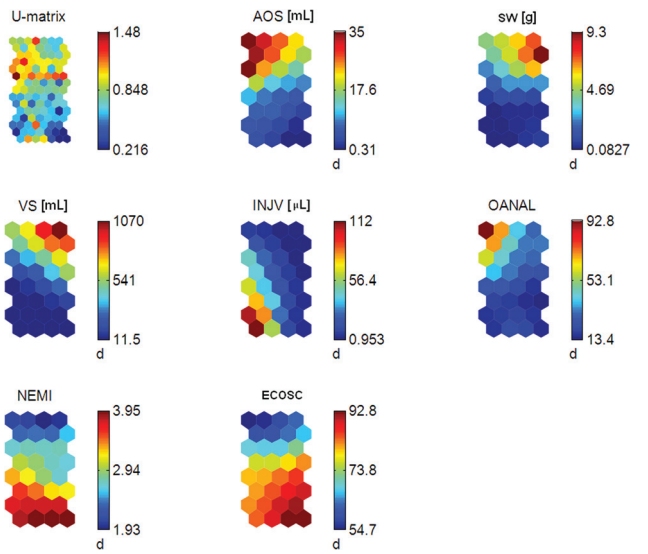

Fig. 6 SOMs for pre-selected variables.

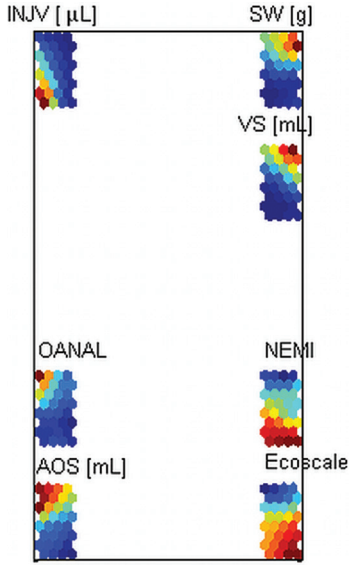

Fig. 7 Component planes for the pre-selected variables.

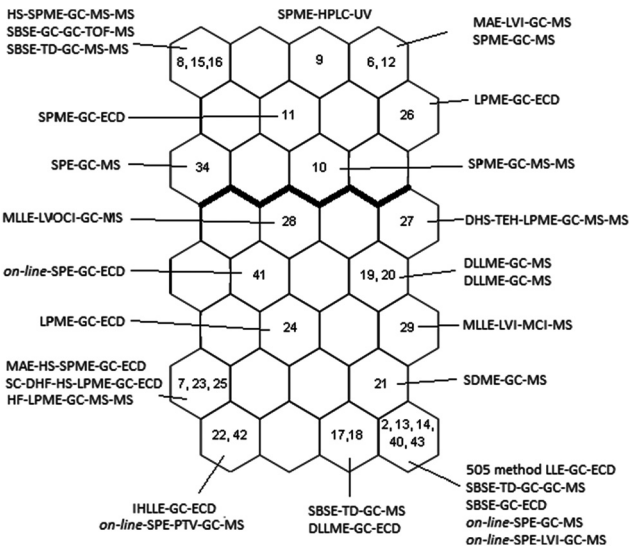

Fig. 8 Hit diagram of pre-selected "green" analytical protocols. Thick line shows the border between the two clusters.

amounts of analytes. Methodologies grouped into the lower cluster are characterized by high NEMI and eco-scale scores and high injection volume. In Fig. 8 the hit diagram is shown allowing identification of the clusters content.

The grouping of objects is not very clear (according to the technique involved), however some trends can be observed. The lower cluster mainly consists of DLLME methodologies, together with *on-line* SPE protocols. LPME and SBSE methodologies are shared among the two clusters. The upper cluster, ranked as "less green", mainly consists of SPME methodologies, which have zero consumption of solvents and waste generation, together with methodologies that clearly have negative environmental impact. It seems that the number of other analytes determined during chromatographic run is the factor that grouped these methodologies together.

Table 3 presents the mean values for the two clusters formed during SOM classification of "green" methodologies, together with the set of methodologies rejected after cluster analysis. The fine ranking of methodologies does not show so much difference in "green" parameters, as it was observed during the preliminary classification. There is no difference in NEMI and eco-scale scores for clusters formed after SOM fine classification. It may be concluded that eco-scale and NEMI are good enough for distinguishing large differences between methodologies environmental impact. When these differences are more subtle the assessment methods do not perform very well. NEMI and eco-scale scoring do not include the amount of solvents, when it is below 10 mL. For example, DLLME methodologies require application of ~300 µL of organic solvents and they are scored by eco-scale in the same way as SPE methodology where the amount of each solvent may be also below 10 mL. On the other hand NEMI does not take into consideration volume of the solvent at all, if DLLE requires 100 µL of solvent categorized as "toxic", the protocol would lose one NEMI point. Both assessment methods are not good enough in assessing the differences in "greenness" of miniaturized extraction techniques.

3.3 Recommendation

The analysis of the dataset of analytical procedures with multivariate statistics tools can be a basis for making some initial assumptions when developing new analytical methodology. The first assumption should be that the initial volume of the

Table 3 Mean values of parameters for the two clusters formed during SOM analysis and methodologies rejected during data pre-treatment as "non-green"

	Mean values for group ranked as "green" during SOM analysis	Mean values for group ranked as "less-green" during SOM analysis	Mean values for group ranked as "non-green" during cluster analysis
Amount of organic solvent [mL]	2.5	10.1	113
Solid waste (g)	1.2	3.7	16.6
Volume of sample [mL]	20.8	85.9	943
Injection volume [µL]	34.2	12.4	1.9
Other analytes (per analyte concept)	17.7	39.4	46.4
NEMI score	3.4	3.7	2
Eco-scale score	86.8	85.9	59

analyzed should be small, definitely not more than 100 mL. Larger sample volume usually requires high material and energetic inputs. The second assumption is to make the sample injection volume to the instrument high. Large volume injection techniques allow for obtaining low limits of detection and avoiding the use of great amounts of organic solvents. The third observation is that multi-residue methodologies are less "green". Therefore, it is much better to optimize the analytical methodology for one target group of analytes. The results of the study also show that multi-residue methodologies are characterized by worse limits of detection and recoveries and require much more time to perform analysis.

The result of cluster analysis is that standard procedures based on LLE and SPE are ranked as "non-green". It seems obvious that standard procedures for determination of organic pollutants should not require any sophisticated or non-standard equipment. The results of this study suggest that a good choice for the standard method of sample preparation would be the DLLME technique, which meets the abovementioned requirements.

3.4 SOM as an assessment tool

The results of the multivariate analysis show that NEMI and eco-scale are good estimates for assessing analytical methodologies "greenness" but only to a certain degree. Is the SOM technique good for assessment of analytical methodologies environmental impact? Certainly, it has broad areas and according to our experience it cannot be applied for labeling but for comparison of many analytical protocols.

The main advantage of the SOM technique is the possibility of including many parameters, describing greenness as well as a lot of others, like energy consumption per analysis or data obtained from life-cycle assessment analysis. The classification analysis can also be broadened by metrological parameters of methodologies, like those included in the present study or others, like linearity range, reproducibility or ruggedness. Finally, cost per analysis can be also included in SOM analysis. The second advantage of the SOM technique is being perfect choice for comparison of many methodologies simultaneously, including many of their parameters. It also has better performance in comparison with methodologies with similar environmental impact as it requires fuzzy not discrete input data.

The main disadvantage is that the analysis with the SOM technique is tedious. Usually datasets with green analytical chemistry and metrological parameters of analytical methodologies are not available. Extraction of these data from analytical protocols is laborious. Finally, this technique is not adequate for simple labeling of methodologies.

4. Conclusions

The whole group of methods for determination of aldrin could be divided into 3 ranks with respect to the "greenness" of the procedures; thus, it is clear that the choice of methodology could be done not only by the use of chemical or metrological conditions (pretreatment, sample preparation, conditions of signal detection, *etc.*) but with respect to the "greenness" of the methodology.

The major conditions for "greenness" determination should involve different aspects of the analytical procedure; each one of the aspects could be numerically determined and it makes possible the objective reading of the "greenness". Multivariate statistical analysis proved that the major differences between the analytical methodologies lie in the "greenness" parameters, not in the metrological ones. The analysis indicates that most of the standard procedures applied for determination of aldrin in water samples do not meet the requirements of green analytical chemistry. Additionally, the study suggests that SPE cannot be considered a green technique, as during this study SPE was grouped as a nongreen technique.

The main advantage of application of the SOM technique as assessment methodology lies in the possibility of consideration of large amounts of factors during analysis. Another advantage is the possibility of simultaneous identification of discriminators between groups formed during analysis. The SOM technique could be also applied for assessment of chemical processes greenness or data from life-cycle assessment.

References

1 S. Armenta, S. Garrigues and M. de la Guardia, *Trends Anal. Chem.*, 2008, **27**, 497–511.
2 M. Farre, S. Perez, C. Goncalves, M. F. Alpendurada and D. Barcelo, *Trends Anal. Chem.*, 2010, **29**, 1347–1362.
3 M. Tobiszewski, A. Mechlińska, B. Zygmunt and J. Namieśnik, *Trends Anal. Chem.*, 2009, **28**, 943–951.
4 M. Tobiszewski and J. Namieśnik, *Trends Anal. Chem.*, 2012, **35**, 67–73.
5 S. Garrigues, S. Armenta and M. de la Guardia, *Trends Anal. Chem.*, 2010, **29**, 592–601.
6 C. J. Welch, N. Wu, M. Biba, R. Hartman, T. Brkovic, X. Gong, R. Helmy, W. Schafer, J. Cuff, Z. Pirzada and L. Zhou, *Trends Anal. Chem.*, 2010, **29**, 667–680.
7 L. H. Keith, L. U. Gron and J. L. Young, *Chem. Rev.*, 2007, **107**, 2695–2708.
8 A. Gałuszka, Z. M. Migaszewski, P. Konieczka and J. Namieśnik, *Trends Anal. Chem.*, 2012, **37**, 61–72.
9 A. Astel, *Talanta*, 2007, **72**, 1–12.
10 M. M. Mostafa, *Expert Syst. Appl.*, 2009, **36**, 11030–11038.
11 M. M. Mostafa, *Expert Syst. Appl.*, 2010, **37**, 2747–2755.
12 N. Chen, B. Ribeiro, A. Vieira and A. Chen, *Expert Syst. Appl.*, 2013, **40**, 385–393.
13 M. M. Mostafa, *Expert Syst. Appl.*, 2011, **38**, 6906–6915.
14 6630C Standard methods online – Standard methods for the examination of water and wastewater.
15 Methods for the determination of organic compounds in drinking water – Method 505 (EPA/600/R-95-131).
16 Methods for the determination of organic compounds in drinking water – Method 508 (EPA/600/R-95-131).

17 O-1104 Methods for the determination of organic substances in water and fluvial sediments.
18 625 Standard EPA method.
19 B. Hauser, M. Schellin and P. Popp, *Anal. Chem.*, 2004, **76**, 6029–6038.
20 H.-P. Lia, G.-C. Li and J.-F. Jen, *J. Chromatogr., A*, 2003, **1012**, 129–137.
21 A. Derouiche, M. R. Driss, J.-P. Morizur and M.-H. Taphanel, *J. Chromatogr., A*, 2007, **1138**, 231–243.
22 M. E. Torres Padrón, Z. Sosa Ferrera and J. J. Santana Rodríguez, *Anal. Bioanal. Chem.*, 2006, **386**, 332–340.
23 A. Garrido Frenich, R. Romero-González, J. L. Martínez Vidal, R. Martínez Ocaña and P. Baquero Feria, *Anal. Bioanal. Chem.*, 2011, **399**, 2043–2059.
24 C. Goncalves and M. F. Alpendurada, *J. Chromatogr., A*, 2002, **968**, 177–190.
25 E. Beceiro-Gonzalez, E. Concha-Grana, A. Guimaraes, C. Goncalves, M.-S. Lorenzo and M. F. Alpendurada, *J. Chromatogr., A*, 2007, **1141**, 165–173.
26 N. Ochiai, T. Ieda, K. Sasamoto, Y. Takazawa, S. Hashimoto, A. Fushimi and K. Tanabe, *J. Chromatogr., A*, 2011, **1218**, 6851–6860.
27 A. Penalver, V. Garcıa, E. Pocurull, F. Borrull and R. M. Marce, *J. Chromatogr., A*, 2003, **1007**, 1–9.
28 M. J. Gomez, S. Herrera, D. Sole, E. García-Calvo and A. R. Fernandez-Alba, *Anal. Chem.*, 2011, **83**, 2638–2647.
29 F. J. Camino-Sánchez, A. Zafra-Gómez, S. Cantarero-Malagón and J. L. Vílchez, *Talanta*, 2012, **89**, 322–334.
30 P. Tölgyessy, B. Vrana and Z. Kras, *Talanta*, 2011, **87**, 152–160.
31 M.-I. Leong and S.-D. Huang, *J. Chromatogr., A*, 2009, **1216**, 7645–7650.
32 W.-C. Tsai and S.-D. Huang, *J. Chromatogr., A*, 2009, **1216**, 5171–5175.
33 C. Cortada, L. Vidal, R. Pastor, N. Santiago and A. Canals, *Anal. Chim. Acta*, 2009, **649**, 218–221.
34 C. Cortada, L. Vidal, S. Tejada, A. Romo and A. Canals, *Anal. Chim. Acta*, 2009, **638**, 29–35.
35 F. Rezaei and M.-R. Milani Hosseini, *Chromatographia*, 2012, **75**, 379–385.
36 S.-P. Huang and S.-D. Huang, *J. Chromatogr., A*, 2007, **1176**, 19–25.
37 H. Farahani, Y. Yamini, S. Shariati, M. R. Khalili-Zanjani and S. Mansour-Baghahi, *Anal. Chim. Acta*, 2008, **626**, 166–173.
38 A. Garrido Frenich, R. Romero-González, J. L. Martínez Vidal, R. Martínez Ocaña and P. Baquero Feria, *Anal. Bioanal. Chem.*, 2011, **399**, 2043–2059.
39 L. Zhao and H. Kee Lee, *J. Chromatogr., A*, 2001, **919**, 381–388.
40 S.-P. Huang, P.-S. Chen and S.-D. Huang, *J. Chromatogr., A*, 2009, **1216**, 4347–4353.
41 E. Concha-Grana, G. Fernández-Martínez, V. Fernández-Villarrenaga, M. I. Turnes-Carou, S. Muniategui-Lorenzo, P. López-Mahía and D. Prada-Rodríguez, *Talanta*, 2009, **78**, 764–771.
42 S. Nakamura, T. Łamagami and S. Daishima, *Analyst*, 2001, **126**, 1658–1662.
43 G. Famiglini, P. Palma, E. Pierini, H. Trufelli and A. Cappiello, *Anal. Chem.*, 2008, **80**, 3445–3449.
44 G. Famiglini, P. Palma, V. Termopoli, H. Trufelli and A. Cappiello, *Anal. Chem.*, 2009, **81**, 7373–7378.
45 V. Matamoros, E. Jover and J. M. Bayona, *Anal. Chem.*, 2010, **82**, 699–706.
46 C. C. Leandro, D. A. Bishop, R. J. Fussell, F. D. Smith and B. J. Keely, *J. Agric. Food Chem.*, 2006, **54**, 645–649.
47 S. Barreka, C. Cren-Olivéa, L. Wiest, R. Baudot, C. Arnaudguilhema and M. F. Grenier-Loustalot, *Talanta*, 2009, **79**, 712–722.
48 Method 525.2 Methods for the determination of organic compounds in drinking water.
49 Method 508.1 Methods for the determination of organic compounds in drinking water.
50 C. Mansilha, A. Melo, H. Rebelo, I. M. P. L. V. O. Ferreira, O. Pinho, V. Domingues, C. Pinho and P. Gameiro, *J. Chromatogr., A*, 2010, **1217**, 6681–6691.
51 I. Vassilakis, D. Tsipi and M. Scoullos, *J. Chromatogr., A*, 1998, **823**, 49–58.
52 A. M. Carro, J. C. Cobas, J. B. Rodrõguez, R. A. Lorenzo and R. Cela, *J. Anal. At. Spectrom.*, 1999, **14**, 1867–1873.
53 L. Brossa, R. M. Marce, F. Borrull and E. Pocurull, *J. Chromatogr., A*, 2002, **963**, 287–294.
54 A. Colume, S. Cardenas, M. Gallego and M. Valcarcel, *Talanta*, 2001, **54**, 943–951.
55 L. Brossa, R. M. Marce, F. Borrull and E. Pocurull, *J. Chromatogr., A*, 2003, **998**, 41–50.
56 Y. Li, C. L. McCarty and E. J. George, *Front. Environ. Sci. Eng. China*, 2011, **5**, 417–425.
57 P. Trebuňa and J. Halčinová, *Procedia Eng.*, 2012, **48**, 673–678.
58 T. Kohonen, *Self-organizing Maps*, Springer, Berlin, 3rd edn, 2001.
59 M. Koel and M. Kaljurand, *Pure Appl. Chem.*, 2006, **78**, 1993–2002.
60 A. Astel, S. Tsakovski, P. Barbieri and V. Simeonov, *Water Res.*, 2007, **41**, 4566–4578.