VERSITA

Cent. Eur. J. Chem. • 11(4) • 2013 • 610-618
DOI: 10.2478/s11532-012-0196-x

Central European Journal of **Chemistry**

# Selection of mineralised methods to analyse different types of matrices. Applying the Box–Cox transformation to chemometrics study the coexistence of heavy metals in natural samples

**Research Article**

Marcin Sajdak[1*], Celina Pieszko[2]

*[1]Institute for Chemical Processing of Coal,
41-803 Zabrze, Poland*

*[2]Department of Analytical Chemistry,
Faculty of Chemistry,
Silesian University of Technology,
44-100 Gliwice, Poland*

**Received 30 August 2012; Accepted 6 December 2012**

**Abstract:** Chemometric methods are mostly used to optimise technological processes and analytical procedures. Applying chemometric methods in environmental tests may reveal relationships among chemical elements in biomes. Cluster analysis and principal component analysis (PCA) are very helpful for detecting relationships among studied parameters. However, large amounts of data may have a negative effect on this analysis and can lead to misinterpretation of the results. This situation was observed when the samples, taken from several places in the Silesian Province, were used to test the relationship between heavy metals contained in various environmental matrices. Samples were collected from a small area and were characterised by a single biome (pine forest) because direct interpretation of PCA and CA was insufficient to correctly describe such data. The solution to this problem was the use of the Box-Cox transformation, which is a rapid method to normalise input data.

The application of chemometric tools enabled the relationships between sampling sites (industrialised and non-industrialised) to be examined and was very helpful in illustrating the relationship between the methodologies of plant preparation samples. Furthermore, the results may indicate the need for further data analysis. The tools described in this paper can be useful for choosing the optimal mineralisation method according to the type of test matrix.

**Keywords:** *Environmental survey • Chemometric analysis • Box-Cox • Heavy metals • Concentration assessment*

## 1. Introduction

Chemometric methods are increasingly used with the experimental data generated by environmental and chemical analyses of natural samples [1] and chemical processes [2]. Unfortunately, without adequate preparation, experimental data are not useful for chemometric analysis. Application of exponential transformations, such as the Box and Cox transformation [3], significantly improves the quality of the data to allow the application of various chemometric techniques.

The results from the studies presented in this paper were analysed with exponential transformation as a tool to increase the usefulness of the obtained data, such as the concentrations of heavy metals in environmental samples. In this case, the results of our own studies conducted in the area of Upper Silesia were used.

Application of chemometric method analysis together with methods for choosing and elimination variations allow for trustworthy comparisonsof sample preparation. Three methods of plant mineralizations to atomic absorption spectroscopy (AAS) analysis for measurement of heavy metals concentration has been shown in this paper.

---

* *E-mail: marcinsajdak.tch@gmail.com*

The correlation among concentrations of Zn, Cd, Ni, Mn, Cu, Pb and Fe in the different components of pine trees such as pine needles, twigs, pine cones, and litter were studied. Information about the concentrations of these elements was subjected to several chemometric and statistical procedures, such as an analysis of similarities, k-means methods, correlation analysis or principal component analysis. Application of an exponential transformation, such as the Box and Cox transformation, has enhanced the quality of data used in different chemometric techniques.

All statistical and chemometric analyzes were performed in STATISTICA 9 [4] and MS Excel.

This paper is another part of the research conducted in the Province of Silesia. Part of the research results that describe the methods, analysis and sampling sites, as well as describing preliminary results of chemical and chemometric analysis were presented in [5]. Chemometric methods presented in this paper are increasingly being used for chemical analysis of wood biomass for energy used, which allows for a broader view of this type of matrix environment [5-7].

## 2. Experimental procedure

The materials for investigation were 36 samples of cones [C], needles [N], branches [B], and litter [L] of Scots pine which were collected from nine sites in the province of Silesia (Szczekociny, Żarnowiec, Sławniów, Podlesice, Solca, Sławków, Dabrowa Górnicza, Jaworzno, and Sosnowiec).

Three methods of sample preparation were used and are presented in this paper.

**Method 1 (DM):** Collected materials were air-dried, ground in a mill equipped with a titanium blade and then dried at 110°C. After drying and digestion at 350°C, the remaining ash was transferred into a beaker and dissolved in 10% $HNO_3$ (POCH S.A. Poland). After reducing the volume (the solutions were concentrated), the solutions were transferred to 10 mL volumetric flasks.

**Method 2 (WM):** In the second method, ultrasound-assisted (Soltec Sonica 5200MH) wet digestion was used to prepare the samples for analysis. Digestion was performed in a 10% solution of $HNO_3$ at 50°C for 30 minutes. After filtering and reducing the volume (concentrating the solution), the sample solutions were transferred to 10 mL volumetric flasks.

**Method 3 (SE):** Solid-liquid extraction (SLE)was used for the third sample preparation. Extractions were performed in a boiling 10% solution of $HNO_3$ in a Soxhlet extractor. A filter crucible was used, and therefore the solution did not need to be filtered after digestion, and the solution could be analysed immediately after reducing the volume (concentrating the solution).

In all of the collected plant material samples, the concentration of several elements (Zn, Cd, Ni, Mn, Cu, Pb, and Fe) were determined using flame atomic absorption spectrometry. The measurements were performed using an AAS 3 Carl Zeiss Jena spectrophotometer. The values of recoveries for the various materials were estimated based on the analysis of a reference material spiked with standard solutions.

The values for recovery of materials were assessed in comparison with the reference and the model material. The conditions of the AAS analysis are presented in Table 1.

The limits of detection determined by AAS for each of the elements were calculated using the following formula:

$$\mathrm{LOD} = \frac{(3.3 \cdot \mathrm{Sb})}{a} \qquad (1)$$

Where:
Sb- standard deviation of residuals of the calibration curve, a- gradient of the line, value of 3.3 standard deviations result from the use of a dedicated formula to determine the LOD for the application of the calibration curve [8,12].

### 2.1. Box–Cox transformation

Transformations of data were performed to achieve specific purposes, e.g., stability of variance, additivity of effects, and symmetry of the density. Among the many parametric transformations, the transformation presented in Eq. 2 is the most commonly used [3]. Let X describe a random variable on the positive half-linethen, the Box–Cox transformation of X with power parameter λ is defined by:

$$X^{\lambda} = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda} \; for\ \lambda \neq 0 \\ \log X \; for\ \lambda = 0 \end{cases} \qquad (2)$$

The formula $\frac{x^{\lambda}-1}{\lambda}$ is chosen so that $X^{\lambda}$ is continuous when λ tends to zero and monotonically increases with respect to *X* for any λ.

The power parameter λ is estimated by a graphical technique or by the method adopted by Box and Cox [3]. A closed form for the estimator λ can rarely be found. The plot of the maximum likelihood against λ is helpful. The value of λ obtained in this manner is treated as if it were a true value, and the model can be fitted to the transformed data [8].

**Table 1.** Conditions of AAS analysis.

| **Element** | **Temperature characterization** | | **Lamp characterization** | | | **Method characterization** | |
|---|---|---|---|---|---|---|---|
| | **Pyrolysis temperature [°C]** | **Atomization temperature [°C]** | **Lamp current [mA]** | **Analytical line [nm]** | **Spectral gap [nm]** | **Analytical range [$\mu g\ kg^{-1}$]** | **LOD [$\mu g\ kg^{-1}$]** |
| **Zn** | 400 | 1800 | 5.0 | 213.9 | 0.5 | 10-1500 | 5.0 |
| **Cd** | 300 | 1800 | 3.0 | 228.8 | 0.5 | 10-1800 | 4.0 |
| **Ni** | 900 | 2400 | 4.0 | 241.5 | 0.2 | 600-25000 | 9.0 |
| **Mn** | 700 | 2400 | 5.0 | 403.1 | 0.2 | 2700-27000 | 2.0 |
| **Cu** | 800 | 2300 | 3.0 | 222.6 | 1.0 | 700-180000 | 1.0 |
| **Pb** | 400 | 2000 | 5.0 | 217.0 | 1.0 | 200-20000 | 10.0 |
| **Fe** | 800 | 2300 | 7.0 | 372.0 | 0.2 | 900-80000 | 5.0 |

In the literature, Box–Cox transformations are applied to basic distributions, *e.g*. the cubic root transformation of chi-squared variables is used for acceleration to normality (cf. also normal distribution), and the square-root transformation stabilises variances of the Poisson distributions. Box–Cox transformations are also applied to link functions in generalised linear models. The transformations aim mainly to obtain linearity of the effects of covariates [8].

# 3. Results and discussion

## 3.1. Results of Transformation

Experimental data collected during the experiment were used as input quantities for the analysis of principal components. The results of this analysis are shown in Fig. 1.

The first two principal components (PC1 and PC2) define approximately 64% of the total variability of the data. Unfortunately, this result does not permit a correct interpretation because the arrangement of data is characterised by considerable heterogeneity around the point (0, 0). Using the standard normal variance (SNV) transformation (Fig. 1b) on that data set does not improve the results. Therefore, the Box-Cox transformation was used with a change of coefficients (lambda) to optimise the outcome of that transformation. The diagram (Fig. 1c) shows the positive results of the transformation, which significantly improves the form of input data. Applying this transformation improved the statistical distribution of the experimental data, changing the experimental data to a normal distribution. This result is shown in Fig. 2.

The graph shown in Fig. 2 is an exemplary diagram for the copper content in the test samples. Transformation was carried out for all of the variables tested with the same results as for copper.

Such data prepared with this method were subjected to principal component analysis (PCA). The (7×108) matrix includes the average concentrations of Zn, Cd, Ni, Mn, Cu, Pb, and Fe from four different sample types (C, N, B, and L). These samples were prepared for analysis by three different methods of sample preparation (methods: 1-DM, 2-WM, and 3-SE). PCA was applied to the finding of the data structure and the visualisation of correlations among properties. The Box-Cox transformation method was very helpful in obtaining information about the uniformity of the data, the existence of groups and relationships among sampling sites, the types of sample matrices, and the concentrations of the metals.

Fig. 3 shows the results of the PCA analysis of the data after Box-Cox transformation. Fig. 3a shows the cumulative percentage variance of the data described by the first seven principal components.

These PCs explain 71.5% of the total data variance. This compression allows sufficient information about the data set to be described; consequently, it is possible to obtain satisfactory results. PC 1 (Fig. 3b) enables a description of the concentrations of Zn, Mn, Fe, and Cu in the samples. Values for cadmium and lead, relative to PC 1, were verified by applying a third main principal component (Fig. 3c). The concentrations of Cd, Ni, and Pb (Fig. 3c) can be described by PC 3, and it could be useful for describing the anthropogenic origin of these metals. The nickel content, in addition to other metals, can be easily described by the first principal component (PC 1) shown in Fig. 3d. Due to the geographical locations of the sampling sites (Upper Silesian Industrial District) and land use planning above the mentioned area, these sites have been originally divided into two clusters: cities from the less industrialised areas such as Szczękociny (30), Żarnowiec (31), Sławniów (32), Podlesice (36), and Solca (37) and the cities lying in the

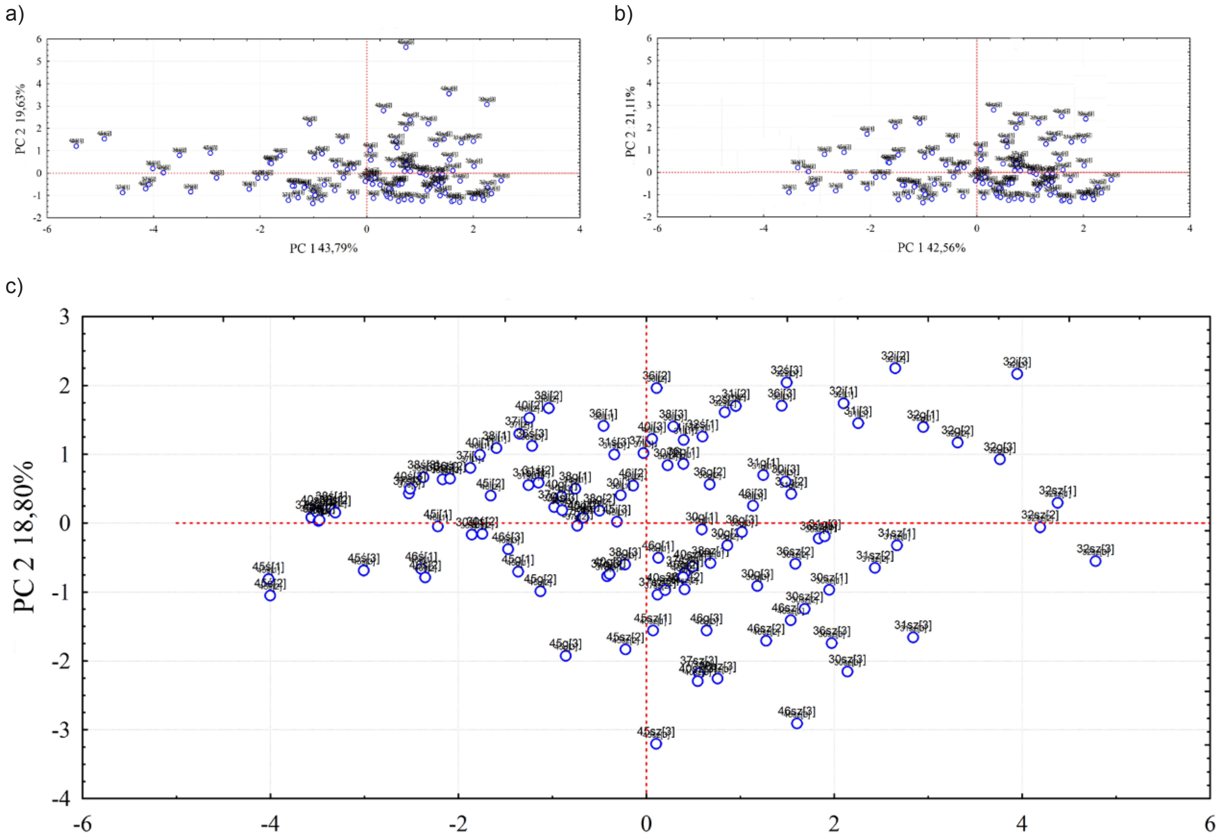

**Figure 1.** Projection of the analysed data on the first plane defined by two main principal components: a) the raw data, b) data after SNV transformation, c) projection of analysed data on the plane defined by the first two principal components and subject to the Box-Cox transformation.

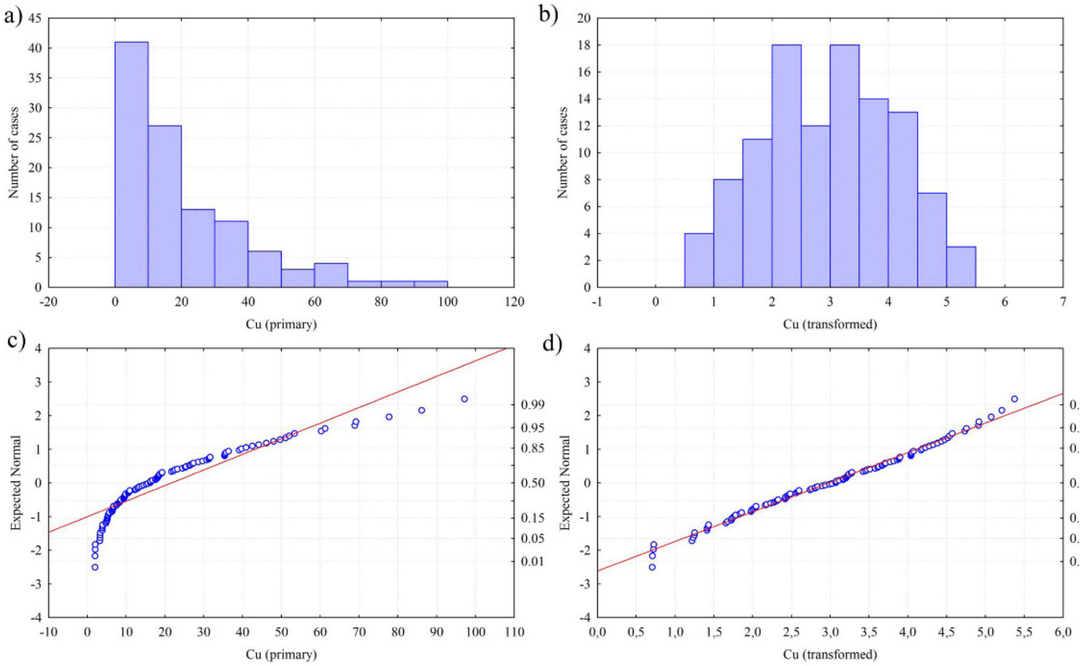

**Figure 2.** Histogram and normal plot of raw data and data subjected to the Box-Cox transformation for Cu.

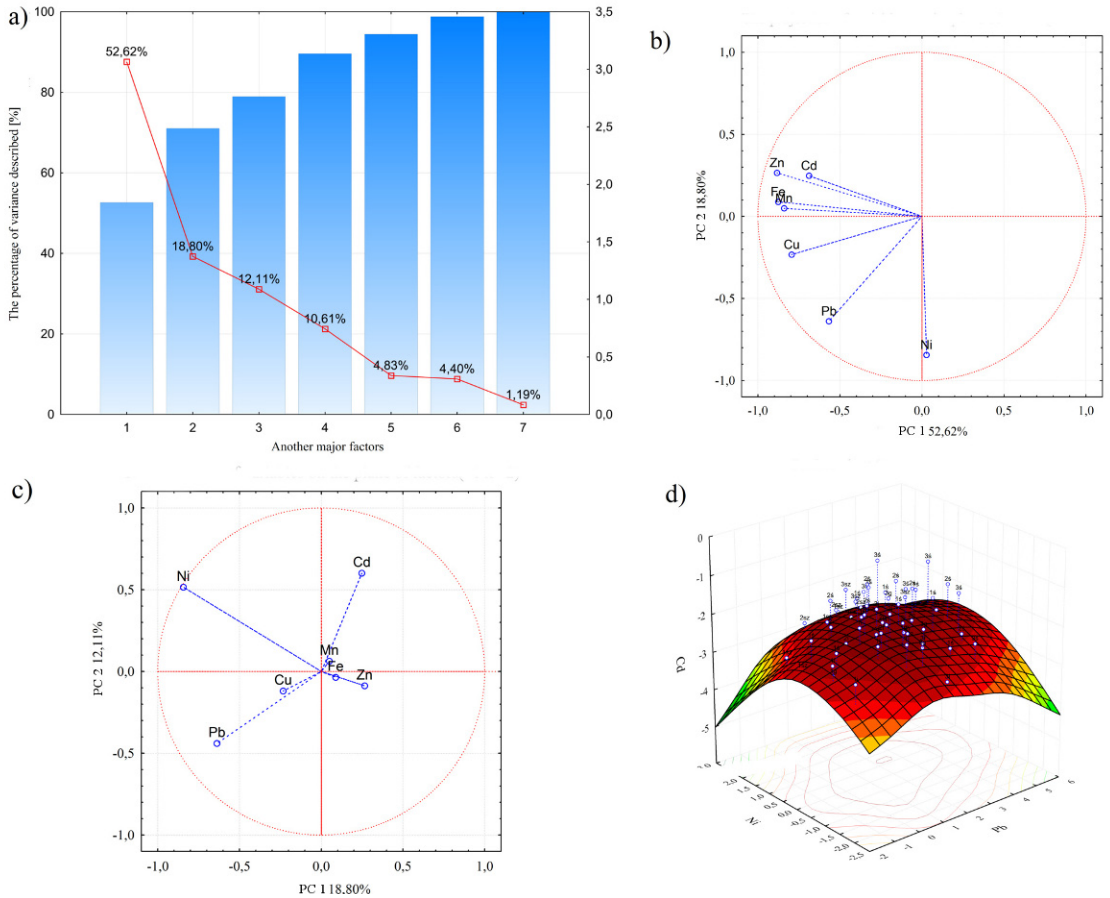

**Figure 3.** a) The cumulative percentage of variance of the data reported in the first seven principal components, b) projection of variables on the plane of the first two principal components, c) the importance of variables in the plane projection of the PC 1 and PC 3, d) the surface chart of Cd, Ni, and Pb in the sampling sites.

more industrialised ones, which include Sławków (38), Dabrowa Górnicza (40), Jaworzno (45), and Sosnowiec (46). The method of k-means clustering was used to verify this assumption. For this purpose, the above-mentioned method was used to create k-clusters with maximum different values from one another. In Fig. 4a, the x-axis displays the test metals, and the y-axis shows their concentrations in different sampling locations.

The urban areas were divided into three groups:

1. Sosnowiec, Żarnowiec, Sławniów, Podlesice
2. Szczekociny, Dąbrowa Górnicza, and Solca
3. Jaworzno and Sławków.

The result of the k-means is shown in Fig. 4a. The data set was divided into three defining groups by the k-means method as a function of the metal concentrations (degree of industrialisation) in the area from which the samples were collected. Sosnowiec was assigned to the first agglomeration, which contains the towns located in non-urbanised and unindustrialised areas, because it showed low average concentrations of zinc, manganese and cadmium. These concentrations were quite similar to the ones found in places considered poorly industrialised. The same situation also applies to the qualification of Dabrowa Górnicza for the second agglomeration.

Fig. 4a shows that certain elements are present in similar quantities and that it is possible to assign them to groups. The k-means method was used to ascertain whether this assignment was possible. The results of this analysis are shown in Fig. 4b.

As expected (Fig. 4b), proportions of the analysed metals can easily be observed: {Cu, Zn}> {Mn, Fe}> {Pb, Ni, Cd}. This accessible presentation allows for visualisation of the result. It can be concluded that the arrangement of these metals is not accidental because of their similar ionic radii, for example, {Mn = $83\times10^3$ nm; Fe = $78\times10^3$ nm} {Cu = $73\times10^3$ nm; Ni = $69\times10^3$ nm}. Therefore, their characteristics and their coexistence suggest that other groups of transition metals can be bioaccumulated on a similar level. Zn, Cd, and Pb were placed in one group mainly due to their low content in the investigated area.

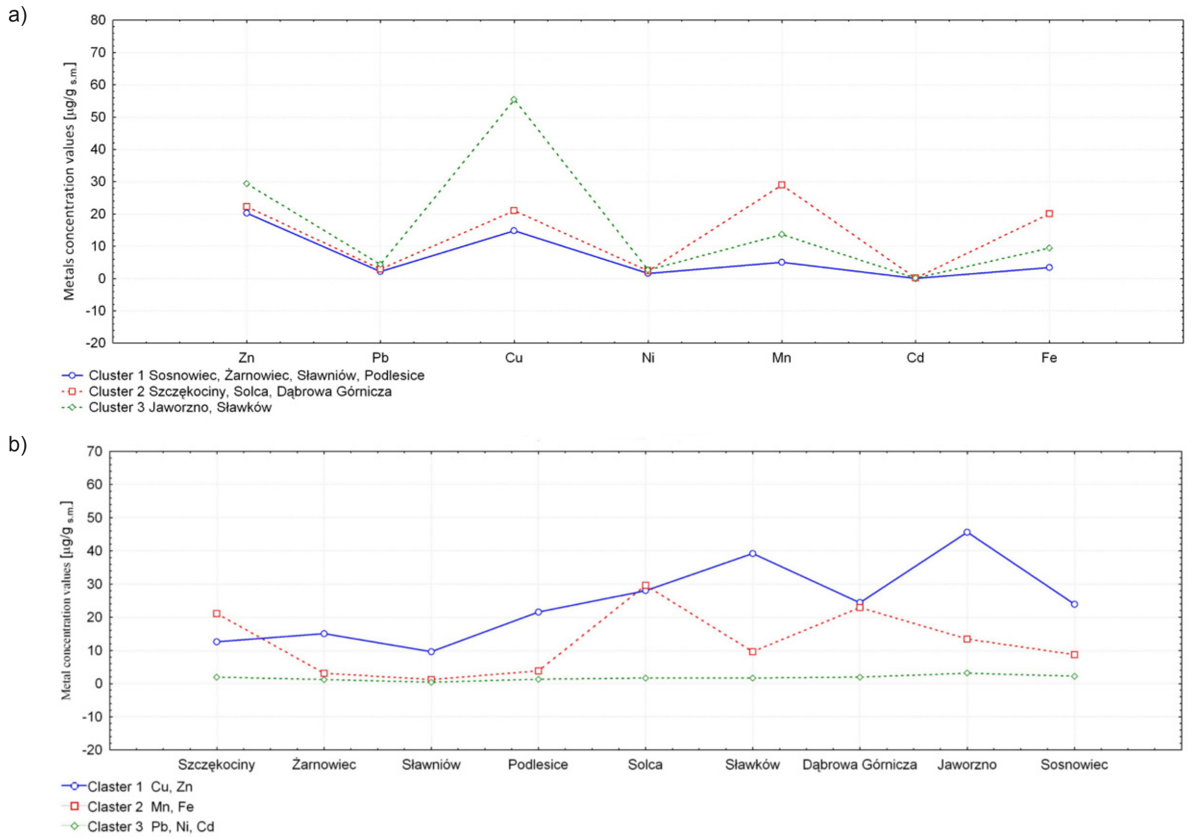

**Figure 4.** a) Diagram of the k-means analysis for metal content in relation to three agglomerations describing the sampling sites, b) diagram of the analysis of k-means for sampling sites in relation to three agglomerations describing the metal concentrations.

Fig. 5 shows the distribution of sampling sites from which samples were collected for analysis. Industrial areas were marked as (1), while non-industrialised places were marked as (2) or (3).

The PCA helped to organise the sampling sites and assign them to specific areas as shown in Fig. 5a. Collected material for analysis was divided into two groups: the red area represents industrialised places, and the blue area represents unindustrialised ones. When new tags (colours) were added to Fig. 5a, the distribution of samples based on their type could be observed (litter of conifer needles and branches). Each of those samples can be attributed to certain values of PC1 and PC2 that determine their type.

The most common method for pre-treating environmental samples, particularly plant samples, is dry mineralisation which is applicable to a wide range of plant matrices [9]. Dry mineralisation served as the most suitable method for the preparation of plant samples containing heavy metals for environmental analysis. To verify the above-mentioned hypothesis, samples were prepared using three methods (including dry mineralisation) and then compared to each other. To verify differences among the mineralisation methods, the cluster analysis method, also called analysis of agglomeration, was used. Cluster analysis allowed for a simple visualisation of the existing relationship among the analysed samples.

Fig. 6a shows a horizontal hierarchical tree plot of the contents of the study samples as a function of the sample preparation method. In this case the Ward's method was superior to the similarity measure method, which is usually used to the squared Euclidean distance.

Each element that was examined was marked with a number corresponding to the type of mineralisation: 1 - ashing of the sample (dry mineralisation), 2 - wet mineralisation, or 3 - wet mineralisation in the Soxhlet apparatus. Fig. 6b shows the graph of the methods' importance for sample preparation relative to the expected values. The higher the F value is, the more important the tested variable. The Sendecora test analysis was performed.

As shown in Fig. 6b, dry mineralisation is not a good method for preparing plant samples when we have to determine, for example, the Pb content in the matrix of Scots pine. According to the results of the chemometric analysis, the best method for preparing the matrix

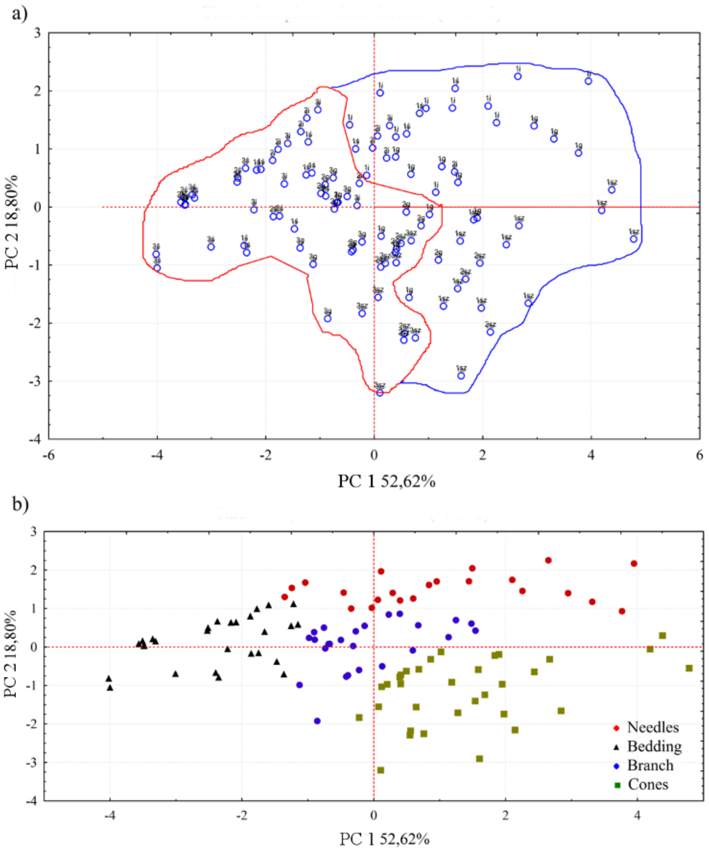

**Figure 5.** a) The projection of samples on the space defined by the first two principal components, b) the projection of samples on the space defined by the first two principal components and selected groups.

**Table 2.** Summary of the best methods for preparing plant samples for analysis and the recovery procedures for the preparation of samples for analysis.

| **Metals** | **The sample preparation[1]** | | | | | | | | | | | | **Summary of the best methods of preparing plant samples considered for analysis** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **C-DM** | **C-WM** | **C-SE** | **N-DM** | **N-WM** | **N-SE** | **L- DM** | **L- WM** | **L- SE** | **B- DM** | **B- WM** | **B- SE** | **C** | **N** | **L** | **B** |
| | **Recovery [%]** | | | | | | | | | | | | | | | |
| **Cu** | 92 | 110 | 57 | 110 | 81 | 42 | 110 | 88 | 50 | 88 | 76 | 101 | SE<br>DM | DM | WM | SE<br>DM |
| **Zn** | 90 | 105 | 95 | 105 | 103 | 83 | 105 | 98 | 85 | 105 | 85 | 87 | SE | DM | SE | SE |
| **Pb** | 42 | 44 | 110 | 91 | 82 | 110 | 64 | 105 | 92 | 54 | 64 | 114 | SE | WM<br>SE | WM<br>SE | WM<br>SE |
| **Mn** | 85 | 105 | 79 | 97 | 99 | 70 | 100 | 97 | 67 | 98 | 89 | 94 | SE | DM<br>SE | WM | WM |
| **Ni** | 39 | 75 | 105 | 105 | 54 | 40 | 110 | 57 | 27 | 76 | 96 | 68 | SE | WM<br>SE | SE | WM |
| **Cd** | 95 | 120 | 28 | 105 | 79 | 16 | 110 | 91 | 22 | 61 | 101 | 11 | DM<br>WM<br>SE | DM<br>WM<br>SE | WM<br>SE | WM<br>SE |
| **Fe** | 96 | 61 | 105 | 105 | 34 | 15 | 118 | 29 | 24 | 120 | 50 | 43 | SE<br>WM | SE | SE | DM |

*1 C- pine cones, N- pine needles, L-pine litter, B- pine branches, DM- dry mineralisation, WM- wet mineralisation, SE- soxhlet extraction,*

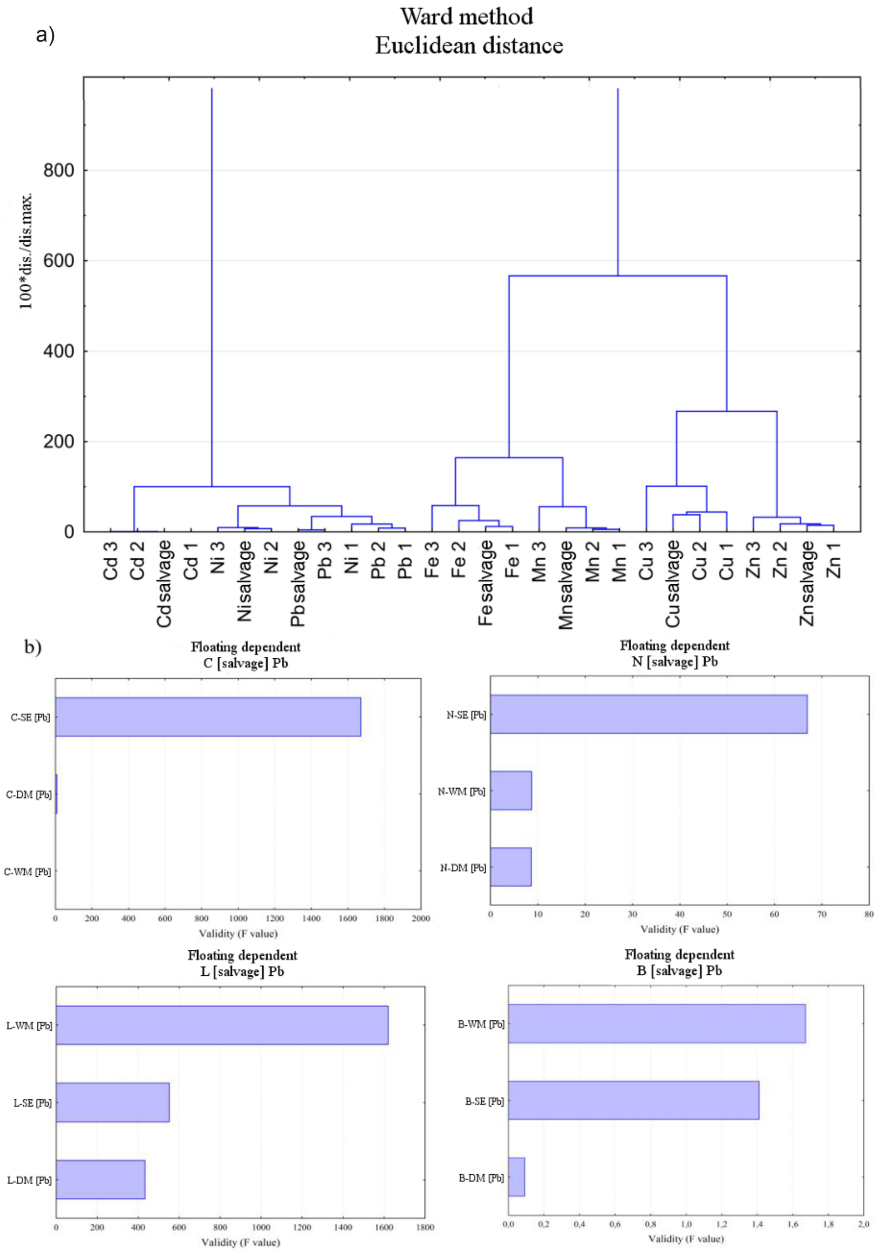

**Figure 6.** a) Horizontal Hierarchical Tree Plot of the cluster analysis of the examined analyses and methods of preparing samples, b) A variables' importance plot for sample preparation in relation to the expected value.

samples to analyse Pb was found to be an ultrasound-assisted digestion. These samples are characterised by the highest correlation coefficients: 0.999 (C SE), 0.852 (N WM), 0.999 (L WM), and 0.776 (B WM). Table 2 shows the comparison of the mineralisation methods. This table contains both the type of mineralisation and the type of the tested matrix samples; it also shows the percentage value of the enrichment methods for the studied metals.

The criterion for choosing the best method was to determine the value of the F-test for the best predictor variable, which in this case was kind of prepared sample. The higher the value of the F-statistic, the better suited is the method for the preparation of the sample in the direction of the analyte. Sample selection chart variables (predictors) are shown in Fig. 6b.

Interpreting this table before the analysis may affect the method of sample preparation. The effective methods of recovery of metals may be chosen by focusing on better linear relationships between measured and expected values. These results indicate that the wet mineralisation and liquid-solid extraction are more efficient methods for preparing plant samples, such as the litter of conifer needles, duff, branches, or cones, in comparison to dry mineralisation. Dry mineralisation is the easiest and the least complicated method, but may not always be the best one (as was shown in Table 2, as it characterises low enrichment levels).

Performance analysis determined during tests of approximately 105% to 120% due to used reagents and the characteristics of the methods of enrichment sample. These results are acceptable because they repeat in carrying out analyzes and may be mathematically compensated in relation to the expected value.

# 4. Conclusion

The use of power series transformations, such as the Box-Cox transformation of experimental data, has enabled us to perform an accurate chemometric analysis and has allowed for correct interpretation of the results. Principal component analysis allowed for the separation of metals and led to the discovery of whether the tested samples were collected from an industrialised area. Thus, this analysis indicates the existence of relationships between sampling sites. Cluster analysis helped to illustrate the relationship between the methodologies for preparing plant samples and indicated the need for further data analysis.

These methods make it possible to reduce the number of analyzes by finding the function of regression between co-occurring elements in these matrices. So that the analysis of one element will evaluate the content of the following in this material. The use of these techniques not only allows to discover sources of pollution, but also helps to reduce the amount of their entry into the environment.

Using the F test allowed for the selection and elimination of variables; it showed which of the studied variables was more important.

All the chemometric techniques used in this paper are very functional for extracting useful information from environmental data.

## References

[1] M. Żelazny, A. Astel, A. Wolanina, S. Małek, Environmental Pollution 159, 1048 (2011)

[2] T. Togkalidou, H. Tung, Y. Sun, A. Andrews, R.D. Braatz, Organic Process Research & Development 6, 317 (2002)

[3] G.E.P. Box, D.R. Cox, J. Roy. Statist. Soc. B 26, 211 (1964)

[4] STATISTICA - data analysis software system, version 9 (StatSoft, Inc., USA, 2009)

[5] M. Sajdak, C. Pieszko, Eur. J. Chem. 10(5), 1696 (2012)

[6] M. Sajdak, Eur. J. Chem. 11(2), 151 (2013)

[7] M. Sajdak, O. Piotrowski, Eur. J. Chem. 11(2), 259 (2013)

[8] M. Tobiszewski, J. Namieśnik, Anal. Bioanal. Chem. 399, 3565 (2011)

[9] P. McCullagh, J.A Nelder, Generalized Linear Models, 2nd edition (Chapman & Hall, London, 1989)

[10] G. Romanik, E. Gilgenast, A. Przyjazny, M. Kamiński, Dear Sirce 70, 253 (2007)

[11] P. Konieczka, J. Namieśnik, Chem. Anal. (Warsaw) 53, 785 (2008)

[12] C.H. Lochmuller (Ed.), Chemical Laboratory - a Practical Approach (CRC/Taylor & Francis, Boca Raton, USA, 2009) 131-216